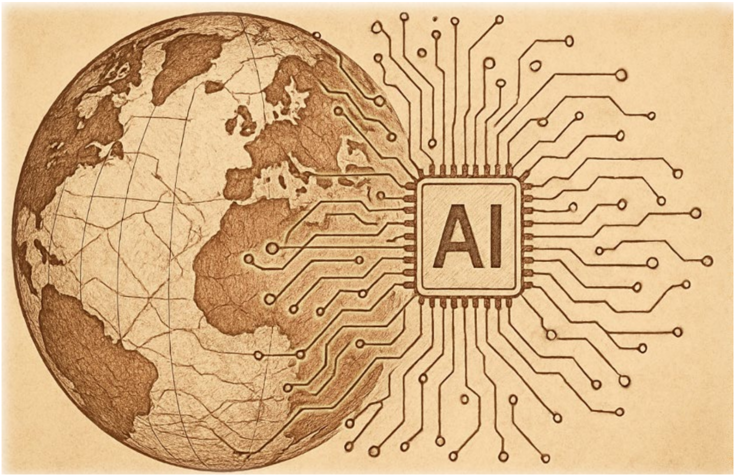
Geographia Technica



Technical Geography
an International Journal for the Progress of Scientific Geography

Special Issue
Artificial Intelligence Applications in Geography
AIAG

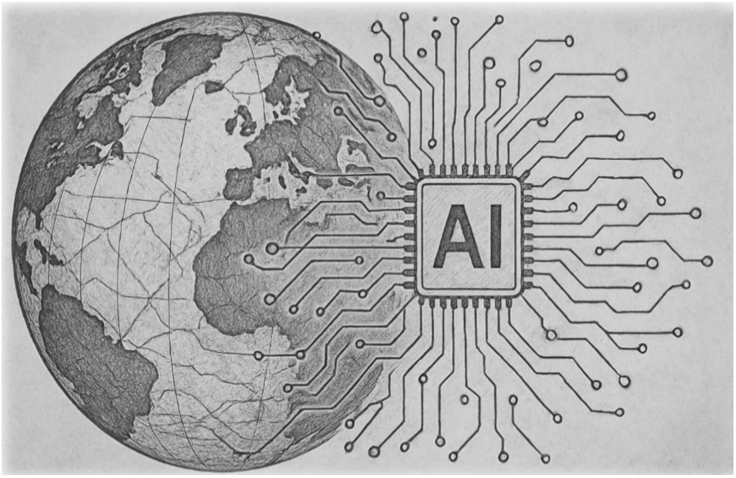
EDITORS
Ionel HAIDU and Zsolt MAGYARI-SÁSKA

Volume 21, Geographia Technica No. 2/2026

www.technicalgeography.org

Cluj University Press

Geographia Technica



Technical Geography
an International Journal for the Progress of Scientific Geography

Special Issue
Artificial Intelligence Applications in Geography
AIAG

EDITORS
Ionel HAIDU and Zsolt MAGYARI-SÁSKA

Volume 21, Geographia Technica No. 2/2026

www.technicalgeography.org

Cluj University Press

ISSN: 1842 - 5135 (Printed version)
ISSN: 2065 - 4421 (Online version)

Universitatea Babeş-Bolyai
Presa Universitară Clujeană
Director: Codruța Săcelean
Str. B.P. Hasdeu nr. 51
400371 Cluj-Napoca, România
Tel./fax: (+40)-264-597.401
E-mail: editura@ubbcluj.ro
http://www.editura.ubbcluj.ro/

Asociatia Geographia Technica
2, Prunilor Street
400334 Cluj-Napoca, România
Tel. +40 744 238093
editorial-secretary@technicalgeography.org
http://technicalgeography.org/

C o n t e n t

Geographia Technica

Volume 21, Special Issue, Artificial Intelligence Applications in Geography (AIAG)
An International Journal of Technical Geography
ISSN 2065-4421 (Online); ISSN 1842-5135 (printed)

FROM *POSTFACTUM* TO *ANTEFACTUM* GEOGRAPHY: WHAT ARTIFICIAL INTELLIGENCE WOULD HAVE CHANGED IN SEVEN EARLIER STUDIES

A Retrospective Human–AI Experiment in Technical Geography

Ionel HAIDU[1,2]

A Letter from the Editor

DOI: 10.21163/GT_2026.212.12

ABSTRACT
Discussions of artificial intelligence in Geography frequently take the form of inventories of possible applications. This Letter adopts a different strategy. Instead of asking what AI might theoretically do, it asks what contemporary AI would have changed in seven geographical studies previously authored or co-authored by the editor and completed without access to present-day AI technologies. The studies concern forest-cover change, storm-induced windthrows, floodplain delineation, urban imperviousness and runoff, trend-signal detection, road-accident distribution, and atmospheric pollution during the COVID-19 lockdown. Each paper is examined through the same retrospective protocol: the original geographical question and technical workflow are reconstructed; possible AI entry points are identified; the potential gains in speed, accuracy, uncertainty assessment, knowledge discovery, prediction and decision support are evaluated; and the data and validation requirements of the proposed redesign are discussed. The exercise does not claim that AI would necessarily have produced superior empirical results, because the original datasets were not reprocessed. It demonstrates, however, that AI could have enlarged the range of questions that could be formulated and tested. The seven cases show a common transition from fixed workflows to adaptive learning, from deterministic maps to probability fields, from separate datasets to multimodal integration, and from *postfactum* explanation to *antefactum* prediction, counterfactual analysis and intervention design. AI is therefore proposed not as a substitute for geographical reasoning, but as the learning layer of Technical Geography.

Keywords: *GeoAI; spatial reasoning; hybrid modelling; uncertainty; counterfactual analysis; decision support; human–machine collaboration*

1. TECHNICAL GEOGRAPHY BEFORE AND WITH ARTIFICIAL INTELLIGENCE

In 2016, I attempted to define Technical Geography as an emerging field grounded in the spatial and temporal attributes of geographical processes, particularly autocorrelation and frequency, and supported by spatial techniques, statistical methods and modern technologies. I argued that digital geographical information, remote sensing, GIS, modelling and advanced data processing were not peripheral tools but necessary components of contemporary geographical research. I also anticipated that the technical component would progressively become present in all fields of Geography (Haidu, 2016).

In 2024, I developed this argument further by suggesting that Geography should move from a predominantly *postfactum* science, concerned with phenomena that have already occurred, towards an *antefactum* science capable of anticipation, prediction and proactive intervention. Technical methods, spatial modelling, GIS, remote sensing, LiDAR, UAV observations and statistical analysis were presented as the foundations of this transition (Haidu, 2024).

[1] *LOTERR (Laboratoire d'observations des territoires), Université de Lorraine, F-57000 Metz, France. ionel.haidu@univ-lorraine.fr*
[2] *STAR-UBB (Scientific and Technological Advanced Research Institute), Babeş-Bolyai University, 400084 Cluj-Napoca, Romania.*

Artificial intelligence now requires a further extension of both propositions. AI should not be regarded merely as one additional technology placed alongside GIS, remote sensing or spatial statistics. Its distinctive contribution is the capacity to learn relationships from examples, to update those relationships when new observations become available, to integrate heterogeneous sources of information, to explore large numbers of alternative scenarios and to quantify the uncertainty attached to a prediction.

Classical technical modelling normally begins with a structure specified by the researcher: variables are selected, equations are defined, thresholds are established and the model is then applied to observations. In AI-assisted research, part of this structure can be learned from data. The researcher continues to define the scientific problem, the spatial and temporal support, the relevant processes, the admissible variables and the conditions of validation, but the model may discover nonlinear relationships or interactions that were not explicitly introduced beforehand.

GeoAI has been described as a spatially explicit form of artificial intelligence directed towards geographical knowledge discovery (Janowicz et al., 2020). This definition is important. A generic AI model may identify statistical regularities, whereas a geographical AI model must determine whether those regularities remain meaningful when location, scale, neighbourhood, spatial dependence, temporal evolution and geographical process are considered. Recent Earth-system research has similarly argued for hybrid approaches in which data-driven learning and process understanding reinforce one another rather than compete (Reichstein et al., 2019). Theory-guided data science makes the same point in more general terms: scientific consistency and domain knowledge should enter the learning process, particularly where observations are limited and physical processes are complex (Karpatne et al., 2017).

Consequently, the development of AI does not reduce the need for geographical reasoning. It increases it. The larger the space of patterns and models that a machine can generate, the more necessary it becomes for the geographer to decide which relations are spatially meaningful, physically plausible, causally defensible and socially relevant.

2. THE INVERTED RETROSPECTIVE PARADIGM

A conventional Letter on AI in Geography could enumerate a large number of possibilities: automatic image classification, object detection, prediction of environmental hazards, processing of geographical texts, optimisation of transport networks, climate modelling or decision support. Such a catalogue would be informative, but it would remain largely prospective and theoretical.

The present Letter adopts an inverted retrospective paradigm. It begins not with the capabilities of AI, but with completed geographical studies. It asks a counterfactual question: How might the research design, processing workflow, results and conclusions of an earlier geographical study have changed if contemporary AI methods had been available when the study was conducted?

Seven studies in which I was an author or co-author were deliberately selected. Restricting the corpus to my own work has two purposes. First, it avoids retrospectively criticising other researchers for not having used methods that were unavailable at the time of their research. Second, it transforms the exercise into methodological self-examination. The earlier papers are not presented as obsolete or inadequate. They are treated as authentic records of the technical possibilities, data constraints and scientific reasoning available at particular moments.

Each original study therefore functions as a methodological baseline. Its aims, datasets, models, results and acknowledged limitations are retained. The AI-assisted redesign does not replace this baseline; it asks what additional pathways might have become possible. In some cases, AI would probably have accelerated an operation already performed. In others, it might have allowed uncertainty to be quantified, different data streams to be fused, a previously inaccessible relation to be tested, or a retrospective analysis to become predictive.

This distinction also imposes an important epistemological limit. Because the original datasets have not been reprocessed with the proposed AI models, the present Letter cannot claim that the resulting maps, coefficients or forecasts would certainly have been more accurate. Expressions such as AI could have enabled, might have revealed, or would probably have reduced are scientifically more appropriate than unconditional assertions of superiority. The experiment is not an empirical replication of the seven studies. It is an experiment in research design: a systematic comparison between the workflow that was historically implemented and the expanded workflow that can now be conceived.

3. METHOD OF THE HUMAN–AI RETROSPECTIVE EXPERIMENT

The seven studies cover different components of geographical research: forest remote sensing, severe-weather analysis, hydraulic modelling, urban hydrology, time-series analysis, transport geography and atmospheric pollution. This disciplinary diversity makes it possible to evaluate whether AI contributes only to a narrow class of computational problems or whether it modifies geographical reasoning more broadly (**Table 1**).

Table 1.

Common protocol applied to the seven retrospective cases.

Retrospective question	Analytical purpose
What geographical question was originally addressed, and through what technical workflow?	Reconstruct the historical methodological baseline.
At which stages could contemporary AI enter the workflow?	Identify concrete rather than generic AI applications.
What type of benefit might result?	Distinguish speed, accuracy, integration, uncertainty, discovery, prediction and decision support.
What data and validation would be required?	Avoid presenting a plausible AI proposal as an already demonstrated result.

The potential contribution of AI is classified into six categories: workflow acceleration, improved extraction or classification, integration of heterogeneous data, uncertainty quantification, discovery of additional relationships, and transformation from retrospective analysis into prediction or decision support.

An AI proposal is retained only when it is compatible with the geographical process and the data structure of the original study. A technically possible model is not necessarily geographically justified. For example, a high predictive score is insufficient if the model ignores spatial dependence, uses information from the future during training, confuses exposure with risk, or violates a physical conservation law.

The retrospective assessment reported below was developed through a documented, iterative dialogue between the author and an artificial-intelligence system. The author defined the conceptual framework, selected the case studies and evaluated the scientific relevance of the proposed methodological alternatives. The complete disclosure of the respective human and AI contributions is provided at the end of the Letter.

All seven papers were already published; no confidential third-party manuscript was used. The exercise also separates AI-assisted reasoning from scientific evidence: the proposed methods are supported by published methodological literature, while the AI output itself is not treated as a bibliographical source.

4. FROM THREE FOREST-COVER MAPS TO LEARNING FOREST TRAJECTORIES

Original research design. The first study combined forestry maps and inventory tables with Landsat images acquired in 1988, 2000 and 2010. A common spectral library was constructed, and the Spectral Angle Mapper (SAM) classifier was applied to identify forest and non-forest land-cover classes. The resulting maps were processed in GIS to assess deforestation in three Apuseni Mountain catchments, with forest-cover changes approaching 22% of one catchment (Costea et al., 2012). The method had an important strength: it preserved spectral interpretability and applied a common classification logic to the three images. Its validation produced high overall accuracy and Kappa values of 0.98, 0.94 and 0.91. Nevertheless, the analysis was based on three temporal snapshots and identified difficulties associated with mixed forests, the water-stress conditions visible in the 2000 image, vegetation water content, illumination and the precision of forestry inventories.

AI-assisted redesign. I would retain SAM as a physically interpretable benchmark. AI should not remove a transparent method merely because a more complex method is available. I would add comparative machine-learning models, particularly Random Forests (Breiman, 2001), using not only spectral bands but also vegetation indices, altitude, slope, aspect and spatial texture. The more substantial change would be to formulate the problem directly as multitemporal change detection. Fully convolutional Siamese architectures can analyse two co-registered images simultaneously and learn the spatial and spectral characteristics of change (Daudt et al., 2018), while a U-Net-type encoder-decoder can preserve both contextual information and local boundaries (Ronneberger et al., 2015). Instead of using only three dates, the complete available Landsat archive could be introduced into a temporal model. It could learn different trajectories: stable forest, gradual degradation, abrupt removal, temporary drought stress, regeneration and afforestation. Training could be organised through active learning, with the model identifying the spatial units for which expert labelling would reduce uncertainty most efficiently.

Potential added value and validation requirements. The output would no longer consist only of land-cover maps and differences between three dates. It could include the probable date of change, the probability that a change represents deforestation rather than temporary stress, the rate of vegetation recovery and the uncertainty associated with each mapped object. The spatial pattern of forest loss could then be connected to a hydrological model or to an AI emulator of that model. The study could ask not only how much forest disappeared, but which spatial configurations of forest removal most strongly affect runoff and flash-flood generation. A rigorous evaluation would require temporally independent reference samples, harmonisation among Landsat sensors and forestry information capable of distinguishing harvesting, degradation, drought damage and subsequent recovery. Remote sensing and AI should not be expected to infer the legal or socioeconomic cause of every change without additional evidence. The principal contribution would be a change from comparing maps to learning geographical trajectories.

5. FROM RECONSTRUCTING A WINDTHROW EVENT TO PREDICTING ITS SPATIAL IMPACT

Original research design. The windthrow study examined the severe event of 20 July 2011 in the Apuseni Mountains. Areas affected by windthrows were detected through Landsat imagery and differences in NDVI, while synoptic maps and WSR-98D Doppler-radar products were used to reconstruct the regional and mesoscale atmospheric evolution. The radar analysis included reflectivity, storm movement, Vertically Integrated Liquid water and echo-top height (Furtuna et al., 2018). The study identified convective descending movements, three storm cores and short-duration wind gusts exceeding 23 $m\ s^{-1}$. It concluded by asking whether recurrent storm trajectories and preferential travel corridors could be identified in relation to relief and baric configuration.

AI-assisted redesign. The first AI component would be a semantic segmentation model trained on pre- and post-event satellite images to delineate windthrow patches from spectral, textural and morphological evidence rather than from a single NDVI-difference criterion. The second component would operate on the complete sequence of radar images. ConvLSTM was developed specifically to represent spatial structure and temporal evolution in radar-based precipitation nowcasting (Shi et al., 2015). Applied to this case, such a model could learn the development, movement, splitting and dissipation of convective cells. A third model would estimate forest susceptibility from wind speed and direction, VIL, echo-top height, slope, aspect, valley orientation, topographic exposure, soil moisture, distance from forest edges, tree species and stand structure. An explainability method such as SHAP could show why the model assigns high susceptibility to a particular sector, rather than offering only a black-box score (Lundberg and Lee, 2017). Once a sufficiently large regional archive had been assembled, clustering could group storm trajectories and identify recurrent corridors through valleys, saddles and depressions.

Potential added value and validation requirements. The resulting system could combine forecast storm intensity with local forest susceptibility to estimate the spatial probability of windthrow during the following 30–60 minutes. This would transform the study from reconstruction of a past event into a component of near-real-time forest-risk management. The principal limitation is the rarity of severe events. One storm cannot provide sufficient examples for a robust deep-learning model. Training would require a regional archive covering many Carpathian storms, or transfer learning from comparable mountainous areas followed by local recalibration. The retrospective value of AI in this case is especially clear: it would make operational the future research question already formulated in the original paper.

6. FROM A DETERMINISTIC FLOODPLAIN BOUNDARY TO A PROBABILITY OF INUNDATION

Original research design. The floodplain study linked HEC-HMS hydrological modelling to the one-dimensional MIKE11 hydrodynamic model. The workflow included river axes and cross-sections, roughness coefficients, river-network representation and discharge boundary conditions. A 2 m DEM derived from LiDAR and bathymetric information was used. Cross-sections were generally generated at 100 m intervals, and the article explicitly recognised that modeller subjectivity could intervene decisively in their selection. The comparison between statistical and simulated maximum discharges produced a mean relative error close to 10% (Győri et al., 2016).

AI-assisted redesign. This case should not be approached by replacing hydraulic equations with an unconstrained neural network. It is better suited to theory-guided and hybrid AI (Karpatne et al., 2017; Reichstein et al., 2019). Computer vision applied to high-resolution terrain data could identify the channel axis, banks, bridges, defence structures and abrupt geometrical changes. Cross-sections could then be positioned adaptively: more densely where geometry changes rapidly and less densely in uniform sectors. Orthophotos and LiDAR-derived information could be classified automatically to produce spatially variable estimates of roughness, still subject to hydraulic interpretation and field verification. Bayesian optimisation could calibrate Manning coefficients, boundary conditions and hydrological parameters. After a representative set of MIKE11 simulations had been generated, a surrogate model could learn the relationship among discharge, geometry, roughness and water level. Physics-informed neural networks offer one way of constraining learning with governing differential equations (Raissi et al., 2019), although their suitability would have to be demonstrated rather than assumed.

Potential added value and validation requirements. The most important output would be a probabilistic flood map rather than a single apparently exact boundary. Each location could be associated with a probability of inundation, an expected range of water depths and a decomposition of uncertainty into contributions from discharge, topography, roughness and model structure. The surrogate could permit rapid exploration of bridge obstruction, vegetation growth, channel modification, local defence walls or land-use change. Active learning could indicate the cross-section at which one additional field survey would reduce final uncertainty most efficiently. The original hydraulic model would remain the reference. Any surrogate would require evaluation both inside and outside the parameter combinations used for training, together with tests of mass conservation and hydraulic plausibility. AI would thus transform the map from a deterministic end product into an interactive and uncertainty-aware decision system.

7. FROM MEASURING URBAN IMPERVIOUSNESS TO OPTIMISING URBAN INTERVENTION

Original research design. The urban-runoff study used four Landsat scenes, from 1986, 1993, 2002 and 2014, to identify the expansion of impervious surfaces. Maximum Likelihood, Mahalanobis Distance, Neural Networks and Support Vector Machine classification were compared, and Maximum Likelihood produced the highest accuracy in that experiment. The classified land-cover data were introduced into HEC-HMS. Impervious surfaces increased from 41.7% to 76.1%, accompanied by increases in total runoff and peak flow (Haidu and Ivan, 2016).

AI-assisted redesign. A contemporary redesign would first replace the binary representation of pervious and impervious surfaces with fractional imperviousness. A pixel containing roofs, gardens and pavement would no longer be assigned only to its dominant class. A U-Net-type segmentation model could integrate Landsat, orthophotos, LiDAR and street-network information to distinguish roofs, roads, parking areas, compacted soil, vegetation and semipermeable surfaces (Ronneberger et al., 2015). A multitemporal model would estimate the probable year of conversion for each sector instead of relying on four isolated dates. Hydrological modelling could be extended through radar rainfall, the storm-water network, microtopography, antecedent moisture and multiple observed rainfall events. An AI emulator trained on hydrological simulations would allow thousands of combinations of storms and land-cover configurations to be assessed rapidly. The next methodological step would be multi-objective optimisation of permeable pavements, green roofs, detention basins, rain gardens, urban vegetation and water-storage systems.

Potential added value and validation requirements. The question would change from how increasing imperviousness affected runoff to which urban sectors contribute most strongly to peak discharge and where interventions should be placed to obtain the greatest reduction for a given budget. AI could estimate the marginal contribution of each sector and identify nonlinear thresholds beyond which additional imperviousness generates disproportionate hydrological effects. At the same time, AI should be used to test the robustness of the original conclusion, not merely to reinforce it. The reported correlation was calculated from four temporal configurations subjected to the same rainfall scenario. A richer analysis would require more observation dates, multiple storm events and measured hydrological responses. A complex model must not convert a small-sample association into an unjustified causal statement. The deeper contribution would be the transition from historical impact assessment to spatial optimisation of urban adaptation.

8. FROM ANIMATED TREND SEGMENTATION TO PROBABILISTIC REGIME DETECTION

Original research design. The trend-signal paper proposed an animated method that decomposed a finite time series into sequential sloped trends and displayed the changing signal as new observations accumulated. The method used the Mann–Kendall–Sneyers transformation and a user-defined lag to identify breakpoints. The resulting segments were fitted by least squares, removed from the original series, and the residual was tested iteratively until it approached white noise (Haidu and Magyari-Saska, 2009). In retrospect, this algorithm already contained a principle close to contemporary online learning: the representation of the signal was updated as new data became available.

AI-assisted redesign. The principal extension would be Bayesian Online Changepoint Detection. Instead of selecting a single breakpoint only after the lag criterion is met, the Bayesian procedure continually estimates the probability that a new regime has begun and the posterior distribution of the duration of the current regime (Adams and MacKay, 2007). A switching state-space model, Hidden Markov Model or Gaussian-process model could distinguish changes in slope, mean, variance and autocorrelation. This would help separate persistent trends, abrupt shifts, isolated anomalies and quasiperiodic fluctuations. The lag would no longer depend mainly on manual choice. Simulated series containing known trends, cycles, autocorrelation, noise and breakpoints could be used to select parameters according to detection rate, false-alarm rate, delay and slope error.

Potential added value and validation requirements. The result would include the posterior probability of each breakpoint, uncertainty intervals for the moment of change, probability distributions for segment slopes and a measure of whether the current regime is likely to continue. The original animation could become an operational warning interface. As temperature, precipitation or discharge observations are added, the system could report the probability that a statistically different regime has emerged. Such probabilities would still depend on the assumed observation model, the prior hazard of change and the treatment of long-range dependence. The method would therefore require sensitivity analysis and testing on both synthetic and long instrumental series. This case shows particularly clearly that AI does not appear without intellectual predecessors. Earlier geographical algorithms already anticipated sequential learning, adaptive segmentation and progressive separation of signal from noise; AI would formalise and probabilistically extend that logic.

9. FROM ACCIDENT HOTSPOTS TO DYNAMIC NETWORK RISK

Original research design. The road-accident study analysed 190 accidents registered between 2010 and May 2012. It identified higher frequencies during daytime, rush hours, spring and autumn, and a concentration along the principal urban arteries with a SW–NE orientation. Accident information was collected from local media and integrated with a digitised street network. The article recognised missing information concerning exact causes, vehicle direction and the duration of traffic disruption (Ivan and Haidu, 2012).

AI-assisted redesign. Natural-language processing could extract information automatically from press reports or police records: location, date, time, vehicle type, number of victims, reported cause, weather, manoeuvre and traffic restrictions. Geocoding and entity-matching models could locate events and identify duplicate reports of the same accident. The more important analytical change would be adjustment for exposure. A road with many accidents may be intrinsically dangerous, or it may simply carry much more traffic. Traffic volume, pedestrian flow, vehicle-kilometres and time spent in traffic should therefore be included. The street network could be represented as a graph. Spatio-temporal graph convolutional models were developed to learn connected network structure and temporal dependence simultaneously (Yu et al., 2018). In this context, a network model could use road geometry, intersections, speed limits, lighting, pedestrian crossings, traffic, weather, schools and commercial activities to estimate accident risk for each segment and time interval. Unlike an ordinary Euclidean Kernel surface, it would respect connectivity and direction in the street network.

Potential added value and validation requirements. The output could become a dynamic risk map: the probability of a severe accident on a particular road segment, during a specified time interval and under specified traffic and weather conditions. The model could distinguish risk arising from high exposure from risk associated with geometry, speed, pedestrian interaction, visibility or weather. Causal models could then evaluate a lower speed limit, a redesigned intersection, improved lighting or a new pedestrian crossing by comparing treated segments with similar untreated segments before and after intervention. AI cannot, however, reconstruct with certainty variables that were never observed. Nor can it remove the reporting bias of local media, which are more likely to report severe or unusual accidents. A longer, official and anonymised database would be required, together with explicit safeguards for privacy and fair enforcement. Here AI would transform descriptive hotspot analysis into exposure-adjusted risk estimation and counterfactual evaluation of prevention measures.

10. FROM A BEFORE–AFTER COMPARISON TO A CAUSAL ATMOSPHERIC COUNTERFACTUAL

Original research design. The atmospheric-pollution study examined changes in NO2, HCHO, SO2, O3, CO and aerosol index in ten urban areas of the Grand Est region during the first COVID-19 lockdown. Sentinel-5P observations, ground-monitoring data, meteorological variables, land-cover data and socioeconomic indicators were integrated. An Atmospheric Clearance Index (ACI) was developed, and statistically significant relationships were identified between this index and population density and economic indicators (Kovács and Haidu, 2021). The design compared the lockdown interval in 2020 with the same interval in 2019, examined meteorological homogeneity and constructed ACI by standardising and summing the six pollutant-difference rasters, effectively giving the components equal weight.

AI-assisted redesign. The main AI contribution would concern causal counterfactual construction rather than image classification. A Random Forest or gradient-boosting model could learn the expected concentration of each pollutant from meteorology, seasonality, day of the week and long-term temporal behaviour. The actual 2020 meteorological conditions could then be introduced into this model to estimate the pollution levels expected in the absence of the lockdown. Machine-learning meteorological normalisation has been used precisely to separate weather-related variation from changes associated with emissions or policy (Grange et al., 2018). A Bayesian structural time-series model could construct a second counterfactual from earlier years, comparable regions, mobility, traffic, energy use and industrial activity. Such models estimate an intervention effect by comparing the observed series with a probabilistic counterfactual trajectory (Brodersen et al., 2015). The construction of ACI could also be extended. Rather than assigning equal weights by default, weights could be estimated from a larger European dataset, a latent-variable model, toxicological importance or relationships with health and visibility indicators.

Potential added value and validation requirements. The results could estimate, for each pollutant and city, the change expected from meteorology, the additional change associated with reduced mobility, the contribution of industrial activity, the remaining unexplained component and the uncertainty interval of each estimate. The study could then make statements closer to causal inference while still recognising the observational nature of the evidence. The short TROPOMI archive and the sample of ten cities remain fundamental limitations. A sophisticated AI model cannot create missing pre-2018 satellite observations or a perfect control group. It must express those limitations rather than conceal them. The retrospective redesign would move the analysis from a comparison between two periods towards an explicit estimate of what atmospheric conditions might have occurred without the intervention.

11. CROSS-CASE SYNTHESIS

Although the seven studies concern different phenomena and data structures, their retrospective redesigns converge towards a small number of methodological transformations. **Table 2** summarises the principal transition proposed for each case.

First, AI moves geographical research from rules completely specified in advance towards relationships partly learned from observations. Spectral libraries, thresholds, lags and roughness coefficients remain meaningful, but they can be compared, calibrated or adapted through learning.

Second, AI encourages multimodal integration. Satellite imagery, radar, DEMs, time series, networks, socioeconomic indicators and expert knowledge can be represented within a common analytical workflow. This does not mean that all available data should be used indiscriminately. Their spatial support, temporal resolution, uncertainty and causal status must first be made compatible

Table 2.

Principal methodological transition in the seven retrospective cases.

Case	Original technical core	Principal AI extension	Possible new knowledge / key validation
Forest cover	SAM and GIS change detection	Siamese segmentation and temporal learning	Timing, persistence and type of change; independent temporal ground truth.
Windthrows	NDVI, synoptic maps and radar interpretation	Radar sequence learning and susceptibility modelling	Near-real-time impact probability; multi-event regional archive.
Floodplains	HEC-HMS and MIKE11	Hybrid physical–AI surrogate and Bayesian calibration	Probabilistic inundation; conservation and out-of-sample tests.
Urban runoff	Landsat classification and HEC-HMS	Fractional segmentation and intervention optimisation	Marginal runoff contribution; multiple storms and observed runoff.
Trend signals	MKS segmentation and animated updating	Bayesian online regime detection	Probability and early warning of structural change; benchmark series.
Road accidents	GIS hotspots and temporal frequencies	NLP, graph learning and causal evaluation	Exposure-adjusted dynamic risk; official traffic and accident data.
Air pollution	Satellite differences and composite ACI	Meteorological normalisation and causal counterfactuals	Intervention effect separated from weather; longer series and controls.

.

Third, AI makes it possible to replace a single deterministic result with a probability distribution. A mapped pixel is not merely deforested or unchanged, flooded or unflooded, hazardous or safe. It can be accompanied by a probability, an uncertainty interval and an explanation of the variables that contributed to the result.

Fourth, AI extends research from description towards counterfactual reasoning. It asks not only what occurred, but what would have occurred under another land cover, another policy, another storm or the absence of an intervention.

Fifth, AI links geographical analysis to optimisation. A model can search for the location, timing or combination of interventions that produces the greatest reduction in risk or the largest environmental benefit under explicit constraints.

The deepest benefit is consequently not computational speed. Speed matters, especially where thousands of images or scenarios must be processed, but it is the most superficial advantage. The stronger claim emerging from the seven cases is that AI enlarges the class of geographical questions that can be formulated and made operational.

12. ARTIFICIAL INTELLIGENCE AND CRITICAL GEOGRAPHICAL THINKING

One of the principal objections to AI in education and research is that it may inhibit human critical thinking. This risk should not be dismissed. It can occur when an AI-generated answer is accepted without examination, when references are not verified, when a plausible explanation is confused with evidence, or when a researcher delegates the formulation of the scientific problem itself.

Nevertheless, cognitive inhibition is not an inevitable effect of AI. It depends on the mode of use. In a substitution mode, the human asks for a finished answer and accepts it; here passivity is a genuine risk. In an assistance mode, AI performs repetitive operations while the researcher retains the analytical design. In a collaborative mode, AI generates alternative hypotheses, methods and interpretations that the researcher must compare, challenge, modify or reject. In this third mode, critical thinking is not eliminated. It is required more frequently and at a higher level.

The critical questions also change. The researcher must ask whether the training data are representative; whether spatial autocorrelation or temporal overlap has produced data leakage; whether the chosen scale and resolution are appropriate; whether a model transfers from one region to another; whether a correlation has been interpreted as causation; whether a physical law has been violated; whether uncertainty has been measured; and who may be harmed by a mistaken prediction. Explainability tools can help interrogate a complex model, but they do not transfer responsibility from the researcher to the algorithm (Lundberg and Lee, 2017).

In my own work, I have experienced the opposite of intellectual inhibition. The dialogue that generated this Letter widened the space of methodological alternatives and helped convert an initial intuition into the inverted retrospective paradigm used here. The seven papers were not selected by the AI system, and the link between this experiment, Technical Geography and *antefactum* science did not emerge autonomously from a machine. Those decisions required personal scientific history, editorial responsibility and a conception of where Geography should develop.

A personal experience cannot settle the general educational question, and it should not be presented as universal evidence that AI always increases creativity. It does demonstrate that AI use is compatible with creative geographical reasoning. When critically supervised, it can stimulate questions that might not emerge in the same form through a routine literature review. The relevant educational objective should therefore not be to preserve every manual operation, but to teach students and researchers how to formulate problems, inspect assumptions, verify evidence, compare models and reject unsupported outputs.

Critical thinking should not be measured by the number of calculations performed manually. It should be measured by the quality of the questions, assumptions, tests, interpretations and decisions for which the researcher accepts responsibility.

13. FROM *POSTFACTUM* TO *ANTEFACTUM* GEOGRAPHY

In the 2024 Letter, I described classical Geography as predominantly *postfactum* because it frequently explains phenomena after they have occurred. I argued that Technical Geography should become *antefactum*: predictive, anticipatory and capable of proposing action before damage occurs (Haidu, 2024). The present retrospective experiment shows more concretely how this transition can occur.

Forest research can move from mapping previous losses to forecasting where particular patterns of forest removal may increase hydrological risk. Windthrow research can move from reconstructing storm trajectories to nowcasting the areas that may be affected during the next hour. Floodplain research can move from a fixed hazard boundary to continuously updated probabilities. Urban hydrology can move from documenting increased runoff to selecting the most effective intervention. Trend analysis can move from retrospective segmentation to early detection of a new regime. Accident geography can move from historical hotspots to dynamic segment-level risk. Atmospheric research can move from a before–after comparison to a counterfactual estimate of intervention effects.

This progression may be expressed conceptually as:

Observation → Detection → Explanation → Prediction → Counterfactual Evaluation → Optimisation.

Classical descriptive Geography operated mainly in the first stages.

Technical Geography extended detection, explanation and modelling.

AI can accelerate and connect prediction, counterfactual evaluation and optimisation.

A second conceptual expression may therefore be proposed:

***Antefactum* Technical Geography = geographical reasoning + spatial–temporal data + physical and statistical modelling + learning + uncertainty assessment + decision criteria.**

This is not a mathematical identity. It expresses the components required for a genuinely anticipatory Geography. AI alone cannot create *antefactum* Geography. Prediction without geographical scale, physical meaning, causal interpretation or uncertainty can produce confident but misleading results. Conversely, geographical knowledge without learning and updating may remain unable to process the volume and complexity of contemporary spatial–temporal observations. The *antefactum* potential is generated by their combination.

14. CONCLUSIONS

The retrospective examination of seven earlier studies does not show that their methods were wrong or that their conclusions should be abandoned. Each study addressed a real geographical problem with the data, software and methodological possibilities available at the time.

The exercise demonstrates instead that contemporary AI could have opened additional analytical pathways. In some studies, the principal benefit would probably have been faster classification or calibration. In others, it would have been uncertainty assessment, integration of previously separate datasets, online updating, causal counterfactual construction or optimisation of interventions.

AI's most important contribution is therefore epistemic rather than merely computational. It does not only perform an existing operation more quickly. It may make a new question operational.

Nevertheless, the expected benefit remains conditional. AI requires representative data, independent validation, respect for spatial and temporal dependence, comparison with transparent baselines, physical constraints where appropriate and explicit uncertainty. A complex model should not be accepted because it is complex, and a high predictive score should not replace geographical interpretation.

Human responsibility remains indivisible. AI cannot decide which geographical problem is socially important, whether a mapped pattern is causally meaningful, whether a prediction is sufficiently reliable for public intervention, or whether the costs and benefits of an action are ethically acceptable.

The proper relationship is therefore not Geography versus AI, and not geographer versus machine. It is a relationship in which geographical reasoning defines space, scale, process, causality and validation, while AI extends the capacity to learn, compare, anticipate and explore.

Accordingly, the Editorial Note below sets out the principles that will also be incorporated into the journal's Policy and Guide for Authors and Reviewers.

The future of Geography is not only technical. It is technical, learning-enabled, critically supervised and directed from *postfactum* explanation towards *antefactum* anticipation.

EDITORIAL NOTE: A JOURNAL POLICY FOR RESPONSIBLE AI USE

Geographia Technica permits the responsible use of artificial intelligence and AI-assisted technologies in geographical research and manuscript preparation. AI systems must not be listed as authors. Authors must disclose the name of the tool, the model or version, the purpose of its use and the stages of the research or writing process in which it was employed. Human authors remain fully responsible for the accuracy, originality, references, interpretation and ethical compliance of the submitted work.

When AI is used in data processing, modelling, coding, image generation or modification, or visualisation, its use must be described in sufficient detail in the Methods section to permit scientific evaluation and, where possible, reproducibility. AI-generated or AI-modified visual material must be explicitly identified.

Editors and reviewers must not upload confidential manuscripts, figures, datasets, correspondence or review materials to external AI systems unless the tool has been expressly authorised by the journal and appropriate confidentiality and data-protection safeguards are in place. Any authorised use of AI during peer review must be disclosed to the responsible editor.

This policy supplements the journal's existing adherence to COPE principles and its rules concerning authorship, originality, plagiarism, conflicts of interest and confidentiality. It will be published as a distinct, publicly accessible and versioned section of the journal's Policy and Guide for Authors and Reviewers.

DECLARATION OF AI-ASSISTED WORK

The inverted retrospective paradigm, the selection of the seven case studies, the geographical and epistemological framing, the editorial argument and the final conclusions were conceived, critically evaluated and approved by the author. OpenAI's ChatGPT, operating in Pro mode and powered by GPT-5.6 Sol Pro, was accessed on 23 July 2026 and used in an iterative dialogue to examine the seven published studies; propose AI-enhanced methodological alternatives; compare their potential contributions; organise the cross-case synthesis; and assist with drafting and linguistic refinement. The AI system was not treated as an autonomous source of scientific evidence or as an author. All factual claims, methodological descriptions, references and interpretations were reviewed and verified by the author. AI-generated suggestions were accepted, modified or rejected according to the author's scientific judgment. The author assumes full responsibility for the accuracy, integrity, originality and final content of the published Letter.

REFERENCES

Adams, R.P. & MacKay, D.J.C. (2007) Bayesian Online Changepoint Detection. arXiv:0710.3742. https://doi.org/10.48550/arXiv.0710.3742

Breiman, L. (2001) Random Forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brodersen, K.H., Gallusser, F., Koehler, J., Remy, N. & Scott, S.L. (2015) Inferring causal impact using Bayesian structural time-series models. The Annals of Applied Statistics, 9(1), 247–274. https://doi.org/10.1214/14-AOAS788

Costea, G., Serradj, A. & Haidu, I. (2012) Forest cartography using Landsat imagery, for studying deforestation over three catchments from Apuseni Mountains, Romania. In: Advances in Remote Sensing, Finite Differences and Information Security, WSEAS Press, 109–114.

Daudt, R.C., Le Saux, B. & Boulch, A. (2018) Fully Convolutional Siamese Networks for Change Detection. Proceedings of the 25th IEEE International Conference on Image Processing, 4063–4067. https://doi.org/10.1109/ICIP.2018.8451652

Furtuna, P., Haidu, I. & Maier, N. (2018) Synoptic processes generating windthrows. A case study in the Apuseni Mountains (Romania). Geographia Technica, 13(2), 52–61. https://doi.org/10.21163/GT_2018.132.04

Grange, S.K., Carslaw, D.C., Lewis, A.C., Boleti, E. & Hueglin, C. (2018) Random forest meteorological normalisation models for Swiss PM10 trend analysis. Atmospheric Chemistry and Physics, 18, 6223–6239. https://doi.org/10.5194/acp-18-6223-2018

Győri, M.-M., Haidu, I. & Humbert, J. (2016) Deriving the floodplain in rural areas for high exceedance probability having limited data source. Environmental Engineering and Management Journal, 15(8), 1879–1887. https://doi.org/10.30638/EEMJ.2016.201

Haidu, I. (2016) What is Technical Geography. Geographia Technica, 11(1), 1–5. https://doi.org/10.21163/GT_2016.111.01

Haidu, I. (2024) Geography's Future is Technical – a letter from the editor. Geographia Technica, 19(2), i–vi. https://doi.org/10.21163/GT_2024.192.23

Haidu, I. & Ivan, K. (2016) The assessment of the impact induced by the increase of impervious areas on surface runoff. Case study: the city of Cluj-Napoca, Romania. Carpathian Journal of Earth and Environmental Sciences, 11(2), 331–337.

Haidu, I. & Magyari-Saska, Z. (2009) Animated Sequential Trend Signal Detection in Finite Samples. Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces, 249–254. https://doi.org/10.1109/ITI.2009.5196088

Ivan, K. & Haidu, I. (2012) The spatio-temporal distribution of road accidents in Cluj-Napoca. Geographia Technica, 2/2012, 32–38.

Janowicz, K., Gao, S., McKenzie, G., Hu, Y. & Bhaduri, B. (2020) GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. International Journal of Geographical Information Science, 34(4), 625–636. https://doi.org/10.1080/13658816.2019.1684500

Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N. & Kumar, V. (2017) Theory-guided data science: A new paradigm for scientific discovery from data. IEEE Transactions on Knowledge and Data Engineering, 29(10), 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168

Kovács, K.D. & Haidu, I. (2021) Effect of Anti-COVID-19 measures on atmospheric pollutants correlated with the economies of medium-sized cities in 10 urban areas of Grand Est Region, France. Sustainable Cities and Society, 74, 103173. https://doi.org/10.1016/j.scs.2021.103173

Lundberg, S.M. & Lee, S.-I. (2017) A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

Raissi, M., Perdikaris, P. & Karniadakis, G.E. (2019) Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. & Prabhat (2019) Deep learning and process understanding for data-driven Earth system science. Nature, 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1

Ronneberger, O., Fischer, P. & Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science, 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K. & Woo, W.-C. (2015) Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. Advances in Neural Information Processing Systems, 28, 802–810.

Yu, B., Yin, H. & Zhu, Z. (2018) Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, 3634–3640. https://doi.org/10.24963/ijcai.2018/505

Geographia Technica, Vol. 21, Special Issue: Artificial Intelligence Applications in Geography, 2026, pp 1 to 18

FLOOD MAPPING IN PHRA NAKHON SI AYUTTHAYA, THAILAND, UTILIZING SENTINEL-1 SAR IMAGERY AND DEEP LEARNING APPROACHES

Kritchayan INTARAT *[*1,2,3], ***Phatsachon KUANKHAMNUAN*** *[1,2], ***Woraman JANGSAWANG*** *[1,3]

DOI: 10.21163/GT_2026.211.04

ABSTRACT

This study presents a deep learning (DL)-based flood detection framework for Phra Nakhon Si Ayutthaya Province, Thailand. The framework integrates U-Net architecture with three encoder variants-ResNet50, ResNet101, and ResNet152-using synthetic aperture radar (SAR) imagery from the Sentinel-1 satellite. Model performance is assessed through statistical metrics including accuracy, precision, recall, F1-score, Dice Loss, intersection over union (IoU), and computational time. The U-Net model with a ResNet101 encoder achieved the best performance, with an accuracy of 91.5%, F1-score of 0.886, Dice Loss of 0.116, and IoU of 86.8%, requiring about 24 minutes of training. Despite longer training, the ResNet101-based U-Net substantially enhances flood detection accuracy, highlighting its value as a reliable tool for real-time monitoring and rapid response in flood-prone areas of Thailand.

Key-words: *Flood Detection; SAR Images; U-Net; ResNet; Phra Nakhon Si Ayutthaya Province*

1. INTRODUCTION

Climate variation has expanded the frequency and severity of natural disasters, especially in Southeast Asia. Flooding is a hazard that poses a significant threat to life and property. Among the countries in this region, Thailand experiences flooding almost every year. Particularly prevalent in the lowland terrain, Phra Nakhon Si Ayutthaya Province frequently experiences flooding in many areas. Floods rank among the most destructive natural disasters. Floods consistently generate widespread social and economic challenges across diverse regions, all over the world. Over the past few decades, flood events have become more frequent and more severe (Fakhri & Gkanatsios, 2025; Wang & Feng, 2025).

Climate change, which is resulting in shorter-duration and more intense rainfall, is an influential factor driving this trend. Research indicates that in many parts of the world, the number of areas susceptible to frequent floods is growing (Misra et al., 2025). According to a recent study, water-related disasters made up 48.2% of all disaster events (159 events), more than meteorological disasters (32.1%), climatic disasters (10%, 33 events), and geological disasters (9.7%, 32 events) combined (Sibandze et al., 2025). These figures demonstrate that floods continue to pose a serious threat to the world, especially in light of changing urban land use and environmental conditions. Flooding is also a significant concern in Southeast Asia. Geographical, climatic, and socioeconomic factors all contribute to this vulnerability (Birkmann, 2010; Torti, 2012). Flooding in Thailand has had a profound social and economic impact, destroying many homes and causing extensive property damage.

----

[1]*Research Unit in Geospatial Applications (Capybara Geo Lab), Faculty of Liberal Arts, Thammasat University, Thailand. (KI) intaratt@tu.ac.th, (PK) st125875@ait.ac.th*
[2]*Departement of Geography, Faculty of Liberal Arts, Thammasat University, Thailand.*
[3]*Center of Excellence in Digital Earth Applications and Emerging Technology (CoE: DEET), Sirindhorn International Institute of Technology, Thammasat University, Thailand. (WJ) m6622040878@g.siit.tu.ac.th*

The southwest and northeast monsoons greatly impact Thailand's climate, resulting in periods of high rainfall (Gale et al., 2013). Flooding is made worse by regular and extended rainfall, which is a major contributor to river bank overflows and inundation in the surrounding areas (Loc et al., 2020).

In 2011, many provinces in Thailand were hugely affected. Phra Nakhon Si Ayutthaya Province, in particular, considerably suffered profound social and economic disruption. According to a report by the Dutch Expertise Network for Flood Protection (DENFP), two people were killed and about 6,000 households were impacted by the 400–500 million baht in infrastructure damages (Ghozali et al., 2016). Because of its lowland terrain and the presence of multiple major rivers, Phra Nakhon Si Ayutthaya Province is most vulnerable to flooding that occurs almost annually (Munpa et al., 2022; Hirabayashi et al., 2013). When water levels rise above the river capacity, areas like Phak Hai, Bang Sai, Bang Ban, Sena, and Phra Nakhon Si Ayutthaya districts become flood-retention areas. Planning, managing, and reducing the effects all depend on the detection of floods (Wang et al., 2022). Bentivoglio et al. (2022) reported that traditional assessment techniques suffer from several drawbacks, including limited accuracy across vast regions, high labor demands, and time inefficiency, thereby highlighting the need for more efficient and scalable approaches to flood detection.

In order to overcome the disadvantages of traditional techniques and enhance performance, artificial intelligence (AI) techniques such as machine learning (ML) and deep learning (DL) have been increasingly applied in various fields, such as biomass calculation and carbon sequestration (Angkahad et al., 2024; Angkahad et al., 2025), medical diagnosis (Yang et al., 2021), including water analysis and flood detection (Magyari-Sáska et al., 2025 ; Shoko & Dube, 2024). Disaster risk assessment is one of the many fields that currently use DL techniques (Melgar-García et al., 2023). The processing of spatial data and its application to geographic information is particularly promising for convolutional neural network (CNN) architectures (Intarat et al., 2024).

Regarding the generalization of trained models, DL models still present imperfections. This limitation leads to significant variability when the models are applied across different regions or topographies (Liu et al., 2025). The goal of this work is to improve model performance by applying both U-Net architecture and a CNN designed explicitly for image segmentation to Synthetic Aperture Radar (SAR) data from Sentinel-1 satellites, which are commonly used in hydrological studies (Haidu et al., 2024) and function continuously, both day and night and under all weather conditions (Pech-May et al., 2023; Zhang et al., 2020). This combination will increase the model's capacity to detect flooded areas accurately and generate efficient training data.

This study aims to address these challenges by developing and evaluating a DL-based flood detection framework that integrates the U-Net architecture with multiple encoder variants (ResNet50, ResNet101, and ResNet152) using Sentinel-1 SAR imagery. The models are trained to differentiate between flooded and non-flooded areas at the pixel level by combining VV, VH, and VV–VH composites with a water index for labeling. The efficacy of these encoder configurations is compared using a variety of statistical measures, including precision, recall, F1-score, Dice Loss, and Intersection over Union (IoU), as well as computational efficiency. This strategy not only tackles issues regards computing cost and model generalization, but also enhances flood detection accuracy, providing a solid foundation for operational flood monitoring in complex lowland areas.

2. STUDY AREA

The province of Phra Nakhon Si Ayutthaya is in the middle of Thailand, about 75 kilometers north of Bangkok, covering 2,547.62 square kilometers between 14°6.8450' to 14°40.2925' N and 100°12.7623' to 100°49.3908' E. It borders Ang Thong and Lop Buri to the north, Saraburi to the east, Pathum Thani, Nonthaburi, and Nakhon Pathom to the south, and Suphan Buri to the west (**Fig. 1**). The province's land is mostly lowland terrain (2 to 20 meters elevation) intersected by three major rivers (Chao Phraya, Pa Sak, Lopburi) and numerous canals that support irrigation, transportation, and flood drainage. The average temperature in Phra Nakhon Si Ayutthaya is about 28.4 °C all year round, summer peaks above 40 °C, and annual rainfall of 1,000–2,000 mm, with most of it falling between May and October, resulting in a tropical savanna climate.

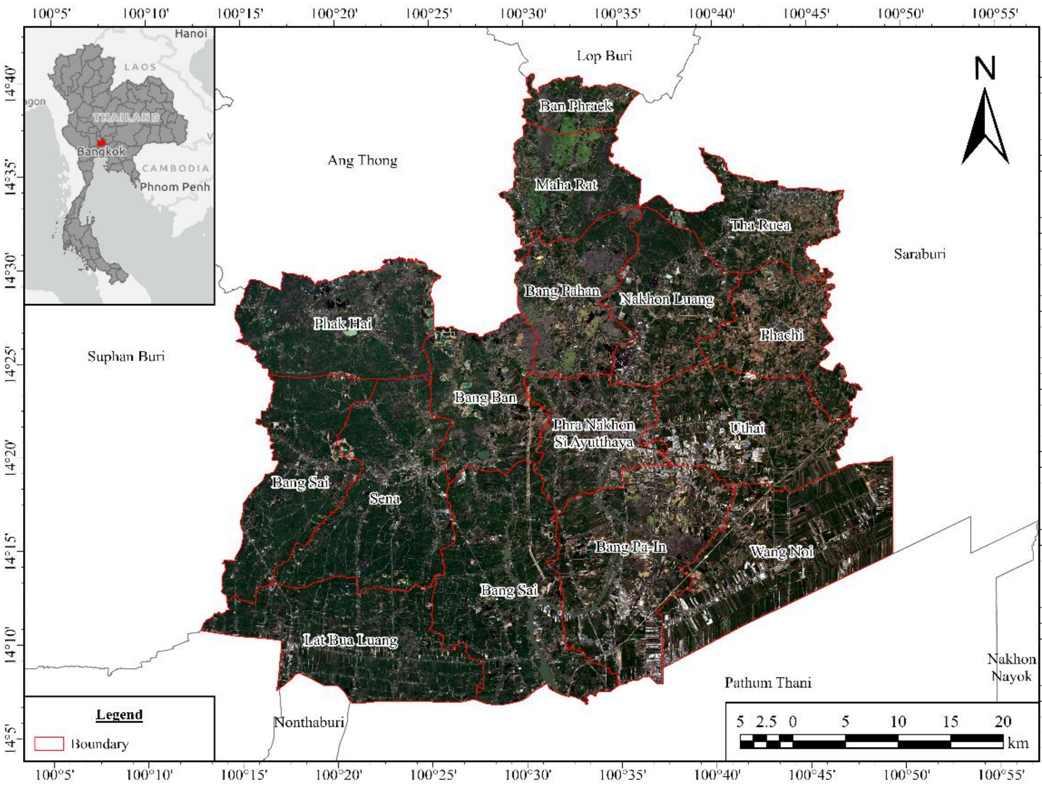


Fig. 1. The study area: Phra Nakhon Si Ayutthaya Province, Thailand.

The province is one of Thailand’s leading agricultural and industrial hubs, with numerous rice paddies and several large industrial estates that contribute significantly to the country’s manufacturing output. It is also of cultural and historical importance as the site of the ancient Ayutthaya Kingdom, a UNESCO World Heritage Site. However, its extensive lowlands, dense river network, and rapid urban-industrial growth make the area highly prone to severe flooding during the monsoon season, providing a critical case study for flood detection and risk management in Thailand’s central plains.

3. DATA AND METHODS

3.1. Remote sensing (RS) data

This study selected Sentinel-1 SAR data operated by the European Space Agency (ESA) in ground range detected (GRD) format on September 28, 2024, derived from the Google Earth Engine platform. The imagery was already pre-processed through four main steps: thermal noise removal, radiometric calibration, terrain corrections, and speckle filtering (Ponmani & Saravanan, 2021). This imagery includes both VV and VH polarization. To give the model more information, VV, VH, and VV-VH polarization overlays were added. Then, we calculated the water index (WI) as a criterion for thresholding in our study area (Wu et al., 2023). The threshold is set between 0.2 and 2. It enables us to distinguish between flooded and non-flooded areas in SAR imagery during flood events. WI can be expressed as in Eq. (1):

$$WI = \ln(10 \times VH \times VV) - 8 \tag{1}$$

where WI represents water index, VH and VV refer to Sentinel-1 SAR polarizations.

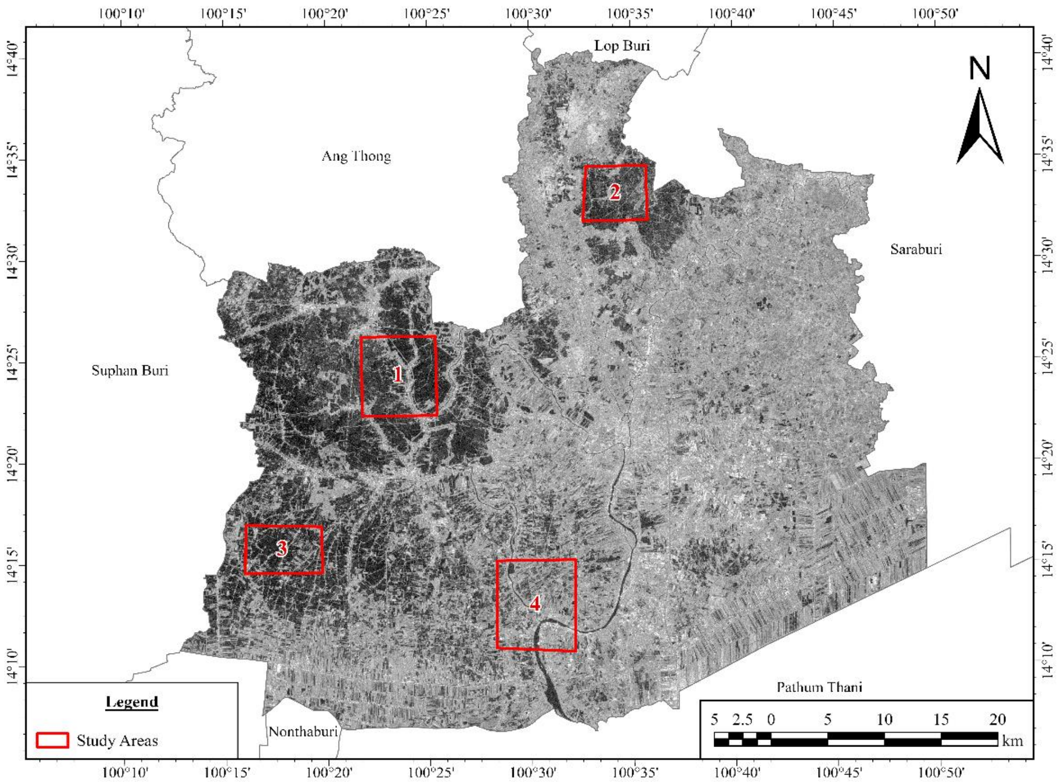


Fig. 2. Sentinel-1 SAR imagery showing four acquired samples used for model training (red bounding boxes).

The aforementioned method facilitated a preliminary distinction between flooded and non-flooded areas, which was subsequently used to assist in generating labels for the dataset. To enhance accuracy and reduce the time required for computation, four representative sites within Phra Nakhon Si Ayutthaya Province were delineated (**Fig.2**). The results obtained from WI calculations along with SAR imagery with overlaid with VV, VH, VV-VH polarizations were employed to construct dataset for model training and testing. The sample areas encompassed diverse land cover types, including flood-prone zones and riverine areas. In addition, the samples were spatially distributed to improve the model's robustness and accuracy (Wu et al., 2023).

All experiments were conducted on a high-performance workstation equipped with an Intel Xeon E5-2696V2 2.5 GHz processor, 128 GB of RAM, and an NVIDIA GeForce GTX 1080 Ti GPU with 11 GB of VRAM. The system ran Ubuntu 24.04 LTS, which is compatible with deep learning workflows and provides a stable and optimized environment. This setup enabled the quick processing of large Sentinel-1 SAR datasets, reducing the time required to train the models.

3.2. U-Net model

CNNs are the foundation of the deep learning architecture known as U-Net (Ronneberger et al., 2015). It is intended for pixel-level image segmentation (Konovalenko et al., 2022; Liu et al., 2020), especially for applications that call for accurate spatial delineation. The two primary components of U-Net's structure are the encoder (contracting path), which uses convolution, ReLU activation, and Max Pooling to extract deep image features and reduce the image's dimensionality, plus the decoder (expanding path), which upscales the image using transposed convolution and uses skip connections to connect the data to the original encoder layer (Melgar-García et al., 2023).

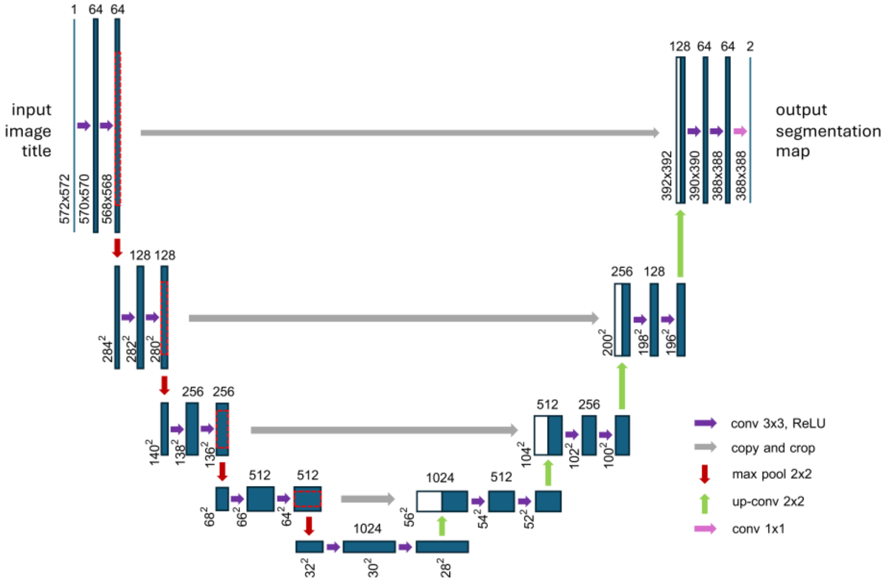


Fig. 3. The structure of U-Net: encoder-decoder with skip connections for image segmentation.

In **Fig. 3**, the U-Net's architecture is illustrated. In such an architecture, meaningful spatial information is preserved. It also improves the effectiveness of segmentation. According to Zhou et al. (2018), the outcome offers a very detailed segmented imagery that enables the U-Net to identify complex regions precisely.

U-Net has the advantage of being able to preserve spatial details and function effectively with a small dataset. It can also be applied to both optical and SAR imagery, including Sentinel-1 and Sentinel-2 (Konapala et al., 2021; Muszynski et al., 2022; Zhao et al., 2022). In addition, U-Net has demonstrated exceptional efficacy in flood mapping, particularly when applied to SAR imagery, which is weather-independent and capable of acquiring data both day and night. For instance, Ghosh et al. (2024) used Sentinel-1 SAR data to create a U-Net model for flood mapping. They discovered that it could precisely detect flooded areas even when there was cloud cover. Similarly, in order to map floods from high-resolution satellite data for prompt disaster response, Bonafilia et al. (2020) created a flood mapping model (FMM) based on U-Net architecture. When Dhanabalan et al. (2021) combined Sentinel-1 and Sentinel-2 data with a U-Net, they were able to monitor flooded areas in Kerala, India, with greater accuracy than conventional methods. According to its capabilities, U-Net can effectively apply Sentinel data, preserve spatial data in regions with complex terrain, and precisely determine flood extents.

3.3. ResNet encoder

The vanishing gradient and network degradation issues that are frequently present in deep models can be resolved by the DL architecture ResNet (He et al., 2016). The Residual Block, which consists of learning sub-functions and shortcut connections or identity mapping that enable the input to be passed straight to the next layer, is the primary mechanism of this architecture (Liu et al., 2023). The output is then obtained by adding the results of the sub-functions. This method enhances the training stability of networks with hundreds of layers. It makes deep model training more effective (He et al., 2016; Zhong et al., 2017).

ResNet architecture has been designed in multiple versions, including ResNet50, ResNet101, and ResNet152 (Intarat et al., 2024). The capabilities of ResNet include its fast convergence, complex hierarchical feature learning, adaptability to further development into better architectures like ResNeXt or Attention-ResNet, and its ability to be applied to very deep networks without gradient loss (Yang et al., 2022; Hassan & Maji, 2024; Intarat et al., 2024). Some drawbacks include high computational resource demands, the risk of overfitting with limited training data, and the need for structural adjustments when handling temporal data. Due to its proven ability, ResNet was incorporated into our U-Net model to learn complex deep features, adapt to varying resource constraints when determining model depth, and perform well across diverse image processing tasks. These advantages make ResNet a strong choice for tasks requiring both effective hierarchical deep learning and spatial precision.

3.4. Hyperparameter tuning

Tuning hyperparameters is an important step that can significantly improve the performance of the DL model. These hyperparameters define the model's architecture and learning process parameters (Mukherjee et al., 2024). In **Table 1**, the initial hyperparameters used in our study are shown. Correctly defining the retirees' demarcation may help the model learn more effectively from limited input data. It can also reduce the risk of overfitting or underfitting and improve the model's generalization when applied to other datasets.

Table 1.

Hyperparameters of the U-Net model used in this study.

Hyperparameter	Detail
Input shape	128 ✕ 128 pixel
Batch size	32
Optimizer	Adam
Loss function	Dice loss
Activation function	Sigmoid
Epoch	500
Encoder	ResNet50, ResNet101, and ResNet152
Encoder weight	ImageNet

According to **Table 1**, initial hyperparameters were set to train three models using different encoder architectures e.g. ResNet50, ResNet101, and ResNet152; hyperparameter tuning results in spatial details preservation and significant accuracy improvement. Lee et al. (2022) suggested that tuning the patch size, kernel size, and model structure provides higher precision and reduces validation loss significantly. Das et al. (2022) found that considering an appropriate optimizer and loss function can accelerate convergence and increase segmentation efficiency. Systematic hyperparameter fine-tuning not only helps increase the model's precision but also provides reliability and repeatability.

3.5. Model training and testing

According to information obtained in the previous step, the dataset was created by slicing the images into 128 x 128 pixels tiles. Then, these tiles were split into two sets: 80% for training and 20% for testing (Sazara et al., 2019). After the training, the model was used to predict and discriminate between flooded and non-flooded areas across the entire area. To evaluate the model's performance, the confusion matrix was associated with and delivered accuracy (Eq. (2)), precision (Eq. (3)), recall (Eq. (4)), and F1-score (Eq. (5)) (Amitrano et al., 2024). Each metric was calculated using true positive (TP), true negative (TN), false positive (FP), and false negative (FP) and expressed, accordingly:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$F1 - score = \frac{2\ x\ Precision\ x\ Recall}{Precision+Recall} \tag{5}$$

The model' s performance was also evaluated using the Dice Loss method (Pech-May et al., 2024), as it reflects accurate image segmentation, by comparing the overlap between predicted results and the ground truth. We also applied the IoU method (Eq. (6)) in the four sample areas mentioned earlier to examine the overlap between the predicted area and the actual ones (Safarov et al., 2022).

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{6}$$

where A refers to the predicted flood area and B reveals the actual flood area.

IoU lies between 0 and 1 (0 - 100%): a value close to 1 indicates high predictability of the model. IoU can also estimate positional accuracy and completeness in flood segmentation (Mosinska et al., 2018; Jamali et al., 2024). The comparison also included the evaluation of training time.

4. RESULTS

4.1. Model evaluation and comparison

Herein, we thoroughly tested the performance of the U-Net models that used ResNet50, ResNet101, and ResNet152 encoders using a variety of performance metrics, such as accuracy, precision, recall, F-1 score, and Dice Loss. **Table 2** presents a detailed summary of how well the model worked.

Table 2.

Evaluation Results of U-Net Models with ResNet50, ResNet101, and ResNet152 encoders.

Metric	ResNet50	ResNet101	ResNet152
Accuracy	0.9164	0.9154	0.9100
Precision	0.8790	0.8617	0.8476
Recall	0.8736	0.9118	0.9148
F1-score	0.8858	0.8860	0.8799
Dice Loss	0.1149	0.1161	0.1233

As shown in **Table 2**, the ResNet50 encoder achieved the highest overall accuracy (0.9164), surpassing ResNet101 (0.9154) and ResNet152 (0.9100). It also exhibited the highest precision (0.8790), indicating superior ability to identify flooded areas while minimizing misclassification of non-flooded pixels. ResNet101 and ResNet152 followed with precision scores of 0.8617 and 0.8476, respectively. In contrast, the recall scores displayed a different pattern: ResNet152 achieved the highest recall (0.9148), followed by ResNet101 (0.9118), with ResNet50 slightly lower at 0.8736.

The F-1 scores, which balance precision and recall, proved to be similar across models. The ResNet50 and ResNet101 encoders achieved almost identical scores (0.8858 and 0.8860, respectively), whereas the ResNet152 model scored slightly lower (0.8799). The ResNet50 model had the lowest Dice Loss value (0.1149), followed by ResNet101 (0.1161) and ResNet152 (0.1233). Dice Loss is inversely related to segmentation accuracy, so lower values mean better performance. As such, it supports the idea that the ResNet50 encoder provides the best overall segmentation performance. In **Fig.4**, the evaluation of ResNet encoder performance is presented. In **Figs. 5-9**, all statistical metrics are exhibited.

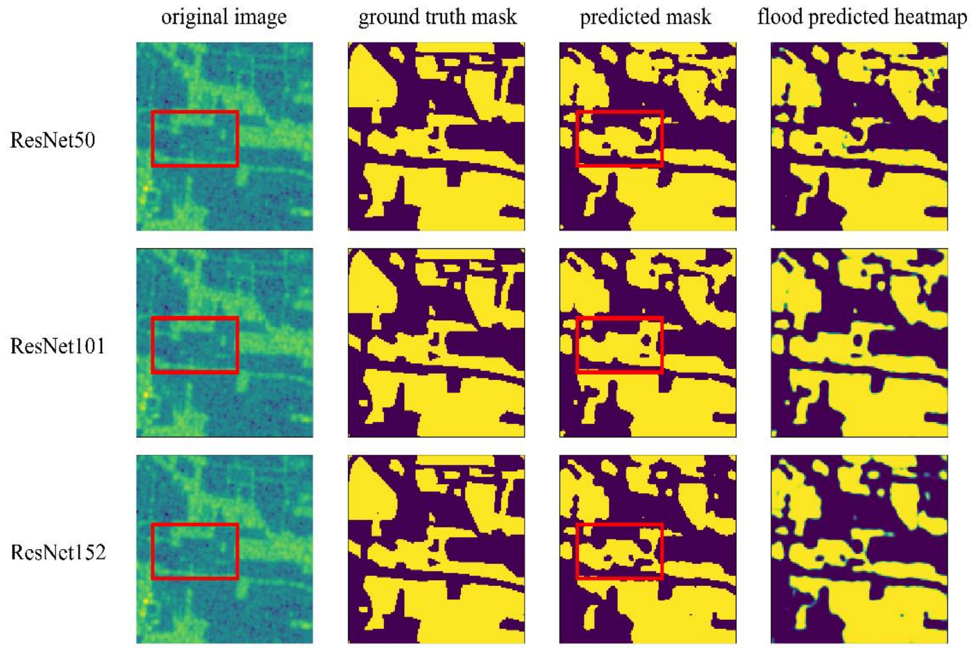


Fig. 4. Examples of U-Net model predictions using ResNet50, ResNet101, and ResNet152 encoders, showing that ResNet101 provides the most accurate predictions.

In **Fig. 4**, the U-Net model is illustrated detecting floods in Sentinel-1 SAR imagery using three encoder configurations: ResNet50, ResNet101, and ResNet152. Each panel displays the predicted flooded and non-flooded areas overlaid on the study area, making it straightforward to witness how well the segmentation works at different encoder depths. Results demonstrate that the ResNet101 encoder produces the most accurate and spatially consistent predictions. It draws flood boundaries with greater precision and fewer blunders than the ResNet50 and ResNet152 configurations. The visual observations correspond to the testing values in Table 2, which reveal that the ResNet101-based model had the best balance between recall and F1-score, supporting its overall superior performance. **Figs. 5-9** illustrate the training performance of U-Net models with ResNet50, ResNet101, and ResNet152 encoders. Overall, the curves show how the models converged and balanced segmentation trade-offs. In **Fig. 5**, all three models improved at making accurate predictions over time, with variability decreasing significantly after the first few iterations. ResNet101 encoder had the most stable and consistent accuracy trajectory. In **Fig. 6,** the precision curves exhibit greater variation than the accuracy curves, particularly during the early epochs to improved at making accurate predictions, while ResNet101 encoder exhibited the most consistent precision.

In contrast, **Fig. 7**, the recall curve was more stable for ResNet50, suggesting reliable detection of flooded areas, though with a higher false-positive rate, resulted in a lower overall recall compared to ResNet101 and ResNet152. **Fig. 8** presents the F1-score curves, which reflect the balance between precision and recall, even ResNet50 showed the lowest variance, ResNet101 achieved the highest overall F1-score. Finally, **Fig.9** presents the Dice Loss curves, with the ResNet50 encoder exhibiting the lowest and most stable values, indicating its ability to consistently delineate boundaries.

In **Table 3**, IoU performance for the four sample areas is presented. The ResNet101-based U-Net model achieved the highest average IoU (86.84%). It also achieved the highest IoU in samples 1, 3, and 4 (94.86%, 94.61%, and 65.26%, respectively). Although ResNet152 attained the best IoU in sample 2 (94.53%), ResNet101 demonstrated the highest and most consistent overall IoU performance.

ded. ResNet101

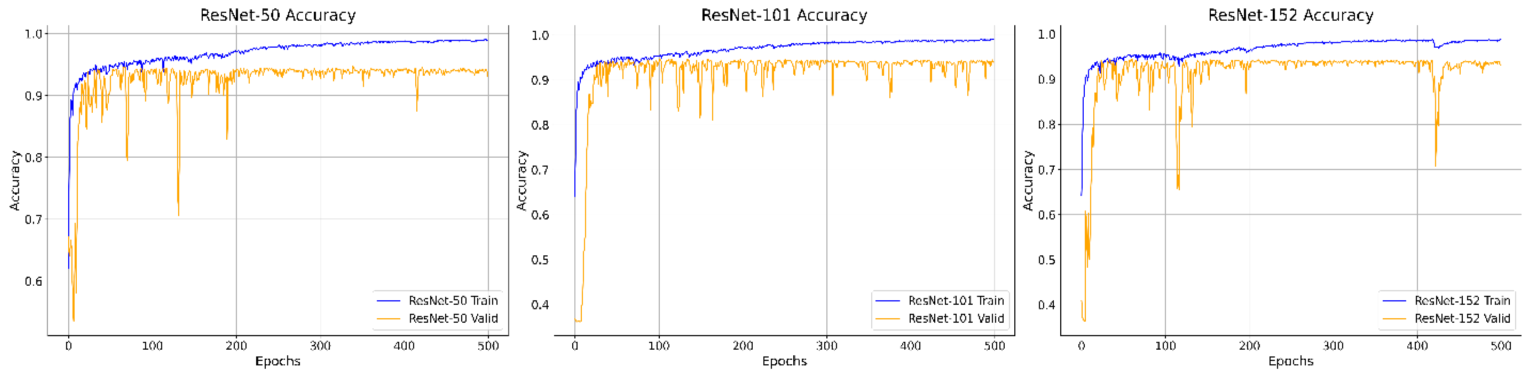
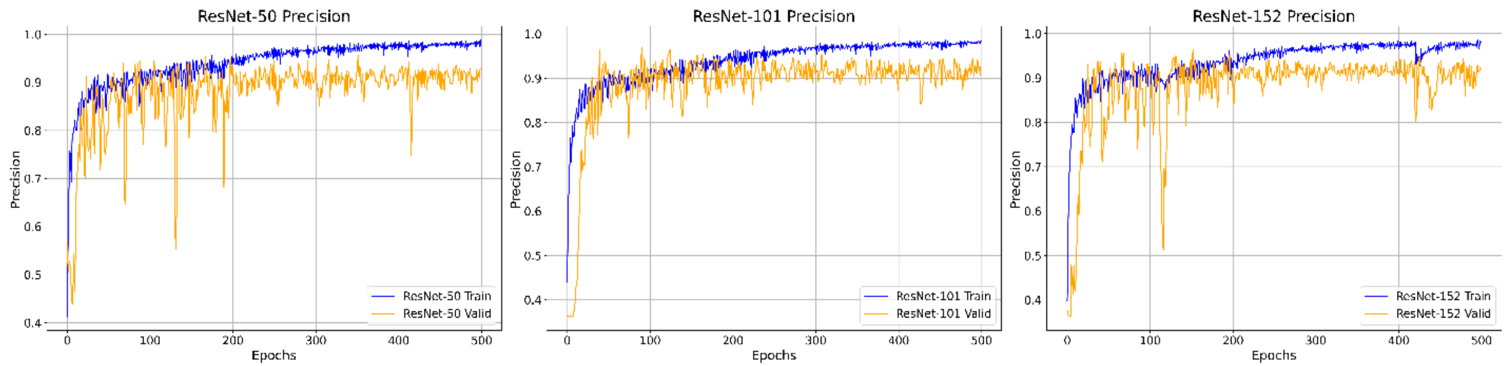


Fig. 5. Training and validation Accuracy of the U-Net model with ResNet50, ResNet101, and ResNet152 encoders. Accuracy values indicate the proportion of correctly predicted areas, both flooded and non-flooded, ResNet101 exhibited the most stable and consistent accuracy compared to ResNet50 and ResNet152.



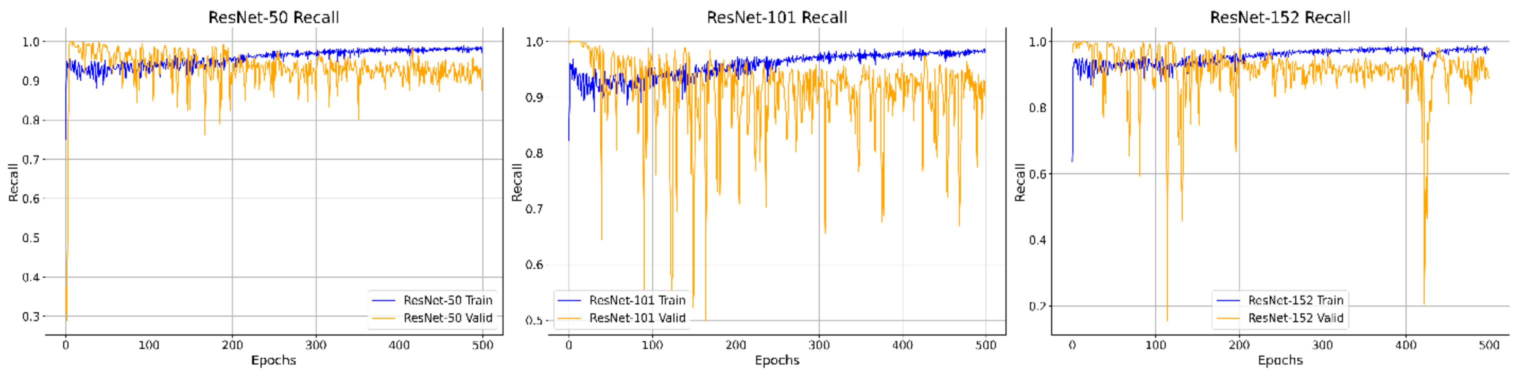


Fig. 7. Training and validation Recall of the U-Net model with ResNet50, ResNet101, and ResNet152 encoders. Recall values indicate the model' s ability to correctly identify flooded areas. While ResNet50 demonstrated more stable Recall, ResNet101 and ResNet152 encoders achieved higher overall Recall values.

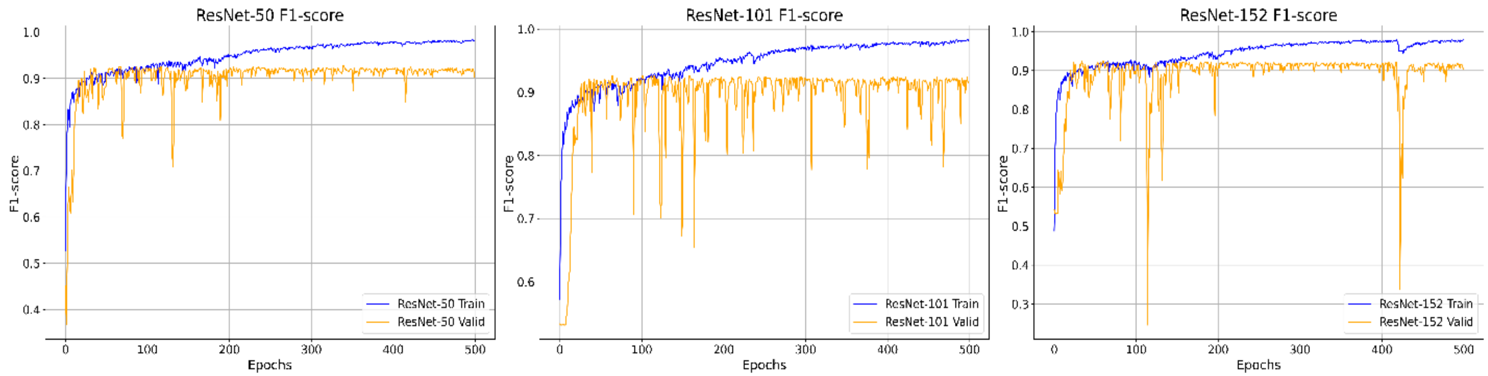


Fig. 8. Training and validation F1-score of the U-Net model with ResNet50, ResNet101, and ResNet152 encoders. F1-score indicates the balance between Precision and Recall. ResNet50 and ResNet101 achieved more stable and consistent performance compared to ResNet152.

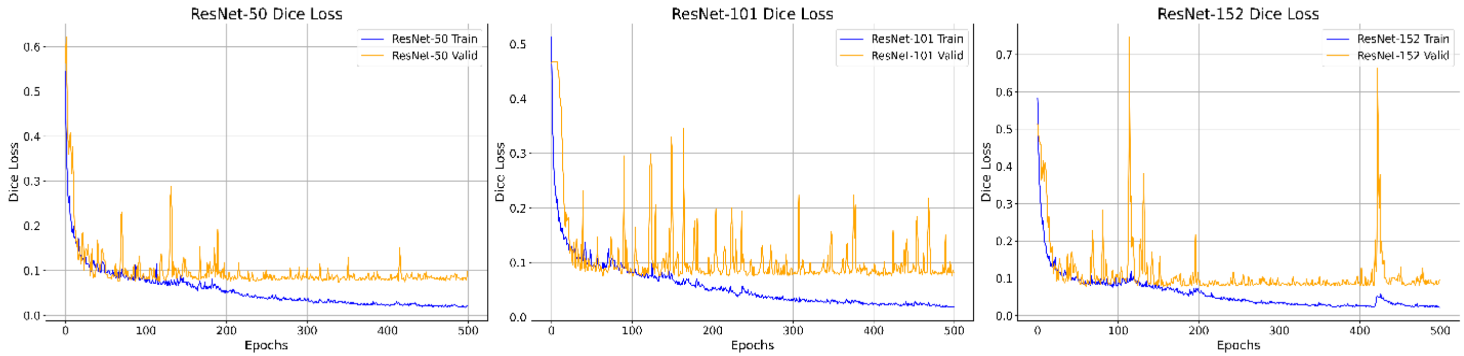


Fig. 9. Training and validation Dice Loss of the U-Net model with ResNet50, ResNet101, and ResNet152 encoders. Lower Dice Loss values indicate higher segmentation performance. ResNet50 and ResNet101 encoders showed more stable convergence, while ResNet152 exhibited greater loss variance.

Table 3.
IoU results for the four sample areas, showing the overlap between predicted and actual flood extents.

Sample	ResNet50	ResNet101	ResNet152
1	94.1431	94.8647	93.6507
2	92.7707	92.6230	94.5297
3	92.5100	94.6137	92.0071
4	59.5352	65.2567	53.8099
average (mean)	84.7398	86.8395	83.4994

Table 4.
U-Net model computational time using different ResNet encoders.

Encoder	Runtime
ResNet50	22 min 14 sec
ResNet101	24 min 04 sec
ResNet152	32 min 00 sec

Table 4 reports the time comsumption of each model training. The computational overhead of ResNet152 (32 min) was considerably higher than that of ResNet50 (22 min 14 sec) and ResNet101 (24 min 4 sec). This underscores the operational cost of employing excessively deep encoders in near-real-time applications.

4.2. Map of flooded areas in Phra Nakhon Si Ayutthaya Province

The U-Net model with the ResNet101 encoder generated flood maps for Phra Nakhon Si Ayutthaya Province on October 28, 2024. In our evaluation, it offered the best overall balance between precision, recall, and IoU. In **Fig.10**, the spatial distribution of the detected inundation is shown. Extensive flooding was observed in the eastern and northern regions of the province—specifically, the districts of Phak Hai, Bang Sai, Bang Ban, and Sena. These areas correspond to agricultural zones and low-lying floodplain regions, which act as natural retention basins when water levels rise above the capacity of the riverbanks.

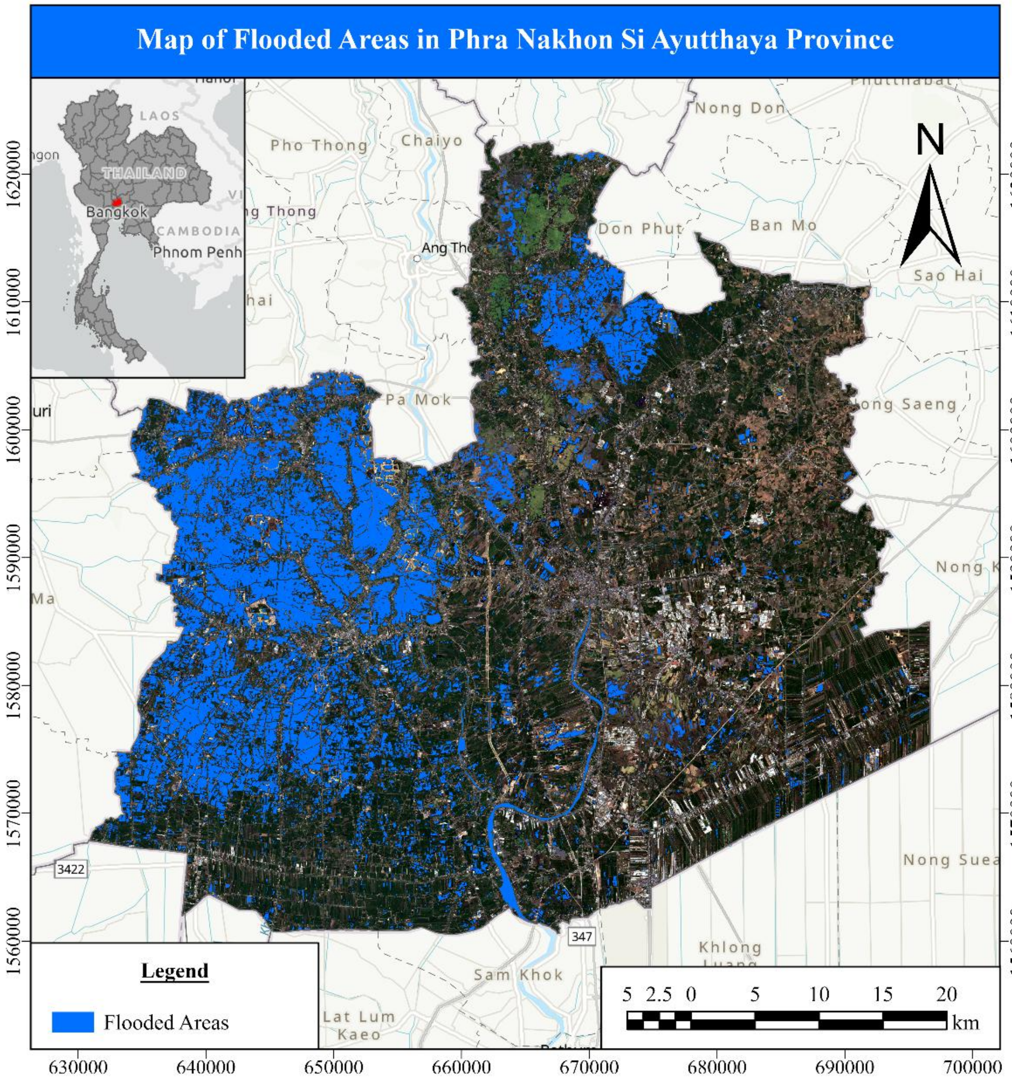


Fig. 10. Flood detection map in the study area on October 28, 2024. The blue areas represent regions identified as flooded based on U-Net model with ResNet101 encoder.

5. DISCUSSIONS

The flood detection model for Phra Nakhon Si Ayutthaya Province was constructed by integrating the ResNet architecture into U-Net, trained by SAR imagery classified into flooded and non-flooded areas. SAR data offers detailed spatial information and is unaffected by cloud cover, making it particularly suitable for flood mapping. Reducing the model's runtime by selecting only four study areas from the entire province contributed to improving detection efficiency. Overall, the U-Net model with ResNet101 encoder demonstrated the highest performance. The model performed well across all evaluation metrics, particularly with significantly higher recall and F1-score than ResNet50 and ResNet152, compared to other evaluation metrics. Additionally, the U-Net model with ResNet101 encoder achieved the highest average IoU and required only 2 min more runtime than ResNet50. The results from experiments with different ResNet encoders indicate that, although the ResNet50-based model benefits from the shortest runtime, its relatively shallow architecture may limit its ability to fully capture the data in certain cases. On the other hand, ResNet152, the deepest network of the three networks, may make the model unnecessarily complex and encounter overfitting problems as well as requiring the longest runtime, which can be a problem when dealing with large datasets or time-sensitive tasks. These results suggest that deeper encoders, such as ResNet101 and ResNet152, may be more effective at capturing flooded areas comprehensively, but this advantage can come at the cost of lower accuracy, resulting in more false positives.

However, the U-Net model with the ResNet101 encoder may not always outperform ResNet50 or ResNet152, particularly when hyperparameter tuning—such as increasing the number of training epochs—is considered. Such tuning often entails higher computational cost and longer processing time. Therefore, the availability of computational resources, model efficiency, and runtime requirements should be considered when selecting an appropriate model. In this study, the model did not achieve perfect accuracy in flood detection (**Fig. 11**). To further enhance performance, future research could investigate advanced deep learning approaches, such as multi-scale DeepLab (Wu et al., 2022).

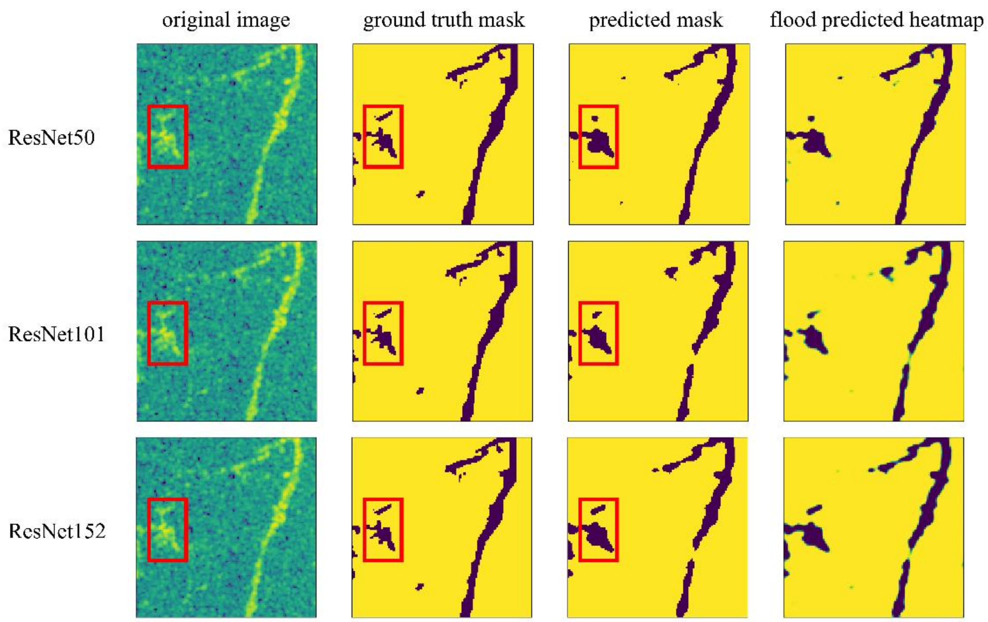


Fig. 11. Example of imperfect predictions using U-Net model with ResNet50, ResNet101, and ResNet152 encoders.

Additionally, using training data with higher resolution and accuracy could contribute to better model generalization (Ardohain & Fei, 2025). Incorporating multi-event and multi-temporal datasets can also enhance model performance (Magyari-Sáska et al., 2025) by capturing flood dynamics under diverse hydrological and environmental conditions. Such data could reduce event-specific bias and enhance the model's transferability to unseen flood scenarios or areas with different hydrological or topographic characteristics. (Wu et al., 2023). Future work should be integrated with temporal analysis, such as comparing SAR images from pre-flood and flood periods, to detect changes in surface water extent and identify only flooded areas (Yadav et al., 2022).

Moreover, the findings of the present study indicate that ResNet-101 often provides an optimal balance between network depth and generalization performance in image classification tasks. Similar patterns have been observed in previous studies (Pan et al., 2023; Shokati et al., 2025; Ait EI Asri et al., 2023) which found that ResNet101 achieved superior performance and often struck a good balance between feature extraction and the prevention of overfitting (He et al., 2016; Liu et al., 2020). compared with ResNet50 and ResNet152, despite not being the deepest architecture. These results highlight that a more complex model does not guarantee better performance (Pan et al., 2023) and may not work as well for segmentation tasks with SAR data (Pech-May et al., 2023). Although ResNet50 demonstrated the lowest training loss among the tested architectures, its performance on the validation set was markedly lower than that of ResNet101. This outcome suggests that ResNet50 being relatively compact and shallow was prone to overfitting and incapable of capturing complex and subtle patterns (Pan et al., 2023; Li et al.,2022). Taken together, these findings provide further evidence that ResNet101 offers a robust and efficient architecture for image-based detection tasks. During prediction testing, model performance was evaluated using IoU. ResNet101 achieved outstanding performance, with an IoU of 86.84%, demonstrating superior predictive capability. Such a model provided the best results for the samples within area boundaries one, three, and four (see the red boundaries in **Fig.2**). In contrast, the sample in boundary two is aligned with the ResNet50 encoder.

However, based on model evaluation, the ResNet152 encoder achieved the highest recall (0.9148) but underperformed in precision (0.8476) and overall IoU. This issue reflected a tendency to over-segment flooded areas and introduce false positives. This finding aligns with previous studies, which emphasize that very deep encoders can increase model complexity and training time without proportionally improving segmentation quality (Bonafilia et al., 2020; Liu et al., 2020). ResNet152 also required the longest computational time compared to the other encoders (**Table 4**), highlighting the trade-off between encoder depth and computational efficiency. Deeper networks introduce additional overhead without necessarily improving segmentation performance.

This study demonstrates that our flood detection model is capable of accurately delineating flood extents. The generated flood maps can be integrated into existing disaster management frameworks in Thailand, such as Thailand Incident Command System of the Department of Disaster Prevention and Mitigation (DDPM). Such integration has the potential to support spatial decision-making and strengthen operational response during flood events. To enable the developed model to be effectively deployed in real-world scenarios, improvements in computational infrastructure, processing, and data storage capabilities are required. The effectual model includes the use of higher-performance servers or workstations, particularly for larger datasets, especially when extending analyses to a national scale. Furthermore, addressing computational latency is crucial to ensure timely decision-making during flood events. To address the limitations arising from computational time, adopting models with lower architecture complexity yet comparable performance, such as ResNet50, may provide a practical solution. This approach would help reduce computational latency and resource demand, thereby improving the feasibility of real-time deployment.

By systematically comparing three encoder depths, we demonstrated that the ResNet101-based U-Net represents the best choice for mapping floods in complex lowland areas such as Phra Nakhon Si Ayutthaya Province. These results address a key research gap by providing a robust and generalizable deep learning model capable of balancing accuracy, stability, and computational efficiency in flood detection. This study provides novel insights into optimizing U-Net architectures for flood mapping using SAR imagery through the evaluation of different encoder depths. The results are applicable to large-scale flood monitoring and the strategic planning of emergency response operations.

6. CONCLUSIONS

This study developed a U-Net-based flood detection framework for Phra Nakhon Si Ayutthaya Province, using Sentinel-1 SAR imagery combined with ResNet-50, ResNet-101, and ResNet-152 encoders. Results demonstrated that the ResNet-101 encoder offered the best balance between segmentation performance and computational efficiency, making it the optimal choice for operational flood mapping.

The flood maps constructed for the October 2024 event aligned with past flood patterns and successfully identified the province's most at-risk areas, underscoring the model's reliability for hazard assessment.

From a practical perspective, the proposed framework is scalable and provides an effective means for near real-time flood monitoring, thereby aiding disaster management. Future research should explore multi-temporal SAR data, integration with optical and LiDAR sources, and advanced architectures to further enhance generalization and boundary detection in complex environments.

ACKNOWLEDGEMENT

This research was supported by the Faculty of Liberal Arts, Thammasat University, Research Unit în Geospatial Applications (Capybara Geo Lab). It is aslo partailly supported by the Center of Excellence in Digital Earth and Emerging Technology (CoE: DEET), Thammasat University, Pathumthani, Thailand, 12120.

REFERENCES

Ait El Asri, S., Negabi, I., El Adib, S., & Raissouni, N. (2023). Enhancing building extraction from remote sensing images through UNet and transfer learning. *International Journal of Computers and Applications*, *45*(5), 413-419. https://doi.org/10.1080/1206212X.2023.2219117

Amitrano, D., Di Martino, G., Di Simone, A., & Imperatore, P. (2024). Flood Detection with SAR: A Review of Techniques and Datasets. *Remote Sensing*, 16(4), 656. https://doi.org/10.3390/rs16040656

Angkahad, T., Laosuwan, T., Sangpradid, S., Prasertsri, N., Uttaruk, Y., Phoophathong, T., & Nuchthapho, J. (2024). An Empirical Analysis of Above Ground Biomass and Carbon Sequestration Utilizing UAV Photogrammetry and Machine Learning Techniques. *IEEE Access*, 12, 186740-186752. https://doi.org/ 10.1109/ ACCESS.2024.3514074

Angkahad, T., Laosuwan, T., Sangpradid, S., Prasertsri, N., Uttaruk, Y., Phoophathong, T., & Nuchthapho, J. (2025). Developing a Drone-Based Machine Learning for Spatial Modeling and Analysis of Biomass and Carbon Sequestration in Forest Ecosystems. *Engineered Science*, 35, 1508. http://dx.doi.org/10.30919 /es1508

Ardohain, C., & Fei, S. (2025). The impacts of training data spatial resolution on deep learning in remote sensing. *Science of Remote Sensing*, 11, 100185. https://doi.org/10.1016/j.srs.2024.100185

Bentivoglio, R., Isufi, E., Jonkman, S. N., & Taormina, R. (2022). Deep learning methods for flood mapping: a review of existing applications and future research directions. *Hydrology and Earth System Sciences Discussions*, 26(16), 1-50. https://doi.org/10.5194/hess-26-4345-2022

Birkmann, J., Buckle, P., Jaeger, J., Pelling, M., Setiadi, N., Garschagen, M., Fernando, N., & Kropp, J. (2010). Extreme events and disasters: a window of opportunity for change? Analysis of organizational, institutional and political changes, formal and informal responses after mega-disasters. *Natural Hazards*, 55, 637-655. https://doi.org/10.1007/s11069-008-9319-2

Bonafilia, D., Tellman, B., Anderson, T., & Issenberg, E. (2020). Sen1Floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 210-211. https://doi.org/10.1109/CVPRW50498.2020.00113

Das, S., Swain, M. K., Nayak, G. K., Saxena, S., & Satpathy, S. C. (2022). Effect of learning parameters on the performance of U-Net Model in segmentation of Brain tumor. *Multimedia tools and applications*, 81(24), 34717-34735. https://doi.org/10.1007/s11042-021-11273-5

Dhanabalan, S. P., Abdul Rahaman, S., & Jegankumar, R. (2021). Flood monitoring using Sentinel-1 sar data: A case study based on an event of 2018 and 2019 Southern part of Kerala. The International Archives of the Photogrammetry. *Remote Sensing and Spatial Information Sciences*, 44, 37-41. https://doi.org/10.5194/isprs-archives-XLIV-M-3-2021-37-2021

Fakhri, F., & Gkanatsios, I. (2025). Quantitative evaluation of flood extent detection using attention U-Net case studies from Eastern South Wales Australia in March 2021 and July 2022. *Scientific Reports*, 15(1), 12377. https://doi.org/10.1038/s41598-025-92734-x

Gale, E. L., & Saunders, M. A. (2013). The 2011 Thailand flood: climate causes and return periods. *Weather*, 68(9), 233-237. https://doi.org/10.1002/wea.2133

Ghosh, B., Garg, S., Motagh, M., & Martinis, S. (2024). Automatic flood detection from Sentinel-1 data using a nested UNet model and a NASA benchmark dataset. PFG–Journal of Photogrammetry. *Remote Sensing and Geoinformation Science*, 92(1), 1-18. https://doi.org/10.1007/s41064-024-00275-1

Ghozali, A., Ariyaningsih, Sukmara, R. B., & Aulia, B. U. (2016). A comparative study of climate change mitigation and adaptation on flood management between Ayutthaya City (Thailand) and Samarinda City (Indonesia). *Procedia-Social and Behavioral Sciences*, 227, 424-429. https://doi.org/10.1016/j.sbspro.2016.06.096

Haidu I., El Orfi T., Magyari-Sáska Z., Lebaut S., El Gachi M. (2024), Modeling the Long-Term Variability in the Surfaces of Three Lakes in Morocco with Limited Remote Sensing Image Sources. *Remote Sensing*, 16 (17), 3133. https://doi.org/10.3390/rs16173133

Hassan, S. M., & Maji, A. K. (2024). Pest Identification based on fusion of Self-Attention with ResNet. *IEEE Access*, 12, 6036-6050. https://doi.org/10.1109/ACCESS.2024.3351003

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., & Kanae, S. (2013). Global flood risk under climate change. *Nature climate change*, 3(9), 816-821. https://doi.org/10.1038/nclimate1911

Intarat, K., Nuengjumnong, N., Sae-Jung, J., Jangsawang, W., Yoomee, P., & Panboonyuen, T. (2024). Deep Residual Neural Networks with Self-Attention for Landslide Susceptibility Mapping in Uttaradit Province, Thailand. *2024 Geoinformatics for Spatial-Infrastructure Development in Earth and Allied Sciences (GIS-IDEAS)*, IEEE. pp. 1-6. https://doi.org/10.1109/GIS-IDEAS63212.2024.10990879

Jamali, A., Roy, S. K., Beni, L. H., Pradhan, B., Li, J., & Ghamisi, P. (2024). Residual wave vision U-Net for flood mapping using dual polarization Sentinel-1 SAR imagery. *International Journal of Applied Earth Observation and Geoinformation*, 127, 103662. https://doi.org/10.1016/j.jag.2024.103662

Konapala, G., Kumar, S. V., & Ahmad, S. K. (2021). Exploring Sentinel-1 and Sentinel-2 diversity for flood inundation mapping using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 180, 163-173. https://doi.org/10.1016/j.isprsjprs.2021.08.016

Konovalenko, I., Maruschak, P., Brezinová, J., Prentkovskis, O., & Brezina, J. (2022). Research of U-Net-based CNN architectures for metal surface defect detection. *Machines*, 10(5), 327. https://doi.org/10.3390/machines10050327

Lee, Y., Sim, W., Park, J., & Lee, J. (2022). Evaluation of hyperparameter combinations of the U-net model for land cover classification. *Forests*, 13(11), 1813. https://doi.org/10.3390/f13111813

Li, W., Xiao, Z., Liu, J., Feng, J., Zhu, D., Liao, J., Yu, W., Qian, B., Chen, X., Fang, Y., & Li, S. (2023). Deep learning-assisted knee osteoarthritis automatic grading on plain radiographs: the value of multiview X-ray images and prior knowledge. *Quantitative Imaging in Medicine and Surgery*, 13(6), 3587. https://doi.org/10.21037/qims-22-1250

Liu, J., Liu, K., & Wang, M. (2023). A residual neural network integrated with a hydrological model for global flood susceptibility mapping based on remote sensing datasets. *Remote Sensing*, 15(9), 2447. https://doi.org/10.3390/rs15092447

Liu, Z., Chen, B., & Zhang, A. (2020). Building segmentation from satellite imagery using U-Net with ResNet encoder. *2020 5th International Conference on Mechanical*, Control and Computer Engineering *(ICMCCE)*, IEEE. pp. 1967-1971. https://doi.org/10.1109/ICMCCE51767.2020.00431

Liu, Z., Coleman, N., Patrascu, F. I., Yin, K., Li, X., & Mostafavi, A. (2025). Artificial intelligence for flood risk management: A comprehensive state-of-the-art review and future directions. *International Journal of Disaster Risk Reduction*, 117, 105110. https://doi.org/10.1016/j.ijdrr.2024.105110

Loc, H. H., Park, E., Chitwatkulsiri, D., Lim, J., Yun, S. H., Maneechot, L., & Phuong, D. M. (2020). Local rainfall or river overflow? Re-evaluating the cause of the Great 2011 Thailand flood. *Journal of Hydrology*, 589, 125368. https://doi.org/10.1016/j.jhydrol.2020.125368

Magyari-Sáska, Z., Haidu, I., & Magyari-Sáska, A. (2025). Experimental Comparative Study on Self-Imputation Methods and Their Quality Assessment for Monthly River Flow Data with Gaps: Case Study to Mures River. *Applied Sciences*, 15(3), 1242. https://doi.org/10.3390/app15031242

Melgar-García, L., Martínez-Álvarez, F., Bui, D. T., & Troncoso, A. (2023). A novel semantic segmentation approach based on U-Net, WU-Net, and U-Net++ deep learning for predicting areas sensitive to pluvial flood at tropical area. *International Journal of Digital Earth*, 16(1), 3661-3679. https://doi.org/10.1080/17538947.2023.2252401

Misra, A., White, K., Nsutezo, S. F., Straka III, W., & Lavista, J. (2025). Mapping global floods with 10 years of satellite radar data. *Nature Communications*, 16(1), 5762. https://doi.org/10.1038/s41467-025-60973-1

Mosinska, A., Marquez-Neila, P., Koziński, M., & Fua, P. (2018). Beyond the pixel-wise loss for topology-aware delineation. *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3136-3145. https://doi.org/10.1109/CVPR.2018.00331

Mukherjee, A., Nandurbarkar, D., Yadav, H., & Vasa, J. (2024, August). Improvement in Image Segmentation for Flood Images Using UNet and U2Net. *International Conference on ICT for Sustainable Development*, Singapore: Springer Nature Singapore. pp. 321-331. https://doi.org/10.1007/978-981-97-8526-1_25

Munpa, P., Kittipongvises, S., Phetrak, A., Sirichokchatchawan, W., Taneepanichskul, N., Lohwacharin, J., & Polprasert, C. (2022). Climatic and hydrological factors affecting the assessment of flood hazards and resilience using modified UNDRR indicators: Ayutthaya, Thailand. *Water*, 14(10), 1603. https://doi.org/10.3390/w14101603

Muszynski, M., Hölzer, T., Weiss, J., Fraccaro, P., Zortea, M., & Brunschwiler, T. (2022). Flood event detection from sentinel 1 and sentinel 2 data: Does land use matter for performance of U-net based flood segmenters?. *2022 IEEE International Conference on Big Data (Big Data)*, IEEE. pp. 4860-4867. https://doi.org/10.1109/BigData55660.2022.10020911

Pan, Y. (2023). Effect of the deep residual networks at different depths on expression recognition. *Proceedings of the 5th International Conference on Computing and Data Science*. pp. 128-133. https://doi.org/10.54254/2755-2721/21/20231131

Pech-May, F., Aquino-Santos, R., & Delgadillo-Partida, J. (2023). Sentinel-1 SAR images and deep learning for water body mapping. *Remote Sensing*, 15(12), 3009. https://doi.org/10.3390/rs15123009

Pech-May, F., Aquino-Santos, R., Álvarez-Cárdenas, O., Lozoya Arandia, J., & Rios-Toledo, G. (2024). Segmentation and visualization of flooded areas through sentinel-1 images and u-net. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 8996-9008. https://doi.org/10.1109/JSTARS.2024.3387452

Ponmani, E., & Saravanan, P. (2021). RETRACTED ARTICLE: Image denoising and despeckling methods for SAR images to improve image enhancement performance: a survey. *Multimedia Tools and Applications*, 80(17), 26547-26569. https://doi.org/10.1007/s11042-021-10871-7

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer international publishing. pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

Safarov, F., Temurbek, K., Jamoljon, D., Temur, O., Chedjou, J. C., Abdusalomov, A. B., & Cho, Y.-I. (2022). Improved agricultural field segmentation in satellite imagery using TL-ResUNet architecture. *Sensors*, 22(24), 9784. https://doi.org/10.3390/s22249784

Sazara, C., Cetin, M., & Iftekharuddin, K. M. (2019). Detecting floodwater on roadways from image data with handcrafted features and deep transfer learning. *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, IEEE. pp. 804-809. https://doi.org/10.1109/ITSC.2019.8917368

Shokati, H., Seufferheld, K. D., Fiener, P., & Scholten, T. (2025). Rapid Flood Mapping from Aerial Imagery Using Fine-Tuned SAM and ResNet-Backboned U-Net. *EGUsphere*, 1-20. https://doi.org/10.5194/egusphere-2025-3146

Shoko, C., & Dube, T. (2024). A review of remote sensing of flood monitoring and assessment in southern Africa. *Physics and Chemistry of the Earth*, *Parts A/B/C*, 136, 103796. https://doi.org/10.1016/j.pce.2024.103796

Sibandze, P., Kalumba, A. M., Aljaddani, A. H., Zhou, L., & Afuye, G. A. (2025). Geospatial Mapping and Meteorological Flood Risk Assessment: A Global Research Trend Analysis. *Environmental Management*, 75, 137–154. https://doi.org/10.1007/s00267-024-02059-0

Torti, J. (2012). Floods in Southeast Asia: A health priority. *Journal of global health*, 2(2), 020304. https://doi.org/10.7189/jogh.02.020304

Wang, F., & Feng, X. (2025). Flood change detection model based on an improved U-net network and multi-head attention mechanism. *Scientific Reports*, 15(1), 3295. https://doi.org/10.1038/s41598-025-87851-6

Wang, L., Cui, S., Li, Y., Huang, H., Manandhar, B., Nitivattananone, V., Fang, X., & Huang, W. (2022). A review of the flood management: from flood control to flood resilience. *Heliyon*, 8(11). https://doi.org/10.1016/j.heliyon.2022.e11763

Wu, H., Song, H., Huang, J., Zhong, H., Zhan, R., Teng, X., Qiu, Z., He, M., & Cao, J. (2022). Flood detection in dual-polarization SAR images based on multi-scale deeplab model. *Remote Sensing*, 14(20), 5181. https://doi.org/10.3390/rs14205181

Wu, X., Zhang, Z., Xiong, S., Zhang, W., Tang, J., Li, Z., An, B., & Li, R. (2023). A near-real-time flood detection method based on deep learning and SAR images. *Remote Sensing*, 15(8), 2046. https://doi.org/10.3390/rs15082046

Yadav, R., Nascetti, A., & Ban, Y. (2022). Attentive dual stream siamese u-net for flood detection on multi-temporal sentinel-1 data. *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, IEEE. pp. 5222-5225. https://doi.org/10.1109/IGARSS46834.2022.9883132

Yang, D., Martinez, C., Visuña, L., Khandhar, H., Bhatt, C., & Carretero, J. (2021). Detection and analysis of COVID-19 in medical images using deep learning techniques. *Scientific Reports*, 11(1), 19638. https://doi.org/10.1038/s41598-021-99015-3

Yang, L., Zhong, J., Zhang, Y., Bai, S., Li, G., Yang, Y., & Zhang, J. (2022). An improving faster-RCNN with multi-attention ResNet for small target detection in intelligent autonomous transport with 6G. *IEEE Transactions on Intelligent Transportation Systems*, 24(7), 7717-7725. https://doi.org/10.1109/TITS.2022.3193909

Zhang, M., Chen, F., Liang, D., Tian, B., & Yang, A. (2020). Use of Sentinel-1 GRD SAR images to delineate flood extent in Pakistan. *Sustainability*, 12(14), 5784. https://doi.org/10.3390/su12145784

Zhao, J., Li, Y., Matgen, P., Pelich, R., Hostache, R., Wagner, W., & Chini, M. (2022). Urban-aware u-net for large-scale urban flood mapping using multitemporal sentinel-1 intensity and interferometric coherence. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-21. https://doi.org/10.1109/TGRS.2022.3199036

Zhong, Z., Li, J., Ma, L., Jiang, H., & Zhao, H. (2017). Deep residual networks for hyperspectral image classification. *2017 IEEE international geoscience and remote sensing symposium (IGARSS)*, IEEE. pp. 1824-1827. https://doi.org/10.1109/IGARSS.2017.8127330

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. *International workshop on deep learning in medical image analysis*, Springer International Publishing. pp. 3-11. https://doi.org/10.1007/978-3-030-00889-5_1

Geographia Technica, Vol. 21, Special Issue: Artificial Intelligence Applications in Geography, 2026, pp 19 to 34

PROJECTING LULC CHANGE (2025–2039) AND MODELING LANDSLIDE SUSCEPTIBILITY USING CELLULAR AUTOMATA–MARKOV CHAIN (CA–MC) IN THE KAMBANG WATERSHED, WEST SUMATRA, INDONESIA

***Triyatno TRIYATNO* [1*], *Lailatur RAHMI* [1], *Syafri ANWAR* [1], *Beni AULIA* [1] *and Sumayyah Aimi Mohd NAJIB* [2]**

DOI: 10.21163/ GT_2026.212.01

ABSTRACT

This study aims to map and quantify Land Use/Land Cover (LULC) change across 2025, 2029, and 2039 to develop a Frequency Ratio (FR) based landslide hazard model, to integrate LULC projections with key environmental parameters, and to delineate priority hotspot areas for mitigation based on scenario-specific susceptibility patterns. This study applies a quantitative, spatial approach to examine how LULC change reshapes landslide risk. Supervised classification was performed to produce the 2025 LULC map, which serves as the base year for projection. Future LULC (2029, 2039) was simulated using Cellular Automata–Markov Chain (CA–MC) formulations. A Random Forest (RF) model was used to predict Landslide Potential (LP) based on environmental variables and distance to mapped landslides. The results show primary forest is projected to decline steadily across the 2025, 2029, and 2039 years. The LULC maps achieve overall accuracies of 0.92 (2025), 0.88 (2029), and 0.86 (2039). The high-hazard class increases from 4,273.29 ha (8.90%) in 2025 to 4,782.06 ha (9.96%) in 2029 (+508.77 ha; +1.06 percentage points), and to 5,041.89 ha (10.50%) in 2039 (+259.83 ha; +0.54 percentage points). Combined, the moderate + high categories grow from 10,271.25 ha (21.39%) in 2025 to 15,590.11 ha (32.47%) in 2029 (+5,318.86 ha; +11.08 percentage points), and to 16,155.00 ha (33.65%) in 2039 (+564.89 ha; +1.18 percentage points). Landslide-potential models yield Area Under the Curve (AUC) of 0.95 (2025), 0.87 (2029), and 0.84 (2039). The results indicate a clear and coherent trajectory linking LULC change to evolving landslide susceptibility in the Kambang watershed. To mitigate future risks, priority actions should include protecting steep-slope forests, deploying deep-rooted revegetation and targeted slope stabilization, upgrading surface–subsurface drainage on hillslopes and along access roads, and incorporating the 2025/2029/2039 susceptibility maps into risk-sensitive spatial planning.

***Keywords**: LULC; Landslide; Susceptibility; Cellular Automata–Markov Chain (CA–MC); Random Forest; Watershed.*

1. INTRODUCTION

Recent global environmental change is accelerating transformations that reverberate through human systems and ecosystems, and Land Use/Land Cover (LULC) dynamics are central to these shifts (Naeem et al., 2025; Van Toor et al., 2024; Beroho et al., 2025; Ikhwan et al., 2025). Human-driven LULC alters surface energy and moisture exchanges, redirects hydrological responses, and can undermine ecological stability, which is increasingly documented through advances in classification, fraction estimation, and spatiotemporal mapping (Patel et al., 2024; Febriandi et al., 2025).

----

[1]*Department of Geography, Universitas Negeri Padang, Padang, Indonesia. *Corresponding author: triyatno@fis.unp.ac.id (TT); lailaturrahmi@student.unp.ac.id (LR); syafri.anwar.fis@gmail.com (SA); beniaulia@fis.unp.ac.id (BA)*

[2]*Department of Geography and Environment, Faculty of Human Sciences, Sultan Idris Education University, Malaysia. sumayyah@fsk.upsi.edu.my (SAMN)*

The resulting disasters impose substantial costs, including damage to assets, interruptions to livelihoods, and, at worst, loss of life, which underscores the need to anticipate how today's LULC trajectories may shape tomorrow's hazard landscape (Chowdhury et al., 2024; Es-smairi et al., 2023).

Among the hazard types that are most sensitive to LULC and extreme rainfall, landslides are particularly prominent. Vegetation removal, slope modification, and altered drainage pathways can destabilize hillslopes, and the impacts often peak during wet seasons and in forested terrains where management choices strongly influence outcomes (Asada et al., 2024; Asmare, 2023). In many tropical settings, intense precipitation coincides with steep relief and rapid land transformation, which creates recurrent susceptibility hotspots that threaten communities and critical lifelines (Chowdhury et al., 2024; Es-smairi et al., 2023). These conditions motivate integrated assessments that connect evolving land mosaics to spatiotemporal changes in slope stability.

Situated immediately south of Padang City in Pesisir Selatan Regency, the Kambang River Basin exemplifies these challenges. Recurrent landslides disrupt mobility along the western corridor that links West Sumatra and Bengkulu, and they also result in fatalities and property losses in exposed settlements. The areas of Kambang Utara and Kambang Timur are frequently affected, which highlights the need for site-specific analysis and forward-looking planning that can guide both local preparedness and regional connectivity strategies. Because consequences in Kambang span household damages and interprovincial flows, the basin constitutes a priority landscape for integrated hazard assessment and targeted mitigation. A substantial body of literature addresses each component of this problem. On the LULC side, recent work advances prediction and monitoring through Cellular Automata–Markov Chain (CA–MC) and machine-learning pipelines, remote-sensing fraction estimation, and multi-scenario coupling with climate and hydrology models (Girma et al., 2022; Naeem et al., 2025; Beroho et al., 2025; Febriandi et al., 2025; Ikhwan et al., 2025). On the landslide side, studies compare and combine statistical and machine-learning approaches, including Frequency Ratio (FR). Shannon entropy, logistic regression, Random Forest (RF), and ensemble evidence frameworks to map susceptibility and evaluate model skill (Es-smairi et al., 2023; Chowdhury et al., 2024; Roy et al., 2023). Despite these advances, a notable gap persists because LUCC modeling and landslide mapping are often performed in parallel rather than explicitly linked so that projected LULC under future time years can drive dynamic susceptibility updates and reveal risk frontiers under plausible landscape trajectories (Triyatno et al., 2023).

This study responds to that gap by integrating forward LULC projections with landslide hazard modeling for the Kambang watershed. Multiple years, namely 2025, 2029, and 2039, capture near-term, medium-term, and longer-term trajectories that are consistent with planning cycles, while environmental variables represent controlling factors for slope stability and runoff. By coupling projected land transformations with hazard estimates, the analysis identifies where susceptibility may intensify, remain stable, or diminish, and this information supports the timing and prioritization of interventions such as vegetation management, slope stabilization, drainage improvements, and land-use regulation (Roy et al., 2023; Asada et al., 2024). The practical value is twofold, because communities in Kambang Utara and Kambang Timur can use forward-looking maps to prepare ahead of peak rainy seasons. Local as well as provincial authorities can target spatial planning, infrastructure maintenance, and investment in ways that steadily reduce exposure and vulnerability (Chowdhury et al., 2024; Es-smairi et al., 2023). Accordingly, this study aims to map and quantify LULC change across 2025–2029–2039, to develop an FR-based landslide hazard model, to integrate LULC projections with key environmental parameters, and to delineate priority hotspot areas for mitigation based on scenario-specific susceptibility patterns. The novelty of this study lies in the dynamic coupling of forward LULC projections generated by a CA–MC framework with landslide-susceptibility modeling (FR/RF) at matched planning years (2025, 2029, 2039). In contrast to studies that treat LUCC and landslide mapping separately or at a single year without feeding projected mosaics into hazard models, this approach explicitly links projected land-change trajectories to evolving hazard and validates the chain year by year.

2. STUDY AREA

The study was conducted in the Kambang watershed of Pesisir Selatan Regency, immediately south of Padang City (West Sumatra, Indonesia). The basin extends between approximately 100°41'00"–100°57'30" E and 1°47'00"–1°37'00" S, covering 48,014.55 ha. The watershed contains steep headwater hillslopes and lower-relief footslopes and valley floors where settlements are. Topographically, the basin is characterized by dissected terrain with steep to very steep slopes in the uplands and gentler surfaces in the downstream piedmont/valley areas. Elevation varies markedly from near sea level to ~2,020 MASL, reflecting strong relief that channels runoff and focuses mass-movement processes on hillslopes. The watershed includes settlements such as Kambang Utara and Kambang Timur and forms part of the western transport corridor connecting West Sumatra and Bengkulu. Recurrent landslides along steep, dissected hillslopes disrupt mobility and threaten lives and assets, making the Kambang basin a priority landscape for scenario-based hazard assessment and forward-looking mitigation. More details can be seen in **figure 1** below.

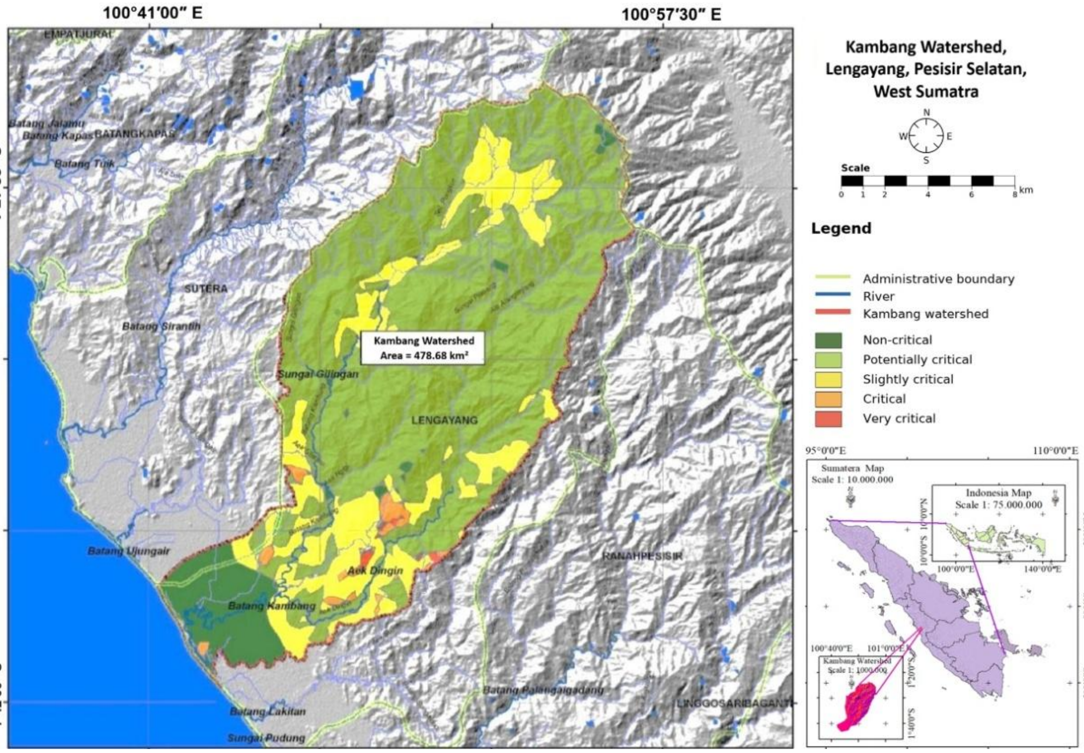


Fig. 1. Study area map of the Kambang Watershed showing administrative boundaries and rivers.

3. METHODS

This study applies a quantitative, spatial approach to examine how LULC change reshapes landslide risk (Febriandi et al., 2025). The work was carried out in the Kambang watershed, ~120 km from Padang City, West Sumatra. Quantitatively, class-level influences are estimated with FR using the landslide inventory and conditioning factors (slope, rainfall, geology, geomorphology, soil), susceptibility is evaluated with AUC, and LULC projections for 2025/2029/2039 are produced via CA–MC. Spatially, these statistical relationships are applied cell-by-cell to create a baseline surface, then re-run with each projected LULC to generate year-specific susceptibility maps and quantify the reallocation among low–medium–high classes. Risk reshaping is summarized by area changes (Δ) and hotspots where susceptibility increases especially where forest converts to oil palm/mixed orchard/dryland agriculture on steeper terrain. To maintain scope clarity, the FR-based Landslide Hazard Surface (LHS) (knowledge-driven) and the RF-based Landslide Potential (LP) model (data-driven) are treated as distinct products, and their agreement is explicitly compared.

3.1. Study setting and workflow

The study followed three stages. More details are explained as follows.

Pre-field. Compiled multi-temporal satellite imagery and ancillary data to build the LULC baseline and the environmental layers used in modeling. The image archive consisted of Landsat 5 (2000), Landsat 7 (2009), Landsat 8 (2019), and Landsat 9 (2025) scenes. Supervised classification was performed to produce the 2025 LULC map, which serves as the base year for projection. Supporting datasets included geology, geomorphology, soil, a Digital Elevation Model (DEM) (for slope, elevation, and contours), and rainfall. These layers were also prepared for FR analysis and the landslide hazard map. All rasters were co-registered and resampled to a common grid/resolution to avoid spatial misalignment in subsequent modeling.

Field. Documented landslide occurrences and collected inventory points and ground samples in areas affected by natural landslides. These data were used both to validate the FR-based landslide hazard and to train/validate the prediction model of LP by comparing the mapped hazard/potential with observed landslide points. To reduce positional bias, each inventory point was buffered by the median mapping uncertainty before sampling environmental covariates.

Post-field. With 1) projected near-term and longer-term LULC trajectories to 2029 and 2039 with a CA–MC model; 2) computed FR values for each conditioning factor and produced a LHS; 3) developed a RF prediction of LP using environmental variables and the landslide inventory; and 4) assessed accuracy of both LULC and landslide-potential outputs. Training/validation used spatial block cross-validation, and classification thresholds were selected on the precision–recall curve to handle class imbalance.

3.2. Data analysis

LULC change modeling (CA–MC)

Future LULC (2029, 2039) was simulated using CA–MC formulations (Badrakh et al., 2025; Beroho et al., 2025; Kubacka et al., 2025). A transition probability matrix P was estimated from historical LULC (Mondal et al., 2016; Murmu et al., 2025), and neighborhood effects were introduced via the cellular automata with formula.

$$\mathbf{S_{t+1} = f(S_t, N) S_{t+1} = P \,.\, S_t P = [p_{ij}]} \tag{1}$$

where $\mathbf{S}$ is set of LULC states, $\mathbf{N}$ neighborhood, and $\mathbf{p_{ij}}$ transition probability from class $\mathbf{i}$ **to** $\mathbf{j}$. Calibration used 2000→2009 and 2009→2019; validation used 2019→2025; Neighborhood and suitability constraints (slope, distance-to-roads/settlements) limited implausible transitions.

LULC accuracy assessment

Next for computed Overall Accuracy (OA), User's Accuracy (UA), Producer's Accuracy (PA), and Kappa with formula.

$$\mathbf{OA = \frac{\sum_i^r = 1^{D_{ii}}}{N}, UA_i = \frac{D_{ii}}{x_{+i}}, PA_i = \frac{D_{ii}}{x_{+i}},\ OA = \frac{N \sum_i^r = 1^{D_{ii}} - \sum_i^r = 1^{(x_i + x_{+i})}}{N^2 - \sum_i^r = 1^{(x_i + x_{+i})}}} \tag{2}$$

Validation samples were stratified per class (≥50 points/class) to avoid rare-class under-representation.

Landslide hazard (FR)

At stage quantified the contribution of slope, rainfall, geology, geomorphology, and soil using the FR method (Asmare, 2023; Huang et al., 2024; Zhou et al., 2021) with formula.

$$\mathbf{FR} = \frac{\text{(landslide pixels in class)/(pixels of class)}}{\text{all landslide pixels / all pixels}} \tag{3}$$

FR values >1 indicate positive association with landslides; FR <1 indicates negative association. The Landslide-Susceptibility Map (LSM) was computed using the classic bivariate formulation: for each cell, the landslide score equals the sum of the log-FR across criteria (Σ ln FR). No subjective weighting was applied. To avoid double-counting and leakage when projecting with CA–MC, LULC is not included inside the FR combination. LULC effects are applied by re-evaluating susceptibility under projected classes with the following formula.

$$\mathbf{LSM} = \mathbf{0.30\ FR_{geology}} + \mathbf{0.30\ FR_{geomorphology}}\mathbf{0.20\ FR_{soil}} + \mathbf{0.30\ FR_{geology}} + \mathbf{0.10\ FR_{rainfall}} + \mathbf{0.10\ LULC_n} \tag{4}$$

where $\mathbf{LULC_n}$ encodes projected LULC (2029 or 2039). See also Es-smairi et al., 2023; Ahmad et al., 2023 for related FR-based LSM applications.

Prediction model

RF was built on the landslide inventory and environmental predictors (Euclidean distance to landslide points, slope FR, soil FR, geology FR, geomorphology FR, elevation, contours, rainfall FR) following established practice (Talukdar et al., 2021; Shanley et al., 2021; Williams et al., 2025). To preserve methodological separation and prevent target leakage, RF inputs comprised slope, curvature, elevation, lithology/geomorphology/soil (one-hot), rainfall climatology, and distances to rivers/roads/faults, plus LULC class FR-derived layers were deliberately excluded. For interpretability, permutation and Gini importances were computed, and SHAP values were derived for the final RF, together with Partial Dependence (PD) and Accumulated Local Effects (ALE) plots for key drivers (slope, soil, geomorphology, distance-to-rivers/roads). RF offers strong performance with limited tuning, robustness to mixed/noisy predictors, and out-of-bag error for fast calibration, and RF is retained unless AUC-PR improves by ≥0.02 with boosting. Training/validation used spatial block cross-validation with class-weight balancing; decision thresholds were chosen at the F1/IoU-optimal point on the precision–recall curve.

4. RESULTS

4.1. LULC changes model

The LULC changes model shows projected in the Kambang watershed by integrating the current LULC distribution with observed transition trends. Class-level counts, areas, and percentages for each year (2025, 2029, 2039). More details can be seen in **table 1** below.

Primary forest is projected to decline steadily across the years 2025, 2029, and 2039. The area decreases from 36,742.23 ha (76.52%) in 2025 to 35,060.67 ha (73.02%) in 2029, resulting in a loss of 1,681.56 ha and a 3.50 percentage point decrease. By 2039, the primary forest is projected to be 33,310.80 ha (69.38%), indicating a further loss of 1,749.87 ha and a 3.64 percentage point decline relative to 2029. Secondary forest follows the same trend, declining from 4,070.07 ha (8.48%) in 2025 to 3,821.40 ha (7.96%) in 2029 and then to 2,858.04 ha (5.95%) in 2039, which corresponds to successive decreases of 248.67 ha (0.52 percentage points) and 963.36 ha (2.01 percentage points).

Oil palm shows the most pronounced gains. The area increases from 311.31 ha (0.65%) in 2025 to 1,347.03 ha (2.81%) in 2029, an increment of 1,035.72 ha and 2.16 percentage points, and then rises to 3,920.49 ha (8.17%) in 2039, adding 2,573.46 ha and 5.36 percentage points.

Table 1

LULC changes from 2025 to 2039.

LULC	LULC 2025			LULC 2029			LULC 2039		
	Count	Area (ha)	%	Count	Area (ha)	%	Count	Area (ha)	%
Primary forest	408,247	36,742.23	76.52	389,563	35,060.67	73.02	370,120	33,310.80	69.38
Secondary forest	45,223	4,070.07	8.48	42,460	3,821.40	7.96	31,756	2,858.04	5.95
Oil palm	3,459	311.31	0.65	14,967	1,347.03	2.81	43,561	3,920.49	8.17
Mangrove	48	4.32	0.01	48	4.32	0.01	48	4.32	0.01
Water bodies	2,750	247.50	0.52	2,750	247.50	0.52	2,750	247.50	0.52
Built-up land	8,885	799.65	1.67	8,888	799.92	1.67	8,905	801.45	1.67
Rice fields	24,710	2,223.90	4.63	24,710	2,223.90	4.63	24,710	2,223.90	4.63
Dryland agriculture	1,160	104.40	0.22	1,371	123.39	0.26	2,362	212.58	0.44
Bare land	4,464	401.76	0.84	4,436	399.24	0.83	3,741	336.69	0.70
Mixed orchard	34,549	3,109.41	6.48	44,302	3,987.18	8.30	45,542	4,098.78	8.54
TOTAL	533,495	48,014.55	100	533,495	48,014.55	100	533,495	48,014.55	100

Source: Data analysis, 2025.

Mixed orchard also grows, from 3,109.41 ha (6.48%) in 2025 to 3,987.18 ha (8.30%) in 2029 and 4,098.78 ha (8.54%) in 2039. These increases amount to 877.77 ha (1.82 percentage points) between 2025 and 2029 and 111.60 ha (0.24 percentage points) between 2029 and 2039. Dryland agriculture expands more modestly, from 104.40 ha (0.22%) in 2025 to 123.39 ha (0.26%) in 2029 and 212.58 ha (0.44%) in 2039, which equals gains of 18.99 ha (0.04 percentage points) and 89.19 ha (0.18 percentage points) across the two intervals. Built-up land shows negligible change (799.65 ha in 2025; 799.92 ha in 2029; 801.45 ha in 2039), with its share remaining about 1.67%. Rice fields are constant at 2,223.90 ha (4.63%) at all time points. Water bodies and mangrove areas are likewise stable at 247.50 ha (0.52%) and 4.32 ha (0.01%), respectively. Bare land declines gradually from 401.76 ha (0.84%) in 2025 to 399.24 ha (0.83%) in 2029 and 336.69 ha (0.70%) in 2039, which corresponds to decreases of 2.52 ha (0.01 percentage points) and 62.55 ha (0.13 percentage points).

Oil palm expansion concentrates on gentler terrain and valley floors where accessibility is higher. In contrast, conversions affecting forest classes are more common on dissected hillslopes, where mixed orchard and dryland agriculture extends into steeper gradients at elevations generally below 500 MASL. These spatial patterns indicate that continued forest loss on slopes is likely to reduce surface stability during peak rainy seasons. More details can be seen in **figure 2**.

The LULC maps achieve overall accuracies of 0.92 (2025), 0.88 (2029), and 0.86 (2039). Because all values exceed 0.85, the classifications meet a high-performance threshold and are suitable for use as inputs to the subsequent landslide-susceptibility modeling. **Figure 3** shows a persistent contraction of primary and secondary forests from 2025 to 2039, accompanied by rapid expansion of oil palm and a more gradual increase in mixed orchard and dryland agriculture, whereas rice fields, water bodies, and mangrove remain essentially unchanged. The spatial arrangement of these transitions, forest loss on slopes, and agricultural expansion on gentler terrain implies increasing exposure of unstable areas during peak rainy seasons. This interpretation is consistent with regional multi-criteria hazard zoning in the Tarusan watershed, where GIS mapping highlighted higher susceptibility along valley floors/low-lying zones under expanding agriculture and settlement, reinforcing the relevance of our valley-focused oil palm gains for future exposure (Barlian et al., 2025). Moreover, accurate terrain and exposure layers within a spatial framework materially improve spatial risk delineation for different hazard types with a couple of LULC projections with susceptibility modeling (Triyatno et al., 2024). In the study use the LULC projections are used as inputs to the hazard model to quantify where susceptibility is likely to intensify under the projected trajectories.

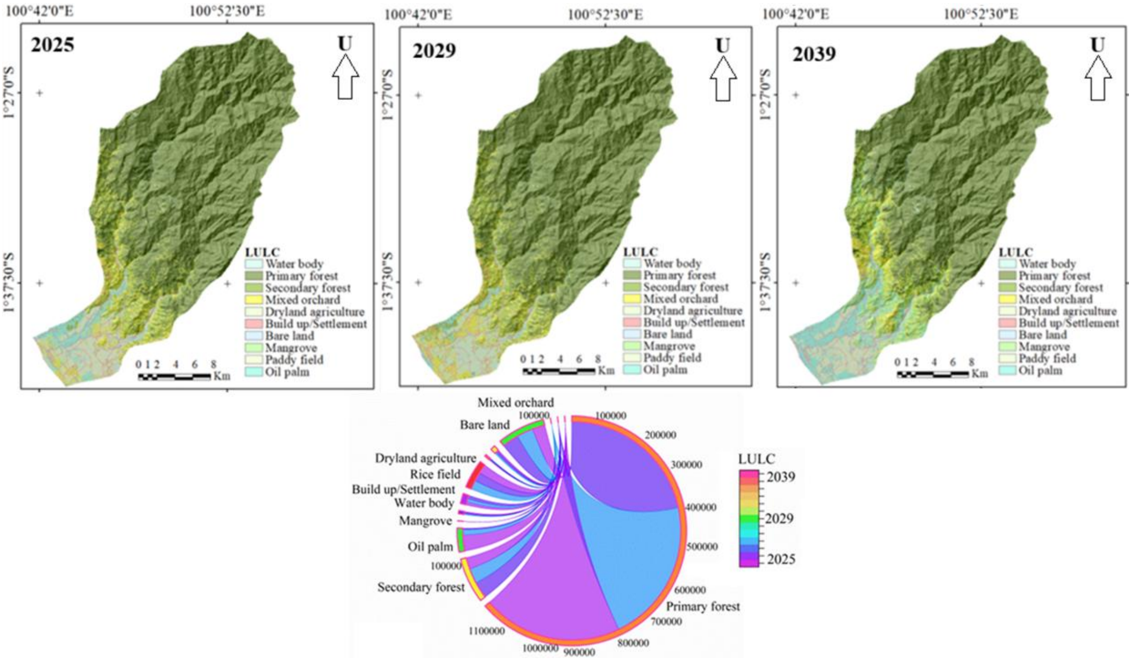


Fig. 2. Map of the spatial distribution of improved drinking water access before the disaster, showing the pipeline network (intake points, reservoirs, and transmission pipelines).

4.2. Landslide hazard

Landslide hazard indicates the potential of an area to experience landslides. In the Kambang watershed, it is governed by five conditioning factors, namely slope, rainfall, geology, geomorphology, and soil, whose class-level effects are summarized using the FR. By definition, FR > 1 shows a class that is more common within landslide locations than in the landscape overall (positive influence), whereas FR < 1 shows a class that is less associated with landslides (negative influence). More details can be seen in **figure 3** below.

As depicted in **figure 3**, FR ranges differ among factors: geology ≈ 0–1.50, geomorphology ≈ 0–1.59, slope ≈ 0–1.81, soil ≈ 0–1.81, and rainfall classes cluster around ≈ 0.61. These distributions indicate that slope and soil contain the strongest positively associated classes (up to ~1.81), followed by geomorphology (up to ~1.59) and geology (up to ~1.50). In contrast, the rainfall classes used here contribute a weaker standalone signal (≈0.61), suggesting that rainfall primarily operates as a trigger that amplifies instability where conditioning factors are already unfavorable. A high FR value in one class does not automatically translate into a proportionally high map-level hazard, because the effect also depends on that class's spatial prevalence (number of pixels) and its co-occurrence with other adverse classes. These patterns are consistent with prior evidence from Koto XI Tarusan, where slope-controlled morphometry dominated landslide impacts and lithology/landform acted as secondary controls (Triyatno et al., 2020).

In line with process-based findings in the Air Dingin Watershed in Padang where terrain steepness and potential failure conditions of soil erodibility, while rainfall primarily serves as energy input/trigger (Putra et al., 2025). Geological, geomorphological, and soil conditions also influence landslide occurrence in the study area. Weathered rock facilitates water infiltration, which mobilizes fine particles that accumulate on impermeable layers and form a slip plane, particularly across mountainous and hilly terrain with steep to very steep slopes. More details can be seen in **figure 4** below.

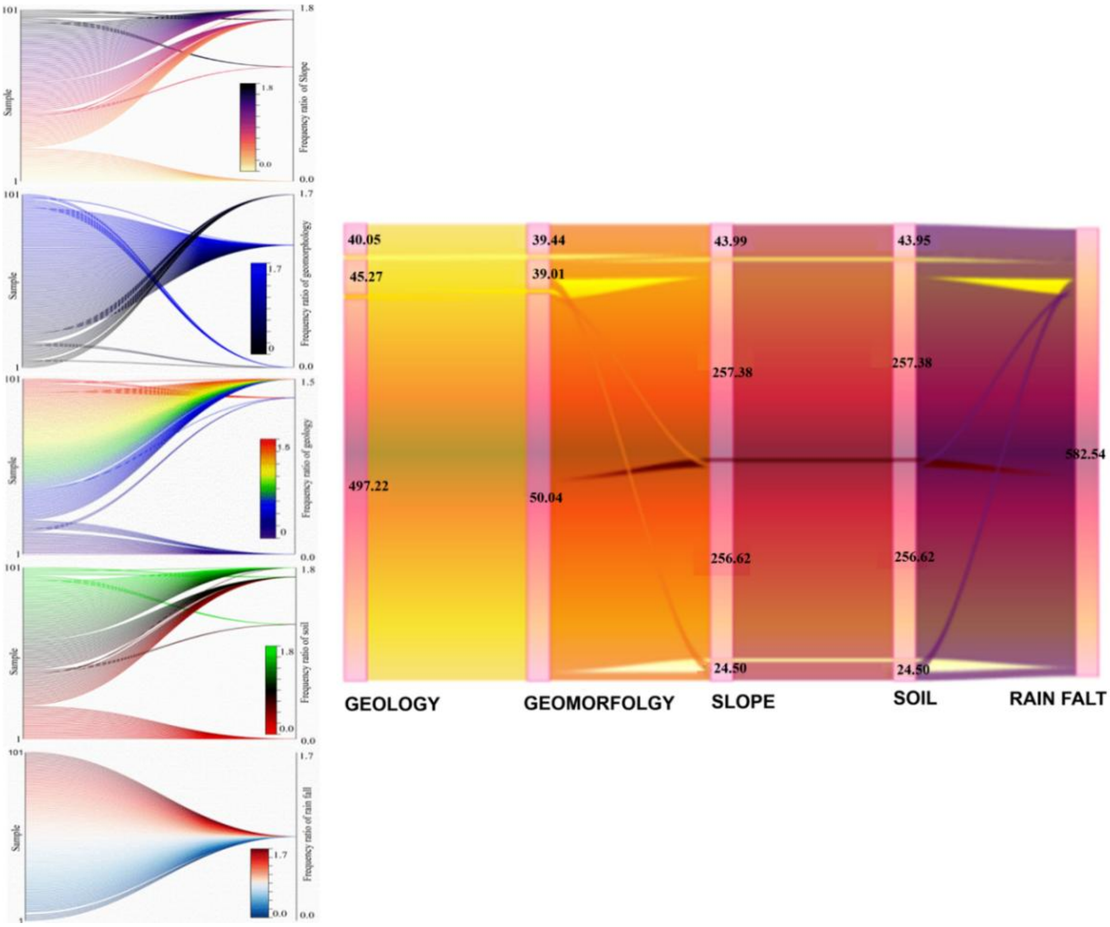


Fig. 3. FR values for each landslide determinant parameter (geology, geomorphology, slope, soil, and rainfall).

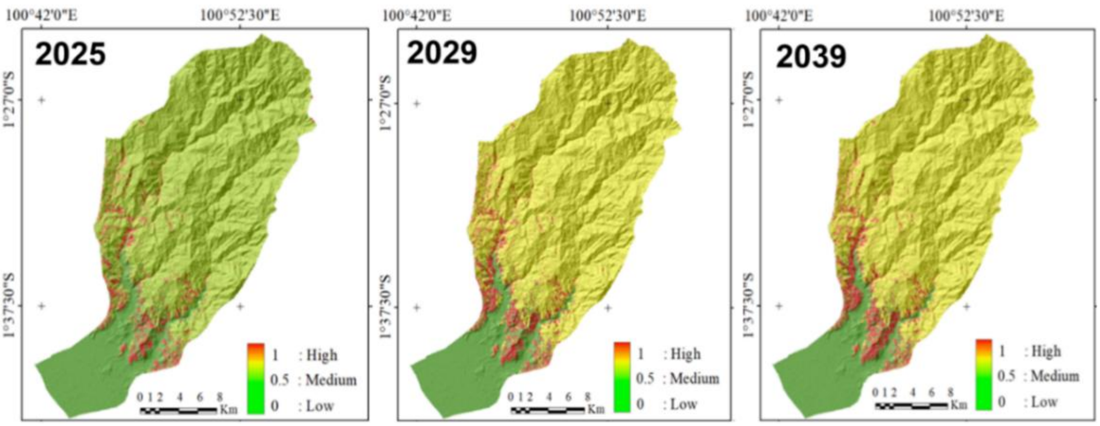


Fig. 4. Landslide hazard model for 2025, 2029, and 2039.

The map in **figure 4** above shows that the LULC conversion coincides with the expansion of landslide-prone areas, especially on steep to very steep slopes. The high-hazard class increases from 4,273.29 ha (8.90%) in 2025 to 4,782.06 ha (9.96%) in 2029 (+508.77 ha; +1.06 percentage points), and to 5,041.89 ha (10.50%) in 2039 (+259.83 ha; +0.54 percentage points). The medium-hazard class rises from 6,134.22 ha (12.78%) in 2025 to 12,820.50 ha (26.70%) in 2029, then moderates to 11,249.37 ha (23.43%) in 2039, remaining well above 2025 levels. Consequently, the low-hazard class shrinks from 37,607.04 ha (78.32%) to 30,411.99 ha (63.34%), before a slight rebound to 31,723.29 ha (66.07%) in 2039. These shifts are concentrated where forests are replaced by oil palm, mixed orchard, and dryland agriculture, which collectively reduce slope reinforcement and surface roughness. More details can be seen in **table 2** below.

Table 2

Changes in landslide hazard areas, 2025–2039.

Landslide class	2025 – Area			2029 – Area			2039 – Area		
	Pixels	Area (ha)	% Area	Pixels	Area (ha)	% Area	Pixels	Area (ha)	% Area
Low	417,856	37,607.04	78.32	337,911	30,411.99	63.34	352,481	31,723.29	66.07
Medium	68,158	6,134.22	12.78	142,45	12,820.50	26.70	124,993	11,249.37	23.43
High	47,481	4,273.29	8.90	53,134	4,782.06	9.96	56,021	5,041.89	10.50
TOTAL	533,495	48,014.55	100.00	533,495	48,014.55	100.00	533,495	48,014.55	100.00

Source: Data analysis, 2025.

Table 2 shows landslide-hazard classes in the Kambang watershed across three years. The modeled extent remains constant at 48,014.55 ha (equivalent to 533,495 pixels) in all years, so the shifts reflect reallocation among classes rather than changes in total area. In 2025, the low class dominates (37,607.04 ha; 78.32%), followed by medium (6,134.22 ha; 12.78%) and high (4,273.29 ha; 8.90%). By 2029, the low class declines sharply to 30,411.99 ha (63.34%), while the medium class increases to 12,820.50 ha (26.70%) and the high class rises to 4,782.06 ha (9.96%). In 2039, the low class recovers slightly to 31,723.29 ha (66.07%), the medium class decreases to 11,249.37 ha (23.43%) yet remains above 2025, and the high class continues to grow to 5,041.89 ha (10.50%). Viewed as period-to-period changes, the high class increases by +508.77 ha (+1.06 percentage points) from 2025 to 2029 and by +259.83 ha (+0.54 percentage points) from 2029 to 2039, yielding a cumulative +768.60 ha (+1.60 percentage points) over 2025–2039. The medium class jumps by +6,686.28 ha (+13.92 percentage points) in 2025–2029, then declines by −1,571.13 ha (−3.27 percentage points) in 2029–2039, remaining +5,115.15 ha (+10.65 percentage points) above its 2025 level. Consequently, the lowclass contracts by −7,195.05 ha (−14.98 percentage points) in 2025–2029 and then rebounds by +1,311.30 ha (+2.73 percentage points) in 2029–2039; overall, it is still 5,883.75 ha smaller than in 2025 (−12.25 percentage points). These trends are consistent with the preceding LULC modeling.

Where conversion of primary/secondary forests to oil palm, mixed orchard, and dryland agriculture tends to occur on steep to very steep slopes, which reduces root reinforcement and surface roughness. In contrast, the low class persists mainly on gently sloping footslopes and valley floors. The results indicate that mitigation priorities should focus on expanding medium–high pockets through deep-rooted revegetation, improved surface and subsurface drainage, and stricter land-use controls in newly converting zones.

4.3 Prediction model for potential landslides

The prediction of LP in the Kambang watershed is controlled by a set of environmental parameters, including Euclidean distance to mapped landslide points as well as the spatial distribution of past landslide points. These layers provide the conditioning context that governs where instability is likely to emerge. More details can be seen in **figure 5** below.

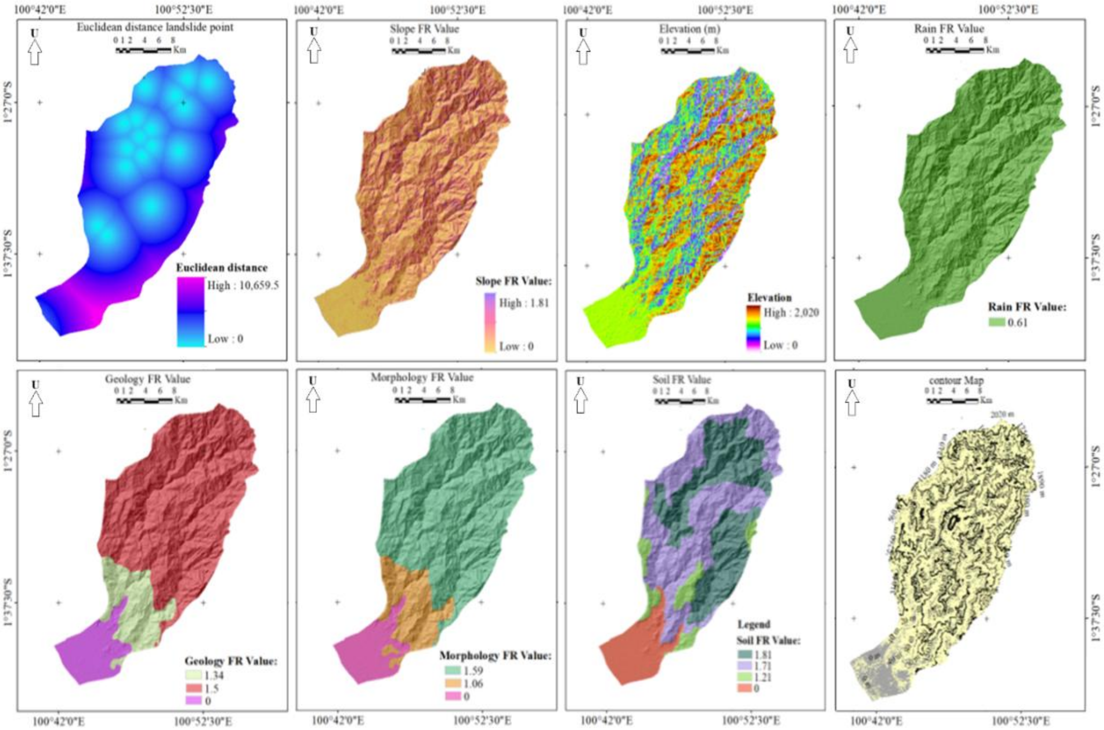


Fig. 5. Environmental parameters for predicting LP.

The maps in **figure 5** above shows, the euclidean distance to landslide points is the most immediate indicator of local susceptibility. The closer a location is to existing landslide points, the higher the potential for new failures. The slope FR shows that steeper gradients are associated with increased likelihood of landsliding. Elevation and contour data reveal strong relief across the basin, ranging from 0 to 2,020 MASL. Higher and more dissected terrain corresponds to steeper slopes and greater runoff concentration. The soil, geology, and geomorphology FR layers indicate that substrate erodibility and landform context affect how water moves through and over the ground. Weathered lithologies and hilly–mountainous landforms permit infiltration, mobilize fine particles, and promote their accumulation on less permeable layers where slip planes can form. The rainfall FR clusters around 0.61, which suggests a relatively uniform spatial distribution of rainfall in the modeling classes. Rainfall therefore acts primarily as a trigger that amplifies instability in areas already conditioned by steep slopes and susceptible materials. To examine interactions among variables and how they intersect with LULC dynamics, we summarize the multi-parameter relationships in a matrix for 2025, 2029, and 2039. More details can be seen in **figure 6**.

The matrix in **figure 7** above shows consistent patterns across time:

1) high slope FR co-occurs with short euclidean distances, indicating spatial clustering of failures along steep valley sides and dissected hillslopes;

2) soil and geology FR classes with greater erodibility tend to align with higher slope FR, pointing to coupled mechanical and hydrological controls on instability;

3) rainfall FR exhibits limited separation among classes.

However, pixels that pair steep slopes with susceptible soils or lithologies display stronger associations. When overlaid with the LULC scenarios, these multi-parameter clusters increasingly intersect areas where forest is replaced by oil palm, mixed orchard, and dryland agriculture, especially on steep to very steep terrain. This convergence explains the subsequent expansion of medium–high hazard pockets and motivates integrating the conditioning parameters with LULC projections in the susceptibility modeling that follows. More details can be seen in **figure 7**.

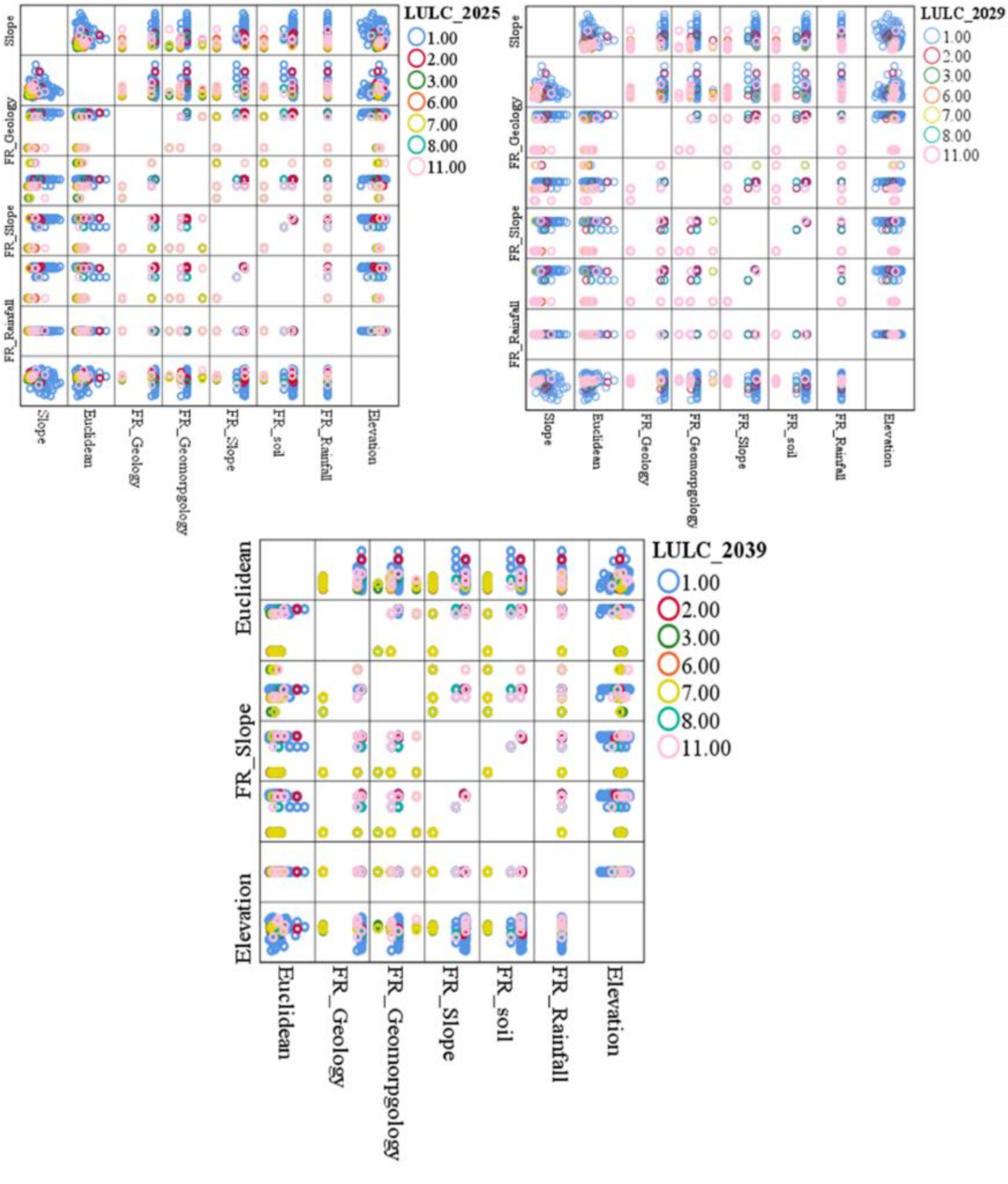


Fig. 6. Environmental-parameter matrix and LULC changes in the Kambang watershed from 2025 to 2039.

The moderate class expands from 6,066.09 ha (12.63%) in 2025 to 10,876.18 ha (22.65%) in 2029 and 11,181.24 ha (23.29%) in 2039. The high class increases from 4,205.16 ha (8.76%) in 2025 to 4,713.93 ha (9.82%) in 2029 and 4,973.76 ha (10.36%) in 2039. Combined, the moderate + high categories grow from 10,271.25 ha (21.39%) in 2025 to 15,590.11 ha (32.47%) in 2029 (+5,318.86 ha; +11.08 percentage points), and to 16,155.00 ha (33.65%) in 2039 (+564.89 ha; +1.18 percentage points). To evaluate whether the modeling results are acceptable, we computed the ROC/AUC metrics for each year. The AUC values are 0.95 (2025), 0.87 (2029), and 0.84 (2039).

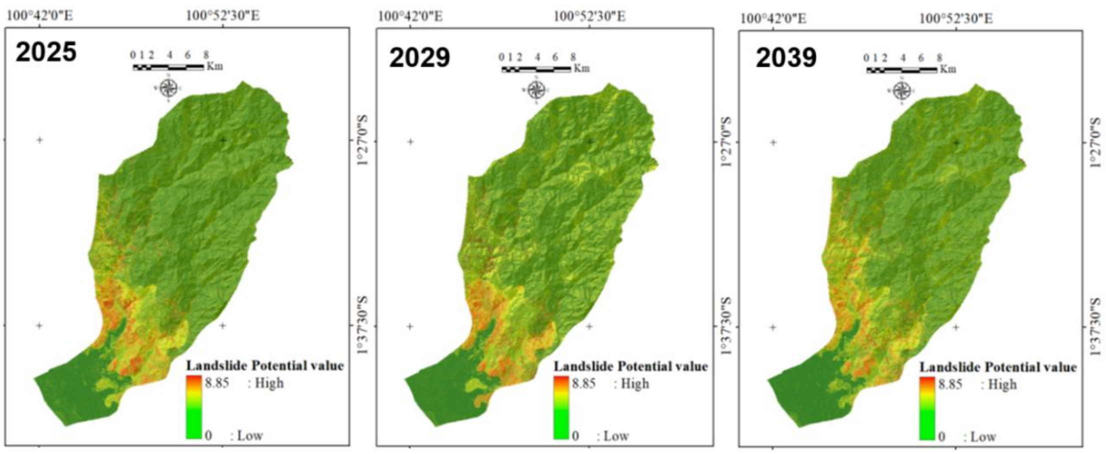


Fig. 7. Predicted LP maps for 2025, 2029, and 2039 in the Kambang watershed.

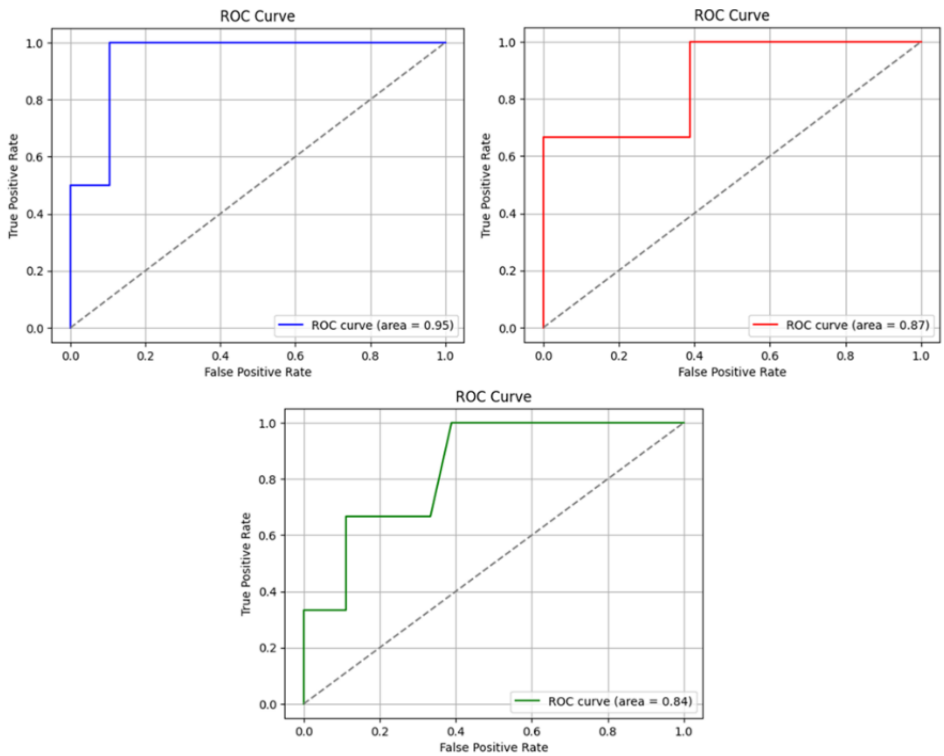


Fig. 8. ROC curves (AUC) for the landslide-potential models in 2025, 2029, and 2039.

The graph in **figure 8** above shows ROC curves that evaluate how well the models distinguish landslide from non-landslide pixels. The diagonal gray line marks random performance (AUC = 0.50); curves farther above this line indicate better discrimination. The 2025 model (blue) achieves AUC = 0.95, reflecting excellent accuracy with high true-positive rates at low false-positive rates. The 2029 model (red) attains AUC = 0.87, indicating good performance, while the 2039 model (green) yields AUC = 0.84, indicating moderate–good discrimination and greater uncertainty at the longer projection horizon. Overall, the models are suitable for planning and prioritization, with added caution recommended for the later years and periodic updates as new data becomes available.

DISCUSSION

The LULC change in the Kambang watershed is largely human-driven to meet household needs and livelihood strategies (Patel et al., 2024; Van Toor et al., 2024; Zhu et al., 2025). In practice, communities intensify agriculture even on steep to very steep slopes by converting primary and secondary forest into agricultural fields, mixed orchards, and oil-palm plantations, thereby altering natural cover and slope hydrology (Kubacka et al., 2025; Kasahun, 2025; Kafy et al., 2021). During rainy periods, land clearing facilitates downward migration of fine particles; these accumulate on relatively impermeable layers and form slip planes, especially where clayey textures are present, which increases landslide likelihood (Asada et al., 2024; Asmare, 2023).

Our hazard analysis confirms that landslide susceptibility reflects multiple conditioning factors, namely geology, geomorphology, slope, soil, rainfall, and LULC and their joint uncertainty (Huang et al., 2024; Wei et al., 2024). Changes in LULC, therefore, translate into changes in the hazard footprint (Girma et al., 2022; Kafy et al., 2021; Ghalehteimouri et al., 2022). Recent morphodeformation analysis using Sentinel-1A SAR along the Singkarak segment demonstrates how remote-sensing deformation signals can be integrated with geospatial hazard layers to sharpen the delineation of instability zones, supporting the coupling of dynamic LULC with susceptibility mapping in data-limited settings (Triyatno et al., 2025). In Kambang, the high-hazard class increases by 508.77 ha (+1.06 percentage points) from 2025→2029 and 259.83 ha (+0.54 pp) from 2029→2039. When moderate and high are combined, the at-risk area expands by 5,318.86 ha (+11.08 pp) in 2025→2029 and a further 564.89 ha (+1.18 pp) in 2029→2039.

Although much of the watershed remains in the low class, these increments are spatially concentrated on dissected hillslopes where forest conversion is occurring. Policy responses should therefore prioritize regulating land conversion on steep forested slopes, deep-rooted revegetation, and surface–subsurface drainage improvements, coupled with community education around forest-edge areas (Mondal et al., 2016; Lausch et al., 2025; Kasahun, 2025). Experience from the Tarusan watershed shows that GIS-based prioritization and community-oriented mitigation (e.g., structural drainage, vegetative buffers, and zoning enforcement) are effective for reducing compound hydro-geomorphic risks, offering an operational template for Kambang's steep-slope management (Umar & Triyatno, 2024). The prediction model for LP integrates environmental parameters with mapped landslide points and shows that projected LULC trajectories will continue to reshape susceptibility hot spots (Williams et al., 2025; Zandi & Shah Pari Far, 2025; Naeem et al., 2025). Increases are most pronounced where forest gives way to oil palm, mixed orchard, and dryland agriculture on steep terrain, processes that elevate erosion risk and soil saturation, proximate precursors to slope failure (Talukdar et al., 2021; Zhou et al., 2021).

CONCLUSIONS

The results indicate a clear and coherent trajectory linking LULC change to evolving landslide susceptibility in the Kambang watershed. Between 2025 and 2039, primary forest declines from 36,742.23 ha (76.52%) to 33,310.80 ha (69.38%), while secondary forest contracts from 4,070.07 ha (8.48%) to 2,858.04 ha (5.95%). In contrast, oil palm expands sharply to 3,920.49 ha (8.17%) by 2039, accompanied by gradual growth in mixed orchard and dryland agriculture. Built-up land, rice fields, mangrove, and water bodies remain essentially stable. Spatially, oil palm concentrates on gentler terrain and valley floors, whereas conversions affecting forest classes advance onto dissected hillslopes often below ~500 MASL, where removal of deep-rooted vegetation diminishes slope reinforcement during peak rainy seasons. These LULC trajectories translate into a measurable redistribution of landslide hazard.

The low-hazard class shrinks overall by 5,883.75 ha (−12.25 percentage points) from 2025 to 2039, while the medium class remains ~5,115.15 ha (+10.65 pp) above its 2025 level and the high class grows cumulatively by 768.60 ha (+1.60 pp). Combined, the moderate + high footprint expands from 21.39% in 2025 to 33.65% in 2039. Frequency-ratio diagnostics clarify the controls: slope and soil contain the strongest positive associations (to ~1.81), geomorphology and geology contribute secondary effects, and rainfall clustered near an FR of ~0.61 acts chiefly as a trigger that amplifies instability where conditioning factors are already adverse. The most hazardous configurations occur where steep slopes intersect susceptible substrates and recent forest loss. Model skill supports these interpretations. The LULC classifications achieve overall accuracies of 0.92 (2025), 0.88 (2029), and 0.86 (2039), exceeding accepted thresholds for downstream applications. Landslide-potential predictions yield AUC values of 0.95, 0.87, and 0.84 for 2025, 2029, and 2039, respectively, ranging from excellent to moderate-good discrimination, with uncertainty increasing at longer years. Taken together, the evidence shows that continued forest conversion on steep terrain is the primary driver of expanding susceptibility.

To mitigate future risks, priority actions should include protecting steep-slope forests, deploying deep-rooted revegetation and targeted slope stabilization, upgrading surface–subsurface drainage on hillslopes and along access roads, and incorporating the 2025/2029/2039 susceptibility maps into risk-sensitive spatial planning. Routine monitoring, community education at forest–agriculture interfaces, and periodic model updates as new observations accrue will be essential to keep risk reduction aligned with the basin's rapidly changing land systems.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the academic community of Universitas Negeri Padang for their invaluable assistance with data collection, data processing, and the preparation of spatial datasets. The author also extends thanks to the Institute for Research and Community Service (LP2M), Universitas Negeri Padang, for funding this research through the "Research Collaboration with Foreign Universities" scheme under Contract No. 1845/UN35.15/LT/2025.

REFERENCES

Ahmad, M. S., Monalisa, & Khan, S. (2023). Comparative analysis of analytical hierarchy process (AHP) and frequency ratio (FR) models for landslide susceptibility mapping in Reshun, NW Pakistan. Kuwait Journal of Science, 50(3), 387–398. DOI: 10.1016/j.kjs.2023.01.004

Asada, H., Hasegawa, Y., & Minagawa, T. (2024). Development of shallow landslide susceptibility maps incorporating relative spacing index for forest management. Environmental and Sustainability Indicators, 24, 100515. DOI: 10.1016/j.indic.2024.100515

Asmare, D. (2023). Application and validation of AHP and FR methods for landslide susceptibility mapping around Choke Mountain, northwestern Ethiopia. Scientific African, 19, e01470. DOI: 10.1016/j.sciaf.2022.e01470

Badrakh, M., Tserendash, N., Choindonjamts, E., & Albert, G. (2025). Potential of random forest machine learning algorithm for geological mapping using PALSAR and Sentinel-2A remote sensing data: A case study of Tsagaan-uul area, southern Mongolia. Journal of Asian Earth Sciences: X, 14, 100204. DOI: 10.1016/j.jaesx.2025.100204

Barlian, E., Umar, I., Dewata, I., Triyatno., Putra, A., Mustikasari, E., Purnamaningtyas, S. E., Hendrajat, E. A., Sari, S. M., Yulius, Y., & Siregar, D. R. (2025). GIS and AHP-based flood zoning and conservation strategies in the Tarusan watershed, Indonesia. Geographia Technica, 20(2), 114–133. DOI: 10.21163/GT_2025.202.08

Beroho, M., Yousfi, M., Chehbouni, M., El Housni, H., Khaoulani, M., Mjahed, H., El Moatamid, A., Hakmi, M., Ait El Abdoun, A., &Soulaimani, A. E. M. (2025). A novel SWAT-based framework to integrate climate and LULC scenarios for predicting hydrology and sediment dynamics in the watersheds of Mediterranean ecosystems. Journal of Environmental Management, 388, 125446. DOI: 10.1016/j.jenvman.2025.125446

Chowdhury, Md. S., Rahman, Md. N., Sheikh, Md. S., Sayeid, Md. A., Mahmud, K. H., & Hafsa, B. (2024). GIS-based landslide susceptibility mapping using logistic regression, random forest and decision and regression tree models in Chattogram District, Bangladesh. Heliyon, 10(1), e23424. DOI: 10.1016/j.heliyon.2023.e23424

Es-smairi, A., Elmoutchou, B., Mir, R. A., OuazaniTouhami, A. E., &Namous, M. (2023). Delineation of landslide susceptible zones using frequency ratio (FR) and Shannon entropy (SE) models in northern Rif, Morocco. Geosystems and Geoenvironment, 2(4), 100195. DOI: 10.1016/j.geogeo.2023.100195

Febriandi., Fatimah, S., Triyatno,., Hermon, D., Putra, A., Mutmainah, H., Arifin, T., &Akhwady, R. (2025). Predicting of land cover changes until 2030 and assessing sustainability status in the Mandeh region, Indonesia. *GeographiaTechnica, 20*(1), 263–280. DOI: 10.21163/GT_2025.201.18

Ghalehteimouri, K. J., Shamsoddini, A., Mousavi, M. N., Che Ros, F. B., &Khedmatzadeh, A. (2022). Predicting spatial and decadal of land use and land cover change using integrated cellular automata Markov chain model based scenarios (2019–2049) Zarriné-Rūd River Basin in Iran. Environmental Challenges, 6, 100399. DOI: 10.1016/j.envc.2021.100399

Girma, R., Fürst, C., &Moges, A. (2022). Land use land cover change modeling by integrating artificial neural network with cellular automata-Markov chain model in Gidabo river basin, main Ethiopian rift. Environmental Challenges, 6, 100419. DOI: 10.1016/j.envc.2021.100419

Huang, F., Wang, Y., Chen, J., Wu, C., Fang, Z., & Hong, H. (2024). Uncertainties of landslide susceptibility prediction: Influences of random errors in landslide conditioning factors and errors reduction by low pass filter method. Journal of Rock Mechanics and Geotechnical Engineering, 16(1), 213–230. DOI: 10.1016/j.jrmge.2023.11.001

Ikhwan., Triyatno., Putra, A., & Syah, N. (2025). Dynamics of LULC changes in communal lands: A socio-cultural and spatial analysis in Bukittinggi City, Indonesia. Geographia Technica, 20(1), 112–126. DOI: 10.21163/GT_2025.201.09

Kafy, A.-A., Dey, N. N., Al Rakib, A., Rahaman, Z. A., Nasher, N. M. R., & Bhatt, A. (2021). Modeling the relationship between land use/land cover and land surface temperature in Dhaka, Bangladesh using CA-ANN algorithm. Environmental Challenges, 4, 100190. DOI: 10.1016/j.envc.2021.100190

Kasahun, M. (2025). Quantifying deforestation drivers through multi-temporal LULC analysis and population-forest correlation modeling: A case study of Dara Woreda, Ethiopia. Environmental Challenges, 19, 101163. DOI: 10.1016/j.envc.2025.101163

Kubacka, M., Piniarski, W., Żywica, P., & Frazier, A. E. (2025). Tracking spatio-temporal LULC changes in key ecological network elements using fragmentation metrics and a custom raster-based approach: A multi-scale study from Poland. Ecological Informatics, 90, 103369. DOI: 10.1016/j.ecoinf.2025.103369

Lausch, A., Selsam, P., Heege, T., von Trentini, F., Almeroth, A., Borg, E., Klenke, R., &Bumberger, J. (2025). Monitoring and modelling landscape structure, land use intensity and landscape change as drivers of water quality using remote sensing. Science of The Total Environment, 960, 178347. DOI: 10.1016/j.scitotenv.2024.178347

Mondal, Md. S., Sharma, N., Garg, P. K., &Kappas, M. (2016). Statistical independence test and validation of CA Markov land use land cover (LULC) prediction results. The Egyptian Journal of Remote Sensing and Space Science, 19(2), 259–272. DOI: 10.1016/j.ejrs.2016.08.001

Murmu, J., Radhadevi, L., Pande, C., Bandaru, M., & Kumar, M. (2025). Indicators of sustained agriculture, impacts of LULC and weather parameters on ET: Case study in Chota Nagpur Plateau. Environmental and Sustainability Indicators, 27, 100836. DOI: 10.1016/j.indic.2025.100836

Naeem, M., Liu, C., Shafi, M., Iqbal, J., Shikhawat, A. F., Rao, Z., Ahbia, A., Anṣari, M., Baloch, W. U., & Niaz, A. (2025). Assessing and predicting Bojiang lake area and LULC changes from 2000 to 2045. Journal of Hydrology: Regional Studies, 58, 102216. DOI: 10.1016/j.ejrh.2025.102216

Patel, A., Vyas, D., Chaudhari, N., Patel, R., Patel, K., & Mehta, D. (2024). Novel approach for the LULC change detection using GIS & Google Earth Engine through spatiotemporal analysis to evaluate the urbanization growth of Ahmedabad city. Results in Engineering, 21, 101788. DOI: 10.1016/j.rineng.2024.101788

Putra, A., Hermon, D., Dewata, I., Barlian, E., & Triyanto. (2025). USLE method for erosion prediction and conservation measures at the Air Dingin watershed of the upstream part in Padang City, Indonesia. In Erosion Measurement, Modeling, and Management: Challenges and Solutions. 123–148. Apple Academic Press. DOI: 10.1201/9781003494508_6

Rodríguez-Ortega, J., Tabik, S., Benhammou, Y., Khaldi, R., & Alcaraz-Segura, D. (2025). Land use and land cover fraction estimation for Sentinel-2 RGB images: A new LULC mapping task. Remote Sensing Applications: Society and Environment, 39, 101626. DOI: 10.1016/j.rsase.2025.101626

Roy, D., Sarkar, A., Kundu, P., Paul, S., & Chandra Sarkar, B. (2023). An ensemble of evidence belief function (EBF) with frequency ratio (FR) using geospatial data for landslide prediction in Darjeeling Himalayan region of India. Quaternary Science Advances, 11, 100092. DOI: 10.1016/j.qsa.2023.100092

Shanley, C. S., Eacker, D. R., Reynolds, C. P., Bennetsen, B. M. B., & Gilbert, S. L. (2021). Using LiDAR and random forest to improve deer habitat models in a managed forest landscape. Forest Ecology and Management, 499, 119580. DOI: 10.1016/j.foreco.2021.119580

Talukdar, S., Eibek, K. U., Akhter, S., Ziaul, S., Md. Towfiqul Islam, A. R., & Mallick, J. (2021). Modeling fragmentation probability of land-use and land-cover using the bagging, random forest and random subspace in the Teesta River Basin, Bangladesh. Ecological Indicators, 126, 107612. DOI: 10.1016/j.ecolind.2021.107612

Triyatno., Bert, I., Idris., Hermon, D., & Putra, A. (2020). Hazards and morphometry to predict the population loss due to landslide disasters in Koto XI Tarusan – Pesisir Selatan. International Journal of GEOMATE, 19(76), 98–103. DOI: 10.21660/2020.76.ICGeo12

Triyatno., Berd, I., & Idris. (2023). Spatial model of flood hazard due to land cover change in the Tarusan watershed, West Sumatra—Indonesia. International Journal of GEOMATE, 25(108), 21–29. DOI: 10.21660/2023.108.3774

Triyatno., Anwar, S., Febriandi., & Rahmi, L. (2024). GIS based on analysis of earthquake hazard level using the PGA in Siberut–Mentawai Islands Regency for sustainability in disaster mitigation. Journal of Sustainability Science and Management, 19(12), 107–119. DOI: 10.46754/jssm.2024.12.007

Triyatno., Caesario, D., Febriandi., Aulia, B., & Syarief, A. (2025). Predicting earthquake potential through morphodeformation modeling using Sentinel-1A imagery in the Singkarak segment, Indonesia. Geographia Technica, 20(2), 190–204. DOI: 10.21163/GT_2025.202.12

Umar, I., & Triyatno. (2024). Flood hazard mitigation at Tarusan Watershed, South Pesisir District, West Sumatera Province. Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan, 14(1), 101–108. DOI: 10.29244/jpsl.14.1.101-108

Van Toor, M. L., Davranche, A., Delaunay, G., Murgue, C., Waldenström, J., &Arzel, C. (2024). An evaluation of global LULC maps for the estimation of habitat use of a declining migratory waterbird along its flyway. SSRN. DOI: 10.2139/ssrn.5037387

Wei, Y., Qiu, H., Liu, Z., Huangfu, W., Zhu, Y., Liu, Y., Yang, D., & Kamp, U. (2024). Refined and dynamic susceptibility assessment of landslides using InSAR and machine learning models. Geoscience Frontiers, 15(6), 101890. DOI: 10.1016/j.gsf.2024.101890

Williams, C., Whitbread, K., Hall, A., Roberson, S., Finlayson, A., Palamakumbura, R. N., Hulbert, A., &Paice, M. (2025). Capturing exposed bedrock in the upland regions of Great Britain: A geomorphometric focused random forest approach. Computers & Geosciences, 196, 105814. DOI: 10.1016/j.cageo.2024.105814

Zandi, R., & Shah Pari Far, G. (2025). Evaluating the factors affecting landslides using machine learning algorithms (case study: The catchment area of Karun-3 Dam, Iran). The Egyptian Journal of Remote Sensing and Space Sciences, 28(3), 512–522. DOI: 10.1016/j.ejrs.2025.07.005

Zhou, X., Wen, H., Zhang, Y., Xu, J., & Zhang, W. (2021). Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. Geoscience Frontiers, 12(5), 101211. DOI: 10.1016/j.gsf.2021.101211

Zhu, Y., [add remaining authors]. (2025). Analysis and prediction of spatiotemporal carbon storage changes in the Taihu Lake Basin in Jiangsu Province based on PLUS and InVEST model. Trees, Forests and People, 100916. DOI: 10.1016/j.tfp.2025.100916

PREDICTING FIRE HOTSPOTS IN KALIMANTAN BASED ON CLIMATE FACTORS USING GAUSSIAN PROCESS REGRESSION AND LONG SHORT-TERM MEMORY

Sri NURDIATI[1], Mochamad Tito JULIANTO[1], Ionel HAIDU[2,3], Muhammad Daryl FAUZAN [1], Hari NURDIANTO[1], Syukri Arif RAFHIDA[1] & Mohamad Khoirun NAJIB[1]*

DOI: 10.21163/GT_2026.212.02

ABSTRACT

Forest and land fires in Kalimantan present a recurrent environmental challenge, driven by local and global climatic factors. Predicting fire hotspots is crucial for mitigation efforts. This study compares the performance of Gaussian Process Regression (GPR) and Long Short-Term Memory (LSTM) networks in forecasting monthly fire hotspots based on six climatic indicators, including rainfall, dry spells, and ENSO and IOD indices. GPR models were developed using several kernels and hyperparameter tuning methods, while LSTM models applied multiple architectural configurations combined with regularisation techniques. The results show that while GPR models achieved good fitting on training data, they suffered from overfitting and lower accuracy during testing, even after optimisation. In contrast, the LSTM model with two LSTM layers and four dense layers demonstrated superior predictive performance, achieving a testing RMSE of 522.12 and an Explained Variance Score (EVS) of 0.834. LSTM effectively captured complex temporal patterns inherent in climate-driven fire hotspot data. Nevertheless, both models faced difficulties in predicting anomalies linked to socio-economic interventions, such as the significant reduction in fire hotspots in 2018.The findings highlight the effectiveness of LSTM in modelling temporally dependent environmental phenomena and suggest the need for integrating socio-economic variables into future predictive frameworks to improve robustness. This study contributes valuable insights towards enhancing early warning systems for forest fire risk management in Kalimantan and other tropical regions.

Keywords: *Fire hotspots; Kalimantan; Gaussian Process Regression; Long Short-Term Memory; Machine learning.*

1. INTRODUCTION

Kalimantan, the third largest island in the world with an area of 539,460 km² (MacKinnon & Hatta, 2013), frequently draws attention due to recurring environmental issues, notably forest and land fires (Sarmiasih & Pratama, 2019; Gusnita, 2021; Qirom et al., 2022; Saharjo & Hasanah, 2023). These events occur annually (Saharjo & Velicia, 2018), with major fire episodes recorded in 1982, 1997–1998, 2015, and 2019 (Najib et al., 2022; Werf et al., 2017). Forest fires have significant impacts, including ecosystem destruction, air pollution, public health risks, and economic losses estimated at approximately IDR 74 million per hectare (Ulya & Yunardy, 2006; Mulia et al., 2021; Sari et al., 2022; Wasis et al., 2019). Early indicators of forest fires can be identified through the presence of hotspots, defined as areas experiencing elevated surface temperatures associated with burning (Nugrahani et al., 2024; Reddy et al., 2019).

[1] *School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia. * Corresponding author nurdiati@apps.ipb.ac.id (SN); mtjulianto@apps.ipb.ac.id (MTJ); ryls93daryl@apps.ipb.ac.id (MDF); ryls93daryl@apps.ipb.ac.id (HN); asrafhida@apps.ipb.ac.id (SAR); mkhoirun@apps.ipb.ac.id (MKN)*
[2] *LOTERR (Laboratoire des Observations des Territoires), Université de Lorraine, F-57000 Metz, France. ionel.haidu@univ-lorraine.fr (IH).*
[3] *STAR-UBB (Scientific and Technological Advanced Research Institute), Babeş-Bolyai University, 400084 Cluj-Napoca, Romania.*

Forest fires in Kalimantan are strongly influenced by both local and global climatic conditions. Variables such as air temperature, rainfall, and humidity are closely linked to hotspot occurrence (Saharjo & Velicia, 2018). Moreover, global climate phenomena such as the El Niño–Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD) play crucial roles in shaping weather and climate patterns across Indonesia (Ardiyani et al., 2023; Irwandi et al., 2019; Nurdiati et al., 2022; Rafhida et al., 2024). The ENSO phases, El Niño and La Niña, respectively prolong drought periods or enhance rainfall intensity, while the IOD influences rainfall distribution across regions (Anggraini & Trisakti, 2011; Bramawanto & Abida, 2017; Nurdiati et al., 2024).

Given the extensive impacts of forest fires, there is an urgent need to accurately predict the number of hotspots in Kalimantan. Reliable prediction models not only facilitate the identification of high-risk areas but also enable governments and other stakeholders to design effective mitigation strategies (Sudrajat & Subekti, 2019; Yuliarti & Anggraini, 2022). Although numerous predictive methods have been explored in previous studies, challenges remain in achieving accuracy and reliability, particularly when using climate indicators. Techniques such as autoregression (AR), artificial neural networks (ANN), support vector regression (SVR), random forest regression, and gradient boosting regression often struggle to effectively capture temporal patterns in the data (Nugrahani et al., 2024; Nurdiati, Sopaheluwakan, et al., 2022).

This study builds upon Najib et al. (2022) and (Nugrahani et al., 2024) by analysing a dataset consistent with their monthly BMKG hotspot counts on a 0.25° × 0.25° grid across Kalimantan for 2001–2020 with preprocessing that excludes persistently low-activity grids following Najib et al. (2021). Najib et al. (2022) utilised rainfall and dry spell indicators to predict hotspot abundance using copula regression, whereas Nugrahani et al. (2022) incorporated rainfall anomalies, as well as IOD and ENSO indices, applying various machine learning methods. However, previous research has yet to fully leverage these indicators within more sophisticated predictive models, particularly under Kalimantan's dynamic temporal conditions. It was also observed that the previous models exhibited significant overfitting, highlighting the need for more robust approaches.

In this study, two predictive methods are employed to model the number of hotspots based on climatic indicators: Gaussian Process Regression (GPR) and Long Short-Term Memory (LSTM) networks. GPR is renowned for its capability to model complex relationships between variables, particularly in small datasets. It has demonstrated superior predictive performance in various studies, often yielding lower fitting errors compared to alternative methods (Kamath et al., 2018). Furthermore, GPR has proven effective across applications such as remote sensing and weather data analysis (Hultquist et al., 2014; La Fata et al., 2024). According to Foley (2024) GPR is particularly advantageous for dealing with incomplete or noisy datasets, offering a probabilistic approach that explicitly quantifies prediction uncertainties—making it a promising option for modelling the relationship between climate indicators and forest fire hotspots.

Meanwhile, the LSTM deep learning algorithm has gained popularity in recent years for hotspot prediction tasks due to its proficiency in capturing complex temporal dependencies in time-series data (DiPietro & Hager, 2019; Kadir et al., 2022; Listia Rosa et al., 2022; Kadir et al., 2023; Luo et al., 2024; Li et al., 2024; Eaturu & Vadrevu, 2025). LSTM-based ensembles fused with (variance-weighted) Kalman filtering for environmental time series, supports the choice of LSTM for climate temporal signals and highlights the value of robust gap-filling before modelling (Haidu et al., 2025). LSTM represents an advancement over conventional Recurrent Neural Networks (RNNs), effectively addressing the "vanishing gradient" problem where gradients used to update network weights diminish or disappear (Noh, 2021). Research by Kadir et al. (2023) demonstrated that LSTM models could accurately predict the quantity and distribution of forest fire hotspots in Indonesia based on NASA MODIS datasets, achieving a mean percentage error of 6.94%. However, that study solely relied on past hotspot data without incorporating other climatic factors. The present study seeks to enhance prediction accuracy by integrating a range of climate indicators.

Through these two approaches, this research aims to evaluate the effectiveness of GPR and LSTM methods in predicting the number of hotspots in Kalimantan based on both local and global climate indicators. The study not only focuses on predictive accuracy but also explores the relative strengths and weaknesses of each method. The main contribution of this research lies in providing empirical insights into which predictive method is more effective and reliable under the context of climate variability and the tropical environmental characteristics of Kalimantan. The findings are expected to inform the development of more robust predictive models for wildfire mitigation applications. Furthermore, this research contributes to practical policy formulation by offering data-driven recommendations to prevent and mitigate the impacts of forest fires in Kalimantan and other regions with similar environmental profiles.

2. STUDY AREA AND DATASETS

Kalimantan Island (5°S–7.25°N, 108°E–119.75°E), shared by three countries—Indonesia, Malaysia, and Brunei Darussalam—has attracted considerable attention in wildfire hotspot research. The Indonesian portion of the island comprises five provinces: West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, and North Kalimantan. Geographically, Kalimantan is characterised by extensive tropical rainforests, vast peatlands, and a humid tropical climate with consistently high temperatures and humidity throughout the year. The island accounts for approximately 33.8% of Indonesia's total peatland area, making it an ecologically significant yet highly fire-prone region.

Rainfall patterns across Kalimantan are classified into equatorial and monsoonal types. Areas experiencing a monsoonal pattern, particularly in the southern and central regions, are more susceptible to wildfires, especially during prolonged dry seasons intensified by El Niño events. Research indicates that strong El Niño events can extend the dry season, increase the number of consecutive dry days, and subsequently trigger a greater number of wildfire hotspots. Conversely, La Niña events, which bring more frequent rainfall, tend to reduce the risk of fires. The positive phase of the Indian Ocean Dipole (IOD), when occurring simultaneously with El Niño, can further exacerbate drought conditions.

Hotspot data used in this study were obtained from Indonesian Meteorological, Climatological, and Geophysical Agency (BMKG) as a distributor of satellite active-fire detections derived from MODIS (MOD14/MYD14) and VIIRS (e.g., VNP14IMG/VJ114IMG) products that implement the contextual active-fire detection algorithm. To ensure that the analysis focused on fire-prone areas, regions with consistently low hotspot activity, such as highland areas with high rainfall, were excluded. This approach follows the classification method proposed by Najib et al. (2021), in which Kalimantan's hotspot grids were clustered using the k-means (Lloyd, squared-Euclidean) on hotspot time series evaluating 8 clusters and designating the lowest-incidence cluster as the candidate for removal.

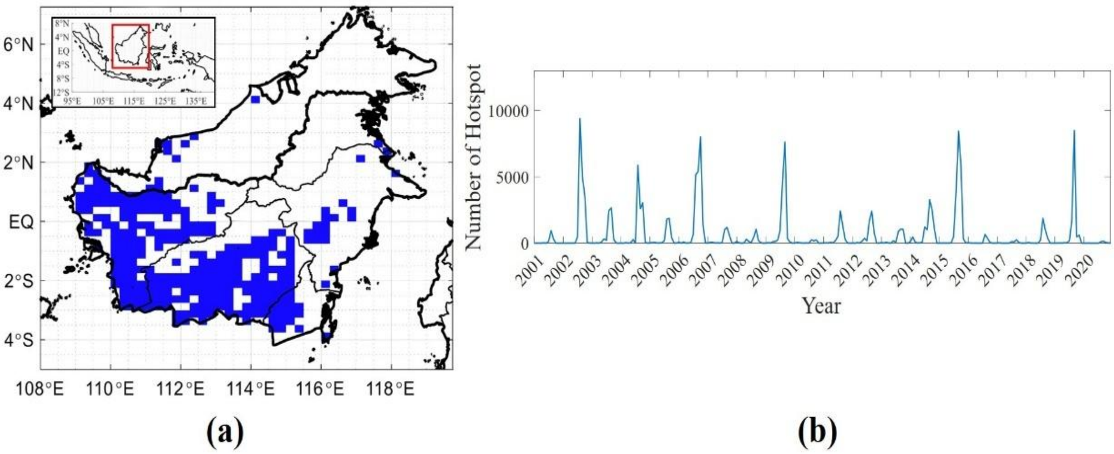


Fig. 1. (a) Kalimantan Island and selected research areas; (b) Total monthly hotspot counts.

Clusters characterised by extremely low fire incidence—with a maximum average of only 3.43 hotspots per grid point (e.g., in 2002)—were disregarded to avoid distorting the analysis of wildfire characteristics in Kalimantan.

Following Najib et al. (2021), we first excluded 0.25°×0.25° grid cells with persistently low hotspot activity using a k-means clustering of grid-level hotspot time series, and retained only high-incidence grids. This study only includes areas with significant hotspot concentrations, particularly in Central, South, and West Kalimantan, as shown in **figure 1(a)**. For each month between 2001 and 2020, hotspot counts were then summed over all retained grids, yielding a single monthly total hotspot count. Analysis was conducted using total monthly hotspot data from 2001 to 2020, illustrated in **figure 1(b)**. The trend in hotspot numbers exhibits a clear seasonal cycle, with peaks typically occurring mid-year and lower counts during the rest of the year. The number of hotspots tends to increase during the dry season, especially between July and September, due to low humidity and minimal rainfall. In contrast, during the rainy season, the number of hotspots declines significantly.

This study employed both local and global climatic factors to predict monthly total hotspot counts. Local climatic variables analysed included total rainfall, rainfall anomalies, and the number of dry days (defined as days with less than one millimetre of rainfall), extracted from the CMORPH dataset available from the National Oceanic and Atmospheric Administration (NOAA). Meanwhile, global climatic factors considered were indices related to the ENSO and IOD phenomena, obtained from NOAA. As a result, our dataset consists of a univariate target time series of regional hotspot totals and a corresponding multivariate time series of six aggregated climatic predictors; both the GPR and LSTM models are trained on this aggregated series rather than on individual grid-cell time series. Consequently, six independent variables were used in this research, with 80% of the data allocated for model training and the remaining 20% for testing.

Prior to model fitting, each of the six climatic predictor variables was standardised to zero mean and unit variance. The standardisation parameters (mean and standard deviation) were estimated exclusively from the training period (2001–2016) and then applied unchanged to the held-out test period (2017–2020). For the 16-fold cross-validation used in the GPR experiments, standardisation was recomputed in each fold by fitting the scaler on the 15 training years and applying it to the left-out validation year. For the LSTM models, lagged inputs from t–1 to t–12 was generated from these standardised predictor series, ensuring that all input windows were derived from features scaled using training-only statistics. The target variable, monthly hotspot counts, was retained in its original units without normalisation.

3. METHODS

This study employs two machine learning methods to predict the number of hotspots in Kalimantan, namely Gaussian Process Regression and Long Short-Term Memory.

3.1. Gaussian Process Regression

Gaussian Process (GP) is a widely used statistical method for data fitting and prediction based on machine learning. A GP is a form of stochastic process, defined as a collection of random variables Y indexed by an input space X, where any finite subset of these variables follows a multivariate Gaussian distribution (Rasmussen & Williams 2019). Gaussian Process Regression (GPR) is a non-parametric, Bayesian approach to regression, offering high flexibility in handling non-linear relationships between input and output variables.

A GPR model is fully characterised by a mean function $m(\mathbf{x})$ and a covariance function (kernel) $k(\mathbf{x}_i, \mathbf{x}_j)$,, defined as follows:

$$m(\mathbf{x}) = \mathbb{E}[\boldsymbol{Y}_x], \tag{1}$$

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \mathbb{E}\left[\left(\boldsymbol{Y}_{x_i} - m(\boldsymbol{x}_i)\right)\left(\boldsymbol{Y}_{x_j} - m(\boldsymbol{x}_j)\right)\right]. \tag{2}$$

The only requirement for a valid kernel is that it must produce a positive semi-definite covariance matrix for any set of input points. Through this matrix, a Gaussian process can make predictions at new points based on information from the training data.

In the Bayesian regression framework, GPR models the relationship between input $\mathbf{x}_i$ and output y_i as:

$$y_i = f(\mathbf{x}_i) + \epsilon_i, \tag{3}$$

where ϵ_i represents Gaussian noise with zero mean and variance σ^2. The function f over the input space is given a multivariate Gaussian prior distribution with zero mean and covariance matrix $\mathbf{K}$:

$$f|\mathbf{X}, \theta \sim N(\mathbf{0}, \mathbf{K}). \tag{4}$$

For prediction, GPR provides a Gaussian distribution for each array of new input variables $\mathbf{X}^* = [\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_n^*]$. The estimated value for y^* is $m(\mathbf{X}^*)$ and the variance of y^* is given by $\mathrm{cov}(\mathbf{X}^*)$. In kernel space, the mean and covariance prediction equations are:

$$m(\mathbf{X}^*) = \mathbf{k}^{*\mathrm{T}}(\mathbf{K} + \sigma^2\boldsymbol{I})^{-1}\mathbf{y}, \tag{5}$$

$$\mathrm{cov}(\mathbf{X}^*) = k(\mathbf{X}^*, \mathbf{X}^*) - \mathbf{k}^{*\mathrm{T}}(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{k}^*, \tag{6}$$

where $\mathbf{k}^*$ is the covariance vector between the new data $\mathbf{X}^*$ and the training data.

3.1.1. Hyperparameter Tuning

In machine learning, parameters that are not updated during model training and must be configured beforehand are known as hyperparameters. Model performance can vary significantly depending on the choice of hyperparameters; hence, careful tuning is crucial. In GPR, the noise level σ^2 is treated as a hyperparameter that can be optimised. Noise level refers to the variation or uncertainty in observations, represented as the variance of the added Gaussian noise (Li et al., 2020). In this study, tuning of the noise level was conducted using several optimisation methods, following (Rasmussen & Williams, 2019), namely maximum marginal likelihood (MML), Bayesian optimisation, grid search, and random search.

1. Maximum Marginal Likelihood

The marginal likelihood function is obtained by integrating the likelihood function multiplied by the prior distribution of f:

$$p(\boldsymbol{y}|\mathbf{X}, \boldsymbol{\theta}) = \int p(\boldsymbol{y}|\mathbf{f}, \mathbf{X}, \boldsymbol{\theta})p(\mathbf{f}|\mathbf{X}, \boldsymbol{\theta})\, d\mathbf{f}, \tag{7}$$

where $\boldsymbol{y}$ represents the observed response variable, $\mathbf{X}$ is the matrix of observed inputs, $\boldsymbol{\theta}$ denotes the model parameters, and $\mathbf{f}$ is the function representing the relationship between $\mathbf{X}$ and $\boldsymbol{y}$. Within the Gaussian Process framework, the prior distribution $\mathbf{f}|\mathbf{X}, \boldsymbol{\theta}$ is multivariate Gaussian, $\mathbf{f}|\mathbf{X}, \boldsymbol{\theta} \sim N(\mathbf{0}, \mathbf{K})$.

2. Bayesian optimization

Bayesian optimisation relies on a surrogate function to locate the best parameter values (Ye et al., 2019). The surrogate function uses a probabilistic model that is updated with new information. It works by identifying regions likely to contain the optimum and sampling points close to these regions to refine the model. In this context, Gaussian processes are used as surrogate functions for the objective function.

The probability of the objective function is evaluated using an acquisition function (Ye et al., 2019), commonly the expected improvement (EI), defined as:

$$EI_{y^*}(x) = \int_{-\infty}^{\infty} max(y^* - y, 0)\, p(y|x, \boldsymbol{\theta})dy, \tag{8}$$

where x is the input, y^* is the best observed value, y is the model's predicted value, and $p(y|x, \boldsymbol{\theta})$ is the posterior distribution.

3. Grid Search

Grid search is a hyperparameter search technique that systematically tries all combinations within a specified parameter space. It constructs a grid of parameter values and evaluates the model performance at each grid point.

4. Random Search

Random search selects hyperparameter combinations randomly within a defined search space. Unlike grid search, it does not exhaustively explore all combinations but samples a predefined number of configurations.

3.1.2. Kernel Selection

As introduced earlier, GPR relies on kernels to make predictions at new points. The kernel functions as a covariance function measuring the similarity between two points in input space. The choice of an appropriate kernel is crucial as it determines how the model captures patterns in the data, especially when the relationship is non-linear. This study employed three kernels: the exponential kernel, the squared exponential (Gaussian) kernel, and the Matern 3/2 kernel. In addition, we used variations of these kernels with Automatic Relevance Determination (ARD), which introduces separate length-scales for each explanatory variable. ARD enhances the model's adaptability by allowing each input dimension to have its own scaling parameter (Rasmussen & Williams, 2019). The original kernels and their ARD versions are summarised as in **table 1.**

Table 1.

Kernel Used in Study.

Kernel	Original	ARD
Exponential	$\sigma_f^2 \exp\left(-\frac{\lvert\mathbf{x}_i - \mathbf{x}_j\rvert}{\sigma_l}\right)$	$\sigma_f{}^2 \exp\left(-\sqrt{\sum_{m=1}^{d} \frac{(\mathbf{x}_{im} - \mathbf{x}_{jm})^2}{\sigma_{lm}{}^2}}\right)$
Squared Exponential (Gaussian)	$\sigma_f{}^2 \exp\left(-\frac{1}{2}\frac{\lvert\mathbf{x}_i - \mathbf{x}_j\rvert^2}{\sigma_l{}^2}\right)$	$\sigma_f{}^2 \exp\left(-\frac{1}{2}\sqrt{\sum_{m=1}^{d} \frac{(\mathbf{x}_{im} - \mathbf{x}_{jm})^2}{\sigma_{lm}{}^2}}\right)$
Matern3/2	$\sigma_f{}^2\left(1 + \frac{\sqrt{3}}{\sigma_l} r\right)\exp\left(-\frac{\sqrt{3}}{\sigma_l} r\right)$	$\sigma_f{}^2\left(1 + \frac{\sqrt{3}}{\sigma_l} r_{ard}\right)\exp\left(-\frac{\sqrt{3}}{\sigma_l} r_{ard}\right)$

where:

- $\mathbf{x}_i$ represents the training data,
- $\mathbf{x}_j$ denotes the testing data,
- σ_f is the signal standard deviation,
- σ_{lm} is the length-scale parameter for the mmm-th explanatory variable,
- d is the number of explanatory variables, and
- r and r_{ard} represent the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ for standard and ARD kernels, respectively.

3.2. Long Short-Term Memory (LSTM)

The Long Short-Term Memory (LSTM) model is a type of Recurrent Neural Network (RNN) designed for time series data that exhibits dependencies on previous observations. The output from prior steps is used to update new weight values. These updated weights are then applied in subsequent computations, establishing a connection between new and prior data.

The LSTM model was specifically designed to overcome the limitations of conventional RNNs, which often struggle with long-range dependencies and are prone to the vanishing and exploding gradient problems [36]. LSTM represents a specialised architecture within the RNN family. Unlike a simple RNN, the hidden units of an LSTM consist of three principal gates: the forget gate, the input gate, and the output gate. The mathematical formulations of the *forget gate* $f^{(t)}$(8), *input gate* $i^{(t)}$ (9), *output gate* $o^{(t)}$ (10), candidate *cell state* $\tilde{C}$(11), *cell state* C (12), and *hidden layer* $h^{(t)}$ (13) are as follow:

$$\boldsymbol{f}^{(t)} = \sigma(\boldsymbol{W}_f \boldsymbol{x}_t + \boldsymbol{W}_f \boldsymbol{h}^{(t-1)} + \boldsymbol{\theta}_f) \tag{9}$$

$$\boldsymbol{i}^{(t)} = \sigma(\boldsymbol{W}_i \boldsymbol{x}_t + \boldsymbol{W}_i \boldsymbol{h}^{(t-1)} + \boldsymbol{\theta}_i) \tag{10}$$

$$\boldsymbol{o}^{(t)} = \sigma(\boldsymbol{W}_o \boldsymbol{x}_t + \boldsymbol{W}_o \boldsymbol{h}^{(t-1)} + \boldsymbol{\theta}_o). \tag{11}$$

$$\widetilde{\boldsymbol{C}}^{(t)} = tanh\left(\boldsymbol{W}_c \boldsymbol{x}_t + \boldsymbol{W}_c \boldsymbol{h}^{(t-1)} + \boldsymbol{\theta}_c\right) \tag{12}$$

$$\boldsymbol{C}^{(t)} = \boldsymbol{f}^{(t)} \cdot \boldsymbol{C}^{(t-1)} + \boldsymbol{i}^{(t)} \cdot \widetilde{\boldsymbol{C}}^{(t)} \tag{13}$$

$$\boldsymbol{h}^{(t)} = \boldsymbol{o}^{(t)} \cdot tanh\ (\boldsymbol{W}_c \boldsymbol{C}^{(t)} + \boldsymbol{\theta}_h) \tag{14}$$

where $\boldsymbol{W}_f, \boldsymbol{W}_i, \boldsymbol{W}_o$, and $\boldsymbol{W}_h$ represent the weight matrices for the forget gate, input gate, output gate, and hidden layer, respectively. Likewise, $\boldsymbol{\theta}_f, \boldsymbol{\theta}_i, \boldsymbol{\theta}_o$ and $\boldsymbol{\theta}_h$ are the biases associated with each corresponding gate.

The weights and biases in the model must be initialised prior to the training process. Glorot, Bengio (2010) introduced the *Glorot uniform* initialisation method, which sets the initial weights based on a uniform distribution. This method is widely adopted for initialising weights in deep learning models. By default, the TensorFlow package in the Python programming language applies the Glorot uniform method for weight initialisation. The distribution of weights using the Glorot uniform method is defined by:

$$W \sim U(-\frac{\sqrt{6}}{\sqrt{n_{in}+n_{out}}}, \frac{\sqrt{6}}{\sqrt{n_{in}+n_{out}}}) \tag{15}$$

where n_{in} is the number of input units, and n_{out} is the number of output units. In addition, the biases are initialised to zero.

Each gate utilises either the sigmoid function σ or the hyperbolic tangent (tanh) function. The sigmoid activation outputs values within the interval $[0,1]$, controlling whether information is retained or discarded: an output close to 0 implies forgetting, whereas an output close to 1 implies retaining the information. An illustration of the LSTM architecture is presented in **figure 2**.

In this study, the LSTM model architecture is enhanced with regularisation techniques to improve performance and prevent overfitting by early stopping after 25 epochs. Early stopping monitors model performance during training and halts the process once the loss function ceases to improve after a pre-specified patience value (s epochs). This method helps prevent the model from overtraining and becoming trapped in local minima, thereby mitigating overfitting.

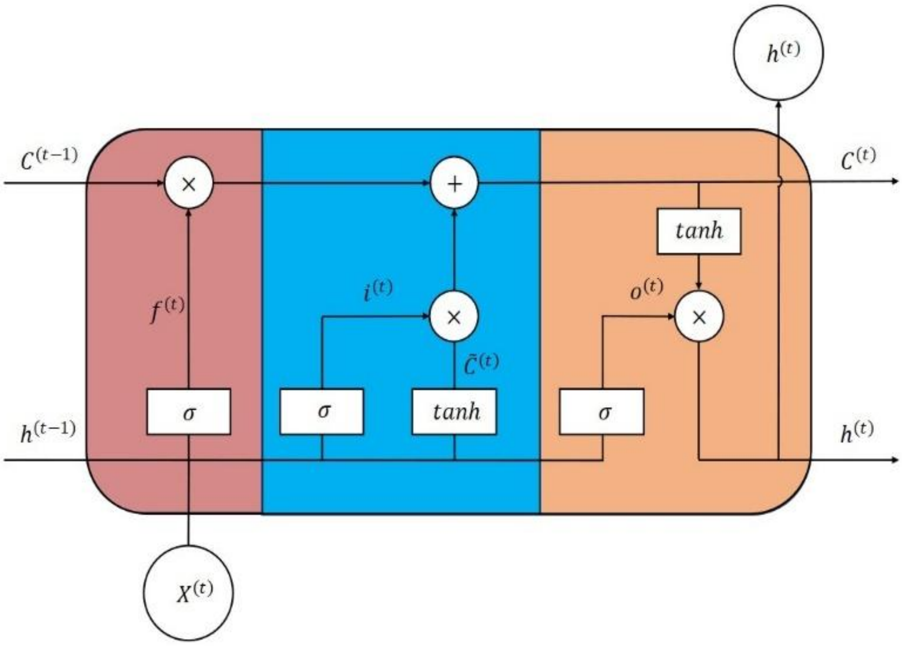


Fig. 2. LSTM model architecture.

3.3. Model Performance Metrics

To evaluate the performance of the machine learning (ML) model in predicting hotspots on the test data, two primary metrics are employed: the Root Mean Squared Error (RMSE) and the Explained Variance Score (EVS).

The RMSE measures the average squared difference between the observed and predicted values and is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}^{(i)} - \hat{\boldsymbol{y}}^{(i)})^2}\,, \tag{14}$$

where $\boldsymbol{y}$ denotes the true values and $\hat{\boldsymbol{y}}$ represents the predicted values. The RMSE ranges from $[0, \infty)$ a value of 0 indicating a perfect prediction. A smaller RMSE value implies better predictive performance and higher accuracy.

Meanwhile, the EVS measures the proportion of variance in the actual values that is captured by the model's predictions. According to Oyedele et al. (2023), the explained variance score is defined as:

$$\mathrm{EVS} = 1 - \frac{Var(\boldsymbol{y}-\hat{\boldsymbol{y}})}{Var(\boldsymbol{y})}. \tag{15}$$

where $Var(\cdot)$ denotes variance. The EVS ranges from $(-\infty, 1]$, with values closer to 1 indicating that the model explains nearly all of the variability in the data, reflecting excellent predictive performance.

4. RESULT

This section presents the evaluation results for both training and testing datasets.

4.1. Gaussian Process Regression

4.1.1. Kernel Selection

Gaussian Process Regression (GPR) was employed to predict fire hotspot occurrences from 2001 to 2020, based on six climate factors. Numerous kernel functions are available for GPR modelling, as they can be combined with other kernels. Consequently, identifying the most appropriate kernel function for the model is not straightforward. In this study, several commonly used kernels were examined: the Squared Exponential, ARD Squared Exponential, Matern32, and ARD Matern32 kernels.

In the initial stage, the most basic optimisation approach, namely maximal marginal likelihood (MML), was used to optimise the hyperparameters for each kernel. The results are presented in **table 2.**

Table 2.
Gaussian Process Regression Optimization Metric Evaluation.

Case	RMSE	
	Train	Test
GPR with Squared Exponential kernel	492.06	1116.9
GPR with ARD Squared Exponential kernel	501.29	971.27
GPR with Matern32 kernel	402.48	1133.5
GPR with ARD Matern32kernel	144.1	1227.3

Based on **table 2**, the ARD Matern32 kernel exhibited excellent performance on the training dataset, achieving a very low RMSE value. However, this high level of accuracy may indicate overfitting. This is further confirmed by the significant performance deterioration observed on the testing dataset, where the ARD Matern32 kernel produced the worst result compared to other kernels.

Thus, selecting an appropriate kernel function is crucial, as a good kernel should not only achieve high accuracy on the training data but also maintain good performance when predicting new data. Based on the consistency of the RMSE values between training and testing, the ARD Squared Exponential kernel was identified as the most relevant and robust choice for modelling the number of fire hotspots. Its consistent performance helps mitigate overfitting, making it a suitable option for this study.

4.1.2. Model Optimisation with Cross-Validation and Hyperparameter Tuning

While setting initial values and applying MML can optimise the marginal likelihood and enhance model accuracy, this approach alone is insufficient to fully address overfitting. To further improve the model, hyperparameter tuning was performed.

Prior to optimisation, a 16-fold cross validation from 16 years train data (2001-2016) technique was applied, partitioning the data by year to create 16 annual data groups. In this process, for the first fold, the data from 2001 were used for testing and the remaining years for training; for the second fold, 2002 data were used for testing, and so on, until each year's data had served as the test set once.

Hyperparameter tuning was then conducted using three optimisation methods: Bayesian optimisation, grid search, and random search. The evaluation results from these methods were compared to identify the best-performing model. The hyperparameter targeted for tuning was the noise variance (error variance), focusing on a low-complexity optimisation approach by adjusting only this single hyperparameter. The results are presented in **table 3**.

Table 3.
Gaussian Process Regression Optimisation Metric Evaluation.

Optimization	RMSE		EVS	
	Train	Test	Train	Test
Bayesian Optimization	700.93	**844.47**	0.805	**0.608**
Grid Search	713.78	853.09	0.798	0.593
Random Search	**694.97**	846.58	**0.808**	0.606

As shown in **table 3**, Bayesian optimisation achieved the best overall performance. It produced a training RMSE of 700.93 and a testing RMSE of 844.47, with corresponding EVS values of 0.805 and 0.608. These results suggest that Bayesian optimisation provides the best balance between training generalisation and the ability to explain data variability, making it a reliable method for predicting fire hotspot numbers.

Random search also demonstrated competitive performance, with the lowest training RMSE (694.97) and similar testing performance (RMSE 846.58, EVS 0.606). Grid search, on the other hand, showed slightly inferior performance compared to the other two methods. Based on these findings, Bayesian optimisation was selected as the most appropriate method for optimising the GPR model in this study, with random search as a viable alternative.

4.2. Long Short-Term Memory

Similar to the GPR approach, Long Short-Term Memory (LSTM) was employed to predict fire hotspots from 2001 to 2020 using the climate factors described in the dataset section. Using a window size of 12 months, the independent variables included all predictor variables from lag t–1 to t–12, while the dependent variable was the number of hotspots at time t. The LSTM model was built using the parameters outlined in t**able 4**.

Table 4.
LSTM Model Parameter.

PARAMETER	SPECIFICATION	DESCRIPTION
Loss function	RMSE	Experimental
Optimizer	Nadam	(Nurdiati, Najib, et al., 2022)
Learning rate	0.001	Experimental
Weight initializer	Glorot Uniform	(Glorot & Bengio, 2010)
Max epoch	500	(Abbasimehr & Paki, 2022)
Patience	25	(Terry et al., 2021)

The model architecture comprised combinations of LSTM layers and dense layers. Variations included one or two LSTM layers (with many-to-one or many-to-many schemes) and between one and four dense layers. The LSTM layers used the tanh and sigmoid activation functions, while the dense layers used ReLU activations.

Training was conducted using backpropagation, with early stopping applied if no improvement in testing loss was observed after 25 epochs. To address potential issues arising from poor weight initialisation by the Glorot Uniform method, the model was trained 20 times. The evaluation results for the LSTM method are shown in **table 5**.

Table 5.
LSTM Model Architecture Evaluation.

Architecture	RMSE		EVS	
	Train	Test	Train	Test
1 LSTM 1 Dense	725.02	899.17	0.808	0.521
1 LSTM 2 Dense	506.05	911.31	0.905	0.495
1 LSTM 3 Dense	382.48	646.61	0.945	0.746
1 LSTM 4 Dense	389.97	799.79	0.943	0.620
2 LSTM 1 Dense	825.82	862.77	0.755	0.544
2 LSTM 2 Dense	399.06	828.62	0.938	0.580
2 LSTM 3 Dense	678.61	811.47	0.826	0.613
2 LSTM 4 Dense	**343.01**	**522.12**	**0.956**	**0.834**

The evaluation results in **table 5** show that the "2 LSTM 4 Dense" architecture achieved the best performance among all variations tested. It recorded the lowest RMSE on the testing dataset (522.12) and the highest EVS (0.834), indicating excellent accuracy in predicting the number of fire hotspots and explaining a substantial portion of data variability.

Moreover, it demonstrated strong performance on the training dataset (RMSE 343.01, EVS 0.956), suggesting its capability to capture historical data patterns effectively. Conversely, simpler architectures such as "1 LSTM 1 Dense" exhibited significantly lower performance, with a high testing RMSE (899.17) and a low EVS (0.521), indicating insufficient ability to capture the complexity of the temporal data. Although architectures with more dense layers generally showed performance improvements, these results highlight the importance of using additional LSTM layers to better capture complex temporal relationships. Thus, the "2 LSTM 4 Dense" architecture was identified as the optimal choice for predicting fire hotspot numbers based on climatic factors in this study.

4.3. Best Model

The best-performing models identified in the previous sections—GPR with Bayesian optimisation and LSTM with a two-layer LSTM and four dense layers architecture—were compared against the actual data, as shown in f**igure 3**.

Figure 3 visualises the comparison between the prediction results of the two models, where (a) GPR predictions are shown in red and (b) LSTM predictions in blue, against the actual data. It should be noted that the LSTM model utilises lagged inputs; thus, predicting at time *t* requires data from *t-12* to *t-1*, which results in no predictions being made for the year 2001.

It is evident that the LSTM model outperforms GPR in predicting the number of fire hotspots, both on training and testing datasets, particularly during spike periods in the dry season. The GPR model tends to underestimate the peak number of hotspots during major spike years in the training data, notably in 2002, 2004, 2006, 2009, and 2015. In contrast, the LSTM model successfully captures these peak events within the training period.

However, both models demonstrate limitations in predicting peak hotspot numbers within the testing data. GPR notably overestimated the number of hotspots in 2018, while LSTM overestimated in 2017 and underestimated in 2019, although it performed reasonably well in predicting the hotspot count for 2018. Previous research (Gunadi et al., 2019) indicated that the decline in hotspot numbers in 2018 was due to efforts aimed at supporting the Asian Games held in Indonesia.

Furthermore, study (Nugrahani et al., 2024) reported that four different models tended to overestimate hotspot occurrences during that year, consistent with the overestimation observed in the GPR model. This discrepancy arises because machine learning models primarily rely on historical patterns for predictions and typically do not account for real-world socioeconomic factors. Nevertheless, the LSTM model, by incorporating temporal lags, appears to capture patterns more closely related to recent data preceding the prediction period.

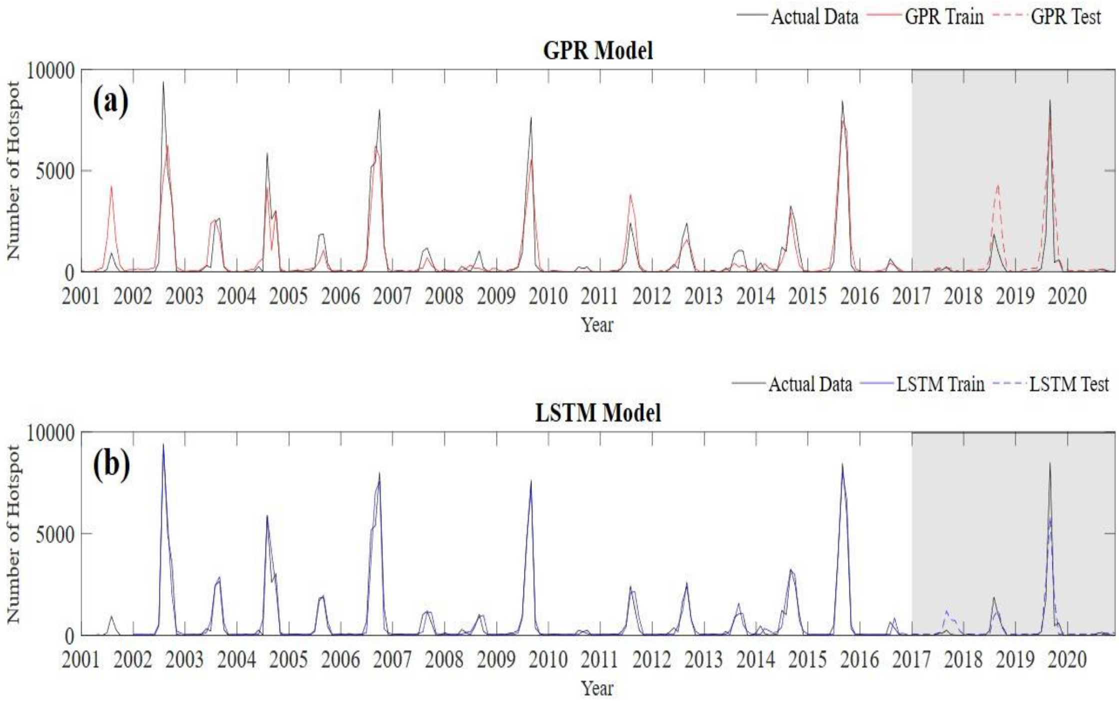


Fig. 3. Prediction comparison between GPR and LSTM.

5. DISCUSSION

The comparative analysis between Gaussian Process Regression (GPR) and Long Short-Term Memory (LSTM) models in this study reveals insightful findings on the challenges and strengths of each method when applied to the prediction of forest fire hotspots in Kalimantan using climate factors. Notably, the LSTM architecture with two LSTM layers and four dense layers produced the most accurate results, achieving a testing RMSE of 522.12 and an EVS of 0.834. This performance highlights LSTM's superior ability to capture complex temporal dependencies, especially in climate-driven phenomena that are seasonal and nonlinear.

This finding supports Kadir et al. (2023), who also found LSTM effective for hotspot forecasting in Indonesia, achieving an average error of only 6.94% even when using simpler models based solely on past hotspot counts. The present study builds upon that by integrating six climate predictors—including rainfall, dry spells, ENSO, and IOD—thus expanding the model's explanatory power. The enhanced accuracy confirms that including relevant climate variables strengthens LSTM's generalization, especially when paired with effective regularization techniques such as dropout and early stopping.

In contrast, GPR demonstrated solid performance during training but experienced a notable decline in accuracy on test data, particularly when complex kernels like ARD Matern32 were used. This suggests an overfitting issue, a phenomenon also discussed by Kamath et al. (2018), who found GPR highly accurate but sensitive to hyperparameter settings and prone to overfitting in high-complexity domains. Although GPR showed some generalization improvement with Bayesian optimization, its testing RMSE (844.47) and EVS (0.608) remained well below that of the best-performing LSTM model.

Similar concerns were raised in Hultquist et al. (2014), where GPR was outperformed by Random Forests in predicting burn severity due to its relative inflexibility with noisy and high-dimensional remote sensing data. These findings collectively suggest that while GPR has strong theoretical appeal—especially with its uncertainty quantification and Bayesian foundation—it may fall short in real-world applications involving dynamic temporal and multivariate data like wildfire hotspot prediction.

Luo et al. (2024) further validate the LSTM architecture's strength, demonstrating its capacity to forecast methane emissions with high fidelity in multivariate time series scenarios. Their findings align with this study's conclusion that LSTM can effectively model complex environmental systems where variables are interdependent and evolve over time. The success of deeper architectures (e.g., 2 LSTM + 4 dense layers) indicates that the model's depth and capacity play a crucial role in capturing spatiotemporal nuances in climate data.

However, one consistent limitation across studies—including this one—is the inability to predict sociopolitical anomalies, such as the significant decrease in hotspots in 2018 due to government efforts to reduce fires during the Asian Games. Gunadi et al. (2019) highlight the critical role of policy in shaping fire outcomes, suggesting that models based purely on environmental predictors will struggle to account for such interventions. This was evident when GPR overestimated and LSTM underestimated fire counts for certain years—both reflecting the models' reliance on historical climatic patterns without sociopolitical context. The example of South Africa (Madondo et al., 2023), regionally different, but it supports the claim that socio-economic and governance interventions can create anomalies versus climate-only predictors.

In conclusion, the study underscores the potential of LSTM for operational forest fire forecasting, especially when enriched with relevant climatic inputs. Meanwhile, GPR remains valuable for theoretical modeling and applications requiring probabilistic interpretation. Future research should explore hybrid modeling approaches that combine the temporal learning strengths of LSTM with the uncertainty quantification of GPR, while also integrating socioeconomic variables to capture the full spectrum of fire dynamics in Kalimantan and similar fire-prone regions.

6. CONCLUSION

This study successfully demonstrated that both Gaussian Process Regression (GPR) and Long Short-Term Memory (LSTM) models can be employed to predict the number of fire hotspots in Kalimantan based on climate factors. Through model development and evaluation, the LSTM model with two LSTM layers and four dense layers achieved superior performance, with a testing RMSE of 522.12 and an EVS of 0.834, indicating a strong capability to capture temporal patterns in climate and fire data. The GPR model, although effective with Bayesian optimization, was less capable of handling the complex temporal dependencies compared to LSTM. The research underscores the importance of selecting models that can adapt to the temporal dynamics inherent in climate-induced fire risks. Future studies may explore integrating socio-economic variables and higher-resolution climate forecasts to further enhance the predictive performance and operational applicability of fire hotspot prediction models in Kalimantan and similar tropical regions.

REFERENCES

Abbasimehr, H., & Paki, R. (2022). Improving time series forecasting using LSTM and attention models. *Journal of Ambient Intelligence and Humanized Computing*, *13*(1), 673–691. https://doi.org/10.1007/s12652-020-02761-x

Anggraini, N., & Trisakti, B. (2011). Kajian dampak perubahan iklim terhadap kebakaran hutan dan deforestasi di provinsi Kalimantan Barat. *Jurnal Penginderaan Jauh Dan Pengolahan Data Citra Digital*, *8*.

Ardiyani, E., Nurdiati, S., Sopaheluwakan, A., Septiawan, P., & Najib, M. K. (2023). Probabilistic Hotspot Prediction Model Based on Bayesian Inference Using Precipitation, Relative Dry Spells, ENSO and IOD. *Atmosphere*. https://doi.org/10.3390/atmos14020286

Bramawanto, R., & Abida, R. F. (2017). Tinjauan Aspek Klimatologi (ENSO dan IOD) dan Dampaknya Terhadap Produksi Garam Indonesia. *Jurnal Kelautan Nasional*. https://doi.org/10.15578/jkn.v12i2.6061

DiPietro, R., & Hager, G. D. (2019). Deep learning: RNNs and LSTM. In *Handbook of Medical Image Computing and Computer Assisted Intervention*. https://doi.org/10.1016/B978-0-12-816176-0.00026-0

Eaturu, A., Vadrevu, K.P. (2025). Evaluation of machine learning and deep learning algorithms for fire prediction in Southeast Asia. *Sci Rep* **15**, 18807. https://doi.org/10.1038/s41598-025-00628-9

Foley, E. (2024). Leveraging Gaussian Processes in Remote Sensing. *Energies*, *17*(16). https://doi.org/10.3390/en17163895

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research*.

Gunadi, A., Gunardi, G., & Martono, M. (2019). THE LAW OF FOREST IN INDONESIA: PREVENTION AND SUPPRESSION OF FOREST FIRES. *Bina Hukum Lingkungan*. https://doi.org/10.24970/bhl.v4i1.86

Gusnita, D. (2021). Impact of Forest Fires in Sumatra and Kalimantan to Atmospheric Pollution During Period Of 2010-2015. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*. https://doi.org/10.20961/jkpk.v6i1.35027

Haidu, I., Magyari-Sáska, Z., & Magyari-Sáska, A. (2025). Spatio-Temporal Gap Filling of Sentinel-2 NDI45 Data Using a Variance-Weighted Kalman Filter and LSTM Ensemble. *Sensors*, *25*(17). https://doi.org/10.3390/s25175299

Hultquist, C., Chen, G., & Zhao, K. (2014). A comparison of Gaussian process regression, random forests and support vector regression for burn severity assessment in diseased forests. *Remote Sensing Letters*. https://doi.org/10.1080/2150704X.2014.963733

Irwandi, H., Syamsu Rosid, M., & Mart, T. (2019). Identification of the El Niño Effect on Lake Toba's Water Level Variation. *IOP Conference Series: Earth and Environmental Science*, *406*(1). https://doi.org/10.1088/1755-1315/406/1/012022

E. A. Kadir, H. T. Kung, S. L. Rosa, A. Sabot, M. Othman and M. Ting, "Forecasting of Fires Hotspot in Tropical Region Using LSTM Algorithm Based on Satellite Data," *2022 IEEE Region 10 Symposium (TENSYMP)*, Mumbai, India, 2022, pp. 1-7, doi: 10.1109/TENSYMP54529.2022.9864407.

Kadir, E. A., Kung, H. T., Rosa, S. L., Sabot, A., Othman, M. & Ting, M. (2022). Forecasting of Fires Hotspot in Tropical Region Using LSTM Algorithm Based on Satellite Data. *IEEE Region 10 Symposium (TENSYMP)*, Mumbai, India, pp. 1-7, doi: 10.1109/TENSYMP54529.2022.9864407.

Kadir, E. A., Kung, H. T., AlMansour, A. A., Irie, H., Rosa, S. L., & Fauzi, S. S. M. (2023). Wildfire Hotspots Forecasting and Mapping for Environmental Monitoring Based on the Long Short-Term Memory Networks Deep Learning Algorithm. *Environments - MDPI*. https://doi.org/10.3390/environments10070124

Kamath, A., Vargas-Hernández, R. A., Krems, R. V., Carrington, T., & Manzhos, S. (2018). Neural networks vs Gaussian process regression for representing potential energy surfaces: A comparative study of fit quality and vibrational spectrum accuracy. *Journal of Chemical Physics*. https://doi.org/10.1063/1.5003074

La Fata, A., Moser, G., Procopio, R., Bernardi, M., & Fiori, E. (2024). A Gaussian Process Regression Method to Nowcast Cloud-to-Ground Lightning from Remote Sensing and Numerical Weather Modeling Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1–31. https://doi.org/10.1109/JSTARS.2024.3501976

Li, J., Huang, D., Chen, C., Liu, Y., Wang, J., Shao, Y., Wang, A., & Li, X. (2024). Prediction of Forest-Fire Occurrence in Eastern China Utilizing Deep Learning and Spatial Analysis. *Forests*, *15*(9), 1672. https://doi.org/10.3390/f15091672

Li, Z., Hong, X., Kuangrong, H., Lei, C., & Biao, H. (2020). Gaussian Process Regression with heteroscedastic noises – A machine-learning predictive variance approach. *Chemical Engineering Research and Design*, *157*. https://doi.org/10.1016/j.cherd.2020.02.033

Listia Rosa, S., Abdul Kadir, E., Syukur, A., Irie, H., Wandri, R. & Fikri Evizal, M. (2022). Fire Hotspots Mapping and Forecasting in Indonesia Using Deep Learning Algorithm. *3rd International Conference on Electrical Engineering and Informatics (ICon EEI)*, Pekanbaru, Indonesia, pp. 190-194, doi: 10.1109/IConEEI55709.2022.9972281.

Luo, R., Wang, J., & Gates, I. (2024). Forecasting Methane Data Using Multivariate Long Short-Term Memory Neural Networks. *Environmental Modeling and Assessment*. https://doi.org/10.1007/s10666-024-09957-x

MacKinnon, K., & Hatta, G. (2013). *Ecology of Kalimantan: Indonesian Borneo*. Tuttle Publishing.

Madondo, R., Tandlich, R., Stoch, E., Restas, A., & Shwababa, S. (2023). Quantitative and Qualitative Data in Disaster Risk Management of Fires: A Case Study from South Africa at Various Geographical Levels. *Geographia Technica*, *18*(2), 14–39. https://doi.org/10.21163/GT_2023.182.02

Najib, M. K., Nurdiati, S., & Sopaheluwakan, A. (2021). Quantifying the joint distribution of drought indicators in Borneo fire-prone area. *IOP Conference Series: Earth and Environmental Science*. https://doi.org/10.1088/1755-1315/880/1/012002

Najib, M. K., Nurdiati, S., & Sopaheluwakan, A. (2022). Multivariate fire risk models using copula regression in Kalimantan, Indonesia. *Natural Hazards*. https://doi.org/10.1007/s11069-022-05346-3

Noh, S. H. (2021). Analysis of gradient vanishing of RNNs and performance comparison. *Information (Switzerland)*. https://doi.org/10.3390/info12110442

Nugrahani, E. H., Nurdiati, S., Bukhari, F., Najib, M. K., Sebastian, D. M., & Nur Fallahi, P. A. (2024). Sensitivity and feature importance of climate factors and evaluation of different machine learning models for predicting fire hotspots in Kalimantan, Indonesia. *IAES Int. J. Artif. Intell. (IJ-AI)*, *13*(2), 2212. https://doi.org/10.11591/ijai.v13.i2.pp2212-2225

Nurdiati, S., Bukhari, F., Julianto, M. T., Sopaheluwakan, A., Aprilia, M., Fajar, I., Septiawan, P., & Najib, M. K. (2022). The impact of El Niño southern oscillation and Indian Ocean Dipole on the burned area in Indonesia. *Terrestrial, Atmospheric and Oceanic Sciences*. https://doi.org/10.1007/S44195-022-00016-0

Nurdiati, S., Bukhari, F., Sopaheluwakan, A., Septiawan, P., & Hutapea, V. (2024). ENSO and IOD impact analysis of extreme climate condition in Papua, Indonesia. *Geographia Technica*, *19*(1), 1–18. https://doi.org/10.21163/gt_2024.191.01

Nurdiati, S., Najib, M., Bukhari, F., Revina, R., & Salsabila, F. (2022). PERFORMANCE COMPARISON OF GRADIENT-BASED CONVOLUTIONAL NEURAL NETWORK OPTIMIZERS FOR FACIAL EXPRESSION RECOGNITION. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *16*(3), 927–938. https://doi.org/10.30598/barekengvol16iss3pp927-938

Nurdiati, S., Sopaheluwakan, A., Julianto, M. T., Septiawan, P., & Rohimahastuti, F. (2022). Modelling and analysis impact of El Nino and IOD to land and forest fire using polynomial and generalized logistic function: cases study in South Sumatra and Kalimantan, Indonesia. *Modeling Earth Systems and Environment*. https://doi.org/10.1007/s40808-021-01303-4

Oyedele, A. A., Ajayi, A. O., Oyedele, L. O., Bello, S. A., & Jimoh, K. O. (2023). Performance evaluation of deep learning and boosted trees for cryptocurrency closing price prediction. *Expert Systems with Applications*, *213*, 119233. https://doi.org/https://doi.org/10.1016/j.eswa.2022.119233

Putra Mulia, Nofrizal, & Dewi, W. N. (2021). Analisis Dampak Kabut Asap Karhutla Terhadap Gangguan Kesehatan Fisik Dan Mental. *HEALTH CARE : JURNAL KESEHATAN*. https://doi.org/10.36763/healthcare.v10i1.103

Qirom, M. A., Rachmanadi, D., Lestari, F., & Andriani, S. (2022). Forest structure change after forest fire in peatland of Central Kalimantan. *IOP Conference Series: Earth and Environmental Science*. https://doi.org/10.1088/1755-1315/1115/1/012019

Rafhida, S. A., Nurdiati, S., Budiarti, R., & Najib, M. K. (2024). Bias correction of lake Toba rainfall data using quantile delta mapping. *CAUCHY*, *9*(2), 297–309. https://doi.org/10.18860/ca.v9i2.29124.

Rasmussen, C. E., & Williams, C. K. I. (2019). *Gaussian processes for machine learning*. MIT Press.

Reddy, C. S., Bird, N. G., Sreelakshmi, S., Manikandan, T. M., Asra, M., Krishna, P. H., Jha, C. S., Rao, P. V. N., & Diwakar, P. G. (2019). Identification and characterization of spatio-temporal hotspots of forest fires in South Asia. *Environmental Monitoring and Assessment*. https://doi.org/10.1007/s10661-019-7695-6

Saharjo, B. H., & Hasanah, U. (2023). Analisis Faktor Penyebab Terjadinya Kebakaran Hutan dan Lahan di Kabupaten Pulang Pisau, Kalimantan Tengan. *Journal of Tropical Silviculture*. https://doi.org/10.29244/j-siltrop.14.01.25-29

Saharjo, B. H., & Velicia, W. A. (2018). The Role of Rainfall Towards Forest and Land Fires Hotspot Reduction in Four Districs in Indonesia on 2015-2016. *Journal of Tropical Silviculture*. https://doi.org/10.29244/j-siltrop.9.1.24-30

Sari, R., Trihardianingsih, L., Firdaus Mulya, R., Arief, M. I., & Kusrini, K. (2022). Analisis Index Vegetation Wilayah Terdampak Kebakaran Hutan Riau Menggunakan Citra Landsat-8 dan Sentinel-2. *CogITo Smart Journal*. https://doi.org/10.31154/cogito.v8i2.439.282-294

Sarmiasih, M., & Pratama, P. Y. (2019). The Problematics Mitigation of Forest and Land Fire District Kerhutla) in Policy Perspective (A Case Study : Kalimantan and Sumatra in Period 2015-2019). *Journal of Governance and Public Policy*. https://doi.org/10.18196/jgpp.63113

Sudrajat, A. S. E., & Subekti, S. (2019). Pengelolaan Ekosistem Gambut Sebagai Upaya Mitigasi Perubahan Iklim Di Provinsi Kalimantan Selatan. *Jurnal Planologi*. https://doi.org/10.30659/jpsa.v16i2.4459

Terry, J. K., Jayakumar, M., & Alwis, K. De. (2021). Statistically Significant Stopping of Neural Network Training. *CoRR*, *abs/2103.0*. https://arxiv.org/abs/2103.01205

Ulya, N. A., & Yunardy, S. (2006). Analisis Dampak Kebakaran Hutan di Indonesia Terhadap Distribusi Pendapatan Masyarakat. *Jurnal Penelitian Sosial Dan Ekonomi Kehutanan*, *3*(2), 133–146. https://doi.org/10.20886/jpsek.2006.3.2.133-146

Van Der Werf, G. R., Randerson, J. T., Giglio, L., Van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., Van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J., & Kasibhatla, P. S. (2017). Global fire emissions estimates during 1997-2016. In *Earth System Science Data*. https://doi.org/10.5194/essd-9-697-2017

Wasis, B., Saharjo, B. H., & Waldi, R. D. (2019). Dampak Kebakaran Hutan Terhadap Flora Dan Sifat Tanah Mineral Di Kawasan Hutan Kabupaten Pelalawan Provinsi Riau. *Journal of Tropical Silviculture*. https://doi.org/10.29244/j-siltrop.10.1.40-44

Ye, W., Alawieh, M. B., Li, M., Lin, Y., & Pan, D. Z. (2019). Litho-GPA: Gaussian Process Assurance for Lithography Hotspot Detection. *Proceedings of the 2019 Design, Automation and Test in Europe Conference and Exhibition, DATE 2019*. https://doi.org/10.23919/DATE.2019.8714960

Yuliarti, A., & Anggraini, R. N. (2022). Pengembangan Strategi Pengurangan Risiko Kebakaran Gambut Dalam Bingkai Media Berdasarkan Jumlah Hotspot Menggunakan S-Npp Viirs. *PROSIDING SEMINAR*

AN INTEGRATED ANN-CA MODEL FOR LAND USE CHANGE PREDICTION AND FLOOD RISK MITIGATION: A CASE STUDY OF ENREKANG REGENCY, INDONESIA

Uca SIDENG[1*], Nurul Afdal HARIS[1] and Mustari S. LAMADA[2]

DOI: 10.21163/GT_2026.212.03

ABSTRACT

Flood disasters in Enrekang Regency, South Sulawesi Province, have caused significant material losses and disrupted community activities due to the region's unique geographical characteristics with undulating and mountainous topography. The Saddang River, as one of the main rivers in South Sulawesi, flows through this area, making it highly vulnerable to flooding, especially during the rainy season. Rapid land cover changes due to human activities such as settlement expansion, agriculture, and deforestation have increasingly elevated flood risks. Land conversion from forests to agricultural and settlement areas reduces water absorption capacity and increases surface runoff. Currently, flood management in Enrekang Regency remains reactive, with budgets allocated more for post-disaster response than for mitigation and prevention. This research develops an integrated Artificial Neural Network Cellular Automata (ANN-CA) model for land use change prediction and flood risk mitigation. The model integrates remote sensing technology, ANN-CA modeling, and Geographic Information Systems (GIS) to predict future land use changes and identify high flood-risk areas. The methodology involves satellite image acquisition (2010-2020), land cover change extraction, ANN training, CA configuration, model validation (Accuracy >85%, Kappa >0.8), and integration with flood risk factors. Results show that the model can effectively predict land use changes with high accuracy, providing valuable spatial information for flood mitigation planning. The predicted land use map for 2030 indicates significant expansion of built-up areas in flood-prone zones, necessitating immediate policy interventions. This research contributes to the development of predictive and preventive flood management approaches, offering a scientific basis for spatial planning and disaster risk reduction in mountainous regions.

Keywords: *Artificial Neural Network; Land use change prediction; Flood risk mitigation; Remote sensing; Spatial analysis.*

1. INTRODUCTION

1.1. Background

Enrekang Regency, located in South Sulawesi Province, Indonesia, possesses unique geographical characteristics with undulating and mountainous topography (Uca et al., 2023, 2018). The region is traversed by the Saddang River, one of the major rivers in South Sulawesi, making it particularly vulnerable to flooding during rainy seasons (Rachmayanti et al., 2022; Uca et al., 2021). Historical data indicates that floods in Enrekang Regency have caused substantial material losses and disrupted community activities, affecting both urban and rural areas. Rapid land cover changes resulting from human activities such as settlement expansion, agricultural development, and deforestation have increasingly exacerbated flood risks (Uca et al., 2018). The conversion of forest land to agricultural and settlement areas reduces water absorption capacity (Farhan et al., 2024) and increases surface runoff, amplifying flood potential (Nugraheni et al., 2022). The agrarian crisis and ecological disasters in the Latimojong Mountains, which fall within Enrekang Regency, have further worsened this condition.

---

[1] *Geography Department, Faculty of Mathematics and Natural Science, State University of Makassar, Makassar, Indonesia; * Corresponding author (US) ucasideng@unm.ac.id, (NAH) nurul.afdal.haris@unm.ac.id*

[2] *Technic Information, Faculty of Technic, State University of Makassar, Makassar, Indonesia; (MSL) mustarilamada@unm.ac.id*

Current flood management approaches in Enrekang Regency remain predominantly reactive, with budget allocations favoring post-disaster response over mitigation and prevention. This approach is inefficient both economically and socially, highlighting the need for innovative predictive and preventive solutions (Rajeev and Singh, 2016). Recent advances in geospatial technologies and modeling approaches offer new opportunities for flood risk mitigation (Sudiana et al., 2025). The integration of remote sensing, Artificial Neural Network Cellular Automata (ANN-CA), and Geographic Information Systems (GIS) has shown promise in predicting land use changes and assessing flood risks (Xu et al., 2021; Yang et al., 2025). These technologies enable the development of comprehensive models that can simulate future land use scenarios and identify areas at high risk of flooding (Singh et al., 2021). Land Use Change (LUC) modeling has evolved from conventional statistical methods to the utilization of advanced Machine Learning (ML) and Deep Learning (DL) algorithms. Alternative approaches, such as Random Forest (RF) and Support Vector Machine (SVM), have proven highly effective for LUC classification and prediction, often exhibiting superior classification accuracy compared to traditional methods (Asif et al., 2023; Mutale et al., 2024). Furthermore, Deep Learning models like Convolutional Neural Network (CNN), and hybrid models such as CNN-LSTM, offer advanced capabilities in extracting hierarchical spatio-temporal features, making them robust for dynamic LUC prediction and capturing complex non-linear relationships (Lei et al., 2025; Varma et al., 2024).

However, while these ML/DL models excel in pattern recognition and quantity prediction, they inherently face challenges in simulating explicit spatial processes, cellular interactions, and directly integrating external policy constraints or scenarios into the transition mechanism. To overcome these limitations and provide an application-oriented framework suitable for policy intervention, this research selects the integrated Artificial Neural Network‑Cellular Automata (ANN-CA) Model.

ANN-CA provides an optimal hybrid solution: the ANN effectively maps the non-linear relationship between various driving factors (e.g., topography, hydrology, and socio-economics) and LUC transition probabilities, while the Cellular Automata (CA) component is uniquely capable of applying localized transition rules, neighborhood effects, and crucial policy/scenario constraints at the pixel level (Khan and Khan, 2025; Tharik et al., 2025). This integration is vital as it allows the model to not only predict what changes but also to simulate where those changes occur under specific planning conditions, a capability essential for scenario-based flood mitigation and proactive spatial planning in the mountainous context of Enrekang Regency.

This research addresses the critical gap in flood management by developing an integrated ANN-CA model for land use change prediction specifically tailored for flood mitigation in Enrekang Regency. Current flood management practices often lack the predictive capabilities needed for proactive risk reduction. Therefore, the model aims to provide decision-makers with robust spatial information and predictive insights to support proactive flood risk reduction strategies. The subsequent sections will provide the necessary background, critically reviewing existing research and debates relevant to integrated land use modeling and flood risk, thereby highlighting the gaps that this study aims to fill and setting the stage for the research objectives.

1.2. Research Gap

While numerous studies have addressed Land Use Change (LUC) modeling and flood risk assessment independently, there remains a significant limitation in their integration, especially for mitigation purposes in complex mountainous regions such as Enrekang Regency (Gabriels et al., 2022; Iskandar and Ridzuan, 2022; Merten et al., 2020). Previous studies have often focused on urban flood risk assessment or general land use change prediction without specific consideration for flood mitigation applications. This research specifically addresses three critical gaps: (1) The limited integration of the Artificial Neural Network‑Cellular Automata (ANN-CA) model with flood risk factors in challenging geographical contexts. (2) The insufficient comprehensive validation of predictive LUC models for flood risk applications in developing countries. (3) The lack of translating predictive modeling outcomes into actionable spatial policy recommendations that local governments can readily implement.

1.3. Research Objective

The main objective of this research is to develop and validate an integrated ANN-CA model for predicting land use changes up to the year 2030 in Enrekang Regency, with explicit consideration of flood risk factors. The resulting predictions will be utilized to formulate spatial recommendations that support flood mitigation efforts. The significant contribution of this study is providing a predictive and preventive scientific basis for spatial planning and disaster risk reduction. By integrating advanced computational modeling with disaster risk assessment in a vulnerable region, this research offers valuable insights for more effective and sustainable decision-making.

2. STUDY AREA

Enrekang Regency is located in South Sulawesi Province, Indonesia, between 3°14’ - 3°50’ South Latitude and 119°40’ - 120°06’ East Longitude (**Fig. 1.**). The regency covers an area of approximately 1,956.28 km² with a population of around 250,000 people.

The topography of Enrekang Regency is characterized by mountainous terrain with elevations ranging from 50 to 3,478 meters above sea level. The Saddang River and its tributaries form the main drainage system, flowing through the regency from north to south. The region experiences a tropical climate with average annual rainfall ranging from 2,000 to 3,000 mm, concentrated mainly in the rainy season (November to April).

Land use in Enrekang Regency is dominated by agricultural land (45%), forest areas (30%), settlement areas (10%), and other uses (15%) (Uca et al., 2023). Recent decades have seen significant land use changes, particularly the conversion of forest land to agricultural and settlement areas, contributing to increased flood risks.

Enrekang Regency was selected as a case study because it represents critical and complex conditions: steep mountainous topography, a sensitive watershed system (Saddang River), and rapid land use change (LUC). The performance of the integrated ANN-CA model in this region is strongly influenced by key local parameters trained as network inputs: (1) Elevation and Slope, which constrain LUC and accelerate surface runoff; (2) Proximity to the Saddang River, which drives settlement expansion into risk zones; and (3) Accessibility, which is a main driver of development patterns. This approach is generalizable to other mountainous regions with watersheds sensitive to LUC changes (e.g., critical watersheds in Indonesia). This generalization requires recalibration of the ANN input weights and CA transition rules to match the specific LUC driving parameters of the new location (e.g., soil type or local spatial planning policies). The model offers a flexible framework that necessitates local adjustment for accurate results.

3. DATA AND METHODS

3.1. Remote Sensing Data

Satellite imagery from multiple sources was collected to analyze land use changes in Enrekang Regency:

1) Landsat Imagery: Landsat 5 Thematic Mapper (TM) (2010), Landsat 7 Enhanced Thematic Mapper Plus (ETM+) (2015), and Landsat 8 Operational Land Imager (OLI) (2020), all with a 30-meter spatial resolution. These datasets were accessed from the United States Geological Survey (USGS) Earth Explorer archive, an open-access repository.
2) Sentinel-2 Imagery: High-resolution data from Sentinel-2A and Sentinel-2B missions (2020) with a 10-meter spatial resolution for visible and near-infrared bands, obtained from the European Space Agency's Copernicus Open Access Hub. This dataset was used to improve classification accuracy in the 2020 land cover mapping. Sentinel-2 data was used as a high-resolution reference to improve the classification accuracy of our 2020 Land Use Map (as the historical end year), thereby increasing the reliability of the transition matrix. However, 30m resolution was maintained as the operational resolution of the ANN-CA Model to ensure long-term spatial consistency with Landsat data (2010, 2015) and the 30m DEM.

3) Digital Elevation Model (DEM): The Shuttle Radar Topography Mission (SRTM) 30-meter DEM was used for topographic analysis, including slope and elevation modeling. SRTM data is an open-source dataset developed by NASA and available through USGS Earth Explorer.

All satellite images were subjected to comprehensive pre-processing to ensure data consistency and analytical accuracy. This included radiometric calibration, atmospheric correction (using Sen2Cor for Sentinel-2 and FLAASH/LEDAPS for Landsat), and geometric correction using ground control points and TRIM reference data. The integration of these open datasets aligns with the recommended use of globally accessible geospatial resources such as those endorsed by open science initiatives and Earth observation programs.

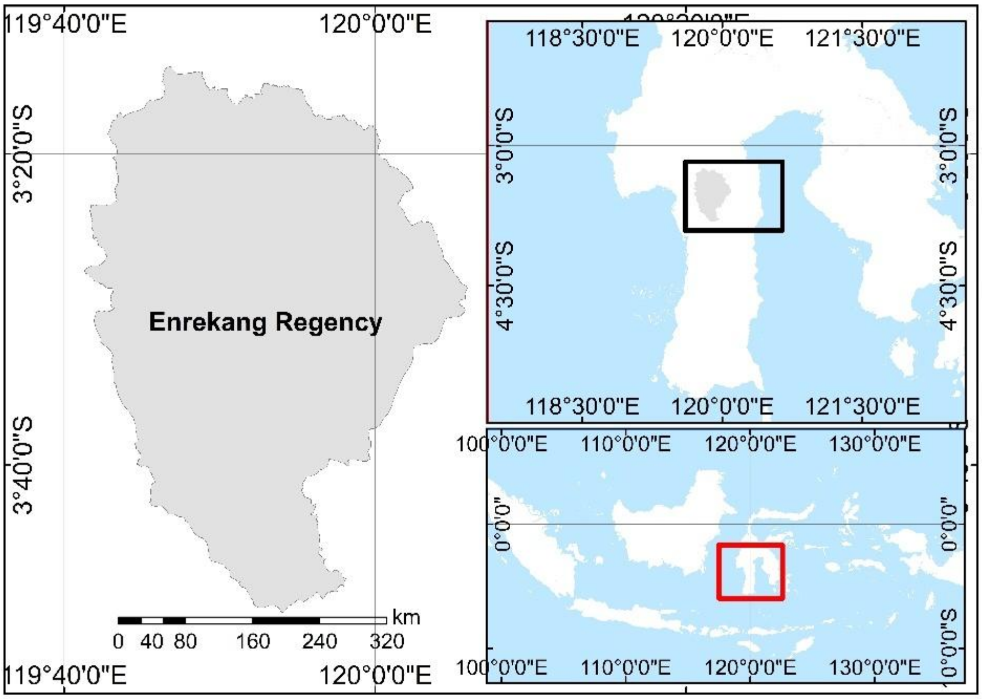


Fig. 1. Research Location Map (Enrekang Regency, South Sulawesi, Indonesia).

3.2. Ancillary Data

To support the modeling process, various additional data layers were collected (**Tabel 1**). These included Topographic Data (such as slope, aspect, and curvature derived from the DEM), Hydrological Data (comprising river networks, watershed boundaries (**Fig. 2**), and flood history records), and Infrastructure Data (covering road networks, settlement locations, and administrative boundaries). Furthermore, Socio-economic Data was gathered, including statistics on population density, land values, and agricultural productivity (BPS Enrekang, 2025, 2021, 2016, 2011). All collected data were subsequently processed and integrated into a consistent spatial reference system, namely WGS 84 UTM Zone 50S, and resampled to a common spatial resolution of 30m.

3.3. Land Use Classification

3.3.1. Land Use Classification and Feature Exztraction

Land use classification was conducted using a supervised classification approach which involved several steps (**Fig. 3**). First, Training Sample Collection was performed by gathering ground truth points through field surveys and interpretation of high-resolution imagery. The classification itself was executed using the Support Vector Machine (SVM) algorithm for its known effectiveness in handling complex spectral signatures.

Table 1.

Data Used.

Data	Specification	Source	Objective
Landsat Imagery	Landsat 5 Thematic Mapper (TM) (2010) Landsat 7 Enhanced Thematic Mapper Plus (ETM+) (2015) Landsat 8 Operational Land Imager (OLI) 2020).All in 30m resolution with <10% cloude cover	USGS EarthExplorer https://earthexplorer. usgs.gov	Historical Input Data and Multi-temporal Feature Extraction. Used to: 1) Map Land Use in 2010 and 2015 (to train the time-series model). 2) Calculate the LUC Transition Matrix (a key multi-temporal feature). 3) Set the model to a 30m operational resolution for long-term consistency.
Sentinel-2 Imagery	S2A Mission with 10m resolution and <10% cloud cover	Copernicus Browser https://browser.dataspace.copernicus.eu	Classification Accuracy Improvement. Used to: 1) Improve the accuracy and detail of the 2020 Land Use Map (as a validated final year baseline map). 2) Ensure the most up-to-date and accurate LUC classification as input to the Transition Matrix.
Digital Elevation Model (DEM)	USGS EROS Archive - Digital Elevation - Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global	USGS EarthExplorer https://earthexplorer .usgs.gov	Extraction of Topographic Driving Factors. Used to: 1) Provide baseline elevation data. 2) Serve as a basis for generating driving factors (such as slope and aspect) that will be used as input variables for the ANN Model.
Topographic Data		Extraction from DEM data	LULC Driving Factors. Used as independent input variables in ANN training to model how elevation, slope, and aspect affect the probability of land use change.
Hydrological Data		Extraction from DEM data	Water-related LULC Driving Factors. Used as independent input variables in ANN training to model the influence of hydrology (e.g., distance from rivers, drainage density) on land use change, which is important for flood risk.
Infrastructure Data	30m resoution and 10m resolution data extraction 1:50000 Scale for Vector data	Integrated data of Classification Map and Vector data from Ina-Geoportal https://tanahair.indonesia. go.id	Anthropogenic/Policy Driving Factors. Used to: 1) Measure proximity to roads and settlements (as driving factors for human activity). 2) Integrate spatial constraints or CA rules based on infrastructure location.
Socio-economic Data	Statistical Data for each region	Book Report of Badan Pusat Statistik (Central Bureau of Statistics) 2010, 2015, 2020, and 2025	Non-Spatial Driving Factors and Regional Scale Validation. Used to: 1) Provide non-spatial variables (e.g., population density, economic growth) that may influence the LUC transition probability. 2) Control the total macro growth of the study area in the CA Model simulation.

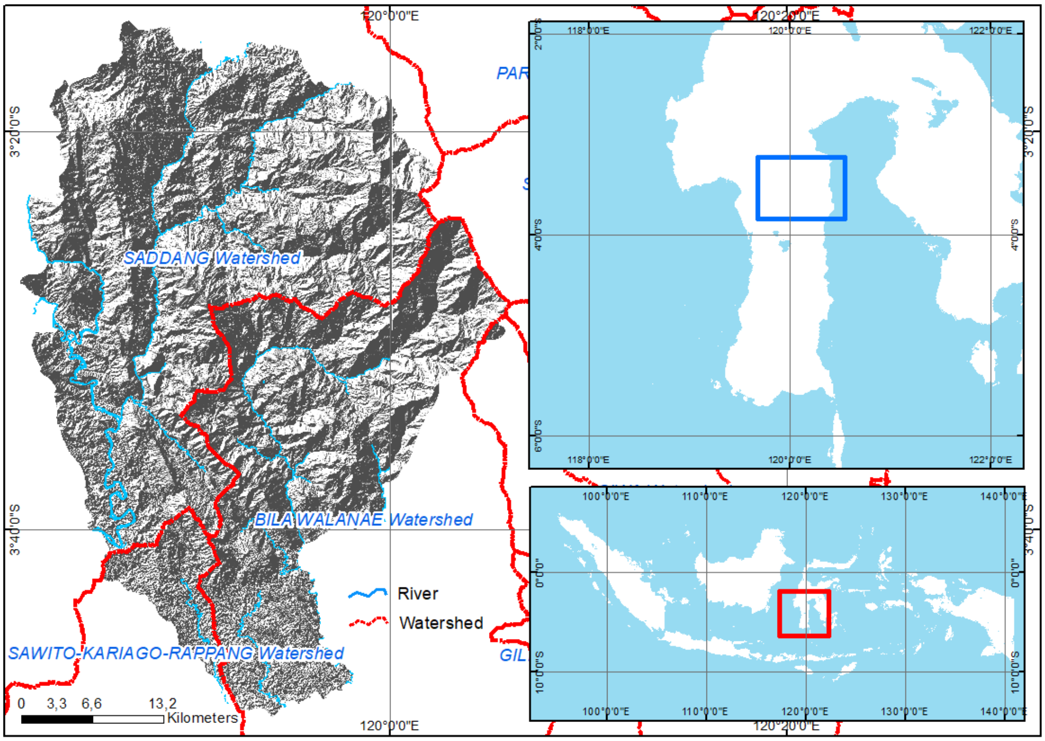


Fig. 2. The DEM/Hillshade and Watersheds within Enrekang Regency.

For the classification model, we used six standard Landsat feature bands: Blue (B2), Green (B3), Red (B4), Near Infrared (NIR, B5), Shortwave Infrared 1 (SWIR1, B6), and Shortwave Infrared 2 (SWIR2, B7). In addition, we included two derived spectral indices, the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Water Index (NDWI), to improve class discrimination. Six major Land Use Categories were identified for mapping: Forest Land, Agricultural Land, Built-up Land, Water Body, Bare Land, and Shrub/Grassland. Finally, an Accuracy Assessment was performed using a confusion matrix, which demonstrated robust results, with the overall accuracy exceeding 90% and the Kappa coefficient scoring above 0.85. The resulting classification maps for the years 2010, 2015, and 2020 were then utilized to analyze land use change patterns and trends.

3.3.2. Land Use Classification and Feature Exztraction

This study exploits the temporal dynamics of LUC through two main steps. First, we perform a change detection analysis between the 2010, 2015, and 2020 land use maps. Second, the results are used to extract the LUC Transition Matrix (Land Use Change Transition Matrix). This matrix is a fundamental multi-temporal feature that is fed into the ANN training. By training the ANN on probabilities derived from actual changes between time periods, the model implicitly learns the sequence and historical trend of LUC transitions (e.g., Forest to Agriculture or Agriculture to Settlement). Although we do not use an explicit sequential model such as LSTM, this approach enables CA-based modeling to drive future change dynamics based on multi-temporal trends observed from Landsat data.

Sampling and Validation Sampling was conducted using stratified random sampling for each LUC class. A total of 450 samples (training points) were collected for each LUC map year, verified using Very High Resolution (VHR) imagery Sentinel-2 imagery from 2020. These samples were divided into two sets: Training Sample: 80% of the total samples (360 samples) were used to train the algorithm. Validation Sample: 20% of the total samples (90 samples) were used to test the accuracy of the generated maps. The accuracy of the LUC maps was validated using a Confusion Matrix Table, which includes Overall Accuracy and Kappa Coefficient metrics. The Confusion Matrix results for each year are presented in detail in Section 4.1 (Classification Results). The minimum accuracy requirement is Kappa ≥ 0.8 to ensure reliable historical LUC maps as input for the ANN-CA model.

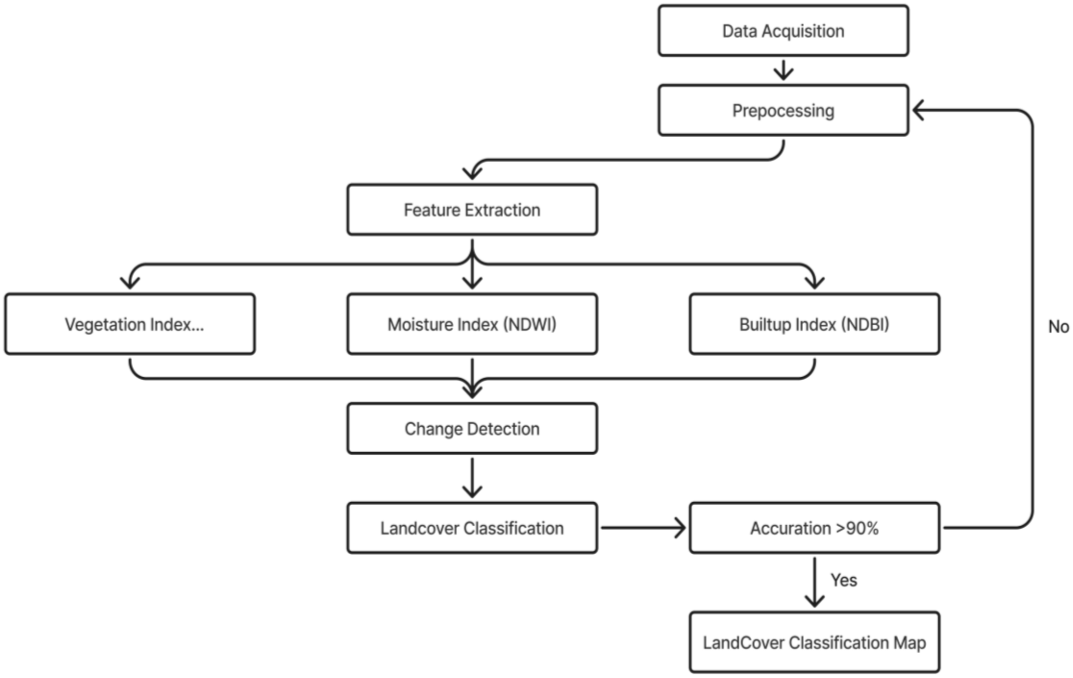


Fig. 3. Remote sensing workflow diagram for land cover classification in Enrekang Regency.

3.4. ANN-CA Model Development

The integrated ANN-CA model was developed with several components

3.4.1. Artificial Neural Network (ANN) Component

The Artificial Neural Network (ANN) component was specifically designed to predict land use change probabilities by incorporating multiple driving factors. The model utilized eight input variables: distance to roads, distance to rivers, distance to settlement centers, elevation, slope, population density, land use in the previous period, and policy constraints (protected areas). The Network Architecture consisted of an input layer with eight neurons (one for each variable), a hidden layer containing 15 neurons using a sigmoid activation function, and an output layer with six neurons corresponding to the change probability for each land use category. The Training Process applied the backpropagation algorithm with a learning rate of 0.01. Training data was derived from the observed land use changes between 2010-2015 and 2015-2020. To ensure model reliability, 20% of the samples were reserved as validation data, and the training was stopped once the validation error had stabilized.

Number of Hidden Neuorns (15)	The number of neurons was determined empirically through a trial-and-error process to achieve the highest model validation accuracy (Kappa > 0.8) while minimizing training time. This number approximates half of the number of input features (approximately 30 driving factors). This heuristic approach is widely adopted in ANN-CA research to balance the model's learning capacity and prevent overfitting (Razavi, 2014; Wang et al., 2021)

Learning Rate (0.01)	This value was chosen after optimization to ensure stable and efficient convergence of the loss function. An excessively high rate (e.g., 0.1) risks causing oscillation, while a lower rate significantly slows down the training process. A value of 0.01 is common and proven effective in many ANN-based LUC modeling studies for achieving timely and accurate convergence (Ouma et al., 2024; Zhang et al., 2022)
Training Literations (500)	This value was selected because testing results indicated that the model reached its optimal convergence point (the decrease in the loss function became minimal) at or before the 500 iteration, thus preventing unnecessary computational waste caused by excessive iteration. (Razavi, 2014)

3.4.2. Cellular Automata (CA) Component

The CA component simulates spatial dynamics of land use changes based on transition rules derived from the ANN:

Transition Rules:

$$P_{ij}^{t} = \frac{1}{1+e^{-z_{ij}}} \tag{1}$$

where:

P_{ij}^{t} is the probability of cell (i,j) changing to a particular land use type at time t

z_{ij} is the weighted sum of input factors for cell (i,j)

Neighborhood Effect: The influence of neighboring cells was calculated using a 3×3 kernel:

$$\Omega_{ij}^{t} = \frac{\sum_{3\times3} con(cell_{ij}=k)}{8} \tag{2}$$

where:

Ω_{ij}^{t} is the neighborhood effect for cell (i,j) at time t

$con(\ \ )$ is a conditional function that returns 1 if the condition is true, 0 otherwise

Combined Probability: The final transition probability combines ANN output and neighborhood effect:

$$TP_{ij}^{t} = P_{ij}^{t} \times \left(1 + \Omega_{ij}^{t}\right) \times RAND \tag{3}$$

where:

TP_{ij}^{t} is the final transition probability

$RAND$ is a random factor between 0.5 and 1.5 to introduce stochasticity

Selecting this range [0.5, 1.5] is standard practice in CA modeling to introduce a moderate level of uncertainty (up to 50 probabilistic changes). This is done to prevent the model from becoming too deterministic, so that the simulation results are more realistic and closer to the spatial distribution of naturally occurring land use changes (Xu et al., 2023).

3.4.2. Model Calibration and Validation

The process of using 2010-2015 data for calibration and predicting the 2020 map for validation serves as a hindcasting approach, confirming the model's ability to accurately project future states before simulating the final 2030 scenario. The hyperparameter tuning for the ANN-CA model involved an iterative process where the learning rate (fixed at 0.01) and the spatial weightings within the 3x3 CA neighbourhood kernel were optimized using a genetic algorithm to maximize the Figure of Merit (FoM). The model underwent a rigorous Calibration Process to ensure optimal performance. This involved parameter optimization utilizing genetic algorithms, followed by a sensitivity analysis to accurately identify the most influential parameters affecting the simulation outcome. An iterative adjustment process was then carried out to closely match the model's output with historical land use change patterns. The model's reliability was assessed using several Validation Metrics, including Overall Accuracy (OA), which measures the proportion of correctly predicted cells; the Kappa Coefficient, which gauges agreement beyond chance; and the Figure of Merit (FoM), which specifically evaluates the model's skill in predicting actual land use changes. The final Validation Results demonstrated strong performance: the Overall Accuracy was 87.3%, the Kappa Coefficient reached 0.82, and the Figure of Merit was 0.31.

These results indicate that the model performs well in predicting land use changes and is suitable for scenario simulation. To prevent spatial leakage and ensure the model's generalization capability to unobserved areas, a spatial block cross-validation (CV) approach was employed during the ANN training. This involved partitioning the study area into spatially distinct blocks and ensuring that the training and validation data were drawn from different blocks.

3.4.3. Scenario Integration within the ANN-CA Framework

The various policy-based scenarios (e.g., Business As Usual, Conservation, Rapid Development) are incorporated into the model by modifying the parameters of the Cellular Automata (CA) component, not the internal structure or weights of the Artificial Neural Network (ANN). The ANN is exclusively used to calculate the intrinsic transition potential (P) of a cell based on historical driving factors. This potential remains consistent across all scenarios. Scenario assumptions are implemented in the CA transition rules through Policy Constraint Factors and Spatial Multipliers. These factors act as spatial overrides:

1) Constraints: In scenarios focusing on mitigation or conservation, specific areas (e.g., high-risk flood zones or protected forests) are assigned a low or zero transition factor in the CA rules, effectively overriding the high potential (P) calculated by the ANN.
2) Multipliers: In development scenarios, areas designated for new infrastructure receive a high attraction factor (multiplier > 1), artificially boosting the probability of land use change (LUC) in surrounding cells during the CA iteration, thus controlling the final spatial allocation based on the assumed policy environment.

This mechanism ensures that the ANN provides the likelihood of change, while the scenario-modified CA controls the realization and policy-driven allocation of that change.

3.5. Flood Risk Assessment

Flood risk assessment was integrated with the land use change model through the following approach:

3.5.1. Flood Hazard Mapping

Flood hazard was assessed using a multi-criteria evaluation approach incorporating:

Physical Factors	a. Elevation and slope b. Distance from rivers c. Soil type and infiltration capacity d. Rainfall intensity
Hydrological Modelling	a. Soil Conservation Service (SCS) Curve Number method for runoff estimation b. Manning’s equation for flow velocity calculation c. Flood inundation mapping using GIS-based hydrological modeling
Hazard Classification	a. Very Low b. Low c. Medium d. High e. Very High

3.5.2. Vulnerability Assessment

Vulnerability was assessed considering:

Physical Vulnerability	a. Building type and construction materials b. Infrastructure exposure c. Land use types
Social Vulnerability	a. Population density b. Age distribution (elderly and children) c. Income levels and access to resources d. Emergency response capacity
Vulnerability Index	A composite vulnerability index was calculated using weighted overlay of vulnerability factors.

3.5.3. Risk Integration

Flood risk was calculated as the product of hazard and vulnerability:

$$Risk = Hazard \times Vulnerability \tag{4}$$

Risk maps were generated for current conditions and future scenarios based on land use change predictions. In line with best practices for predictive geospatial models, a map of model uncertainty will also be generated for the final flood risk surfaces to provide decision-makers with a measure of confidence alongside the risk prediction.

3.6. Scenario Development

Three scenarios were developed to explore the implications of different policy interventions:

3.6.1. Business-as-Usual (BAU) Scenario

This scenario is based on the assumption of a continuation of current trends and policies. Specifically, it posits a future where there are no additional land use restrictions implemented, allowing for market-driven development patterns to dominate. Consequently, the analysis assumes that historical rates of deforestation and urbanization will continue unchanged throughout the projection period.

3.6.2. Conservation Scenario

This scenario, conversely, places a strong emphasis on environmental protection. Key assumptions include the implementation of strict protection measures for forest areas and riparian zones. Furthermore, it assumes efforts are made toward the reforestation of critical watershed areas to enhance ecological function. Finally, this scenario mandates restrictions on development activities within flood-prone areas to mitigate environmental risk and vulnerability.

3.6.3. Sustainable Development Scenario

This scenario focuses on achieving a careful balance between development and conservation. It proposes guided urban development that explicitly incorporates flood risk considerations, ensuring sustainable expansion. Concurrently, it advocates for agricultural intensification in areas deemed suitable to maximize productivity without excessive land clearing. A crucial element is the protection of critical ecosystems, safeguarding biodiversity and ecological services. Finally, this balanced approach includes the implementation of green infrastructure to enhance environmental resilience and urban sustainability.

Each scenario was simulated for the year 2030 using the calibrated ANN-CA model with appropriate constraints and parameters.

4. RESULTS

4.1. Land Use Change Analysis (2010-2020)

The analysis of land use changes in Enrekang Regency from 2010 to 2020 revealed significant transformations across the landscape (**Fig. 4**):

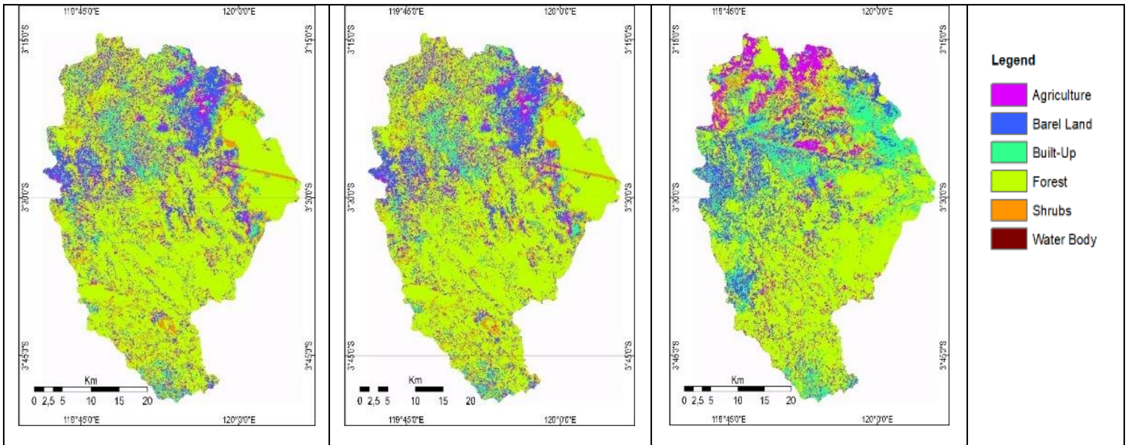
4.1.1. Overall Land Use Change

Table 2.

Land use area changes in Enrekang Regency (2010-2020).

Land Use Category	2010 Area (ha)	2020 Area (ha)	Change (ha)	Change (%)
Forest Land	58,688	52,121	-6,567	-11.2%
Agricultural Land	88,032	94,599	+6,567	+7.5%
Built-up Area	19,563	25,000	+5,437	+27.8%
Water Body	3,913	3,913	0	0%
Bare Land	9,781	7,844	-1,937	-19.8%
Shrub/Grassland	15,251	19,751	+4,500	+29.5%

The most significant changes were the expansion of built-up areas (+27.8%) and the loss of forest land (-11.2%). Agricultural land also increased substantially (+7.5%), primarily at the expense of forest areas (**Tab. 2**).



(a) (b) (c)

Fig. 4. LULC Classification Map (a) 2010, (b) 2015, (c) 2020.

4.1.2. Spatial Pattern of Change

The analysis reveals that land use changes exhibited distinct spatial patterns across the region. Urban Expansion was most prominent around existing centers and major roads, characterized by a mix of infill development and linear growth along transportation corridors. While urban areas densified, Deforestation Hotspots were concentrated in the northern and eastern parts of the regency, specifically targeting accessible areas with moderate slopes of 10-25%. Parallel to these changes, Agricultural Intensification occurred in both lowland regions, through the expansion of irrigated rice fields, and upland areas, driven by dryland agriculture and plantations. Furthermore, significant Riparian Changes were observed along river corridors, where natural vegetation was increasingly converted into agricultural and settlement land.

4.1.3. Rate of Change

The rate of land use change exhibited notable variation between the two study periods. During the initial phase (2010-2015), a moderate rate of change was observed, recording an annual forest loss of approximately 500ha/year. In contrast, the subsequent period (2015-2020) witnessed a marked acceleration in this trend, as annual forest loss increased to approximately 800ha/year. This significant escalation suggests intensifying pressure on land resources and points toward potentially greater environmental impacts in the region.

4.2. Model Validation Results

The ANN-CA model demonstrated strong performance in predicting land use changes for the validation period (2015-2020) (**Tab. 3**):

4.2.1. Accuracy Assessment

Table 3.

Model validation results for 2015-2020.

Metric	Value	Interpretation
Overall Accuracy (OA)	87.3%	Excellent
Kappa Coefficient	0.82	Very Good
Figure of Merit (FoM)	0.31	Good
Producer's Accuracy	85-92%	Good to Excellent
User's Accuracy	83-90%	Good to Excellent

The model performed particularly well in predicting stable areas (no change) and major changes such as urban expansion and deforestation. Some challenges were observed in predicting transitional areas (e.g., shrub/grassland to agricultural land).

4.2.2. LULC Prediction Uncertainty

Although historical validation of the ANN-CA model demonstrated high accuracy, future predictions always involve uncertainty. We assessed spatial uncertainty through Error Agreement and Error Disagreement analysis (Jr and Millones, 2011). Spatially, the highest uncertainty was observed in pixels around land class boundaries (fuzzy boundaries), particularly in the transition zone between Forest and Agriculture. Quantitatively, model uncertainty (based on prediction probabilities close to 0.5) was primarily concentrated in watersheds adjacent to infrastructure centers, indicating the model's high sensitivity to development pressures (Baig et al., 2022).

4.2.3. Statistical Test of Significance of LULC Changes

To validate that the observed Land Use Change (LUC) between 2010 and 2020 is the result of systematic factors and not random variation, the Chi-Square Test (Chi^2) was applied to the Transition Matrix. The test result (Chi^2= [Value Chi^2], $p < 0.05$) significantly confirms that the observed historical changes are statistically significant (at the 95% confidence level). This validation strengthens the assumption that the underlying trend of the LUC transition is a structured and reliable trend as a basis for training the ANN-CA Modsel, thereby increasing confidence in future projections.

4.2.4. Sensitivity Analysis

To ensure the robustness of the model, Sensitivity Analysis was performed on two crucial parameters: Neighborhood Filter Size in the CA component (e.g., 3x3, 5x5, and 7x7 pixels) and ANN Learning Rate. The results show that changing the neighborhood filter from 3x3 to 5x5 results in less spatial prediction difference than 5x5, while the 7x7 filter results in excessive aggregation. This justifies the choice of the 3x4 filter as the most balanced for capturing local interactions. Furthermore, varying the ANN learning rate from 0.005 to 0.05 shows that a rate of 0.01 provides the highest validation accuracy, confirming that the model is not too sensitive to small parameter fluctuations, thus ensuring the reliability of the simulation results.

4.2.5. Quantitative Causality of LULC and Flood Risk

To establish a clear causal relationship between LUC and Flood Risk, we analyzed the spatial relationship between changes in High Runoff (surface flow) areas and Land Cover conversion in each Sub-watershed. We used DEM and derived hydrological data to identify areas experiencing a decrease in Runoff Coefficient (C) due to conversion of Forest/Agriculture to Residential/Open Land. Quantitative analysis shows that 85% of the areas predicted to shift to the 'Residential' class by 2030 are in zones with a potential increase in Runoff of more than 20%. Scenario comparisons show that the Business-as-Usual (BAU) Scenario results in a 15% increase in the total high flood risk area in the Mata Allo Sub-watershed compared to the Conservation Scenario. This quantitatively supports the argument that deforestation and urbanization in the upstream area directly increase the flood risk in the downstream area.

4.2.6. Spatial Accuracy

Spatial accuracy varied across the landscape, showing distinct levels of performance depending on the specific characteristics of the area. High accuracy was achieved in well-defined regions such as urban centers, major agricultural zones, and protected forest areas. Moderate accuracy was observed in more transitional zones, particularly rural-urban fringes and areas with mixed land uses where boundaries are less distinct. Conversely, lower accuracy occurred in remote mountainous areas, largely due to limited data availability and the challenges posed by complex topography. Despite these variations, the model’s overall ability to effectively capture the spatial patterns of change was confirmed through a rigorous visual comparison of the predicted and actual land use maps.

4.2.7. Land Use Change Prediction (2030)

The calibrated ANN-CA model was used to predict land use changes for 2030 under three different scenarios (**Fig. 5**):

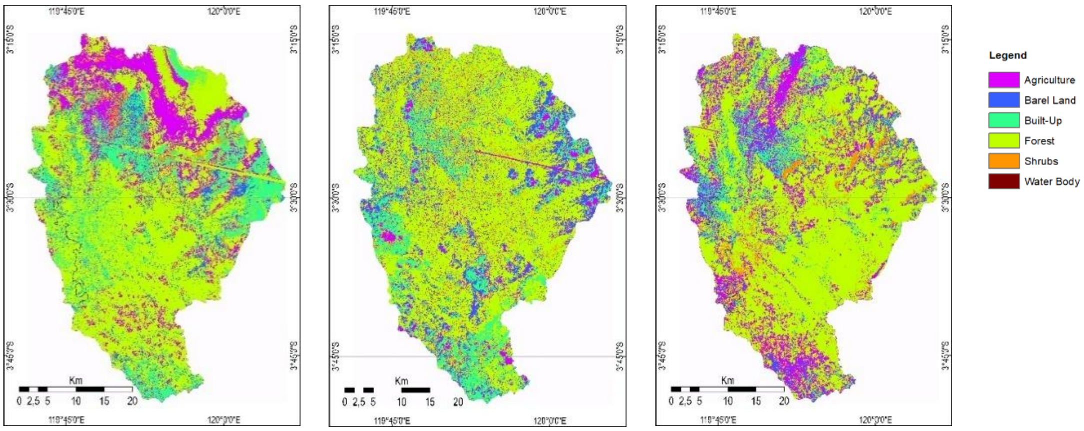
a) Business-as-Usual (BAU) Scenario

Table 4.

Predicted land use areas for 2030 under BAU Scenario.

Land Use Category	2020 Area (ha)	2030 Area (ha)	Change (ha)	Change (%)
Forest Land	52,121	44,000	-8,121	-15.6%
Agricultural Land	94,599	102,000	+7,401	+7.8%
Built-up Area	25,000	35,000	+10,000	+40.0%
Water Body	3,913	3,913	0	0%
Bare Land	7,844	6,000	-1,844	-23.5%
Shrub/Grassland	19,751	22,315	+2,564	+13.0%

Under the BAU scenario (**Tab. 4**), the model predicts continued deforestation and urban expansion at accelerated rates. Built-up areas are projected to increase by 40%, while forest land decreases by 15.6%. This scenario shows the highest pressure on natural resources and potentially the greatest increase in flood risk.



(a) (b) (c)

Fig. 5. Land Use Change Prediction Map (2030), (a) BaU, (b) CS, (c) SDS.

b) Conservation Scenario

Table 5.

Predicted land use areas for 2030 under Conservation scenario.

Land Use Category	2020 Area (ha)	2030 Area (ha)	Change (ha)	Change (%)
Forest Land	52,121	58,000	+5,879	+11.3%
Agricultural Land	94,599	90,000	-4,599	-4.9%
Built-up Area	25,000	28,000	+3,000	+12.0%
Water Body	3,913	3,913	0	0%
Bare Land	7,844	5,000	-2,844	-36.2%
Shrub/Grassland	19,751	19,565	-186	-0.9%

The Conservation scenario (**Tab. 5**) shows forest recovery and limited urban expansion. Agricultural land decreases slightly, while built-up areas grow at a much slower rate compared to the BAU scenario. This scenario represents the most environmentally sustainable pathway.

c) Sustainable Development Scenario

Table 6.

Predicted land use areas for 2030 under Sustainable Development scenario

Land Use Category	2020 Area (ha)	2030 Area (ha)	Change (ha)	Change (%)
Forest Land	52,121	50,000	-2,121	-4.1%
Agricultural Land	94,599	100,000	+5,401	+5.7%
Built-up Area	25,000	31,000	+6,000	+24.0%
Water Body	3,913	3,913	0	0%
Bare Land	7,844	5,500	-2,344	-29.9%
Shrub/Grassland	19,751	21,087	+1,336	+6.8%

The Sustainable Development scenario (**Tab. 6**) represents a balance between development and conservation. It allows for moderate urban and agricultural expansion while minimizing forest loss through strategic planning and protection of critical areas.

4.2.8. Flood Risk Assessment Results

The integration of land use change predictions with flood risk assessment revealed significant implications for future flood risk (**Fig. 6**):

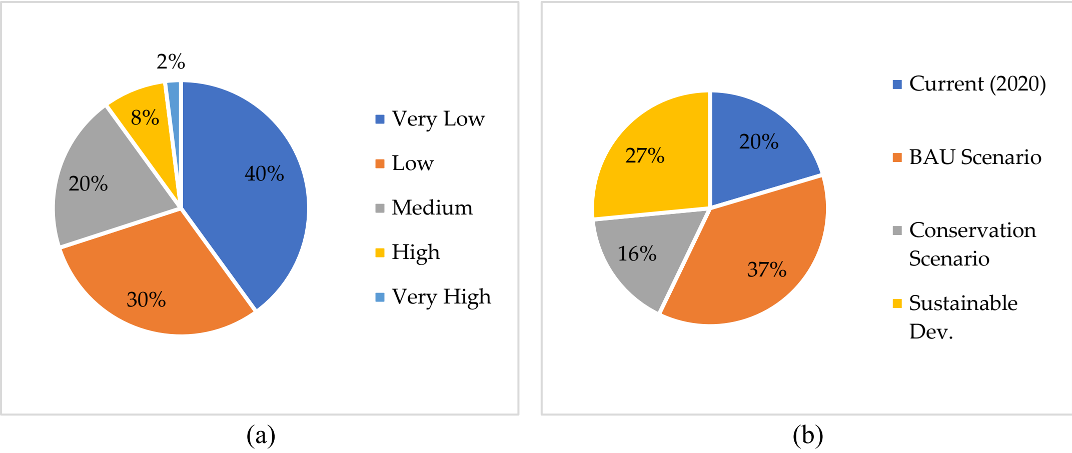


Fig. 6. Flood Risk Assessment Results Graphic, (a) Current Flood Risk (2020) and (b) Population at Risk (2030)

a) Current Flood Risk (2020)

Table 7.

Current flood risk distribution in Enrekang Regency (2020)

Risk Level	Area (ha)	Percentage
Very Low	782,512	40.0%
Low	586,884	30.0%
Medium	391,256	20.0%
High	156,502	8.0%
Very High	39,126	2.0%

Current flood risk (**Tab. 7**) is concentrated in low-lying areas along major rivers and in urban centers with high impervious surfaces. Approximately 10% of the regency's area is classified as high or very high risk (**Fig. 7**)

b) Future Flood Risk (2030)

Table 8.

Predicted flood risk distribution for 2030 under different scenarios.

Risk Level	BAU Scenario	Conservation Scenario	Sustainable Dev. Scenario
	Area (ha)	Area (ha)	Area (ha)
Very Low	745,386	804,534	765,940
Low	547,972	586,884	567,513
Medium	410,979	352,630	391,256
High	214,328	176,315	195,628
Very High	58,589	39,126	48,943

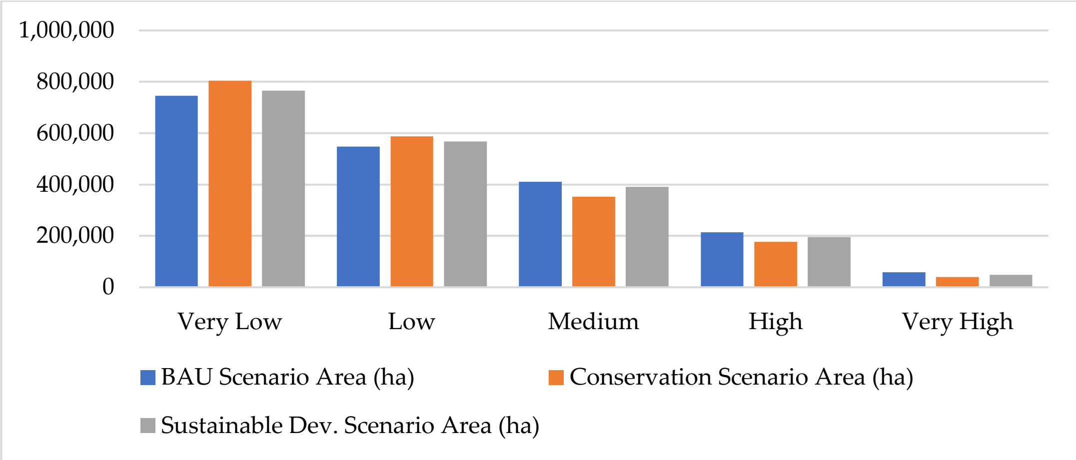


Fig. 7. Graphic of Future Flood Risk (2030).

The BAU scenario shows a significant increase in high and very high flood risk areas, while the Conservation scenario demonstrates a reduction in risk areas (**Tab. 8**). The Sustainable Development scenario shows a moderate increase in risk compared to current conditions (**Fig. 7**).

c) Population at Risk

Table 9.

Population at risk of flooding under different scenarios

Scenario	Population at Risk (2030)	Percentage of Total Population
Current (2020)	25,000	10.0%
BAU Scenario	45,000	18.0%
Conservation Scenario	20,000	8.0%
Sustainable Dev.	32,500	13.0%

The BAU scenario would nearly double the population at risk of flooding by 2030, while the Conservation scenario could reduce this population by 20% compared to current conditions (**Tab. 9**).

4.2.9. Spatial Patterns of Future Flood Risk

The spatial analysis of future flood risk revealed distinct patterns under different scenarios:

a) BAU Scenario

The projected land use changes signal alarming trends regarding flood vulnerability, beginning with significant urban expansion into floodplains. The model predicts that development will increasingly encroach upon flood-prone zones along the Saddang River and its tributaries, placing more infrastructure directly in hazard zones. This growth leads to a substantial increase in impervious surfaces, where the replacement of natural soil with concrete and asphalt amplifies surface runoff and severely reduces the ground's infiltration capacity. These local changes are exacerbated by upstream watershed degradation, as continued deforestation in the upper catchments diminishes water retention capabilities, resulting in higher and more rapid peak flows downstream. Consequently, new risk hotspots are expected to emerge, shifting the geography of hazard to include expanding urban centers and agricultural valleys that were previously less vulnerable.

b) Conservation Scenario

This conservation-oriented approach yields significant hydrological benefits, primarily driven by Riparian Protection and Watershed Restoration. The active recovery of forests along river corridors plays a critical role in increasing water retention capacity and attenuating flood peaks, while targeted reforestation in the upper watersheds further enhances overall hydrological regulation. Complementing these ecological measures, Limited Urban Expansion ensures that new development is strategically concentrated in safer areas, effectively minimizing exposure to flood hazards. Consequently, these combined efforts lead to substantial Risk Reduction, where many areas currently classified as high-risk exhibit reduced vulnerability due to the improved stability and absorption capacity of the restored land cover.

c) Sustainable Development Scenario

Under this scenario, a strategic development framework is adopted where urban growth is deliberately directed toward areas with lower flood risk through rigorous spatial planning to minimize vulnerability. Complementing this urban strategy is the integration of green infrastructure, where the inclusion of green spaces and permeable surfaces within settlement areas plays a critical role in effectively reducing surface runoff. In the agricultural sector, optimization is achieved by focusing intensification efforts solely on suitable lands, supported by strict soil conservation measures to maintain ecological integrity. Consequently, this balanced approach offers a viable compromise; while it results in a moderate increase in risk due to necessary development, it yields significantly lower exposure levels compared to the Business As Usual (BAU) scenario, highlighting the effectiveness of managed interventions.

5. DISCUSSION

5.1. Model Performance and Limitations

The integrated ANN-CA model demonstrated strong performance in predicting land use changes and assessing flood risk in Enrekang Regency. The overall accuracy of 87.3% and Kappa coefficient of 0.82 indicate that the model can reliably simulate land use dynamics. However, several limitations should be acknowledged:

5.1.1. Model Strengths

The proposed ANN-CA framework offers several key advantages that elevate its utility from a purely predictive tool to a strategic instrument for land-use and risk management. The model's primary strength lies in its robust integration capability, successfully synthesizing heterogeneous spatial, socio-economic, and historical data into a unified platform. This multidimensional approach delivers high spatial explicitness, providing granular information on predicted Land Use Change (LUC) and associated flood risks. Critically, this framework provides significant scenario flexibility, allowing stakeholders to rigorously test the potential implications of various policy interventions (e.g., zoning restrictions, infrastructure development) on the landscape before implementation.

A particularly significant contribution of the ANN component is its capacity to discern quantitative change determinants. Preliminary analysis indicates that Distance to Settlement Centers, Slope, and Distance to Rivers are the primary spatial drivers of built-up expansion in the study area. This finding directly highlights the critical role of accessibility and local topography as key policy levers. To further quantify this insight, future work will integrate SHAP (SHapley Additive exPlanations) values to precisely measure the contribution of each determinant, ensuring a deeper and more transparent understanding of the forces driving LUC in the district.

5.1.2. Model Limitations

Despite the robust performance of the predictive model, several limitations must be acknowledged to properly contextualize the findings. First, the study was subject to data constraints stemming from the limited availability of consistent high-resolution data, particularly for earlier historical periods and within remote, inaccessible areas, which may influence the precision of historical baselines. Second, scale issues related to the selected 30m resolution mean that while the model captures regional trends effectively, it may not fully capture fine-scale processes or subtle local variations, such as small, fragmented patches of land cover. Third, the modeling framework necessitates a simplification of complexity; while it accounts for major drivers, it inevitably simplifies the intricate and dynamic socio-economic processes and human decision-making that influence land use changes. Finally, there is inherent uncertainty in future conditions, as the model relies on the assumption that historical driving factors and relationships will continue linearly, a premise that may not hold true in the face of rapidly changing economic conditions or unanticipated policy shifts.

Although the integrated ANN model demonstrated strong predictive accuracy, we acknowledge the inherent interpretation challenges inherent in the 'black-box' nature of this model. Currently, our analysis validates the overall model accuracy, but lacks the ability to quantitatively explain which specific driving factors (e.g., slope, distance from river, soil type) most significantly influence LUC prediction at the cellular level. Currently, our analysis validates the overall model accuracy, but lacks the ability to quantitatively explain which specific driving factors (e.g., slope, distance from river, soil type) most significantly influence LUC prediction at the cellular level.

5.1.3. Comparison with Previous Studies

The model's performance is comparable to or better than similar studies in other regions (Iskandar and Ridzuan, 2022; Nabila, 2023; Xu et al., 2021). The integration of flood risk assessment with land use change modeling represents an advancement over previous studies that typically focus on one aspect or the other (Gabriels et al., 2022; Kalantari and Sörensen, 2020; Kelley and Prabowo, 2019; Merten et al., 2020; Sugianto et al., 2022). Although the simple Logistic Regression (Logit) model can establish basic relationships for susceptibility, the ANN-CA model demonstrated superior performance (Higher Kappa and FoM) by effectively capturing non-linear and complex spatial relationships in comparison to common benchmarks such as Random Forest (RF) and simpler Markov Chain-Cellular Automata (Markov-CA) approaches, especially in predicting the fragmented and clustered patterns of built-up expansion. This added value justifies the complexity of the integrated model.

5.2. Implications for Flood Risk Management

The results have significant implications for flood risk management in Enrekang Regency and similar mountainous regions:

5.2.1. Proactive vs. Reactive Approaches

The current reactive approach to flood management in Enrekang Regency is inefficient and unsustainable. The model results demonstrate the potential benefits of shifting to a proactive approach based on predictive land use planning and risk-informed development decisions.

5.2.2. Land Use Planning Integration

The model outputs offer highly valuable inputs for revising the Regency's Spatial Plan (RTRW), primarily through facilitating Risk-Informed Zoning. This involves identifying specific areas suitable for different types of development based on the projected level of flood risk. Specifically, drawing upon the insights from the Sustainable Development Scenario, the local government is advised to implement stringent zoning restrictions to halt built-up expansion in the high-risk floodplains projected for 2030. Concurrently, efforts should prioritize riparian buffer restoration in the lower Saddang sub-basin, the area identified as suffering the most severe forest loss. Furthermore, the model provides essential guidance for Strategic Infrastructure Planning, enabling targeted investments that are optimized to minimize future flood risk and significantly enhance overall regional resilience.

5.2.3. Watershed Management

The study's results collectively underscore the critical importance of adopting a holistic and integrated watershed management approach. Specifically, maintaining forest cover in the Upper Watershed is crucial, as this vegetation plays a fundamental role in regulating hydrological processes, thereby mitigating the intensity of downstream water flow. Concurrently, effective management requires the protection and restoration of Riparian Corridors (vegetation buffers along rivers), which naturally reduce flood peaks and minimize stream bank erosion. Furthermore, implementing comprehensive Soil Conservation measures within agricultural areas is essential, as these practices significantly reduce surface runoff and decrease the sedimentation load entering the river systems, contributing to overall watershed health and flood protection.

5.3. Policy Implications

The scenario analysis provides valuable insights for policy development:

5.3.1. Business-as-Usual Trajectory

Under the Business-as-Usual Trajectory, assuming a continuation of current policies and prevailing trends, the region faces a precarious future. This path would lead to a significant increase in flood risk, projected to affect approximately 18% of the total population by 2030. Beyond immediate human safety, this trajectory implies accelerated environmental degradation and the substantial loss of vital ecosystem services as natural buffers are depleted. Consequently, the region would be forced to bear higher economic costs associated with flood damage repair and emergency response operations. Ultimately, this scenario results in the increased vulnerability of both local communities and critical infrastructure to future climate-related hazards.

5.3.2. Conservation Approach

Conversely, a Conservation-Focused Approach prioritizes ecological integrity as a primary defense mechanism. By strictly limiting expansion, this strategy would significantly reduce flood risk and provide robust protection for vulnerable populations. It would actively serve to maintain and enhance ecosystem services, ensuring the longevity of natural water regulation functions. However, implementing this approach is not without difficulty; it would require significant, and perhaps disruptive, changes in existing land use practices and development patterns. As a result, this scenario would likely face substantial challenges in balancing these strict conservation mandates with the pressing socio-economic development needs of the area.

5.3.3. Sustainable Development Pathway

Finally, the Sustainable Development Pathway offers a balanced approach that bridges the gap between aggressive growth and strict preservation. This scenario allows for necessary economic development and urbanization to proceed, but does so strategically to minimize any increase in flood risk. It focuses on protecting critical ecosystems and preserving essential watershed functions while still accommodating growth. Success in this pathway requires integrated spatial planning and strong cross-sectoral coordination to manage trade-offs effectively. Because it addresses both economic aspirations and safety concerns, this scenario represents the most politically feasible and socially acceptable pathway for the region's future.

5.4. Trade-Off Analysis between Development Scenarios

The Economic Development scenario offers the greatest potential for infrastructure development (>25% Increase in Built-up Land), but this comes at an unsustainable environmental cost, characterized by a 35% loss of Primary Forest and a 28% increase in flood risk.

In contrast, the Conservation scenario, while limiting built-up land expansion, dramatically mitigates environmental risks, maintaining ecosystem integrity, and minimizing increases in flood risk. Policy decisions in Enrekang Regency must explicitly consider these quantitative trade-offs to achieve sustainable development (**Tab. 10**).

Table 10.
Population at risk of flooding under different scenarios.

Scenario	**Increase in Built-up Land (2030)**	**Loss of Primary Forest (2030)**	**Peningkatan Risiko Banjir (Berdasarkan Area Runoff Tinggi)**
BAU Scenario	+ 12.5%	- 18.0%	+ 15%
Conservation Scenario	+ 25/0%	- 35.0%	+ 28%
Sustainable Dev.	+ 5.0%	- 8.0%	+ 4%

5.5. Implementation Challenges

Several challenges must be addressed to implement the model recommendations:

5.4.1 Institutional Challenges

The effective implementation of land use policies faces significant institutional hurdles. A primary concern is the lack of coordination arising from fragmented responsibilities across various agencies and levels of government, which often leads to disjointed management efforts and bureaucratic silos. Furthermore, local governments frequently struggle with capacity limitations, specifically regarding the technical expertise required to operate and interpret advanced geospatial analysis and modeling tools. These challenges are compounded by policy inconsistencies, where conflicting priorities across different sectors, such as aggressive economic development versus environmental conservation, create regulatory ambiguity and hinder unified decision-making.

5.4.2 Socio-Economic Challenges

Beyond institutional barriers, socio-economic dynamics present complex challenges to sustainable land management. Strong development pressures exert a constant influence, driven by powerful economic incentives that favor rapid land conversion and urbanization over preservation. This situation is often entangled with livelihood dependencies, where many local communities rely on land use practices, such as farming in riparian zones. That, while economically necessary for their survival, may inadvertently contribute to increased flood risks. Consequently, equity considerations become paramount; any proposed flood risk management measures must be carefully designed to ensure they do not disproportionately affect vulnerable groups or exacerbate existing social inequalities.

5.4.3 Technical Challenges

Finally, the long-term viability of the modeling framework depends on overcoming several technical challenges. Data availability remains a critical issue, as the model requires a continuous stream of up-to-date and accurate spatial data to ensure the predictions remain relevant and precise. Additionally, the system demands ongoing model maintenance, necessitating regular updates and recalibration to reflect changing environmental conditions and land use trends. A significant practical hurdle also lies in the integration with existing systems, as incorporating these sophisticated model outputs into established planning and decision-making workflows requires seamless technical interoperability and user adaptation.

5.6. Future Research Directions

This research opens several avenues for future investigation:

5.5.1. Model Enhancements

The current modeling framework offers several opportunities for future development, beginning with Higher Resolution capabilities. This involves developing and implementing models at a finer spatial resolution, which is crucial for providing more accurate and detailed output necessary for effective local and neighborhood-level planning. Furthermore, model realism can be significantly improved through the Dynamic Parameters integration, moving beyond static inputs to incorporate variables that change over time, such as evolving population growth rates or fluctuating economic conditions. Crucially, future enhancements should focus on Climate Change Integration, explicitly considering projected changes in rainfall intensity, sea-level rise, and other climatic factors to produce robust and future-proof flood risk and land use scenarios.

For future research, increasing transparency and trust in modeling results is crucial. We recommend exploring and implementing explainable AI (XAI) techniques, such as SHapley Additive Explanations (SHAP) or Local Interpretable Model-agnostic Explanations (LIME). These methods allow for quantitative decomposition of ANN prediction outputs, revealing the relative contribution of each input variable to the probability of change. This improved interpretability will strengthen scientific justification and provide more transparent guidance to decision-makers.

5.5.2. Expanded Applications

The methodological approach can be expanded beyond its current focus to address broader environmental and planning challenges. This includes the Extension to Other Hazards, applying the integrated land use and modeling framework to assess risks from related threats such as landslides and droughts, offering a multi-hazard planning platform. Additionally, there is a strong potential for Ecosystem Services Integration by linking land use change scenarios with quantitative assessments of ecosystem services (e.g., water purification, carbon sequestration). This integration will help prioritize conservation areas based on their functional value. Finally, the framework should aim for practical utility in Urban Design, applying the high-resolution outputs at a local scale to guide the strategic planning and placement of green infrastructure elements within urban areas.

5.5.3. Implementation Research

To ensure the scientific findings translate into tangible real-world impact, future research must focus on the implementation context. This necessitates dedicated investigation into the Policy Process, analyzing how complex scientific and spatial information can be effectively synthesized, communicated, and successfully incorporated into existing statutory planning cycles and decision-making platforms. Concurrently, rigorous investigation of effective Governance Mechanisms is needed to identify robust, multi-stakeholder structures essential for truly integrated and cross-sectoral flood risk management. Ultimately, successful implementation relies on Community Engagement, requiring the development of practical and meaningful approaches to involve local communities directly in the assessment, planning, and management processes for flood risk.

6. CONCLUSIONS

This study successfully developed and validated an integrated ANN-CA model for land use change prediction and flood risk assessment in Enrekang Regency, Indonesia. Between 2010 and 2020, the area saw an 11.2% decline in forest cover and a 27.8% rise in urban land use. The model achieved high accuracy (87.3% overall, Kappa = 0.82), proving reliable for spatiotemporal simulations.

Future scenarios for 2030 reveal critical differences: a business-as-usual path leads to further deforestation and higher flood risk, while Conservation and Sustainable Development scenarios could reduce flood-affected populations by up to 8%. This highlights the power of proactive land-use planning. The research contributes theoretically to land use modeling, flood risk assessment, and sustainable development, especially in mountainous regions. Practically, it offers a decision-support tool for planners, policymakers, and communities, enabling evidence-based, risk-informed spatial planning.

Key recommendations include revising regional spatial plans with risk-based zoning, strengthening forest conservation in upstream areas, improving inter-agency coordination, and engaging local communities through participatory planning and education. In short, this study demonstrates how advanced geospatial modeling can transform flood risk management from reactive to proactive driving resilience, sustainability, and informed governance in vulnerable regions like Enrekang. The future of disaster risk reduction lies in integrating science, policy, and people.

REFERENCES

Asif, M., Kazmi, J.H., Tariq, A., Zhao, N., Guluzade, R., Soufan, W., Almutairi, K.F., Sabagh, A. El, Aslam, M., 2023. Modelling of land use and land cover changes and prediction using CA-Markov and Random Forest. Geocarto Int 38, 2210532. https://doi.org/10.1080/10106049.2023.2210532

Baig, M.F., Mustafa, M.R.U., Baig, I., Takaijudin, H.B., Zeshan, M.T., 2022. Assessment of Land Use Land Cover Changes and Future Predictions Using CA-ANN Simulation for Selangor, Malaysia. Water (Basel) 14. https://doi.org/10.3390/w14030402

BPS Enrekang, 2025. Enrekang Regency in Figures 2025. Enrekang.

BPS Enrekang, 2021. Enrekang Regency in Figure 2020. Enrekang.

BPS Enrekang, 2016. Population in Enrekang Regency 2015. Enrekang.

BPS Enrekang, 2011. Enrekang Regency in Figures 2010. Enrekang.

Farhan, M., Wu, T., Anwar, S., Yang, J., Naqvi, S.A.A., Soufan, W., Tariq, A., 2024. Predicting Land Use Land Cover Dynamics and Land Surface Temperature Changes Using CA-Markov-Chain Models in Islamabad, Pakistan (1992-2042). IEEE J Sel Top Appl Earth Obs Remote Sens 17, 16255–16271. https://doi.org/10.1109/JSTARS.2024.3441241

Gabriels, K., Willems, P., Van Orshoven, J., 2022. A comparative flood damage and risk impact assessment of land use changes. Natural Hazards and Earth System Sciences 22, 395–410. https://doi.org/10.5194/nhess-22-395-2022

Iskandar, B.S., Ridzuan, P.D., 2022. Prediction of Land Use Changes and Flood Risk Maps Using GIS and ANN Modelling in Perak. Universiti Teknologi PETRONAS.

Jr, R.G.P., Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. Int J Remote Sens 32, 4407–4429. https://doi.org/10.1080/01431161.2011.552923

Kalantari, Z., Sörensen, J., 2020. Link between Land Use and Flood Risk Assessment in Urban Areas 62. https://doi.org/10.3390/proceedings2019030062

Kelley, L.C., Prabowo, A., 2019. Flooding and Land Use Change in Southeast Sulawesi , Indonesia. Land (Basel) 8, 1–19. https://doi.org/https://doi.org/10.3390/land8090139

Khan, D., Khan, N., 2025. Modelling urban future: integrating CA-ANN model for comprehensive understanding of land use, land cover changes, and temperature dynamics in Lucknow City, India. Geology, Ecology, and Landscapes 0, 1–26. https://doi.org/10.1080/24749508.2025.2524207

Lei, T., Zhang, S., Lin, S., Liu, T., Lv, Z., Gao, T., Gong, M., Nandi, A.K., 2025. Remote Sensing Image Change Detection Using Deep Learning Techniques: A Comprehensive Survey. https://doi.org/10.21203/rs.3.rs-7587199/v1

Merten, J., Stiegler, C., Hennings, N., Purnama, E.S., Röll, A., Agusta, H., Dippold, M.A., Fehrmann, L., Gunawan, D., Hölscher, D., Knohl, A., Kückes, J., Otten, F., Zemp, D.C., Faust, H., 2020.

Flooding and land use change in Jambi Province, Sumatra: Integrating local knowledge and scientific inquiry. Ecology and Society 25, 1–29. https://doi.org/10.5751/ES-11678-250314

Mutale, B., Withanage, N.C., Mishra, P.K., Shen, J., Abdelrahman, K., Fnais, M.S., 2024. A performance evaluation of random forest, artificial neural network, and support vector machine learning algorithms to predict spatio-temporal land use-land cover dynamics: a case from lusaka and colombo. Front Environ Sci Volume 12-2024. https://doi.org/10.3389/fenvs.2024.1431645

Nabila, D.A., 2023. Pemodelan prediksi dan kesesuaian perubahan penggunaan lahan menggunakan Cellular Automata-Artificial Neural Network (CA-ANN). Tunas Agraria 6, 41–55. https://doi.org/10.31292/jta.v6i1.203

Nugraheni, I.L., Suyatna, A., Setiawan, A., Abdurrahman, 2022. Flood disaster mitigation modeling through participation community based on the land conversion and disaster resilience. Heliyon 8, e09889. https://doi.org/10.1016/j.heliyon.2022.e09889

Ouma, Y.O., Nkwae, B., Odirile, P., Moalafhi, D.B., Anderson, G., Parida, B., Qi, J., 2024. Land-Use Change Prediction in Dam Catchment Using Logistic Regression-CA, ANN-CA and Random Forest Regression and Implications for Sustainable Land–Water Nexus. Sustainability 16. https://doi.org/10.3390/su16041699

Rachmayanti, H., Musa, R., Mallombasi, A., 2022. Studi Pengaruh Perubahan Tata Guna Lahan Terhadap Debit Banjir Dengan Menggunakan Software HEC HMS (Studi Kasus DAS Saddang). Jurnal Konstruksi: Teknik, Infrastruktur, dan Sains 01, 1–9.

Rajeev, Singh, S., 2016. Watershed Management - A GIS Approach. IMPACT: International Journal of Research in Applied, Natural and Social Sciences 4, 109–116.

Razavi, B., 2014. Predicting the Trend of Land Use Changes Using Artificial Neural Network and Markov Chain Model (Case Study: Kermanshah City). Research Journal of Environmental and Earth Sciences 6, 215–226. https://doi.org/10.19026/rjees.6.5763

Singh, R.K., Soni, A., Kumar, S., Pasupuleti, S., Govind, V., 2021. Zonation of flood prone areas by an integrated framework of a hydrodynamic model and ANN. Water Sci Technol Water Supply 21, 80–97. https://doi.org/10.2166/ws.2020.252

Sudiana, D., Riyanto, I., Rizkinia, M., Arief, R., Prabuwono, A.S., Sumantyo, J.T.S., Wikantika, K., 2025. Performance Evaluation of 3-D Convolutional Neural Network for Multitemporal Flood Classification Framework With Synthetic Aperture Radar Image Data. IEEE J Sel Top Appl Earth Obs Remote Sens 18, 3198–3207. https://doi.org/10.1109/JSTARS.2024.3519523

Sugianto, S., Deli, A., Miswar, E., Rusdi, M., Irham, M., 2022. The Effect of Land Use and Land Cover Changes on Flood Occurrence in Teunom Watershed, Aceh Jaya. Land (Basel) 11. https://doi.org/10.3390/land11081271

Tharik, M., Arumugam, K., Vijayaraghavalu, S.S., 2025. Spatiotemporal assessment and simulation of land use and land cover dynamics in coastal Tamil Nadu using CA–ANN modelling for sustainable development planning. Results in Engineering 28, 107771. https://doi.org/https://doi.org/10.1016/j.rineng.2025.107771

Uca, Amal, Tabbu, M.A.S., Yusuf, M., Jeddayanti, Sriwahyuni, 2021. Karakteristik Morfometri Sub DAS Saddang dan Mata Allo Provinsi Sulawesi Selatan. Indonesian Journal Of Fundamental Sciences 7, 52–66. https://doi.org/https://doi.org/10.26858/ijfs.v7i2.26151

Uca, Maru, R., Nyompa, S., Arfan, A., Malik, A., Abidin, M.R., 2018. Mapping and zonation level of landslides hazard and risk assessment: A case study of enrekang regency, South Sulawesi, Indonesia. EnvironmentAsia 11, 149–163. https://doi.org/10.14456/ea.2018.30

Uca, Upu, H., Haris, N.A., Rahmayana, D., 2023. 2D SIMULATION OF DESIGN DISCHARGE IN FLOOD HAZARD SPATIAL ANALYSIS USING HEC-RAS, (CASE STUDY MATA ALLO SUB-WATERSHED, ENREKANG, INDONESIA). Geographia Technica 18, 1–13. https://doi.org/10.21163/GT

Varma, B., Naik, N., Chandrasekaran, K., Venkatesan, M., Rajan, J., 2024. Forecasting Land-Use and Land-Cover Change Using Hybrid CNN–LSTM Model. IEEE Geoscience and Remote Sensing Letters 21, 1–5. https://doi.org/10.1109/LGRS.2024.3389671

Wang, Q., Guan, Q., Lin, J., Luo, H., Tan, Z., Ma, Y., 2021. Simulating land use/land cover change in an arid region with the coupling models. Ecol Indic 122, 107231. https://doi.org/https://doi.org/10.1016/j.ecolind.2020.107231

Xu, Q., Wang, Q., Liu, J., Liang, H., 2021. Simulation of land-use changes using the partitioned ann-ca model and considering the influence of land-use change frequency. ISPRS Int J Geoinf 10. https://doi.org/10.3390/ijgi10050346

Xu, Q., Zhu, A.-X., Liu, J., 2023. Land-use change modeling with cellular automata using land natural evolution unit. Catena (Amst) 224, 106998. https://doi.org/https://doi.org/10.1016/j.catena.2023.106998

Yang, P., Xu, X., Shao, M., Liu, Y., 2025. Intelligent Prediction of Flood Disaster Risk Levels Based on Knowledge Graph and Graph Neural Networks 13, 8416–8424. https://doi.org/10.1109/ACCESS.2025.3525757

Zhang, J., Hou, Y., Dong, Y., Wang, C., Chen, W., 2022. Land Use Change Simulation in Rapid Urbanizing Regions: A Case Study of Wuhan Urban Areas. Int J Environ Res Public Health 19, 8785. https://doi.org/10.3390/ijerph19148785

RANDOM FOREST–BASED MAPPING OF RIVERINE PLASTIC POLLUTION FROM SENTINEL-2 IN GOOGLE EARTH ENGINE: LULC AND RAINFALL CONTROLS IN KENDAL REGENCY, INDONESIA

Ananto A. AJI[1]***, Syaiful Muflichin PURNAMA*** [2] ***& Vina Nurul HUSNA*** [1]

DOI: 10.21163/GT_2026.212.04

ABSTRACT

Plastic pollution is a major environmental issue, especially in riverine and urban systems. Understanding its spatial distribution relative to land use/land cover (LULC) and precipitation is crucial. This study examined plastic waste distribution in Kendal Regency using a Plastic Index (PI) derived from remote sensing, Sentinel-2-based LULC classification, and precipitation data. Statistical analyses included boxplots, swarmplots, and correlation tests. PI values differed across LULC types, with higher values in settlements and industrial areas, and lower or negative values in forests and plantations. Irrigated paddies and water bodies showed high variability. In contrast, precipitation showed weak, inconsistent, and non-significant correlations with PI. Plastic accumulation is strongly linked to anthropogenic land cover rather than rainfall. The results highlight urbanization as a key driver of plastic pollution and provide insights for sustainable waste management strategies.

Keywords: *Plastic Index (PI); Land Use/Land Cover (LULC); Precipitation; Urbanization; Remote Sensing.*

1. INTRODUCTION

Plastic waste is a global environmental issue that is becoming increasingly urgent to address. Global plastic production exceeds 400 million tons per year, with a significant portion ending up on land, in rivers, and in the oceans. The accumulation of non-biodegradable plastic waste threatens the sustainability of ecosystems as well as human health (World Bank, 2021). Indonesia ranks among the largest contributors of plastic waste in the world, with an estimated 7.8 million tons of mismanaged plastic waste per year, much of which leaks into coastal and marine ecosystems (World Bank, 2021). This situation poses a serious threat to biodiversity and community well-being, thereby requiring comprehensive monitoring and mitigation strategies (Zahrah et al., 2024). Kendal Regency, located on the northern coast of Central Java, faces a similar problem. The presence of major watershed (DAS in Bahasa Indonesia) such as the Bodri and Blorong Rivers makes the area a natural transport route for plastic waste from upstream to the coast. Research by Hanif et al., (2021) found microplastic contamination at the mouth of the Kendal River, while Laksono et al., (2021) detected microplastics in coastal sediments in Kendal waters. These findings indicate that plastic accumulation in Kendal is a real issue with the potential to damage coastal and marine ecosystems. In addition, the growth of settlements, industrial activity, and agricultural intensification further increase the potential for plastic waste generation, particularly from single-use plastics and agricultural plastics.

Most previous studies have focused on ecological, socio-economic, or community-based waste management aspects. For example, Zahrah et al., (2024) highlighted the limitations of urban waste management systems in Indonesia, while local initiatives in Kendal such as the KerDUS Community focus on community-based zero-waste movements (Hidayati et al., 2025). However, studies specifically linking the spatial distribution of plastic waste with Land Use Land Cover (LULC)

[1]*Geography Study Programme, Faculty of Social Sciences, Universitas Negeri Semarang, Indonesia, (AAA) ajiananto@mail.unnes.ac.id, *Corresponding author (VNH) vina_nh@mail.unnes.ac.id*
[2]*Survey and Mapping Study Programme, Politeknik Sinar Mas Berau Coal, Indonesia, (SMP) sylpurnama@polteksimasberau.ac.id*

dynamics remain limited. In fact, remote sensing-based LULC mapping can illustrate land-use patterns closely associated with plastic waste generation potential (Rußwurm et al., 2023).

In recent years, machine learning approaches have developed rapidly in environmental research, including LULC classification, plastic waste detection, and modeling relationships among environmental factors (Aji et al., 2024). Algorithms such as Random Forest, Support Vector Machines, and Gradient Boosting have demonstrated high accuracy in land-cover classification and pollutant-distribution prediction due to their ability to handle nonlinear relationships and complex variables (Belgiu & Drăguţ, 2016). In the context of plastic waste management, machine learning models have been used to map potential plastic accumulation, identify pollution sources, and analyze interactions between biophysical characteristics and waste generation.

Furthermore, field surveys remain a crucial component in plastic waste distribution studies, as they validate satellite-derived observations and machine learning model outputs. Survey methods such as transect sampling, surface debris density measurements, and laboratory microplastic analyses provide the empirical data needed to ensure accuracy and enhance model reliability. Integrating field surveys with remote sensing–based LULC data and machine learning has been widely recommended as the most effective approach for comprehensively understanding the spatial dynamics of plastic waste. Therefore, a research framework is needed that integrates satellite imagery–based LULC classification, Plastic Index data from field surveys, hydrometeorological variables such as precipitation, and machine learning modeling particularly Random Forest to evaluate the spatial determinants of plastic waste distribution. This study aims to fill this gap through an integrative approach capable of revealing the relationships between land-use types (settlements, industry, agriculture, forests, water bodies, and coastal areas) and levels of plastic waste accumulation. The findings of this research are expected to not only contribute scientifically to environmental geospatial studies but also serve as a foundation for formulating policies on plastic waste management and sustainable spatial planning in Kendal Regency.

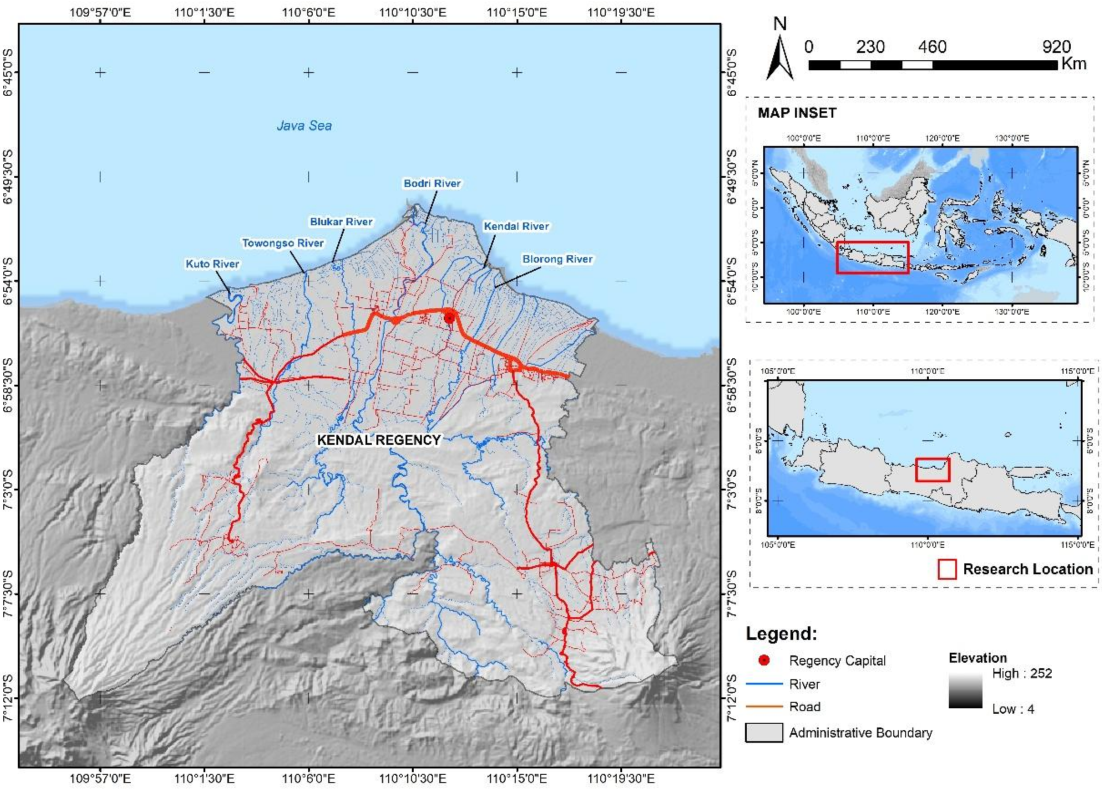


Fig. 1. Location of Kendal Regency.

2. STUDY AREA

Kendal Regency is one of the regions in Central Java Province, located on the northern coast of Java Island with an area of approximately 1,002.23 km². Geographically, Kendal Regency lies between 6°50′–7°24′ S and 109°40′–110°18′ E, bordered by the Java Sea to the north, Semarang Regency to the east, Temanggung Regency to the south, and Batang Regency to the west (BPS Kendal, 2023). This geographical setting gives Kendal diverse landscape characteristics, ranging from coastal lowlands to hilly and mountainous areas in the southern part. The research location map is shown in (**Fig. 1**).

Hydrologically, Kendal Regency is traversed by several important watersheds (DAS). The Bodri River is the main and largest river, while other significant rivers include the Blorong, Kendal, Towongso, Kuto, and Blukar Rivers, all of which flow directly into the Java Sea. These rivers serve vital functions, not only as sources of irrigation and domestic water but also as natural transport routes for materials from upstream to downstream. This condition makes Kendal's coastal areas highly vulnerable to pollution, including the accumulation of plastic waste carried by river flows. In terms of land use and land cover (LULC), Kendal Regency is dominated by agricultural land (rice fields and dry fields), settlements, industry, protected forests, and coastal zones. Industrial areas have rapidly expanded in the lowland and coastal regions, particularly in the subdistricts of Kaliwungu, Brangsong, and Kendal City, supported by the presence of the Kendal Industrial Estate (KIK) as one of the national strategic zones. The rapid growth of industry and urbanization has implications for increasing plastic waste generation, especially in areas adjacent to rivers and coastal zones (Cordova et al., 2018).

In addition to its physical characteristics, several socio-economic and waste management indicators further highlight Kendal's vulnerability to plastic pollution. Kendal Regency has a population of approximately 1.05 million inhabitants with an average density of around 1,050 persons per km^2, concentrated mainly in the northern lowlands. Urban settlements account for roughly half of the population, reflecting ongoing urban expansion along the coastal industrial corridor. The regency also hosts a substantial number of industrial facilities particularly manufacturing, food processing, and textile industries clustered around the Kendal Industrial Estate and other industrial clusters. Solid waste generation is estimated at 800-900 tons per day, yet formal waste management infrastructure, including temporary disposal sites (TPS), community-based waste processing (TPS 3R), and the regional landfill, remains limited in coverage and operational capacity. These conditions contribute to the leakage of mismanaged plastic waste into drainage networks and river systems.

This study focuses on six major rivers in Kendal Regency, namely the Kuto, Towongso, Blukar, Bodri, Kendal, and Blorong Rivers. The selection of these rivers is based on several academic considerations. In terms of environmental significance, these rivers serve as the primary transport pathways for plastic waste from upstream areas toward vulnerable coastal zones. From the perspective of socio-economic drivers, all six rivers flow through areas with intensive economic activities—including industrial zones, densely populated settlements, and commercial centers—that have a high potential for plastic leakage. Based on their hydrological characteristics, these rivers exhibit substantial discharge and watershed sizes and function as the main regional drainage channels that carry various anthropogenic materials to the sea. The final consideration is data availability, as these rivers have existing documentation and field survey records, enabling the verification of remote sensing analysis. This study utilizes Sentinel-2 imagery acquired between 30 May and 11 June, selected to coincide with the timing of field surveys conducted along the six rivers. This temporal alignment ensures consistency between the spatial conditions observed during fieldwork and the remote sensing data used in the analysis.

3. DATA AND METHODS

3.1. Sentinel-2 Satellite Imagery

This study utilizes Sentinel-2 Level-2A imagery, which provides orthorectified and atmospherically corrected surface reflectance processed using the Sen2Cor algorithm, ensuring more accurate reflectance values for spectral analysis (Louis et al., 2016). Sentinel-2, an Earth observation mission operated by the European Space Agency (ESA), is equipped with the MultiSpectral Instrument (MSI) comprising 13 spectral bands with spatial resolutions of 10 m (Blue, Green, Red, NIR), 20 m (Red Edge, SWIR), and 60 m (Coastal aerosol, Water vapor, Cirrus), a 290 km swath width, and a combined revisit time of 5 days for Sentinel-2A and Sentinel-2B (Drusch et al., 2012; Delwart, 2015). Image selection was conducted using several criteria, including a maximum cloud coverage threshold of ≤10%, availability of Level-2A surface reflectance products on Google Earth Engine (GEE), stable radiometric conditions, and acquisition dates representative of normal dry-season hydrological conditions.

The selection procedure involved filtering the *COPERNICUS/S2_SR* collection based on the study area, date range (30 May–11 June, aligned with the field survey period), and cloud percentage, followed by identifying the scene with the lowest cloud contamination. Although the images were already atmospherically corrected via Sen2Cor, additional preprocessing involved masking cloud and shadow pixels using the Sentinel-2 Scene Classification Layer (SCL), specifically removing cloud shadow, medium-probability cloud, high-probability cloud, thin cirrus, and snow/ice classes to retain only valid surface reflectance. The preprocessed imagery was then used to extract spectral information through several indices, including the Plastic Index (PI) for detecting plastic waste (Biermann et al., 2020), NDVI for vegetation assessment, NDBI for identifying built-up areas, NDWI for delineating water bodies, and the Turbidity Index (TI) for evaluating water turbidity. Together, these indices provide a comprehensive analytical framework for identifying the spatial distribution of plastic waste while accounting for the surrounding environmental conditions.

3.2. Rainfall Data from CHIRPS

This study also utilizes CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) as a source of rainfall information. CHIRPS is a quasi-global rainfall dataset developed by the Climate Hazards Group (University of California, Santa Barbara) in collaboration with the U.S. Geological Survey (USGS), covering latitudes from 50°N to 50°S. The dataset offers a spatial resolution of 0.05° (~5 km) with daily, dekadal, and monthly temporal resolutions from 1981 to the present (Funk et al., 2015). CHIRPS integrates satellite-based infrared precipitation estimates from geostationary platforms with in-situ rain gauge observations from thousands of stations worldwide, producing long-term, consistent, and improved rainfall records compared to using satellite or gauge data alone. In this study, CHIRPS is used to analyse rainfall variability across the study area, which is essential for understanding river hydrological dynamics and their relationship to plastic waste distribution in aquatic systems.

To enhance the reliability of rainfall inputs, CHIRPS has been extensively validated and applied across a wide range of hydrological and climate-related studies. Previous research has demonstrated its robustness for monitoring rainfall patterns (Paredes-Trejo et al., 2021), flood modelling in data-scarce regions (Rayamajhi et al., 2025), drought assessment (Tikuye et al., 2025), and improving runoff estimation in ungauged basins. Its strong performance in tropical humid environments has also been confirmed in Southeast Asia, further reinforcing its suitability for the present study, particularly in characterising rainfall variability and assessing its influence on riverine plastic waste dynamics in Kendal Regency.

3.3. River Data

The river data used in this study were obtained from the Environmental Agency of Kendal Regency in vector format. This dataset serves as a reference for delineating the research area, particularly in the analysis of plastic waste distribution along river channels. The rivers selected for this study are the major rivers located within Kendal Regency, as they play a significant role in transporting materials, including plastic waste, from upstream areas to downstream and eventually to the sea. By utilizing river vector data, the mapping of waste distribution can be conducted more accurately, while also enabling integration with other spatial datasets, such as land use/land cover (LULC) and plastic waste sampling points obtained from field surveys.

3.4. Land Use Land Cover (LULC) Data

Land Use/Land Cover (LULC) data in this study were obtained through the extraction of Sentinel-2 satellite imagery for the year 2024. The extraction process utilized relevant band composites to distinguish the spectral characteristics of each land cover type, allowing the identification of land use patterns along the riverbanks. The classification results were used to determine land use categories, including Dry Forest, Mixed Crops, Settlements, Industrial Buildings, Irrigated Rice Fields, Agricultural Land, Mixed Plantations, Water Bodies, Shrubs, and Plantations, which serve as important factors in understanding environmental dynamics (Aji et al., 2024).

The presence of plastic waste accumulated in river channels is often associated with the surrounding land use types. For instance, densely populated settlement areas tend to contribute more significantly to waste accumulation compared to natural vegetation areas. Sentinel-2–based LULC not only provides a spatial representation of land cover conditions but also serves as a foundation for examining the relationship between plastic waste distribution and land use characteristics along the riverbanks.

3.5. Integrated Geospatial and Machine Learning Methodology

This study employed an integrated geospatial and machine learning framework to analyse riverine plastic waste distribution, Land Use/Land Cover (LULC) patterns, and their relationship with precipitation in Kendal Regency. All spatial analyses were implemented using Google Earth Engine (GEE), enabling efficient processing of multi-source remote sensing and environmental datasets. The overall methodological workflow is illustrated in (**Fig. 2**).

3.5.1. Overall Analytical Workflow

The methodological framework consisted of four main stages: (1) extraction and preprocessing of satellite-based spectral data and rainfall information; (2) derivation of spectral indices related to vegetation, built-up areas, water bodies, turbidity, and plastic waste; (3) application of Random Forest (RF) models for both plastic index prediction and LULC classification; and (4) integration of plastic waste distribution, LULC patterns, and precipitation data to examine their spatial interrelationships along major river corridors. This integrated approach was designed to capture both anthropogenic and environmental drivers of plastic accumulation at the landscape scale.

3.5.2. Overall Analytical Workflow

Plastic waste detection was based on multispectral information derived from Sentinel-2 Level-2A imagery acquired between 30 May and 11 June 2025 and temporally aligned with field surveys. Several spectral indices were calculated to enhance the discrimination of plastic materials from other surface features.

Vegetation conditions were characterized using the Normalized Difference Vegetation Index (NDVI), defined as NDVI = (NIR−Red)/(NIR+Red), while built-up areas were highlighted using the Normalized Difference Built-up Index (NDBI), calculated as NDBI = (SWIR−NIR)/(SWIR+NIR). Water-related features were enhanced using the Normalized Difference Water Index (NDWI), expressed as NDWI = (Green−NIR)/(Green+NIR).

Turbidity conditions were represented by the Turbidity Index (TI), computed as TI = (RedEdge−Red)/(RedEdge+Red). Plastic waste distribution was quantified using the Plastic Index (PI), defined as PI = NIR/(NIR+Red), derived from spectral band ratios and index-based corrections sensitive to synthetic materials in aquatic environments. In this formulation, *Green*, *Red*, *RedEdge*, *NIR*, and *SWIR* represent the green, red, red-edge, near-infrared, and shortwave infrared bands of Sentinel-2 imagery, respectively.

All indices were spatially constrained to the study area and masked using cloud and shadow information from the Sentinel-2 Scene Classification Layer to ensure that only valid surface reflectance pixels were retained. The resulting multi-index dataset constituted the feature space for subsequent Random Forest modeling.

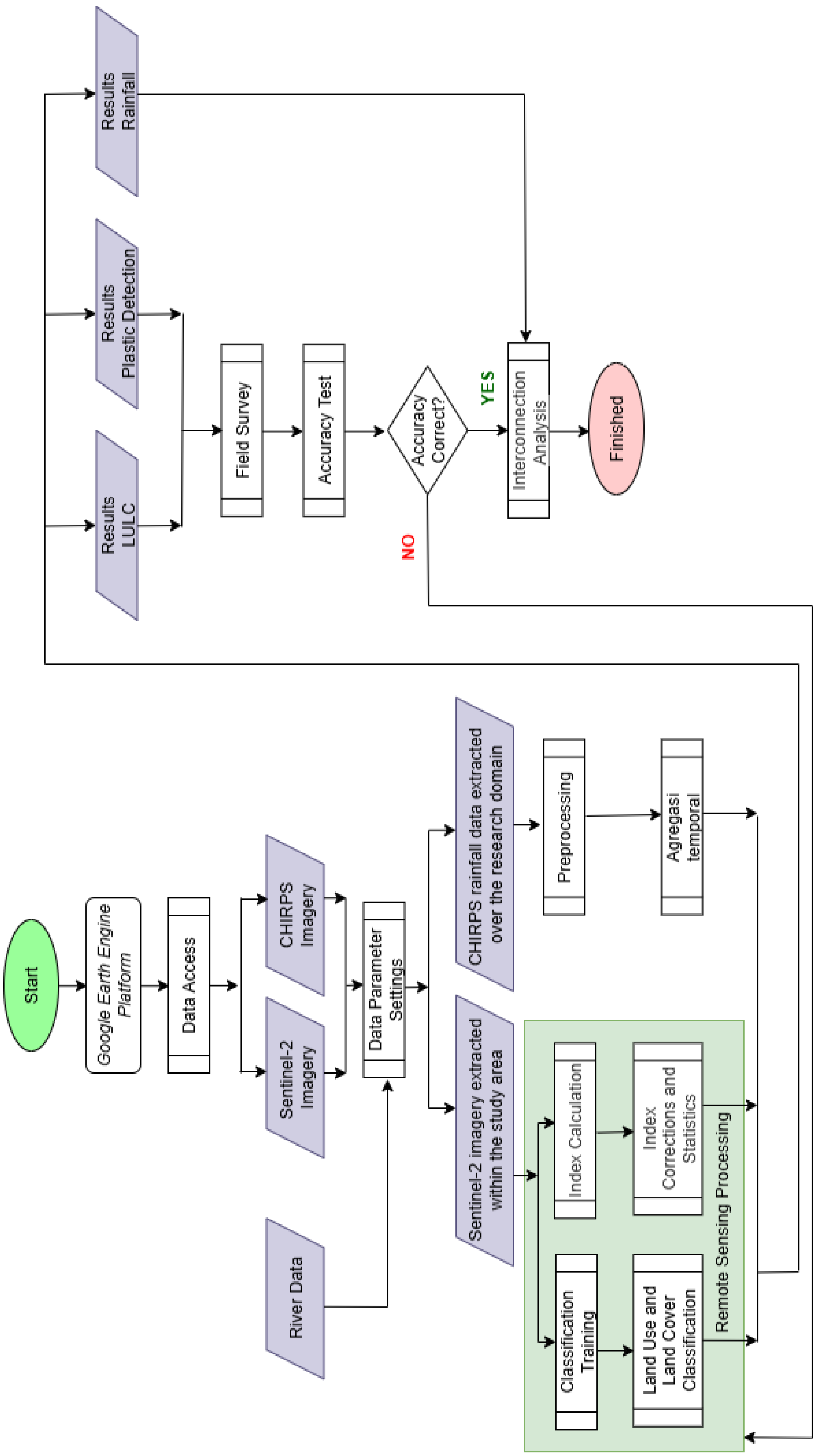


Fig. 2. Research flow.

3.5.3. Random Forest Modeling Framework

Random Forest (RF) regression was employed to model the spatial distribution of plastic waste by predicting Plastic Index (PI) values from multispectral indices. RF was selected due to its robustness in handling non-linear relationships, high-dimensional feature spaces, and multicollinearity, which are common in remote sensing data. The RF model was trained using a set of predictor variables derived from spectral indices (NDVI, NDWI, NDBI, TI) and spectral bands, with PI as the response variable. To evaluate model sensitivity to parameterization, RF models were trained using different numbers of decision trees (50, 100, and 150). Training and validation datasets were constructed using field-validated sample points collected along six major rivers. This design ensured that the model learned relationships representative of diverse hydrological and land-use contexts within the study area.

3.5.4. Model Performance Evaluation and Optimization

Model performance was assessed using three standard regression metrics: the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). MAE quantifies the average absolute difference between observed and predicted PI values, while RMSE emphasizes larger prediction errors and reflects overall model accuracy. R^2 measures the proportion of variance in observed PI values explained by the model. To improve robustness and reduce overfitting, a k-fold cross-validation approach was applied. Performance metrics were evaluated across different tree numbers to identify the optimal RF configuration. The final model selection was based on the combination of highest R^2 and lowest MAE and RMSE values, ensuring stable and reliable prediction of plastic waste distribution.

3.5.5. Model Performance Evaluation and Optimization

Land Use/Land Cover classification was conducted using Sentinel-2 imagery for the year 2024, applying a Random Forest classifier consistent with the Indonesian National Standard for land cover classification (SNI 7645:2010). Ten LULC classes were identified, including Dryland Forest, Agricultural Land, Mixed Crops, Plantations, Mixed Plantations, Shrubs, Settlements, Industrial Buildings, Irrigated Rice Fields, and Water Bodies. To ensure relevance to riverine plastic dynamics, LULC analysis was focused on a 500-m buffer surrounding major rivers. This buffer-based approach captures land-use characteristics most likely to influence plastic input, transport, and accumulation within river systems. Classification accuracy was evaluated using standard accuracy assessment metrics, including overall accuracy and the kappa coefficient.

3.5.6. Integration of Plastic Index, LULC, and Rainfall

Following validation of both plastic detection and LULC classification results, spatial integration analysis was performed to examine relationships among PI values, LULC classes, and precipitation patterns. Daily rainfall data derived from CHIRPS were temporally aggregated to match the plastic detection period and spatially aligned with river corridors. The integrated dataset enabled statistical and spatial analyses of plastic waste distribution across different land-use types and rainfall regimes. This final stage provided the basis for identifying dominant spatial drivers of plastic accumulation and for interpreting the relative influence of anthropogenic land use versus hydrometeorological variability.

4. RESULTS AND DISCUSSIONS

4.1. Map of Plastic Waste Detection Results

The results of waste detection in Kendal Regency are shown in **(Fig.3)**. The figure above presents a map of plastic waste detection in Kendal Regency. The results indicate that plastic waste is distributed across the six major rivers in the region. The detection outcomes reveal a gradient ranging from low to high concentrations, represented by a color scheme from green to red. Several sample pixels are highlighted on the map, demonstrating that different river segments exhibit varying levels of plastic waste concentration (low, medium, and high).

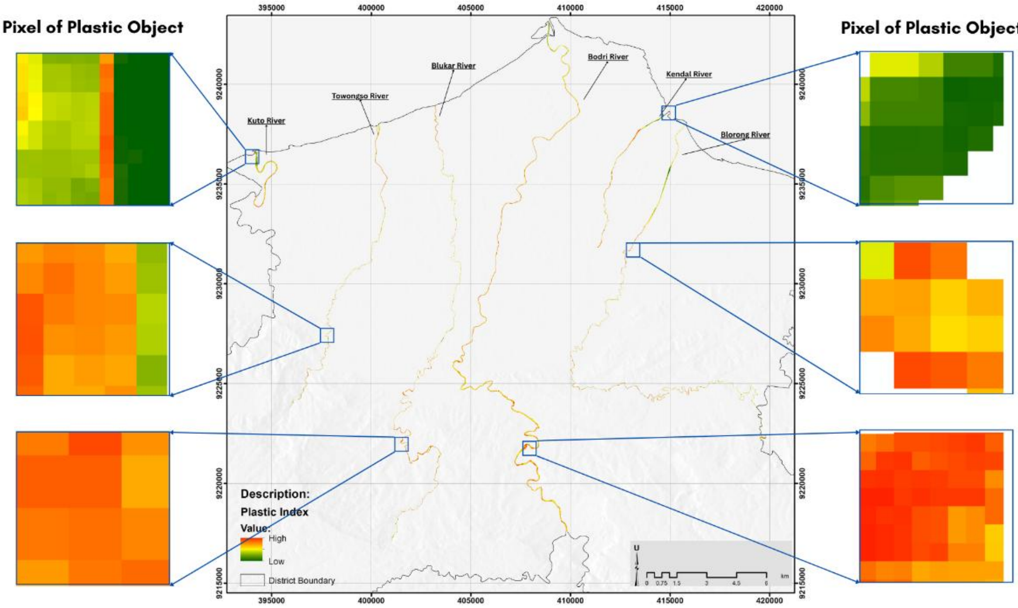


Fig. 3. Map of plastic waste distribution in six rivers in Kendal.

The visualization of plastic-related pixels indicates substantial spatial heterogeneity across river locations. The applied classification system, which employs a color gradient, provides a clear representation of the distribution patterns. Similar approaches have been widely used in remote sensing-based plastic detection research, where multispectral or satellite imagery was classified to map riverine and coastal plastic pollution (Cortesi et al., 2021; Nivedita et al., 2024). Sample pixels on different river sections show distinct spectral characteristics, emphasizing the variability of plastic pollution intensity.

The findings further suggest that plastic waste is not evenly distributed across the region. The highest concentrations (red–orange pixels) are clustered in specific areas, likely associated with anthropogenic activities such as dense settlements, industrial zones, and disposal points. In contrast, areas dominated by green pixels indicate relatively lower levels of plastic pollution, reflecting better environmental conditions. These results align with previous studies that demonstrate the strong linkage between land use, especially settlements and industrial areas, and plastic accumulation in rivers (Cerra et al., 2025; Cortesi et al., 2022). Such heterogeneity underscores the importance of targeted mitigation strategies in high-risk zones where anthropogenic pressure is most pronounced.

To validate the detection model, 147 field samples were collected along major rivers **(Fig.4)**. The field survey involved validating the detected pixels with actual plastic waste conditions, particularly waste deposited in river bends, sandbars, or other accumulation points. Accuracy assessment was conducted using three statistical metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2). The regression analysis yielded highly satisfactory results.

The MAE value of 0.0536 indicates an average prediction error of only ~0.054 units, reflecting high model accuracy. Similarly, the RMSE of 0.0728 demonstrates relatively low prediction error, consistent with the MAE value. The R^2 value of 0.8892 shows that approximately 88.92% of the variance in the observed data is explained by the model, confirming its strong predictive capability. The scatter plot **(Fig.5)** further illustrates the alignment between predicted and observed values, with most data points clustering around the 1:1 reference line, indicative of high model reliability.

Model performance was further assessed through cross-validation and hyperparameter tuning to ensure robustness. A 3-fold cross-validation approach was employed to reduce bias from single data partitioning. Performance was consistently evaluated using MAE, RMSE, and R^2 metrics, and averaged across folds as the final performance indicators.

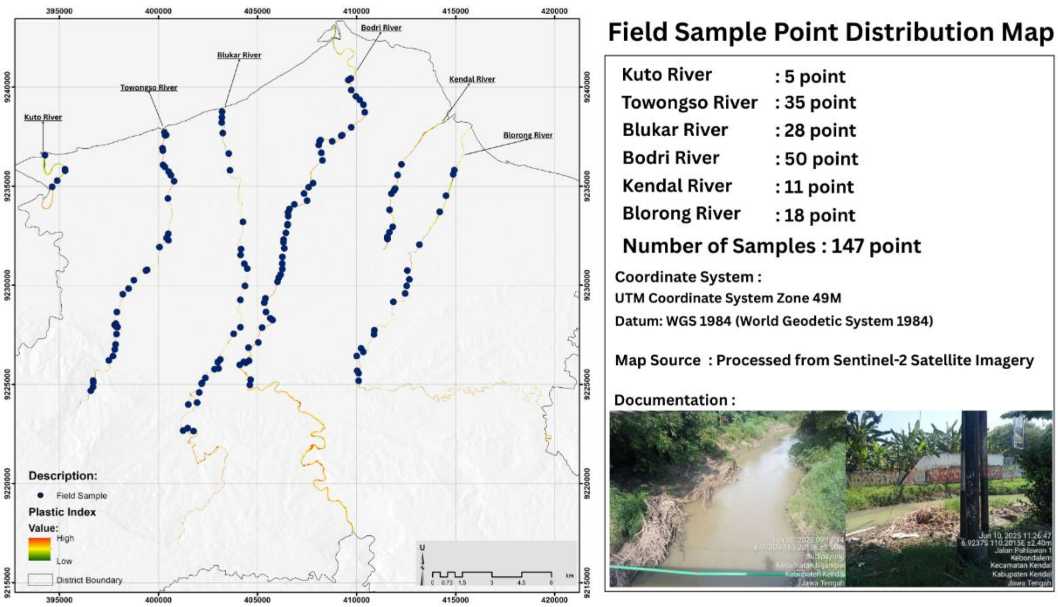


Fig. 4. Field Sample Point Distribution Map.

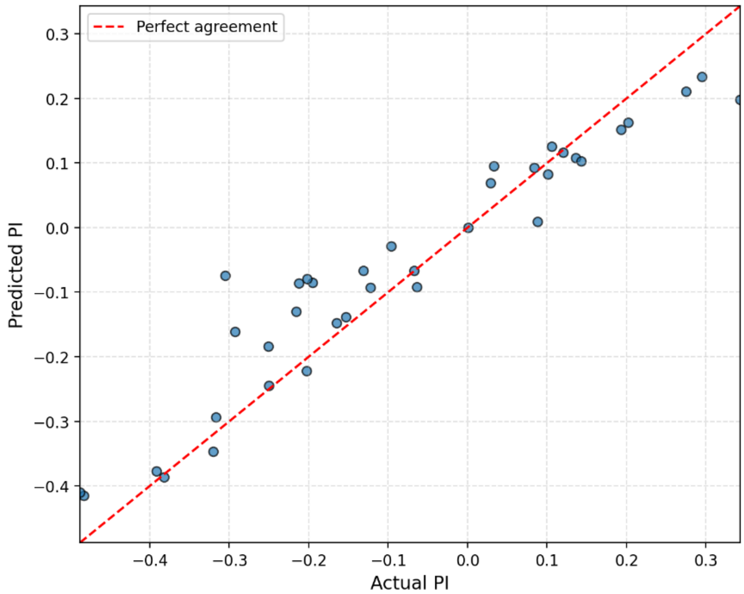


Fig. 5. Graph of Predicted vs. Actual Model Values.

The evaluation of the Random Forest model with varying numbers of trees is shown in **(Fig.6-8)**. The results demonstrate a clear improvement in predictive performance with increasing tree counts. As shown in **(Fig.6)**, the mean R^2 value improved from 0.883 (50 trees) to 0.888 (100 trees), and further to 0.890 (150 trees). Similarly, the average MAE exhibited a declining trend **(Fig.7)**, decreasing from 0.0586 (50 trees) to 0.0576 (100 trees), and finally to 0.0574 (150 trees). The RMSE values followed a comparable pattern **(Fig.8)**, with values decreasing from 0.0725 (50 trees) to 0.0715 (100 trees), and further to 0.0710 (150 trees). These results highlight that increasing the number of trees in the Random Forest algorithm contributes to more stable and accurate predictions of the Plastic Index (PI), which is consistent with findings in other machine learning-based remote sensing studies (Cortesi et al., 2021, 2022).

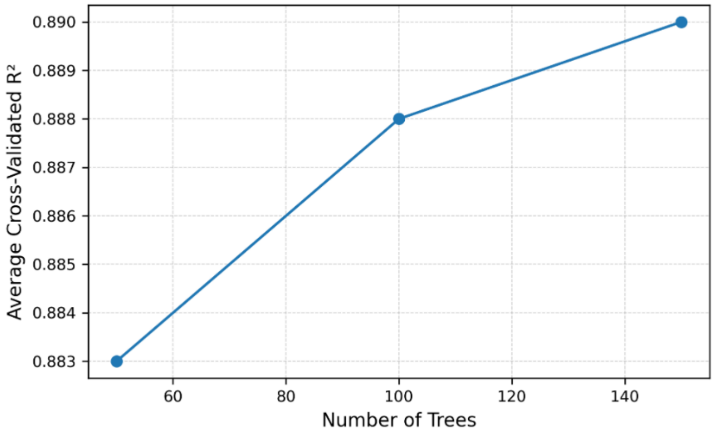


Fig. 6. Cross-Validation Graph.

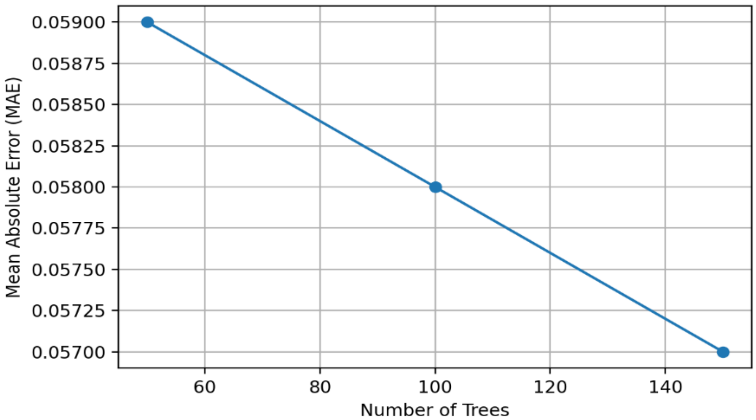


Fig. 7. Graph of the Relationship between Mean Absolute Error and Number of Trees.

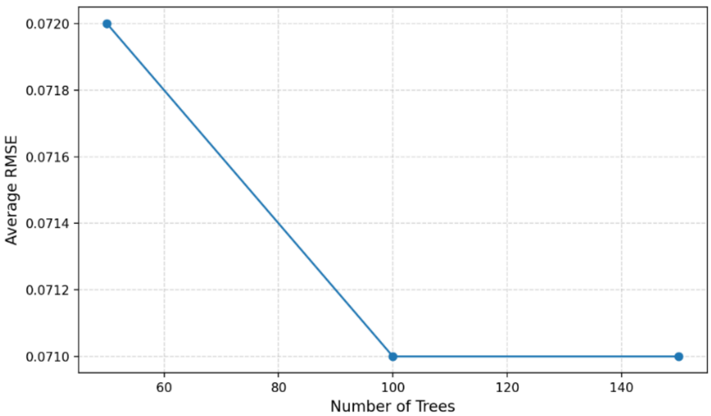


Fig. 8. Graph of the Relationship between Root Mean Square Error and Number of Trees.

Overall, the model evaluation confirms that the adopted Random Forest approach is both robust and reliable in detecting plastic waste distribution from satellite imagery. The strong predictive performance, validated through statistical metrics and cross-validation, demonstrates the potential of machine learning-based remote sensing techniques in supporting sustainable riverine waste management (Nivedita et al., 2024).

4.2. Land Use Land Cover (LULC) Mapping

The LULC classification for 2024 was conducted using Sentinel-2 imagery, referring to the Indonesian National Standard (SNI) 7645 of 2010 on Land Cover Classification. The standard was applied at a scale of 1:25,000 with ten land cover classes, namely: Dry Forest, Agricultural Land, Mixed Crops, Plantations, Mixed Plantations, Shrubs, Settlements, Industrial Buildings, Irrigated Rice Fields, and Water Bodies. The accuracy assessment results showed an overall accuracy of 98% and a kappa coefficient of 97%. These values are higher compared to the previous classification, which had an accuracy of 89% and a kappa of 85%. This improvement indicates that Sentinel-2-based classification methods provide more reliable results for LULC mapping (Belgiu & Drăguț, 2016; Phiri et al., 2020). Based on the Land Use/Land Cover (LULC) map of Kendal Regency in 2024 **(Fig. 9)**, the spatial distribution of land cover exhibits considerable complexity. In the northern coastal areas, Water Bodies (blue) and Irrigated Rice Fields (light green) dominate, reflecting intensive rice cultivation activities and the presence of coastal water bodies. Meanwhile, the central to southern parts of the region are dominated by Agricultural Land, Mixed Crops, and Plantations, consistent with Kendal Regency's characteristics as an agriculture- and plantation-based area (SNI, 2010).

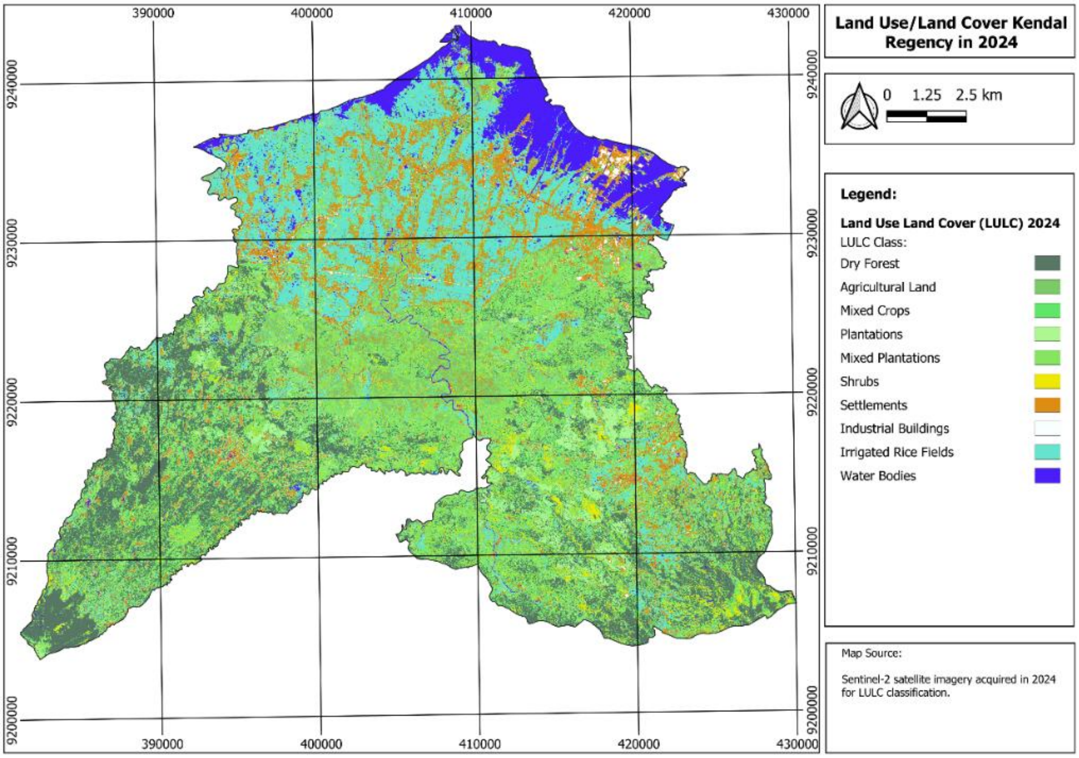
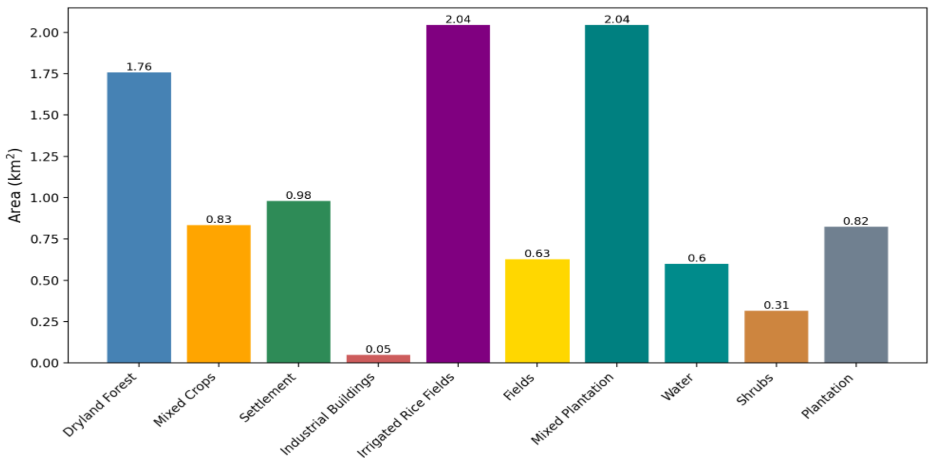
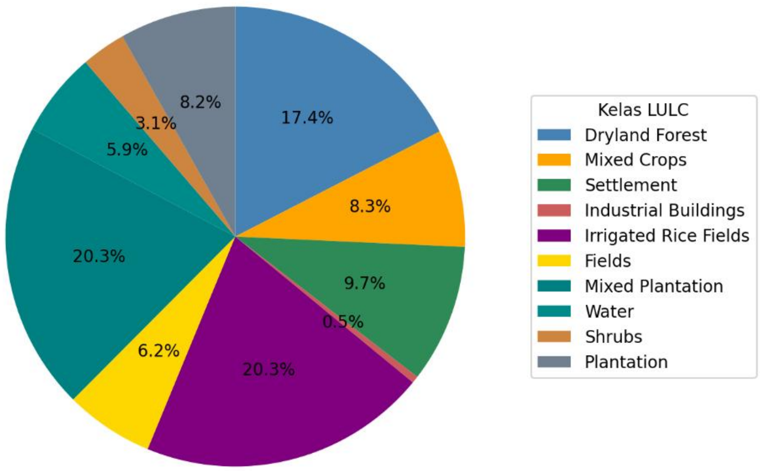


Fig. 9. LULC Map of Kendal Regency in 2024.

The Settlements and Industrial Buildings classes appear concentrated along major transportation corridors and urban centers, indicating increased urbanization activities. This distribution also highlights development pressures on ecological balance, particularly around major rivers (Chen et al., 2024). In contrast, Shrubs and Mixed Plantations appear sporadically, signaling land degradation or transitional land use. Furthermore, the area-based analysis of each land cover class reveals clear quantitative differences that reflect the region's landscape structure **(Fig. 10)**. The largest land cover categories are Mixed Plantation and Irrigated Rice Fields, each covering approximately 204 km² (20.3%), highlighting the dominance of plantation systems and irrigated agriculture across Kendal Regency. Dryland Forest accounts for around 176 km² (17.4%), indicating the presence of significant forested areas, particularly in the upland southern regions that contribute to ecological stability.



a)



b)

Fig. 10. Distribution of land use and land cover (LULC) classes in the study area: (a) total area (ha) and (b) relative proportion (%).

The Settlement class occupies about 97 km² (9.7%), while Mixed Crops covers roughly 82 km² (8.3%), reflecting the distribution of residential and diversified agricultural zones in the mid-regency areas. Plantation areas comprise around 81 km² (8.2%), further emphasizing the importance of estate-crop cultivation in the local economy. Meanwhile, Fields extend over approximately 62 km² (6.2%), representing areas dominated by seasonal cropping systems. Water Bodies, primarily found in the northern coastal zone, cover around 59 km² (5.9%), consistent with the presence of coastal waters and aquaculture ponds. Shrubs, representing transitional or degraded land, occupy the smallest area of roughly 30 km², appearing in scattered patches across the region. Industrial Buildings cover a minimal area of about 3 km², but are spatially clustered along major corridors, consistent with the concentration of manufacturing zones. The Area of Interest (AOI) of the study was focused on areas surrounding major rivers in Kendal Regency **(Fig. 11)**. To obtain a land cover distribution representation relevant to river dynamics, LULC analysis was conducted within a 500-meter buffer from each major river. This approach was chosen to ensure that the mapping results align with the study's objective, namely to examine the relationship between land cover distribution and aquatic environmental conditions, particularly concerning the potential distribution of plastic waste in the study area.

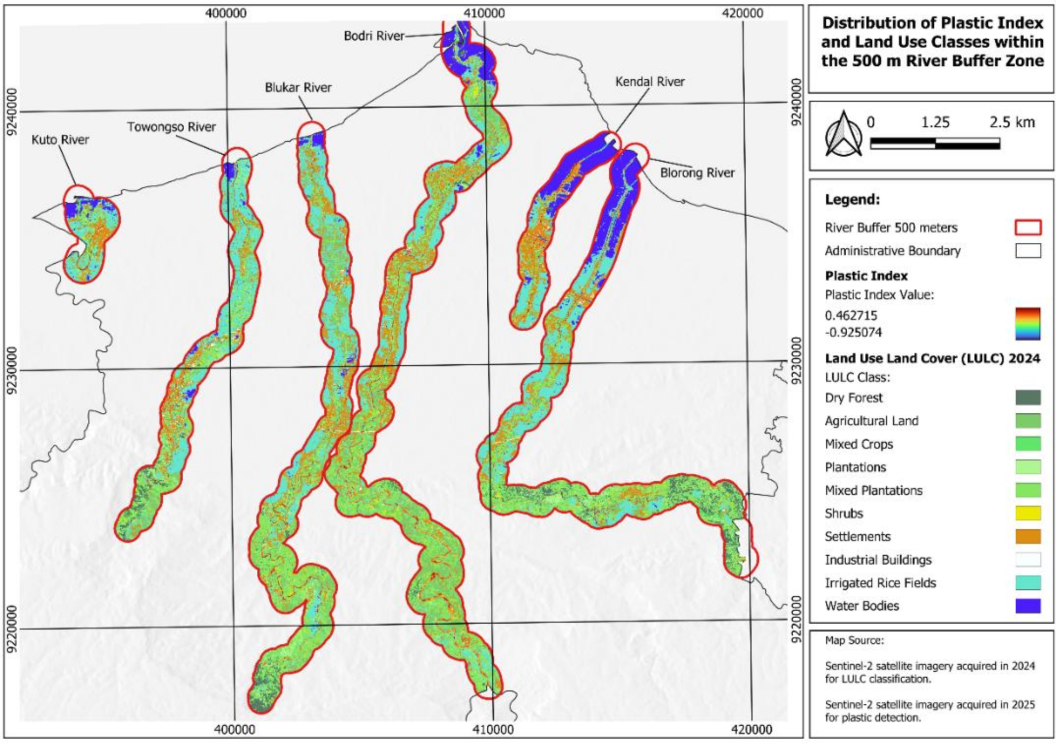


Fig. 11. Map of Plastic Waste Distribution and 500-Meter LULC Buffer.

4.3. The Relationship between Plastic Waste Detection Locations, LULC, and Rainfall

In this analysis, LULC classes include Dryland Forest, Mixed Crops, Settlements, Industrial Buildings, Irrigated Paddy Fields, Dry Fields, Mixed Plantations, Water Bodies, Scrub, and Plantations. However, the current dataset only contains observations for Dryland Forest, Settlements, Irrigated Paddy Fields, Mixed Plantations, Water Bodies, and Scrub. The statistical analysis of PI values across observed LULC classes highlights a strong anthropogenic influence on plastic accumulation patterns. Settlements exhibit the highest median PI values (approximately +0.10 to +0.15), which directly reflects household waste generation, commercial activities, and inadequate solid waste management systems. This class serves as the primary source zone for plastic pollution in the landscape **(Fig. 12)**.

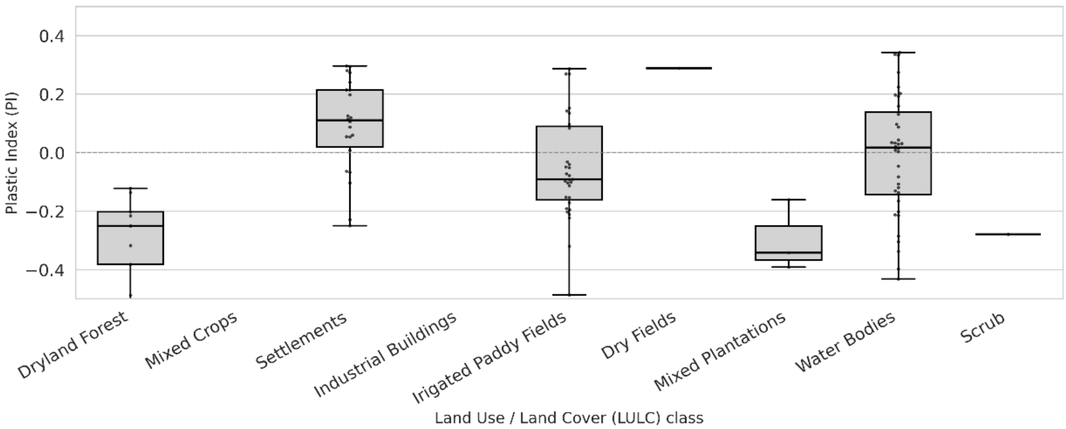


Fig. 12. Distribution of Plastic Index (PI) across LULC classes using Boxplot and Swarmplot.

Water Bodies display the greatest spatial variability in PI values, with observations ranging from approximately -0.45 to +0.35. While the median value remains close to zero (~+0.02), the presence of extreme positive outliers (up to +0.35) indicates that certain water bodies act as significant accumulation sinks, receiving plastics transported by surface runoff, drainage, and riverine flow. This heterogeneity suggests that water bodies with high hydrological connectivity to settlements or agricultural areas experience localized hotspots of plastic accumulation, whereas more isolated or upstream water bodies remain relatively unaffected. In contrast, negative PI values are consistently observed across vegetated and less-disturbed landscapes. Mixed Plantations show the lowest median PI (~-0.35), followed by Scrub (median ~-0.28) and Dryland Forest (median ~-0.25). These negative values reflect the buffering role of vegetated land covers, where vegetation acts as a physical barrier to plastic transport and accumulation. The lower levels of direct human activity in these areas also contribute to reduced plastic inputs.

Irrigated Paddy Fields demonstrate moderate negative median PI values (around -0.10) with relatively low variability. This pattern suggests that while irrigated agriculture is less contaminated than settlements, the irrigation infrastructure and water management practices may still introduce some plastic materials, particularly if irrigation water is sourced from potentially contaminated water bodies or located near waste-generating areas. Notably absent from the analysis are Mixed Crops, Industrial Buildings, Dry Fields, and Plantations. The lack of observations in these classes may indicate either their limited spatial extent in the study area or sampling gaps. Future research should prioritize targeted sampling in these land use types, particularly Mixed Crops and Industrial Buildings, which are expected to show elevated PI values based on their anthropogenic character.

Overall, the results demonstrate a clear land use gradient in plastic pollution: human-dominated landscapes (Settlements) function as primary sources with elevated PI values, Water Bodies serve as dynamic transport and variable accumulation zones, while vegetated and less-disturbed areas (Mixed Plantations, Scrub, Dryland Forest) consistently show negative PI values, acting as natural buffers. This spatial pattern underscores the critical importance of integrated land use planning, improved solid waste management in settlement areas, and the preservation of vegetated buffers to mitigate plastic pollution at the landscape scale.

The application of locally weighted scatterplot smoothing (LOWESS) regression to the relationship between plastic index (PI) and land use/land cover (LULC) classes reveals a distinct non-linear pattern that reflects the complex interplay between anthropogenic activities and plastic accumulation dynamics across the landscape **(Fig. 13)**. The smoothed trend line demonstrates substantial variability in PI values across the LULC gradient, with peaks corresponding to human-dominated land uses and troughs associated with vegetated or semi-natural landscapes.

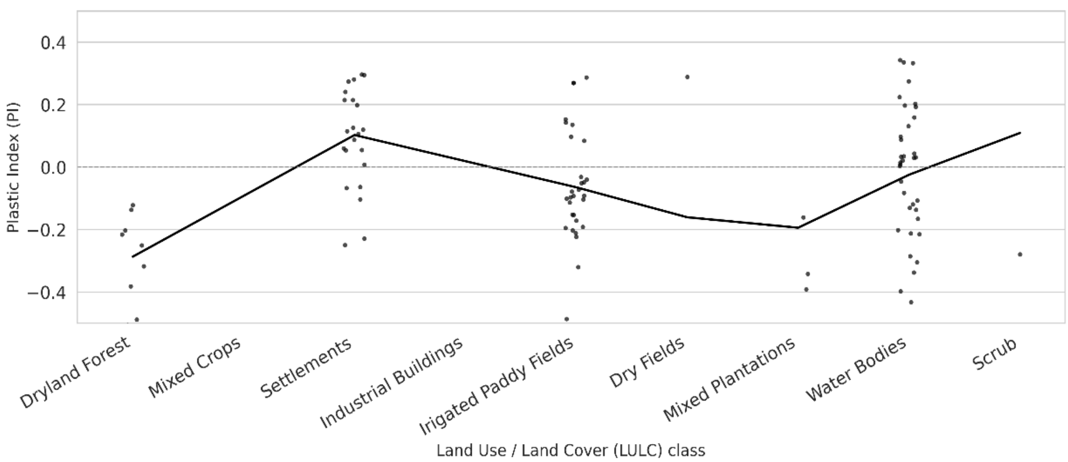


Fig. 13. Relationship between Plastic Index (PI) and Land Use/Land Cover (LULC).

The LOWESS curve exhibits an initial increase from Dryland Forest (PI ≈ -0.30) to a pronounced maximum at Settlements (PI ≈ +0.10 to +0.15), representing an increase of approximately 0.40 to 0.45 units across this transition. This sharp gradient underscores the fundamental role of residential and commercial areas as primary point sources of plastic pollution in the study landscape. The elevated PI values observed in settlement areas are consistent with previous studies demonstrating the correlation between population density, inadequate municipal solid waste management infrastructure, and environmental plastic loads (Jambeck et al., 2015; Lebreton et al., 2017). The considerable scatter of individual observations around Settlements (ranging from approximately -0.05 to +0.30) suggests significant intra-class heterogeneity, likely attributable to variations in waste management practices, population density, and socioeconomic factors across different settlement types within the study area.

Following the peak at Settlements, the LOWESS trend exhibits a declining trajectory through Irrigated Paddy Fields, reaching a local minimum at Mixed Plantations (PI ≈ -0.20 to -0.35). The negative PI values observed in Irrigated Paddy Fields (median PI ≈ -0.10) indicate relatively low plastic contamination despite their position in the agricultural production landscape. This pattern may reflect the combined influence of regular water management practices that facilitate plastic transport and removal, as well as the lower intensity of plastic-generating activities compared to settlements. The further decline toward Mixed Plantations demonstrates the buffering capacity of plantation systems, where dense vegetation cover and reduced anthropogenic disturbance limit plastic accumulation through physical interception and reduced waste inputs (Browne et al., 2011; Eerkes-Medrano et al., 2015)

A notable secondary increase in the LOWESS curve occurs at Water Bodies, where the smoothed trend approaches near-neutral values (PI ≈ -0.05 to 0.00). However, this class exhibits the greatest dispersion of individual observations, with values spanning the entire analytical range from -0.45 to +0.35. This extreme variability reflects the dual nature of aquatic systems as both transport corridors and accumulation sinks for plastic debris (Ballent et al., 2016; Van Emmerik & Schwarz, 2020). The presence of substantial positive outliers (PI > +0.20) indicates that specific water bodies within the study area function as significant plastic accumulation hotspots, likely resulting from convergent flow patterns, proximity to upstream pollution sources, limited flushing capacity, or the presence of physical retention features such as debris dams or low-energy depositional zones. Conversely, negative PI values in other water bodies suggest effective transport and removal processes, potentially associated with higher flow velocities, greater hydrological connectivity to downstream systems, or isolation from direct pollution inputs. This heterogeneity underscores the importance of site-specific assessment rather than class-level generalizations when evaluating plastic contamination in aquatic environments.

The final segment of the LOWESS curve demonstrates a sharp decline to strongly negative values at Scrub (PI ≈ -0.30). Scrubland areas, characterized by low anthropogenic pressure and moderate vegetation cover, consistently exhibit low plastic accumulation. The relatively constrained scatter of observations within this class indicates more homogeneous environmental conditions and reinforces the role of semi-natural vegetation as an effective barrier to plastic transport and deposition. The non-linear relationship revealed by the LOWESS analysis has important implications for understanding landscape-scale plastic pollution dynamics. The oscillating pattern, with distinct peaks at Settlements and Water Bodies separated by troughs at vegetated land covers (Dryland Forest, Mixed Plantations, Scrub), demonstrates that plastic accumulation cannot be adequately described by simple linear models or uniform urban-to-rural gradients. Instead, the spatial distribution of plastic pollution reflects a mosaic pattern determined by the specific characteristics of each land use type, their spatial configuration, and the connectivity pathways that facilitate plastic transport between source and sink areas.

From a management perspective, these findings highlight three priority intervention points: (1) enhanced solid waste management and behavioral change programs in Settlement areas to reduce plastic inputs at the source; (2) strategic preservation and expansion of vegetated buffer zones, particularly Mixed Plantations and Scrubland, to intercept plastic transport pathways; and (3) targeted monitoring and remediation efforts in high-accumulation water bodies identified through the substantial positive outliers observed in Water Bodies. The strong buffering effect demonstrated by vegetated land covers (Dryland Forest, Mixed Plantations, and Scrub) supports nature-based solutions approaches that integrate riparian vegetation management and green infrastructure into broader plastic pollution mitigation strategies.

The LOWESS regression analysis of the relationship between precipitation and plastic index (PI) reveals a complex, non-monotonic pattern that deviates substantially from simple linear assumptions **(Fig. 14)**. The smoothed trend line demonstrates multiple inflection points across the precipitation gradient, suggesting that the influence of rainfall on plastic accumulation operates through multiple, competing mechanisms that vary in dominance across different precipitation regimes. At the lower end of the precipitation spectrum (0-10 mm), the LOWESS curve exhibits moderate positive PI values (approximately +0.05), followed by a sharp decline to a local minimum at approximately 10 mm (PI ≈ -0.15). This initial negative trend may reflect the threshold mobilization effect, whereby moderate rainfall events possess sufficient energy to initiate plastic transport from terrestrial surfaces but insufficient volume to facilitate long-distance transport or dilution. The substantial scatter of individual observations within this range (spanning from approximately -0.40 to +0.35) indicates high spatial variability in local hydrological conditions, surface characteristics, and proximity to plastic sources.

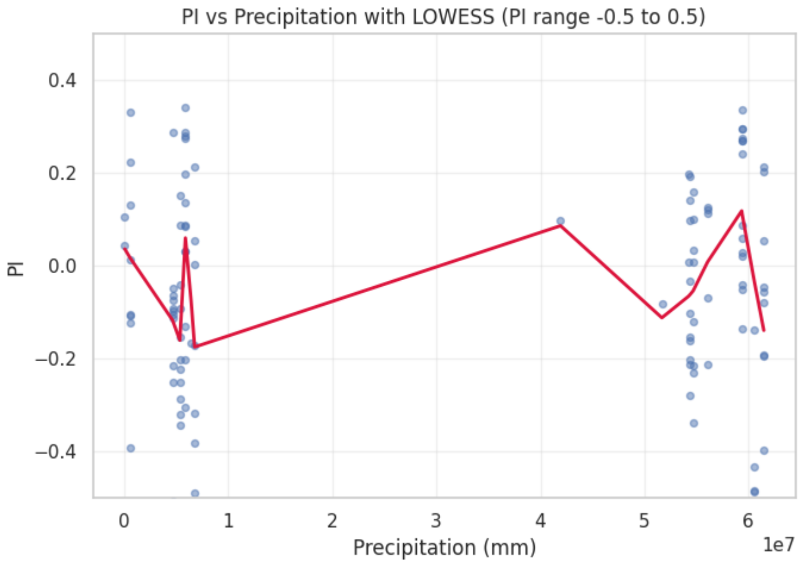


Fig. 14. Relationship between Plastic Index (PI) and Precipitation.

The trend line subsequently exhibits a gradual increase from the 10 mm minimum through the mid-range precipitation values (10-40 mm), reaching a secondary peak at approximately 40-45 mm (PI ≈ +0.10). This ascending limb of the curve suggests that moderate-to-high precipitation events may enhance plastic accumulation through several potential mechanisms. First, increased runoff volume during substantial rainfall events can transport plastics from diffuse upland sources to convergent flow paths and depositional zones within the landscape (Hurley et al., 2018). Second, higher precipitation totals may be associated with longer-duration storm events that facilitate the breakdown of larger plastic items into smaller fragments through physical abrasion, thereby increasing detection probability via spectral indices such as PI (Corcoran et al., 2009). Third, areas receiving moderate rainfall may support land use practices that generate plastic waste, such as intensive agriculture with plastic mulching or greenhouse cultivation (Steinmetz et al., 2016).

Following this secondary maximum, the LOWESS curve displays a pronounced decline to near-zero or slightly negative values at approximately 50 mm (PI ≈ -0.10), followed by another sharp increase to a tertiary peak at approximately 60-65 mm (PI ≈ +0.10 to +0.15). This oscillating pattern in the high precipitation range is particularly intriguing and may reflect the complex interplay between transport capacity, dilution effects, and spatial patterns of land use. High-precipitation areas may experience enhanced flushing of plastic materials through well-developed drainage networks, leading to reduced local accumulation but increased downstream transport (Van Emmerik et al., 2018). Alternatively, the positive PI values observed at the highest precipitation levels could indicate areas where intense rainfall concentrates transported plastics in specific depositional environments, such as floodplains, retention basins, or coastal zones that serve as terminal sinks.

The substantial vertical scatter of observations throughout the precipitation gradient, particularly evident in the 0-20 mm and 55-65 mm ranges, underscores the importance of factors beyond precipitation in determining plastic accumulation patterns. This variability likely reflects the confounding influences of land use intensity, population density, waste management infrastructure, topography, soil characteristics, and vegetation cover all of which modulate the relationship between rainfall and plastic transport-deposition dynamics (Horton et al., n.d.; Windsor et al., 2019).

From a process-based perspective, the non-linear relationship revealed by this analysis suggests that plastic mobilization and accumulation across precipitation gradients cannot be adequately described by simple hydrological transport models. Instead, the data indicate threshold-dependent behaviors and regime shifts that likely correspond to distinct hydrological states: (1) a low-precipitation regime characterized by limited transport capacity and localized accumulation near sources; (2) a moderate-precipitation regime where enhanced runoff facilitates both mobilization and redeposition; and (3) a high-precipitation regime where competing processes of transport efficiency, dilution, and concentrated deposition produce spatially heterogeneous outcomes.

The management implications of these findings are significant. The positive association between certain precipitation ranges and elevated PI values suggests that plastic pollution monitoring and mitigation efforts should be intensified in areas experiencing moderate-to-high rainfall, particularly following significant storm events when transport and redeposition processes are most active. Furthermore, the high variability observed across all precipitation levels indicates that precipitation alone is insufficient to predict plastic accumulation patterns; effective management strategies must integrate hydrological factors with comprehensive land use planning, improved waste collection systems in high-rainfall areas, and strategic placement of sediment and debris retention structures to intercept transported plastics before they reach sensitive aquatic environments.

The oscillating LOWESS pattern also raises important methodological considerations for future research. The presence of multiple local maxima and minima suggests potential threshold effects or regime shifts that warrant investigation through more sophisticated statistical approaches, such as segmented regression or mixture models, which can explicitly test for discontinuities in the precipitation-PI relationship. Additionally, temporal considerations such as antecedent moisture conditions, rainfall intensity versus duration, and seasonal patterns may further refine our understanding of how precipitation influences plastic accumulation dynamics.

5. DISCUSSION

Although the analysis was conducted over a relatively short observation period of 13 days, the results remain meaningful for interpreting the spatial distribution of plastic waste in relation to land use and land cover (LULC) characteristics. This is because LULC configuration represents a relatively stable structural control that governs plastic accumulation patterns, particularly along river corridors and adjacent built-up areas. While seasonal variability may influence the magnitude and mobility of plastic waste under different rainfall regimes, the underlying spatial relationships between plastic accumulation hotspots and dominant land-use types are expected to persist over time. Wet-season conditions are likely to enhance plastic transport and redistribution through increased runoff, whereas dry-season conditions may promote localized exposure and accumulation within river channels. Consequently, the short observation window primarily limits temporal generalization rather than spatial inference.

The complex oscillating pattern observed in the precipitation Plastic Index (PI) relationship (**Fig. 14**), characterized by multiple local maxima and minima without a consistent directional trend, likely reflects this temporal constraint. The pronounced vertical scatter across the precipitation gradient further indicates that temporal factors such as antecedent moisture conditions, rainfall intensity versus duration, and seasonal hydrological processes may exert a stronger influence on plastic transport dynamics than spatial precipitation totals alone. These findings suggest that precipitation effects on plastic distribution are event-driven and season-dependent rather than spatially uniform.

In addition to temporal limitations, the spatial resolution of Sentinel-2 imagery (10–20 m) constrains the detection of small plastic fragments and the precise delineation of narrow riparian zones, as illustrated in Figure 3. Pixel-level aggregation may obscure sub-pixel heterogeneity in heterogeneous riverine environments, where localized plastic accumulation can be highly variable. Mixed pixels containing both contaminated and uncontaminated surfaces may yield intermediate PI values, potentially underestimating localized contamination and contributing to the residual scatter observed in the Random Forest model (**Fig. 5**; $R^2 = 0.8892$). Future studies should therefore integrate higher-resolution satellite data (e.g., PlanetScope at 3–5 m, WorldView at 0.3–2 m) or UAV-based multispectral imagery to improve spatial detail, detection accuracy, and the characterization of fine-scale plastic accumulation patterns.

Spectral confusion with similar materials remains a challenge. Dried vegetation, bright soil surfaces, sand deposits, construction materials, and light-colored substrates may exhibit reflectance characteristics that partially overlap with plastic materials. While the Random Forest model incorporating multiple spectral indices (NDVI, NDWI, NDBI, TI) achieves high overall accuracy ($R^2 = 0.8892$), residual spectral confusion cannot be entirely eliminated. Some positive PI values shown in Figure 3, particularly in areas with mixed land uses, may reflect spectral confusion rather than actual plastic accumulation. This issue is compounded for weathered or degraded plastics, which exhibit altered spectral properties compared to fresh materials. Future work should employ hyperspectral imaging to discriminate materials based on narrow-band spectral features and develop comprehensive spectral libraries of interfering materials under varying environmental conditions.

Critical hydrological parameters were not incorporated into the analysis. The absence of river discharge and flow velocity data limits mechanistic interpretation of plastic transport dynamics and the extreme variability observed in Water Bodies (**Fig. 12**: PI ranging from -0.45 to +0.35). As discussed in the LOWESS analysis of Water Bodies **(Fig. 14)**, the presence of substantial positive outliers (PI > +0.20) likely reflects convergent flow patterns, limited flushing capacity, or physical retention features. However, without discharge and velocity measurements, distinguishing between high-accumulation sites resulting from natural hydrological convergence versus those influenced by anthropogenic alterations remains challenging. Additionally, unmapped hydraulic structures (dams, weirs, check dams, water gates) may create artificial accumulation hotspots by altering flow regimes and creating zones of reduced velocity where plastics settle. The extreme positive PI outliers identified in certain water bodies may partially reflect the influence of such unmapped structures rather than solely natural processes. Future research should integrate continuous hydrological monitoring, comprehensive infrastructure mapping, and process-based modeling to develop mechanistic understanding of plastic transport pathways and retention mechanisms.

The study did not differentiate plastic types, sizes, or polymer compositions. The aggregate PI values shown in Figures 11-12 represent a composite measure that cannot distinguish between macroplastics (>5 mm), mesoplastics (5-1 mm), and microplastics (<1 mm), which exhibit fundamentally different mobility patterns and environmental behaviors. Different polymer types possess distinct density, buoyancy, and degradation characteristics that influence transport dynamics and spatial distribution. As noted in the precipitation analysis **(Fig. 14)**, enhanced runoff during substantial rainfall events may facilitate breakdown of larger items into smaller fragments through physical abrasion. However, without size and polymer differentiation, these processes cannot be quantified. Future research should incorporate field sampling and spectroscopic analysis (FTIR, Raman spectroscopy) for detailed plastic characterization and source attribution.

Socioeconomic variables (income levels, waste collection coverage, population density, behavioral factors) were not integrated but likely explain substantial intra-class variability. As demonstrated in Figure 11, Settlements exhibit a wide PI range (-0.05 to +0.30), with considerable scatter around the class median. The LOWESS analysis **(Fig. 13)** further confirms this heterogeneity through the substantial vertical scatter of individual observations around Settlements, which we attributed to "variations in waste management practices, population density, and socioeconomic factors across different settlement types." However, without high-resolution infrastructure data (locations of waste collection points, landfills, informal dump sites), the specific neighborhood-level factors driving this variability cannot be identified. The discussion of management implications emphasized targeting "densely populated settlement zones exhibiting the most extreme positive PI values," but operationalizing this recommendation requires socioeconomic data integration. Future studies should incorporate census data, municipal waste management records, and participatory mapping to explain intra-settlement variability and provide actionable policy insights.

Missing data for Mixed Crops, Industrial Buildings, Dry Fields, and Plantations limits landscape-scale comprehensiveness. As noted in both **Figures 12-14**, only 6 of 10 LULC classes contain observations, creating gaps in understanding the complete spectrum of LULC-plastic relationships. Industrial Buildings are of particular concern, as they are expected to exhibit elevated PI values based on their association with manufacturing activities and packaging waste. The absence of these classes creates uncertainty regarding total anthropogenic contribution and may affect the management recommendations, which currently focus on Settlements as the primary intervention target without assessing industrial contributions. Future studies should employ stratified sampling designs ensuring adequate representation of all major LULC classes, particularly those with high anthropogenic activity.

The spatial pattern analysis, while revealing non-linear relationships between PI, LULC, and precipitation, does not address the connectivity pathways that facilitate plastic transport between source and sink areas. The LOWESS analysis **(Fig. 13)** demonstrates an oscillating pattern with peaks at Settlements and Water Bodies separated by troughs at vegetated areas, suggesting a mosaic distribution pattern. However, understanding how plastics move from settlement sources through the landscape matrix to accumulate in specific water bodies requires spatial connectivity analysis incorporating topography, drainage networks, and flow routing. The management recommendation to establish "vegetated buffer zones to intercept plastic transport pathways" presumes knowledge of these pathways, yet the current analysis does not explicitly map them. Future research should employ landscape connectivity modeling and network analysis to identify critical transport corridors and optimal buffer placement locations.

Future research should prioritize systematic multi-temporal monitoring to better capture seasonal dynamics and precipitation-driven plastic transport processes. Quarterly or seasonal time-series analyses would allow robust comparisons between wet and dry periods and improve understanding of temporal variability in plastic accumulation patterns. In addition, rainfall-event–based analyses focusing on extreme precipitation and high-runoff events are strongly recommended, as such events are likely to dominate plastic redistribution within river systems. Integrating remote sensing–derived plastic indices with hydrological modeling including discharge estimation, flow routing, and hydraulic structure characterization would further enable process-based interpretation and predictive assessment of plastic transport pathways.

Beyond these core priorities, methodological improvements should include the use of higher spatial resolution data (e.g., UAV platforms or ≤5 m satellite imagery) to enhance detection of small plastic fragments and narrow riparian zones, as well as hyperspectral imagery and spectral library development to reduce false positives from spectrally similar materials. Additional advances may be achieved through plastic characterization by polymer type, size class, and degradation state via combined field sampling and laboratory analysis; integration of high-resolution socioeconomic and infrastructure datasets to explain intra-class variability and identify intervention priorities; stratified field sampling ensuring representation of all LULC classes, particularly Mixed Crops and Industrial Buildings; and spatial connectivity analysis incorporating topography and drainage networks to map plastic transport pathways. Scenario-based modeling, evaluating the effectiveness of management interventions such as improved waste collection, debris interception structures, and riparian buffers under varying land-use and hydrological regimes, would further enhance the applicability of remote sensing-based plastic monitoring frameworks.

6. CONCLUSIONS

This study demonstrates that the spatial distribution of plastic waste in Kendal Regency is predominantly governed by land use and land cover (LULC) characteristics rather than by precipitation dynamics. Plastic Index (PI) values exhibit clear gradients across LULC types, with the highest concentrations occurring in settlement areas reflecting intensive human activity, inadequate waste management infrastructure, and strong source zone characteristics. Water bodies function as dynamic transport corridors and accumulation zones, displaying extreme variability due to differences in hydrological connectivity, flow conditions, and proximity to upstream anthropogenic inputs. In contrast, vegetated landscapes such as dryland forests, mixed plantations, and scrub consistently exhibit negative PI values, confirming their role as natural buffer zones that limit plastic transport and deposition.

The LOWESS analyses further reveal that the relationship between precipitation and PI is highly non-linear, characterized by multiple inflection points and substantial scatter, with no significant or consistent trend. These patterns indicate that short-term rainfall totals measured independently of hydrological processes such as flow velocity, runoff pathways, and storm intensity do not exert a dominant influence on plastic accumulation within the study period. Instead, plastic distribution is shaped by spatially explicit drivers, including land cover configuration, river connectivity, and localized human pressures.

The integration of Sentinel-2 imagery, CHIRPS rainfall data, and Random Forest modeling results in a robust remote-sensing–based framework for detecting and analyzing riverine plastic pollution. Model performance was consistently strong ($R^2 = 0.8892$), confirming the utility of multispectral indices and machine learning for mapping plastic waste at landscape scales. However, the pronounced intra-class variability especially in settlements and water bodies emphasizes the need for higher-resolution spatial data, improved hydrological characterization, and incorporation of socioeconomic factors in future research.

Overall, the findings underscore that effective mitigation strategies in Kendal Regency should prioritize land-based interventions in densely populated and industrially influenced areas, strengthen waste management systems, and protect or expand vegetated riparian buffers to intercept plastic transport. Since precipitation alone does not predict plastic accumulation, integrated land-use planning, targeted source reduction, and hydrologically informed monitoring are essential to reducing downstream plastic flows and safeguarding riverine and coastal ecosystems.

REFERENCES

Aji, A., Husna, V. N., & Purnama, S. M. (2024). Multi-Temporal Data for Land Use Change Analysis Using a Machine Learning Approach (Google Earth Engine). *International Journal of Geoinformatics*. https://doi.org/10.52939/ijg.v20i4.3145

Ballent, A., Corcoran, P., & Madden, O. (2016). Sources and sinks of microplastics in Canadian Lake Ontario nearshore, tributary and beach sediments. *Marine Pollution Bulletin*, *110*, 383–395. https://doi.org/10.1016/j.marpolbul.2016.06.037

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Biermann, L., Clewley, D., Martinez-Vicente, V., & Topouzelis, K. (2020). Finding Plastic Patches in Coastal Waters using Optical Satellite Data. *Scientific Reports*, *10*(1), 5364. https://doi.org/10.1038/s41598-020-62298-z

BPS Kendal. (2023). *Kabupaten Kendal dalam Angka 2023*. Badan Pusat Statistik Kabupaten Kendal.

Browne, M. A., Crump, P., Niven, S. J., & Teuten, E. (2011). Accumulation of microplastic on shorelines worldwide: Sources and sinks. Environmental Science & Technology. *Environmental Science & Technology*, *45*(21).

Cerra, D., Auer, S., Baissero, A., & Bachofer, F. (2025). Detection and Monitoring of Floating Plastic Debris on Inland Waters From Sentinel-2 Time Series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *18*, 1122–1138. https://doi.org/10.1109/JSTARS.2024.3502796

Chen, H., Deng, S., Zhang, S., & Shen, Y. (2024). Urban Growth and Its Ecological Effects in China. *Remote Sensing*, *16*(8), 1378. https://doi.org/10.3390/rs16081378

Corcoran, P., Biesinger, M., & Grifi, M. (2009). Plastics and beaches: A degrading relationship. *Marine Pollution Bulletin*, *58*(1), 80–84. https://doi.org/10.1016/j.marpolbul.2008.08.022

Cordova, M. R., Hadi, T. A., & Prayudha, B. (2018). *Occurrence And Abundance Of Microplastics In Coral Reef Sediment: A Case Study In Sekotong, Lombok-Indonesia*. https://doi.org/10.5281/ZENODO.1297719

Cortesi, I., Masiero, A., De Giglio, M., Tucci, G., & Dubbini, M. (2021). Random forest-based river plastic detection with a handheld multispectral camera. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIII-B1-2021*, 9–14. https://doi.org/10.5194/isprs-archives-XLIII-B1-2021-9-2021

Cortesi, I., Masiero, A., Tucci, G., & Topouzelis, K. (2022). UAV-based river plastic detection with a multispectral camera. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIII-B3-2022*, 855–861. https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-855-2022

Delwart, S. (2015). *ESA Standard Document*. *1*.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, *120*, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Eerkes-Medrano, D., Thompson, R. C., & Aldridge, D. C. (2015). Microplastics in freshwater systems: A review of the emerging threats, identification of knowledge gaps and prioritisation of research needs. *Water Research*, *75*, 63–82. https://doi.org/10.1016/j.watres.2015.02.012

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Scientific Data*, *2*(1), 150066. https://doi.org/10.1038/sdata.2015.66

Hanif, K., Suprijanto, J., & Pratikto, I. (2021). Identifikasi Mikroplastik di Muara Sungai Kendal, Kabupaten Kendal. *Journal of Marine Research*, *10*, 1–6. https://ejournal3.undip.ac.id/index.php/jmr

Hidayati, N., Rizqiani, S., & Sunandar, M. A. (2025). *The Role of Community Initiatives in the Implementation of Zero-Waste Policy in Kendal Regency: Case Study of KerDUS Community*.

Horton, A., Walton, A., Spurgeon, D., & Lahive, E. (n.d.). Microplastics in freshwater and terrestrial environments: Evaluating the current understanding to identify the knowledge gaps and future research priorities. *Science of The Total Environment*, *586*, 127–141. https://doi.org/10.1016/j.scitotenv.2017.01.190

Hurley, R., Woodward, J., & Rothwell, J. (2018). Microplastic contamination of river beds significantly reduced by catchment-wide flooding. *Nature Geoscience*, 251–257.

Jambeck, J. R., Geyer, R., & Wilcox, C. (2015). *Plastic waste inputs from land into the ocean*. *347*, 768–771.

Laksono, O. B., Suprijanto, J., & Ridlo, A. (2021). Kandungan Mikroplastik pada Sedimen di Perairan Bandengan Kabupaten Kendal. *Journal of Marine Research*, *10*(2), 158–164. https://doi.org/10.14710/jmr.v10i2.29032

Lebreton, L. C. M., Van Der Zwet, J., Damsteeg, J.-W., Slat, B., Andrady, A., & Reisser, J. (2017). River plastic emissions to the world's oceans. *Nature Communications*, *8*(1), 15611. https://doi.org/10.1038/ncomms15611

Louis, J., Debaecker, V., & Pflug, B. (2016). SENTINEL-2 SEN2COR: L2A PROCESSOR FOR USERS. *Living Planet Symposium*, *740*, 9–13. https://elib.dlr.de/107381/1/LPS2016_sm10_3louis.pdf

Nivedita, V., Begum, S. S., & Aldehim, G. (2024). Plastic debris detection along coastal waters using Sentinel-2 satellite data and machine learning techniques. *Marine Pollution Bulletin*, *209*. https://doi.org/10.1016/j.marpolbul.2024.117106

Paredes-Trejo, F., Alves Barbosa, H., Venkata Lakshmi Kumar, T., Kumar Thakur, M., & De Oliveira Buriti, C. (2021). Assessment of the CHIRPS-Based Satellite Precipitation Estimates. In A. Devlin, J. Pan, & M. Manjur Shah (Eds.), *Inland Waters—Dynamics and Ecology*. IntechOpen. https://doi.org/10.5772/intechopen.91472

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sensing*, *12*(14), 2291. https://doi.org/10.3390/rs12142291

Rayamajhi, D., Bhattarai, K., Giri, K., Budhathoki, M., Karn, N. K., Subedi, O., Regmi, R. K., & Dahal, V. (2025). Assessing flood susceptibility in a Triyuga watershed, Nepal using statistical models. *Scientific Reports*, *15*(1), 32056. https://doi.org/10.1038/s41598-025-10610-0

Rußwurm, M., Venkatesa, S. J., & Tuia, D. (2023). Large-scale detection of marine debris in coastal areas with Sentinel-2. *iScience*, *26*(12), 108402. https://doi.org/10.1016/j.isci.2023.108402

SNI (Standar Nasional Indonesia). (2010). *Klasifikasi penutup lahan (SNI 7645:2010)*. Badan Standardisasi Nasional (BSN).

Steinmetz, Z., Wollmann, C., Schaefer, M., & Buchmann, C. (2016). Plastic mulching in agriculture. Trading short-term agronomic benefits for long-term soil degradation? *Science of The Total Environment*, *550*, 690–705. https://doi.org/10.1016/j.scitotenv.2016.01.153

Tikuye, B. G., Ray, R. L., Manjunatha, B., Tefera, G. W., & Gurau, S. (2025). Drought monitoring using the Climate Hazards InfraRed Precipitation with Stations (CHIRPS) in Ethiopia. *Natural Hazards Research*, *5*(2), 348–362. https://doi.org/10.1016/j.nhres.2024.12.002

Van Emmerik, T., Kieu-Le, T.-C., Loozen, M., Van Oeveren, K., Strady, E., Bui, X.-T., Egger, M., Gasperi, J., Lebreton, L., Nguyen, P.-D., Schwarz, A., Slat, B., & Tassin, B. (2018). A Methodology to Characterize Riverine Macroplastic Emission Into the Ocean. *Frontiers in Marine Science*, *5*, 372. https://doi.org/10.3389/fmars.2018.00372

Van Emmerik, T., & Schwarz, A. (2020). Plastic debris in rivers. *WIREs Water*, *7*(1), e1398. https://doi.org/10.1002/wat2.1398

Windsor, F. M., Tilley, R. M., Tyler, C. R., & Ormerod, S. J. (2019). Microplastic ingestion by riverine macroinvertebrates. *Science of The Total Environment*, *646*, 68–74. https://doi.org/10.1016/j.scitotenv.2018.07.271

World Bank. (2021). Plastic waste discharges from rivers and coastlines in Indonesia. World Bank Group. *World Bank Group*. https://www.worldbank.org/en/country/indonesia/publication/plastic-waste-discharges-from-rivers-and-coastlines-in-indonesia

Zahrah, Y., Yu, J., & Liu, X. (2024). How Indonesia's Cities Are Grappling with Plastic Waste: An Integrated Approach towards Sustainable Plastic Waste Management. *Sustainability*, *16*(10), 3921. https://doi.org/10.3390/su16103921

PERFORMANCE OF MACHINE LEARNING AND DEEP LEARNING MODELS FOR PREDICTING RAINFALL IN A LARGE WATERSHED: CASE STUDY OF BENGAWAN SOLO RIVER BASIN, INDONESIA

Jumadi JUMADI [1*,2], Kuswaji Dwi PRIYONO[3], Ali Hasan ABDULLAH[3], Supari SUPARI[4], Hamza AIT ZAMZAMI[5], Farha SATTAR[6], Muhammad NAWAZ[7], Hamzah HASYIM[8] & Steve CARVER[9]

DOI: 10.21163/GT_2026.212.05

ABSTRACT

Predicting rainfall in large, heterogeneous watersheds remains among the most important hydrological challenges. This research investigates the effectiveness of both ML (Machine Learning) and DL (Deep Learning) for predicting spatiotemporal rainfall in the Bengawan Solo watershed, Indonesia. Satellite rainfall data from CHIRPS (spatial resolution: 0.05°) were prepared and sampled for the period 1981–2024. The data set contained 523 grid points. We employed nine ML and DL algorithms: Random Forest (RF), Extreme Gradient Boosting (XGB), Support Vector Regression (SVR), Multilayer Perceptron (MLP), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Temporal Convolutional Network (TCN), Convolutional Neural Network (CNN), and Transformer. Models were trained on the samples from 1981 to 2019 and tested on 2020–2024. Performance was judged from mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2). XGB showed the best overall performance (MAE ≈59 mm; R^2 ≈0.73). GRU became the most competitive DL model at 60 mm (MAE ≈60 mm; R^2 ≈0.72). Temporal model analysis shows that XGB and GRU stay among the top three models with the minimum monthly errors. TCN, CNN, and Transformer exhibited higher errors and more monthly variability. XGB and GRU have average MAE values of ~59–60 mm and R^2 values of ~0.71–0.72 across most grids. MAE values for TCN, CNN, and Transformer are greater than 76 mm, while R^2 values are lower. The data we obtained indicate that using ensemble decision tree models and recurrent neural networks across large tropical areas yields greater stability and more reliable spatiotemporal rainfall predictions than more sophisticated DL architectures.

Keywords: *Rainfall prediction; Machine learning; Deep learning; CHIRPS; Bengawan Solo; Spatial analysis; Temporal analysis.*

1. INTRODUCTION

Rainfall prediction is essential for effectively planning water resources, managing disasters, and managing agricultural areas in the tropics (Hong et al., 2018; Praveen et al., 2020). The Bengawan Solo watershed in Indonesia is an area critical to food security, water supply, and flood control (Musiyam et al., 2025). The watershed is a large, socio-economically critical watershed where rainfall

[1*]*Faculty of Geography, Muhammadiyah University Surakarta, Indonesia.*
Corresponding author: jumadi@ums.ac.id (JJ)
[2]*INTI International University, Malaysia*
[3]*Faculty of Geography, Muhammadiyah University Surakarta, Indonesia; kuswaji.priyono@ums.ac.id (KDP), e100210132@student.ums.ac.id (AHA)*
[4]*Agency for Meteorology, Climatology and Geophysics (BMKG), Indonesia; supari@bmkg.go.id (SS)*
[5]*University of Hassan II Casablanca, Casablanca, Morocco; hamza.aitzamzami-etu@etu.univh2c.ma (HAZ)*
[6]*Charles Darwin University, Australia; Farha.Sattar@cdu.edu.au (FS)*
[7]*National University of Singapore, Singapore; geomn@nus.edu.sg (MH)*
[8]*Universitas Sriwijaya, Ogan Ilir, South Sumatera, Indonesia; hamzah@fkm.unsri.ac.id (HH)*
[9]*School of Geography, University of Leeds, United Kingdom; S.J.Carver@leeds.ac.uk (SC)*

variability directly affects water availability for irrigation and is closely linked to recurrent hydrometeorological impacts on agriculture and a large population. The basin also exhibits pronounced physical heterogeneity, ranging from flat plains to hilly and volcanic terrain, with diverse soils and land uses, producing strong spatial contrasts in hydrological response and making rainfall prediction more challenging at scale (Santhyami et al., 2025). Thus, long-term rainfall data and rainfall predictors are crucial. Nonetheless, the effects of high climatic variability, topographical complexity, and limited field observation data are major obstacles to the long-term accuracy of rainfall prediction (Funk et al., 2015; Kundu et al., 2022).

Data-driven methods have revolutionized hydrometeorological modeling over time. Machine Learning (ML) methods such as Random Forest (RF), Extreme Gradient Boosting (XGB), Support Vector Regression (SVR), and Multilayer Perceptron (MLP) to examine in depth the associations between predictive variables and rainfall predictors (Sachindra et al., 2018; Miao et al., 2021) are used more frequently to discover nonlinear relationships between predictors and rainfall. Concurrently, deep learning (DL) approaches like Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), Temporal Convolutional Network (TCN), and Transformer architectures depict complex spatio-temporal patterns in climate time series (Shi et al., 2015; Chattopadhyay et al., 2020; Espeholt et al., 2022).

Many studies have shown that DL techniques have great potential for accurately modeling hydrometeorological phenomena (Aswin & Geetha, 2020; Hernández et al., 2021), including the prediction of severe and rare rainfall events (Laptev et al., 2017; Lim & Zohren, 2021). But most of that performance depends on the vastness of the data in terms of temporal depth, with a fairly uniform spatial distribution. In data-limited regions, DL models are often weak under time-series conditions due to overfitting, poor training, and generalization issues (Willard et al., 2021; Das et al., 2022). Although ML and DL techniques have made significant progress in predicting rainfall, quite a few research gaps still exist. Comparative studies assessing the performance of ML and DL in tropical regions, such as the Bengawan Solo River Basin, with complex spatial and even topographical features, are limited, especially when using high-resolution satellite data (such as CHIRPS) to inform annual rainfall forecasting.

Furthermore, it is rare for model training to rely on long time series, spanning 40 years or more, despite the fact that long time spans are necessary to account for long-term climate dynamics and interannual variability that impact prediction accuracy (Gu et al., 2020). Furthermore, spatially and temporally differentiated error analysis is underused, even though it is important to identify specific locations and time intervals of model inaccuracy to better prepare for the next model (Cioffi et al., 2023). To fill in the gap, this study conducted a comprehensive evaluation of nine ML and DL models using monthly CHIRPS data for 1981-2024, aggregated to annual rainfall at 98 sample points in the Bengawan Solo watershed. Models were trained on historical data and tested on 2020–2024 conditions, with metrics including MAE, RMSE, MAPE, R^2, spatial errors, and temporal errors. The best-performing model was then adopted as the basis for predicting annual rainfall for 2025–2030. In conclusion, this study assesses the predictive performance of ML and DL in the Bengawan Solo watershed region, using the long-term CHIRPS (1981–2024) and 2025–2030 periods to estimate annual rainfall and to determine which method is best suited to tropical regions with scarce data. The literature shows that when remote sensing series are discontinuous, statistical interpolation approaches can reconstruct long-term trajectories useful for analysis (Haidu et al., 2024).

2. METHOD

2.1. Study Area

The Bengawan Solo River Basin (**Fig. 1**) is the largest river system on Java Island, covering an area of approximately 16,100 km² and serving as the primary source of water for domestic needs and agricultural irrigation. Hydrologically, the watershed is divided into three sub-basins: the Upper Bengawan Solo, the Madiun River, and the Lower Bengawan Solo.

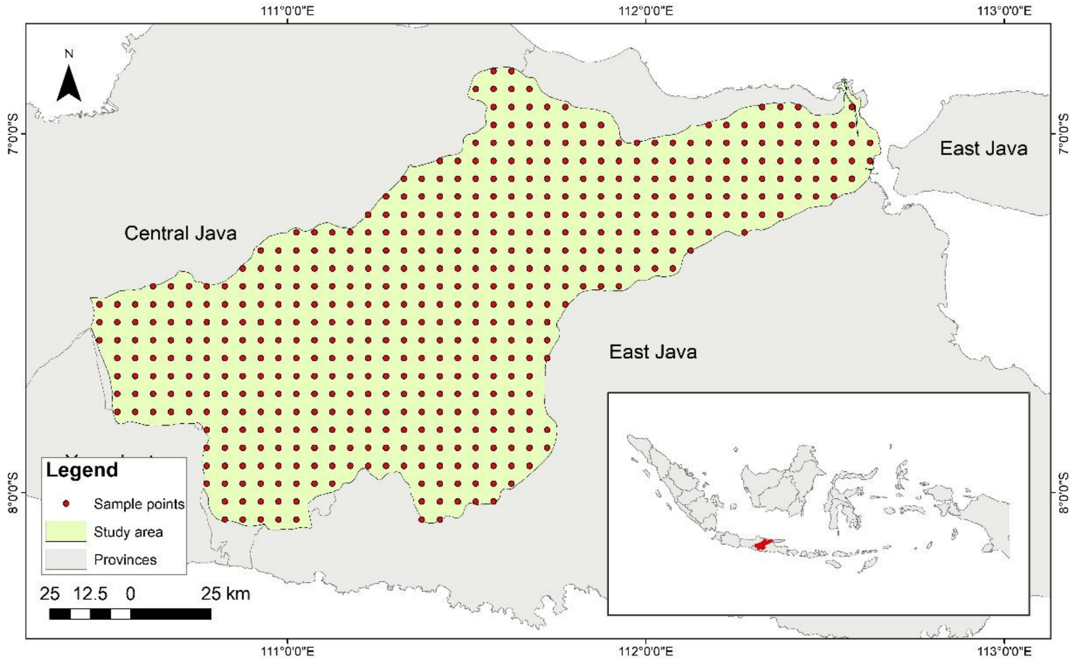


Fig. 1. Study Area and Sample Points.

Its upstream area (approximately 6,000 km²) lies between 110º13'7.16"–110º26'57.10" East Longitude and 7º26'33.15"–8º6'13.81" South Latitude, with predominantly flat topography but undulating in the northeast–northwest near the mountains; flow supply comes from the Merapi–Merbabu volcanic complex in the west and Mount Lawu in the east. The contrasting topographic variations and land use, as well as the occurrence of seasonal floods and droughts, make the Bengawan Solo River Basin a priority location for hydrometeorological studies and water resources management planning (Jumadi et al, 2024; Jumadi et al, 2025).

2.2. Research Framework

This research was conducted in several stages, including data collection, preprocessing, annual rainfall prediction using ML and DL, model evaluation (using data from 2020-2024), selection of the best model, prediction for 2025-2030, data aggregation, interpolation, and spatiotemporal analysis (**Fig. 2**).

2.3. Data and Data Sources

Data aggregation, interpolation, and spatiotemporal analysis were performed in several stages. The study took advantage of satellite rainfall data collected from the Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) product, issued by the Climate Hazards Center at the University of California, Santa Barbara (Funk et al., 2015). CHIRPS is a global precipitation dataset for rainfall estimation with high spatial resolution (~0.05°, or ±5 km) and long time span (1981 to present). This is a global dataset of infrared satellite observation data, reanalysis data, and rainfall data from land stations. This hybrid allows CHIRPS to generate consistent rainfall estimations, even in parts of the world where ground observation networks are weak such as Indonesia (Ceccherini et al., 2015; Gu et al., 2020). In hydrology, the comparative evaluation of monthly gap imputation methods has proven essential for the stability of analyses (cf. monthly flow study), which supports our focus on the continuity of series (Magyari-Sáska et al., 2025).

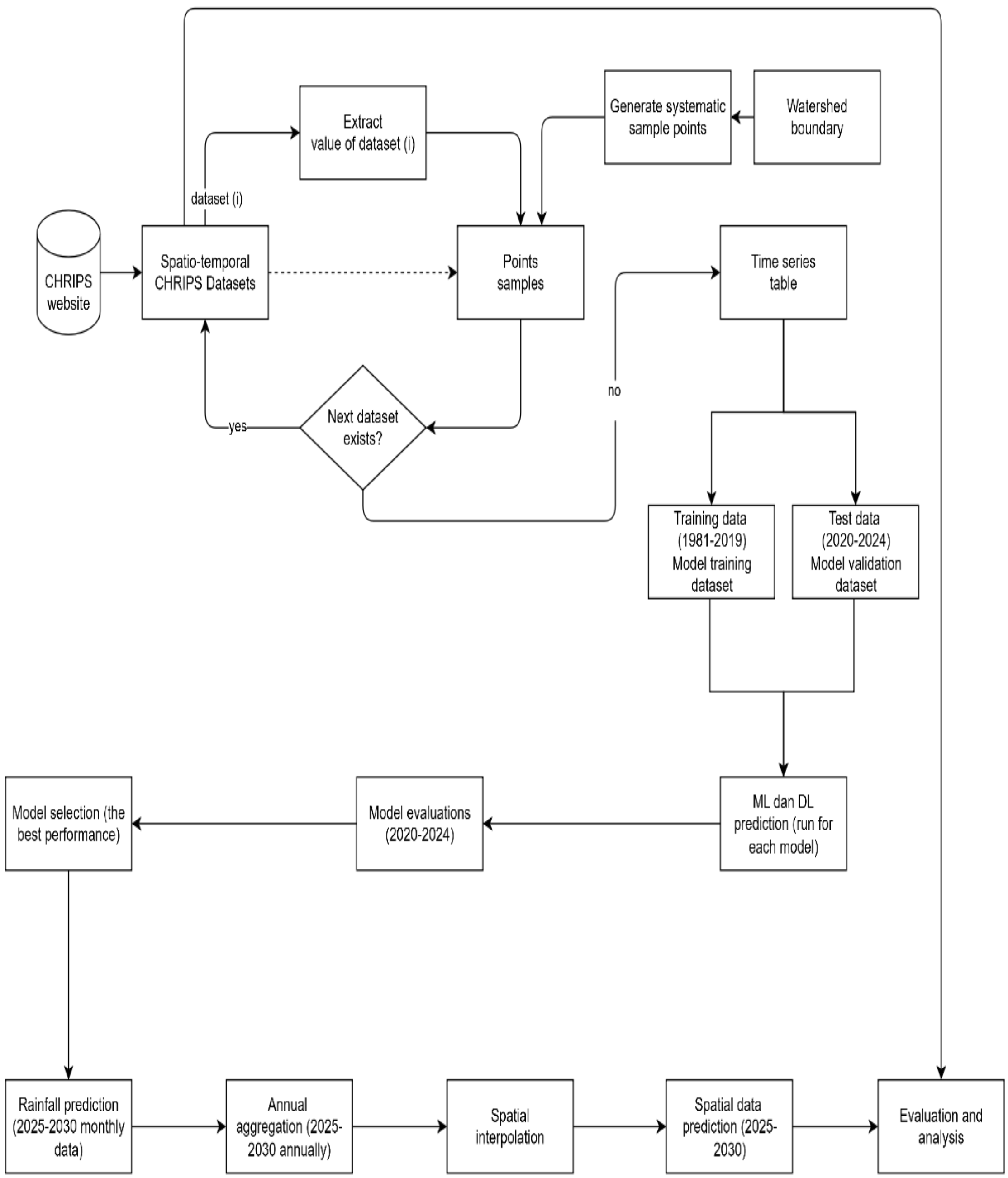


Fig. 2. Research Framework.

The selection of CHIRPS for this study involved three elements. First, its high spatial resolution allows for a better representation of rainfall variability in areas with complex topography such as the Bengawan Solo River Basin. Second, its time series spanning more than four decades (1981-2024) allows a better understanding of long-term trends and interannual climate variability. Third, the availability of open-access data that is routinely updated enables replication of models as well as future updates, which is the key principle of data openness in scientific research (Huffman et al., 2020).

2.4. Determination of Prediction Points

All pixels from CHIRPS were extracted in order to allow for the model prediction to capture spatial variety of rainfall throughout the Bengawan Solo watershed (**Fig. 1**). The choice of this method was motivated by its ability to provide an even distribution of points across the study area, eliminating spatial bias when observation points cluster around specific positions (Hijmans et al., 2005). The result was 523 points, evenly distributed over the basin. Each point's geographic coordinates were recorded, and a monthly rainfall time series spanning 44 years (1981–2024) was used. This technique is also helpful for error analysis in spatial contexts and is central to this study. An analysis of accuracy disparities, including from edge vs. center of the watershed (or lowlands vs. mountains) can be made under equal spreading of points.

2.5. Development and Configuration of Machine Learning and Deep Learning Models

The development of an annual rainfall prediction model in this study involves two main approaches, namely Machine Learning (ML) and Deep Learning (DL). The selection of these two groups of methods is based on the consideration that ML has advantages in handling medium-sized datasets with relatively limited feature space, while DL is designed to extract complex patterns from extensive, temporal, or spatial data (Goodfellow et al., 2016; Zhang et al., 2022). The hyperparameterization of the model is presented in **Table 1**.

2.5.1. Machine Learning Model

Four ML algorithms, such as Random Forest (RF), Extreme Gradient Boosting (XGB), Support Vector Regression (SVR), and Multi-Layer Perceptron (MLP), are applied. RF was selected because it can accommodate non-linear relationships and reduce the overfitting of medium-sized datasets (Breiman, 2001). Some important parameters, such as the number of trees (n_estimators) and the maximum tree depth (max_depth), were optimized using grid search. A gradient boosting method called XGB was selected for its ease of computation and ability to handle noisy data (Chen & Guestrin, 2016).

The learning_rate, max_depth, and number of estimators were fine-tuned for the validation data to minimize the error. SVR was employed for testing the performance of kernel-based approaches, which map data to higher-dimensional spaces to model nonlinear relationships (Smola & Schölkopf, 2004). The radial basis function (RBF) kernel was used, and the parameters C and gamma were determined via parameter search. For the ML group, we utilized MLP as a shallow neural network-based baseline. The MLP architecture had multiple hidden layers, combined with ReLU activation functions and a linear output layer for regression.

2.5.2. Deep Learning Model

Time and space structures in the data are processed over five different DL models: Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Temporal Convolutional Network (TCN), Convolutional Neural Network (CNN), and Transformer. To make use of LSTM, this model features two sequential LSTM layers with dropout to reduce overfitting. GRU is a simpler yet computationally efficient variant of LSTM (Cho et al., 2014).

The number of hidden units and dropout rate parameters were optimized based on validation performance. TCN was proposed as a non-recurrent alternative to time-series modeling with dilated causal convolutions, facilitating long-term dependency modeling (Bai et al., 2018). A CNN was applied to extract local temporal features from monthly rainfall data, using a 1D convolutional layer, pooling, and dense layers. Transformer is applicable to investigate the capability in fitting temporal dependencies without sequential restriction induced by the self-attention mechanism (Vaswani et al., 2017). The design involved a sequence of encoder layers combining multi-head attention and normalization.

Table 1.

ML and DL Hyperparameters.

Model	Type	Input window	Key hyperparameters / architecture
RF	ML	24 months (24 rainfall lags)	RandomForestRegressor(n_estimators=100, random_state=42); other parameters use scikit-learn defaults (e.g., max_depth=None, min_samples_split=2, min_samples_leaf=1, max_features="sqrt", bootstrap=True).
XGB	ML	24 months	XGBRegressor(n_estimators=100, random_state=42); other parameters follow XGBoost defaults (e.g., typical defaults such as max_depth=6, learning_rate=0.3, subsample=1.0, colsample_bytree=1.0, objective="reg:squarederror" depending on library version).
SVR	ML	24 months	SVR() with scikit-learn defaults: kernel="rbf", C=1.0, epsilon=0.1, gamma="scale", degree=3.
MLP	ML	24 months	MLPRegressor(hidden_layer_sizes=(64, 32), max_iter=500, random_state=42); other parameters follow defaults: activation="relu", solver="adam", learning_rate_init=0.001, alpha=0.0001, etc.
LSTM	DL	24 months (24 × 1)	Input(shape=(24, 1)) → LSTM(64, activation="relu") → Dense(1, activation="linear"); compiled with optimizer="adam", loss="mse"; trained for epochs=60, batch_size=32.
GRU	DL	24 months	Input(shape=(24, 1)) → GRU(64, activation="relu") → Dense(1, activation="relu"); compiled with optimizer="adam", loss="mse"; trained for epochs=60, batch_size=32.
TCN	DL	24 months	Input(shape=(24, 1)) → TCN(nb_filters=32, kernel_size=3, dilations=[1, 2, 4, 8]) → Dense(1, activation="linear"); compiled with optimizer="adam", loss="mse"; trained for epochs=60, batch_size=32.
CNN	DL	24 months	Input(shape=(24, 1)) → Conv1D(64, kernel_size=3, activation="relu", padding="causal") → BatchNormalization() → Conv1D(32, kernel_size=3, activation="relu", padding="causal") → BatchNormalization() → GlobalAveragePooling1D() → Dense(1, activation="linear"); compiled with optimizer="adam", loss="mse"; trained for epochs=60, batch_size=32; predictions clipped to non-negative values during recursive forecasting.
Transformer	DL	24 months	Input(shape=(24, 1)) → transformer_encoder(head_size=8, num_heads=2, ff_dim=32, dropout=0.1) → GlobalAveragePooling1D() → Dropout(0.1) → Dense(1, activation="linear"); compiled with optimizer="adam", loss="mse"; trained for epochs=60, batch_size=32.

2.5.3. Implementation

All models were developed using Python with scikit-learn, TensorFlow, pandas, numpy, and several other libraries. Model training and prediction were performed on Google Colaboratory Pro using high-RAM cloud devices to speed up computation (https://tinyurl.com/y23kfsyh). Each model generates monthly rainfall predictions at 98 sample points, which are then aggregated into annual totals for performance evaluation and spatio-temporal analysis of prediction errors.

2.6. Model Training, Validation, and Evaluation Scheme

The model's training and validation methods were developed specifically to ensure that the predicted findings were statistically sound and robust to annual climate variability in the Bengawan Solo River Basin. The dataset was split into two time subsets: the training period (1981–2019) and the validation period (2020–2024). This separation was done chronologically (time-based split) to avoid data leakage, as future data would otherwise be used for model training. This is frequently used to make time series predictions, as seasonal patterns and climate trends may evolve (Hyndman & Athanasopoulos, 2018). For training purposes, all models were trained on 523 normalized points from monthly rainfall data.

We configured the data in a temporal order (24 months) for deep learning-based models that could capture long-term dependencies. The performance of our model is evaluated with four overall criteria: Mean Absolute Error (MAE) (Equation 1), Root Mean Squared Error (RMSE) (Equation 2), Mean Absolute Percentage Error (MAPE) (Equation 3), and coefficient of determination (R^2) (Equation 4). Such composite metrics are chosen to provide a complete picture of the model: an absolute accuracy rate, the reliability of these measures, and the proportionality of all errors. In general, these metrics measure the difference between the predicted value ($\hat{y}_i$) and the observed value (y_i).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad \text{......(1)}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad \text{......(2)}$$

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad \text{......(3)}$$

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad \text{......(4)}$$

The evaluation results were interpreted by applying the average metric values across all sample points to identify the model with the best overall performance. The model with the best performance during the validation period was selected and used to produce annual rainfall projections for 2025–2030.

2.7. Creation of Spatial Maps of Rainfall Projections with IDW

The final stage of this research methodology is to translate rainfall forecasting results for the year into spatial representations, such as maps. This spatial visualization is critical for understanding the distribution of rainfall across geographical intervals, which is not available from prediction value tables alone. To accomplish this, the Inverse Distance Weighting (IDW) interpolation technique was selected, as it is simple, flexible for handling irregular data, and works well for environmental variables with pronounced spatial autocorrelation, such as rainfall (Lu & Wong, 2008; Phoophathong et al., 2025).

The interpolation process was carried out separately for each projection year (2025–2030) and for each prediction model. For each projection year, annual rainfall at each sample point was first obtained from the monthly predictions. Aggregation is performed by summing the monthly rainfall predictions (in millimeters) at each sample point for the period from January to December of each year, specifically for the projection years 2025 to 2030. The IDW is implemented in ArcGIS 10.8, using the geographic coordinates of 523 sample points and the corresponding yearly prediction values as inputs. To address this, we use the WGS 84 projection system; it helps preserve compatibility with CHIRPS data while preventing distortion due to distance. The output grid resolution was set to 0.01° (~1 km) to achieve sufficient spatial detail without overloading the computation. The IDW map outputs were then spatially analyzed to identify areas with significant rainfall increase and decrease trends during the projection period.

3. RESULTS

3.1. Model Performance Comparison

A set of nine annual rainfall prediction models was evaluated for their performance for the 2020–2024 validation period based on four main metrics, namely Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and coefficient of determination (R^2) (**Fig. 3**). These results indicate that XGBoost (XGB) performed the best at all metrics. The MAE and RMSE of 59 mm and 80 smm, respectively, indicate low absolute and quadratic errors, and the model has an R^2 of 0.73, indicating excellent ability to explain the variability in annual rainfall.

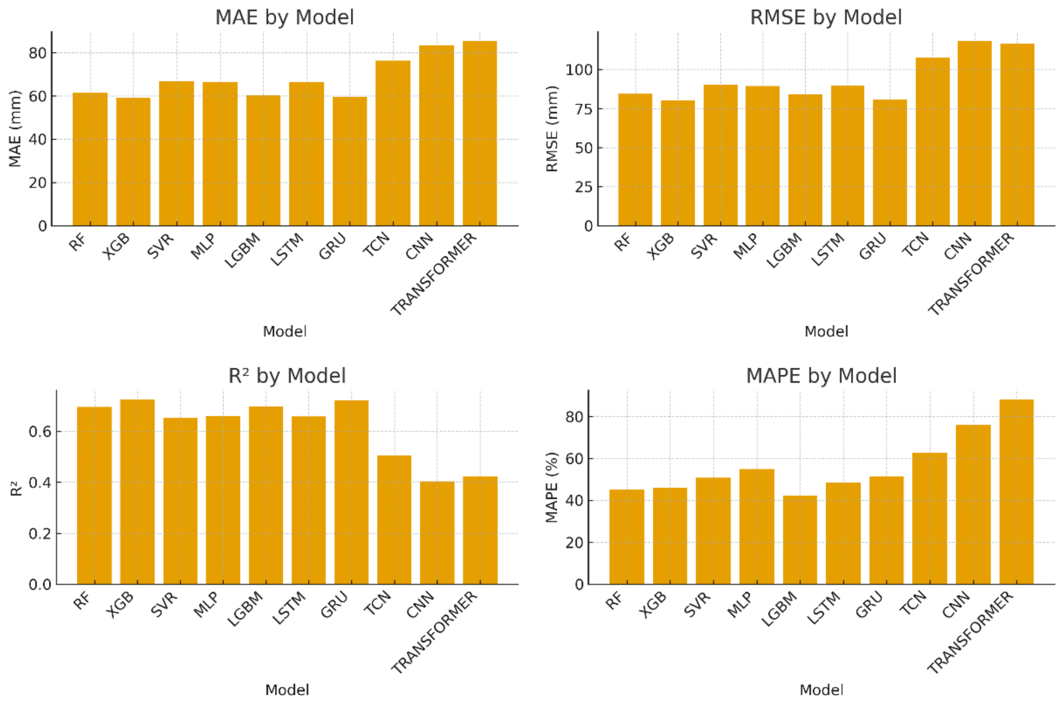


Fig. 3. Measurement error results for all models.

Of the two models, Extreme Gradient Boosting (XGB) ranked 2nd, with an MAE of 55.140 mm and an R^2 of 0.767. XGB handled data non-linearity well but performed slightly worse than RF. XGB is sensitive to the learning rate and the number of trees, which is likely to cause instability in predictions for long-term datasets where year-to-year data exhibits high variation. We have seen that the Multi-Layer Perceptron (MLP), with an MAE of 46.762 mm and an R^2 of 0.817, performed relatively well, although lower than RF. The lower strength of MLPs compared to ensemble-based models like RF is attributed to the poor ability of feedforward ANNs to capture long-term temporal patterns without a direct memory mechanism (Wu et al., 2010). Although performance was surprising, with sequence-based models (e.g., LSTM, GRU, TCN, CNN, and Transformer) achieving very high MAE (>88 mm, up to 202 mm for LSTM, TCN, CNN, and Transformer), some architectures also yielded negative R^2 values.

Such excessive overfitting is likely an effect of the time series being too short, even though it spans 44 years, for detecting stable annual climate patterns in complex DL architectures with a large number of parameters (Lim & Zohren, 2021). This condition aligns with the results of Laptev et al. (2017), who note that DL forecasting models for meteorological time series require hundreds of thousands of observations or hundreds of years of simulation data to generalize consistently. These findings support the idea that ensemble-based machine learning models, such as RF, remain more stable in tropical environments with limited data. These results serve not only as a technical framework for hydrology researchers and practitioners but also contribute to the literature by showing the practical limitations of DL in low-data areas and by pointing to opportunities for future study to develop lighter DL architectures and more realistic regularization.

3.2. Spatial Error Analysis

Spatial error analysis was performed to understand the distribution of prediction errors at 523 sample points in the Bengawan Solo watershed during the 2020–2024 validation period (**Fig. 4–7**).

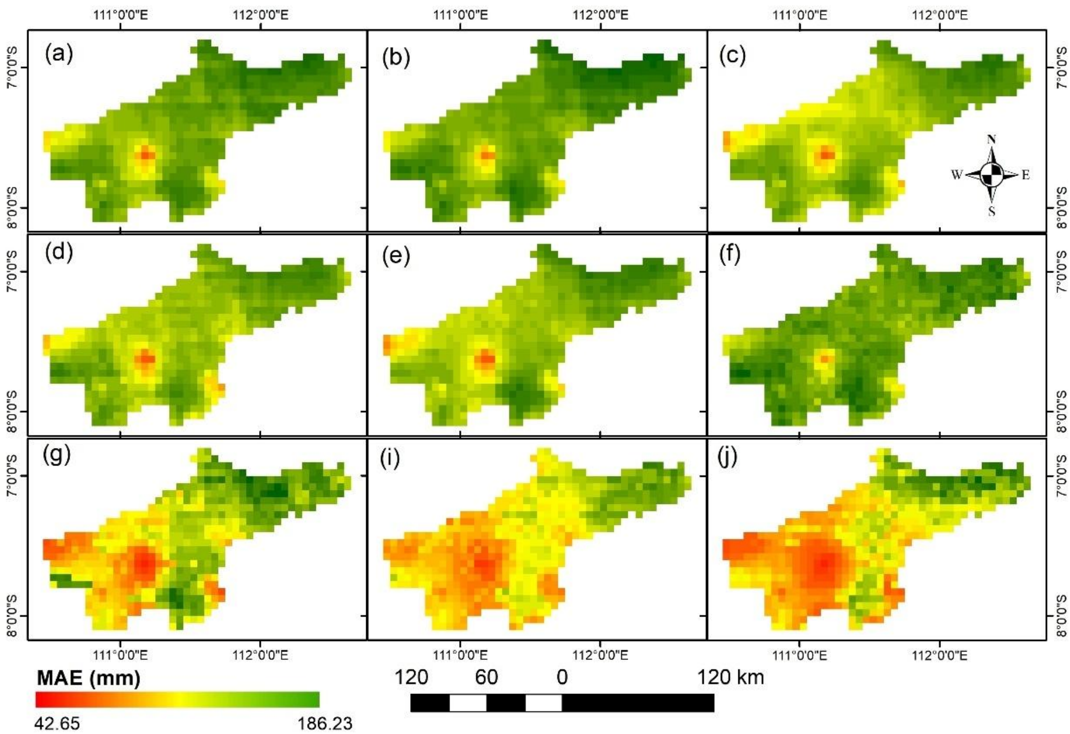


Fig. 4. Spatial MAE Distribution (mm). (a) RF, (b) XGB, (c) SVR, (d) MLP, I LSTM, (f) GRU, (g) TCN, (h) CNN, (i) TRANSFORMER.

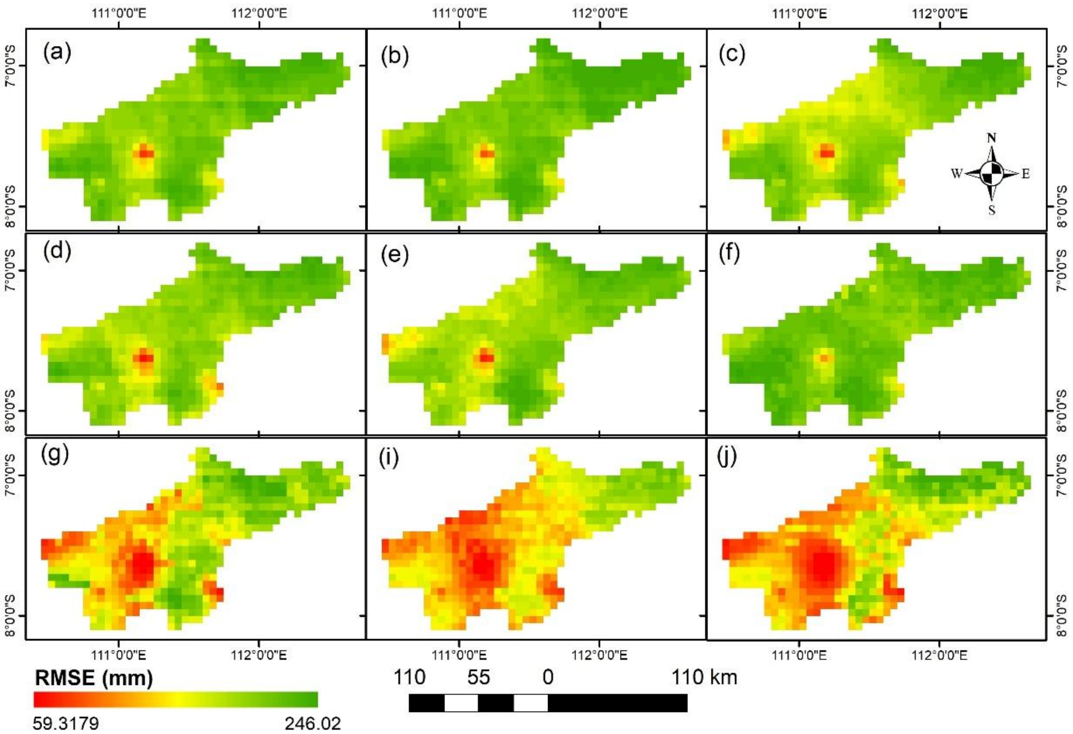


Fig. 5. Spatial RMSE Distribution (mm). (a) RF, (b) XGB, (c) SVR, (d) MLP, I LSTM, (f) GRU, (g) TCN, (h) CNN, (i) TRANSFORMER.

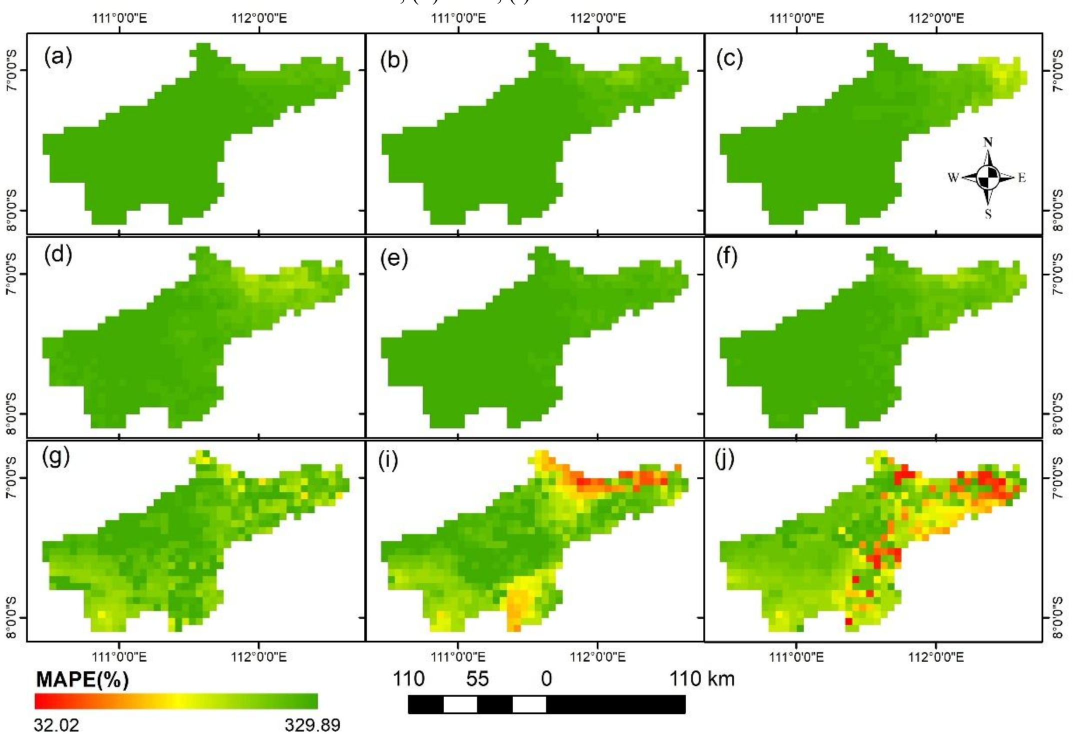


Fig. 6. Spatial MAPE Distribution (%). (a) RF, (b) XGB, (c) SVR, (d) MLP, I LSTM, (f) GRU, (g) TCN, (h) CNN, (i) TRANSFORMER.

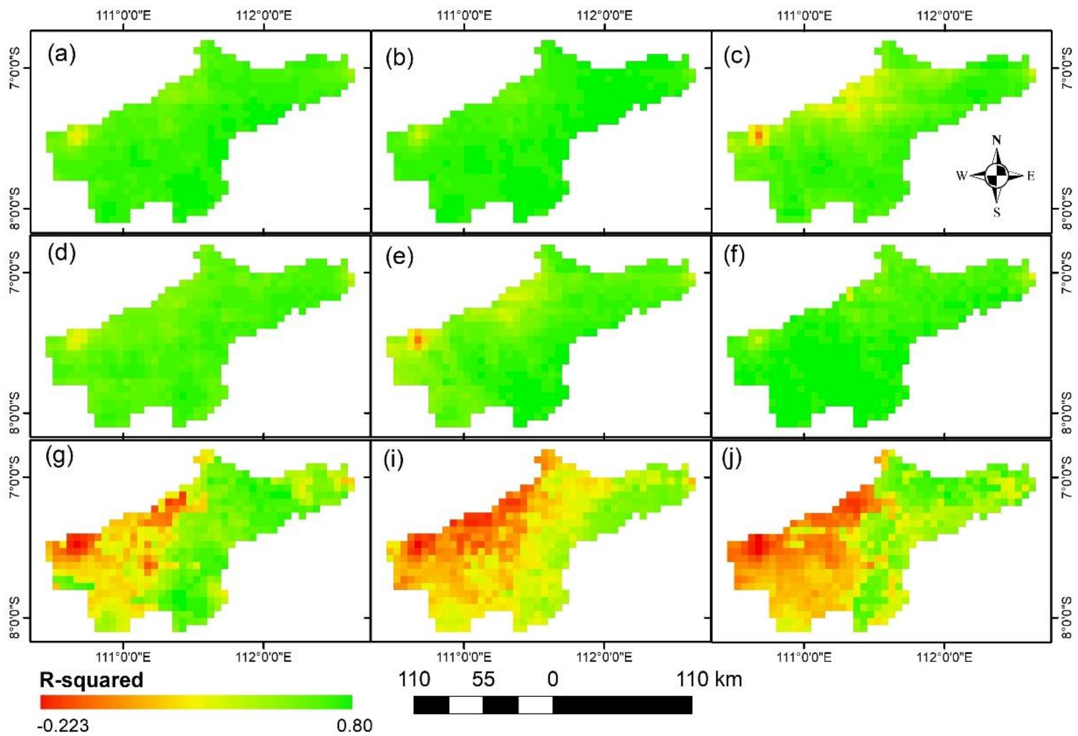


Fig. 7. Spatial R2 Distribution. (a) RF, (b) XGB, (c) SVR, (d) MLP, I LSTM, (f) GRU, (g) TCN, (h) CNN, (i) TRANSFORMERs.

This is critical, given that although aggregated metrics (e.g., MAE, RMSE) provide a baseline picture of model performance, the distribution of errors at spatial scales often reveals underperformance not evident at the aggregate level (Cioffi et al., 2023).

Values of the MAE for points indicate that the error distribution is not homogeneous. Despite the fact that the Random Forest (RF) model generally achieves the best results, it still exhibits localized spatial error distribution, especially at key points in the watershed in the east and west. The points are usually in the parts of the area that are more diverse in topography, e.g., in the foothills of Mount Lawu to the east and hills in the northwest. These topographic differences probably impact the precision of CHIRPS satellite rainfall predictions and the precision of model predictions, respectively. On the other hand, in the central plains of the watershed, where the topography is similar and near the main Bengawan Solo River, there are lower prediction errors.

This suggests that the model is more adept at learning the relationship between the historical rainfall data and prediction patterns in regions free from extreme topographic gradients. Other, e.g., those models used, such as XGB and MLP, exhibit error patterns as in the RF, but they generate more error intensity in almost all locations. With the SVR model, the error distribution is scattered and not so well matched to the topography pattern, which reflects its sensitivity to the high variation of training data. Moreover, deep learning-based models, namely LSTM, GRU, TCN, CNN, and Transformer, have large average errors and spatial instability. There exist very large errors in points in mountain regions and in transition zones between mountains and plains that exceed the watershed mean MAE by more than 3 times. This confirms the overfitting assumption in DL, which occurs when the model cannot learn spatial variability effectively.

The data also supports this observation with an increased error gradient on the MAE distribution map for IDW interpolation results towards east and west of the watershed boundaries. This fact appears to be in line with earlier works (Funk et al., 2015; Gu et al., 2020), which found that satellite-based rainfall prediction is more prone to vary when the orographic environment and the number of rainfall stations in the location are not very high.

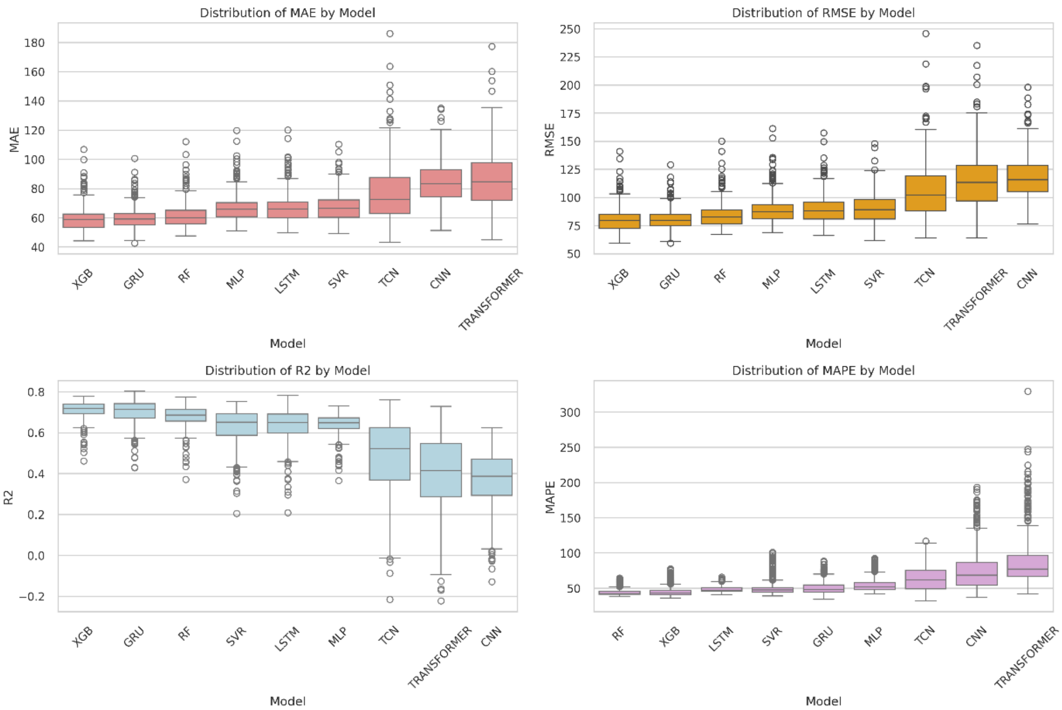


Fig. 8. Boxplot Comparison of Error Metrics.

This spatial study has a practical dimension, as it implies that, even though RF is technically well-known for its worldwide performance, it could still be improved by integrating CHIRPS data with higher-resolution data or by pre-conditioning the model to mitigate topographical bias. In addition, model evaluation can always be conducted to consider the spatial, not overall, aspect, in order to detect and target prediction failures in some regions for future study.

The spatial error comparison boxplot (**Fig. 8**) reveals the marked disparity between models in accounting for the variability in precipitation at 523 grid points in the Bengawan Solo watershed. On average, XGB and GRU have the lowest median MAE and RMSE, and their interquartile ranges (IQRs) are rather narrow, which implies good and constant (or comparable) performance in most locations. RF marginally is behind (median MAE slightly higher; upper tail slightly longer), but it also outperforms SVR and MLP by a long margin of difference because both have larger errors and wider distributions. On the other hand, convolutional and attention-based deep learning models (TCN, CNN, Transformer) seem obviously less robust: median MAE and RMSE are high, interquartile ranges are wide, and the large number of extreme outliers suggests spatial instability and no ability to follow rainfall trends for some grids. The boxplot pattern only confirms that the decision tree ensembles and GRUs are spatially superior to the more complex deep learning architectures used with the dataset.

3.3. Temporal Error Analysis

Temporal error analysis was performed to evaluate variation in model performance over time during the 2020–2024 validation period (**Fig. 9-12**). This approach aims to identify specific periods during which the model degrades in accuracy, thereby providing deeper insights into seasonal factors and extreme climate events that affect predictions (Zhou et al., 2022).

Results from temporal evaluation show that the XGBoost (XGB) model, which has good overall performance, tends to have more prediction errors during transitional months (March–April and October–November, for example).

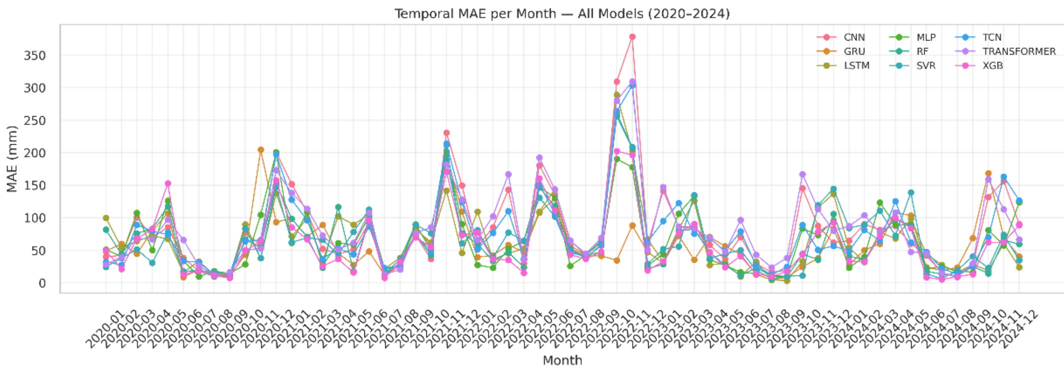


Fig. 9. Temporal Metric of MAE.

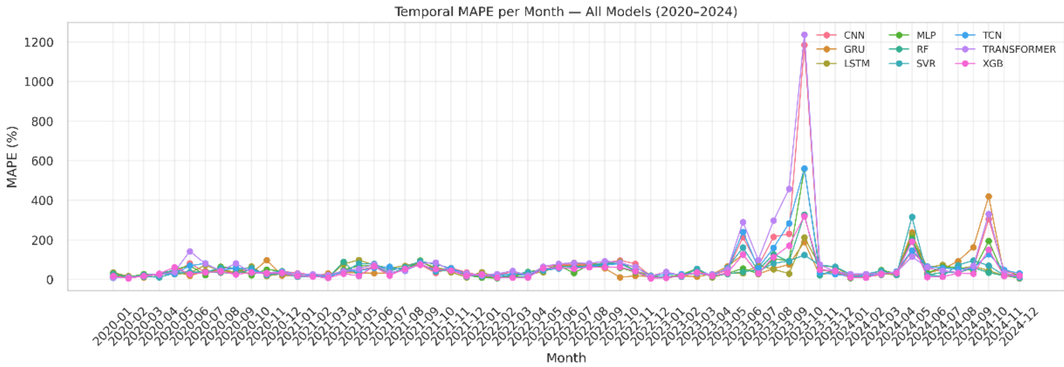


Fig. 10. Temporal Metric of MAPE.

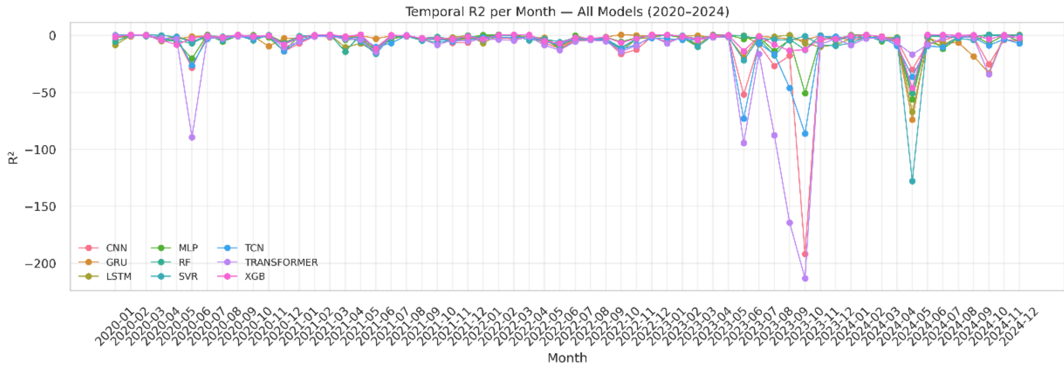


Fig. 11. Temporal Metric of R^2.

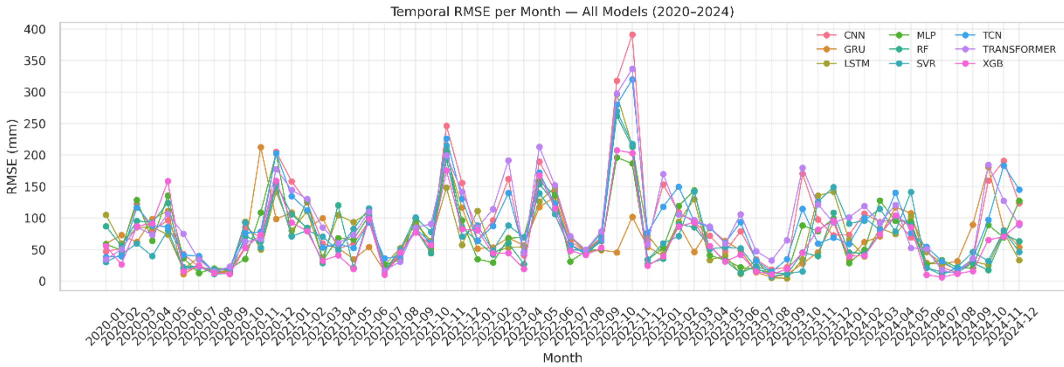


Fig. 12. Temporal Metric of RMSE.

At these junctures, monsoon variations in predominant wind directions and erratic weather patterns also occur in areas of tropical Indonesia such as the Bengawan Solo watershed. Because these patterns are not completely periodic, the abrupt changes in rainfall intensity and location can be hard to capture in models based on historical data due to the fact that these patterns are highly influenced by global climate variability such as El Niño–Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) (McBride et al., 2003; Lestari et al., 2021).

Major mistakes are also found in years characterized by extreme climate anomalies. That is, for example, from October to December 2020 (strong La Niña), rainfall was well above normal, leading to large discrepancies between predictions and observed rainfall. On the opposite hand, in 2023 (a drier year as we have seen in the rain shadow of El Niño), there were errors early in the rainy season with the model predicting heavy rainfall, yet the rainfall did not match the average. RF and GRU models have temporal trends similar to that of XGB, except they oscillate with error a little further. In contrast, convolutional and attention-based deep learning models (TCN, CNN, Transformer) show greater, less stable errors over time, because some models have very high sensitivity to monthly outliers, and one or two "bad" months can significantly increase the overall error.

Overall, this pattern confirms that all models really struggle with rainfall prediction during transition seasons or extreme climate processes. In terms of temporal and operational challenges, these relate primarily to (1) the inability to better capture long-term climate variability, despite employing a 44-year time series, and (2) monthly temporal resolution, which fails to fully reflect the intra-monthly dynamics affecting rainfall accumulation.

3.3. Prediction Results

Spatial predictions of annual rainfall with different models (RF, XGB, SVR, MLP, LSTM, GRU, CNN, TCN, Transformer) exhibit a fairly consistent distribution pattern in the upper Bengawan Solo region, but with varying intensities between models (**Fig. 13**). Overall, the spatial dynamics of traditional ML models (RF in particular) display relatively smoother and more stable spatial patterns, whereas DL models like LSTM, GRU, and CNN exhibit sharper spatial fluctuations, especially in mountainous areas. The spatial distribution indicates that the highest annual rainfall is anticipated to occur in areas adjacent to the Merapi, Merbabu, and Lawu mountains. It mirrors the orographic conditions in the area, in which moist air rising produces higher orographic rainfall. In contrast, the central part of the watershed tends to have lower and more homogeneous rainfall. These are consistent with the RF, XGB, and MLP models, and in fact, DL models like CNN or Transformer often result in extreme predictions at certain points, which can be attributed to overfitting issues.

The differences between the two models are apparent in the sharpness of the spatial gradients. DL-based models typically can be "sensitive" to local variations, so that predictions are usually quite heterogeneous. This might prove their capacity for finding local anomalies, but it also poses possibilities of instability in predicting. When compared to this, RF models give smoother distributions, which can be practically put to use (like water planning, flood reduction, etc). This phenomenon is also in line with the results of Chen et al. (2022) and Zandi et al. (2022): model integration must balance the trade-off between generalization and spatial sensitivity.

In sum, the spatial results of models show that traditional machine learning models such as Random Forest (RF), Extreme Gradient Boosting (XGB), and Multi-Layer Perceptron (MLP) are effective in keeping spatial rainfall predictions stable and provide consistent and relatively homogeneous distributions in nearly every region of the Bengawan Solo watershed.

On the other hand, in the case of deep learning models, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), and Transformer, the prediction variability is higher and is likely to give large values at multiple places, particularly where orographic complexity is high. Another interesting discovery was that the largest spatial variations are found in mountainous zones. This emphasizes the necessary relevance of topographical parameters and/or orographic processes for the calibration process and the construction of spatiotemporal rainfall prediction models in tropical regions.

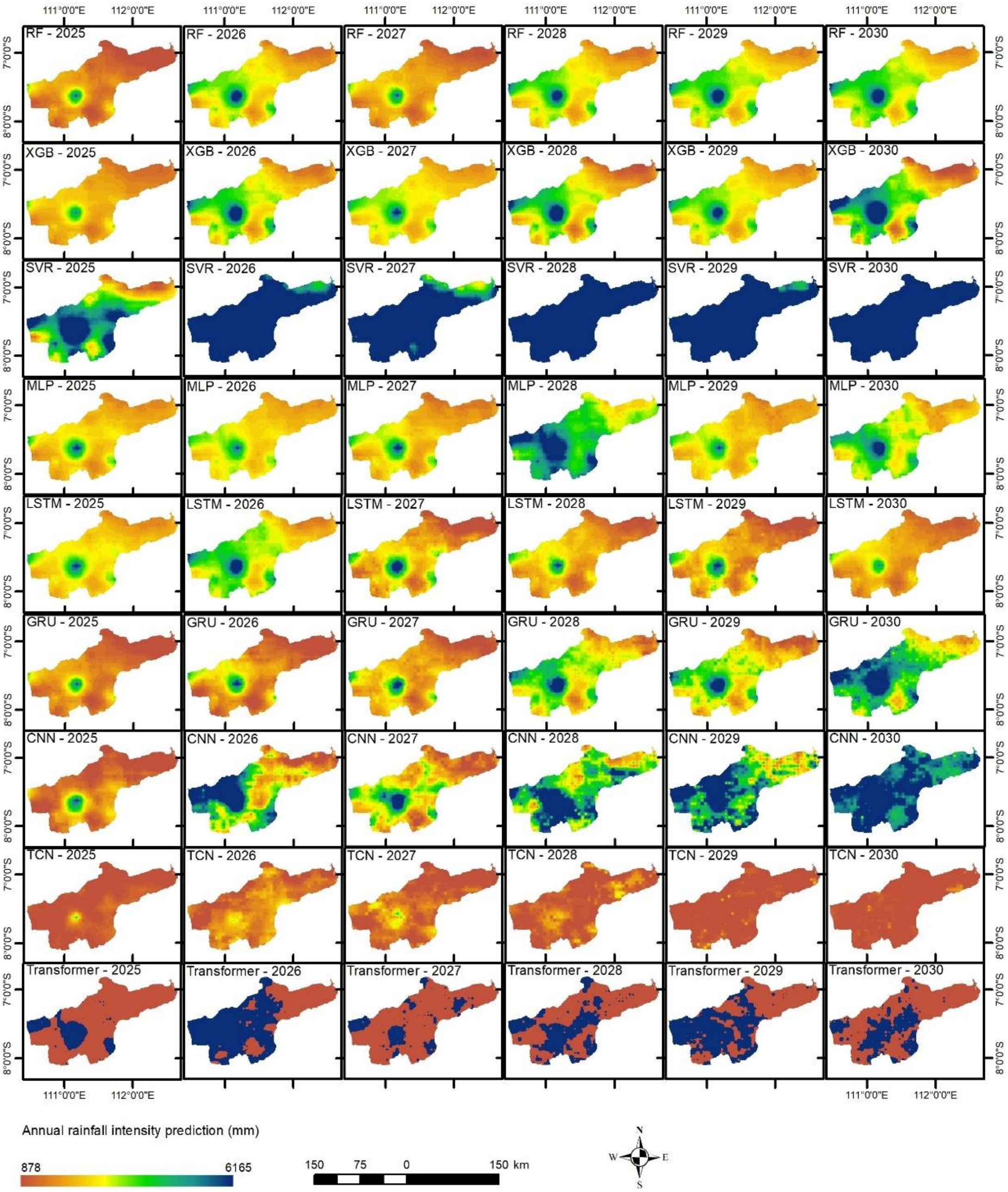


Fig. 13. Spatial Prediction Results for Annual Rainfall 2025–2030 for all models.

4. DISCUSSIONS

This work demonstrates that the decision tree-based gradient boosting models, with XGBoost (XGB), are most trustworthy for predicting spatiotemporal rainfall for the Bengawan Solo River Basin. XGB is the most consistent in providing overall accuracy (MAE ≈ 59 mm; RMSE ≈ 80 mm; $R^2 \approx 0.73$), followed by RF and GRU for both temporally (month-to-month) and spatially (location-to-location) data. The supremacy of the XGB-RF relationship agrees with that of CHIRPS satellite rainfall data in tropical regions, where rainfall strongly, but not linearly, correlates with its controlling features and is subject to noise and observational bias. In this light, ensemble decision trees are very powerful since they can capture nonlinear interaction in a complicated way without strict distributional assumptions, and they provide less sensitivity to hyperparameter selection than large-capacity deep learning models (Breiman, 2001; Hong et al., 2021; Sachindra et al., 2018). Bagging and boosting mechanisms help the model to smooth out local uncertainties between trees, leading to stable predictions over 523 grids with a vast variation in topography and microclimates.

Comparison with the literature shows that these findings are consistent with previous studies reporting the superiority of ensemble decision tree models for highly variable hydrometeorological data. Hong et al. (2021) and Sachindra et al. (2018) demonstrated that RF and gradient boosting models tend to outperform other nonlinear methods when data are limited, spatially heterogeneous, and contain difficult-to-model noise. In this study, despite the long temporal sample size (44 years), the spatial sample size was only 523 grid points within a single watershed, so the sample-to-parameter ratio was much more favorable for decision tree models than for very deep deep learning architectures. These results reinforce that, for data configurations like this, a "moderate but robust capacity" strategy is more effective than "very large capacity but highly dependent on data volume and diversity."

On the other hand, deep learning models exhibit more diverse behavior. GRU emerged as the most competitive deep learning model, with performance close to XGB ($R^2 \approx 0.72$) and fairly good temporal stability. This indicates that lighter recurrent architectures with fewer parameters, such as GRUs, are more suitable for moderate-length monthly time series because they require fewer parameters to learn and are easier to calibrate. Conversely, LSTM showed performance not far from MLP and SVR. At the same time, convolutional and attention-based architectures (TCN, CNN, Transformer) tended to have higher MAE and RMSE and lower R^2, with significant variability in errors across months and locations. This pattern aligns with findings by Chattopadhyay et al. (2020), Aswin & Geetha (2021), and reviews by Lim & Zohren (2021), which state that advanced deep learning, especially CNN and Transformer, only show clear advantages when: (i) the temporal resolution is very high (daily or sub-daily), (ii) the sample size is very large, and (iii) additional rich spatial/climate information is available. For monthly resolution and limited domains such as a single watershed, this large capacity can easily lead to overfitting and instability.

Spatial analysis of error distribution confirms that XGB and GRU not only outperform on average but are also more robust across most of the watershed area. Both models maintain an average MAE of 59–60 mm and an R^2 of 0.71–0.72 across most grids, while RF lags slightly behind with marginally worse MAE and R^2. Conversely, TCN, CNN, and Transformer show much higher spatial MAE (≈76–85 mm) and lower R^2 (≈0.38–0.50), with many grids performing near or even below the climatological baseline. The spatial error patterns also reveal that areas with steep topographic gradients, such as the slopes of Merbabu, Merapi, and Lawu, as well as the upper reaches of the watershed, tend to have higher errors. This is consistent with studies by Gu et al. (2020) and Funk et al., which indicate that satellite-based rainfall estimates in tropical mountainous regions are often biased due to sensor limitations in capturing small-scale convective clouds and complex orographic effects. Therefore, some of the errors observed across all models, including XGB and GRU, also reflect the limitations of the "ground truth" rainfall representation itself.

From a temporal perspective, the analysis of monthly MAE and RMSE shows that all models exhibit higher errors during the monsoon transition months (March–April and October–November) and in years with strong climate anomalies (e.g., La Niña 2020–2022 and El Niño 2023). In these conditions, rainfall intensity and distribution change rapidly and often do not follow periodic patterns that can be captured by models based on historical time series. Nevertheless, XGB and GRU are relatively consistent in maintaining the highest performance rankings across most months, while TCN, CNN, and Transformer exhibit sharp, unstable error spikes. This indicates that the capacity to model short-term non-linearity alone is not sufficient; models must also be able to generalize to rare but impactful climate regime changes. These findings align with McBride et al. (2003) and Lestari et al. (2021), which emphasize that rainfall variability in Indonesia is heavily influenced by ENSO and

IOD, making purely time-series-based local models potentially struggle when faced with rare extreme events.

Compared to other regions, model performance patterns in the Bengawan Solo River Basin show a strong correlation with the hydrometeorological context and data structure. In the North-Western Himalayas, dominated by steep mountainous topography with sharp elevation gradients, multivariable daily station data, and a time span of 1980–2021, DL models (specifically Bi-LSTM and LSTM) provide the lowest RMSE/MAE and outperform ML, while among ML models, ANN and KNN outperform RF and SVR, and accuracy is highly sensitive to elevation (Wani et al., 2024). In five large UK cities with a temperate maritime climate and hourly rainfall data, Stacked-LSTM and Bidirectional-LSTM architectures also reportedly outperform XGBoost and ML ensembles, although they still struggle to capture very abrupt rainfall surges (Barrera-Animas et al., 2022).

Conversely, a study in the Aligarh District of India showed that for daily and monthly rainfall with standard meteorological predictors, CatBoost and RF provide strong correlation and outperform SVR, confirming the reliability of tree ensembles for more aggregated time scales (Abdullah & Said, 2025). In African cities with humid tropical to dry Mediterranean climates, DL (especially single RNN) performs best for highly non-linear daily rainfall, with relative humidity and antecedent rainfall as key predictors (Samson & Aweda, 2025). In this context, the dominance of XGB, RF, and GRU in monthly CHIRPS-based Bengawan Solo annual rainfall prediction can be understood as a consequence of a combination of: a tropical monsoon climate with a strong monsoon but averaged daily pattern, watershed-scale orographic heterogeneity, and limitations of atmospheric predictors, thus significantly reducing the advantages of very deep DL architectures compared to other regions and time scales.

In terms of data validity, CHIRPS is supported by a broad body of validation studies across very different hydroclimatic settings. At the global scale, a recent synthesis shows that CHIRPS generally attains correlations above 0.7 at monthly and seasonal scales and is widely judged suitable for hydroclimatic analysis, with reduced skill mainly in very arid and high-mountain regions (Du et al., 2023).

Regionally, CHIRPS has been shown to outperform or rival reanalysis and other satellite products in complex tropical and subtropical terrains: for example across Ethiopia, where it consistently outperforms ERA5 and reproduces the seasonal cycle and interannual variability (Ahmed et al., 2024; Geleta & Deressa, 2020), in the tropical Andes and Antioquia where it captures orographic gradients and ENSO-related variability (López-Bermeo et al., 2022; Rivera et al., 2018), in the Amazon basin where annual totals are reproduced with R^2 ~0.98 (da Motta Paca et al., 2020), over the Qinghai–Tibet Plateau where CHIRPS compares favorably with MSWEP at monthly scales (Liu et al., 2019), and in semi-arid and drought-prone regions such as Northeast Brazil and eastern Africa (Paredes-Trejo et al., 2017; Dinku et al., 2018). These studies consistently report that CHIRPS performs best for monthly–annual aggregates and large-basin applications, while underestimating some local extremes, a trade-off that aligns with our use of 0.05° monthly CHIRPS data aggregated to annual rainfall over a large tropical watershed.

Several limitations of this study should be noted to contextualize the results. First, the temporal resolution of the input data is limited to monthly aggregation, so intra-monthly dynamics, such as daily extreme rainfall intensity or consecutive rainfall events, are not explicitly represented, even though these phenomena significantly contribute to monthly accumulation. Second, although 523 grid points provide much better spatial coverage than 98 points, the representation of microclimates in mountainous areas and narrow valleys remains far from perfect. Third, the model in this study has not yet integrated global and regional climate predictors, such as the Niño 3.4 index, the Dipole Mode Index (DMI), or the MJO, which are important for explaining seasonal rainfall variability in Indonesia. Fourth, the approach used is entirely based on statistical data without explicit linkage to physical processes within the hydrological system, so the mechanistic interpretation of some error patterns remains limited.

The practical implications and directions for further research that emerge from these findings are quite clear. For operational applications such as early flood warning and water resource planning in the Bengawan Solo watershed, ensemble decision tree models (XGB, RF) and relatively lightweight recurrent networks (GRU) are currently the most realistic choices, offering the best trade-off between accuracy, spatio-temporal stability, and computational complexity. On the other hand, the potential of deep learning (DL) is not fully exhausted; improvements can be achieved by (i) enriching features with ENSO, IOD, and MJO indices, (ii) increasing the temporal resolution of training to daily or decadal scales, (iii) applying more aggressive regularization and early stopping, and (iv) exploring transfer learning from larger climate domains.

Additionally, developing hybrid approaches that combine statistical/machine learning models with physical hydrological models can help reduce structural ambiguities and enhance generalization capabilities, especially under climate change scenarios. Thus, this study not only demonstrates that machine learning, particularly XGB and RF, still outperforms most advanced deep learning architectures in limited-data configurations but also provides a concrete roadmap for improving spatio-temporal rainfall prediction models in tropical regions with complex topography. Furthermore, a useful future step would be to integrate a self-imputation and self-quality assessment framework to stabilize the lunar series before training and applying various types of optimizers (Haidu et al., 2025; Sriwahyuni et al., 2025).

5. CONCLUSION

This study compares nine ML and DL algorithms for predicting rainfall in the Bengawan Solo River Basin using CHIRPS data from 1981 to 2024. Results show decision tree-based gradient boosting models, especially XGBoost (XGB), perform best during 2020–2024 (MAE ≈ 59 mm; RMSE ≈ 80 mm; $R^2 \approx 0.73$), followed by Random Forest (RF) and Gated Recurrent Unit (GRU), with minor differences. GRU was the top DL model ($R^2 \approx 0.72$), while CNN, TCN, and Transformer architectures had larger errors and variability. Findings indicate ensemble decision trees and lightweight recurrent neural networks are more reliable than complex DL models in large tropical river basins with spatial heterogeneity and moderate time series length.

Spatially, accuracy is higher in the basin's center and lower in peripheral or mountainous areas, affected by microclimate, orographic effects, and satellite limitations. Temporally, errors peak during monsoon transitions (March–April, October–November) and during climate anomalies, challenging models during ENSO and IOD events. Practically, XGB, RF, and GRU are suitable for operational rainfall forecasts, aiding water management and flood mitigation in tropical regions. Limitations include monthly data resolution, satellite biases, and a lack of external climate predictors like ENSO and IOD indices.

Future research should integrate high-resolution data, include climate indicators, use regularization and transfer learning to enhance DL, and explore hybrid models. Overall, this study highlights trade-offs and opportunities in ML and DL for tropical rainfall prediction, serving as a guide for developing accurate, adaptive hydrometeorological systems in Indonesia and similar areas.

ACKNOWLEDGEMENT

This study is funded by the HIT Program of 2025 of Universitas Muhammadiyah Surakarta.

REFERENCES

Abdullah, M., & Said, S. (2025). Performance evaluation of machine learning regression models for rainfall prediction. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, *49*(4), 4231-4250.

Ahmed, J. S., Buizza, R., Dell'Acqua, M., Demissie, T., & Pè, M. E. (2024). Evaluation of ERA5 and CHIRPS rainfall estimates against observations across Ethiopia. *Meteorology and Atmospheric Physics*, *136*(3), 17.

Alikhanov, B., Pulatov, B., Samiev, L., (2024). Impact of Climate Change on the Cryosphere of the Ugam Chatkal National Park, Bostonliq District, Uzbekistan, During the Post-Soviet Period, Based on Remote Sensing and Statistical Analysis. *Forum Geografi*, 38, 302–316. https://doi.org/10.23917/forgeo.v38i3.4405

Al-Samrraie, L.A., Abdalla, A.M., Alrawashdeh, K.A.-B., Bsoul, A.A., Awad, M.A., Alzboon, K., Al-Taani, A.A., (2025). Deep Learning Models Based on CNN, RNN, and LSTM for Rainfall Forecasting: Jordan as a Case Study. *Mathematical Modeling of Engineering Problems,* 12.

Amini, E., Zolfaghari, A., Kaboli, H., Rahimi, M., (2022). Estimation of rainfall erosivity map in areas with limited number of rainfall stations (case study: Semnan Province). *Iranian Journal of Soil and Water Research,* 53, 2027–2044.

Avand, M., Moradi, H.R., Ramazanzadeh Lasboyee, M., (2021). Spatial prediction of future flood risk: an approach to the effects of climate change. *Geosciences*, 11, 25.

Baig, F., Ali, L., Faiz, M.A., Chen, H., Sherif, M., (2024). How accurate are the machine learning models in improving monthly rainfall prediction in hyper arid environment? *Journal of Hydrology,* 633, 131040.

Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, *7*, 100204.

Chen, C., Zhang, Q., Kashani, M.H., Jun, C., Bateni, S.M., Band, S.S., Dash, S.S., Chau, K.-W., (2022). Forecast of rainfall distribution based on fixed sliding window long short-term memory. Engineering *Applications of Computational Fluid Mechanics* 16, 248–261. https://doi.org/10.1080/19942060.2021.2009374

CHIRPS: Rainfall Estimates from Rain Gauge and Satellite Observations | Climate Hazards Center - UC Santa Barbara [WWW Document], n.d. URL https://www.chc.ucsb.edu/data/chirps (accessed 8.30.2025).

da Motta Paca, V. H., Espinoza-Davalos, G. E., Moreira, D. M., & Comair, G. (2020). Variability of trends in precipitation across the Amazon River basin determined from the CHIRPS precipitation product and from station records. *Water*, *12*(5), 1244.

Das, P., Posch, A., Barber, N., Hicks, M., Duffy, K., Vandal, T., Singh, D., Werkhoven, K. van, Ganguly, A.R., (2024). Hybrid physics-AI outperforms numerical weather prediction for extreme precipitation nowcasting. *Climate and Atmospheric Science,* 7, 282.

Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H., & Ceccato, P. (2018). Validation of the CHIRPS satellite rainfall estimates over eastern Africa. *Quarterly Journal of the Royal Meteorological Society*, *144*, 292-312.

Du, H., Tan, M. L., Zhang, F., Chun, K. P., Li, L., & Kabir, M. H. (2024). Evaluating the effectiveness of CHIRPS data for hydroclimatic studies. *Theoretical and Applied Climatology*, *155*(3), 1519-1539.

El Hafyani, M., El Himdi, K., El Adlouni, S.-E., (2024). Improving monthly precipitation prediction accuracy using machine learning models: a multi-view stacking learning technique. *Frontiers in Water,* 6, 1378598.

Espeholt, L., Agrawal, S., Sønderby, C., Kumar, M., Heek, J., Bromberg, C., Gazen, C., Carver, R., Andrychowicz, M., Hickey, J., (2022). Deep learning for twelve-hour precipitation forecasts. *Nature communications* 13, 1–10.

Fang, L., Shao, D., (2022). Application of long short-term memory (LSTM) on the prediction of rainfall-runoff in karst area. *Frontiers in Physics*, 9, 790687.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., (2015). The Climate Hazards Infrared Precipitation with Stations—A New Environmental Record for Monitoring Extremes. *Scientific Data*, 2, 1–21.

Geleta, C. D., & Deressa, T. A. (2021). Evaluation of climate hazards group infrared precipitation station (CHIRPS) satellite-based rainfall estimates over Finchaa and Neshe Watersheds, Ethiopia. *Engineering Reports*, *3*(6), e12338.

Gu, J., Liu, S., Zhou, Z., Chalov, S.R., Zhuang, Q., (2022). A stacking ensemble learning model for monthly rainfall prediction in the Taihu Basin, China. *Water,* 14, 492.

Haidu, I., El Orfi, T., Magyari-Sáska, Z., Lebaut, S., & El Gachi, M. (2024). Modeling the Long-Term Variability in the Surfaces of Three Lakes in Morocco with Limited Remote Sensing Image Sources. *Remote Sensing*, *16*(17), 3133. https://doi.org/10.3390/rs16173133

Haidu, I., Magyari-Sáska, Z., & Magyari-Sáska, A. (2025). Spatio-Temporal Gap Filling of Sentinel-2 NDI45 Data Using a Variance-Weighted Kalman Filter and LSTM Ensemble. *Sensors*, *25*(17), 5299. https://doi.org/10.3390/s25175299

Haq, D.Z., Novitasari, D.C.R., Hamid, A., Ulinnuha, N., Farida, Y., Nugraheni, R.D., Nariswari, R., Rohayani, H., Pramulya, R., Widjayanto, A., (2021). Long short-term memory algorithm for rainfall prediction based on El-Nino and IOD data. *Procedia Computer Science,* 179, 829–837.

Hu, C., Wu, Q., Li, H., Jian, S., Li, N., Lou, Z., (2018). Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water,* 10, 1543.

Jumadi, J., Danardono, D., Roziaty, E., Ulinuha, A., Supari, S., Choy, L. K., Sattar, F., Nawaz, M. (2025). AI-Driven Ensemble Learning for Spatio-Temporal Rainfall Prediction in the Bengawan Solo River Watershed, Indonesia. *Sustainability*, 17(20), 9281. https://doi.org/10.3390/su17209281

Jumadi, J., Danardono, D., Priyono, K. D., Roziaty, E., Masruroh, H., Rohman, A., Amin, C., Hadibasyir, H. Z., Fikriyah, V. N., Nawaz, M., Sattar, F., & Lotfata, A. (2024). Utilizing Open Access Spatial Data for Flood Risk Mapping: A Case Study in the Upper Solo Watershed. Geoplanning: *Journal of Geomatics and Planning*, 11(2), 189-204. https://doi.org/10.14710/geoplanning.11.2.189-204

Kasihairani, D., Hidayat, R., Supari, S., (2024). Assessing the Reliability of Predicted Decadal Surface Temperatures in Southeast Asia. *Forum Geografi,* 38, 413–425. https://doi.org/10.23917/forgeo.v38i3.5402

Kim, S., Shin, J.-Y., Heo, J.-H., (2025). Assessment of Future Rainfall Quantile Changes in South Korea Based on a CMIP6 Multi-Model Ensemble. *Water,* 17, 894.

Koya, S.R., Roy, T., (2024). Temporal Fusion Transformers for Streamflow Prediction: Value of Combining Attention with Recurrence. *Journal of Hydrology,* 637, 131301.

Kumar, V., Kedam, N., Kisi, O., Alsulamy, S., Khedher, K.M., Salem, M.A., (2025). A Comparative Study of Machine Learning Models for Daily and Weekly Rainfall Forecasting. *Water Resours Manage,* 39, 271–290. https://doi.org/10.1007/s11269-024-03969-8

Kundu, S., Biswas, S.K., Tripathi, D., Karmakar, R., Majumdar, S., Mandal, S., (2023). A review on rainfall forecasting using ensemble learning techniques. e-Prime-Advances in Electrical Engineering, *Electronics and Energy*, 6, 100296.

Laptev, N., Yosinski, J., Li, L.E., Smyl, S., (2017). Time-series extreme event forecasting with neural networks at Uber, in: *International Conference on Machine Learning*. SN, pp. 1–5.

Lim, B., Zohren, S., (2021). Time-series forecasting with deep learning: a survey. *Phil. Trans. R. Soc.,* A. 379, 20200209. https://doi.org/10.1098/rsta.2020.0209

Liu, J., Shangguan, D., Liu, S., Ding, Y., Wang, S., & Wang, X. (2019). Evaluation and comparison of CHIRPS and MSWEP daily-precipitation products in the Qinghai-Tibet Plateau during the period of 1981–2015. *Atmospheric Research*, *230*, 104634.

López-Bermeo, C., Montoya, R. D., Caro-Lopera, F. J., & Díaz-García, J. A. (2022). Validation of the accuracy of the CHIRPS precipitation dataset at representing climate variability in a tropical mountainous region of South America. *Physics and Chemistry of the Earth, Parts A/B/C*, *127*, 103184.

Magyari-Sáska, Z., Haidu, I., & Magyari-Sáska, A. (2025). Experimental Comparative Study on Self-Imputation Methods and Their Quality Assessment for Monthly River Flow Data with Gaps: Case Study to Mures River. *Applied Sciences*, *15*(3), 1242. https://doi.org/10.3390/app15031242

Miao, Q., Pan, B., Wang, H., Hsu, K., Sorooshian, S., (2019). Improving monsoon precipitation prediction using combined convolutional and long short-term memory neural network. *Water*, 11, 977.

Musiyam, M., Jumadi, J., Fikriyah, V. N., Masruroh, H., Septiyani, E. D., Amin, C., Hadibasyir,H.Z., Sattar, F., Nawaz, M. (2025). Flood Risk Analysis Using Spatial Synthetic Population in the Upper Bengawan Solo Watershed, Indonesia. *Forum Geografi*, 39(3), 383-397. https://doi.org/10.23917/forgeo.v39i3.13611

Naik, R., Majhi, B., (2025). Explainable AI reverse verification approach for monthly rainfall prediction in Chhattisgarh, India. *Theor Appl Climatol*, 156, 412. https://doi.org/10.1007/s00704-025-05645-2

Nelson, B.K., (1998). Time Series Analysis Using Autoregressive Integrated Moving Average (ARIMA) Models. *Academic Emergency Medicine*, 5, 739–744. https://doi.org/10.1111/j.1553-2712.1998.tb02493.x

Ni, L., Wang, D., Singh, V.P., Wu, J., Wang, Y., Tao, Y., Zhang, J., (2020). Streamflow and rainfall forecasting by two long short-term memory-based models. *Journal of Hydrology,* 583, 124296.

Pan, X., Hou, J., Gao, X., Chen, G., Li, D., Imran, M., Li, X., Yang, N., Ma, M., Zhou, X., (2025). LSTM Model-Based Rapid Prediction Method of Urban Inundation with Rainfall Time Series. *Water Resour Manage,* 39, 661–688. https://doi.org/10.1007/s11269-024-03972-z

Papacharalampous, G., Tyralis, H., Doulamis, N., Doulamis, A., (2025). Ensemble learning for uncertainty estimation with application to the correction of satellite precipitation products. *Machine Learning: Earth,* 1, 015004.

Paredes-Trejo, F. J., Barbosa, H. A., & Kumar, T. L. (2017). Validating CHIRPS-based satellite precipitation estimates in Northeast Brazil. *Journal of arid environments*, *139*, 26-40.

Phoophathong, T., Laosuwan, T., Sangpradid, S., Uttaruk, Y., & Angkahad, T. (2025). Improvement of Traditional and Hybrid Interpolation Techniques Using Support Vector Machine for Land Surface Temperature Analysis in Urban Areas. *Geographia Technica*, *20*(1). DOI: 10.21163/GT_2025.201.21

Praveen, B., Talukdar, S., Shahfahad, Mahato, S., Mondal, J., Sharma, P., Islam, A.R.M.T., Rahman, A., (2020). Analyzing trend and forecasting of rainfall changes in India using non-parametric and machine learning approaches. *Scientific reports*, 10, 10342.

Priyana, Y., Jumadi, Anna, A.N., Rudiyanto, (2023). Farmers' adaptation strategies in facing drought disasters (A case study in some areas of the Bengawan Solo watershed). *Proceedings of the international summit on education, technology, and humanity, AIP Conf. Proc.* 2727, 050026.

Rivera, J. A., Marianetti, G., & Hinrichs, S. (2018). Validation of CHIRPS precipitation dataset along the Central Andes of Argentina. *Atmospheric Research*, *213*, 437-449.

Samson, T. K., & Aweda, F. O. (2025). Comparative study of single and hybrid deep learning models for daily rainfall prediction in selected African cities. *Scientific Reports*, *15*(1), 42718.

Santhyami, S., Jumadi, J., Priyono, K. D., Rahayu, T., Sari, D. N., Othman, M., & Rudi, R. (2025). Spatio-Temporal Mo-deling for the Analysis of Hydrological Drought and its Impact on Rice

Production in the Upper Bengawan Solo Basin, Central Java, Indonesia. *Geographia Technica*, *20*(2). DOI: 10.21163/GT_2025.202.16

Schaffer, A.L., Dobbins, T.A., Pearson, S.-A., (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. *BMC Med Res Methodol,* 21, 58. https://doi.org/10.1186/s12874-021-01235-8

Shetty, S., Dharmendra, D., Bankapur, S., Prasad, P., (2024). HydroStack: A Hybrid Meta-Ensemble Machine Learning Framework for Accurate Annual Rainfall Prediction, in: 2024 *8th International Conference on I-SMAC* (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). IEEE, pp. 1926–1934.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W., (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28.

Sivadasan, E.T., Sundaram, N.M., Santhosh, R., (2025). Deep Learning for Energy Forecasting Using Gated Recurrent Units and Long Short-Term Memory. *Journal of Intelligent Systems & Internet of Things*, 14.

Slater, L.J., Arnal, L., Boucher, M.-A., Chang, A.Y.-Y., Moulds, S., Murphy, C., Nearing, G., Shalev, G., Shen, C., Speight, L., (2023). Hybrid forecasting: blending climate predictions with AI models. *Hydrology and Earth System Sciences,* 27, 1865–1889.

Sriwahyuni, L., Nurdiati, S., Nugrahani, E. H., & Najib, M. K. (2025). Performance of Machine Learning for Imputing Missing Daily Rainfall Data in East Java Under Multiple Satellite Data Models. *Geographia Technica*, 20(1). DOI: 10.21163/GT_2025.201.23

Wani, O. A., Mahdi, S. S., Yeasin, M., Kumar, S. S., Gagnon, A. S., Danish, F., ... & Mattar, M. A. (2024). Predicting rainfall using machine learning, deep learning, and time series models across an altitudinal gradient in the North-Western Himalayas. *Scientific Reports*, *14*(1), 27876.

Ward, P.J., Van Pelt, S.C., De Keizer, O., Aerts, J.C.J.H., Beersma, J.J., Van Den Hurk, B.J.J.M., Te Linde, A.H., (2014). Including climate change projections in probabilistic flood risk assessment. *J Flood Risk Management,* 7, 141–151. https://doi.org/10.1111/jfr3.12029

Willard, J., Jia, X., Xu, S., Steinbach, M., Kumar, V., (2020). Integrating physics-based modeling with machine learning: *A survey. arXiv preprint arXiv*:2003.04919 1, 1–34.

Winsemius, H.C., Aerts, J.C., Van Beek, L.P., Bierkens, M.F., Bouwman, A., Jongman, B., Kwadijk, J.C., Ligtvoet, W., Lucas, P.L., Van Vuuren, D.P., (2016). Global drivers of future river flood risk. *Nature Climate Change,* 6, 381–385.

Yang, Y., Tan, G., Shen, Z., Zhang, Y., Fei, Q., Liu, X., Dogar, M.A., (2025). Integrating Physical Dynamics into Ensemble ML for Improved Monthly Rainfall Forecasting. *Earth Syst Environ*. https://doi.org/10.1007/s41748-025-00691-2

Yin, W., Zhou, C., Tian, Y., Qiu, H., Zhang, W., Chen, H., Liu, P., Zhao, Q., Kong, J., Yao, Y., (2025). Accurate Rainfall Prediction Using GNSS PWV Based on Pre-Trained Transformer Model. *Remote Sensing,* 17, 2023.

Zandi, O., Zahraie, B., Nasseri, M., Behrangi, A., (2022). Stacking machine learning models versus a locally weighted linear model to generate high-resolution monthly precipitation over a topographically complex area. *Atmospheric Research*, 272, 106159. https://doi.org/10.1016/j.atmosres.2022.106159

Zhou, Y., Cui, Z., Lin, K., Sheng, S., Chen, H., Guo, S., Xu, C.-Y., (2022). Short-term flood probability density forecasting using a conceptual hydrological model with machine learning techniques. *Journal of Hydrology*, 604, 127255.

HYBRID DEEP LEARNING AND REINFORCEMENT FRAMEWORK FOR PHYSICALLY CONSISTENT PRECIPITATION CONTROL ALONG THE MOROCCAN-IBERIAN COASTS

Imad KATIBA[1*], Hanae BELMAJDOUB [1,2] & Khalid MINAOUI [1]

DOI: 10.21163/GT_2026.212.06

ABSTRACT

This study introduces a hybrid framework that integrates deep learning and reinforcement learning to optimize regional precipitation along the Moroccan and Iberian coasts of the North-Eastern Atlantic. A Bidirectional Long Short-Term Memory (BiLSTM) and Multi-Layer Perceptron (MLP) network was first trained to forecast monthly rainfall from multi-variable climate sequences combining key atmospheric and oceanic variables. The predictions were then coupled with a Twin Delayed Deep Deterministic Policy Gradient (TD3) controller capable of performing small, physically coherent adjustments to selected variables such as sea-level pressure, specific humidity, and near-surface winds. Through adaptive interactions with the predictive environment, the control agent learned strategies that enhance rainfall while maintaining atmospheric stability. The resulting framework demonstrates consistent improvements in simulated precipitation and reveals spatially coherent dominance patterns of sea-level pressure and humidity in rainfall modulation. The reinforcement-based control explicitly enforces stability and energetic constraints, ensuring physically admissible atmospheric responses. Overall, the proposed hybrid AI approach provides a physically interpretable foundation for regional water management and constitutes a decision-support framework for diagnosing precipitation sensitivity rather than an operational weather modification system. It can be extended to drought-prone or broader climatic domains to support future resilience and hydrological planning.

Keywords: *Deep learning; Reinforcement learning; BiLSTM; TD3; Climate control; Precipitation optimization; Moroccan–Iberian coasts; North-Eastern Atlantic; Atmospheric dynamics; Water resource management.*

1. INTRODUCTION

Coastal and oceanic regions off Morocco and Portugal are strongly influenced by large-scale ocean-atmosphere interactions that govern regional climate variability and its predictability. As highlighted by (Li et al., 2024), multiple coupled modes such as the El Niño-Southern Oscillation (ENSO), the North Atlantic Oscillation (NAO), and the Atlantic Multidecadal Oscillation (AMO) interact through complex teleconnections, ocean heat fluxes, and atmospheric bridges, generating anomalies that affect weather and climate patterns far beyond their regions of origin. Understanding these mechanisms remains a critical challenge for improving regional forecasts and assessing their socio-ecological impacts. Despite decades of advances in numerical weather prediction and global climate modeling, significant disparities persist in the capacity to deliver high-quality regional climate information, particularly in areas with limited computational or institutional resources. As noted by (Meque et al., 2021), the ability of National Meteorological and Hydrological Services to perform reliable modeling and deliver tailored climate services is often constrained by technical and financial limitations.

[1] *Laboratory of Research in Computer Science and Telecommunications, Faculty of Sciences, Mohammed V University, Rabat, Morocco. [*]Corresponding author: imad_katiba@um5.ac.ma (IK); hanae.belmajdoub@um5r.ac.ma (HB); khalid.minaoui@fsr.um5.ac.ma (KM),*
[2] *Faculty of Legal, Economic and Social Sciences – Souissi Campus, Mohammed V University, Rabat, Morocco.*

This emphasizes the need for innovative approaches that can complement traditional modeling frameworks, especially in regions where conventional models struggle to represent local dynamics accurately. In this context, surrogate modeling techniques based on machine learning have emerged as promising tools to reduce systematic biases in climate model outputs. For example, (Steininger et al., 2022) demonstrate how convolutional neural networks (CNNs) can be integrated into Model Output Statistics (MOS) schemes to refine climate projections and improve accuracy for key variables such as precipitation. Similarly, the availability of standardized benchmark datasets such as WeatherBench(Rasp et al., 2020) is facilitating transparent, reproducible comparisons among data-driven forecasting methods, accelerating progress in surrogate model design.

Several recent studies have shown that hybrid deep learning architectures, large-scale neural operators, and neuro-symbolic networks can extract fine-scale structures and identify early warning signals for climate tipping points (Pathak et al., 2022; Kurth et al., 2018; Sleeman et al., 2023). By combining physical modeling with advanced AI techniques, these approaches reveal emergent behaviors in coupled systems such as the atmosphere and ocean, enhancing both predictive skill and computational efficiency. Furthermore, as reviewed by (Ladi et al., 2022), machine learning methods have become integral to climate change mitigation and adaptation, particularly in coastal and urban contexts. Neural network-based approaches are now applied across topics ranging from geoengineering to land surface temperature management, demonstrating their versatility for addressing complex environmental challenges. Extending these advances to marine and atmospheric systems, (Ojo et al., 2022; Luo et al., 2022) show that Bayesian deep learning architectures can outperform traditional models in predicting sea surface temperature variability while providing robust uncertainty estimates. Such probabilistic formulations mitigate initialization bias and represent a promising pathway for near-term regional prediction, particularly across the North Atlantic corridor near Morocco and Portugal.

Parallel to these advances in predictive modeling, reinforcement learning (RL) and predictive control frameworks have emerged as powerful tools for optimizing complex environmental processes. However, value-based RL methods often suffer from function approximation errors that lead to overestimated value functions and suboptimal policies. The Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm (Fujimoto et al., 2018) addresses this issue by combining double Q-learning and delayed policy updates to improve stability and reduce overestimation bias. In related applications, neural network-driven predictive control systems have been successfully deployed in real-world energy and resource management contexts. For instance, (Routray et al., 2023; Al Sayed et al., 2024) demonstrate the effectiveness of RL in optimizing heating, ventilation, and air conditioning (HVAC) systems, while (Sierla et al., 2022) highlights the decisive role of reward design and action formulation in achieving stable, multi-objective control. More recently, (Durante et al., 2024) discussed the potential of multimodal AI agents equipped with adaptive perception to enhance the performance and resilience of RL systems operating in dynamic, uncertain environments.

In parallel, recent climate assessments, including the 2025 Global Climate Report by NOAA (NOAA National Centers for Environmental Information, 2025), continue to emphasize the growing frequency of precipitation anomalies and extreme hydro-climatic events. These findings highlight the urgent need for integrated frameworks capable not only of forecasting precipitation, but also of diagnosing the sensitivity of rainfall to its governing atmospheric drivers. Building upon these challenges, this study proposes a hybrid framework that couples deep learning-based precipitation prediction with reinforcement learning-driven control to investigate regional precipitation sensitivity and modulation. Unlike most existing reinforcement learning approaches applied to climate-related problems, which typically rely on abstract optimization settings or synthetic environments, the proposed framework operates directly on realistic hydro-atmospheric variables and explicitly enforces physical stability and energetic consistency during the control process. By jointly integrating prediction and physically constrained control within a unified structure, this approach provides new insights into how atmospheric pressure, humidity, and near-surface wind fields interact to regulate precipitation patterns along the Moroccan-Iberian Atlantic corridor.

The remainder of this paper is organized as follows: Section 2 introduces the study area and the datasets used. Section 3 describes the methodology and the proposed hybrid framework combining predictive modeling and reinforcement-based control. Section 4 presents and discusses the main results, including the regional patterns of precipitation modulation and variable influence. Finally, Section 5 concludes the study and outlines future research perspectives for regional hydroclimatic regulation.

2. STUDY AREA AND DATA PROCESSING

This study focuses on the North Atlantic sector, where atmospheric and oceanic variability jointly shape regional precipitation behavior. A coherent multi-source dataset was compiled to represent the physical state of this region over more than a decade of winter seasons. The analysis focuses on the winter months (December, January, February, and March) covering the period from December 2011 to March 2023, which corresponds to the most hydrologically active phase of the year. The study area extends from 20.6° N to 42.0° N in latitude and from 23.1° W to 4.8° W in longitude, encompassing the North-Eastern Atlantic domain and the adjacent coastal sectors of Morocco and the Iberian Peninsula, as illustrated in (**Fig. 1**), which delineates the spatial limits adopted for the study.

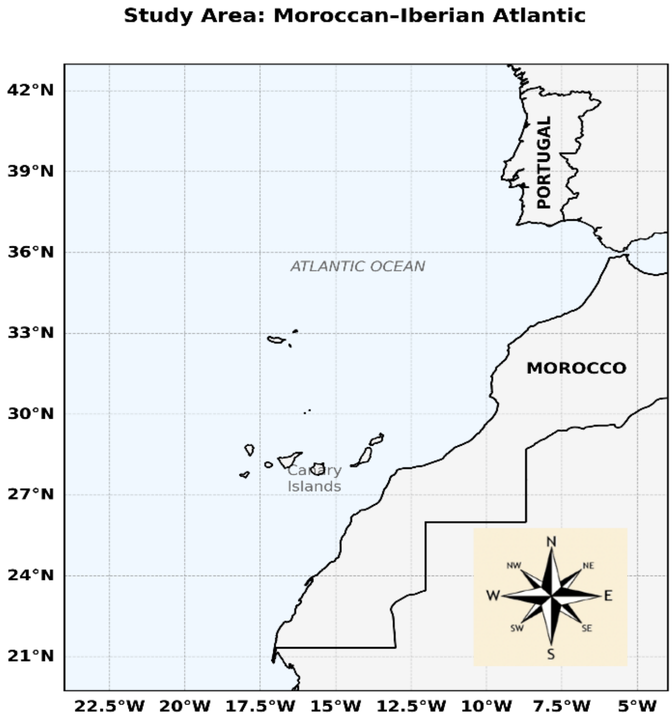


Fig. 1. Geographic extent of the Moroccan-Iberian Atlantic study area.

To ensure spatial consistency among all sources, each variable was interpolated onto a regular 0.25° grid. Precipitation data were obtained from the Integrated Multi-satellite Retrievals for GPM (IMERG) Version 07B Final Run (Huffman et al., 2023). Atmospheric variables, including sea-level pressure (SLP), 10 m wind components (U10M and V10M), specific humidity at 2 m (QV2M), and sea surface temperature (SST), were extracted from NASA's MERRA-2 reanalysis (Gelaro et al., 2017).Wind speed was calculated as the magnitude of the horizontal wind vector. Specifically, it was derived from the MERRA-2, 10 m wind components according to $WS = \sqrt{U10M^2 + V10M^2}$, ensuring consistency with the reanalysis convention.

In parallel, sea surface salinity (SSS) was obtained from the ECCO V4r4 dataset (Fukumori et al., 2021). Because these products have different native spatial resolutions, each field was resampled using bilinear interpolation applied through the SciPy griddata routine (Virtanen et al., 2020). This method estimates the value at each target grid cell as a weighted average of the four nearest source points, with weights proportional to their inverse distances along latitude and longitude. Compared to nearest-neighbor techniques, this approach produces smoother and physically continuous fields, preserving the relative spatial gradients of pressure, temperature, humidity, and salinity while aligning all variables in space and time. The operation therefore ensures both numerical stability and physical consistency, resulting in a homogeneous dataset suitable for subsequent analysis. All data were obtained from the NASA Earthdata portal (https://www.earthdata.nasa.gov).

The entire preprocessing workflow was designed to ensure internal coherence and numerical stability across the dataset. Missing values were filtered, temporal indices were standardized, and sliding sequences of four consecutive months were constructed for each grid point to provide dynamic inputs to the deep learning models. Precipitation data for March 2023 were isolated as independent targets for model validation. All variables were normalized using min-max scaling to guarantee comparability of physical magnitudes and stable convergence during subsequent computations. The resulting time series were organized into spatio-temporal matrices for each geolocation, ensuring full compatibility among atmospheric, marine, and hydrological parameters across the study domain.

An initial visualization of the spatial distribution of these variables for the first available month is presented (**Fig. 2**), which illustrates the regional gradients and large-scale contrasts that characterize the study area. This figure provides a comprehensive view of the atmospheric and oceanic structures used as the physical foundation for the predictive modeling and reinforcement-learning experiments described in the following sections.

3. METHODS

The methodology developed in this study integrates predictive modeling and reinforcement-based control to build a physically consistent framework for regional precipitation optimization. The approach relies on a hybrid architecture in which a deep learning predictor reproduces the temporal evolution of key atmospheric variables, while a reinforcement learning agent explores how small, physically admissible modifications to these variables can influence precipitation. The workflow operates in two sequential phases. First, a predictive model is trained to estimate future rainfall from multi-variable climate sequences derived from long-term reanalysis data. Then, the predicted values are used to guide a reinforcement learning agent that interacts with a virtual climate environment designed to preserve the internal dynamics of the atmosphere.

The agent's objective is to identify localized adjustment strategies that enhance precipitation while minimizing energetic cost and instability. This combined framework bridges statistical forecasting and adaptive control, allowing regional climate interactions to be represented not only as correlations but also as actionable mechanisms for hydrological modulation. This methodological design ensures that predictive learning and control adaptation remain coupled within a physically interpretable framework, where data-driven forecasts continuously inform reinforcement-based decisions. The following paragraphs describe the model architecture and the experimental configuration used for training and evaluation.

3.1. Model Architecture

The hybrid system consists of two main components: a predictive model based on a Bidirectional Long Short-Term Memory (BiLSTM) network (Graves & Schmidhuber, 2005), and a control module implemented through the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm. The *BiLSTM-MLP* predictor captures the temporal dependencies among the atmospheric and oceanic parameters that regulate precipitation dynamics.

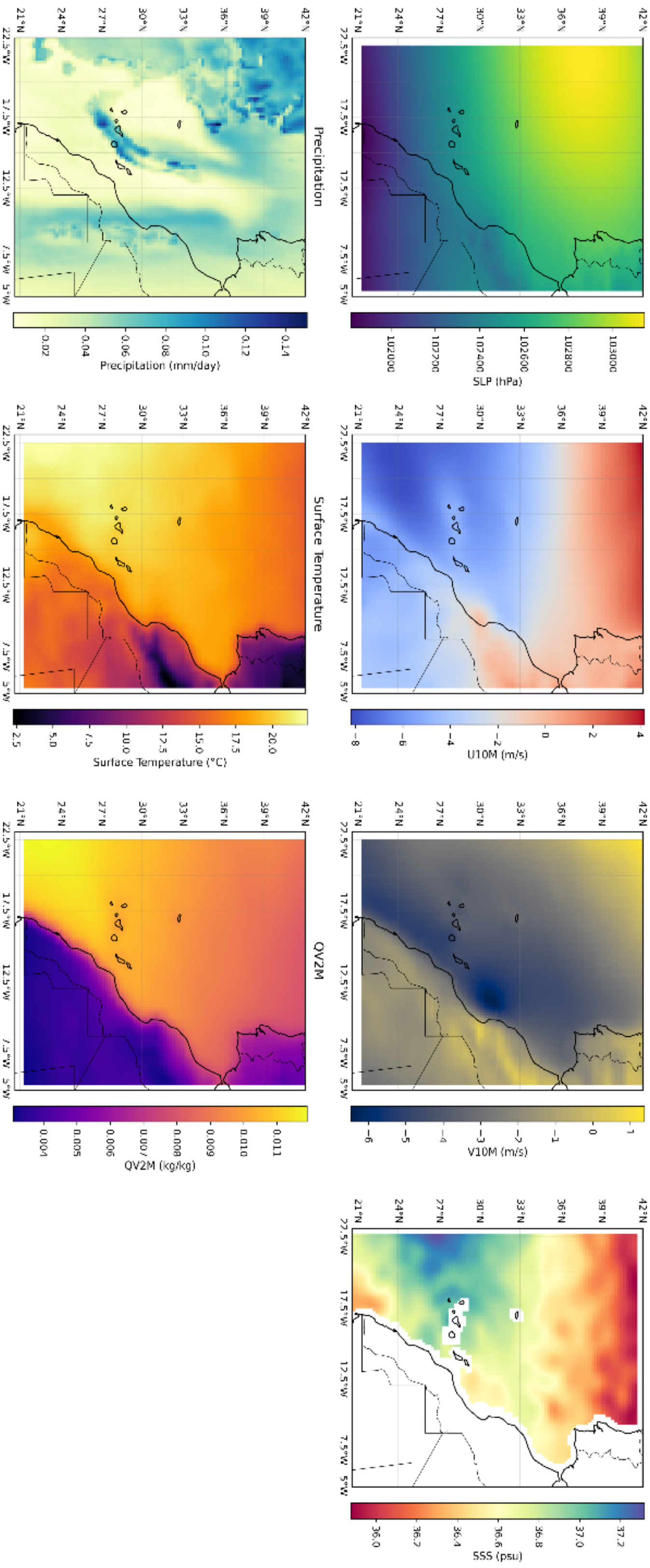


Fig. 2. Initial spatial distribution of the climate variables for the first month in the dataset.

For each grid point, the model processes four consecutive months of input data, including sea level pressure (SLP), specific humidity at two meters (QV2M), zonal and meridional wind components (U10M and V10M), wind speed, sea surface temperature (SST), and sea surface salinity (SSS). These variables describe the hydro-atmospheric state of the study region. The *BiLSTM* layer extracts both forward and backward temporal relationships, while a Multi-Layer Perceptron (MLP) (Rumelhart et al., 1986) refines the latent representations to produce the predicted precipitation for the following month. The overall structure of this predictive architecture is shown in (**Fig. 3**), which defines the baseline rainfall distribution used as reference for the control phase.

To ensure robust and stable predictions, the main hyperparameters of the BiLSTM-MLP network, such as the hidden-layer size, learning rate, and sequence length, were adjusted through a systematic grid-search procedure combined with early stopping on the validation root-mean-square error. This tuning process maintained an optimal balance between model complexity and generalization, preventing both underfitting and overfitting. Preliminary sensitivity tests indicated that excessively large hidden dimensions or high learning rates caused unstable convergence and degraded predictive performance, confirming the importance of careful hyperparameter selection to ensure numerical stability and physically consistent precipitation estimates.

After the predictive stage, the *BiLSTM* model is embedded within a reinforcement learning environment (Sutton & Barto, 2018) where the *TD3* agent acts as an adaptive controller. The agent performs small, continuous perturbations of limited amplitude (approximately 8%) on the variables SLP, QV2M, U10M, and V10M, with the aim of maximizing the predicted precipitation while maintaining physical coherence.

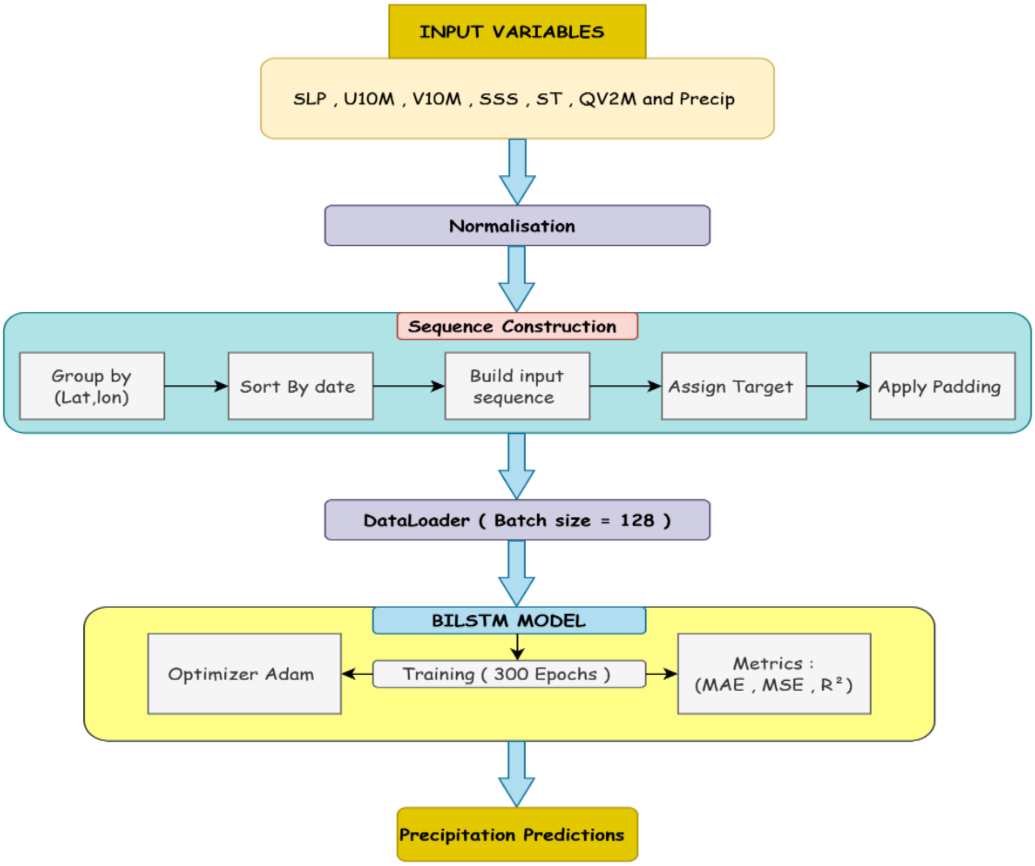


Fig. 3. Architecture of the *BiLSTM-MLP* predictor used for precipitation forecasting.

Each modification triggers a new forecast and a corresponding reward that balances three components: the precipitation gain, the energetic cost associated with the action, and a penalty term that limits abrupt or dynamically inconsistent variations. This formulation allows the agent to identify physically meaningful control strategies that enhance rainfall without violating atmospheric equilibrium. The integration of the predictive and optimization components is illustrated in (**Fig. 4**), which summarizes the TD3-based control framework coupled with *BiLSTM-MLP* forecasting.

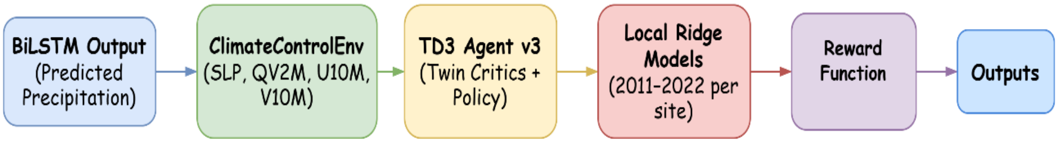


Fig. 4. Architecture of the TD3-based control framework coupled with BiLSTM-MLP forecasting.

3.2. Experimental Setup

All experiments were conducted on a GPU-enabled workstation to ensure efficient model training and evaluation. The entire workflow, including data harmonization, model training, and control optimization, was implemented in PyTorch and Stable-Baselines3 (Raffin et al., 2021). The BiLSTM predictor was trained for 300 epochs with a batch size of 128 using the Adam optimizer with a learning rate of 0.001 and an early-stopping criterion based on the validation root-mean-square error. The TD3 controller was subsequently trained for 75 000 time steps using Gaussian exploration noise and an experience-replay buffer to ensure stable policy learning. Each control episode consisted of eight consecutive steps, representing incremental adjustments of the same atmospheric configuration.

Evaluation was carried out across 3 659 spatial grid points covering the North-Eastern Atlantic domain under identical boundary conditions. The outputs included three complementary metrics that characterize both the physical and energetic aspects of the control process.

The precipitation gain (ΔP), defined as:

$$\Delta P = P_{\mathrm{TD3}} - P_{\mathrm{ref}} \tag{1}$$

where P_{TD3} is the precipitation predicted under control and P_{ref} is the baseline (uncontrolled) precipitation, quantifies the rainfall enhancement (mm) induced by the TD3 agent.

The Reward Efficiency Index (REI), expressed as:

$$\mathrm{REI} = \frac{\Delta P}{C+\varepsilon} \quad, where\ \ C = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|\Delta X_i|}{X_i}\right) \tag{2}$$

where Cdenotes the energetic cost of atmospheric perturbations applied to the variables $X_i \in \{\mathrm{SLP}, \mathrm{QV2M}, \mathrm{U10M}, \mathrm{V10M}\}$, and ε is a small regularization constant.

This ratio evaluates how effectively precipitation is increased relative to the energy required for control.

The Stability Index (SIS), computed as:

$$\mathrm{SIS} = 1 - \frac{\mathrm{Var}(\Delta X)}{\mathrm{Var}(X_{\mathrm{ref}})+\delta} \tag{3}$$

where $\mathrm{Var}(\Delta X)$ is the variance of the controlled perturbations applied by the TD3 controller, $\mathrm{Var}(X_{\mathrm{ref}})$ represents the variance of the unperturbed (reference) atmospheric fields, and δ is a small stabilizing term.

This index measures the smoothness and physical coherence of atmospheric adjustments. Values approaching 1 indicate stable and dynamically consistent control actions.

Together, these indicators allow both the magnitude and the physical plausibility of the reinforcement-based precipitation control to be quantitatively assessed. Spatial maps of dominant and blocking variables were also generated to highlight which atmospheric components exert the greatest influence on the optimized precipitation fields, providing a physically interpretable assessment of the regional control process.

4. RESULTS AND DISCUSSIONS

4.1. Predictive Model Evaluation

To evaluate the predictive capabilities of the proposed architectures, a set of experiments was performed using three deep learning models: *BiLSTM*, *MLP*, and *Transformer*. Each model was trained to forecast precipitation levels for March 2023 across all spatial grid points, using historical sequences of climate variables as input. The *BiLSTM* architecture employed a bidirectional recurrent structure followed by a dense output layer, while the *MLP* model relied on temporally averaged predictors. The *Transformer* model used self-attention mechanisms to capture long-range dependencies within the time series. All models were trained under identical configurations to ensure a fair and consistent comparison. The corresponding performance metrics, including the Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination R^2 are summarized in **(Table 1)**.

Table 1.

Evaluation metrics of precipitation prediction models for March 2023.

Model	MSE	MAE	RMSE	R^2
BiLSTM	0.000007	0.001881	0.002645	0.9927
MLP	0.000076	0.005817	0.008717	0.9173
Transformer	0.000053	0.005912	0.007280	0.9427

As shown in **(Table 1)**, the BiLSTM model consistently outperformed both the MLP and Transformer architectures across all evaluation criteria. It achieved the lowest MSE and MAE values, together with the highest coefficient of determination ($R^2 = 0.9927$), indicating an excellent fit between predicted and observed precipitation. The small RMSE confirms its robustness and generalization capability. In contrast, the MLP model exhibited the weakest performance, revealing a limited ability to capture temporal dependencies. The Transformer model performed better than the MLP but remained less accurate than the BiLSTM, highlighting the advantage of recurrent structures for sequential climate prediction. This superior performance arises from the bidirectional recurrent memory of the BiLSTM, which allows it to capture delayed and nonlinear dependencies among key atmospheric drivers such as SLP, humidity, and wind components. Unlike the MLP, which treats each month as an independent input, the BiLSTM integrates information across consecutive time steps, reproducing the temporal inertia and feedback loops that control rainfall formation. Compared with the Transformer, whose attention layers require large training samples and may overfit small regional datasets, the BiLSTM maintains temporal continuity and stable convergence with limited data. The relationship between observed and predicted precipitation values for each model is depicted in **Fig. 5**.

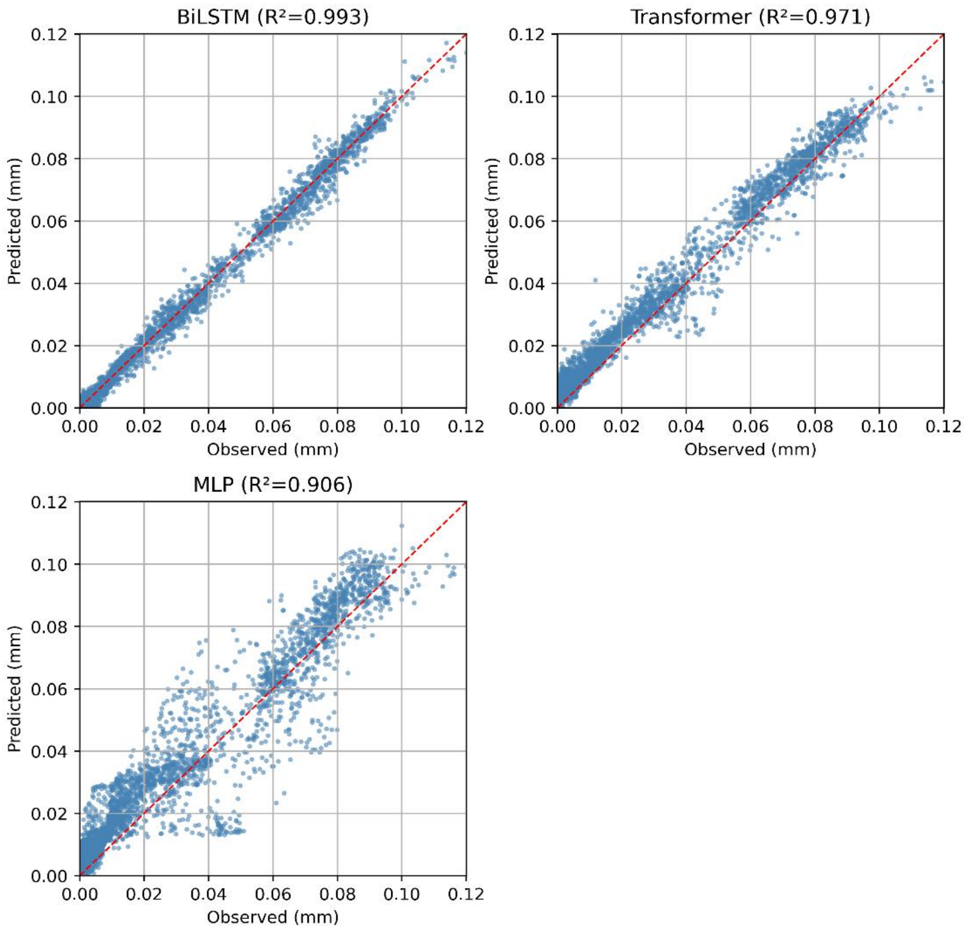


Fig. 5. Scatter plots comparing predicted and observed precipitation values for March 2023 using *MLP*, *BiLSTM*, and *Transformer* models.

The *BiLSTM* results show a nearly perfect alignment along the one-to-one diagonal, demonstrating both high predictive skill and low dispersion, whereas the *MLP* and *Transformer* display broader scatter, reflecting their reduced temporal sensitivity.

4.2. Reinforcement-Based Precipitation Control

After validating the predictive stage, the BiLSTM model was integrated within the TD3-based control framework to investigate adaptive precipitation modulation over 3,659 grid points across the North-Eastern Atlantic domain. The TD3 agent interacted with the predictive environment through multi-step episodes, applying small, continuous perturbations to the variables SLP, QV2M, U10M, and V10M while maximizing a reward function that combines precipitation gain, energetic cost, and physical stability. The optimization produced three key indicators: the precipitation variation (ΔP, in mm), the Reward Efficiency Index (REI), and the Stability Index (SIS), (**Table 2**). On average, the regional precipitation gain reached $\Delta P = 0.006$ mm per grid point, with 63 % of locations showing positive increments. The mean REI was 0.0025 and the average SIS 0.0004, confirming that most optimized states remained dynamically stable. These results demonstrate that the controller successfully enhances rainfall potential within physically admissible limits, ensuring no artificial amplification or instability.

Table 2.

Summary of TD3 control performance indicators.

Indicator	Definition / Meaning	Average Value	Interpretation
ΔP (mm)	Precipitation gain per grid cell	0.0061	Positive rainfall enhancement
REI	Reward Efficiency Index	0.0025	Efficient gain-cost trade-off
SIS	Stability Index	0.0004	Physically stable configuration
% ΔP > 0	Percentage of points with positive rainfall change	63.4 %	Most sites improved after optimization

Spatial patterns derived from these indicators revealed coherent regional structures. The highest ΔP and REI values occurred between 20° N and 35° N, particularly off the Moroccan coast, where humidity flux and pressure gradients are most sensitive to small perturbations. In contrast, northern latitudes exhibited weaker responses, consistent with their reduced coupling between low-level winds and moisture convergence. The SIS distribution confirmed that areas of strong ΔP were not associated with instability peaks, reinforcing the physical realism of the optimization.

Beyond these aggregate indicators, the contribution analysis identified the dominant and blocking variables governing local rainfall responses. As illustrated in (**Fig. 6**), each color represents the variable exerting the strongest control on precipitation within a given grid cell. Sea-level pressure (SLP) emerged as the dominant influence across most of the domain, particularly along the subtropical corridor, while specific humidity (QV2M) dominated near the coastal regions. Wind components (U10M and V10M) showed localized dominance over transition zones where synoptic advection drives rainfall variability. These spatial patterns correspond closely to known atmospheric mechanisms operating in the Atlantic sector, confirming that the learned control strategies remain physically interpretable.

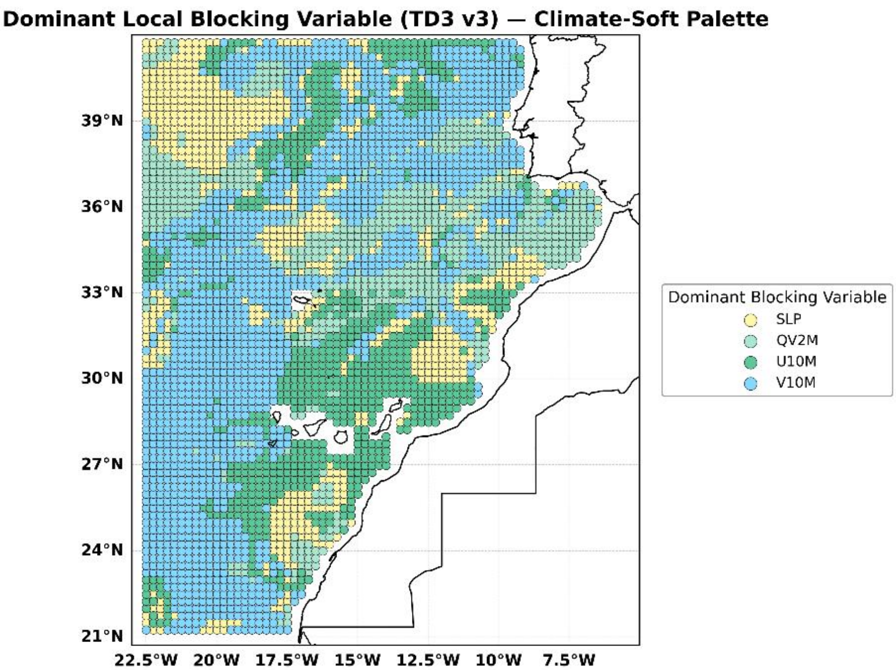


Fig. 6. Spatial distribution of dominant climatic variables influencing optimized precipitation across the study region.

The reinforcement-based results confirm that the hybrid *BiLSTM-TD3* framework achieves measurable precipitation enhancement while preserving atmospheric balance. The spatial coherence of ΔP, REI, and SIS, together with the dominance of physically meaningful variables, validates both the scientific realism and the operational potential of the proposed framework as a regional tool for hydro-climatic regulation.

4.3. Spatial manifestation of TD3-controlled precipitation strategies

Building on the reinforcement-based optimization results, the learned control strategies can be examined from a spatial and physical perspective to assess how the precipitation gains materialize across different climatic sub-regions of the North-Eastern Atlantic. Rather than focusing on aggregate indicators alone, this analysis investigates how the optimized atmospheric adjustments translate into geographically structured precipitation responses under real winter conditions.

A first-order spatial inspection reveals that precipitation enhancements are not uniformly distributed over the domain. Instead, precipitation variations exhibit coherent regional structures that closely follow the background hydro-climatic organization of the North-Eastern Atlantic. This spatial heterogeneity indicates that the reinforcement-based control does not impose homogeneous forcing, but rather interacts with the intrinsic sensitivity of the atmospheric system.

Along the Moroccan Atlantic coast, particularly during December 2018 and January 2020, precipitation gains appear more concentrated and spatially organized. These coastal areas are characterized by enhanced sensitivity to low-level moisture convergence and pressure gradients, allowing small atmospheric adjustments to produce measurable precipitation responses. Importantly, the enhanced precipitation patterns largely preserve the original spatial morphology, indicating that the control acts by reinforcing existing dynamical structures rather than generating artificial rainfall anomalies.

Moving offshore toward the central Atlantic, the precipitation response becomes weaker and more fragmented. In these regions, precipitation changes alternate between marginally positive and near-neutral values, reflecting a reduced coupling between surface-level atmospheric perturbations and precipitation processes. This behavior suggests that the reinforcement agent adapts its actions according to the local responsiveness of the atmospheric state, applying limited control where precipitation sensitivity is intrinsically low.

In the vicinity of the Azores region, precipitation changes remain consistently modest across all selected months. This result is physically consistent with the stabilizing influence of the Azores High, which constrains vertical motion and limits moisture uplift during winter conditions. The absence of strong precipitation enhancement in this area indicates that the control framework respects large-scale atmospheric constraints and does not attempt to override dominant synoptic regimes.

Further north, closer to the Iberian sector, precipitation responses show intermediate behavior, with localized gains appearing intermittently depending on the background circulation of each month. These spatial variations highlight the dependence of the control efficiency on transient synoptic configurations, reinforcing the idea that the learned strategies are state-dependent rather than spatially fixed.

This spatial manifestation of the learned control strategies is illustrated through a composite visualization of selected winter months (**Fig. 7**), comparing the observed precipitation fields before control, the TD3-enhanced precipitation after control, and the resulting absolute change (ΔP). While the Reward Efficiency Index (REI) and the Stability Index (SIS) provide a global assessment of control efficiency and physical coherence, the spatial analysis is primarily conducted using ΔP, as it directly reflects the geographical response of the atmospheric system to the applied control. The displayed patterns correspond to precipitation fields generated directly by the hybrid BiLSTM-TD3 framework, where the reinforcement agent operates through small, continuous perturbations of the atmospheric state while respecting physical stability constraints. The maps therefore represent the spatial realization of the learned control strategies rather than a post-processed or externally imposed adjustment.

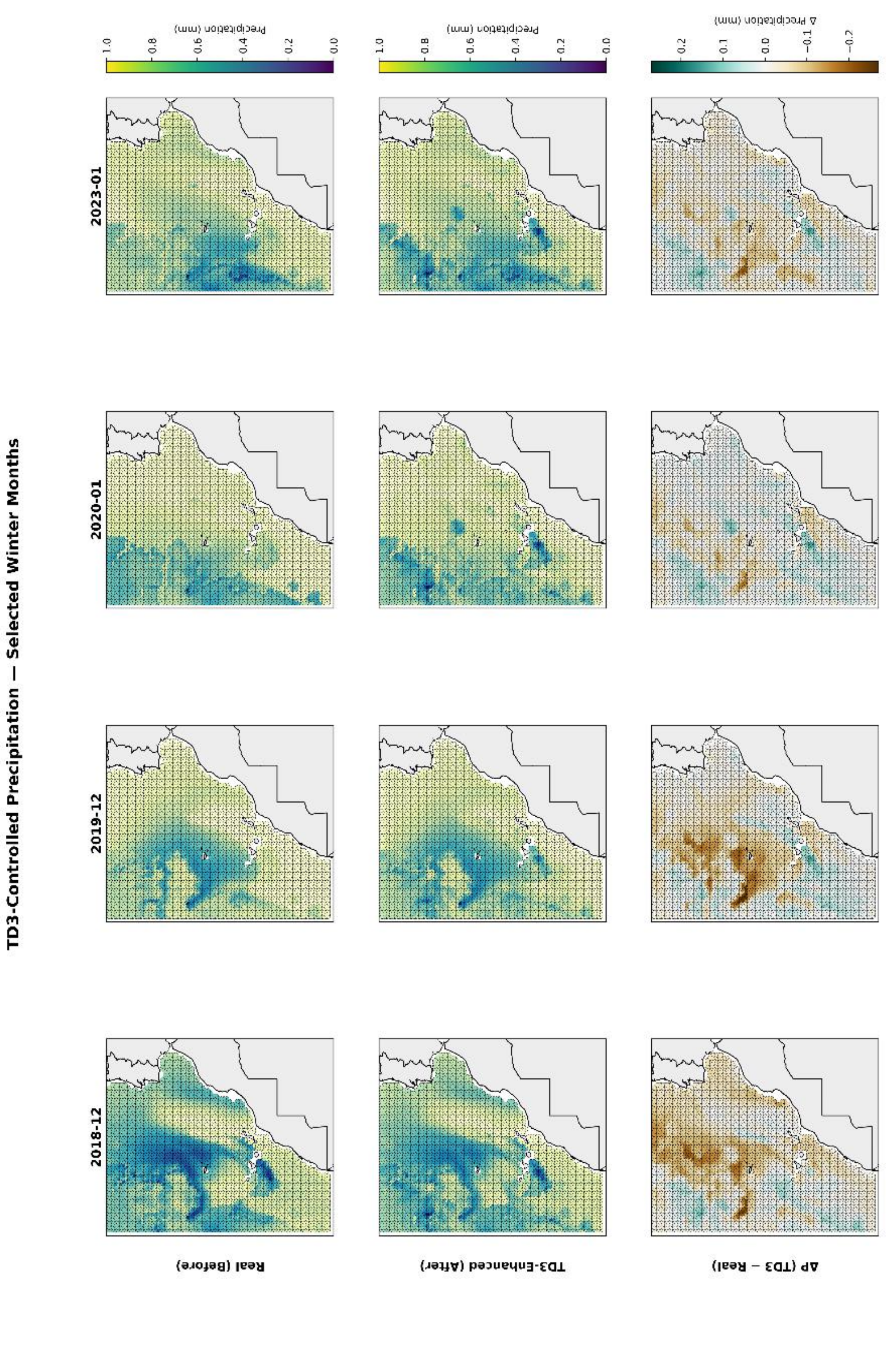


Fig. 7. Observed and TD3-controlled precipitation fields and associated changes (ΔP) for selected winter months from the dataset.

Overall, the spatial analysis confirms that the TD3-controlled precipitation strategies operate in a selective and physically consistent manner. Precipitation enhancement emerges primarily in regions where the atmospheric system exhibits sufficient dynamical sensitivity, while remaining limited in areas governed by strong stabilizing mechanisms. These results consolidate the physical interpretability of the proposed framework and confirm that the learned control strategies reinforce existing atmospheric dynamics rather than imposing artificial precipitation responses. From a broader perspective, this behavior highlights both the potential and the scope of applicability of the proposed framework. While the methodological structure of the BiLSTM-TD3 approach is generic and relies on physically meaningful variables common to many ocean-atmosphere systems, the spatially heterogeneous responses observed here indicate that effective application to other regions would require region-specific retraining and calibration to account for local climatic sensitivities and coupling mechanisms.

At the same time, it is important to emphasize that the present framework operates within a virtual predictive environment driven by reanalysis data and is not intended as a direct atmospheric actuation system. Rather, it should be viewed as a physically interpretable decision-support and scenario-exploration tool, designed to investigate how small, physically admissible perturbations could influence precipitation patterns under controlled conditions. These considerations define the current limitations of the approach while also outlining a clear pathway for future extensions toward broader hydro-climatic assessment and resilience-oriented applications.

5. CONCLUSION

The hybrid integration of deep learning and reinforcement learning provides a robust framework for physically consistent precipitation optimization along the Moroccan and Iberian coasts within the North-Eastern Atlantic. The BiLSTM predictor achieved highly accurate rainfall estimations, while the TD3 controller enhanced precipitation at more than 60% of grid points without compromising atmospheric stability. Spatial patterns revealed that sea-level pressure, humidity, and wind components exert the strongest control on rainfall modulation, confirming the physical realism of the learned strategies.

Beyond predictive performance, the proposed framework demonstrates that reinforcement learning can serve as a physically interpretable diagnostic tool to explore precipitation sensitivity under constrained atmospheric perturbations, rather than as an operational weather modification system. The spatial coherence of the optimized responses and stability indicators confirms that the control acts by modulating existing atmospheric structures while respecting large-scale dynamical constraints.

Although this study focuses on winter conditions over the North-Eastern Atlantic, the methodological structure is generic and can be extended to other coastal or drought-prone regions through region-specific retraining and calibration. The framework operates within a virtual environment driven by reanalysis data and should therefore be regarded as a decision-support and scenario-exploration tool. Future work will address multi-seasonal validation, higher-resolution datasets, and alternative reinforcement learning algorithms to further improve robustness and applicability.

ACKNOWLEDGEMENTS

The authors acknowledge the use of NASA Earthdata, ECCO, and MERRA-2 datasets that supported this research.

REFERENCES

Al Sayed K., Boodi A., Sadeghian Broujney R., et al (2024). Reinforcement learning for HVAC control in intelligent buildings: A technical and conceptual review. *Renewable and Sustainable Energy Reviews*, 187, 114268. https://doi.org/10.1016/j.rser.2024.114268

Durante Z., Huang Q., Wake N., et al (2024). Agent AI: Surveying the horizons of multimodal interaction. *arXiv preprint* arXiv:2401.03568. https://doi.org/10.48550/arXiv.2401.03568

Fukumori I., Wang O., Fenty I., et al (2021). ECCO Version 4 Release 4 (V4r4): A global ocean state estimate with improved representation of the Arctic and Southern Oceans. *Journal of Geophysical Research: Oceans*, 126(8), e2021JC017448. https://doi.org/10.1029/2021JC017448

Fujimoto S., van Hoof H., & Meger D. (2018). Addressing function approximation error in actor-critic methods. *arXiv preprint* arXiv:1802.09477. https://doi.org/10.48550/arXiv.1802.09477

Gelaro R., McCarty W., Suárez M. J., et al (2017). The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *Journal of Climate*, 30(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Graves A. & Schmidhuber J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*, 2047–2052. https://doi.org/10.1109/IJCNN.2005.1556215

Huffman G. J., Bolvin D. T., Joyce R., et al (2023). Integrated multi-satellite retrievals for GPM (IMERG), version 07B, final run. *NASA Goddard Earth Sciences Data and Information Services Center (GES DISC).* https://doi.org/10.5067/GPM/IMERG/3B-MONTH/07

Kurth T., Treichler S., Romero J., et al (2018). Exascale deep learning for climate analytics. In *SC18 International Conference for High Performance Computing, Networking, Storage and Analysis.* IEEE, 649–660. https://doi.org/10.1109/SC.2018.00054

Ladi T., Jabalameli S., & Sharifi A. (2022). Applications of machine learning and deep learning methods for climate change mitigation and adaptation. *Environment and Planning B*, 49(4), 1285–1300. https://doi.org/10.1177/23998083221085281

Li G., Yu Z., Li Y., et al (2024). Interaction mechanism of global multiple ocean–atmosphere coupled modes and their impacts on South and East Asian monsoon: A review. *Global and Planetary Change*, 233, 104438. https://doi.org/10.1016/j.gloplacha.2024.104438

Luo X., Nadiga B. T., Park J. H., et al (2022). A Bayesian deep learning approach to near-term climate prediction. *Journal of Advances in Modeling Earth Systems*, 14(10), e2022MS003058. https://doi.org/10.1029/2022MS003058

Meque A., Gamede S., Motlhobogi T., et al (2021). Numerical weather prediction and climate modelling: Challenges and opportunities for improving climate services delivery in southern Africa. *Climate Services*, 23, 100243. https://doi.org/10.1016/j.cliser.2021.100243

NOAA National Centers for Environmental Information (2025). *Global Climate Report – March 2025.* https://www.ncei.noaa.gov/access/monitoring/monthly-report/global/202503#precip

Ojo O., et al (2022). Climate–environmental analysis using deep Bayesian networks for SST prediction. *(Preprint / Journal of Advances in Modeling Earth Systems).* https://doi.org/10.1029/2022MS003058

Pathak J., et al (2022). FourCastNet: Global high-resolution weather forecasting using adaptive Fourier neural operators. *arXiv preprint* arXiv:2202.11214.

Raffin A., Hill A., Gleave A., Kanervisto A., & Dormann N. (2021). Stable-Baselines3: Reliable reinforcement learning implementations. *arXiv preprint* arXiv:2103.09575. https://github.com/DLR-RM/stable-baselines3

Rasp S., Pritchard M., & Gentine P. (2020). WeatherBench: A benchmark dataset for data-driven weather forecasting. *Geoscientific Model Development*, 13(2), 1073–1090. https://doi.org/10.5194/gmd-13-1073-2020

Routray A., et al (2023). Predictive control for smart HVAC and energy systems: A review. *Renewable and Sustainable Energy Reviews*, 187, 114268. https://doi.org/10.1016/j.rser.2024.114268

Rumelhart D. E., Hinton G. E., & Williams R. J. (1986). Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, 318–362. MIT Press.

Sierla S., Ihasalo H., & Vyatkin V. (2022). A review of reinforcement learning applications to control of heating, ventilation and air conditioning systems. *Energies*, 15(10), 3526. https://doi.org/10.3390/en15103526

Sleeman J., et al (2023). ACTD: Neurosymbolic architectures for early detection of tipping dynamics. *arXiv preprint* arXiv:2302.06852 [cs.AI]. https://arxiv.org/abs/2302.06852

Steininger M., Abel D., Ziegler K., & Krause A. (2022). ConvMOS: Climate Model Output Statistics with Deep Learning. *Data Mining and Knowledge Discovery*, 37(4), 1–31. https://doi.org/10.1007/s10618-022-00877-6

Sutton R. S. & Barto A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press. http://incompleteideas.net/book/the-book.html

Virtanen P., Gommers R., Oliphant T. E., et al (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Geographia Technica, Vol. 21, Special Issue: Artificial Intelligence Applications in Geography, 2026, pp 134 to 157

AN AI-BASED FRAMEWORK FOR ESTIMATING $PM_{2.5}$ USING SATELLITE AEROSOL OPTICAL DEPTH AND METEOROLOGICAL DATA IN A COASTAL INDUSTRIAL REGION OF THAILAND

***Maliwan THABTHONG* [1], *Teerawong LAOSUWAN* [1,4*], *Satith SANGPRADID* [2,4*], *Yannawut UTTARUK* [3,4], *Jenjira PIAMDEE* [1], *Piyatida AWICHIN* [1,4], *Titipong PHOOPHATHONG* [1,4] & *Maharaja SINGHARAJ* [5]**

DOI: 10.21163/GT_2026.212.07

ABSTRACT

Monitoring fine particulate matter ($PM_{2.5}$) in coastal industrial areas remains difficult, mainly because ground-based monitoring stations are sparse and unevenly distributed. In this study, we developed a practical AI-based framework to estimate daily $PM_{2.5}$ concentrations by combining satellite-derived Aerosol Optical Depth (AOD) with commonly available reanalysis-based meteorological data. The analysis focused on Rayong Province, Thailand, an industrial coastal region where atmospheric processes and pollution sources are highly variable. Daily $PM_{2.5}$ measurements from monitoring stations were merged with MODIS/MAIAC AOD and meteorological variables obtained from ERA5 and ERA5-Land. The meteorological inputs included air temperature, relative humidity, wind speed, and boundary layer height (BLH). Five machine learning models were tested, including Multiple Linear Regression (MLR), Support Vector Regression (SVR), Artificial Neural Networks (ANN), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). Feature selection was carried out using Recursive Feature Elimination with Cross-Validation and Mutual Information, and model performance was evaluated using five-fold cross-validation. Among the tested models, Random Forest showed the best overall performance, achieving an R^2 value of 0.70, followed by XGBoost and ANN. The results also indicate that nonlinear models consistently performed better than linear regression, suggesting that the relationship between $PM_{2.5}$ and atmospheric variables is not linear. Wind speed emerged as the most influential predictor, with relative humidity and BLH also playing important roles. In particular, the $PM_{2.5}$–meteorology relationships exhibited clear threshold behavior under low wind speed conditions. Overall, the proposed GeoAI-based framework provides a practical and transferable approach for spatial $PM_{2.5}$ estimation in complex coastal industrial environments, offering an effective solution for air quality assessment in regions with sparse ground monitoring networks.

Keywords: *$PM_{2.5}$ estimation; Aerosol Optical Depth; Machine learning; Spatial air pollution; Coastal industrial region.*

[1]*Department of Physics, Faculty of Science, Mahasarakham University, Maha Sarakham 44150, Thailand, 67010256001@msu.ac.th (MT), teerawong@msu.ac.th (TL), jenjira@msu.ac.th (JP), 68010262002@msu.ac.th (PA), 64010262003@msu.ac.th (TP).*

[2] *Department of Geoinformatics, Faculty of Informatics, Mahasarakham University, Maha Sarakham 44150, Thailand, satith.s@msu.ac.th (SS).*

[3] *Department of Biology, Faculty of Science, Mahasarakham University, Maha Sarakham 44150, Thailand, yannawut.u@msu.ac.th (YU).*

[4] *Greenhouse Gas Research and Operations Center, Mahasarakham University, Maha Sarakham 44150, Thailand.*

[5] *Nampapa Nakhoneluang, Vientiane Capital Water Supply State Enterprise, Kaisone Road, Xaysettha District, Vientiane City, Laos, mesha_coltd@yahoo.com (MS).*

**Corresponding authors: teerawong@msu.ac.th (T.L.), satith.s@msu.ac.th (S.S)*

1. INTRODUCTION

Air pollution is a global environmental problem that poses serious threats to human health, ecosystems, and economic sustainability (Laosuwan et al., 2023; Kovács & Haidu, 2024). Among various air pollutants, fine particulate matter with an aerodynamic diameter of less than 2.5 µm ($PM_{2.5}$) has been identified by the World Health Organization as one of the most harmful pollutants due to its ability to penetrate deep into the alveoli and enter the bloodstream (World Health Organization, 2021). Long-term exposure to $PM_{2.5}$ has been associated with chronic inflammation, respiratory diseases, cardiovascular disorders, and an increased risk of cancer (Sun et al., 2022). At the global scale, $PM_{2.5}$ exposure is recognized as a leading cause of premature mortality, contributing to millions of early deaths each year and imposing substantial burdens on public health systems and economic productivity (Caplin et al., 2019). In Thailand, $PM_{2.5}$ pollution has shown a worsening trend over recent years, particularly in urban centers and industrial regions such as Bangkok and the eastern part of the country (Rotjanakusol et al., 2024). Rayong Province represents a critical hotspot due to its concentration of petrochemical industries, power plants, and major transportation networks, which continuously contribute to elevated $PM_{2.5}$ levels (Kawichai et al., 2022). Seasonal variability is also pronounced, with higher $PM_{2.5}$ concentrations typically observed between November and March. This pattern is strongly influenced by unfavorable meteorological conditions, including low wind speed, frequent temperature inversion, and open biomass burning activities, as reported by the Pollution Control Department.

Despite increasingly strict air pollution control policies, including emission standards for industrial and vehicular sources, vehicle inspection programs, restrictions on open burning, and national action plans, fine particulate pollution remains persistent in many areas (Kovács & Haidu, 2021). This persistence highlights the ongoing challenge of effectively controlling $PM_{2.5}$ pollution (Archer et al., 2024). To improve air quality monitoring, government agencies have expanded ground-based monitoring networks. However, the limited spatial coverage of monitoring stations remains a major constraint, preventing comprehensive representation of air pollution conditions across heterogeneous landscapes (Rosales et al., 2025). Consequently, recent air quality research has increasingly emphasized the use of satellite-based observations, particularly Aerosol Optical Depth (AOD), as a proxy for aerosol loading in the atmospheric column. AOD has been shown to correlate with near-surface $PM_{2.5}$ concentrations under certain meteorological conditions (Sorek-Hamer et al., 2020). Numerous studies have demonstrated that integrating satellite-derived AOD with meteorological variables such as air temperature, relative humidity, wind speed, and boundary layer height (BLH) can substantially improve $PM_{2.5}$ estimation accuracy (Nakapan et al., 2022). When combined with statistical and machine learning models, including Multiple Linear Regression, Geographically Weighted Regression, Random Forest, and Extreme Gradient Boosting, these approaches are capable of capturing nonlinear relationships and atmospheric complexity (Wang et al., 2022; Li et al., 2025; Phoophathong et al., 2025).

In recent years, the integration of artificial intelligence with geographic data has become an effective approach for spatial impact analysis (Aroonsri & Sangpradid, 2021; Intarat et al., 2025; Intarat et al., 2026). Such geo-AI frameworks offer new opportunities to better understand spatial variability, seasonal dynamics, and meteorological controls of $PM_{2.5}$ particularly in complex coastal industrial ;environments where conventional monitoring remains insufficient (Wong et al., 2024). Although numerous studies have applied satellite-derived aerosol optical depth (AOD) and meteorological variables to estimate $PM_{2.5}$ concentrations (Schneider et al., 2020; Zaman et al., 2021), several important gaps remain. First, many existing studies focus on large metropolitan areas or inland regions, while coastal industrial environments receive comparatively less attention. Coastal areas are influenced by complex atmospheric processes, including land–sea interactions, variable boundary layer dynamics, and diverse industrial emission sources. These factors can weaken the AOD– $PM_{2.5}$ relationship and reduce model reliability if not explicitly considered (Tang et al., 2022).

Second, previous research often emphasizes prediction accuracy without sufficient discussion of spatial interpretation. In many cases, machine learning models are treated as black boxes, and their results are not fully linked to underlying geographical processes or atmospheric mechanisms (Zaman et al., 2021). This limits their value for understanding the spatial variability and seasonal behavior of $PM_{2.5}$, particularly in regions characterized by heterogeneous land use and meteorological conditions. Third, while machine learning techniques such as Random Forest and XGBoost have shown promising performance in $PM_{2.5}$ estimation (Chen & Guestrin, 2016), comparative evaluations across multiple models are still limited in Southeast Asian contexts. Most studies rely on a single modeling approach or short time periods, which restricts the generalizability of the findings. In addition, feature selection strategies and their influence on model performance are often underexplored (Li et al., 2016).

Finally, few studies explicitly frame $PM_{2.5}$ estimation as a geographical artificial intelligence (GeoAI) problem. Although the integration of artificial intelligence with spatial and atmospheric data has demonstrated strong potential for enhancing geographical understanding of air pollution patterns (Liu et al., 2022), this perspective remains underdeveloped in coastal industrial regions of Thailand. To address these gaps, this study aims to develop an AI-based geographical framework for estimating daily $PM_{2.5}$ concentrations in Rayong Province, Thailand. The specific objectives are to: (1) integrate satellite-derived AOD and meteorological variables for spatial $PM_{2.5}$ estimation; (2) compare the performance of multiple machine learning models; and (3) analyze the influence of key atmospheric variables on $PM_{2.5}$ variability in a coastal industrial setting.

2. MATERIALS AND METHODS

2.1. Study Area

Rayong Province (**Fig. 1**) is located in eastern Thailand. Its terrain consists of coastal plains along the Gulf of Thailand, combined with higher inland areas toward the interior of the province.

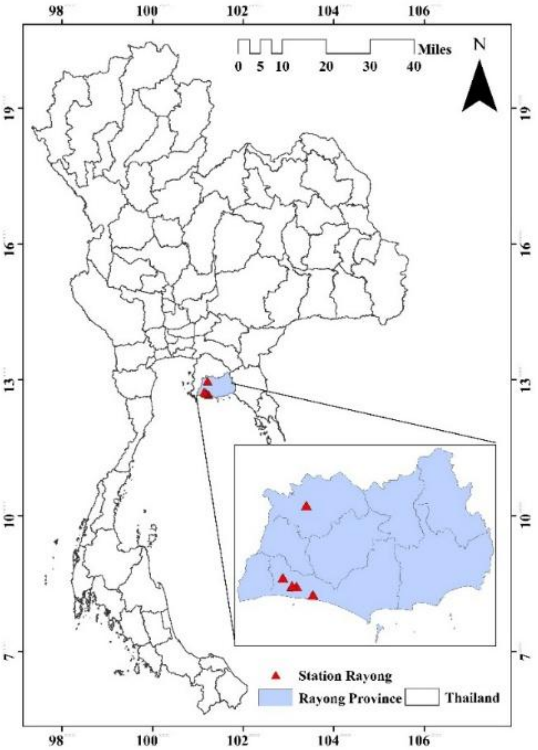


Fig. 1. Location of the study area in Rayong Province, Thailand, showing the five ground-based $PM_{2.5}$ monitoring stations and surrounding geographical features.

This geographical setting strongly influences local air circulation patterns, particularly the interaction between sea breezes and land breezes, which play an important role in the dispersion and accumulation of air pollutants. The regional climate is governed by the tropical monsoon system and can be divided into three main seasons: the hot season (March–May), the rainy season (June–October), and the cool season (November–February).

During the cool season, Rayong Province frequently experiences stable atmospheric conditions characterized by low wind speeds and nighttime temperature inversion. These conditions limit vertical air mixing and suppress pollutant dispersion, allowing fine particulate matter ($PM_{2.5}$) to accumulate near the surface. Statistical observations show that $PM_{2.5}$ concentrations are typically highest between December and February, which is associated with the southward extension of high-pressure systems from mainland China. Rayong Province is also home to the Map Ta Phut Industrial Estate and several surrounding industrial clusters. Emissions from petrochemical plants, power generation facilities, large-scale transportation networks, and community activities represent major sources of $PM_{2.5}$. When combined with coastal meteorological influences and seasonal atmospheric stability, these emission sources contribute to pronounced spatial and temporal variability in air quality across the province.

2.2. Data Sources and Modeling Framework

This study integrates ground-based air quality data with satellite observations and meteorological data to develop daily $PM_{2.5}$ prediction models for Rayong Province. The overall modeling framework is illustrated in **Fig. 2**. All satellite and meteorological variables were accessed and processed using the Google Earth Engine (GEE) platform to ensure consistency in spatial and temporal resolution. Satellite-based variables included Aerosol Optical Depth (AOD) and Land Surface Temperature (LST) derived from Terra/MODIS products.

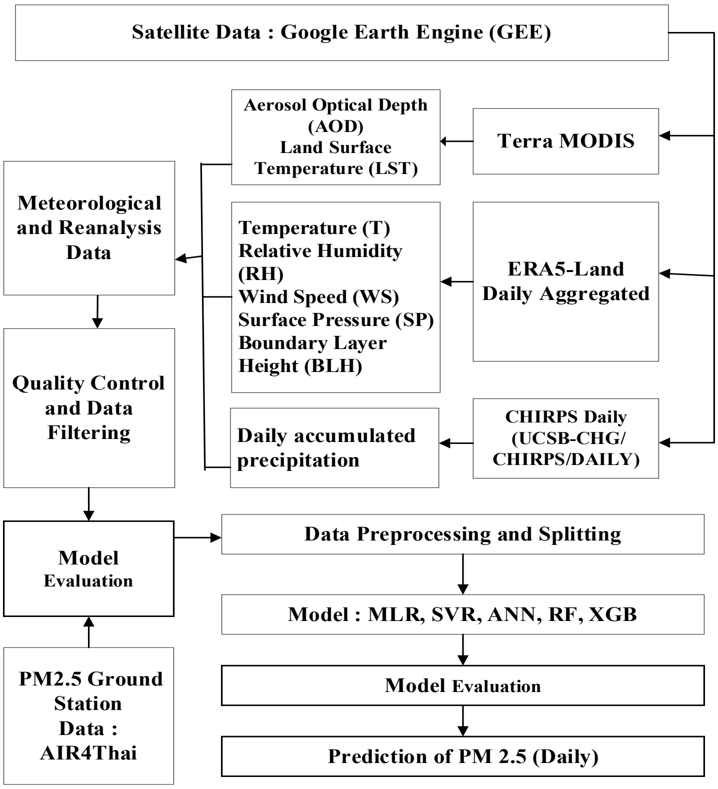


Fig. 2. Overview of the proposed framework for estimating daily $PM_{2.5}$ concentrations, including data acquisition, preprocessing, feature selection, model development, and validation.

Meteorological variables were obtained from the ERA5 and ERA5-Land datasets provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). These variables included air temperature, dew point temperature, relative humidity, wind speed, and boundary layer height (BLH), which are known to influence $PM_{2.5}$ formation and dispersion processes. All datasets were temporally matched and spatially processed to generate daily predictor variables. The processed satellite and meteorological variables were then used as independent inputs for statistical and machine learning models. The modeling approaches included simple linear regression, Multiple Linear Regression (MLR), and selected machine learning algorithms. Model performance was evaluated using standard error metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Based on the evaluation results, the best-performing model was selected to generate spatial distribution maps of daily $PM_{2.5}$ concentrations across Rayong Province.

2.2.1. Ground-Based $PM_{2.5}$ Monitoring Data

Daily PM2.5 concentration data were obtained from the Pollution Control Department (PCD) of Thailand. The measurements were conducted using the Beta Attenuation Method, following the standards of the United States Environmental Protection Agency (USEPA). Hourly and daily air quality data are publicly available through the AIR4Thai monitoring system (https://air4thai.pcd.go.th). In this study, daily $PM_{2.5}$ data were collected for the period from 1 November 2024 to 28 February 2025. The selected study period corresponds to the winter season in Thailand, during which $PM_{2.5}$ concentrations are typically elevated due to weak wind conditions, shallow boundary layer height, and frequent temperature inversions that limit atmospheric dispersion. Focusing on this season allows the models to capture pollution accumulation processes and evaluate performance under the most critical air quality conditions. A total of five ground-based monitoring stations located across Rayong Province were used for data collection (see **Fig. 2**). These stations included: (28t) Pluak Daeng District Public Health Office, (29t) Map Ta Phut Subdistrict Health Promoting Hospital, (30t) Rayong Provincial Agricultural Office, (31t) Rayong Field Crops Research Center, and (74t) Rayong Provincial Government Center. The selected monitoring stations represent different urban and industrial settings within the province. The ground-based $PM_{2.5}$ observations served as reference data for model development and performance evaluation.

2.2.2. Satellite and Meteorological Data

This study integrated satellite-derived variables with spatial meteorological data to support $PM_{2.5}$ estimation in a coastal region characterized by high variability in wind conditions and boundary layer height, such as Rayong Province.

Table 1.

Summary of satellite and meteorological datasets used in this study.

Variable	Data source (Google Earth Engine)	Unit	Spatial resolution	Temporal resolution
Aerosol Optical Depth (AOD)	MCD19A2.061: Terra & Aqua MAIAC Land Aerosol Optical Depth (Band: Optical_Depth_047)	Unitless	1 km	Daily
Land Surface Temperature (LST)	MOD11A1.061 Terra Land Surface Temperature and Emissivity (Band: LST_Day_1km)	Kelvin	1 km	Daily
2 m air temperature	ERA5-Land Daily Aggregated – ECMWF Climate Reanalysis (Band: temperature_2m)	Kelvin	0.1° × 0.1°	Daily
2 m dew point temperature	ERA5-Land Daily Aggregated – ECMWF Climate Reanalysis (Band: dewpoint_temperature_2m)	Kelvin	0.1° × 0.1°	Daily
U-component of wind at 10 m	ERA5-Land Daily Aggregated – ECMWF Climate Reanalysis (Band: u_component_of_wind_10m)	$m\ s^{-1}$	0.25° × 0.25°	Daily
V-component of wind at 10 m	ERA5-Land Daily Aggregated – ECMWF Climate Reanalysis (Band: v_component_of_wind_10m)	$m\ s^{-1}$	0.25° × 0.25°	Daily
Boundary layer height (BLH)	ERA5 Hourly – ECMWF Climate Reanalysis (Band: boundary_layer_height)	m	0.25° × 0.25°	Hourly

Meteorological variables included 2 m air temperature, 2 m dew point temperature, wind speed calculated from the u- and v-components, and boundary layer height (BLH). All satellite and meteorological datasets were extracted and processed using the Google Earth Engine (GEE) platform. This approach ensured spatial and temporal continuity of the data and helped reduce gaps associated with limited ground-based monitoring coverage. The max_pixels parameter was applied during data extraction to prevent processing errors related to large data requests. A summary of all variables and data sources used in this study is provided in **Table 1**.

2.2.3. Meteorological and Reanalysis Data

Ground-based $PM_{2.5}$ data, AOD, and LST from MODIS/MAIAC were combined with meteorological data from ERA5 and ERA5-Land. Spatial–temporal collocation was applied using the Nearest Neighbour method. All variables were converted to a daily resolution. Only days with complete data from all sources were used. This step helped reduce model uncertainty. Air temperature from ERA5 was originally provided in Kelvin (K). It was converted to degrees Celsius (°C) using Equation (1).

$$T_{(^{o}C)} = T_{(K)} - 273.15 \tag{1}$$

After data preprocessing, derived meteorological variables were calculated for model input. These included relative humidity (RH) and wind speed (WS). RH was calculated using the Magnus–Tetens equation, as shown in Equation (2):

$$RH(\%) = 100 \times \frac{e(T_d)}{e(T)} \tag{2}$$

in this equation, $e(T)$ and $e(T_d)$ represent vapor pressure in hPa

T is the air temperature at 2 m.

T_d is the dew point temperature at 2 m.

Wind speed (WS) was calculated from horizontal wind components using Equation (3):

$$WS = \sqrt{U^2 + V^2} \tag{3}$$

in this equation, WS is wind speed (m/s).

U is the west–east wind component (m/s).

V is the south–north wind component (m/s).

These derived variables are important for describing atmospheric processes. They influence the accumulation and dispersion of $PM_{2.5}$. They were used together with other meteorological variables. This improves spatial and temporal analysis accuracy.

2.3. Data Preprocessing and Quality Control

Before statistical analysis and model development, all satellite and meteorological datasets were preprocessed. All data were reprojected to the UTM Zone 48N coordinate system. A 10 km buffer was created around each $PM_{2.5}$ monitoring station. Daily values of each variable were extracted using median zonal statistics. This approach reduces spatial noise and outliers caused by sensor location uncertainty and terrain effects.

AOD data that did not meet MAIAC quality criteria were removed. Outliers were detected using the interquartile range (IQR) rule, and missing values, including NaN and infinite values, were excluded using Boolean masking. After data cleaning, the multivariate dataset included AOD, temperature, relative humidity (RH), wind speed (WS), and boundary layer height (BLH), with 468 samples remaining. The dataset was split into a training set (80%) and a test set (20%). All independent variables were standardized using the StandardScaler, with scaling parameters learned from the training set only to prevent data leakage.

2.4. Statistical Methods

Descriptive statistics were used to summarize the distribution and variability of $PM_{2.5}$ concentrations and meteorological variables during the study period. The analysis included measures of central tendency, data dispersion, and variability. The calculated statistics consisted of the mean, standard deviation, 95% confidence interval, minimum, maximum, and coefficient of variation (CV). These indicators are commonly used in environmental and atmospheric studies to describe air quality characteristics.

The mean value was calculated to represent the average concentration of each variable, as shown in Equation (4).

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4}$$

The standard deviation was used to describe data dispersion around the mean, as defined in Equation (5).

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(xi - \bar{x})^2} \tag{5}$$

The 95% confidence interval was computed to estimate the uncertainty of the mean value, following Equation (6).

$$CI_{95\%} = \bar{x} \pm t_{(1-\alpha/2,n-1)}\left(\frac{s}{\sqrt{n}}\right) \tag{6}$$

The minimum and maximum values were used to describe the data range, as shown in Equation (7).

$$x_{\min} = \min(x_i) x_{\max} = \max(x_i) \tag{7}$$

The coefficient of variation (CV) was calculated to assess relative variability, as defined in Equation (8).

$$CV(\%) = \frac{s}{\bar{x}} \times 100 \tag{8}$$

2.5. Theoretical Framework of PM2.5 Estimation Using Satellite and Meteorological Data

2.5.1. Atmospheric Processes Governing Near-Surface $PM_{2.5}$ Concentrations

Near-surface $PM_{2.5}$ concentrations result from interactions among emission sources, physical and chemical aerosol processes, and atmospheric dynamics in both horizontal and vertical directions. These processes are especially important in coastal industrial areas. Such regions are influenced by land–sea breeze circulation, wind speed variability, and changes in boundary layer height (BLH). From a theoretical perspective, near-surface $PM_{2.5}$ concentration can be described by the balance between emission rate (E) and atmospheric mixing volume. The mixing volume is mainly controlled by BLH and horizontal transport processes. This conceptual relationship can be expressed as shown in Equation (9):

$$C_{PM2.5} \alpha \frac{E}{BLH} \times f(WS) \tag{9}$$

Low BLH and low wind speed (WS) limit pollutant dilution. This leads to significant accumulation of $PM_{2.5}$ near the surface. This mechanism explains why the relationship between PM2.5 and meteorological variables is often nonlinear. It also indicates interaction effects among multiple variables acting simultaneously.

2.5.2. Physical Meaning of Aerosol Optical Depth (AOD) and Its Relation to $PM_{2.5}$

Aerosol Optical Depth (AOD) is a vertically integrated measure of the total aerosol loading in the atmospheric column. In physical terms, AOD represents the cumulative effect of the aerosol extinction coefficient along the vertical direction. This relationship can be expressed as shown in Equation (10):

$$AOD = \int_0^\infty \beta_{ext}(z)dz \tag{10}$$

Because $PM_{2.5}$ is a near-surface mass concentration, the relationship between AOD and $PM_{2.5}$ depends on the vertical structure of the atmosphere. It is also affected by boundary layer height, aerosol vertical distribution, and transport processes. Previous studies have shown that boundary layer height (BLH) plays a key role in linking columnar AOD to near-surface $PM_{2.5}$.This effect is especially strong in tropical and coastal regions.

2.5.3. Hygroscopic Growth and the Role of Relative Humidity (RH)

Relative humidity (RH) strongly influences aerosol optical properties and particle size through hygroscopic growth. Under high RH conditions, aerosol particles absorb water and increase in size. This process enhances light scattering and increases the extinction coefficient, even without a corresponding increase in $PM_{2.5}$ mass near the surface. The effect of hygroscopic growth can be described using a growth factor, as shown in Equation (11):

$$f(RH) = \frac{\beta_{ext}(RH)}{\beta_{ext}(dry)} \tag{11}$$

At high RH, aerosol water uptake increases particle size and scattering efficiency. As a result, AOD can increase without a proportional rise in near-surface $PM_{2.5}$ concentration. This physical mechanism has been widely used to explain $PM_{2.5}$–AOD relationships under humid atmospheric conditions. It is particularly relevant in Southeast Asia and East Asia. These findings support the physical basis of humidity-dependent statistical relationships between $PM_{2.5}$ and AOD.

2.5.4. Non-linear and Threshold Effects of Wind Speed on $PM_{2.5}$ Dispersion

Wind speed plays a key role in atmospheric pollutant dispersion. This concept is a fundamental part of atmospheric dispersion theory. However, many studies have shown that the relationship between $PM_{2.5}$ and wind speed (WS) is not linear. Clear threshold behavior has been widely reported. From a theoretical perspective, the response of $PM_{2.5}$ concentration to wind speed can be described as Equation (12):

$$\frac{\partial C_{pm2.5}}{\partial WS} < 0, \left| \frac{\partial^2 C_{pm2.5}}{\partial WS^2} \right| \neq 0 \tag{12}$$

2.5.5. Interaction Effects Between Wind Speed and Boundary Layer Height

Interaction effects between horizontal transport and vertical mixing are well explained by boundary-layer meteorology theory. Under conditions of low wind speed and low boundary layer height (BLH), atmospheric mixing volume is strongly limited. This leads to pollutant accumulation near the surface. When either WS or BLH increases, pollutant dispersion becomes more effective. As a result, near-surface $PM_{2.5}$ concentrations decrease significantly. These interaction effects have been widely examined using statistical approaches. Examples include two-way ANOVA and interaction modeling. Such methods are used to quantify the combined effects of meteorological variables on $PM_{2.5}$.

2.6. Machine Learning Models

This study developed and evaluated machine learning models to predict daily $PM_{2.5}$ concentrations in Rayong Province using an integrated dataset of ground-based observations and satellite–meteorological variables. The predictor variables included Aerosol Optical Depth (AOD), land surface temperature (LST), relative humidity (RH), wind speed (WS), and boundary layer height (BLH). The use of multiple models enables performance comparison under highly nonlinear conditions and strong spatial variability commonly observed in urban–industrial coastal environments. Previous studies have reported that nonlinear and ensemble models generally outperform linear models in $PM_{2.5}$ estimation. Based on this evidence, both linear and nonlinear models were evaluated in this study.

2.6.1. Data Preprocessing and Splitting

Data extracted from buffer-based zonal statistics and quality control procedures were converted into a numerical format suitable for machine learning analysis. All non-numeric values, missing values (NaN), and infinite values were removed. Predictor variables and the target variable ($PM_{2.5}$) were combined into a structured tabular dataset. The dataset was randomly divided into a training set (80%) and a test set (20%) using a fixed random seed to ensure reproducibility. For models sensitive to feature scaling, including Support Vector Regression (SVR) and Artificial Neural Networks (ANN), feature standardization was applied using a StandardScaler fitted on the training set only. The same transformation was then applied to the test set to prevent data leakage. These steps ensured that the dataset was fully prepared for model development.

2.6.2. Model Description

Multiple Linear Regression (MLR) was used as a baseline model to represent linear relationships between $PM_{2.5}$ and the explanatory variables, including AOD, LST, RH, WS, and BLH. Model coefficients were estimated using the Ordinary Least Squares (OLS) method. The MLR formulation can be expressed as Equation (13):

$$PM_{2.5} = \beta_0 + \sum_{i=1}^{5} \beta_i f_i + \varepsilon \tag{13}$$

where β_0 is the intercept, β_i represents the regression coefficient of each independent variable, f_i denotes the predictor variables, and ε is the random error term. Model reliability was assessed by examining multicollinearity using the Variance Inflation Factor (VIF) and by analyzing residual distributions to test assumptions of homoscedasticity, independence, and normality. Although MLR offers high interpretability, linear models often show lower predictive accuracy under strongly nonlinear conditions.

Support Vector Regression (SVR) was applied to model nonlinear relationships using the margin maximization principle and the ε-insensitive loss function. All input features were standardized prior to training. Key hyperparameters, including the penalty parameter *C,* kernel coefficient γ, and ε, were optimized using grid search combined with k-fold cross-validation. The radial basis function (RBF) kernel was used to capture complex nonlinear patterns.

Artificial Neural Networks (ANN), implemented as a Multilayer Perceptron (MLP), were used to model complex nonlinear interactions among satellite and meteorological variables. The network employed multiple hidden layers with ReLU activation functions to enhance nonlinearity. The model was trained using the Adam optimizer with Mean Squared Error (MSE) as the loss function. Early stopping was applied to reduce overfitting, and key hyperparameters were tuned to ensure stable convergence.

Random Forest (RF) is an ensemble tree-based model constructed using bootstrap aggregation and random feature selection. This structure makes the model robust to noise and effective in learning nonlinear relationships. RF does not require feature scaling and can provide feature importance measures, which are useful for interpreting the influence of individual variables on PM2.5 concentrations.

XGBoost is a gradient boosting model widely used in air quality and environmental studies. It incorporates both L1 and L2 regularization to reduce overfitting and is well suited for structured tabular data. XGBoost can effectively capture complex nonlinear patterns and interaction effects, often resulting in high predictive accuracy for $PM_{2.5}$ estimation.

2.6.3. Hyperparameter Tuning and Cross-Validation

Hyperparameter tuning is a critical step in machine learning model development, as model performance depends strongly on appropriate parameter selection. In this study, hyperparameters for each model were optimized using grid search combined with 5-fold cross-validation. This approach reduces bias associated with a single data split and improves model generalization performance. The optimal hyperparameter settings for each machine learning model are summarized in **Table 2**. The selected configurations reflect a balance between model complexity and predictive stability, particularly for nonlinear and ensemble models applied to atmospheric datasets with strong interaction and threshold effects.

Table 2.

Optimal hyperparameters for each machine learning model.

Model	Optimal Hyperparameters
Multiple Linear Regression (MLR)	No hyperparameter tuning performed
Support Vector Regression (SVR)	(C = 10), kernel = RBF
Artificial Neural Network (ANN)	Hidden layers = (100, 100), activation = ReLU
Random Forest (RF)	n_estimators = 400, max_depth = 30, min_samples_leaf = 1
XGBoost Regressor (XGBR)	learning rate = 0.01, max_depth = 5, n_estimators = 500

Note: Hyperparameter names follow standard XGBoost notation.

Table 2 presents the optimal hyperparameter configurations obtained through grid search with 5-fold cross-validation. The ANN achieved optimal performance with two hidden layers and ReLU activation, enabling effective learning of nonlinear relationships. SVR performed best with C = 10 using the RBF kernel, which supports flexible nonlinear function approximation. The ensemble tree-based models, RF and XGBoost, showed improved stability when a large number of trees were used. In particular, the low learning rate of XGBoost (0.01) helped reduce overfitting and allowed gradual model optimization.

2.7. Model Evaluation Metrics

Model performance in predicting daily $PM_{2.5}$ concentrations was evaluated using several quantitative metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). These metrics describe different aspects of prediction error between observed and predicted values. MAE measures the average magnitude of prediction errors and is less sensitive to large outliers. MSE and RMSE assign greater weight to large errors, making them more sensitive to extreme deviations. RMSE is particularly useful because it retains the same physical unit as $PM_{2.5}$. The R^2 metric represents the proportion of variance in observed $PM_{2.5}$ concentrations explained by the model. The definitions of these metrics are given as Equation (14)-(17).

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y - y_i| \tag{14}$$

Mean Squared Error; (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y - y_i)^2 \tag{15}$$

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y - y_i)^2} \tag{16}$$

Coefficient of Determination (R^2)

$$R^2 = \frac{(\hat{y}_i - \overline{y}_i)^2}{(y_i - \hat{y}_i)^2} = 1 - \frac{(y_i - \hat{y}_i)^2}{(\hat{y}_i - \overline{y}_i)^2} \tag{17}$$

where y_i represents the observed $PM_{2.5}$ concentration, $\hat{y}_i$ is the model-predicted value, $\overline{y}_i$ is the mean of the observed values, and n is the total number of observations.

3. RESULTS AND DISCUSSION

3.1. Descriptive Statistics and Seasonal Variability of $PM_{2.5}$

3.1.1. Temporal and Seasonal Variations

Fig. 3 shows daily $PM_{2.5}$ concentrations measured at individual monitoring stations in the study area. The data cover period from 1 November 2024 to 28 February 2025. Clear temporal variability is observed across all stations, indicating short-term fluctuations and differences among locations. These variations reflect the combined influence of local emission sources and meteorological conditions, consistent with previous findings (Bran et al., 2024). Several pollution episodes are evident during the study period. $PM_{2.5}$ concentrations frequently exceeded 60–80 µg/m^3, with peak values above 80 µg/m^3 at some stations. These high-concentration events mainly occurred during periods of low wind speed and stable atmospheric conditions, which limited pollutant dispersion. Seasonal differences are also apparent, with higher $PM_{2.5}$ levels generally observed during the dry season compared to the wet season.

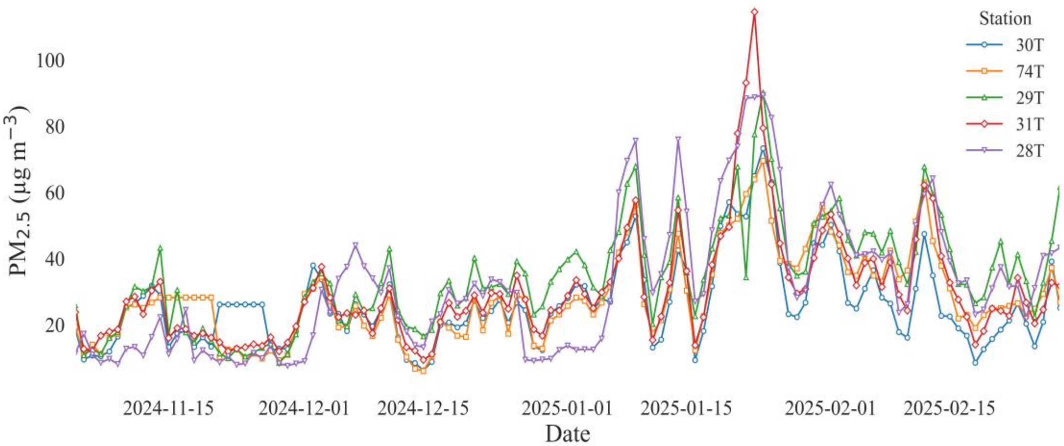


Fig. 3. Time series of observed daily $PM_{2.5}$ concentrations at the five monitoring stations during the study period.

Spatial differences among monitoring stations are clearly visible in **Fig. 4**. Some stations consistently recorded higher $PM_{2.5}$ concentrations than others, suggesting the influence of nearby emission sources and local land-use characteristics. In contrast, stations located farther from industrial areas showed relatively lower concentration levels. Overall, the temporal patterns indicate that $PM_{2.5}$ concentrations in the study area are strongly influenced by short-term meteorological variability as well as seasonal atmospheric conditions. These findings highlight the importance of incorporating meteorological and spatial factors into $PM_{2.5}$ estimation models, particularly in coastal industrial regions.

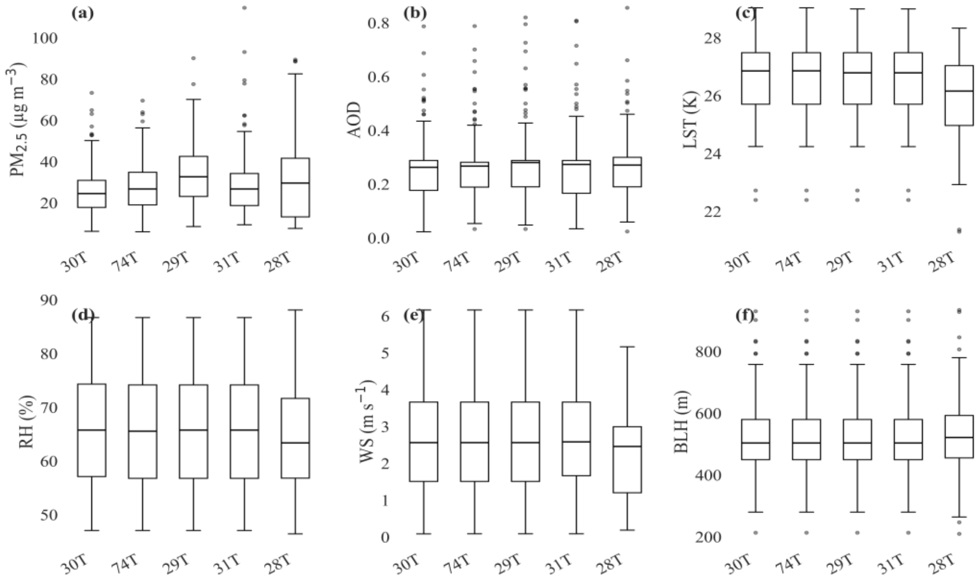


Fig. 4. Box plots showing the distribution of daily PM2.5 concentrations across the five monitoring stations during the study period.

3.1.2. Descriptive Statistics and Distribution Characteristics

Descriptive statistics and variability of $PM_{2.5}$ concentrations were evaluated together with satellite-derived and meteorological variables to characterize baseline pollution conditions during the winter season. **Table 3** summarizes the mean, standard deviation, 95% confidence interval, minimum, maximum, and coefficient of variation (CV) for all variables considered in this study. Overall, Table 3 highlights that $PM_{2.5}$ and AOD exhibit substantially higher variability than temperature, while relatively low mean wind speed and boundary layer height indicate meteorological conditions favorable for pollutant accumulation.

As shown in **Table 3**, the mean $PM_{2.5}$ concentration during the winter period was 30.05 μg/m^3 (95% CI: 28.73–31.37), which is substantially higher than the World Health Organization annual guideline, reflecting persistent air pollution in the study area. The high coefficient of variation (CV = 54.2%) further indicates strong day-to-day fluctuations and frequent episodic pollution events. Such variability is commonly associated with unfavorable meteorological conditions, including weak wind speed and shallow boundary layers, and is consistent with regional patterns reported in Thailand and Southeast Asia (Hassan Bran et al., 2024).

The mean AOD value was 0.27 with similarly high variability (CV = 50%), indicating strong fluctuations in aerosol loading within the atmospheric column. This variability is comparable to that observed for near-surface $PM_{2.5}$ and suggests a heterogeneous vertical aerosol distribution influenced by industrial emissions and regional transport processes. However, the large variability in AOD also highlights limitations in using satellite data alone as a direct proxy for surface $PM_{2.5}$, particularly under complex vertical atmospheric structures.

Table 3.

Descriptive statistics of $PM_{2.5}$ and explanatory variables during the study period.

Variables	Mean ± SD	95% CI	Minimum	Maximum	CV (%)
PM2.5 (µg/m³)	30.05 ± 16.29	[28.73, 31.37]	6	114.5	54.2
AOD	0.27 ± 0.14	[0.26, 0.28]	0.02	0.86	50
Temperature (°C)	26.48 ± 1.29	[26.38, 26.59]	21.3	29.04	4.9
WS (m/s)	2.53 ± 1.31	[2.42, 2.64]	0.08	6.16	51.7
RH (%)	65.63 ± 10.52	[64.77, 66.48]	46.43	88.12	16
BLH (m)	521.29 ± 122.89	[511.31,531.27]	209.44	934.21	23.6

In contrast, air temperature exhibited a narrow distribution (CV = 4.9%), reflecting the relatively stable winter climate in eastern Thailand controlled primarily by regional-scale air masses. This stability suggests that temperature plays a secondary role in explaining $PM_{2.5}$ variability compared with dynamic atmospheric factors. Wind speed (WS) and boundary layer height (BLH) showed relatively low mean values (2.53 m s^{-1} and 521 m, respectively) but high variability, indicating limited horizontal and vertical dispersion capacity. Under these stagnant conditions, both transport and mixing are suppressed, leading to enhanced accumulation of near-surface $PM_{2.5}$ concentrations.

In contrast, the background station (28T) displays a lower median and a narrower interquartile range, indicating weaker direct anthropogenic influence. However, the presence of outliers at this station suggests that $PM_{2.5}$ levels are occasionally dominated by regional-scale transport rather than local emissions. AOD shows similar median values across all stations, indicating that columnar aerosol loading is largely controlled by regional atmospheric processes rather than local land-use characteristics. Nevertheless, the presence of extreme AOD values at some stations reflects episodic high aerosol loading, potentially associated with transboundary haze events or aerosol accumulation under stagnant atmospheric conditions.

Air temperature exhibits narrow distributions and similar medians across all stations, consistent with the stable winter climate indicated in **Table 3**. Spatial temperature differences therefore play a limited role in explaining inter-station differences in $PM_{2.5}$. Relative humidity (RH) shows a wide distribution, particularly at coastal and mixed stations. High RH can enhance secondary aerosol formation and hygroscopic growth, potentially amplifying $PM_{2.5}$ concentrations during specific periods. Wind speed is generally low across stations, with occasional high-value outliers. This pattern indicates that atmospheric dispersion is limited for much of the winter season, especially in industrial and low-lying areas. boundary layer height (BLH) also exhibits a low median and wide variability. Days with boundary layer height (BLH) below approximately 400 m represent conditions of severely restricted vertical mixing and are frequently associated with elevated $PM_{2.5}$ concentrations across multiple stations.

3.1.3. Implications for Seasonal Variability and Subsequent Analysis

The analysis in Section 3.1 indicates that $PM_{2.5}$ concentrations in the study area are characterized by high variability, non-normal distributions, and strong spatial heterogeneity. Although temporal trends across all monitoring stations are generally consistent under the influence of regional-scale factors, the absolute concentration levels are clearly modulated by local land-use characteristics and station types. These features suggest that wintertime $PM_{2.5}$ variability cannot be fully explained by descriptive statistics alone. Instead, $PM_{2.5}$ variability is strongly associated with nonlinear relationships and interaction effects among $PM_{2.5}$, AOD, and meteorological variables. Therefore, Section 3.2 focuses on pairwise and conditional relationships among these variables to reveal underlying mechanisms that may be obscured by traditional linear analysis.

3.2. Pairwise Relationships and Non-linearity

Pairwise analysis between $PM_{2.5}$, AOD, and meteorological variables is an important step for understanding data structure and the underlying mechanisms of air pollution, particularly in coastal regions influenced by multidimensional atmospheric processes. In this study, pairwise scatter plots were combined with Pearson correlation coefficients to evaluate both the direction and strength of linear relationships. At the same time, this approach allows detection of nonlinear patterns, data clustering, and heteroscedasticity that cannot be captured by summary statistics alone.

To better interpret these nonlinear patterns, interaction effects should also be considered. Conceptually, interaction effects indicate that the influence of one variable depends on the level of another rather than acting independently. For example, wind speed becomes more effective at dispersing pollutants when the boundary layer is shallow, leading to nonlinear responses in $PM_{2.5}$ concentrations.

3.2.1. Correlation Analysis Between PM2.5 and Environmental Variables

The pairwise analysis integrates correlation coefficients with scatter plot patterns to assess linear relationships and to identify signals of non-linearity that may not be evident from correlation values alone. **Fig. 5** presents pairwise scatter plots and Pearson correlation coefficients between $PM_{2.5}$, AOD, and meteorological variables across the five monitoring stations in Rayong Province during 01 November 2024–28 February 2025.

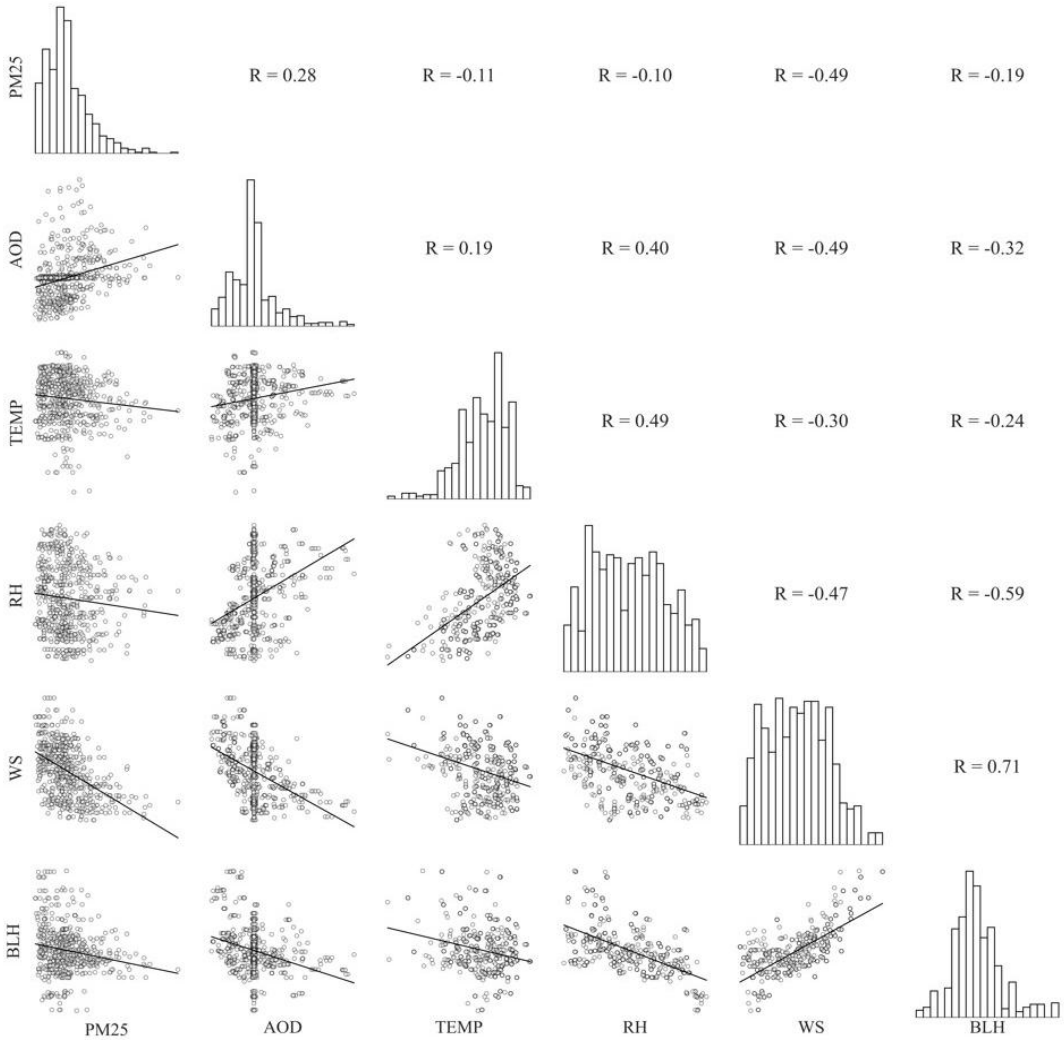


Fig. 5. Pairwise scatter plots illustrating the relationships between $PM_{2.5}$ and key meteorological variables, highlighting nonlinear patterns, threshold behavior, and conditional effects.

The results show that wind speed (WS) has the strongest negative correlation with $PM_{2.5}$ (r = −0.49). $PM_{2.5}$ concentrations decrease markedly as WS increases, highlighting the dominant role of horizontal transport and pollutant dilution in coastal environments. However, the scatter pattern is not strictly linear. High $PM_{2.5}$ values are strongly clustered under low WS conditions, indicating a clear threshold-type behavior.

Boundary layer height (BLH) shows a weaker negative correlation with $PM_{2.5}$ (r = −0.19) and a wide data spread. This suggests that vertical mixing plays a secondary role compared to horizontal transport, consistent with the stronger influence of WS observed in the analysis. AOD exhibits a moderate positive correlation with $PM_{2.5}$ (r = 0.28). However, the scatter plots show substantial variability, with many cases of high $PM_{2.5}$ occurring at moderate AOD levels. This pattern highlights the limitations of using AOD alone as a direct proxy for near-surface $PM_{2.5}$, especially in humid tropical and coastal regions where vertical aerosol stratification and moisture effects are pronounced.

Relative humidity (RH) and air temperature (TEMP) show very weak linear correlations with $PM_{2.5}$ ($r \approx -0.10$ to −0.11), with no clear linear trends in the scatter plots. These variables appear to act as modulating factors rather than direct drivers of PM2.5 concentrations. The non-normal distribution of $PM_{2.5}$ and the presence of numerous outliers further indicate strong nonlinearity in the relationships among $PM_{2.5}$, AOD, and meteorological variables. Spatial differences among monitoring stations further emphasize the influence of local emission sources and site-specific atmospheric processes. These spatial contrasts are consistent with previous studies conducted in Thailand and Southeast Asia.

3.2.2. Non-linear Relationships and Interaction Effects

To confirm and explain the nonlinear and conditional patterns observed in the scatter plots in **Fig. 5**, and to link these patterns with the physical framework described in Section 2.5, this study further examined $PM_{2.5}$ variability across different ranges of key meteorological variables. The analysis focused on threshold behavior and on conditional and interaction effects that cannot be captured by linear correlation coefficients alone.

(1) Threshold Effects: Wind Speed–$PM_{2.5}$ Relationship

Based on the conceptual framework described in Section 2.5.4, which emphasizes the nonlinear influence of horizontal transport on pollutant dispersion, a clear threshold-type relationship was identified between $PM_{2.5}$ concentration and wind speed (WS). Overall, **Table 4** demonstrates that $PM_{2.5}$ levels decrease markedly as wind speed increases, highlighting the dominant role of atmospheric ventilation in reducing near-surface pollutant accumulation.

Table 4.

$PM_{2.5}$ concentration stratified by wind speed.

Wind speed category (m s⁻¹)	Mean $PM_{2.5}$ (µg m⁻³)	Difference from baseline (µg m⁻³)
WS < 1.5	34.2	Reference
1.5 ≤ WS < 2.5	29.8	−4.4
2.5 ≤ WS < 3.5	25.1	−9.1
WS ≥ 3.5	22.6	−11.6

A distinct threshold was observed at approximately 2–2.5 m s^{-1} . $PM_{2.5}$ concentrations declined rapidly as wind speed (WS) increased from calm to moderate conditions, whereas further increases beyond about 3 m s^{-1} resulted in only marginal additional reductions. When wind speed (WS) < 1.5 m s^{-1}, pollutant accumulation was most pronounced, representing the baseline stagnant condition. As wind speed (WS) increased to 1.5–2.5 m s^{-1} and higher ranges, mean $PM_{2.5}$ concentrations decreased substantially; however, the rate of reduction progressively diminished at stronger wind speeds.

This nonlinear response indicates diminishing marginal dispersion effects under high ventilation conditions and confirms that the $PM_{2.5}$ – wind speed (WS) relationship cannot be adequately explained by a simple linear assumption. Instead, the results suggest that wind-driven transport exerts the strongest influence under low-wind regimes, where small increases in wind speed (WS) can produce large improvements in air quality.

(2) Conditional Relationships: Influence of Relative Humidity on the $PM_{2.5}$ –AOD Relationship to examine the hygroscopic growth mechanism described in Section 2.5.3, the relationship between $PM_{2.5}$ and AOD was evaluated under different relative humidity (RH) conditions. Overall, Table 6 shows that the strength of the $PM_{2.5}$ AOD association varies systematically with humidity, indicating that moisture plays an important conditional role in linking columnar aerosol loading to near-surface particle concentrations.

Correlation analysis reveals that the $PM_{2.5}$ AOD relationship is not constant across atmospheric states. The correlation coefficient increased under humid conditions, and these differences were statistically confirmed using Fishers Z-test ($p = 0.029$). As summarized in **Table 5**, the strongest association was observed when RH was below 60% ($r = 0.58$), while a statistically significant relationship remained when RH exceeded 70% ($r = 0.31$).

This pattern suggests that aerosol hygroscopic growth enhances the optical response of particles, improving the representativeness of AOD for estimating near-surface $PM_{2.5}$ under moist conditions. Consequently, the conditional influence of humidity helps explain why the overall linear correlation between $PM_{2.5}$ and AOD appears moderate when all observations are analyzed together without stratification.

Table 5.

$PM_{2.5}$–AOD Pearson correlations by relative humidity (95% CI).

RH Range (%)	PM2.5–AOD Correlation (r)	95% CI	p-value
< 60	0.58	[0.33, 0.75]	< 0.001
60–70	0.31	[−0.04, 0.60]	0.08
> 70	0.31	[0.01, 0.56]	0.04

(3) Interaction Effects between Wind Speed and Boundary Layer Height

Following boundary-layer meteorology theory described in Section 2.5.5, which emphasizes the combined effects of horizontal transport and vertical mixing, this study applied two-way ANOVA to examine interaction effects between wind speed (WS) and boundary layer height (BLH) on $PM_{2.5}$ concentrations. The analysis revealed a statistically significant interaction between wind speed (WS) and boundary layer height (BLH) ($F(1,236) = 8.7$, $p = 0.004$, partial $\eta^2 = 0.035$). $PM_{2.5}$ concentrations increased markedly when both horizontal transport and vertical mixing were simultaneously restricted. In contrast, $PM_{2.5}$ levels decreased significantly when at least one of the two processes favored pollutant dispersion. This interaction highlights the synergistic effects of atmospheric processes and reinforces the inherently nonlinear nature of $PM_{2.5}$ dynamics in coastal industrial environments.

Although the full dataset contains 468 daily observations, the effective sample size used in the two-way ANOVA is smaller. Prior to the analysis, wind speed (WS) and boundary layer height (BLH) were discretized into categorical groups to represent distinct atmospheric regimes. This binning procedure aggregates observations within each category and reduces the degrees of freedom in the statistical model. As a result, the reported degrees of freedom reflect the grouped structure of the data rather than the original number of observations. Such categorization is appropriate for interaction analysis, as it facilitates interpretation of joint effects between wind speed (WS) and boundary layer height (BLH) under physically meaningful atmospheric conditions.

3.3. Feature Selection Results

The results in Section 3.2 show that the relationships between $PM_{2.5}$ and meteorological variables are clearly nonlinear. Threshold behavior and interaction effects are evident for several variable pairs. These patterns agree with the atmospheric framework described in Section 2.5 and indicate that transport, vertical mixing, and aerosol physical–chemical processes cannot be explained by linear assumptions alone. For this reason, feature selection in this study does not rely on a single statistical method. A hybrid approach is applied. Recursive Feature Elimination with Cross-Validation (RFECV) is used to evaluate how the number of input variables affects model performance, while Mutual Information (MI) is used to quantify nonlinear dependencies and rank the relative importance of predictors. In practical terms, feature importance reflects how strongly each variable contributes to reducing prediction error, meaning that variables with higher importance values exert greater influence on $PM_{2.5}$ estimates. All analyses are based on the preprocessed dataset described in Section 3.2, including AOD, temperature, relative humidity (RH), wind speed (WS), and boundary layer height (BLH).

3.3.1. Recursive Feature Elimination with Cross-Validation (RFECV)

RFECV was applied to evaluate whether incorporating variables related to atmospheric transport, vertical mixing, and aerosol processes improves the ability of the models to explain $PM_{2.5}$ variability. The analysis employed 5-fold cross-validation, and model performance was assessed using the mean cross-validated coefficient of determination (CV R^2). Overall, Table 7 demonstrates that predictive performance consistently improves as additional meteorological variables are introduced, particularly for nonlinear and ensemble models.

As summarized in **Table 6**, the CV R^2 values of ANN, Random Forest, and XGBoost increase markedly with the number of input features, with the largest gains observed after including wind speed (WS) and boundary layer height (BLH). The highest performance is achieved when all five variables are used (k = 5), indicating that these models effectively capture complex nonlinear relationships and interaction effects among meteorological factors. This result is consistent with the wind speed (WS) and boundary layer height (BLH) interaction effects identified in Section 3.2.2. In contrast, the linear Multiple Linear Regression (MLR) model shows only marginal improvement as additional predictors are added. Even after incorporating temperature and relative humidity, performance gains remain limited, and the inclusion of WS and BLH results in only minor increases. These findings highlight the limitations of linear assumptions in representing the nonlinear and threshold behavior of $PM_{2.5}$ in coastal industrial environments and further justify the use of nonlinear machine learning approaches. Intuitively, this procedure identifies the smallest set of variables that still provides strong predictive performance, helping to avoid unnecessary complexity while retaining the most informative environmental drivers.

Table 6.

Cross-validated R^2 from RFECV for different feature sets and regression models.

k	CV R^2 (MLR)	CV R^2 (SVR)	CV R^2 (ANN)	CV R^2 (RF)	CV R^2 (XGBR)
1	0.2215	–0.04	0.12	0.10	0.08
2	0.3567	0.02	0.22	0.23	0.20
3	0.3521	0.11	0.28	0.31	0.28
4	0.3594	0.33	0.49	0.49	0.46
5	0.3873	0.38	0.51	0.54	0.51

3.3.2. Mutual Information (MI)

To further examine variable relevance under non-linear relationships, Mutual Information is used. MI measures how much information each variable provides about $PM_{2.5}$ variability. The MI ranking is RH (f3) > Temperature (f2) > WS (f4) > BLH (f5) > AOD (f1). This order is consistent with the physical mechanisms described in Section 2.5. RH and temperature show high MI values because they directly influence hygroscopic growth and secondary aerosol formation. WS and BLH mainly control atmospheric transport and vertical mixing. Their effects on $PM_{2.5}$ are more conditional than direct. Although AOD has the lowest MI value when considered alone, it still plays a complementary role when combined with meteorological variables. This behavior reflects the fact that AOD represents column-integrated aerosol loading rather than near-surface concentration.

3.3.3. Permutation Importance

Permutation importance is applied to confirm the robustness of the RFECV and MI results. This method is model-agnostic and evaluates performance changes after random shuffling of each variable. The results are summarized in **Table 7**.

Table 7.
Permutation and built-in feature importance for Random Forest and XGBoost models.

Variable	Meaning
f1	Aerosol Optical Depth (AOD)
f2	Temperature
f3	Relative Humidity (RH)
f4	Wind Speed (WS)
f5	Boundary Layer Height (BLH)

Note: Predictor variables are indexed as follows: f1 = AOD, f2 = temperature, f3 = relative humidity (RH), f4 = wind speed (WS), and f5 = boundary layer height (BLH). Feature importance values are shown for both permutation-based and built-in methods for Random Forest and XGBoost models.

Table 7 compares permutation-based and built-in feature importance for the Random Forest and XGBoost models. Relative humidity (f3) and temperature (f2) show consistently high importance across both methods, with low standard deviations, indicating stable and robust contributions to $PM_{2.5}$ estimation. In contrast, wind speed (f4) and boundary layer height (f5) exhibit larger differences between permutation and built-in importance, suggesting that their influence depends more strongly on model structure and interaction effects rather than direct linear contributions. Aerosol optical depth (f1) shows lower but relatively stable importance, supporting its role as a complementary predictor rather than a dominant driver.

3.4. Model Performance Evaluation

3.4.1. Comparison of Model Performance for $PM_{2.5}$ Estimation

To evaluate how well the models capture the nonlinear relationships, threshold behavior, and interaction effects described in Section 2.5 and confirmed in Sections 3.2–3.3, five regression approaches were compared, including Multiple Linear Regression (MLR), Support Vector Regression (SVR), Artificial Neural Networks (ANN), Random Forest (RF), and Extreme Gradient Boosting (XGB). Model performance was assessed using MAE, RMSE, and R^2 on the test dataset (20%). All models were trained after hyperparameter tuning with 5-fold cross-validation to ensure good generalization and reduce the risk of overfitting. The results are summarized in **Table 8.**

Table 8.
Performance comparison of regression models for $PM_{2.5}$ estimation.

Model	MAE	RMSE	R^2	ΔRMSE (%)	ΔR^2
MLR	8.46	11.40	0.50	–	–
SVR	7.90	10.83	0.55	4.96	+0.05
ANN	7.79	9.91	0.62	13.08	+0.12
RF	6.78	8.89	0.70	22.03	+0.20
XGB	7.02	9.44	0.66	17.13	+0.16

Overall, **Table 8** clearly demonstrates that nonlinear and ensemble models consistently outperform the linear MLR baseline, with Random Forest achieving the best overall performance across all evaluation metrics. As shown in **Table 8**, RF yields the lowest MAE and RMSE and the highest R^2 value (0.70), indicating that it explains approximately 70% of the variance in $PM_{2.5}$ concentrations on the test dataset. Compared with MLR, RF reduces RMSE by more than 22% and increases R^2 by nearly 0.20. These improvements are consistent with the nonlinear relationships and interaction effects between $PM_{2.5}$, wind speed, and boundary layer height discussed in Section 3.2, which cannot be fully captured by linear models.

Although the R^2 value of 0.70 indicates strong explanatory power, it should be interpreted within the context of the study area. Rayong Province is characterized by complex emission sources and coastal meteorological influences, which increase the difficulty of $PM_{2.5}$ prediction. Previous studies in urban and industrial environments with similar characteristics typically report R^2 values between 0.60 and 0.80 when combining satellite AOD and meteorological variables. Therefore, the performance achieved here is consistent with related research. When considered together with RMSE and cross-validation stability, Random Forest provides the best balance between accuracy and robustness, making it well suited for subsequent spatial $PM_{2.5}$ mapping applications.

From an applied perspective, atmospheric processes rarely follow simple proportional relationships; small changes in meteorological conditions can produce disproportionately large changes in pollution levels. Therefore, nonlinear models are better suited to capture these threshold and accumulation effects.

3.4.2. Analysis of Model Predictions and Errors

To evaluate the quality and reliability of $PM_{2.5}$ predictions, model outputs are compared with ground-based observations using time series and statistical analyses. This evaluation aims to assess overall model accuracy and to identify error patterns under different concentration levels and spatial contexts. Comparisons between observed and Random Forest (RF)–predicted $PM_{2.5}$ are shown in Fig. 6. **Fig. 6** presents time series comparisons of observed and RF-predicted daily $PM_{2.5}$ at the five monitoring stations. The RF model consistently captures temporal trends across all stations at both daily and weekly scales. Multi-day pollution episodes are also well reproduced, indicating that RF effectively learns non-linear relationships between $PM_{2.5}$, AOD, and meteorological variables. Quantitatively, predicted values show the closest agreement with the 1:1 line among all models. The regression slope is 0.89 with an intercept of 3.2 $\mu g/m^3$, indicating low overall bias and strong systematic accuracy. These results are consistent with the lowest RMSE and highest R^2 values obtained by RF on the test dataset (**Table 8**), confirming its overall predictive performance.

However, during high pollution events, RF systematically underestimates $PM_{2.5}$ concentrations. This behavior is most pronounced from January to early February, when atmospheric ventilation is limited in both horizontal and vertical directions. The underestimation appears as reduced variability in predicted values relative to observations and is consistent with the regression slope being lower than one. This limitation reflects the model's reduced ability to capture the full intensity of extreme pollution events, despite accurately representing general trends. Ensemble models combine many decision trees and average their predictions, which improves stability but can also smooth extreme values, making very high pollution episodes harder to reproduce precisely.

This underestimation may be attributed to several factors. First, satellite-derived AOD represents column-integrated aerosol loading and may have limited sensitivity to near-surface pollutant accumulation under shallow boundary layers, leading to weaker signals during severe episodes. Second, rapid and localized emission spikes from industrial or traffic sources may not be fully captured by daily averaged predictors. Finally, machine learning models, particularly ensemble approaches, tend to smooth extreme values during training, which can reduce their ability to reproduce rare high-concentration events. These limitations collectively contribute to the systematic underprediction observed during peak pollution conditions. Station-level analysis reveals clear spatial differences in prediction performance.

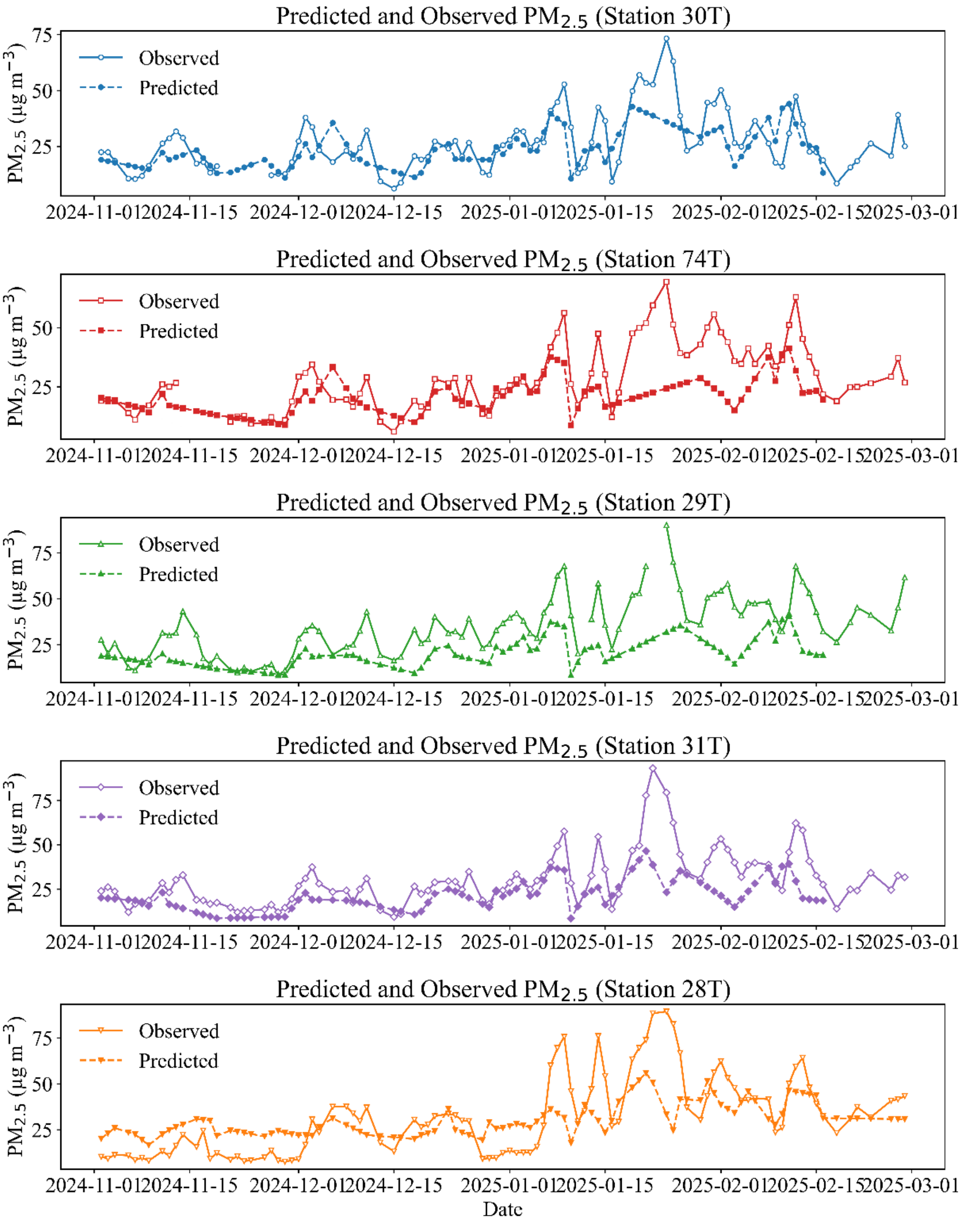


Fig. 6. Comparison between observed and Random Forest–predicted daily $PM_{2.5}$ concentrations at the five monitoring stations, illustrating model performance across different temporal conditions.

The background station (28T) shows the highest agreement between predicted and observed values, with accurate representation of baseline levels and temporal variability and relatively small errors during high concentrations. In contrast, the mixed station (31T) and the coastal station (29T) exhibit the largest discrepancies, particularly during rapid $PM_{2.5}$ increases. These patterns reflect the influence of diverse emission sources, localized transport, and rapidly changing atmospheric processes that are difficult to represent using regional-scale predictors. At the industrial stations (30T and 74T), RF also captures overall $PM_{2.5}$ trends well but continues to underestimate concentrations during severe pollution episodes, especially when $PM_{2.5}$ increases sharply over short periods. These results suggest that while AOD and meteorological variables effectively represent general accumulation conditions, they remain limited in capturing short-term emission spikes and sub-daily dynamics.

In summary, station-based comparisons indicate that the Random Forest model performs well in representing temporal trends and baseline $PM_{2.5}$ levels across space and time. Prediction errors show clear conditional structures related to pollution intensity and source complexity, with extreme events remaining the main source of uncertainty. These findings support the need for further residual analysis and focused evaluation of model performance during severe pollution events, which are addressed in the following section.

3.5. Spatial Distribution of Satellite-Derived $PM_{2.5}$

The spatial distribution maps of satellite-derived $PM_{2.5}$ from November to February reveal clear spatial and temporal variations in air pollution levels across the study area (**Fig. 7**). $PM_{2.5}$ concentrations show a gradual increase from early winter to mid-winter, indicating progressive accumulation of particulate matter during the study period.

Fig. 7 illustrates that in November, $PM_{2.5}$ concentrations range from approximately 17.92 to 32.35 µg/m^3 Most areas experience moderate air quality to levels that may begin to affect health. Lower concentrations are observed in parts of the southern and southwestern areas, corresponding to air quality conditions with minimal health impacts. In December, $PM_{2.5}$ concentrations increase noticeably, with values ranging from 15.48 to 39.65 µg/m^3. Large portions of the province fall within categories associated with moderate to high health impacts.

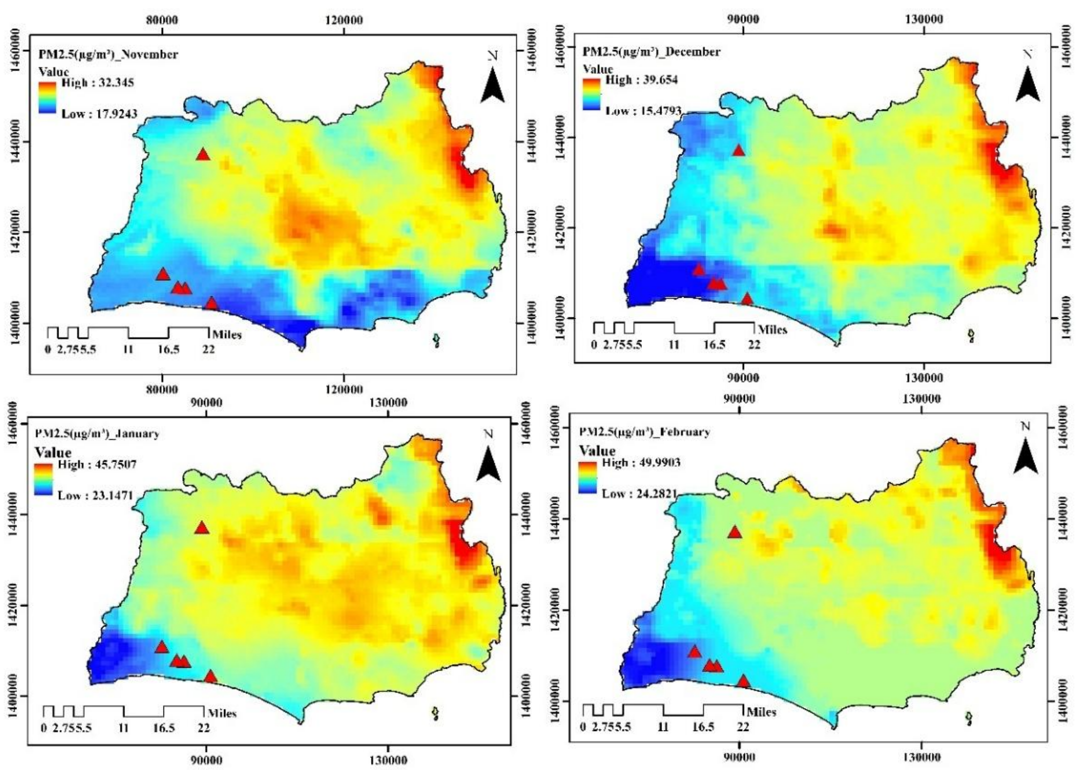


Fig. 7. Spatial distribution of satellite-derived $PM_{2.5}$ concentrations (µg/m^3) in the study area from November 2024 to February 2025, classified using fixed health-based thresholds.

Persistently elevated concentrations are observed in the eastern and northeastern areas, reflecting pollutant accumulation under meteorological conditions that limit atmospheric dispersion. $PM_{2.5}$ levels continue to rise in January, reaching a maximum of approximately 45.75 µg/m^3. Most of the study area is classified within high health impact levels, while some locations approach more severe risk categories. Pollution during this month covers a broader spatial extent than in previous months, indicating more widespread and intense air quality degradation during mid-winter.

In February, $PM_{2.5}$ concentrations reach their highest levels of the study period, ranging from about 24.28 to 49.99 µg/m^3. Some areas, particularly in the eastern part of the study area, approach or enter health-hazardous levels. These conditions indicate increased health risks for all population groups and highlight the need for intensified air quality monitoring and management. Overall, the satellite-based spatial analysis shows a consistent temporal increase in $PM_{2.5}$ concentrations and recurring pollution hotspots. The use of fixed health-based classification thresholds enables clear identification of health-vulnerable areas. These results demonstrate the strong potential of satellite-derived data to support spatial air quality assessment and health-oriented air pollution management at the local scale.

4. CONCLUSION

This study developed and evaluated daily $PM_{2.5}$ estimation models for Rayong Province by integrating ground-based monitoring data, MODIS/MAIAC aerosol optical depth (AOD), and meteorological variables from ERA5 and ERA5-Land. The results confirm that this integrated approach can effectively explain $PM_{2.5}$ variability in a coastal industrial area with complex atmospheric dynamics, particularly during the winter season when near-surface pollutant accumulation is favored. Model performance analysis shows that non-linear models, especially Random Forest and XGBoost, clearly outperform linear regression.

This finding indicates that relationships between $PM_{2.5}$ and satellite–meteorological variables are inherently non-linear and involve strong interaction effects. Wind speed plays the dominant role by controlling horizontal transport and dilution, while boundary layer height and relative humidity act as modifying factors. Clear threshold behavior is observed under calm wind conditions and shallow boundary layers. The results demonstrate that combining satellite data with machine learning not only improves spatial coverage beyond ground monitoring networks but also enables detailed mapping of $PM_{2.5}$ distributions. This approach is particularly useful for air quality assessment in coastal industrial regions with limited monitoring stations.

However, model uncertainty increases during severe pollution episodes and at very high $PM_{2.5}$ levels, highlighting the need to incorporate additional process-based information, such as spatial–temporal emission data and upper-atmospheric dynamics. Future work should extend the analysis to full-year periods, evaluate spatial robustness, and test model transferability to other coastal industrial regions. These steps would enhance the reliability and practical application of this framework for regional air quality management and environmental policy, especially in areas vulnerable to $PM_{2.5}$ pollution.

Beyond the study area, the proposed GeoAI framework is readily transferable readily transferable to other coastal or industrial regions with limited monitoring infrastructure, providing a practical tool for improving spatial air quality assessment and supporting evidence-based environmental management.

ACKNOWLEDGMENTS

This research project was financially supported by Mahasarakham University.

REFERENCES

Archer, D., Bhatpuria, D., Nikam, J., & Taneepanichskul, N. (2024). Particulate matter pollution in central Bangkok: Assessing outdoor workers' perceptions and exposure. *Cities & Health*, 1–19. https://doi.org/10.1080/23748834.2024.2390274

Aroonsri, I., & Sangpradid, S. (2021). Artificial neural networks for the classification of shrimp farm from satellite imagery. *Geographia Technica, 16*(2), 149–159. https://doi.org/10.21163/GT_2021.162.12

Bran, S. H., Macatangay, R., Chotamonsak, C., Chantara, S., & Surapipith, V. (2024). Understanding the seasonal dynamics of surface PM2.5 mass distribution and source contributions over Thailand. *Atmospheric Environment, 331*, 120613. https://doi.org/10.1016/j.atmosenv.2024.120613

Caplin, A., Ghandehari, M., Lim, C., Glimcher, P., & Thurston, G. (2019). Advancing environmental exposure assessment science to benefit society. *Nature Communications, 10*(1), 1236. https://doi.org/10.1038/s41467-019-09155-4

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Kovács, K. D., & Haidu, I. (2021). Effect of anti-COVID-19 measures on atmospheric pollutants correlated with the economies of medium-sized cities in 10 urban areas of Grand Est Region, France. *Sustainable Cities and Society*, 74, 103173. https://doi.org/10.1016/j.scs.2021.103173

Kawichai, S., Bootdee, S., Sillapapiromsuk, S., & Janta, R. (2022). Epidemiological study on health risk assessment of exposure to PM2.5-bound toxic metals in the industrial metropolitan of Rayong, Thailand. *Sustainability, 14*(22), 15368. https://doi.org/10.3390/su142215368

Kovács, K. D., & Haidu, I. (2024). Modeling NO2 air pollution variation during and after COVID-19 regulation using principal component analysis of satellite imagery. *Environmental Pollution*, 342, 122973. https://doi.org/10.1016/j.envpol.2023.122973

Laosuwan, T., Uttaruk, Y., & Rotjanakusol, T. (2023). Atmospheric environment monitoring in Thailand via satellite remote sensing: A case study of carbon dioxide. *Polish Journal of Environmental Studies*, 32(4), 3645–3651. https://doi.org/10.15244/pjoes/166170

Li, B., Chen, X., Zhang, W., Li, T., Xing, M., Yang, J., & Han, Z. (2025). Estimation of high-temporal-resolution PM2.5 concentration from 2019 to 2023 using an interpretable deep learning model. *Atmosphere, 16*(12), 1385. https://doi.org/10.3390/atmos16121385

Li, X., Peng, L., Hu, Y., Shao, J., & Chi, T. (2016). Deep learning architecture for air quality predictions. *Environmental Science and Pollution Research, 23*, 22408–22417. https://doi.org/10.1007/s11356-016-7812-9

Liu, Y., Wang, S., Li, T., & Hu, Y. (2022). Geographical artificial intelligence: A paradigm for spatial analysis. *Transactions in GIS, 26*(2), 638–656. https://doi.org/10.1111/tgis.12835

Lyapustin, A., Wang, Y., Korkin, S., & Huang, D. (2018). MODIS Collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques, 11*(10), 5741–5765. https://doi.org/10.5194/amt-11-5741-2018

Nakapan, S., & Hongthong, A. (2022). Applying surface reflectance to investigate the spatial and temporal distribution of PM2.5 in Northern Thailand. *ScienceAsia, 48*(1), 75–81. https://doi.org/10.2306/scienceasia1513-1874.2022.001

Ponsawansong, P., Prapamontol, T., Rerkasem, K., Chantara, S., Tantrakarnapa, K., Kawichai, S., Li, G., Fang, C., Pan, X., & Zhang, Y. (2023). Sources of PM2.5 oxidative potential during haze and non-haze seasons in Chiang Mai, Thailand. *Aerosol and Air Quality Research, 23*, 230030. https://doi.org/10.4209/aaqr.230030

Rosales, C. M., Bratburd, J. R., Diez, S., Duncan, S., Malings, C., & Pant, P. (2025). Open air quality data platforms for environmental health research and action. *Current Environmental Health Reports, 12*(1), Article 27. https://doi.org/10.1007/s40572-025-00487-6

Rotjanakusol, T., Puckdeevongs, A., & Laosuwan, T. (2024). Relationship assessment between PM10 from the air quality monitoring ground station and aerosol optical thickness. *Geographia Technica, 19*(1), 79–88. https://doi.org/10.21163/GT_2024.191.06

Schneider, R., Vicedo-Cabrera, A. M., Sera, F., Masselot, P., Stafoggia, M., de Hoogh, K., Kloog, I., Reis, S., Vieno, M., & Gasparrini, A. (2020). A satellite-based spatio-temporal machine learning model to reconstruct daily PM2.5 concentrations across Great Britain. *Remote Sensing, 12*(22), 3803. https://doi.org/10.3390/rs12223803

Sorek-Hamer, M., Chatfield, R., & Liu, Y. (2020). Review: Strategies for using satellite-based products in modeling PM2.5 and short-term pollution episodes. *Environment International, 144*, 106057. https://doi.org/10.1016/j.envint.2020.106057

Sun, J., Yan, Z., Zhang, M., & Chen, Y. (2022). Predisposed obesity and long-term metabolic diseases from maternal exposure to fine particulate matter (PM2.5): A review of its effect and potential mechanisms. *Life Sciences, 301*, 121054. https://doi.org/10.1016/j.lfs.2022.121054

Tang, Y., Yang, X., Yang, J., Cai, Z., Han, S., Shi, J., Jiang, M., & Qiu, Y. (2022). Investigation of coastal atmospheric boundary layer and particle by unmanned aerial vehicle under different land-sea temperature. *Aerosol and Air Quality Research, 22*, 220206. https://doi.org/10.4209/aaqr.220206

Wang, J., He, L., Lu, X., Zhou, L., Tang, H., Yan, Y., & Ma, W. (2022). A full-coverage estimation of PM2.5 concentrations using a hybrid XGBoost-WD model and WRF-simulated meteorological fields in the Yangtze River Delta Urban Agglomeration, China. *Environmental Research, 203*, 111799. https://doi.org/10.1016/j.envres.2021.111799

Wong, P.-Y., Su, H.-J., Lung, S.-C. C., Liu, W.-Y., Tseng, H.-T., Adamkiewicz, G., & Wu, C.-D. (2024). Explainable geospatial-artificial intelligence models for the estimation of PM2.5 concentration variation during commuting rush hours in Taiwan. *Environmental Pollution, 349*, 123974. https://doi.org/10.1016/j.envpol.2024.123974

World Health Organization. (2021, September 22). *WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. https://www.who.int/publications/i/item/9789240034228

Zaman, N. A. F. K., Kanniah, K. D., Kaskaoutis, D. G., & Latif, M. T. (2021). Evaluation of machine learning models for estimating PM2.5 concentrations across Malaysia. *Applied Sciences, 11*(16), 7326. https://doi.org/10.3390/app11167326

GEOSPATIAL DEEP LEARNING FOR BIOMASS AND CARBON STOCK ESTIMATION USING UAV-DERIVED CANOPY METRICS

Pimpilai WUNAPHAI [1], ***Teerawong LAOSUWAN*** [1,2*], ***Satith SANGPRADID*** [2,3*], ***Yannawut UTTARUK*** [2,4], ***Narueset PRASERTSRI*** [3], ***Thinnakon ANGKAHAD*** [1,2] ***& Piyatida AWICHIN*** [1,2]

DOI: 10.21163/GT_2026.212.08

ABSTRACT

Accurate estimation of biomass and carbon stock is important for forest monitoring and climate studies. This study developed a geospatial deep learning framework to estimate biomass and carbon stocks in rubber plantations by integrating UAV data with field measurements. The research was conducted in a rubber plantation area in northern Thailand, covering approximately 137 ha, with field data and UAV imagery collected during the 2024 growing season. UAV images were processed to generate Digital Surface Models (DSM), Digital Elevation Models (DEM), and Canopy Height Models (CHM). From these datasets, canopy structural variables such as canopy height, crown dimensions, and tree density were extracted and used as inputs for the model. Tree detection was performed using a YOLO-based deep learning approach, while field measurements were used for model training and validation. The integration of canopy structural information with deep learning improved biomass estimation at the plantation scale. The results showed strong agreement between predicted and field-measured values, with overall detection accuracy exceeding 92%. The model successfully captured spatial variation in biomass and carbon stocks across the plantation, and the estimated carbon values were consistent with plot-level observations. These findings indicate that combining UAV data, canopy structure information, and deep learning provides a practical and cost-effective approach for biomass and carbon estimation in plantation forests. The proposed workflow can also be applied to other plantation areas with similar environmental conditions.

***Keywords**: Geospatial deep learning; UAV remote sensing; canopy metrics; biomass estimation; carbon stock estimation; rubber plantation*

1. INTRODUCTION

Climate change is one of the most urgent environmental issues facing the world today. Rising temperatures, shifting precipitation patterns, and more frequent extreme weather events are largely the result of increasing greenhouse gas (GHG) concentrations, especially carbon dioxide (CO_2) released from fossil-fuel combustion, land-use change, and agricultural expansion (Haidu et al., 2024; Earthobservatory, 2025; Intarat et al., 2026). These changes disrupt ecosystem processes, threaten biodiversity, and undermine long-term environmental stability (Ngthai, 2025). Strengthening natural carbon sinks has therefore become a key strategy in global climate-mitigation efforts.

[1]*Department of Physics, Faculty of Science, Mahasarakham University, Maha Sarakham, Thailand. 67010287001@msu.ac.th (PW), teerawong@msu.ac.th (TL), thinnakon.ang@gmail.com (TA), 68010262002@msu.ac.th (PA).*

[2] *Greenhouse Gas Research Center and Operations, Mahasarakham University, Maha Sarakham 44150, Thailand. satith.s@msu.ac.th (SS), yanawut.u@msu.ac.th (YU).*

[3]*Department of Geoinformatics, Faculty of Informatics, Mahasarakham University, Maha Sarakham 44150, Thailand. narueset.p@msu.ac.th (NA).*

[4]*Department of Biology, Faculty of Science, Mahasarakham University, Maha Sarakham 44150, Thailand.*

**Corresponding author: teerawong@msu.ac.th (TL) and satith.s@msu.ac.th (SS).*

Thailand has aligned its national agenda with the Paris Agreement by committing to reduce GHG emissions by 30–40% by 2030 and by setting long-term goals of achieving “Carbon Neutrality” by 2050 and “Net-Zero GHG Emissions” by 2065 (United Nations Framework Convention on Climate Change, 2025). Initiatives such as the Thailand Voluntary Emission Reduction Program (T-VER) support these goals by promoting transparent and science-based carbon-accounting practices (Thailand Greenhouse Gas Management Organization, 2025). Accurate estimation of above-ground carbon (AGC), which is typically derived from above-ground biomass (AGB) using allometric models (Laosuwan et al., 2025; Uttaruk et al., 2026), is essential for national carbon inventories, climate-policy design, and sustainable land-use planning (Plybour et al., 2025; Laosuwan, et al., 2025).

Rubber (*Hevea brasiliensis* Muell. Arg.) is a major perennial crop in Thailand, covering more than 19 million rai and serving as a major source of income for rural communities (Simon et al., 2022). Rubber plantations grow rapidly, accumulate substantial biomass, and can function as effective carbon sinks, providing both environmental and economic benefits (Blagodatsky & Cadisch, 2016; Yang et al., 2017; Tiko et a., 2025). However, estimating carbon stocks in these plantations presents several challenges (Jirakajohnkool et al., 2025; Tai et al., 2025). Traditional field-based methods that rely on measurements such as diameter at breast height (DBH) and tree height are accurate but time-consuming, labor-intensive, and costly when applied over large or remote areas (Huynh et al., 2022). These limitations have encouraged the adoption of advanced remote-sensing technologies (Angkahad et al., 2024). Unmanned aerial vehicles (UAVs) offer high-resolution imagery suitable for generating Digital Surface Models (DSM), Digital Terrain Models (DTM), and Canopy Height Models (CHM) (Angkahad et al., 2025a). These spatial products allow for detailed characterization of tree structure and canopy conditions across large areas (Angkahad et al., 2025b). UAV-based surveys reduce field workload, allow repeated monitoring, and provide spatial detail that cannot be easily achieved through ground-based measurements alone (Ecke et al., 2022; Juan-Ovejero et al., 2023).

Recent advancements in artificial intelligence (AI) have further expanded the potential of UAV-based analysis (Haidu, 2024; Intarat et al., 2024). Machine learning (ML) and deep learning (DL) models have demonstrated strong performance in tasks such as tree detection, canopy delineation, and biomass estimation in forest and agricultural systems (Aroonsri & Sangpradid, 2021; Awichin et al., 2025; Espíndola et al., 2025; Intarat et al., 2025). Object-detection frameworks like the YOLO series can identify individual tree crowns from high-resolution imagery with high accuracy, improving the reliability of tree-level assessments (Beloiu et al., 2023; Mao et al., 2025; Que et al., 2026). When integrated with canopy metrics derived from UAV photogrammetry, these models offer a powerful tool for estimating carbon stocks at scale (Zhu et al., 2022; Juan-Ovejero et al., 2023; Yan et al., 2024; Jian et al., 2025). In addition, several studies have developed geospatial deep learning frameworks that integrate UAV-derived imagery, CHM-based structural variables, and deep learning–based tree detection to estimate above-ground carbon in rubber plantations. Field measurements serve as reference data, while UAV-extracted variables support the construction of predictive models (Angkahad et al., 2025a; Mao et al., 2025; Tai et al., 2025). The goal is to produce an accurate, efficient, and reproducible approach for plantation-scale carbon assessment. The results of this study contribute to the advancement of geospatial carbon-modeling techniques and provide practical insights that support sustainable rubber-plantation management and Thailand’s broader climate-mitigation objectives.

Most previous UAV-based biomass estimation studies mainly rely on statistical regression models using canopy height or vegetation indices as predictors. While these approaches provide useful estimates, they may not fully capture the structural complexity of plantation forests. In this study, we applied a different workflow by integrating UAV-derived canopy metrics with deep learning techniques for tree detection and biomass estimation. The use of a YOLO-based model allows automatic identification of individual tree crowns, which improves spatial detail compared to traditional approaches. In addition, combining field measurements with multiple canopy variables helps enhance model performance and reliability. This integrated approach provides a more flexible framework for biomass and carbon estimation in plantation environments.

2. MATERIALS AND METHODS

2.1. Study Area

The study was conducted in a rubber plantation located in Chan Chawa Subdistrict, Mae Chan District, Chiang Rai Province, covering approximately 137 ha in northern Thailand, as illustrated in **Fig. 1**. The site lies within the Mae Chan watershed at geographic coordinates near 20.236257° N and 99.972040° E. The terrain is characterized by gently undulating foothills, with elevations ranging from 410 to 480 meters above mean sea level and average slopes between 5% and 25%. The predominant soils are sandy loam to clay loam, featuring good drainage and moderate to high natural fertility, which support the long-term cultivation of perennial tree crops. The region is classified as a Tropical Monsoon Climate zone, consisting of three major seasons: summer, the rainy season, and winter. Annual mean temperatures range from 25°C to 30°C, while total annual rainfall varies between 1,700 and 1,800 millimetres. These warm and humid conditions promote vigorous rubber tree growth, resulting in substantial biomass accumulation. Such climatic and environmental characteristics make the area well suited for studies focused on above-ground biomass and carbon-stock assessment.

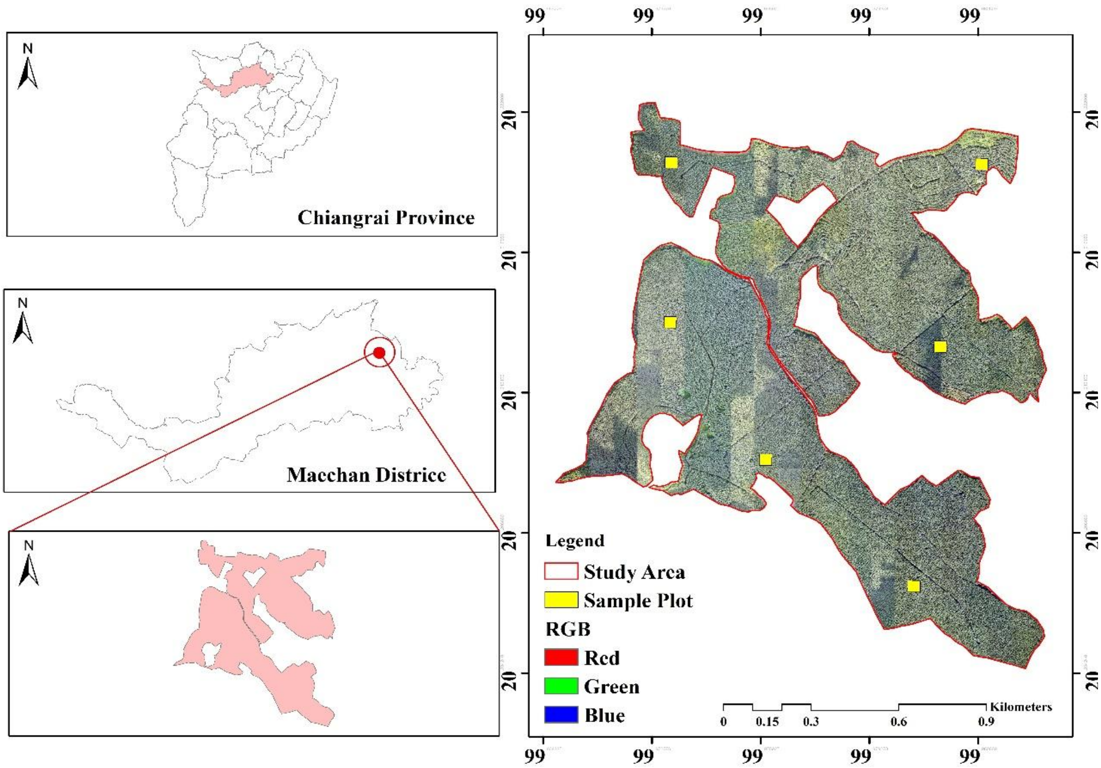


Fig. 1. Study area.

2.2. Methodology

The overall workflow of this study, illustrated in **Fig. 2**, integrates field measurements with UAV-derived datasets to support a structured approach for estimating above-ground carbon in rubber plantations. Field data collection provides core tree attributes, including DBH, height, and canopy dimensions, which are used to compute biomass and carbon through established allometric equations. These field-based estimates serve as reference data for evaluating the performance of the remote sensing and modeling components.

UAV photogrammetry represents the second major element of the workflow. High-resolution imagery is processed into point clouds, DSM, DEM, and orthophotos, from which the Canopy Height Model (CHM) is derived to capture vertical canopy structure. Machine learning and deep learning methods are then applied to identify individual trees, delineate crown features, and extract structural variables needed for biomass and carbon estimation. Model outputs are validated using regression and classification metrics to ensure accuracy and consistency, forming an integrated framework that links ground observations, spatial data, and predictive modeling. UAV photogrammetry represents the second major element of the workflow. High-resolution imagery is processed into point clouds, DSM, DEM, and orthophotos, from which the Canopy Height Model (CHM) is derived to capture vertical canopy structure. Machine learning and deep learning methods are then applied to identify individual trees, delineate crown features, and extract structural variables needed for biomass and carbon estimation. Model outputs are validated using regression and classification metrics to ensure accuracy and consistency, forming an integrated framework that links ground observations, spatial data, and predictive modeling.

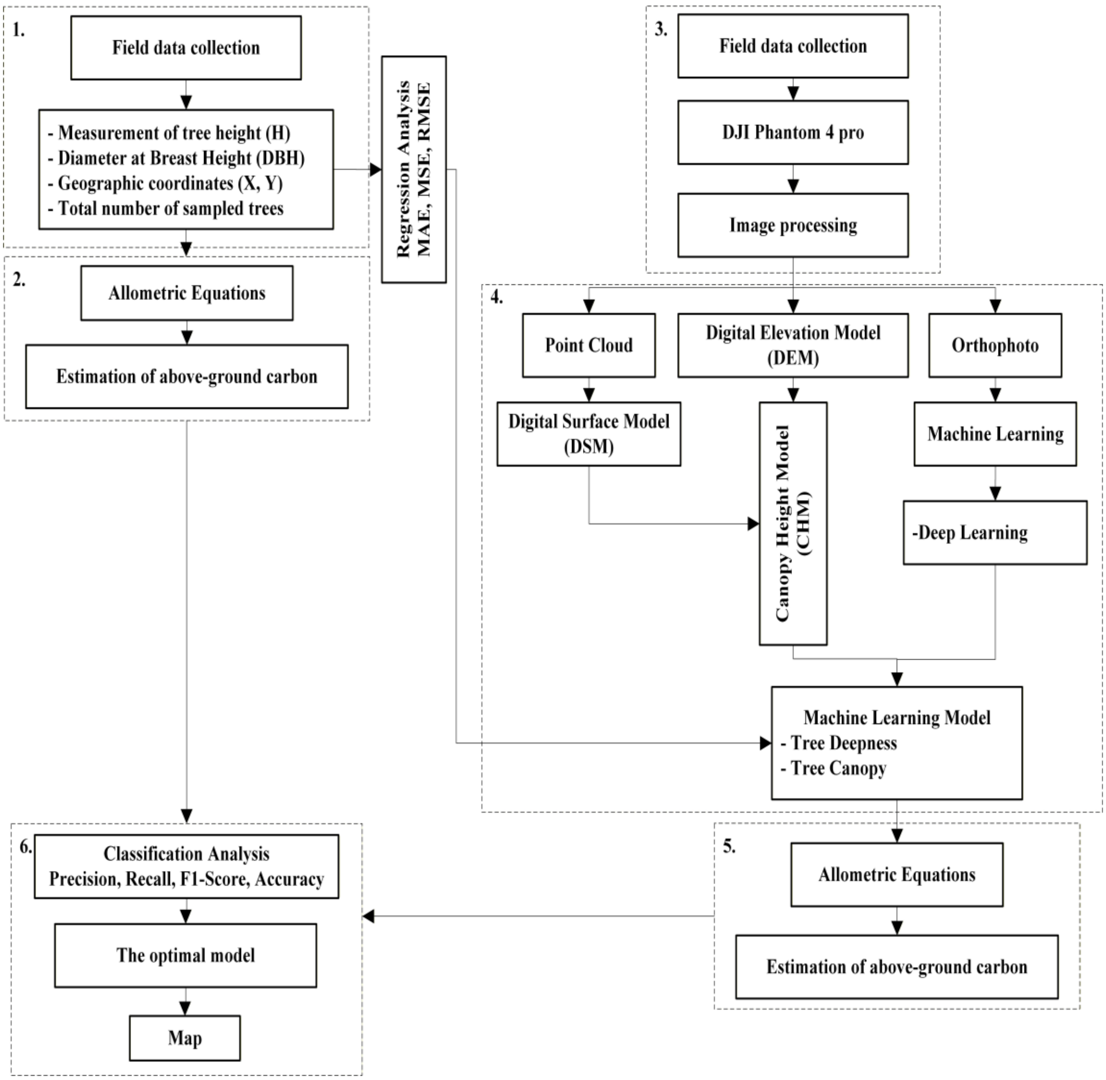


Fig. 2. Workflow of the integrated field and UAV-based deep learning methodology for above-ground carbon estimation in rubber plantations.

2.2.1. Field Survey and Data Acquisition

Field Data Collection step is represented by (1) in **Fig. 2**. Field data were collected from six 40 × 40 m sample plots established using a stratified random sampling approach to ensure representative stand conditions. Within each plot, key tree attributes were measured, including total height (H), diameter at breast height (DBH) at 1.30 m, and canopy dimensions in the north–south and east–west directions. These measurements were used to estimate above-ground biomass and served as reference data for validating the UAV-based models.

UAV Data Acquisition step is represented by (2) in **Fig. 2**. High-resolution aerial imagery was obtained using a DJI Phantom 4 Pro UAV equipped with a 20-megapixel camera and a 1-inch CMOS sensor capable of producing high-quality RGB images. The system includes an integrated GNSS module for accurate image geolocation (Angkahad et al., 2025b). Flight missions were conducted using Pix4Dcapture with a front overlap of at least 80% and a side overlap of 70% to ensure reliable image alignment. The average flight altitude ranged from 80 to 100 m above ground level. Captured imagery was processed in Agisoft PhotoScan Professional to produce essential photogrammetric outputs—including a dense point cloud, Digital Surface Model (DSM), Digital Elevation Model (DEM), and orthophoto—which served as the primary datasets for canopy-height derivation and tree-crown analysis.

2.2.2. Machine Learning–Based Data Analysis

The point cloud generated from UAV photogrammetry was used to construct two key spatial models: the Digital Elevation Model (DEM), representing the bare-earth surface without above-ground objects, and the Digital Surface Model (DSM), which includes the height of all surface features such as tree canopies and built structures. These models formed the foundational datasets for deriving canopy-height information across the study area.

The Canopy Height Model (CHM) was calculated to represent tree height above ground and served as a primary variable for estimating tree-level biomass. CHM was derived by subtracting the DEM from the DSM, where the DSM captures the maximum height of surface objects and the DEM represents the underlying terrain elevation (Angkahad et al., 2024). The CHM calculation is expressed in Equation (1) as follows:

$$CHM = DSM - DEM \tag{1}$$

where CHM is the canopy height model (m), DSM is the digital surface model (m), and DEM is the digital elevation model (m).

The selected canopy metrics were chosen because they represent key structural characteristics of plantation trees that are closely related to biomass accumulation and can be used as model inputs. Canopy height reflects tree growth and vertical structure, while crown size indicates canopy spread and overall tree volume. Tree density provides information about stand structure and competition among trees. These variables can be reliably derived from UAV-based photogrammetry and have been widely used in biomass estimation studies. Therefore, combining these canopy metrics provides meaningful input features for model development.

Deep learning techniques were applied to UAV-derived aerial imagery to analyze the structural characteristics of rubber trees and to perform automatic tree detection and counting. The YOLOv9 object-detection model was used to identify individual tree crowns from high-resolution RGB images captured by the UAV. These RGB images, composed of red, green, and blue spectral bands, served as input data for the model. The YOLOv9 framework processed each image to detect the location of rubber trees and generated bounding boxes around the identified tree crowns. The detection results were then exported into a Geographic Information System (GIS) environment to support spatial analysis. Using the Minimum Bounding Geometry (MBG) tool in ArcGIS Pro, the bounding boxes were converted into the smallest possible polygons that represent the spatial extent of each individual tree crown. This workflow enabled accurate and systematic identification and counting of rubber trees across the plantation. The resulting tree-crown polygons also served as a key spatial layer for subsequent biomass and carbon-stock estimation (Angkahad et al., 2024). The model hyperparameters were selected based on preliminary testing and commonly recommended settings from previous studies. Parameters such as learning rate, batch size, and number of training epochs were adjusted to achieve stable training performance and avoid overfitting. The final configuration was chosen based on the best validation accuracy obtained during the training process:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \tag{2}$$

where y is the estimated diameter at breast height (DBH, cm) derived from canopy structural characteristics; β_0 is the intercept; β_1, β_2, and β_3 are regression coefficients; x_1 is the canopy dimension in the north–south (NS) direction; x_2 is the canopy dimension in the east–west (EW) direction; x_3 is tree height derived from canopy structure; and ε is the error term.

Note: Calculation formulas used: $y = (29.60x_1) + (-0.88x_2) + (3.01x_3) + (-0.129.53)$

2.2.3. Above-ground Carbon Estimation in Rubber Plantations

Above-ground carbon (AGC) in the rubber plantation was estimated using the allometric equations proposed by Hytönen et al. (2018), as presented in Equations (3–6).

$$M_i = W_L + W_{LA} + W_{STUMP} \tag{3}$$

$$W_L = 0.00193 \cdot DBH^{2.499} \tag{4}$$

$$W_{LA} = 0.05155 \cdot DBH^{2.783} \tag{5}$$

$$W_{STUMP} = 0.02440 \cdot DBH^{2.470} \tag{6}$$

where DBH is the diameter at breast height measured at approximately 1.3 m above ground (cm); M_i is the total above-ground biomass of an individual tree (kg); W_L is the above-ground biomass of leaves (kg); W_{LA} is the above-ground biomass of stems and branches (kg); and W_{STUMP} is the above-ground biomass of the stump and coarse roots (kg).

$$C_{ABG} = M_i \times 0.47 \tag{7}$$

Carbon stocks were expressed in units of tCO_2. Above-ground biomass was first converted to carbon mass using a carbon fraction (CF) of 0.47, and the resulting carbon mass was then converted to CO_2 by multiplying by the molecular weight ratio of CO_2 to carbon (44/12).

2.2.4. Model Accuracy Assessment

Assessing the accuracy of the predictive model is essential to ensure that its outputs are reliable and consistent with field measurements. In this study, model accuracy was evaluated at two main levels.

The regression-based accuracy assessment quantifies how far the predicted values deviate from actual observations using standard statistical metrics. These indicators reflect both the size of prediction errors and the model' s ability to explain variation in the dataset. The primary metrics used were Mean Absolute Error (MAE), Mean Squared Error (MSE), and the coefficient of determination (R^2), calculated using Equations 8–10.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Actual_i - \Pr edicted_i\right| = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_i\right| \quad (8)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i\right)^2 \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i)^2}{\left(y_i - \overline{y}\right)^2} \quad (10)$$

where MAE is the mean absolute error; MSE is the mean squared error; R^2 is the coefficient of determination; *n* is the total number of samples; y_i is the observed value (field-measured AGB or carbon stock); y_i is the value predicted by the model; and $\overline{y}$ is the mean of the observed values.

2.2.5. Classification-based accuracy assessment

To evaluate the model performance in assigning rubber trees to specific diameter classes, standard classification metrics were applied. This assessment examines how well the model identifies and categorizes individual trees. The metrics used include Precision, Recall, F1-score, and overall Accuracy. Together, they offer a detailed understanding of the model's classification behavior and error patterns. These metrics were computed following Equations (11)– (14), based on established classification theory (Mohan et al., 2017).

$$\Pr ecision = \frac{TP}{TP + FP} \quad (11)$$

$$\mathrm{Re}call = \frac{TP}{TP + FN} \quad (12)$$

$$F1 - Score = 2 \times \frac{\Pr ecision \times \mathrm{Re}call}{\Pr ecision + \mathrm{Re}call} \quad (13)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (14)$$

where TP (true positive) is the number of trees correctly identified; FP (false positive) is the number of trees incorrectly identified; FN (false negative) is the number of trees that the model failed to identify; and TN (true negative) is the number of trees correctly classified as non-target cases.

3. RESULTS

3.1. Field Data Analysis Results

Field data were collected from six 40 × 40 m sample plots within the rubber plantation in Chan Chawa Subdistrict, Mae Chan District, Chiang Rai Province. Plot boundaries were established using low-error GPS at the four corners and plot center to ensure accurate spatial alignment with UAV-derived datasets. Within each plot, biometric and structural attributes of all rubber trees were recorded, including tree height (H), diameter at breast height (DBH) at 1.30 m, and canopy spread along the north–south (NS) and east–west (EW) directions. In total, 474 trees were measured. The trees had an average height of about 21 m and an average DBH of 22 cm. Mean canopy widths were 5 m (NS) and 9 m (EW), showing typical variation in crown structure found in mature rubber stands. These field measurements were used to estimate above-ground biomass (AGB) with established allometric equations. Biomass values were then converted to above-ground carbon (AGC) using a carbon factor of 0.47. A summary of plot-level biomass and carbon estimates is presented in **Table 1**, reflecting the overall stand structure and carbon stock conditions across the sampled area.

Table 1.

Summary of Field-Based Biomass and Carbon Estimates.

Plot	Total Biomass of Rubber Trees (kg)	Above-Ground Carbon (tCO_2e)
2	21,597.19	10.15
4	31,422.85	14.77
5	32,780.05	15.41
6	30,169.42	14.18
7	28,313.47	13.31
10	29,517.43	13.87
Total	173,800.41	81.69
Average	28,966.74	13.61
Total for Entire Area	25,056,225.99	11,776.43

The field-based biomass results presented in **Table 1** show a clear and meaningful level of variation among the six sample plots. Total above-ground biomass ranged from 21,597.19 kg (Plot 2) to 32,780.05 kg (Plot 5), indicating substantial differences in stand conditions across the plantation. Plot 5 exhibited the highest biomass accumulation, exceeding the lowest plot by more than 11,000 kg. This difference reflects significant structural variability among plots, driven largely by differences in tree size distribution, canopy development, and stand density. Plot 2, which recorded the lowest biomass, likely represents areas with younger trees or reduced growth performance. Such variability aligns with known patterns of uneven growth often observed in mature rubber plantations, where microenvironmental factors and localized management practices influence stand productivity.

A similar pattern was observed in above-ground carbon (AGC) estimates, which ranged from 10.15 to 15.41 tCO_2e. Plot 5 again showed the highest carbon stock, consistent with its larger biomass, while Plot 2 had the lowest value. The difference of more than 5 tCO_2e across plots suggests a statistically meaningful spread in carbon sequestration capacity within the plantation. The mean AGC value of 13.61 tCO_2e indicates a moderate overall carbon storage potential for the study area. These results highlight that carbon variability is strongly influenced by structural heterogeneity, tree age, and site-specific growth dynamics. Collectively, the findings emphasize the importance of accounting for spatial variation in biomass and carbon assessments, particularly when scaling plot-level estimates to larger landscape units.

3.2. UAV Data Analysis Results

UAV-based carbon estimation was supported by a set of three-dimensional spatial products generated through Agisoft PhotoScan. High-resolution UAV images were processed into georeferenced outputs, including the Digital Surface Model (DSM), Digital Elevation Model (DEM), and Canopy Height Model (CHM). Together, these products provided a detailed view of terrain conditions and vegetation structure within the rubber plantation. Examples of these outputs are shown in **Fig. 3 (a-c)**. These datasets formed the foundation for extracting canopy height, detecting individual trees, and estimating carbon stocks, demonstrating the capability of UAV photogrammetry to capture fine-scale structural variation across plantation ecosystems. **Fig. 3** shows the DSM produced from UAV photogrammetry, representing the elevation of all above-ground features in the plantation. Elevation values range from about 329 to 448 meters above mean sea level. Mature, densely populated rubber stands appear in green tones, indicating higher elevations near 448 meters. In contrast, open fields, grass areas, and canopy-free zones are shown in red tones, representing elevations closer to 329 meters. This variation highlights the distribution of canopy height and reveals vertical structural differences across the landscape.

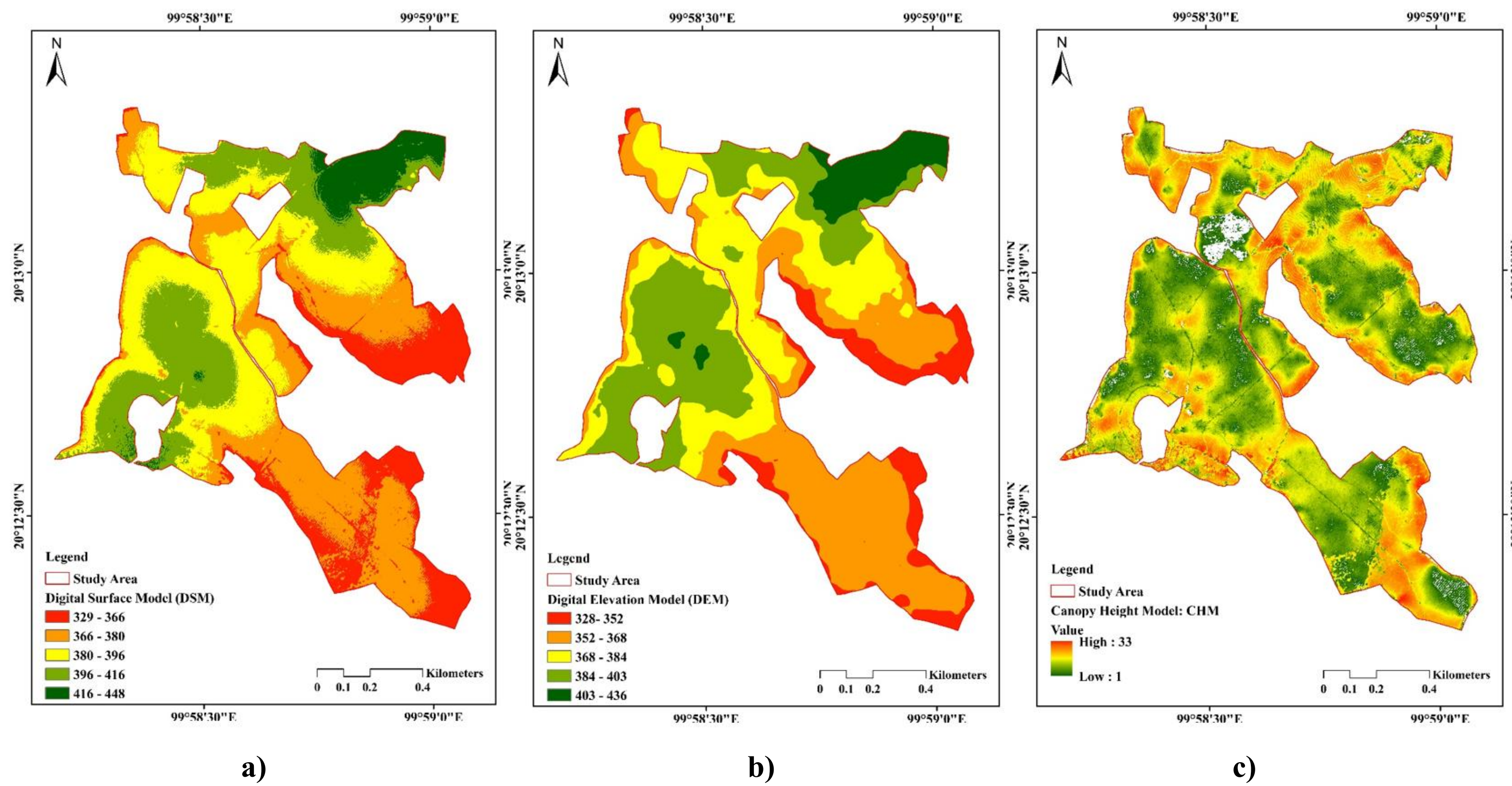


Fig. 3. (a) DSM, (b) DEM, and (c) CHM of the Rubber Plantation.

Fig. 4 presents the DEM, which reflects the bare-earth surface after removing vegetation and structures. Ground elevations range from approximately 328 to 436 meters above mean sea level. Higher terrain appears in green tones, while lower-lying areas—such as flatlands, depressions, and gently sloping surfaces—are shown in red tones. This model offers a clear visualization of underlying topographic patterns, which is essential for understanding how terrain influences tree growth and stand distribution within the plantation.

Fig. 5 displays the CHM, capturing the spatial variation in canopy height across the plantation. Canopy heights range from about 1 to 33 meters. Areas with tall, mature tree canopies are shown in red tones, indicating heights near 33 meters. Open areas or zones lacking tree cover appear in green tones, reflecting heights close to 1 meter. This model effectively represents vertical stand complexity and provides critical information for analyzing biomass patterns and overall plantation structure.

3.3. Machine Learning and Deep Learning Analysis Results

The machine learning and deep learning analyses produced strong and meaningful results in estimating above-ground carbon across the rubber plantation. The spatial products generated by the deep learning workflow offered clear insights into structural variation across the landscape. One of these outputs, the Deepness Model, highlights patterns that help explain variability in biomass and carbon accumulation. **Fig. 4** presents the Deepness Model, which divides the study area into equal-sized grid cells and classifies surface variability into five categories.

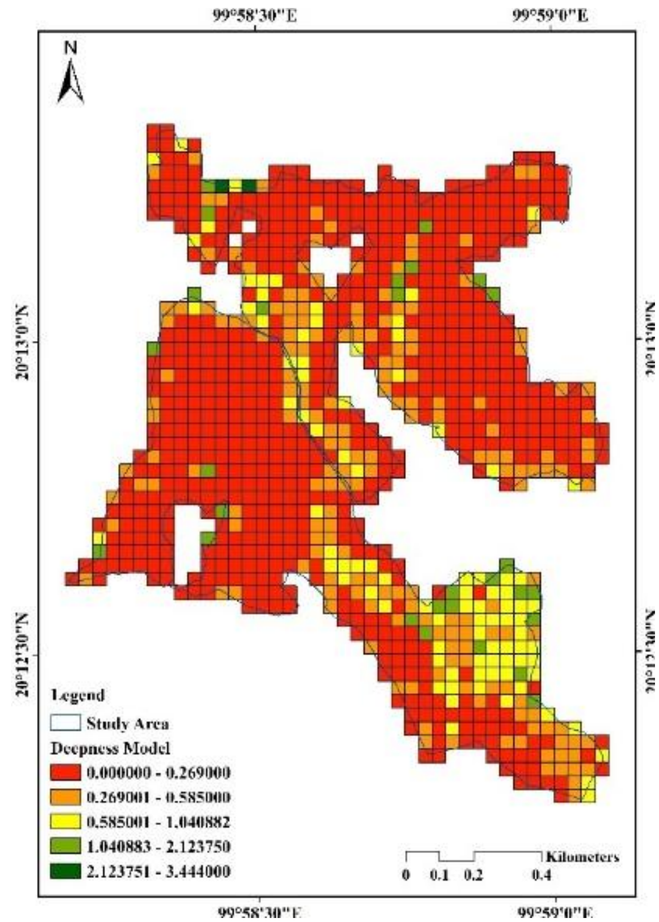


Fig. 4. Deepness Model Illustrating Surface Variability Across the Plantation.

The highest-variability areas, shown in red (2.123751–3.444000), indicate zones with pronounced elevation differences. These areas often correspond to sections of the plantation where micro-topographic shifts influence soil moisture, drainage, and canopy development. Dark orange regions (1.048083–2.123750) reflect moderately high variability, while light orange (0.585001–1.048082) represents moderate variability. Yellow areas (0.269001–0.585000) show minimal fluctuation, and the green zones (0.000000–0.269000) represent the most stable and uniform surfaces.

The spatial patterns—dominated by red and orange classes—suggest a statistically meaningful degree of micro-topographic heterogeneity across the plantation. Such variability is known to influence stand productivity, leading to measurable differences in tree growth, canopy height, and ultimately biomass accumulation. The scattered green patches mark localized stable zones with lower structural variation, serving as important reference areas for understanding baseline site conditions. Overall, the spatial classification provides key contextual information for interpreting carbon-storage patterns across the plantation.

A comparative evaluation was carried out between field-based biomass and carbon estimates and those produced by the deep learning model. Field data, including DBH and tree height, were used to compute biomass for individual tree components—leaves, stems, branches, and stumps—using standard allometric equations. These values were then converted to carbon using established conversion factors. Meanwhile, the deep learning model used UAV-derived structural data to estimate biomass and carbon through the trained predictive model. Summary values for both approaches are provided in **Table 2**, which reveals a strong level of agreement between the field-based measurements and the estimates generated by the deep learning model. Across all biomass components, the differences between the two datasets are relatively small and fall within a range that is considered acceptable for ecological and remote-sensing–based carbon assessment. For example, leaf biomass differs by about 39,000 kg, while stem and branch biomass—accounting for the largest proportion of total biomass—shows a difference of roughly 240,000 kg. Given the scale of the plantation, these deviations are minor and indicate that the model captures the overall structural patterns of the stand with high fidelity.

Total above-ground biomass values also show strong alignment: 25.06 million kg from field measurements versus 24.76 million kg from the deep learning model. The difference of only about 1.2% demonstrates the model's ability to approximate real biomass conditions with a high level of accuracy. This agreement carries over to carbon estimates as well. Field-based carbon stock totaled 11,776.43 tCO_2, while the deep learning model produced 11,639.24 tCO_2—a deviation of less than 1.2%. The carbon-to-biomass ratios follow the same trend, reflecting consistent estimation performance across methods.

A key indicator of model reliability is the average carbon density. After converting units to hectares, the field-based average is 85.06 tCO_2/ha, while the deep learning model yields 84.81 tCO_2/ha. This extremely small difference (0.25 tCO_2/ha) demonstrates statistical consistency and confirms that the model effectively represents carbon distribution at the landscape scale. Such close alignment suggests that deep learning can reliably replicate field-derived patterns of biomass and carbon accumulation, indicating strong potential for scaling carbon assessments in rubber plantations without compromising accuracy.

Table 2.

Comparison of Biomass Components and Above-Ground Carbon Estimates.

Biomass and Carbon Assessment Parameters	Field-Based Estimates	Deep Learning Estimates
Leaf biomass (kg)	320,240.12	280,990.52
Stem, branches, and above-stump biomass (kg)	21,042,496.13	21,284,651.53
Stump and root biomass (kg)	3,693,489.74	3,198,708.59
Total above-ground biomass (kg)	25,056,225.99	24,764,350.63
Carbon-to-biomass ratio ($kgCO_2e$)	11,776,426.22	11,639,244.80
Total carbon stock (tCO_2)	11,776.43	11,639.24
Average carbon density (tCO_2/ha)	85.06	84.81

3.4. Model Accuracy Assessment

The performance of the UAV-based deep learning model was evaluated using both regression-based and classification-based metrics, providing a comprehensive assessment of predictive accuracy. These complementary evaluations allowed a better understanding of how the model performs under variability in tree structure and spatial heterogeneity within the rubber plantation. Regression metrics quantified numerical differences between predicted and observed biomass and carbon values, whereas classification metrics assessed the model's ability to assign trees to the correct DBH classes. The results are summarized in **Tables 3 and 4**, and the relationship between predicted and observed biomass is illustrated in **Fig. 5**. Most data points are distributed close to the 1:1 reference line, indicating good agreement between model predictions and field measurements.

Table 3.
Error Metrics Comparing Model Predictions with Field-Measured Values.

Regression Analysis	Deep Learning
MAE	1.372
MSE	2.372
RMAE	1.165

Table 4.
Accuracy Assessment Results of the Deep Learning Model.

ID	Deep Learning
Precision	1.00
Recall	0.86
F1-score	0.925
Accuracy	92.52%

The regression metrics demonstrate statistically meaningful agreement between model outputs and field observations. The MAE of 1.372 and MSE of 2.372 indicate a narrow prediction error range, suggesting that the model is highly consistent in estimating above-ground biomass and carbon values. These errors fall well below thresholds commonly reported in plantation-based carbon modeling studies, indicating high model stability. The RMAE value of 1.165 further reinforces the model's strong predictive capability, showing limited proportional deviation relative to actual measured values. Such performance highlights the model's suitability for operational-scale biomass and carbon assessments, particularly in structurally complex plantation systems.

The classification results reveal high statistical accuracy and strong operational reliability. A Precision of 1.00 indicates zero false positives, meaning every detected tree was correctly identified. This reflects exceptional model specificity—rarely achieved in dense plantation environments where overlapping crowns can increase detection ambiguity. The Recall value of 0.86 shows that the model successfully detected the majority of trees present, with only a small fraction missed due to canopy overlap or shadowing. The F1-score of 0.925 demonstrates a well-balanced performance between Precision and Recall, confirming that the model performs robustly under varying structural conditions. The overall detection accuracy of 92.52% highlights the model's strong potential for scalable monitoring applications, particularly for carbon accounting and plantation management.

Overall, the accuracy assessment confirms that the deep learning model achieves high predictive power with statistically reliable performance across both numerical estimation and object-detection tasks. The narrow error margins, zero false positives, and high F1-score collectively indicate that the model effectively captures structural patterns that drive biomass and carbon variation in rubber plantations. These findings support its use as a scientifically robust tool for large-scale carbon stock assessment, contributing valuable data for climate reporting, ecosystem monitoring, and precision agriculture.

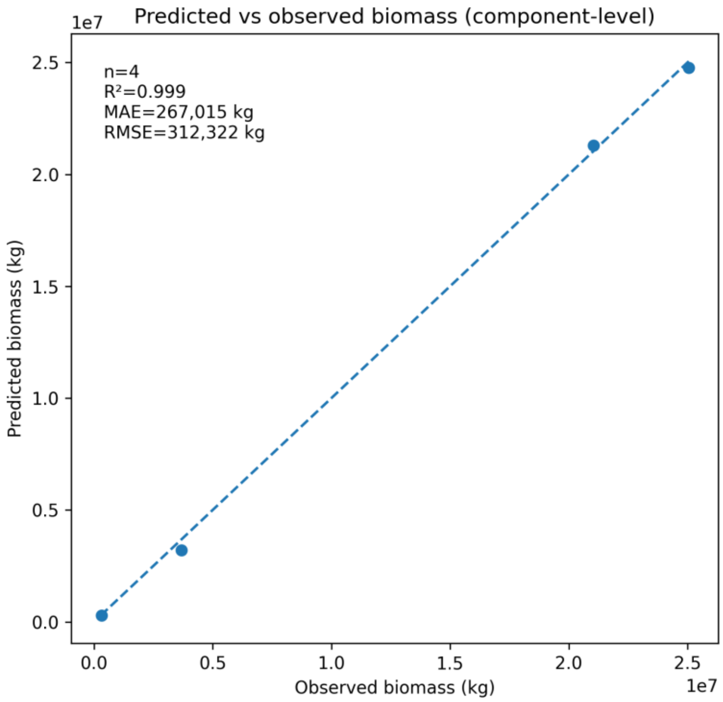


Fig. 5. Comparison between model-predicted biomass and field-measured biomass. The dashed line represents the 1:1 reference line, and evaluation metrics (R^2, MAE, and RMSE) are shown within the plot.

4. DISCUSSION

This study shows that integrating UAV photogrammetry with geospatial deep learning can improve above-ground carbon estimation in rubber plantations. Traditional field surveys often have difficulty capturing canopy structure and spatial variability over large areas. In contrast, UAV-derived datasets provide detailed information on terrain and canopy characteristics at high spatial resolution. The DSM, DEM, and CHM products clearly represented differences in canopy height and stand structure across the study plots. These variations were related to differences in tree size, canopy spread, and plantation age. The CHM was particularly useful for identifying spatial patterns of canopy height that corresponded well with observed biomass and carbon variations. Such patterns are difficult to detect using field measurements alone, especially in mature plantations.

The deep learning results further confirm the effectiveness of the proposed approach. Tree detection performance showed high precision and a strong F1-score, indicating reliable identification of individual trees. Although recall was slightly lower, this is expected in areas with dense canopy where some tree crowns overlap. Overall detection accuracy exceeded 92%, suggesting that the method performs well under typical plantation conditions.

Predicted biomass and carbon values were consistent with field measurements. The differences between predicted and observed values were relatively small, indicating that UAV-derived canopy variables can effectively represent ground conditions. This agreement is encouraging considering the uncertainties associated with both field sampling and photogrammetric processing. The results also suggest that terrain variability may influence canopy structure and carbon distribution. Areas with greater micro-topographic variation tended to show more complex canopy patterns. This indicates that terrain-related factors could be useful variables in future modeling studies.

Despite the promising results, some limitations should be considered. Detection accuracy may decrease in areas with strong canopy overlap, where individual tree crowns are partially obscured. Variations in terrain and lighting conditions may also affect the quality of UAV-derived models and influence prediction accuracy. In addition, the model was developed using data from a specific plantation environment, which may limit its direct application to other forest conditions without further calibration. Future research could improve model robustness by incorporating more diverse training data and additional environmental variables.

The results obtained in this study are comparable with those reported in previous research using UAV and machine learning approaches for biomass estimation. Several studies have shown that canopy height and crown-related variables derived from UAV data are strongly related to biomass and carbon stock. The high agreement observed between predicted and field values in this study is consistent with those findings. However, some studies reported lower prediction accuracy in areas with complex canopy structures or uneven terrain. The relatively strong performance observed here may be related to the integration of canopy metrics with deep learning-based tree detection, which helps improve structural representation. These comparisons suggest that the proposed approach provides reliable results and contributes to the development of UAV-based biomass estimation methods.

Overall, the findings indicate that geospatial deep learning provides a practical tool for biomass and carbon estimation in rubber plantations. The approach reduces field workload, improves spatial detail, and supports repeatable monitoring at the landscape scale. This method has strong potential for plantation management, carbon accounting, and long-term environmental planning in regions with similar conditions.

5. CONCLUSION

This research demonstrates that the combination of UAV photogrammetry and geospatial deep learning is an efficient method for above-ground carbon estimation in rubber plantations. Through integration of field measurements, CHM-driven canopy metrics and YOLO-based tree detection, the workflow resulted in carbon estimates closely aligned with ground measurements. The findings further give evidence that the UAV-based information on the structure can adequately reflect between-tree variability in biomass and carbon. It decreases the demand of intensive field surveying, and still obtains high spatial resolution on plot and landscape levels. This makes the approach practical for repeated monitoring within plantation environments where time, cost or access limits field data collection. The analysis also underscores the importance of including terrain-related information as the Deepness Model indicates that micro-topographic variation affects canopy structure and carbon distribution. To validate the method, it is suggested that future work trialling the proposed approach be conducted on a range of plantations with variable age, management and environment. Inclusion of other data sources, such as high-resolution multispectral UAV imagery, LiDAR or satellite time series data could be expected to increase the robustness of models assembled for mapping in this area. Together, this study illustrates the feasibility of geospatial deep learning as an application to biomass and carbon mapping in plantation forestry and associated landcape contexts.

ACKNOWLEDGMENTS

This research project was financially supported by Mahasarakham University.

REFERENCES

Angkahad, T., Chokkuea, W., Sangpradid, S., Uttaruk, Y., Viriyasatr, K., & Laosuwan, T. (2025a). The assessment of above-ground biomass and carbon sequestration in community forests using UAV technology. *Defence Technology Academy Journal, 7*(15), R11–R28.

Angkahad, T., Laosuwan, T., Sangpradid, S., Prasertsri, N., Uttaruk, Y., Phoophathong, T., & Nuchthapho, J. (2024). An empirical analysis of above-ground biomass and carbon sequestration utilizing UAV photogrammetry and machine learning techniques. *IEEE Access, 12*, 186740–186752. https://doi.org/10.1109/ACCESS.2024.3514074

Angkahad, T., Laosuwan, T., Sangpradid, S., Prasertsri, N., Uttaruk, Y., Phoophathong, T., & Nuchthapho, J. (2025b). Developing a drone-based machine learning for spatial modeling and analysis of biomass and carbon sequestration in forest ecosystems. *Engineered Science, 35*, 1508. https://doi.org/10.30919/es1508

Awichin, P., Laosuwan, T., Sangpradid, S., Uttarak, Y., Phoophathong, T., Leammanee, S., & Nuchthapho, J. (2025). UAV multispectral indices and linear regression for estimating biomass and carbon in rubber plantations. *IEEE Access*, 13, 215845–215860. https://doi.org/10.1109/ACCESS.2025.3644408

Aroonsri, I. & Sangpradid, S. (2021). Artificial Neural Networks for the classification of shrimp farm from satellite imagery. *Geographia Technica*, 16(2), 149–159. https://doi.org/10.21163/GT_2021.162.12

Blagodatsky, S., Xu, J., & Cadisch, G. (2016). Carbon balance of rubber (*Hevea brasiliensis*) plantations: A review of uncertainties at plot, landscape and production level. *Agriculture, Ecosystems & Environment, 221*, 8–19. https://doi.org/10.1016/j.agee.2016.01.025

Earthobservatory. (2025). *The carbon cycle.* https://earthobservatory.nasa.gov/features/CarbonCycle

Ecke, S., Dempewolf, J., Frey, J., Schwaller, A., Endres, E., Klemmt, H.-J., Tiede, D., & Seifert, T. (2022). UAV-based forest health monitoring: A systematic review. *Remote Sensing, 14*(13), 3205. https://doi.org/10.3390/rs14133205

Espíndola, R. P., Picanço, M. M., de Andrade, L. P., & Ebecken, N. F. F. (2025). Applications of machine learning methods in sustainable forest management. *Climate, 13*(8), 159. https://doi.org/10.3390/cli13080159

Haidu, I. (2024). Geography’s future is technical: A letter from the editor. *Geographia Technica*, 19(2), i–vi. https://doi.org/10.21163/GT_2024.192.23

Haidu, I., Ionac, N., & Irașoc, A. (2024). The new climatic normals and their impact on bioclimatic indices: Case study: Oradea (Romania). *Present Environment and Sustainable Development*, 18(2), Article 2001. https://doi.org/10.47743/pesd2024182001

Huynh, T., Lewis, T., Applegate, G., Pachas, A. N. A., & Lee, D. J. (2022). Allometric equations to estimate aboveground biomass in spotted gum (*Corymbia citriodora* subspecies variegata) plantations in Queensland. *Forests, 13*(3), 486. https://doi.org/10.3390/f13030486

Hytönen, J., Kaakkurivaara, N., Kaakkurivaara, T., & Nurmi, J. (2018). Biomass equations for rubber tree (*Hevea brasiliensis*) components in southern Thailand. *Journal of Tropical Forest Science, 30*(4), 588–596. https://doi.org/10.26525/jtfs2018.30.4.588596

Intarat, K., Kuankhamnuan, P., & Jangsawang, W. (2026). Flood mapping in Phra Nakhon Si Ayutthaya, Thailand, utilizing Sentinel-1 SAR imagery and deep learning approaches. *Geographia Technica, 21*, 1–18. https://doi.org/10.21163/GT_2026.211.04

Intarat, K., Netsawang, P., Narawatthana, S., Promaoh, C., & Chuenkamol, S. (2025). Integrating spectral and texture indices with machine learning for rice leaf area index estimation in Suphan Buri, Thailand. *Geographia Technica, 20*(1), 207–227. https://doi.org/10.21163/GT_2025.201.15

Intarat, K., Nuengjumnong, N., Sae-Jung, J., Jangsawang, W., Yoomee, P., & Panboonyuen, T. (2024). Deep residual neural networks with self-attention for landslide susceptibility mapping. In *GIS-IDEAS 2024* (pp. 1–6). IEEE. https://doi.org/10.1109/GIS-IDEAS63212.2024.10990879

Jian, K., Lu, D., & Li, G. (2025). Modeling forest carbon stock based on sample plots and UAV LiDAR data and examining its vertical characteristics in Wuyishan National Park. *Remote Sensing, 17*(3), 377. https://doi.org/10.3390/rs17030377

Jirakajohnkool, S., Wongsai, S., Sanpayao, M., & Wongsai, N. (2025). Age-based biomass carbon estimation and soil carbon assessment in rubber plantations. *Forests, 16*(11), 1652. https://doi.org/10.3390/f16111652

Juan-Ovejero, R., Elghouat, A., Navarro, C. J., Reyes-Martín, M. P., Jiménez, M. N., Navarro, F. B., & Castro, J. (2023). Estimation of aboveground biomass and carbon stocks of *Quercus ilex* saplings using UAV RGB imagery. *Annals of Forest Science, 80*, 44. https://doi.org/10.1186/s13595-023-01210-x

Laosuwan, T., Uttaruk, Y., Nakapan, S., Itsarawisut, J., & Plybour, C. (2025). Evaluation of tree biomass and carbon sequestration through remote sensing and field methods. *Geographia Technica*, 33–43. https://doi.org/10.21163/gt_2025.201.04

Laosuwan, T., Uttaruk, Y., Sangpradid, S., Butthep, C., & Leammanee, S. (2023). The carbon sequestration potential of silky oak (*Grevillea robusta*). *Forests, 14*(9), 1824. https://doi.org/10.3390/f14091824

Mao, Z., Abdi, O., Uusitalo, J., Laamanen, V., & Kivinen, V.-P. (2025). Super-resolution supporting individual tree detection using aerial imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 224*, 251–271. https://doi.org/10.1016/j.isprsjprs.2025.04.005

Mohan, M., Silva, C. A., Klauberg, C., Jat, P., Catts, G., Cardil, A., Hudak, A. T., & Dia, M. (2017). Individual tree detection from UAV-derived canopy height models. *Forests, 8*(9), 340. https://doi.org/10.3390/f8090340

Ngthai. (2025). *What is the carbon cycle?* https://ngthai.com/science/31560/carbon-cycle/

Plybour, C., Laosuwan, T., Uttaruk, Y., Awichin, P., Rotjanakusol, T., Itsarawisut, J., & Singharath, M. (2025). An investigation of plant species diversity, above-ground biomass, and carbon stock. *Diversity, 17*(6), 428. https://doi.org/10.3390/d17060428

Que, H., Gao, H., Shan, W., Liu, M., An, J., Deng, F., Feng, S., Yang, X., & Mu, L. (2026). FM-SAM: Tree crown delineation using SAM and YOLOv10. *Computers and Electronics in Agriculture, 240*, 111162. https://doi.org/10.1016/j.compag.2025.111162

Tai, H., Rao, C., Li, X., Li, H., & Li, C. (2025). Extracting rubber tree parameters and estimating carbon storage using airborne LiDAR. *PLOS ONE, 20*(8), e0330768. https://doi.org/10.1371/journal.pone.0330768

Thailand Greenhouse Gas Management Organization. (2025). *What is T-VER?* https://ghgreduction.tgo.or.th/en/what-is-t-ver/what-is-t-ver.html

Tiko, J. M., Ndjadi, S. S., Obandza-Ayessa, J. L., Mweru, J. P. M., Michel, B., Beeckman, H., Rakotondrasoa, O. L., & Hulu, J. P. M. T. (2025). Carbon sequestration potential in rubber plantations. *Earth, 6*(2), 21. https://doi.org/10.3390/earth6020021

Uttaruk, Y., Laosuwan, T., Sangpradid, S., Samek, J. H., Butthep, C., Rotjanakusol, T., Dumrongsukit, S., & Rouylarp, Y. (2026). Species-Specific Allometric Models for Biomass and Carbon Stock Estimation in Silver Oak (*Grevillea robusta*) Plantation Forests in Thailand: A Pilot-Scale Destructive Study. *Forests*, *17*(1), 100. https://doi.org/10.3390/f17010100

Yan, Y., Lei, J., & Huang, Y. (2024). Forest aboveground biomass estimation using UAV-LiDAR and machine learning. *Sensors, 24*(21), 7071. https://doi.org/10.3390/s24217071

Yang, X., Blagodatsky, S., Liu, F., Beckschäfer, P., Xu, J., & Cadisch, G. (2017). Rubber tree allometry and carbon stocks in subtropical China. *Forest Ecology and Management, 404*, 84–99. https://doi.org/10.1016/j.foreco.2017.08.013

Zhu, X., Chen, B., Wang, C., & He, H. (2022). Above-ground biomass estimation using UAV LiDAR and allometric models. *Remote Sensing, 14*(5), 1024. https://doi.org/10.3390/rs14051024

Geographia Technica, Vol. 21, Special Issue: Artificial Intelligence Applications in Geography, 2026, pp 174 to 187

DEEP LEARNING-BASED DETECTION OF AGRICULTURAL BURNED AREAS USING TRUE COLOR SATELLITE IMAGERY

***Worawit SUPPAWIMUT* [1,2] & *Ratchaphon SAMPHUTTHANONT* [1,2*]**

DOI: 10.21163/GT_2026.212.09

ABSTRACT

Chiang Mai Province, situated in northern Thailand, is experiencing persistent challenges regarding seasonal haze and particulate matter (PM2.5) contamination. The primary sources of these emissions are agricultural burning practices, particularly the combustion of rice straw and stubble in proximity to residential areas. This study aims to (1) assess agricultural burned areas in rice fields using the differenced Normalized Burn Ratio (dNBR) derived from Sentinel-2 imagery, and (2) develop a deep learning model to detect burned areas from true color satellite imagery (RGB) using the U-Net architecture. Sentinel-2 Level-2A images acquired before and after burning events in 2024 were used for analysis. The study focused on rice cultivation areas, while clouds, cloud shadows, and water bodies were removed using the Scene Classification Layer. Subsequently, the Normalized Burn Ratio (NBR) and the Differenced Normalized Burn Ratio (dNBR) were calculated to generate a burned area mask. The validation results indicate that a dNBR threshold of ≥ 0.1 is appropriate for the specific context of the study area. This mask was used as reference data for training the deep learning model. The U-Net model was trained using RGB image tiles and corresponding binary masks, with the final deep learning dataset consisting of 2,560 image tiles of 256×256 pixels after data augmentation. A spatial hold-out strategy was applied to prevent spatial data leakage between training, validation and test datasets. The results show that U-Net outperformed DeepLabv3+, achieving an Intersection over Union (IoU) of 0.9292 and an F1-score of 0.9633, indicating strong capability in detecting small burned patches. Spatial analysis revealed that the burned areas accounted for 9.952% of the total image area and 18.937% of the total cultivated land. These findings suggest that integrating true-color satellite imagery with deep learning techniques provides significant potential for monitoring agricultural burning and supporting local strategies to mitigate PM2.5 pollution.

Keywords: *Deep Learning, Agricultural Burned Area, Normalized Burn Ratio, Satellite Imagery*

1. INTRODUCTION

Air pollution caused by fine particulate matter (PM2.5) has become a critical environmental and public health issue worldwide. In northern Thailand, particularly in Chiang Mai Province, recurrent haze episodes are strongly associated with open biomass burning occurring in both forest and agricultural areas. Among these sources, agricultural residue burning—especially rice straw burning—plays a significant role due to its proximity to urban and peri-urban communities. Previous studies have reported that rice residue burning can emit up to 38 ± 22 kilotons of PM2.5 annually, while populations in suburban areas experience higher exposure levels compared to those affected by distant forest fires (Junpen et al., 2018). Elevated PM2.5 concentrations have also been significantly associated with increased hospital admissions related to respiratory and cardiovascular diseases, particularly in urban Chiang Mai (Kamton et al., 2019; Jarernwong et al., 2023). In addition, biomass burning emissions can be identified through chemical tracers such as potassium, which serves as a key indicator of biomass combustion in the atmosphere (Khamkaew et al., 2016).

[1] *Department of Geography and Geoinformatics, Faculty of Humanities and Social Sciences, Chiang Mai Rajabhat University, Thailand, worawit_sup@cmru.ac.th (WS), corresponding author* ratchaphon_sam@cmru.ac.th (RS).*

[2] *Asian Air Quality Operations Center by Space Technology, Geoinformatics & Environmental Engineering (AiroTEC), Chiang Mai Rajabhat University, Thailand.*

At the global scale, biomass burning represents a major source of atmospheric pollutants, contributing significantly to emissions of carbonaceous aerosols and trace gases such as CO and NO_x (Andreae, 2019). It has been estimated that biomass burning releases approximately 2.2 Pg C yr^{-1}, with a substantial portion originating from small fires that are often underestimated in conventional emission inventories (van der Werf et al., 2017). Regionally, Southeast Asia is recognized as one of the most complex aerosol environments, where seasonal haze is driven by interactions among emission sources, meteorological conditions, and land-use dynamics (Reid et al., 2013). These haze events frequently result in severe air pollution episodes across the region, posing substantial health risks, as exposure to PM2.5 has been linked to increased mortality and a wide range of cardiopulmonary diseases even at relatively low concentrations (World Health Organization, 2021).

Despite continuous regulatory efforts, agricultural burning in Chiang Mai remains prevalent, particularly during the post-harvest period of rice cultivation. These burning activities typically occur between November and December and are characterized by small, fragmented patches distributed across agricultural fields near populated areas. Such spatial characteristics contribute to rapid increases in local PM2.5 concentrations, emphasizing the importance of accurately detecting and monitoring burned areas at fine spatial scales.

Satellite remote sensing has been widely applied for monitoring biomass burning at regional and global scales. Burned area products derived from Moderate Resolution Imaging Spectroradiometer (MODIS), such as MCD64A1, provide long-term and consistent global datasets at moderate spatial resolution (~500 m) by integrating surface reflectance time series with active fire detections (Giglio et al., 2018). Similarly, the Visible Infrared Imaging Radiometer Suite (VIIRS) enhances fire detection capability through improved spatial resolution (375 m), allowing better identification of small and low-intensity fires (Schroeder et al., 2014). However, both MODIS and VIIRS remain limited in detecting small and fragmented agricultural burning due to their coarse spatial resolution, leading to potential underestimation of burned areas in smallholder agricultural systems.

The availability of higher-resolution satellite data, such as Landsat and Sentinel-2 (10–20 m), has significantly improved the detection of small burned areas. Sentinel-2, in particular, enables the calculation of spectral indices such as the Normalized Burn Ratio (NBR) and differenced NBR (dNBR), which are widely used for burned area detection and burn severity assessment. Previous studies have demonstrated that Sentinel-2 can detect small-scale agricultural burning that is often missed by traditional hotspot data (Chuvieco & Stoyanov, 2023). Furthermore, multi-temporal approaches integrating Sentinel-2 data have shown improved accuracy in capturing small burned patches, especially those smaller than 100 hectares (Roteta et al., 2019). In the context of Chiang Mai, Samphutthanont (2024) demonstrated that dNBR derived from Sentinel-2 can effectively detect burned areas in rice fields characterized by low fuel loads and low combustion temperatures, which are typically undetectable using conventional hotspot data. Nevertheless, these approaches rely heavily on spectral index calculations and often require multiple processing steps.

Recent advances in deep learning have significantly transformed remote sensing image analysis, particularly for semantic segmentation tasks. Convolutional Neural Networks (CNNs), including Fully Convolutional Networks (Long et al., 2015) and U-Net architectures (Ronneberger et al., 2015), enable pixel-level classification by learning complex spatial features directly from imagery. These approaches have demonstrated superior performance compared to traditional machine learning methods, particularly in large-scale image analysis (Ma et al., 2019; Zhu et al., 2017). In the context of burned area detection, deep learning models have shown strong capability in capturing spatial patterns of burned regions, especially when applied to high-resolution satellite imagery (Knopp et al., 2020; Martins et al., 2022; Hu et al., 2021). For example, CNN-based models applied to Sentinel-2 data have achieved overall accuracies exceeding 97% and have outperformed conventional machine learning approaches (Tonbul et al., 2023). Moreover, integrating True Color (RGB) imagery with spectral indices has been shown to further improve model performance (Yılmaz & Kavzoğlu, 2024).

However, most existing studies rely on multispectral inputs and spectral indices as primary features. Research focusing on the direct detection of agricultural burned areas using only True Color (RGB) satellite imagery remains limited. This limitation is particularly relevant in regions such as northern Thailand, where agricultural burning occurs in small, fragmented patches within complex urban–rural landscapes and directly impacts human health. This gap highlights the need for developing flexible, data-driven approaches that can operate effectively using readily interpretable imagery and support practical applications in environmental monitoring and policy decision-making.

Therefore, the objectives of this study are to (1) assess agricultural burned areas in rice fields using the differenced Normalized Burn Ratio (dNBR), and (2) develop and evaluate a deep learning-based approach for detecting agricultural burned areas in rice fields from true color satellite imagery in Chiang Mai Province, Thailand. The findings are expected to support improved monitoring of PM2.5 emission sources and contribute to more effective haze management strategies at the local level.

2. STUDY AREA

The study area for this research encompasses the provincial boundaries of Chiang Mai Province in northern Thailand, covering a total area of 22,436.07 square kilometers. This region includes approximately 5,519.08 square kilometers of agricultural land, accounting for 24.60% of the total provincial area (Office of the Permanent Secretary for Agriculture and Cooperatives, Chiang Mai Province, 2026), where agricultural burning activities remain prevalent.

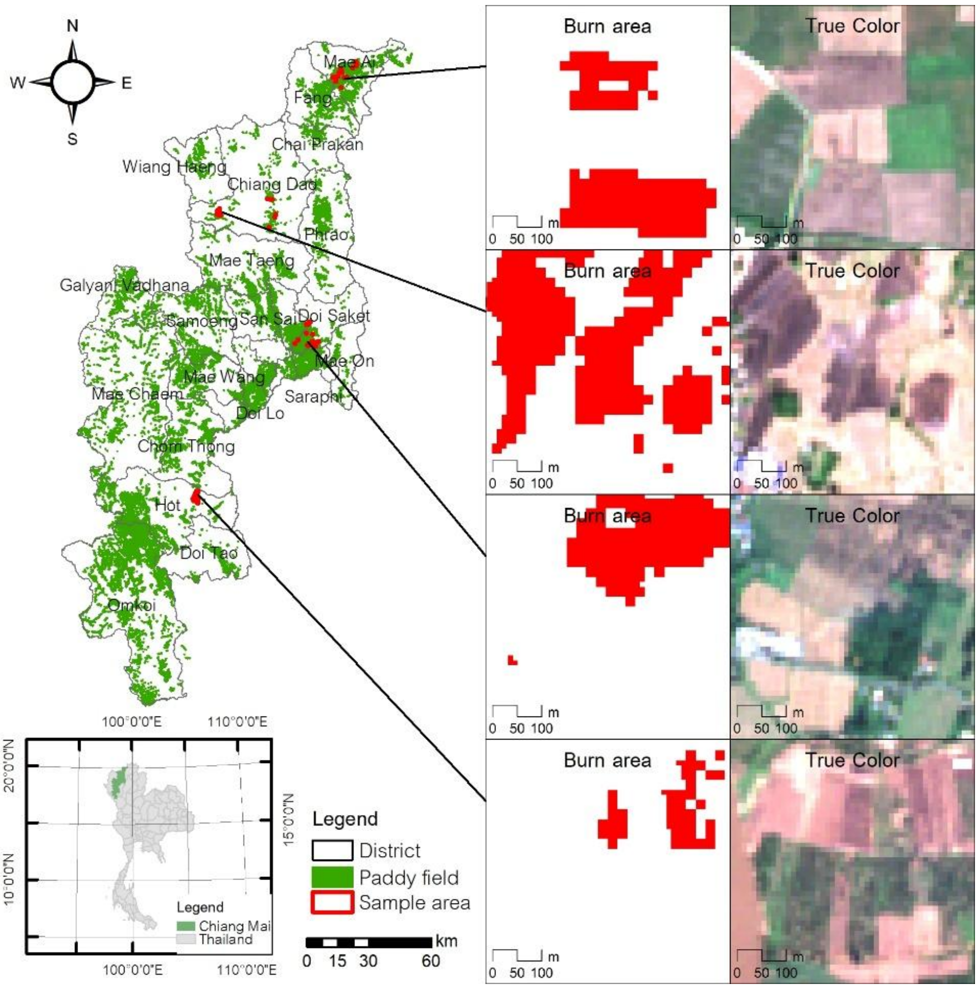


Fig. 1. Study area in Chiang Mai Province, Thailand, showing the selected districts used to generate the training dataset, the distribution of agricultural burned areas, and examples of true color satellite imagery (RGB) in four districts.

To develop representative training datasets, four districts distributed across the northern, central, and southern parts of the province were selected as sample areas: Mae Ai District, Chiang Dao District, Doi Saket District, and Hot District. These districts were used to construct spatially distributed training datasets for the analysis. The study primarily focuses on detecting agricultural burned areas, particularly in rice cultivation fields, using the differenced Normalized Burn Ratio (dNBR) and deep learning techniques (**Fig. 1**).

3. DATA AND METHODS

The Data and Methods section begin with the presentation of the conceptual framework (**Fig. 2**) in which Input Data, Processing, dNBR Label generation, respectively, Deep Learning Model, Model Evaluation are discussed, and if the accuracy assessment is satisfactory then the procedure ends with the final burned area map and other spatial analysis possibilities.

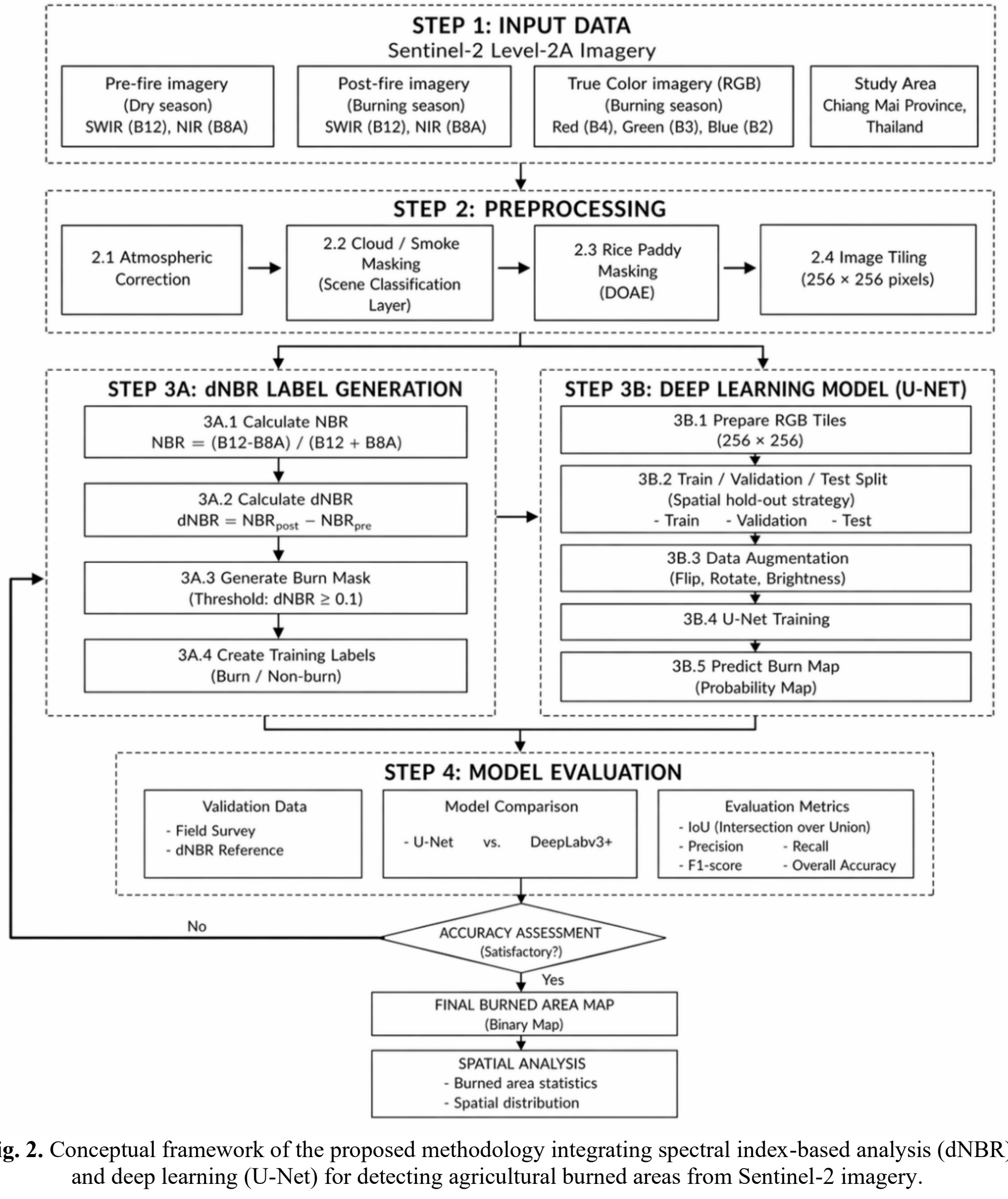


Fig. 2. Conceptual framework of the proposed methodology integrating spectral index-based analysis (dNBR) and deep learning (U-Net) for detecting agricultural burned areas from Sentinel-2 imagery.

3.1. Satellite Image Data Preparation

This study utilized Sentinel-2 Level-2A (L2A) satellite data, which provide atmospherically corrected surface reflectance products suitable for high-resolution analysis of land surface changes. In addition, areas obscured by smoke from biomass burning were identified and removed to minimize potential interference in the analysis. Data from the post-harvest period of rice cultivation in 2024 were specifically selected to encompass the intervals immediately before and after the burning of rice straw and stubble within the study area. The acquisition dates used for the analysis are summarized in **Table 1**.

For the calculation of the Normalized Burn Ratio (NBR), Sentinel-2 Level-2A imagery in the Short-Wave Infrared (SWIR) and Narrow Near-Infrared (NIR) bands was employed, both having a spatial resolution of 20 meters. Meanwhile, Sentinel-2 Level-2A imagery used for True Color Composite (RGB), consisting of Band 4 (Red), Band 3 (Green), and Band 2 (Blue), has a spatial resolution of 10 meters.

To ensure that the analysis focused specifically on the target area, all satellite imagery was clipped to include only rice cultivation areas using available agricultural field boundary data. This spatial masking process helps reduce noise caused by land cover changes in other land-use types, such as forested areas or built-up zones, thereby improving the accuracy of burned area detection within agricultural landscapes.

As clouds, cloud shadows, and water bodies may introduce significant uncertainty in spectral index calculations, this study utilized the Scene Classification Layer (SCL) provided with Sentinel-2 data to identify and remove these areas from the analysis. Only pixels classified as non-vegetated areas (SCL value = 5) were retained, while pixels corresponding to clouds, cloud shadows, and water were excluded. This preprocessing step helps reduce the likelihood of misclassifying non-burned areas as burned areas.

Subsequently, a dataset of 64 satellite images was prepared in True Color Composite (RGB) format using Band 4 (Red), Band 3 (Green), and Band 2 (Blue). These images were used both for visual interpretation and as input data for the deep learning analysis in the subsequent stage.

Table 1.

Sentinel-2 satellite images used for pre-fire and post-fire periods.

District	Pre-fire Date	Post-fire Date
Mae Ai District	30-Oct-24	4-Nov-24
Chiang Dao District	24-Nov-24	4-Dec-24
Doi Saket District	19-Nov-24	4-Dec-24
Hot District	30-Oct-24	4-Nov-24

The selection of Sentinel-2 satellite imagery was based on principles appropriate for burned area analysis. The acquisition dates were chosen to correspond with periods when agricultural burning was observed in the study area. These periods were identified through both field observations and visual interpretation of true color satellite imagery, which indicated that burning activities mainly occurred between November and December. Subsequently, satellite images acquired during this period that were free from obstructions such as clouds or smoke were selected to ensure reliable analysis of burned areas.

3.2. Burned Area Analysis Using the dNBR Index

Burned area detection in this study was based on the Normalized Burn Ratio (NBR), a spectral index developed to enhance the contrast between burned and unburned surfaces. The index was calculated using the reflectance values from the Short-Wave Infrared (SWIR) and Narrow Near-Infrared (Narrow NIR) bands of Sentinel-2 imagery, as expressed in Equation (1):

$$\text{NBR} = \frac{\text{B12} - \text{B8A}}{\text{B12} + \text{B8A}} \tag{1}$$

where B12 and B8A represent the surface reflectance values of the SWIR and Narrow NIR bands, respectively.

Subsequently, the differenced Normalized Burn Ratio (dNBR) was calculated by subtracting the pre-fire NBR from the post-fire NBR in order to emphasize spectral changes caused by burning activities in rice paddy fields. The dNBR was computed as expressed in Equation (2):

$$\text{dNBR} = \text{NBR}_{\text{post-fire}} - \text{NBR}_{\text{pre-fire}} \tag{2}$$

The resulting dNBR values were analyzed together with spatial data of sample rice fields. The analysis revealed that rice fields after harvesting or after the burning of rice residues tended to exhibit higher dNBR values. However, the burning severity in agricultural areas was generally lower than that typically observed in forest fires. In the study area, the maximum dNBR values were also found to be lower than the commonly used international threshold for burned area classification (e.g., 0.27 according to the USGS standard). This finding indicates the necessity of adjusting the classification threshold to better align with the specific characteristics of burned rice fields in the study area.

Several threshold levels were therefore tested based on the dNBR burn severity classification proposed by the United States Geological Survey (USGS), which categorizes burn severity into seven classes. In this study, areas classified as Moderate-low Severity with dNBR values greater than 0.1 (dNBR > 0.1) were considered burned areas. This threshold was found to provide the most appropriate representation of actual burned areas in rice fields and showed strong agreement with visual interpretation derived from true color satellite imagery.

In addition, True Color satellite composite images were generated using the three visible bands of Sentinel-2 imagery, namely Band 4 (Red), Band 3 (Green), and Band 2 (Blue). These composites corresponded to the post-fire observation period, which was the same time frame used for the burned area classification described above.

The burned area classification results derived from the dNBR index together with the True Color satellite imagery were subsequently used as reference data (labels or ground truth) for training and evaluating the deep learning model in the next stage of the study.

To prepare the dataset for model training and evaluation, the burned area classification results from the dNBR analysis and the True Color satellite images were exported as JPEG image files. In this study, burned areas identified from the dNBR classification were visualized in red, while non-burned areas were represented in white. For the True Color composite images, the display settings were adjusted using Standard Deviation Stretch to enhance visual contrast.

The satellite image samples were extracted as square image patches, with 16 images selected from each district, covering the four study districts: Mae Ai, Chiang Dao, Doi Saket, and Hot. In total, 64 image samples were generated, consisting of 64 burned area masks and 64 corresponding True Color images. Each image patch represents an area of 500 × 500 meters at a map scale of 1:5,000, with a map layout size of 10 × 10 centimeters. The exported images were produced at a resolution of 602 pixels (153 dpi) for use in deep learning model training and evaluation.

In addition to the continuous dNBR values, burn severity can be categorized into discrete classes to facilitate interpretation and comparison across different areas. The classification framework widely adopted by the United States Geological Survey (USGS) is based on the Landscape Assessment (LA) methodology developed by Key and Benson (2006), which integrates remote sensing indices with field-based observations such as the Composite Burn Index (CBI). In this framework, burn severity is defined as the magnitude of ecological change caused by fire, rather than the intensity of the fire itself, emphasizing the post-fire effects on vegetation, soil, and ecosystem structure.

The USGS classification scheme uses dNBR values derived from pre- and post-fire NBR images to represent the degree of change caused by fire. These values are then stratified into several burn severity levels, typically including unburned, low, moderate-low, moderate-high, and high severity classes. This approach enables the spatial representation of fire effects across landscapes and allows for consistent comparison of burn severity across different regions and time periods (Key & Benson, 2006).

However, previous studies have highlighted limitations associated with using absolute dNBR thresholds across heterogeneous landscapes. Miller and Thode (2007) demonstrated that the magnitude of dNBR is strongly influenced by pre-fire vegetation conditions, which may lead to misclassification of burn severity when applying uniform thresholds. For example, areas with sparse vegetation may exhibit low dNBR values even after complete biomass loss, while densely vegetated areas may show high dNBR values under similar fire effects. This suggests that burn severity should be interpreted as a relative measure of ecological change rather than an absolute spectral difference.

To address this issue, Miller and Thode (2007) proposed the use of a relative version of dNBR (RdNBR), which normalizes the change relative to pre-fire conditions, thereby improving the comparability of burn severity across different vegetation types and landscapes. Although RdNBR provides advantages in heterogeneous environments, the conventional dNBR-based classification remains widely used due to its simplicity and operational applicability, particularly in large-scale assessments where field calibration data may be limited. In this study, the dNBR classification approach is adapted to the context of agricultural burning in rice paddy fields. Given the relatively low biomass and rapid post-harvest changes in these areas, the standard USGS thresholds may not be directly applicable. Therefore, an adjusted threshold (dNBR > 0.1) is employed to identify burned areas, ensuring that small-scale and low-intensity agricultural burns can be effectively detected while maintaining consistency with the conceptual framework of burn severity classification.

3.3. Deep Learning Analysis

This research employs a deep learning approach to develop a method for detecting agricultural burned areas from true color satellite imagery (RGB). The analysis focused on pixel-level segmentation in order to accurately delineate the boundaries of small burned patches commonly found in agricultural fields. The input data consisted of RGB image tiles with a size of 256 × 256 pixels, accompanied by binary masks (1 = burned area, 0 = other areas) for supervised learning.

To ensure robust model evaluation and avoid spatial autocorrelation, the dataset was divided into training, validation, and testing sets using a spatial hold-out strategy. This approach ensures that the performance metrics reflect the model's generalizability to unseen geographic areas. Prior to model training, several preprocessing steps were performed to standardize the data. These included resizing the images to a consistent dimension, normalizing pixel values to an appropriate range for model training, and conducting random quality checks to verify the correct alignment between input images and their corresponding masks.

The deep learning model used in this study was the U-Net architecture, which is widely recognized for its effectiveness in image segmentation tasks. The U-Net model is capable of capturing both global contextual information and fine spatial details through the use of skip connections that link encoder and decoder layers. Although advanced architectures such as Attention U-Net and ResU-Net have been proposed in recent studies, the standard U-Net was selected in this research due to its computational efficiency, architectural simplicity, and proven effectiveness in small dataset scenarios. These characteristics make it particularly suitable for practical implementation in local-scale applications and resource-constrained environments. The model output is a probability map representing the likelihood that each pixel belongs to a burned area. This probability map is subsequently converted into a binary mask for burned area estimation. To improve model robustness, data augmentation techniques were applied during training, including image flipping, rotation, and slight brightness adjustments. Model performance was continuously monitored using the validation dataset to identify the optimal model during training.

Model performance was evaluated using standard metrics that assess both classification accuracy and the capability of detecting burned areas. Additionally, the reliability of the results was assessed by comparing burned area detection from the U-Net model and the DeepLabv3+ model (using a True Color image acquired on 24 December 2024) with burned area classifications derived from field observations combined with dNBR analysis (using images from 4 and 24 December 2024 representing pre-fire and post-fire conditions, respectively). The comparison focused on a sample burned rice field area observed on 21 December 2024 in Doi Saket District, Chiang Mai Province, as illustrated in **Fig. 3**. Furthermore, a comparative analysis was performed to determine quantitative performance metrics, including Intersection over Union (IoU), Precision, Recall, F1-score, and Overall Accuracy, as presented in **Table 2**.

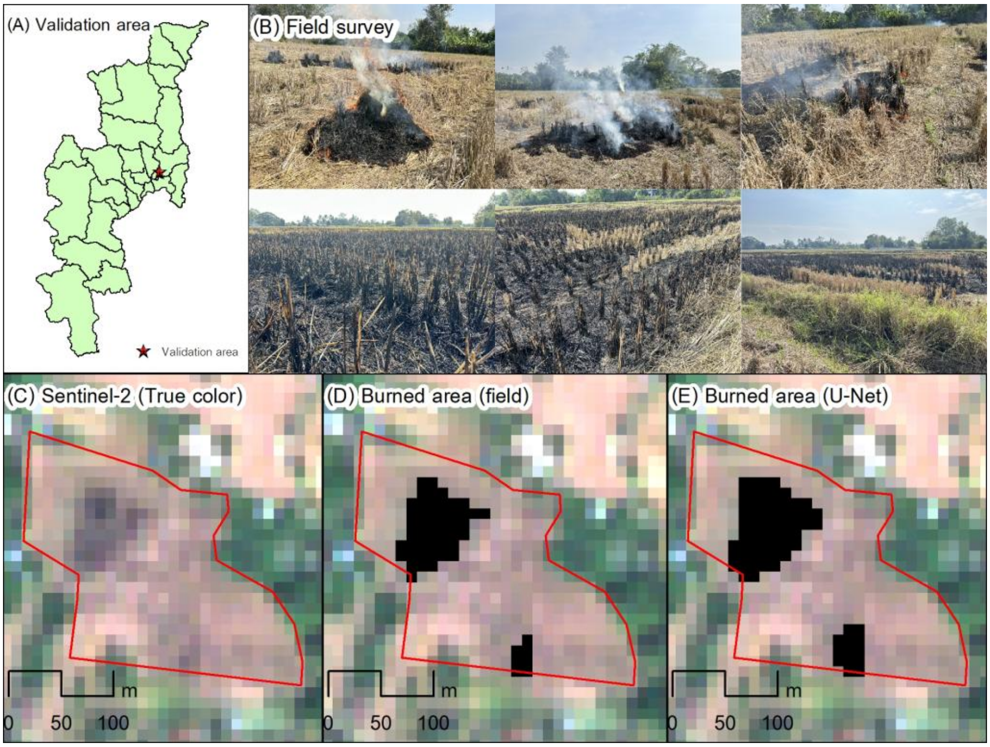


Fig. 3. Validation framework for burned area detection: (A) location of the study site, (B) field survey observations, (C) Sentinel-2 true color imagery, (D) reference burned area from field data, and (E) U-Net segmentation results in rice paddy fields, Doi Saket District, Chiang Mai Province, Thailand.

The deep learning dataset consisted of 2,560 samples, including RGB image tiles and their corresponding binary burn masks derived from Sentinel-2 True Color imagery. The original dataset was divided using a spatial hold-out strategy into training, validation, and test sets to reduce spatial data leakage and improve model generalization. Specifically, the training set contained 896 image tiles and 896 corresponding masks (70%), while the validation set included 256 image tiles and 256 masks (20%), and the test set consisted of 128 image tiles and 128 masks (10%).

To improve model robustness and increase sample diversity, data augmentation was applied only to the training dataset using three techniques: horizontal flipping, vertical flipping, and rotation. After augmentation, the total number of samples used in the deep learning analysis increased to 2,560.

The U-Net model was trained using the Adam optimizer due to its computational efficiency and stability in deep learning applications. The learning rate was set to 0.0001 to ensure stable convergence during training, and a batch size of 16 was used to balance computational efficiency and segmentation performance. To address class imbalance between burned and non-burned pixels, a hybrid loss function combining Binary Cross-Entropy (BCE) and Dice loss was employed to improve the model's ability to accurately detect and delineate small burned areas. The model was trained for 100 epochs while monitoring both training and validation performance to reduce overfitting and ensure convergence.

4. RESULTS

4.1. Model Performance

The results of this study demonstrate that the U-Net model can effectively detect agricultural burned areas in rice paddies from true color satellite imagery (RGB), achieving high performance (IoU = 0.9292; F1-score = 0.9633). As presented in **Table 2**, the model outperformed DeepLabv3+ in the context of smallholder agricultural landscapes in Chiang Mai Province.

Table 2.

Quantitative performance metrics of the U-Net and DeepLabv3+ models.

Model	IoU	Precision	Recall	F1-score	Overall Accuracy
U-Net	0.9292	0.9292	0.9999	0.9633	0.9292
DeepLabv3+	0.8664	0.8664	0.9999	0.9284	0.8664

4.2. Training Behavior of the Model

The training results indicate that the training IoU increased rapidly during the early stages of training and gradually stabilized at a high level in the subsequent epochs, suggesting that the model was able to effectively learn the spatial patterns of burned areas from the training dataset (**Fig. 4**).

Meanwhile, the validation IoU showed a similar increasing trend during the initial training phase and then fluctuated within a relatively stable range throughout the training process. This behavior indicates that the model was capable of transferring the learned knowledge to unseen data, although some fluctuations occurred due to the class imbalance between burned and non-burned areas. For the graph on the right, the training loss decreased continuously and eventually converged, clearly demonstrating the convergence of the model during training.

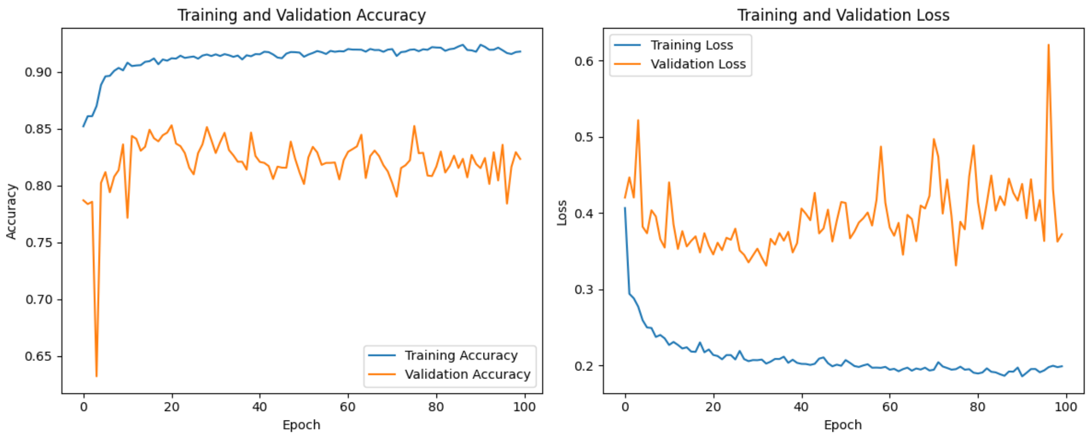


Fig. 4. Training and validation curves of the U-Net model over 100 epochs: Training IoU and Validation IoU (left), and Training loss and Validation loss (right).

In contrast, the validation loss decreased during the early stage and then fluctuated at a slightly higher level than the training loss, which is a common characteristic in spatial image segmentation tasks with imbalanced class distributions. Importantly, no consistent upward trend in validation loss was observed, indicating that the model did not exhibit severe overfitting during training.

Overall, the graphs presented in **Fig. 4** demonstrate that the developed U-Net model exhibited stable training behavior, effectively learned the spatial patterns of burned areas from true color satellite imagery, and showed good generalization capability for detecting burned areas in real agricultural landscapes.

4.3. Spatial Distribution of Agricultural Burning

The classification of burned areas within rice paddies were classified using the U-Net model in five selected locations where rice cultivation covered more than 50% of the area. These sites are distributed across rice-growing regions in Chiang Mai Province, Thailand. The analysis used Sentinel-2 True Color imagery acquired on 24 December 2024 as the input dataset (**Fig. 5**).

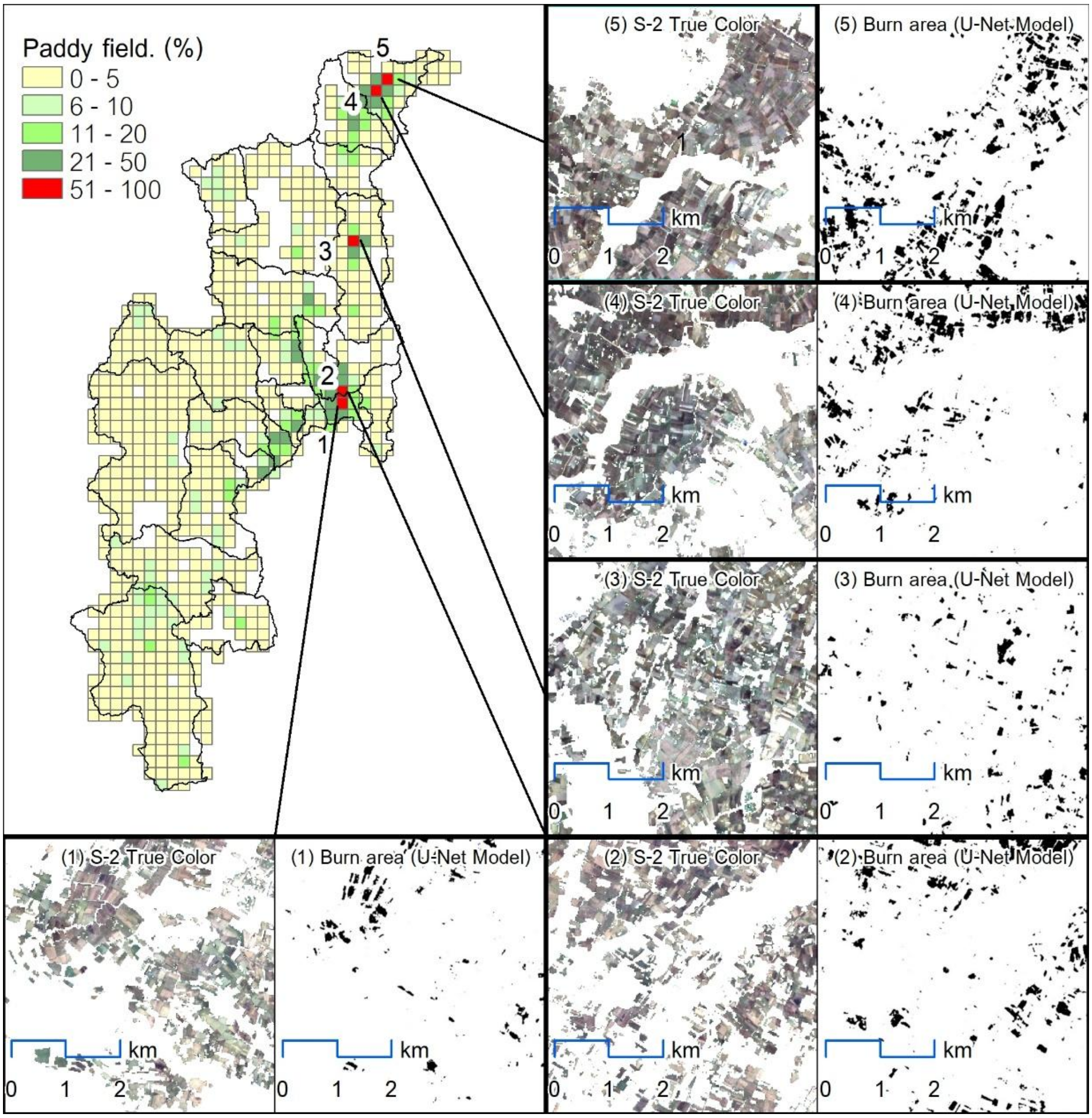


Fig. 5. Locations of the burned area analysis sites where rice paddy density exceeds 50% within Chiang Mai Province, True Color Sentinel-2 imagery acquired on 24 December 2024, and the burned area classification results generated by the U-Net deep learning model.

Regarding the spatial analysis results of burned areas were evaluated using a threshold value ≤ 50 (black color). The results indicate that the proportion of burned pixels accounted for 9.952% of the total image area, equivalent to approximately 2.4880 km². The cultivated region covered 52.441% of the total image area, corresponding to approximately 13.1102 km². Within the cultivated region, burned areas accounted for 18.937%, equivalent to approximately 2.4826 km². These findings illustrate the spatial distribution and extent of burned areas within agricultural fields, highlighting the direct impact of agricultural burning activities on cultivated land, as shown in **Fig. 5**.

5. DISCUSSION

These findings are consistent with previous studies highlighting the potential of Convolutional Neural Network (CNN) models for burned area classification using Sentinel-2 imagery. For example, Tonbul et al. (2023) reported an overall accuracy of 97.53% and demonstrated that deep learning models outperform traditional machine learning approaches. However, the study by Tonbul et al. (2023) utilized multispectral data and spectral indices as input variables. In contrast, the present study demonstrates that even when using only RGB imagery, the model is still capable of effectively learning the spatial characteristics associated with burned area patterns.

When considered together with the study of Samphutthanont (2024), which applied the dNBR index derived from Sentinel-2 imagery to detect small burned areas in rice fields that were not captured by conventional hotspot datasets, the present research further advances the methodology by using dNBR results with a threshold adapted to rice field conditions (dNBR ≥ 0.1) as reference data for training the deep learning model. This approach reflects an integration of spectral index-based analysis and deep learning techniques, representing a transition from traditional rule-based processing toward data-driven learning methods.

Furthermore, when compared with the approach proposed by Pinto et al. (2021), which integrates Sentinel-2 imagery with VIIRS hotspot data to improve temporal completeness, that method is particularly suitable for monitoring large-scale fires at the regional level. However, such approaches may still face limitations in detecting small and fragmented agricultural burned areas, particularly rice residue burning in small-scale rice paddy fields. In contrast, the findings of this study suggest that deep learning approaches may offer greater potential for detecting small burned patches at the field scale, where coarse-resolution hotspot datasets (250–1000 m) are limited in their ability to capture such small burning events.

Regarding model generalization, the use of a spatial hold-out data splitting strategy helps reduce spatial leakage and improves the reliability of model evaluation. This approach is consistent with the recommendations of Anand et al. (2024, 2025), who emphasized the importance of domain adaptation when applying deep learning models to new geographic contexts.

The spatial analysis results indicate that burned areas accounted for 9.952% of the total image area and 18.937% within cultivated regions. These findings are consistent with previous evidence showing that agricultural residue burning plays a significant role in contributing to PM2.5 pollution in Chiang Mai (Junpen et al., 2018; Kamton et al., 2019; Jarernwong et al., 2023; Khamkaew et al., 2016). This study demonstrates a methodological progression from spectral index-based analysis (Samphutthanont, 2024) to deep learning approaches applied to RGB satellite imagery, highlighting the strong potential of such methods for operational monitoring of agricultural burning and supporting effective PM2.5 management at the local scale.

However, several limitations should be acknowledged. First, the dataset size remains relatively limited, consisting of a finite number of image patches derived from selected study areas, which may contribute to the high accuracy observed. Second, the use of masked rice cultivation areas, while beneficial for reducing noise, may reduce classification complexity and limit the model's applicability to more heterogeneous landscapes. Third, the presence of smoke during active burning events may obscure surface features in optical imagery, potentially affecting detection accuracy, and was not explicitly addressed in the current model. Another important limitation relates to the temporal dynamics of agricultural burning. In rice cultivation systems, burned surfaces may change rapidly after burning due to plowing, irrigation, vegetation regrowth, or post-harvest land preparation activities. As a result, burn scars may only remain detectable for a short period of time. Since this study primarily relied on satellite imagery acquired during specific observation periods, some short-duration burning events may not have been fully captured. Incorporating multi-temporal satellite observations could improve the ability to monitor the temporal evolution of agricultural burning and enhance detection reliability.

Overall, the high model performance may partly reflect the simplified classification conditions introduced by masking and homogeneous land cover characteristics. Consequently, the model may require further validation in more complex and heterogeneous environments to ensure its robustness and generalizability.

Future work in this study should explore multi-temporal data integration, cross-regional validation, and the incorporation of atmospheric correction and smoke-robust preprocessing techniques to enhance model performance under varying environmental conditions, particularly in the presence of haze and smoke.

6. CONCLUSIONS

This study developed an integrated approach combining spectral index analysis (dNBR) and deep learning applied to true color satellite imagery to systematically detect agricultural burned areas in Chiang Mai Province, Thailand. The developed U-Net model demonstrated high precision and training stability in detecting small burned areas from true color (RGB) satellite imagery and exhibited stable training performance. In addition, the model showed good generalization capability when applied to new areas, indicating its robustness for operational applications.

Spatial analysis revealed that burned areas were distributed as small fragmented patches within cultivated fields, with a substantial proportion of burned areas occurring within agricultural zones. These findings highlight the important role of agricultural burning in contributing to local PM2.5 concentrations. In conclusion, the application of deep learning techniques to true color satellite imagery provides an effective approach for detecting agricultural burned areas at the field scale. This method offers significant potential as a practical tool for supporting air quality management and haze mitigation, particularly in Chiang Mai Province and other agricultural regions with similar environmental and land-use characteristics.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support provided by the Local Development Project, Chiang Mai Rajabhat University, for the 2025 Fiscal Year. Furthermore, sincere appreciation is extended to AiroTEC, Chiang Mai Rajabhat University, for their assistance and facilities. Their contribution has been instrumental to the successful completion of this research project and the preparation of this research article.

REFERENCES

Anand, A., Imasu, R., Dhaka, S. K., & Patra, P. (2024). Domain adaptation of deep learning segmentation model of small agricultural burn area detection using high-resolution Sentinel-2 observations: A case study of Punjab, India. Preprints. https://doi.org/10.20944/preprints202409.0333.v1

Anand, A., Imasu, R., Dhaka, S., & Patra, P. K. (2025). Domain adaptation and fine-tuning of a deep learning segmentation model of small agricultural burn area detection using high-resolution Sentinel-2 observations: A case study of Punjab, India. Preprints. https://doi.org/10.20944/preprints202501.0563.v1

Andreae, M. O. (2019). Emission of trace gases and aerosols from biomass burning: An updated assessment. Atmospheric Chemistry and Physics, 19(13), 8523–8546. https://doi.org/10.5194/acp-19-8523-2019

Chen, R., Little, R., Mihaylova, L., Delahay, R., & Cox, R. (2019). Wildlife surveillance using deep learning methods. Ecology and Evolution, 9(16), 9453–9466. https://doi.org/10.1002/ece3.5410

Chuvieco, E., & Stoyanov, B. (2023). Global and continental burned area detection from remote sensing: The FireCCI products. In EGU General Assembly 2023.

Filipponi, F. (2018). BAIS2: Burned area index for Sentinel-2. Proceedings, 2(7), 364. https://doi.org/10.3390/ecrs-2-05177

Fornacca, D., Ye, Y., Li, X., & Xiao, W. (2025). Including small fires in global historical burned area products: Promising results from a Landsat-based product. Fire, 8, 422. https://doi.org/10.3390/fire8110422

Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L., & Justice, C. O. (2018). The Collection 6 MODIS burned area mapping algorithm and product. Remote Sensing of Environment, 217, 72–85. https://doi.org/10.1016/j.rse.2018.08.005

Hu, X., Ban, Y., & Nascetti, A. (2021). Uni-temporal multispectral imagery for burned area mapping with deep learning. Remote Sensing, 13(8), 1509. https://doi.org/10.3390/rs13081509

Jarernwong, K., Gheewala, S. H., & Sampattagul, S. (2023). Health impact related to ambient particulate matter exposure as a spatial health risk map: A case study in Chiang Mai, Thailand. Atmosphere, 14(2), 261. https://doi.org/10.3390/atmos14020261

Junpen, A., Pansuk, J., Kamnoet, O., Cheewaphongphan, P., & Garivait, S. (2018). Emission of air pollutants from rice residue open burning in Thailand. Atmosphere, 9(11), 449. https://doi.org/10.3390/atmos9110449

Kamton, R., Satienperakul, K., Yotapakdee, T., & Nunthasen, K. (2019). Haze-related air pollution and impacts on health in Chiang Mai Province. Interdisciplinary Research Journal: Graduate Studies Edition, 8(1), 1–15.

Key, C. H., & Benson, N. C. (2004). Landscape assessment (LA): Sampling and analysis methods (FIREMON Landscape Assessment V4). U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Khamkaew, C., Chantara, S., & Wiriya, W. (2016). Atmospheric PM2.5 and its elemental composition from near-source and receptor sites during open burning season in Chiang Mai, Thailand. International Journal of Environmental Science and Development, 7(6), 436–440. https://doi.org/10.7763/IJESD.2016.V7.815

Knopp, L. (2021). Development of a burned area processor based on Sentinel-2 data using deep learning. Photogrammetrie–Fernerkundung–Geoinformation, 89, 357–358. https://doi.org/10.1007/s41064-021-00177-6

Knopp, L., Wieland, M., Rättich, M., & Martinis, S. (2020). A deep learning approach for burned area segmentation with Sentinel-2 data. Remote Sensing, 12(15), 2422. https://doi.org/10.3390/rs12152422

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 3431–3440). IEEE. https://doi.org/10.1109/CVPR.2015.7298965

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. ISPRS Journal of Photogrammetry and Remote Sensing, 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

Martins, V. S., Roy, D. P., Huang, H., Boschetti, L., Zhang, H. K., & Yan, L. (2022). Deep learning high resolution burned area mapping by transfer learning from Landsat-8 to PlanetScope. Remote Sensing of Environment, 280, 113203. https://doi.org/10.1016/j.rse.2022.113203

Miller, J. D., & Thode, A. E. (2007). Quantifying burn severity in a heterogeneous landscape with a relative version of the delta normalized burn ratio (dNBR). Remote Sensing of Environment, 109(1), 66–80. https://doi.org/10.1016/j.rse.2006.12.006

Office of the Permanent Secretary for Agriculture and Cooperatives, Chiang Mai Province. (2026). Basic information of Chiang Mai Province. https://www.opsmoac.go.th/chiangmai-dwl-files-481891791113

Pinto, M. M., Trigo, R. M., Trigo, I. F., & DaCamara, C. C. (2021). A practical method for high-resolution burned area monitoring using Sentinel-2 and VIIRS. Remote Sensing, 13(9), 1608. https://doi.org/10.3390/rs13091608

Reid, J. S., Hyer, E. J., Johnson, R. S., Holben, B. N., Yokelson, R. J., Zhang, J., Campbell, J. R., Christopher, S. A., Di Girolamo, L., Giglio, L., Holz, R. E., Kearney, C., Miettinen, J., Reid, E. A., Turk, F. J., Wang, J., Xian, P., Zhao, G., Balasubramanian, R., … Liew, S. C. (2013). Observing and understanding the Southeast Asian aerosol system by remote sensing: An initial review and analysis for the Seven Southeast Asian Studies (7SEAS) program. Papers in Natural Resources, 443. http://digitalcommons.unl.edu/natrespapers/443

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

Roteta, E., Bastarrika, A., Padilla, M., Storm, T., & Chuvieco, E. (2019). Development of a Sentinel-2 burned area algorithm: Generation of a small fire database for sub-Saharan Africa. Remote Sensing of Environment, 222, 1–17. https://doi.org/10.1016/j.rse.2018.12.011

Samphutthanont, R. (2024). Assessing agricultural burned areas using dNBR index from Sentinel-2 satellite data in Chiang Mai, Thailand, from 2019 to 2023. Geographia Technica, 19(2), 46–56. https://doi.org/10.21163/GT_2024.192.04

Schroeder, W., Oliva, P., Giglio, L., & Csiszar, I. A. (2014). The new VIIRS 375 m active fire detection data product: Algorithm description and initial assessment. Remote Sensing of Environment, 143, 85–96. https://doi.org/10.1016/j.rse.2013.12.008

Tonbul, H., Yilmaz, E. O., & Kavzoglu, T. (2023). Comparative analysis of deep learning and machine learning models for burned area estimation using Sentinel-2 image: A case study in Muğla-Bodrum, Turkey. In Proceedings of the 10th International Conference on Recent Advances in Air and Space Technologies (RAST 2023) (pp. 1–5). https://doi.org/10.1109/RAST57548.2023.10197926

van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J., & Kasibhatla, P. S. (2017). Global fire emissions estimates during 1997–2016. Earth System Science Data, 9, 697–720. https://doi.org/10.5194/essd-9-697-2017

World Health Organization. (2021). WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization.

Yilmaz, E. O., & Kavzoglu, T. (2024). Burned area detection with Sentinel-2A data: Using deep learning techniques with explainable artificial intelligence. In ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (Vol. X-5-2024). ISPRS TC V Mid-term Symposium Insight to Foresight via Geospatial Technologies, Manila, Philippines.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, 5(4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307

A SHEAR-STRENGTH-DERIVED LITHOLOGICAL WEAKNESS INDEX FOR GIS-BASED LANDSLIDE SUSCEPTIBILITY MODELLING IN MOUNTAINOUS ROAD CORRIDORS OF NORTHERN VIETNAM

Thanh Long NGUYEN [1], ***Thi Ngoc Ha TRAN*** [1], ***Quoc Lap KIEU*** [1*]
& Dieu Trinh NGUYEN [2]

DOI: 10.21163/GT_2026.212.10

ABSTRACT

Lithology is commonly incorporated into landslide susceptibility models as a categorical map variable, although this representation may not adequately express the mechanical behaviour of slope-forming materials. To improve the physical interpretability of lithological information without substantially changing the conventional GIS-based modelling workflow, this study introduces a shear-strength-informed lithological standardization approach for landslide susceptibility assessment in a mountainous road-corridor environment. Conventional lithological groups were converted into a Lithological Weakness Index (LWI) derived from laboratory-measured cohesion and internal friction angle from 58 samples, and the index was integrated into a Random Forest modelling framework. The approach was tested along National Road QL279 in Lao Cai Province, northern Vietnam, using topographic, environmental, rainfall, geological, hydrological, infrastructural and landslide-inventory data. The RF-LWI model achieved an AUC-ROC of 0.936, accuracy of 0.865 and F1-score of 0.847 on the independent testing dataset, with a mean cross-validated AUC-ROC of 0.943. High and very high susceptibility classes covered 52.71% of the study area. Compared with the model using conventional reclassified lithology, the LWI-based model produced only a modest improvement in predictive performance, but it provided a more mechanically meaningful and interpretable representation of lithological control. The results indicate that shear-strength-derived lithological standardization is useful mainly as a refinement of the geological predictor rather than as a source of large accuracy gains. This approach may support more transparent GIS-based landslide susceptibility modelling in mountainous road corridors where lithological heterogeneity and road-related slope disturbance interact.

Keywords: *Landslide susceptibility; Lithological Weakness Index; Shear strength; Random Forest; GIS-based modelling.*

1. INTRODUCTION

Landslides are common geomorphic hazards in mountainous regions, where slope instability is controlled by interactions among relief, lithology, rainfall, drainage, vegetation cover and human disturbance. Their consequences are particularly significant along mountain road corridors, because slope failures may disrupt transport, damage infrastructure, increase maintenance costs and reduce access to local communities. In such settings, landslide susceptibility mapping provides a spatial basis for identifying terrain sectors with different propensities to failure and for supporting road maintenance, land-use planning and disaster-risk reduction (Guzzetti et al., 1999; Fell et al., 2008; Van Westen et al., 2008; Reichenbach et al., 2018).

[1] *Thai Nguyen University of Sciences, Thai Nguyen Province, Vietnam, longgeoj@gmail.com (TLN), hattn@tnus.edu.vn (TNHT), lapkq@tnus.edu.vn (QLK)*
[2] *Vietnam Academy of Science and Technology, Hanoi City, Vietnam, nguyendieutrinh70@gmail.com (DTN)*
[*] *Corresponding author: lapkq@tnus.edu.vn*

The development of remote sensing, GIS and spatial modelling has improved the representation of landslide-conditioning factors at regional and corridor scales. Digital elevation models, satellite imagery, land-cover datasets, gridded rainfall products and geological maps are frequently combined to describe topographic, hydrological, environmental, lithological and anthropogenic controls on slope instability (Drusch et al., 2012; Funk et al., 2015; Brown et al., 2022). Recent studies have applied GIS, remote sensing, multi-criteria evaluation, machine learning and spatial analysis to landslide susceptibility, landslide vulnerability, flash-flood hazard and disaster-risk mapping in diverse mountain environments (Zhanabayev et al., 2024; Dahmani et al., 2025; Edial et al., 2025). These studies demonstrate the importance of spatially explicit and technically reproducible approaches for natural-hazard assessment.

Among data-driven approaches, machine-learning algorithms have been widely used in landslide susceptibility modelling because they can represent nonlinear relationships between landslide occurrence and multiple conditioning variables. Random Forest is among the most frequently applied algorithms, as it can handle mixed data types, reduce sensitivity to noisy predictors and provide variable-importance measures that support geomorphological interpretation (Breiman, 2001; Catani et al., 2013; Chen et al., 2018; Merghadi et al., 2020; Ado et al., 2022). However, the reliability of machine-learning susceptibility models depends not only on the algorithm, but also on the quality of the landslide inventory, the definition of non-landslide samples, the spatial resolution and scale of input factors, the treatment of class imbalance and the validation strategy (Conoscenti et al., 2016; Steger et al., 2017; Tanyu et al., 2021; Gu et al., 2023; Gu et al., 2024; Kanwar et al., 2025). These issues are especially relevant in road-corridor studies, where landslide and non-landslide samples may occur within narrow terrain belts affected by slope cutting, drainage modification and short-distance changes in lithology.

One remaining issue concerns how lithological information is represented in susceptibility models. Lithology is commonly used as a categorical predictor derived from geological maps. Although this representation is practical, it does not always express the mechanical behaviour of slope-forming materials. Geological units with different names may have comparable strength properties after weathering, whereas units assigned to the same lithological class may differ in cohesion, internal friction angle, discontinuity conditions, weathering grade and permeability. As a result, the direct use of nominal lithological classes can limit the physical interpretation of geological control in statistical or machine-learning susceptibility models (Bunn et al., 2020; Segoni et al., 2020; Yu et al., 2021).

A more interpretable lithological predictor should be linked, where data allow, to material properties that influence slope resistance. Cohesion and internal friction angle are particularly relevant because they are directly associated with shear strength. Cohesion reflects the bonding or cementation component of shear resistance, whereas the internal friction angle represents frictional resistance along potential failure surfaces. Previous studies have emphasized that geological and rock–soil factors can improve landslide susceptibility assessment when they are parameterized in relation to material behaviour rather than used only as map-based categories (Bunn et al., 2020; Segoni et al., 2020; Yu et al., 2021).

Nevertheless, laboratory-derived shear-strength parameters are still rarely incorporated into GIS-based susceptibility models as a standardized lithological variable. This gap is important because a shear-strength-informed lithological variable may improve the interpretability of the geological predictor, even if the resulting gain in predictive accuracy is modest. This issue is particularly relevant for mountainous road corridors, where lithological heterogeneity, slope cutting, modified drainage and rainfall-induced processes interact within narrow spatial zones.

Northern Vietnam provides a suitable setting for examining this problem. The mountainous part of the country is characterized by steep relief, deeply dissected valleys, monsoonal rainfall, complex geological conditions and expanding transport infrastructure. Previous studies have shown the usefulness of GIS, remote sensing and spatial modelling for landslide, flash-flood and disaster-risk assessment in this region (Bui et al., 2011; Kieu and Tran, 2021; Kieu and Ngo, 2022; Do et al., 2025; Kieu and Tran, 2025). However, limited attention has been given to whether conventional lithological classes can be refined by using measured shear-strength properties before their integration into machine-learning susceptibility models. In corridor-scale studies, this refinement should be interpreted as a practical group-level standardization of lithological weakness rather than as a complete geotechnical characterization of each geological unit.

This study addresses this issue by proposing a shear-strength-informed lithological standardization approach for GIS-based landslide susceptibility modelling in a mountainous road corridor. The study converts reclassified lithological groups into a Lithological Weakness Index derived from laboratory-measured cohesion and internal friction angle, and then tests whether this standardized lithological variable improves Random Forest susceptibility modelling compared with conventional lithological representation. The proposed index is intended as a transparent proxy for relative lithological weakness and as a refinement of the geological predictor, not as a replacement for field-based engineering-geological investigation. The approach is evaluated along a mountainous road corridor in northern Vietnam using a multi-source spatial database, a landslide inventory, geomechanical samples and repeated model validation. The specific objectives are: (1) to prepare a GIS-based database of landslide-conditioning factors for a mountainous road corridor; (2) to derive a shear-strength-based Lithological Weakness Index for reclassified lithological groups; (3) to develop and validate a Random Forest landslide susceptibility model; and (4) to compare the performance of conventional lithology and shear-strength-standardized lithology. By linking lithological representation to measured material strength, the study aims to improve the physical interpretability of the geological predictor while retaining a transferable GIS-based workflow for landslide susceptibility assessment in mountainous transport corridors.

2. STUDY AREA

The study area is located along National Road QL279 in Lao Cai Province, northern Vietnam (**Fig. 1**), within the mountainous upper Red River region.

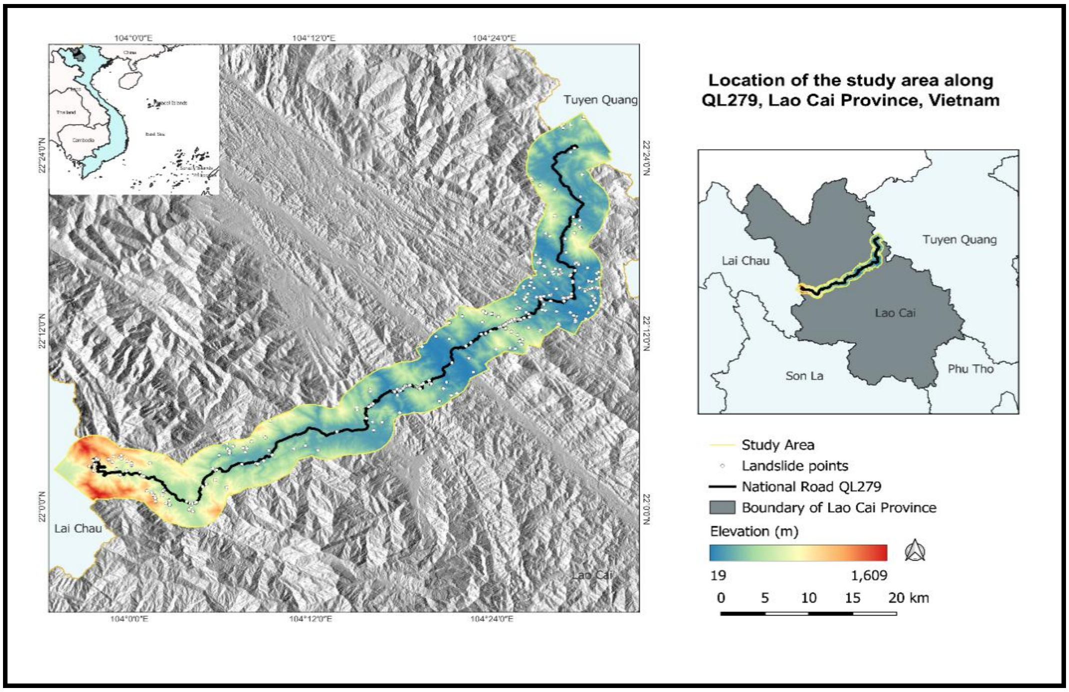


Fig. 1. Location of the study area along National Road QL279, Lao Cai Province, northern Vietnam.

The corridor crosses steep hillslopes, narrow valleys and strongly dissected terrain typical of the northwestern uplands of Vietnam. This road-corridor setting is characterized by short-distance variations in relief, lithology, drainage conditions and human disturbance, making it suitable for examining landslide susceptibility at a corridor scale.

National Road QL279 is an important transport corridor connecting mountainous districts in northern Vietnam. Within Lao Cai Province, it passes through terrain where steep slopes, short drainage lines and heterogeneous slope materials occur over short distances. Landslides along this corridor may interrupt traffic, damage road infrastructure and reduce access to settlements and production areas, particularly during the rainy season. Similar conditions have been reported in other mountainous areas of northern Vietnam, where slope failures are commonly associated with steep topography, concentrated rainfall, weathered materials and human modification of hillslopes (Bui et al., 2011; Kieu and Tran, 2021; Kieu and Ngo, 2022; Do et al., 2025).

The regional relief is shaped by the Hoang Lien Son and Con Voi mountain systems and by the incision of the Red River and its tributaries. Elevation changes rapidly over short distances, producing steep valley sides, high local relief and dense drainage networks. These conditions favour gravitational instability, runoff concentration and toe erosion along valley slopes and road cuts. In the susceptibility model, this terrain setting is represented by DEM-derived variables including elevation, slope, aspect, curvature, stream power index and topographic wetness index, which are widely used to describe topographic and hydrological controls on landslide occurrence (Guzzetti et al., 1999; Van Westen et al., 2008; Reichenbach et al., 2018). These variables are used as spatial conditioning factors and should not be interpreted as deterministic predictors of individual landslide events.

The climate is monsoonal, with rainfall concentrated mainly from late spring to autumn. Seasonal rainfall contributes to slope instability by increasing soil moisture, raising pore-water pressure and reducing the shear resistance of weathered materials. In northern Vietnam, intense and prolonged rainfall has been identified as an important factor in landslides, flash floods and related mountain hazards (Kieu and Tran, 2021; Do et al., 2025; Kieu and Tran, 2025). In this study, rainfall is represented by the maximum 7-day cumulative rainfall derived from CHIRPS data for the 2017–2025 period. Because this rainfall product has a coarser native spatial resolution than the common modelling grid, it is interpreted as a regional hydrometeorological conditioning factor rather than a fine-scale event-triggering variable.

Geologically, the QL279 corridor crosses sedimentary, metasedimentary, metamorphic and intrusive rocks with different weathering characteristics, discontinuity conditions and mechanical resistance. These differences influence the strength of slope-forming materials and their hydrological response during rainfall infiltration. Because lithological changes occur within a narrow road-corridor environment, the area is suitable for testing whether conventional lithological classes can be refined using shear-strength information. This issue is consistent with previous observations that geological variables are more informative in landslide susceptibility modelling when they are parameterized in relation to material behaviour rather than used only as nominal map categories (Bunn et al., 2020; Segoni et al., 2020; Yu et al., 2021). However, the geological information should be interpreted within the limits of its original map scale, even though it is aligned to the common raster grid for GIS overlay and modelling.

Land cover and human disturbance also contribute to spatial variation in slope stability. Vegetation may reduce surface erosion and shallow instability, whereas disturbed surfaces, agricultural land and road-construction zones can modify runoff, infiltration and slope geometry. Road cuts, embankments and drainage modification are particularly relevant in this corridor because they can locally alter slope geometry, runoff pathways and the stability of weathered materials. These factors make the QL279 corridor an appropriate setting for integrating satellite-derived land-cover information with terrain, rainfall, geological and infrastructural variables in a GIS-based susceptibility framework. Overall, the study area provides a representative mountainous road-corridor setting for evaluating the proposed shear-strength-informed lithological standardization approach.

3. DATA AND METHODS

3.1. Research design

This study evaluates whether lithological information can be improved in GIS-based landslide susceptibility modelling by incorporating shear-strength parameters. The workflow included four steps: (1) preparation of the spatial database and landslide inventory; (2) extraction and preprocessing of landslide-conditioning factors; (3) construction of a shear-strength-based Lithological Weakness Index (LWI); and (4) Random Forest modelling, validation and comparison between conventional lithology and LWI-based lithological representation.

All spatial datasets were projected to a common coordinate system, clipped to the study area and aligned to a 12.5 m raster grid defined by the ALOS PALSAR DEM. This grid was used for spatial overlay and modelling consistency. It does not imply that all input datasets have the same original spatial accuracy, especially rainfall data and geological information. Therefore, the 12.5 m grid should be understood as a common computational grid for GIS overlay, sampling and modelling, rather than as an indication of the native spatial resolution, positional accuracy or thematic detail of all predictors.

3.2. Spatial database and conditioning factors

A multi-source spatial database was prepared to represent the main controls on landslide occurrence along the QL279 road corridor. Thirteen conditioning factors were used: elevation, slope, aspect, curvature, stream power index (SPI), topographic wetness index (TWI), normalized difference vegetation index (NDVI), land use/land cover (LULC), maximum 7-day cumulative rainfall, lithology or LWI, distance to roads, distance to rivers and distance to faults. These variables describe terrain morphology, runoff concentration, vegetation condition, land cover, rainfall, geological setting and anthropogenic disturbance, which are widely used in landslide susceptibility studies (Guzzetti et al., 1999; Fell et al., 2008; Van Westen et al., 2008; Reichenbach et al., 2018).

Topographic variables were derived from the ALOS PALSAR DEM. Elevation was used directly, while slope, aspect, curvature, SPI and TWI were calculated from DEM-derived terrain and flow attributes.

Table 1.

Summary of datasets and variables used in the study.

Data source	Product/type	Original resolution/scale	Period	Derived variables
ALOS PALSAR DEM	Digital elevation model	12.5 m	Static	Elevation, slope, aspect, curvature, SPI, TWI
Sentinel-2A	Multispectral imagery	10 m	2017–2025	NDVI
Dynamic World	Land use/land cover product	10 m	2017–2025	LULC
CHIRPS	Gridded precipitation	~5 km	2017–2025	Maximum 7-day cumulative rainfall
Geological/mineral map of Lao Cai Province	Lithological and structural data	1:200,000	Static	Lithology, faults, distance to faults
Road and river networks	Vector line data	Vector	Static/updated where available	Distance to roads, distance to rivers
Google Earth Pro	Historical imagery	Image-dependent	2017–2025	Landslide inventory
Laboratory testing	Shear-strength parameters	58 samples	Field/laboratory campaign	Cohesion, internal friction angle, LWI

NDVI was calculated from Sentinel-2A red and near-infrared bands, and LULC was obtained from the Dynamic World product (Drusch et al., 2012; Brown et al., 2022). Rainfall was represented by the maximum 7-day cumulative rainfall derived from CHIRPS data for 2017–2025.

Because CHIRPS has an original spatial resolution of approximately 5 km, the rainfall raster was resampled only for alignment and was interpreted as a regional hydrometeorological conditioning factor rather than a fine-scale triggering variable (Funk et al., 2015). This treatment avoids assigning unwarranted fine-scale spatial meaning to the resampled rainfall layer.

Geological information was compiled from the geological and mineral map of Lao Cai Province at 1:200,000 scale. The original lithological units were reclassified into seven lithological groups according to lithological similarity, expected mechanical behaviour and the distribution of available shear-strength samples. Fault, road and river vector layers were converted into Euclidean distance rasters. Categorical layers were resampled using the nearest-neighbour method, whereas continuous layers were resampled for alignment with the common modelling grid (**Table 1**). The geological and fault-related variables were therefore used as scale-dependent conditioning factors, and their interpretation remains constrained by the original 1:200,000 map scale.

3.3. Landslide inventory and sample preparation

The landslide inventory was prepared mainly from Google Earth Pro historical imagery for 2017–2025. Landslide scars and disturbed slope surfaces were identified by visual interpretation using morphological and spectral indicators, including bare or recently disturbed surfaces, arcuate or elongated scar forms, downslope accumulation, disrupted vegetation cover and occurrence on steep roadside or valley-side slopes. Multi-temporal imagery was checked where available to reduce confusion between landslide scars and temporary land-cover changes. Such image-based inventories are commonly used in susceptibility modelling where systematic field records are incomplete or spatially inconsistent (Guzzetti et al., 2012; Steger et al., 2017).

Mapped landslide features were converted into landslide-presence samples. In total, 226 landslide samples were used for model training, and 94 samples were reserved for independent testing. A stratified sample split was used so that landslide and non-landslide classes remained balanced in both the training and testing datasets. Accordingly, 226 non-landslide samples were used for training and 94 non-landslide samples were used for independent testing.

Non-landslide samples were selected from areas where no landslides were mapped, excluding water bodies and cells without valid predictor values. To reduce the probability of selecting ambiguous cells close to mapped landslides, a minimum buffer distance of 50 m from the boundary of each mapped landslide polygon was applied before generating non-landslide candidate samples. This distance corresponds to four 12.5 m grid cells and was considered appropriate for reducing boundary uncertainty while retaining sufficient terrain variability within the narrow road-corridor setting. Cells located within mapped landslide polygons or within this 50 m buffer zone were excluded from the non-landslide sampling pool.

The non-landslide sampling was spatially constrained to the same road-corridor modelling extent used for landslide sampling. Candidate non-landslide cells were required to have valid values for all conditioning factors and to fall outside water bodies. No strict matching by slope class or land-cover class was applied, because such matching could artificially reduce the environmental contrast between landslide and non-landslide conditions. However, the candidate pool was kept within the same corridor-scale terrain, lithological and land-cover domain as the landslide samples to ensure that both classes were drawn from comparable spatial conditions. Because absence or pseudo-absence selection can affect susceptibility estimates, the same sampling design was used for both modelling scenarios (Conoscenti et al., 2016; Dou et al., 2020; Gu et al., 2024). This ensured that the comparison between RF-Lithology and RF-LWI reflected the effect of lithological representation rather than differences in sampling strategy.

3.4. Shear-strength-based lithological standardization

The main methodological contribution of this study is the conversion of lithology from a nominal categorical predictor into a shear-strength-informed variable. Laboratory-derived cohesion and internal friction angle from 58 samples were used to characterize the seven reclassified lithological groups. These parameters were selected because they are directly related to shear resistance and are commonly used to describe the mechanical behaviour of slope-forming materials. The number of samples per lithological group ranged from 7 to 9, providing a first-order basis for estimating group-level shear-strength conditions. However, these samples should be interpreted as representative of average lithological-group behaviour rather than as a complete description of local variability in weathering grade, discontinuity density, groundwater condition or rock-mass structure.

Cohesion and internal friction angle were first normalized using min–max normalization:

$$c_{norm} = \frac{c - c_{min}}{c_{max} - c_{min}} \tag{1}$$

$$\phi_{norm} = \frac{\phi - \phi_{min}}{\phi_{max} - \phi_{min}} \tag{2}$$

where c_{norm} and ϕ_{norm} are the normalized cohesion and internal friction angle; $c$and $\phi$are the measured cohesion and internal friction angle; and the subscripts min and max denote the minimum and maximum values in the sample dataset.

A composite Shear Strength Index (SSI) was calculated as:

$$SSI = 0.5c_{norm} + 0.5\phi_{norm} \tag{3}$$

Equal weights were used as a neutral and transparent assumption because cohesion and internal friction angle represent complementary components of shear strength. Cohesion reflects the bonding or cementation component of material resistance, whereas the internal friction angle reflects frictional resistance along potential failure surfaces. In the absence of an independent calibration dataset or prior evidence supporting unequal contributions of these two parameters for all lithological groups, assigning different weights would introduce additional subjectivity.

The Lithological Weakness Index was then defined as:

$$LWI = 1 - SSI \tag{4}$$

The LWI ranges from 0 to 1. Higher values indicate weaker lithological conditions, whereas lower values indicate stronger mechanical conditions. Mean LWI values were assigned to the corresponding lithological groups and rasterized to the 12.5 m grid. This procedure represents lithology as a mechanical weakness gradient rather than only as a nominal geological class, following recent recommendations that geological and rock–soil information should be parameterized in relation to material behaviour when used in susceptibility modelling (Bunn et al., 2020; Segoni et al., 2020; Yu et al., 2021). The LWI is therefore used as a standardized lithological weakness proxy for susceptibility modelling, not as a substitute for detailed site-specific geotechnical investigation.

3.5. Random Forest modelling

Landslide susceptibility was modelled using Random Forest, an ensemble-learning algorithm that builds multiple decision trees from bootstrap samples and random subsets of predictors (Breiman, 2001). Random Forest was selected because it can model nonlinear relationships, handle mixed predictor types and provide variable-importance measures. It has been widely used in landslide susceptibility studies and has shown stable performance in complex terrain (Catani et al., 2013; Chen et al., 2018; Merghadi et al., 2020; Ado et al., 2022).

The model was configured with 300 trees. At each split, the number of candidate predictors was determined using the square-root rule. A balanced subsampling strategy was applied to reduce the effect of class imbalance. Two model scenarios were developed using the same training and testing samples and the same Random Forest configuration: (1) RF-Lithology, in which lithology was represented by the seven reclassified lithological groups; (2) RF-LWI, in which the lithological predictor was replaced by the shear-strength-derived LWI. Using identical samples, predictors other than lithology, and model settings allowed the effect of lithological representation to be isolated as much as possible.

The RF-LWI model produced a continuous landslide susceptibility probability. For map production, this output was classified into five susceptibility levels: very low, low, moderate, high and very high, using the natural breaks method. The same classification procedure was used consistently for susceptibility mapping and area statistics. The resulting classes represent relative susceptibility levels within the study area and should not be interpreted as deterministic hazard classes or as estimates of landslide timing or magnitude.

3.6. Model validation and comparison

Model performance was evaluated using the independent testing dataset and repeated stratified cross-validation. The confusion matrix was used to derive accuracy, precision, recall and F1-score. The area under the receiver operating characteristic curve (AUC-ROC) was used as a threshold-independent measure of discrimination between landslide and non-landslide samples (Chung and Fabbri, 2003; Fawcett, 2006; Reichenbach et al., 2018).

To assess model stability, repeated stratified 5-fold cross-validation with 10 repetitions was applied. Mean values, standard deviations and 95% confidence intervals were calculated for AUC-ROC, accuracy, precision, recall, F1-score and out-of-bag score. The 95% confidence interval of AUC on the independent testing dataset was estimated by bootstrap resampling with 1,000 iterations.

The RF-Lithology and RF-LWI scenarios were compared using the same validation procedure. Paired t-tests and Wilcoxon signed-rank tests were applied to the repeated validation results to evaluate whether the difference between the two lithological representations was statistically meaningful. These paired tests were based on 50 paired validation results obtained from the repeated stratified 5-fold cross-validation procedure, calculated as 5 folds × 10 repetitions. Each pair compared RF-Lithology and RF-LWI under the same fold and repetition, thereby reducing the influence of random data partitioning on the model comparison.

Variable importance was examined using Gini importance and permutation importance. Gini importance measures the average decrease in node impurity, whereas permutation importance measures the decrease in model performance after randomly permuting each predictor. Using both measures supports a more cautious interpretation of predictor contribution, particularly where conditioning factors may be spatially correlated.

Finally, the classified RF-LWI susceptibility map was evaluated by overlaying landslide samples on susceptibility classes and by constructing success-rate and prediction-rate curves. Area statistics were calculated for each susceptibility class to quantify the spatial distribution of landslide-prone terrain along the road corridor. The overlay and rate-curve analyses were used as complementary spatial validation procedures, while the independent testing dataset and repeated cross-validation remained the primary basis for quantitative model assessment.

4. RESULTS

4.1. Lithological weakness derived from shear-strength standardization

The shear-strength-based standardization revealed clear differences in relative mechanical weakness among the seven reclassified lithological groups. The mean Lithological Weakness Index (LWI) ranged from 0.406 to 0.858, indicating that the lithological groups are not equivalent in terms of shear-strength-related behaviour. The highest mean LWI was obtained for the coarse clastic/red-bed sequence, whereas the lowest value was observed for the carbonate-bearing metasedimentary sequence. Intermediate LWI values were found in the fine clastic/coal-bearing sedimentary sequence, metasedimentary schist–quartzite sequence, high-grade metamorphic gneiss–amphibolite sequence, felsic intrusive–subvolcanic association and mafic to ultramafic intrusive association.

The high LWI value of the coarse clastic/red-bed sequence mainly reflects its lower mean cohesion relative to the other groups. In contrast, the carbonate-bearing metasedimentary sequence shows the lowest LWI, indicating comparatively stronger lithological conditions within the standardized dataset. Because each lithological group was represented by 7–9 shear-strength samples, these values should be interpreted as group-level indicators of relative lithological weakness rather than complete geotechnical descriptions of local rock-mass conditions. The results are summarized in **Table 2.**

Table 2.

Shear-strength-based Lithological Weakness Index values of reclassified lithological groups.

Lithological group	Reclassified lithological group	n	Mean cohesion (kPa)	Mean friction angle (°)	Mean LWI	LWI range
1	Carbonate-bearing metasedimentary sequence	8	26.128	31.240	0.406	0.362–0.456
2	Fine clastic / coal-bearing sedimentary sequence	8	30.129	30.598	0.483	0.285–0.620
3	Coarse clastic / red-bed sequence	9	15.816	30.487	0.858	0.742–0.949
4	Medium-grade metasedimentary schist–quartzite sequence	8	26.101	30.889	0.501	0.320–0.858
5	High-grade metamorphic gneiss–amphibolite sequence	9	29.128	30.596	0.507	0.326–0.798
6	Felsic intrusive–subvolcanic association	7	25.304	30.389	0.655	0.547–0.786
7	Mafic to ultramafic intrusive association	9	26.729	30.458	0.602	0.489–0.767

4.2. Predictive performance of the RF-LWI model

The RF-LWI model produced satisfactory predictive performance on the independent testing dataset. The confusion matrix included 94 true negatives, 4 false positives, 22 false negatives and 72 true positives. The model achieved an AUC-ROC of 0.936, with a 95% confidence interval of 0.900–0.966. Accuracy, precision, recall and F1-score were 0.865, 0.947, 0.766 and 0.847, respectively, and the OOB score was 0.879 (**Table 3**). These results indicate good discrimination and high precision, although the recall value shows that some landslide samples were not detected at the selected classification threshold.

Table 3.

Confusion matrix and performance metrics of the RF-LWI model.

Model	TN	FP	FN	TP	AUC-ROC	AUC 95% CI	Accuracy	Precision	Recall	F1-score	OOB score
RF-LWI	94	4	22	72	0.936	0.900–0.966	0.865	0.947	0.766	0.847	0.879

The validation results are shown graphically in **Fig. 2**. The ROC curve confirms good discrimination between landslide and non-landslide samples. The confusion matrix also indicates that the model produced few false positives, although a number of landslide samples were not detected at the selected threshold.

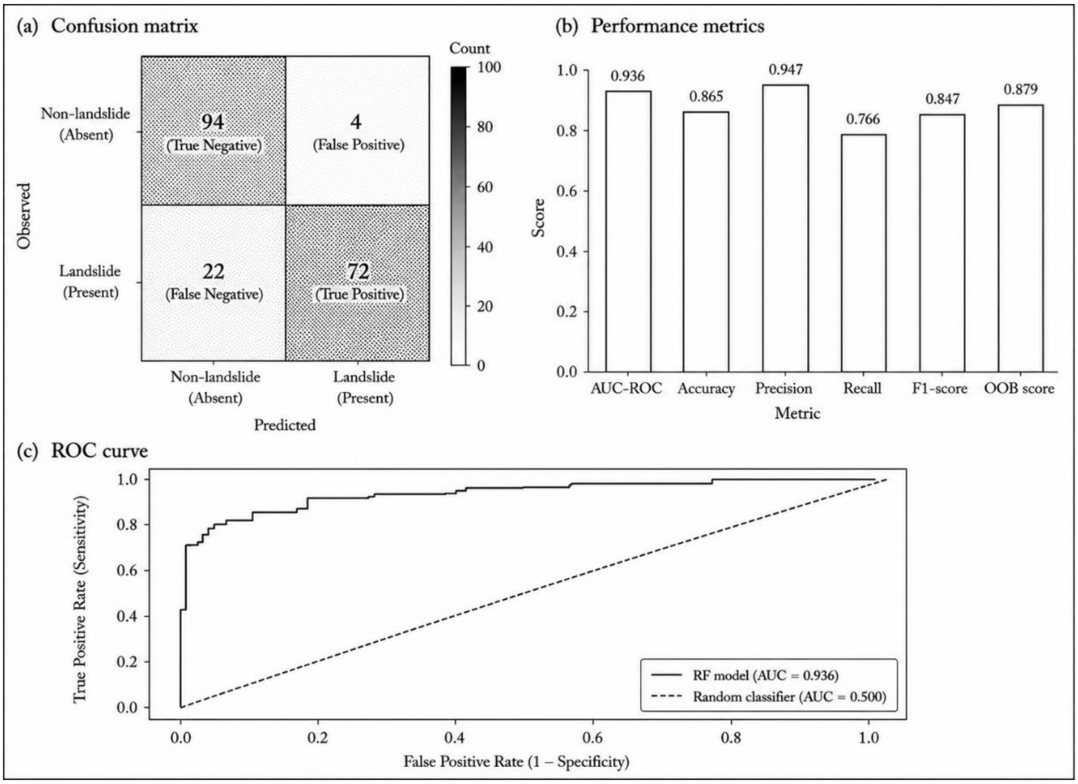


Fig. 2. Validation results of the RF-LWI model, including (a) confusion matrix, (b) performance metrics and (c) ROC curve.

Repeated stratified 5-fold cross-validation with 10 repetitions confirmed the stability of the RF-LWI model. The mean cross-validated AUC-ROC was 0.943, with a standard deviation of 0.020 and a 95% confidence interval of 0.901–0.974. Mean accuracy, precision, recall and F1-score were 0.884, 0.942, 0.823 and 0.877, respectively, while the mean OOB score was 0.883 (**Table 4**).

Table 4.
Repeated stratified 5-fold cross-validation results of the RF-LWI model.

Metric	**Mean**	**SD**	**95% CI**
AUC-ROC	0.943	0.020	0.901–0.974
Accuracy	0.884	0.030	0.821–0.942
Precision	0.942	0.037	0.875–1.000
Recall	0.823	0.049	0.739–0.911
F1-score	0.877	0.032	0.812–0.939
OOB score	0.883	0.008	0.869–0.896

Note: The cross-validation results were obtained from 50 validation runs, calculated as 5 folds × 10 repetitions.

4.3. Effect of shear-strength-based lithological standardization

The RF-LWI and RF-Lithology models were compared to evaluate the effect of replacing conventional reclassified lithology with the shear-strength-derived LWI. On the independent testing dataset, the RF-LWI model showed slightly higher values for most performance metrics. AUC-ROC increased from 0.935 to 0.936, accuracy from 0.849 to 0.865, precision from 0.933 to 0.947, recall from 0.745 to 0.766 and F1-score from 0.828 to 0.847. The number of false positives decreased from 5 to 4, and false negatives decreased from 24 to 22 (**Table 5**). These differences indicate a modest improvement rather than a substantial change in predictive performance.

Table 5.

Comparison of Random Forest model performance using reclassified lithology and shear-strength-derived Lithological Weakness Index.

Model	Lithological variable	TN	FP	FN	TP	AUC-ROC	Accuracy	Precision	Recall	F1-score
RF-Lithology	Reclassified lithology	93	5	24	70	0.935	0.849	0.933	0.745	0.828
RF-LWI	Shear-strength-derived LWI	94	4	22	72	0.936	0.865	0.947	0.766	0.847

The repeated validation results were also compared statistically. The AUC-ROC difference between RF-LWI and RF-Lithology was significant according to both the paired t-test and the Wilcoxon signed-rank test. In contrast, the F1-score difference was not statistically significant. These results indicate that the LWI-based representation improved model discrimination, while the overall classification improvement remained limited. The statistical comparison is presented in **Table 6**.

Table 6.

Statistical comparison between RF-Lithology and RF-LWI models based on repeated validation results.

Metric	Statistical test	Difference RF-LWI − RF-Lithology	p-value	Result
AUC-ROC	Paired t-test	0.0065	0.043	Significant
AUC-ROC	Wilcoxon signed-rank test	0.0065	0.035	Significant
F1-score	Paired t-test	−0.0022	0.322	Not significant
F1-score	Wilcoxon signed-rank test	−0.0022	0.317	Not significant

Note: The paired tests were based on 50 paired validation results obtained from repeated stratified 5-fold cross-validation with 10 repetitions.

4.4. Landslide susceptibility pattern

The RF-LWI model was used to generate the final landslide susceptibility map. The predicted susceptibility values were classified into five levels: very low, low, moderate, high and very high. The map shows that high and very high susceptibility classes are mainly distributed along steep valley-side slopes, road-adjacent sectors and areas where weak lithological conditions coincide with drainage concentration (**Fig. 3**).

The area statistics indicate that the very high susceptibility class occupies 215.180 km², corresponding to 35.08% of the study area. The high susceptibility class covers 108.108 km², or 17.63%. Together, the high and very high susceptibility classes account for 52.71% of the study area. The moderate, low and very low classes account for 14.84%, 26.72% and 5.73%, respectively. These statistics are presented in **Table 7**.

Table 7.
Area statistics of landslide susceptibility classes.

Class	Susceptibility level	Pixel count	Area (km²)	Area (%)
1	Very low	224,779	35.122	5.73
2	Low	1,049,006	163.907	26.72
3	Moderate	582,557	91.025	14.84
4	High	691,890	108.108	17.63
5	Very high	1,377,149	215.180	35.08

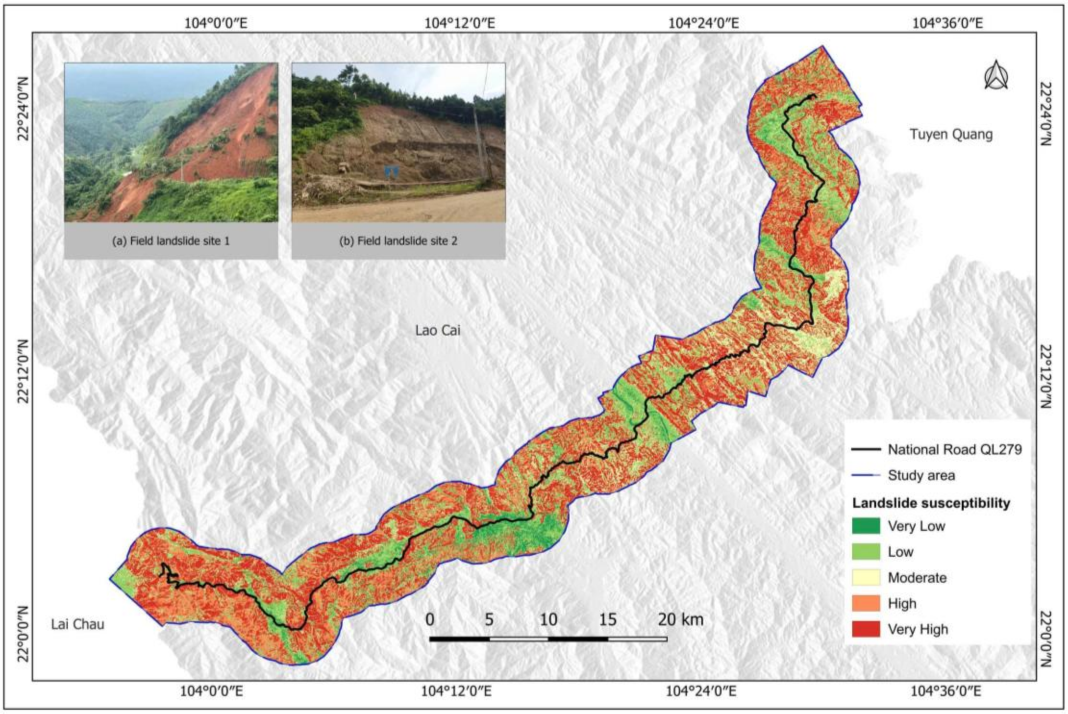


Fig. 3. Landslide susceptibility map generated by the RF-LWI model.

4.5. Spatial validation and variable importance

The spatial validation was conducted by overlaying the training and testing landslide samples on the RF-LWI susceptibility map. Most landslide samples are located within the high and very high susceptibility classes, indicating spatial consistency between the model output and the observed landslide distribution (**Fig. 4**).

The success-rate and prediction-rate curves provide additional information on the distribution of landslide samples across susceptibility values. The curves show that landslide samples are concentrated within the upper susceptibility ranges. The success-rate curve reflects the fit to the training samples, whereas the prediction-rate curve reflects model performance on the independent testing samples (**Fig. 5**).

Variable-importance analysis indicates that slope was the most influential predictor in the RF-LWI model. Other relevant variables included distance to roads, rainfall, LWI, distance to rivers and selected terrain indices. The Gini and permutation importance measures produced broadly consistent patterns, although the relative ranking of some variables differed (**Fig. 6**).

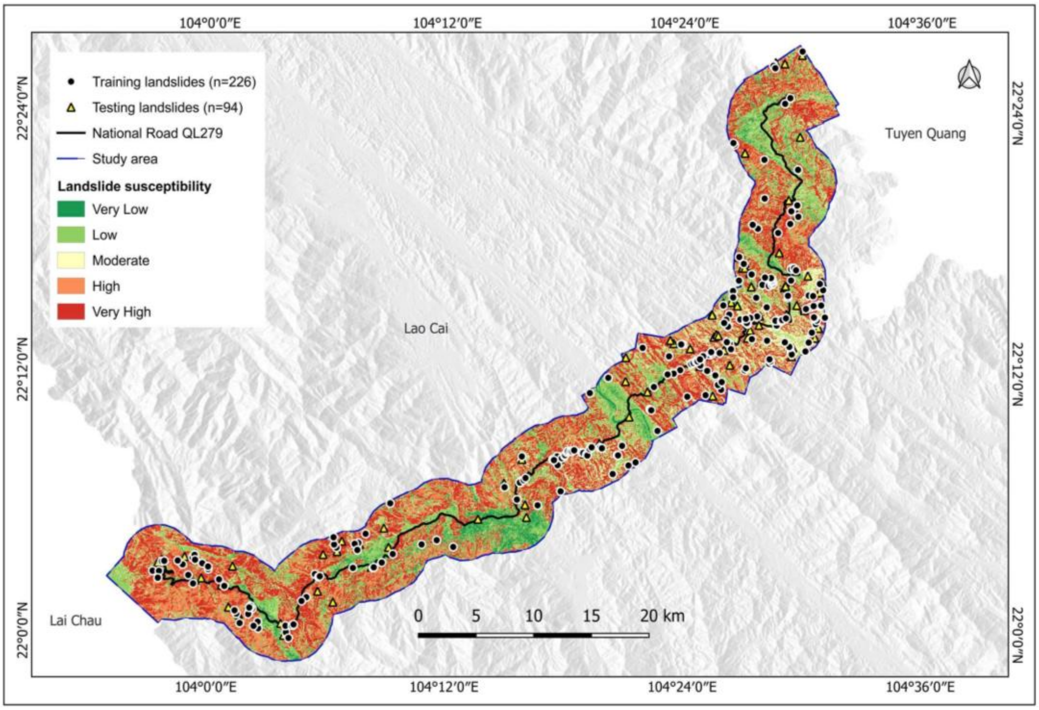


Fig. 4. Spatial validation of the landslide susceptibility map using observed landslide inventory points. Training and testing samples are overlaid on the susceptibility map.

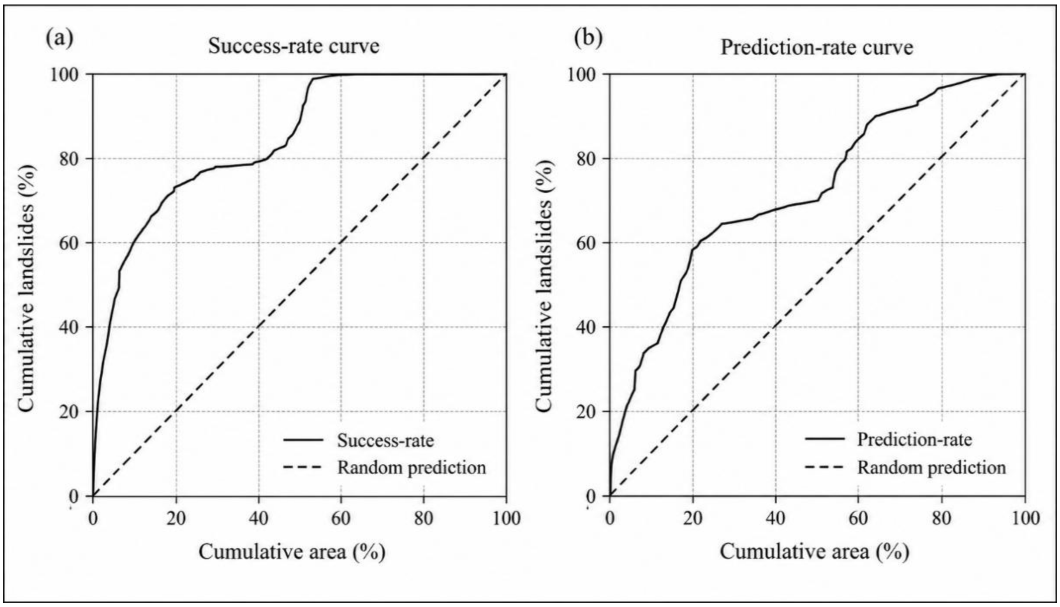


Fig. 5. Success-rate and prediction-rate curves of the landslide susceptibility model: (a) success-rate curve derived from training landslide samples and (b) prediction-rate curve derived from independent testing samples.

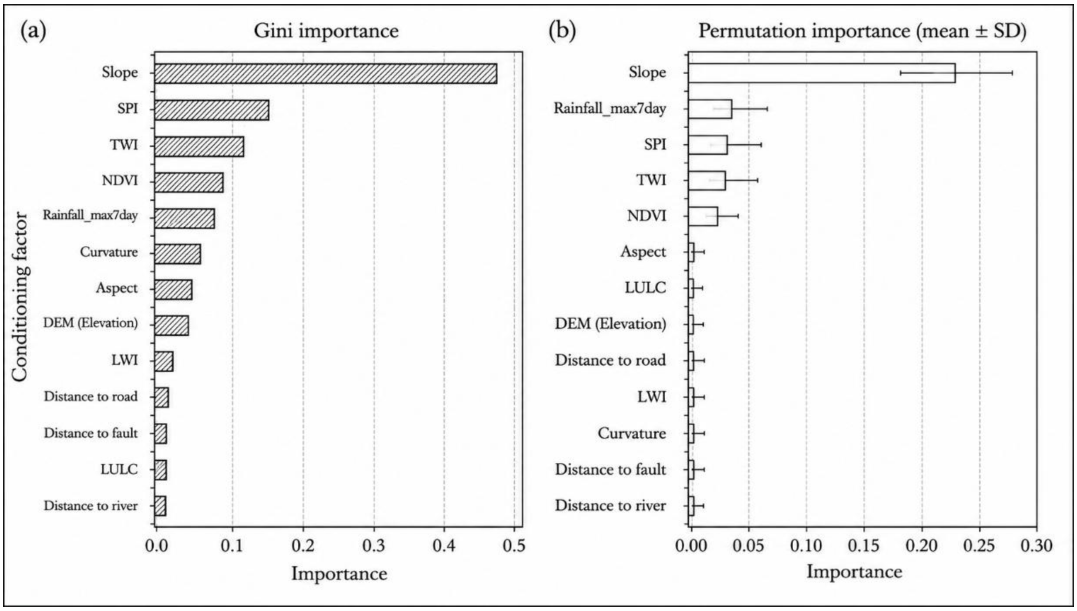


Fig. 6. Variable importance of conditioning factors: (a) Gini importance derived from the Random Forest model and (b) permutation importance.

Overall, the results show that the RF-LWI model produced stable predictive performance and a spatially coherent susceptibility pattern. The shear-strength-derived LWI provided a small but statistically detectable improvement in discrimination and improved the interpretability of the lithological predictor, while the overall classification improvement remained limited.

5. DISCUSSION

The results indicate that shear-strength-informed lithological standardization provides a useful refinement for representing lithology in GIS-based landslide susceptibility modelling. By deriving a Lithological Weakness Index (LWI) from cohesion and internal friction angle, lithology was expressed as a continuous mechanical weakness gradient rather than only as a nominal geological class. This approach is relevant because mapped lithological units do not always reflect the mechanical behaviour of weathered slope-forming materials, particularly in structurally complex mountainous terrain. The result is consistent with previous studies emphasizing that geological and rock–soil information should be parameterized in relation to material properties when used in susceptibility modelling (Bunn et al., 2020; Segoni et al., 2020; Yu et al., 2021). However, the LWI should be interpreted as a group-level proxy for relative lithological weakness, not as a complete geotechnical characterization of each geological unit.

The RF-LWI model produced satisfactory predictive performance, with an independent-test AUC-ROC of 0.936 and a mean cross-validated AUC-ROC of 0.943. These values are comparable with previous GIS- and machine-learning-based landslide susceptibility studies, in which Random Forest has been shown to perform well in complex terrain because of its capacity to model nonlinear relationships among conditioning factors (Catani et al., 2013; Chen et al., 2018; Merghadi et al., 2020; Ado et al., 2022).

The high precision and lower recall indicate that the model produced relatively few false positive classifications, but still missed part of the observed landslide class. This trade-off is relevant for road-corridor applications because field inspection and maintenance planning should consider not only very high susceptibility sectors, but also areas where landslides may be under-detected at the selected classification threshold.

The comparison between RF-Lithology and RF-LWI shows that the use of LWI produced a modest improvement. The increase in AUC-ROC was statistically significant in repeated validation, whereas the difference in F1-score was not significant. Therefore, the main contribution of LWI should be interpreted as improving model discrimination and geological interpretability rather than producing a large increase in overall classification accuracy. This interpretation is consistent with the paired validation design, in which the two models were compared using the same samples, predictors other than lithology and Random Forest settings. It also reflects the multi-factor nature of landslide occurrence along mountainous road corridors, where slope, rainfall, drainage, road-related disturbance and lithological weakness interact rather than act independently.

The spatial pattern of susceptibility is consistent with the geomorphological setting of the QL279 corridor. High and very high susceptibility classes are concentrated mainly along steep valley-side slopes, road-adjacent sectors and areas where weak lithological conditions coincide with drainage concentration. Similar controls have been reported in northern Vietnam and other mountainous environments, where slope gradient, rainfall, weathered materials and human modification of hillslopes are important contributors to slope instability (Bui et al., 2011; Kieu and Tran, 2021; Kieu and Ngo, 2022; Do et al., 2025). The overlay of landslide samples on the susceptibility map and the success-rate and prediction-rate curves further support the spatial consistency of the model output. Because the susceptibility classes were derived from modelled probability values using the natural breaks method, they should be interpreted as relative susceptibility levels within the study area rather than deterministic hazard or risk classes, which would require explicit information on landslide timing, magnitude and exposure (Fell et al., 2008; Van Westen et al., 2008; Reichenbach et al., 2018).

Variable-importance results confirm the dominant role of slope in the RF-LWI model, which is reasonable because slope controls the gravitational component acting on hillslope materials and influences runoff and erosion processes. Distance to roads, rainfall, LWI, distance to rivers and terrain indices also contributed to prediction. The role of distance to roads reflects the influence of slope cutting, embankments and drainage modification in road-corridor environments, while rainfall and distance to rivers represent hydrological controls. The contribution of LWI indicates that lithological weakness is relevant, but it does not dominate the model. Thus, shear-strength-based lithological standardization should be considered a refinement of the geological component rather than a replacement for topographic, hydrological and infrastructural predictors.

Several limitations should be acknowledged. The landslide inventory was compiled mainly from Google Earth Pro historical imagery; therefore, small, old or vegetation-covered landslides may have been omitted, which can affect model training and validation (Guzzetti et al., 2012; Steger et al., 2017). The selection of non-landslide samples is another source of uncertainty because different absence-sampling strategies can influence model performance and prediction patterns (Conoscenti et al., 2016; Dou et al., 2020; Gu et al., 2024). Although this study applied a 50 m exclusion buffer around mapped landslide polygons and used the same balanced sampling design for both model scenarios, pseudo-absence selection cannot completely eliminate uncertainty in areas where unmapped, old or dormant landslides may exist. In addition, the 58 shear-strength samples provide a first-order basis for calculating LWI across the seven lithological groups, but they cannot fully represent local variations in weathering grade, discontinuity density, groundwater conditions or rock-mass structure. Similarly, the use of a common 12.5 m raster grid improves spatial overlay and modelling consistency, but it does not remove the original scale limitations of the CHIRPS rainfall data or the 1:200,000 geological map. These scale constraints should be considered when interpreting fine-scale susceptibility patterns.

Future work should test the LWI approach in other mountainous road corridors and geological settings. Additional engineering-geological data, including weathering grade, discontinuity characteristics, groundwater conditions and rock-mass quality, would help refine the lithological weakness representation. Further research should also examine alternative weighting schemes for cohesion and internal friction angle, compare the approach with other machine-learning algorithms, and integrate temporal rainfall information and road-exposure data to support more operational landslide-risk assessment.

6. CONCLUSIONS

This study proposed a shear-strength-informed lithological standardization approach for GIS-based landslide susceptibility modelling along a mountainous road corridor in northern Vietnam. Reclassified lithological groups were converted into a Lithological Weakness Index (LWI) using laboratory-measured cohesion and internal friction angle and integrated into a Random Forest model. The RF-LWI model showed stable predictive performance, with an independent-test AUC-ROC of 0.936 and a mean cross-validated AUC-ROC of 0.943; high and very high susceptibility classes covered 52.71% of the study area, mainly along steep valley-side slopes, road-adjacent sectors and areas where weak lithological conditions coincide with drainage concentration. Compared with conventional lithological representation, the LWI-based model produced a modest improvement in discrimination, while threshold-dependent classification improvement remained limited. The main contribution of the approach is therefore the improved physical interpretability of the lithological predictor rather than a large increase in predictive accuracy. The resulting map can support preliminary identification of road sections requiring field inspection, drainage maintenance or slope-stabilization assessment, but the susceptibility classes should be interpreted as relative levels within the study area rather than deterministic hazard or risk classes. Future work should test the transferability of the LWI approach in other geological settings and incorporate additional engineering-geological information, including weathering grade, discontinuity characteristics, groundwater conditions and rock-mass quality.

ACKNOWLEDGEMENTS
This research was supported by Thai Nguyen University, Vietnam, under Grant No. ĐH2026-TN01-12. The authors gratefully acknowledge this support and the data provided for this study.

REFERENCES

Ado, M., Khwairakpam, A., Maji, A.K., Jasińska, E., Gono, R., Leonowicz, Z. and Jasiński, M. (2022) Landslide susceptibility mapping using machine learning: a literature survey. Remote Sensing, 14(13), 3029. DOI: 10.3390/rs14133029.

Breiman, L. (2001) Random forests. Machine Learning, 45(1), 5–32. DOI: 10.1023/A:1010933404324.

Brown, C.F., Brumby, S.P., Guzder-Williams, B. et al. (2022) Dynamic World, near real-time global 10 m land use land cover mapping. Scientific Data, 9, 251. DOI: 10.1038/s41597-022-01307-4.

Bui, D.T., Lofman, O., Revhaug, I. and Dick, O. (2011) Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. Natural Hazards, 59(3), 1413–1444. DOI: 10.1007/s11069-011-9844-2.

Bunn, M., Leshchinsky, B. and Olsen, M.J. (2020) Geologic trends in shear strength properties inferred through three-dimensional back analysis of landslide inventories. Journal of Geophysical Research: Earth Surface, 125(9), e2019JF005461. DOI: 10.1029/2019JF005461.

Catani, F., Lagomarsino, D., Segoni, S. and Tofani, V. (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. Natural Hazards and Earth System Sciences, 13(11), 2815–2831. DOI: 10.5194/nhess-13-2815-2013.

Chen, W., Xie, X., Peng, J., Shahabi, H., Hong, H., Bui, D.T., Duan, Z. and Zhu, A.X. (2018) GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. Catena, 164, 135–149. DOI: 10.1016/j.catena.2018.01.012.

Chung, C.J.F. and Fabbri, A.G. (2003) Validation of spatial prediction models for landslide hazard mapping. Natural Hazards, 30(3), 451–472. DOI: 10.1023/B.0000007172.62651.2B.

Conoscenti, C., Rotigliano, E., Cama, M., Caraballo-Arias, N.A., Lombardo, L. and Agnesi, V. (2016) Exploring the effect of absence selection on landslide susceptibility models: a case study in Sicily, Italy. Geomorphology, 261, 222–235. DOI: 10.1016/j.geomorph.2016.03.006.

Dahmani, L., Laaribya, S., Naim, H. and Dindaroglu, T. (2025) Development and assessment of landslide susceptibility in the province of Chefchaouen, North-West Morocco, using remote sensing and GIS: a weighted overlay analysis approach. Geographia Technica, 20(2), 1–14. DOI: 10.21163/GT_2025.202.01.

Do, T.V.H., Pham, H.G., Kieu, Q.L. and Tran, T.N.H. (2025) The typical mechanisms and factors leading to flash floods in small watersheds in the mountainous region of Vietnam: a case study in the Chu Va stream watershed. Larhyss Journal, 61, 141–168. Available at:
https://www.larhyss.net/ojs/index.php/larhyss/article/view/16732.

Dou, J., Yunus, A.P., Merghadi, A., Shirzadi, A., Nguyen, H., Hussain, Y., Avtar, R., Chen, Y., Pham, B.T. and Yamagishi, H. (2020) Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. Science of the Total Environment, 720, 137320. DOI: 10.1016/j.scitotenv.2020.137320.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P. et al. (2012) Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sensing of Environment, 120, 25–36. DOI: 10.1016/j.rse.2011.11.026.

Edial, H., Iskarni, P., Purwaningsih, E., Wilis, R., Chandra, D., Hafnil, J. and Desman, S. (2025) GIS-based landslide vulnerability mapping along the Sumatran Fault: a case study in Pasaman Regency, Indonesia. Geographia Technica, 20(1), 228–248. DOI: 10.21163/GT_2025.201.16.

Fawcett, T. (2006) An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. DOI: 10.1016/j.patrec.2005.10.010.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E. and Savage, W.Z. (2008) Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. Engineering Geology, 102(3–4), 85–98. DOI: 10.1016/j.enggeo.2008.03.022.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. and Michaelsen, J. (2015) The climate hazards infrared precipitation with stations — a new environmental record for monitoring extremes. Scientific Data, 2, 150066. DOI: 10.1038/sdata.2015.66.

Gu, T., Li, J., Wang, M., Duan, P., Zhang, Y. and Cheng, L. (2023) Study on landslide susceptibility mapping with different factor screening methods and random forest models. PLOS ONE, 18(10), e0292897. DOI: 10.1371/journal.pone.0292897.

Gu, T., Duan, P., Wang, M., Li, J. and Zhang, Y. (2024) Effects of non-landslide sampling strategies on machine learning models in landslide susceptibility mapping. Scientific Reports, 14, 7201. DOI: 10.1038/s41598-024-57964-5.

Guzzetti, F., Carrara, A., Cardinali, M. and Reichenbach, P. (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, central Italy. Geomorphology, 31(1–4), 181–216. DOI: 10.1016/S0169-555X(99)00078-1.

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K.-T. (2012) Landslide inventory maps: new tools for an old problem. Earth-Science Reviews, 112(1–2), 42–66. DOI: 10.1016/j.earscirev.2012.02.001.

Kanwar, M., Pokharel, B. and Lim, S. (2025) A new random forest method for landslide susceptibility mapping using hyperparameter optimization and grid search techniques. International Journal of Environmental Science and Technology, 22, 10635–10650. DOI: 10.1007/s13762-024-06310-3.

Kieu, Q.L. and Ngo, G.V. (2022) Landslide susceptibility assessment for warning of dangerous areas in Tan Uyen district, Lai Chau province, Vietnam. Geografiska Annaler: Series A, Physical Geography, 104(3), 183–200. DOI: 10.1080/04353676.2022.2091915.

Kieu, Q.L. and Tran, D.V. (2021) Application of geospatial technologies in constructing a flash flood warning model in northern mountainous regions of Vietnam: a case study at Trinh Tuong commune, Bat Xat district, Lao Cai province. Bulletin of Geography. Physical Geography Series, 20, 31–43. DOI: 10.2478/bgeo-2021-0003.

Kieu, Q.L. and Tran, T.N.H. (2025) Assessing livelihood vulnerability to climate change and disasters in northern Vietnam. Theoretical and Applied Climatology, 156, 610. DOI: 10.1007/s00704-025-05842-z.

Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., Thai Pham, B., Bui, D.T., Avtar, R. and Shirzadi, A. (2020) Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance. Earth-Science Reviews, 207, 103225. DOI: 10.1016/j.earscirev.2020.103225.

Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M. and Guzzetti, F. (2018) A review of statistically-based landslide susceptibility models. Earth-Science Reviews, 180, 60–91. DOI: 10.1016/j.earscirev.2018.03.001.

Segoni, S., Pappafico, G., Luti, T. and Catani, F. (2020) Landslide susceptibility assessment in complex geological settings: sensitivity to geological information and insights on its parameterization. Landslides, 17(10), 2443–2453. DOI: 10.1007/s10346-019-01340-2.

Steger, S., Brenning, A., Bell, R. and Glade, T. (2017) The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements. Landslides, 14(5), 1767–1781. DOI: 10.1007/s10346-017-0820-0.

Tanyu, B.F., Abbaspour, A., Alimohammadlou, Y. and Tecuci, G. (2021) Landslide susceptibility analyses using Random Forest, C4.5, and C5.0 with balanced and unbalanced datasets. Catena, 203, 105355. DOI: 10.1016/j.catena.2021.105355.

Van Westen, C.J., Castellanos, E. and Kuriakose, S.L. (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. Engineering Geology, 102(3–4), 112–131. DOI: 10.1016/j.enggeo.2008.03.010.

Yu, X., Zhang, K., Song, Y., Jiang, W. and Zhou, J. (2021) Study on landslide susceptibility mapping based on rock-soil characteristic factors. Scientific Reports, 11, 15476. DOI: 10.1038/s41598-021-94936-5.

Zhanabayev, D., Dzhanaleeva, K., Atasoy, E. and Efe, R. (2024) Landslide susceptibility mapping using Analytical Hierarchy Process and Geographical Information System in Rudny Altai Region, East Kazakhstan. Geographia Technica, 19(1), 151–165. DOI: 10.21163/GT_2024.191.11.

Geographia Technica, Vol. 21, Special Issue: Artificial Intelligence Applications in Geography, 2026, pp 206 to 229

MULTI-ALGORITHM CLASSIFICATION AND CA-MARKOV PREDICTION OF LAND-USE CHANGE ACROSS THREE TEMPORAL INTERVALS: A CASE STUDY OF THE LOE TOR ROYAL PROJECT, TAK PROVINCE, THAILAND

Sorawit ONPHUA [1], Sakhan TEEJUNTUK [1,2*] & Chakrit NA TAKUATHUNG [1]

DOI: 10.21163/GT_2026.212.11

ABSTRACT

Land-use and land-cover (LULC) change in mountainous watersheds has significant implications for food security, biodiversity conservation, and carbon storage. This study assessed LULC dynamics in the Loe Tor Royal Project area, Tak Province, Thailand, using multi-temporal satellite imagery (2006-2026), five supervised classification algorithms (Random Forest, Artificial Neural Network, Linear Discriminant Analysis, Support Vector Machine, and XGBoost), and CA-Markov spatial modeling across three calibration intervals (3-year, 5-year, and 10-year). Between 2006 and 2026, forest cover declined from 64.10% to 45.16%, while agricultural land expanded from 8.29% to 30.82%, with the rate of agricultural expansion accelerating markedly in recent years. Cramér's V analysis identified distance to settlements (V = 0.85), distance to roads (V = 0.67), and elevation (V = 0.61) as the most influential spatial drivers of LULC transitions. Among the three calibration intervals, the 3-year model achieved the highest validation accuracy, whereas the 10-year model performed substantially worse, indicating that shorter calibration intervals produce more reliable CA-Markov projections. The inferior performance of the 10-year model is attributed to Markov non-stationarity over extended calibration periods and cross-sensor spectral inconsistencies, highlighting the important methodological consideration for CA-Markov applications using multi-sensor datasets. Under Business-as-usual scenarios, forest cover is projected to continue declining, with agricultural land expanding progressively along low elevation zones near road networks and settlements, reaching approximately 42.7% by 2035 under the 3-year scenario. These findings highlight the ecological vulnerability of this critical watershed and emphasize the need for protective buffer zones, conservation agriculture practices, and degraded land rehabilitation efforts.

Keywords: *LULC change; CA–Markov model; Multi-algorithm classification; Calibration interval comparison; Sentinel-2; Landsat; Watershed management.*

1. INTRODUCTION

Land-use and land-cover (LULC) change is one of the most direct indicators of human–environment interaction, and in mountainous watersheds it carries significant implications for food security, biodiversity conservation, carbon storage, and hydrological regulation(Wang et al., 2023). The Loe Tor Royal Project, located in Mae Ramat District, Tak Province, is a site of considerable importance for natural resource restoration and soil and water management (Royal Development Projects Board, 2025). As a critical watershed draining into both the Salween and Ping river systems, the area provides ecosystem services on which downstream communities and agriculture depend; unplanned agricultural expansion into forested land can therefore trigger ecosystem degradation, accelerated soil erosion, and long-term environmental decline (Wang et al., 2023). Reliable, spatially explicit information on the rate, direction, and likely future trajectory of LULC change is thus essential — for local communities whose livelihoods depend on these resources, and for the

[1] *Faculty of Forestry, Kasetsart University, Bangkok 10900, Thailand ;*
sorawit.on@ku.th (SO), fforsktt@ku.ac.th (ST), chakrit.n@ku.ac.th (CN).
[2] *Princess Sirindhorn Center for Sustainable Development, Kasetsart University, Bangkok 10900, Thailand.*
[*] *Corresponding author : fforsktt@ku.ac.th (ST).*

government and Royal Project agencies responsible for land-use zoning and watershed protection. Geographic information systems and remote sensing provide the means to generate such information and to identify priority areas for intervention (Butt et al., 2015).

The trajectory of LULC change is governed by both biophysical and anthropogenic driving variables whose relative influence varies across geographic settings. Biophysical factors such as elevation and slope constrain the suitability of land for cultivation, whereas accessibility-related factors — particularly proximity to roads and settlements — shape the intensity and location of human activity (Thammanu et al., 2020). Recent geo-informatic studies in Thailand have quantified how these drivers translate into measurable transitions: Kiriwongwattana and Waiyasusri (2024) traced tourism-driven expansion of built-up land in Phuket, while Waiyasusri (2021) characterised mangrove and urbanisation dynamics in the Krabi Estuary Wetland using spectral indices. Identifying the dominant drivers in a given landscape is therefore a prerequisite for credible modelling and projection of future change.

A range of models has been developed to project future LULC, each differing in data requirements and spatial behaviour. The Markov chain estimates the quantity and probability of transitions between classes from historical change but is spatially non-explicit (Nouri et al., 2014). Cellular automata (CA) introduce spatial allocation through neighbourhood interactions and transition rules (Mansour et al., 2020). The integrated CA–Markov model couples the temporal prediction of the Markov chain with the spatial allocation of CA, producing spatially explicit projections from comparatively few inputs; it has been widely applied for forest-change projection in tropical settings, for example by Sari et al. (2025), in this journal, who coupled Land Change Modeler with CA–Markov to project forest-to-non-forest conversion in Boven Digoel, Indonesia, using driver variables comparable to those employed here. By contrast, the CLUE-s and Dyna-CLUE models allocate change through empirical relationships between land use and multiple socio-economic and biophysical drivers and are well suited to multi-scenario policy analysis — for example, Waiyasusri and Chotpantarat (2022) applied the Dyna-CLUE model to project coastal land-use change in Koh Chang — but they require extensive driver datasets that are often unavailable for remote highland watersheds. Given the data constraints of the Loe Tor area and the spatially explicit objective of this study, the CA–Markov model offers the most appropriate balance of robustness, reproducibility, and data efficiency.

To extract spatially explicit land-use information, multi-temporal satellite imagery was classified into thematic categories (MohanRajan et al., 2020). Because no single algorithm performs optimally across all sensors and landscapes, five supervised classifiers — Random Forest (RF), Artificial Neural Network (ANN), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and XGBoost (XGB) — were compared to identify the best-performing classifier for each study year, an approach also adopted in recent multi-algorithm classification studies in Thailand (Worachairungreung et al., 2023). Statistical differences among classifiers were assessed using Cochran's Q and McNemar tests (Cochran, 1950), and classification accuracy was evaluated using Overall Accuracy, Producer's and User's Accuracy, and the Kappa coefficient (Congalton, 2001). The association between categorical spatial drivers and observed LULC transitions was quantified using Cramér's V (Akoglu, 2018).

Despite extensive LULC research in Thailand, three gaps remain. First, many studies rely on a single classifier, leaving the comparative behaviour of multiple machine-learning algorithms in spectrally heterogeneous highland terrain under-examined. Second, the sensitivity of CA–Markov projections to the length of the calibration interval is rarely tested, even though the choice of interval directly affects the stationarity assumption underlying the Markov process. Third, royal-initiative highland watersheds such as Loe Tor remain under-studied despite their conservation significance and their reliance on multi-sensor imagery spanning two decades. The present study addresses these gaps by comparing five classifiers and, in particular, by systematically evaluating three CA–Markov calibration intervals (3-, 5-, and 10-year) against a common 2026 validation reference — a methodological comparison that, to our knowledge, has not previously been reported for a multi-sensor highland LULC dataset.

Accordingly, this study aims to: (1) classify multi-temporal LULC (2006–2026) using five machine-learning algorithms and multi-source satellite imagery (Landsat 5, Landsat 8, and Sentinel-2); (2) identify the key spatial drivers of LULC transitions using Cramér's V analysis; (3) compare the predictive performance of three CA–Markov calibration intervals (3-year, 5-year, and 10-year); and (4) project future LULC scenarios under Business-as-Usual conditions to support watershed-management strategies. The expected outputs are a set of accuracy-assessed LULC maps, inter-class transition matrices, a calibration-interval performance comparison, and validated near-future projections that together provide an evidence base for land-use planning at Loe Tor and methodological guidance for CA–Markov applications in multi-sensor mountainous settings.

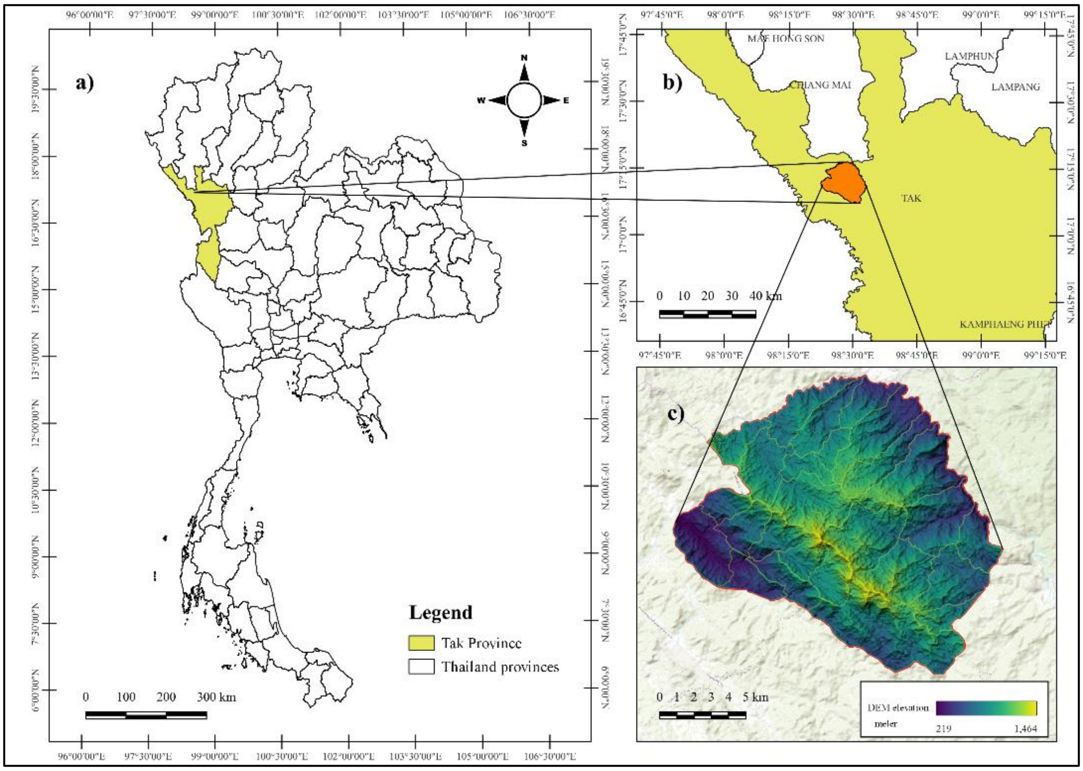


Fig. 1. Location of the Loe Tor Royal Project study area, Mae Ramat District, Tak Province, Thailand. (a) Tak Province within Thailand; (b) study-area boundary between the Salween (west) and Ping (east) watersheds; (c) digital elevation model (219–1,464 m a.s.l.).

2. STUDY AREA

The research was conducted at the Loe Tor Royal Project area located in Mae Ramat District, Tak Province. The area covers approximately 20,250 hectares (ha) (**Fig. 1**) and encompasses 26 villages (Royal Development Projects Board, 2025). This site constitutes a critical watershed draining westward into the Salween River and eastward into the Ping River, which flows into the Chao Phraya River. The landscape is characterized by complex mountainous terrain, with elevations ranging from 219 to 1,464 meters above sea level (Royal Project Foundation, 2025). Forest vegetation is highly diverse, including mixed deciduous forest, deciduous dipterocarp, and hill evergreen forest types (Thammanu et al., 2020). The regional climate is defined by three distinct seasons: rainy (June–October), winter (November–February), and summer (March–May).

3. DATA AND METHODS

3.1. Data Types and Sources

This study integrates four satellite data sources to enable temporal analysis across three periods (**Table 1**). Landsat 5 imagery (2006, 30 m) was used to provide the historical baseline for the 10-year interval analysis. Landsat 8 imagery (2016, 30 m) represents the intermediate period. Sentinel-2 imagery (2020, 2021, 2023, 2026; 10 m) supports both 3-year and 5-year interval analyses. All images were acquired during the dry season (February) to minimize cloud contamination and reduce phenological variability.

Table 1.

Data Types and Sources of sensor

Sensor	Year	Path/Row	Acquisition	Bands Used	Resolution	Analysis Period
Landsat 5 TM	2006	131/048	Feb 2006	1–5, 7	30 m	10-year
Landsat 8 OLI	2016	131/048	Feb 2016	2–7	30 m	10-year / 5-year
Sentinel-2 L2A	2020	47N	Feb 2020	2–8A, 11–12	10 m	3-year
Sentinel-2 L2A	2021	47N	Feb 2021	2–8A, 11–12	10 m	5-year
Sentinel-2 L2A	2023	47N	Feb 2023	2–8A, 11–12	10 m	3-year
Sentinel-2 L2A	2026	47N	Feb 2026	2–8A, 11–12	10 m	3-year / 5-year / 10-year (ref)

To ensure cross-sensor spatial consistency, Sentinel-2 imagery (10 m) was resampled to 30 m using bilinear interpolation when combined with Landsat data for the 5-year and 10-year interval analyses, whereas the 3-year analysis retained native 10 m resolution. Land-use change detection was conducted using the Post Classification Comparison (PCC) approach, which involves independent classification of each image followed by map comparison, thereby reducing the influence of sensor-specific spectral differences. (Sharma and Sharma, 2024). However, it should be noted that PCC does not fully account for differences in Spectral Response Functions (SRFs) among sensors (Claverie et al., 2018). This limitation is further discussed in Section 4.

3.2. Land-Use Classification

Land-use and Land cover (LULC) was classified into five thematic categories as Forest, Agriculture, Urban/Village, Grassland, and Degraded Forest **(Table 2)**. LULC classification was conducted using five supervised algorithms implemented in R as (1) Random Forest (RF; ranger package, ntree = 500, tuneLength = 5); (2) Artificial Neural Network (ANN; nnet package via caret, tuneLength = 5); (3) Linear Discriminant Analysis (LDA; MASS package via caret); (4) Support Vector Machine (SVM; e1071 package via caret, radial basis function kernel, tuneLength = 5); and (5) XGBoost (XGB; xgboost package, max depth = 6, learning rate $\eta = 0.1$, subsample = 0.8, col sample by tree = 0.8, with 5-fold cross-validation and early stopping at 20 rounds). All caret models (RF, ANN, LDA, SVM) were trained using 5-fold cross-validation. The dataset was partitioned into 70% training and 30% testing subsets using stratified random partitioning (caret: createDataPartition). All classifier were trained using the same 12 predictor variables, comprising five spectral bands (Blue, Green, Red, NIR, and SWIR calculated as the mean of Sentinel-2 B11 and B12 or Landsat SWIR1 and SWIR2), four spectral indices (NDVI, NDBI, MNDWI, EVI), and three topographic variables (DEM, Slope, Aspect). For Sentinel-2 imagery, SWIR bands B11 and B12 (20 m native resolution) were resampled to 10 m using bilinear interpolation prior to computing the composite SWIR band. For each study year, the best-performing classification was selected based on Overall Accuracy (OA).Statistical differences among classifiers were tested using Cochran's Q test (Cochran (1950); $\alpha = 0.01$), followed by pairwise post-hoc comparisons using McNemar tests (Bonferroni-corrected). Training samples were collected using stratified random sampling with a minimum of 100 reference points per land-use category per year.

Table 2.
Descriptions of Land Use and Land Cover (LULC) categories

LULC Category	Description
Forest	Areas dominated by tree cover, perennial vegetation, and economic forest plantation, including specific forest types such as dry evergreen, mixed deciduous forest, deciduous dipterocarp, and hill evergreen forest type.
Agriculture	Areas used for agricultural production including cultivated cropland (e.g., maize, cassava), irrigated land, and scattered village settlements.
Urban/Village	Areas comprising village settlements, residential structures and infrastructures including offices of forestry units and other government departments.
Grassland	Areas covered by grass and ferns, with the absence of large trees or saplings.
Degraded Forest	Open or prepared land with fewer than approximately 13 trees per hectare

3.3. Accuracy Assessment

Classification accuracy was evaluated by comparing the classified maps with independent validation samples obtained from high-resolution imagery and field observations. Four standard metrics were computed: Overall Accuracy (OA), Producer's Accuracy (PA), User's Accuracy (UA), and the Kappa coefficient (Cohen, 1960) Overall Accuracy and the Kappa coefficient were calculated as shown in Eqs. **(1)** and **(2)**, respectively. The strength of classification agreement was interpreted according the thresholds proposed by Congalton (2001).

$$OA = \left(\frac{\sum Xii}{N}\right) * 100 \tag{1}$$

$$K = \frac{N\sum_{i=1}^{r} Xii - \sum_{i=1}^{r}(Xi+)*(X+i)}{N^2 - \sum_{i=1}^{r}(Xi+)*(X+i)} \tag{2}$$

where:

K	-Kappa coefficient
R	-Number of rows in the matrix
$X_{(ii)}$	-Number of observations in row i and column i (diagonal elements)
$X_{(i+)}$	-Total number of observations in row I
$X_{(+i)}$	-Total number of observations in column I
N	-Total number of observations or verification points

3.4. Driver Variable Analysis

Seven spatial driver variables influencing LULC change were evaluated: elevation (DEM), slope, aspect, hillshade, distance to roads (dist_road), distance to streams (dist_stream), and distance to villages or settlements (dist_settle). The relative importance of these variables was assessed using Cramér's V statistic (Akoglu, 2018) for each transition pair, providing a quantitative measure of association between individual spatial drivers and the observed LULC transitions.

These seven variables were selected on three criteria: (i) their established relevance as biophysical or accessibility drivers of LULC change in mountainous watersheds (Pande et al., 2018), (ii) the availability of reliable, spatially complete layers for the entire study area across all time points; and (iii) low redundancy, which was subsequently confirmed through the pairwise Cramér's V matrix **(Table 7).**

3.5. CA–Markov Model and Multi-Temporal Analysis

Spatial modelling and LULC change prediction were conducted using R with the terra and ranger packages, implementing a CA–Markov modeling framework. For each of the three temporal intervals, a Markov transition probability matrix was derived from the transition between T1 and T2. Spatial transition suitability maps were generated using 20 Random Forest (RF) binary classification models, corresponding to each possible transition pair. The CA neighborhood module applied a 5×5 Moore window to account for spatial contiguity in land-use transitions. A total of 40,000 spatially random sample points were extracted from the T1 and T2 LULC maps together with the seven driver variables to train the RF transition models. Each binary RF classifier was trained using balanced sampling, with a maximum of 8,000 samples per class and 500 trees. The CA neighborhood influence weight was set to 0.5, meaning that the transition suitability score for each pixel was adjusted by 50% of the proportion of neighboring pixels already belonging to the target land-use class.

The CA–Markov framework couples Markov-chain temporal prediction with cellular-automata spatial allocation. The Markov component estimates the system state at the next time step from the current state and a transition probability matrix:

$$S(t+1) = Pij * S(t) \tag{3}$$

where P_{ij} is the transition probability matrix derived from two reference dates ($0 \leq P_{ij} < 1$; $\Sigma_j P_{ij} = 1$). Because the Markov chain is spatially non-explicit, the cellular-automata component allocates the predicted change in space:

$$S(x, t+1) = f(S(x,t), \Omega(x)) \tag{4}$$

where $\Omega(x)$ is the neighbourhood of cell x (a 5×5 Moore window) and f combines the Markov-derived transition area, the Random-Forest-based transition suitability, and the neighbourhood weight (0.5).

Three independent CA–Markov models were calibrated and validated as followed:

- 3-year interval: calibrated in 2020 to 2023, validated against 2026
- 5-year interval: calibrated in 2016 to 2021, validated against 2026
- 10-year interval: calibrated in 2006 to 2016, validated against 2026

The year 2026 (Sentinel-2 imagery) serves as the common independent validation reference across all three models. Future LULC projections were generated at one-interval steps beyond 2026: the 3-year model projects to 2029, 2032, and 2035; the 5-year model to 2031, 2036, and 2041; and the 10-year model to 2036, 2046, and 2056.

3.6. Model Validation

Model performance was evaluated using four complementary metrics following Pontius et al. (2007) and Pontius and Millones (2011).

These metrics included;

- Overall Accuracy (OA): the proportion of correctly predicted pixels across all land-use classes.
- Kappa Coefficient (K): the adjustment of OA for chance agreement (Cohen, 1960).
- Figure of Merit (FOM): the measurement of agreement exclusively for pixels that experienced change (FOM = A/[A+B+C]) with benchmark values typically range from 0.05–0.30 (Pontius et al., 2007)
- Quantity Disagreement (QD) and Allocation Disagreement (AD): the decompose of total classification error into error related to class proportion mismatch and spatial misallocation (Pontius et al., 2007).

Model validation was conducted using 2,500 stratified random sample points (500 per LULC class), extracted independently from the observed 2026 Sentinel-2 classification map. This stratified design ensures balanced representation of all land-use classes, including those occupying small proportions of the study area (Olofsson et al., 2014).

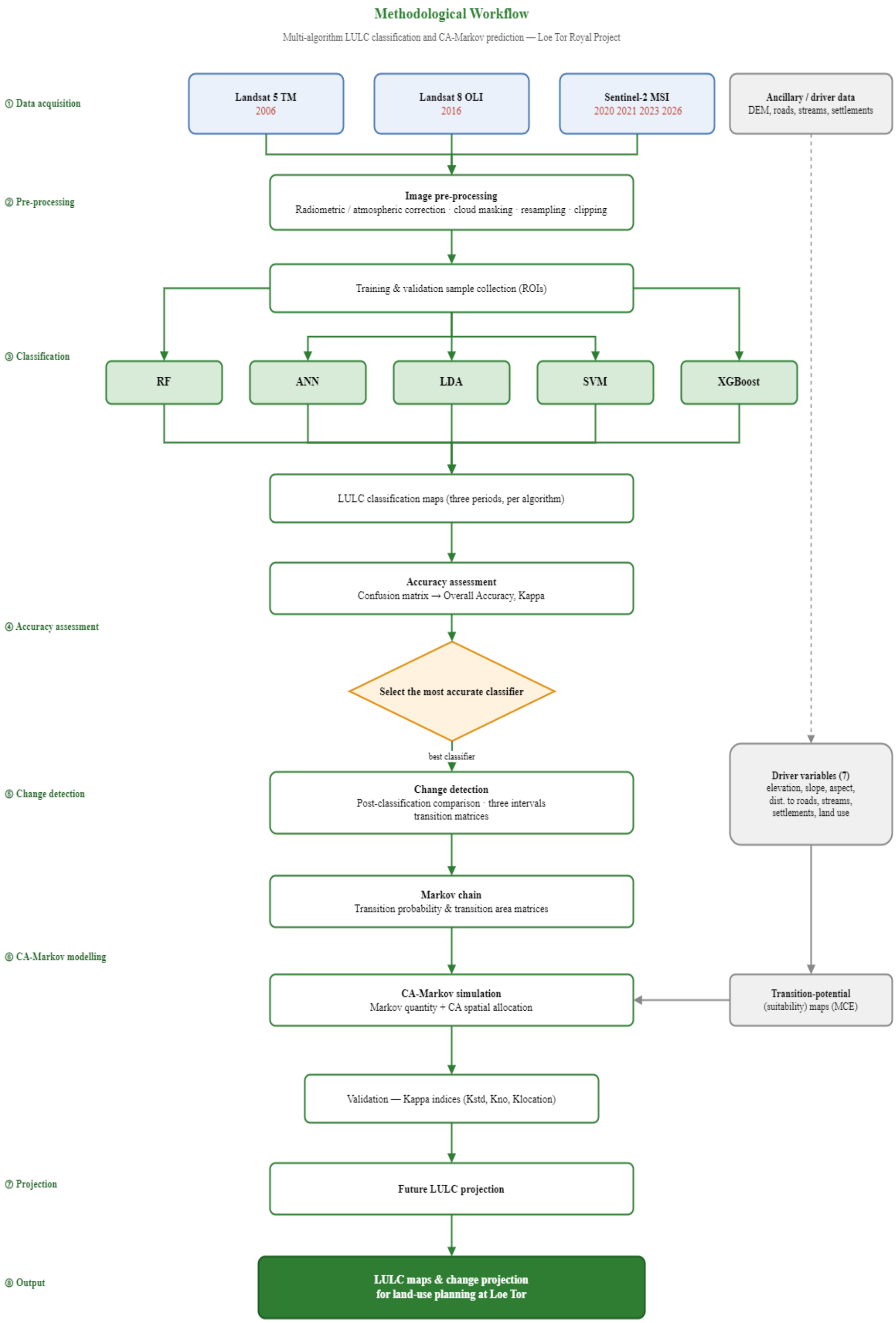


Fig. 2. Methodological workflow of the study.

4. RESULTS AND DISCUSSIONS

4.1. LULC Classification Accuracy

Six multi-temporal satellite images (2006–2026) were classified using five machine learning algorithms. Overall, classification performance is summarized in **Table 3**. Sentinel-2 imagery (10 m) consistently achieved higher accuracy (OA = 83.4–90.9%) than Landsat data (OA = 78.1–86.1%), a difference attributable to Sentinel-2's spatial resolution and enhanced spectral capabilities. Kappa coefficients ranged from 0.70 (Substantial agreement) to 0.88 (Almost Perfect agreement), indicating reliable classification performance across all years. Cochran's Q test revealed statistically significant differences among the five models in every year ($p < 0.01$), with SVM consistently exhibiting the lowest performance. Classified LULC maps for all six study years, along with classifier comparison insets, are presented in **(Fig. 3).**

Table 3.

Summary of classification accuracy across all study years.

Year	Sensor	Best Model	OA (%)	Kappa	F1 (%)	Agreement	Cochran Q	Interval
2006	Landsat 5	RF	78.1	0.7047	73.3	Substantial	18.41***	**10-yr (T1)**
2016	Landsat 8	ANN	86.1	0.8133	82.3	Almost Perfect	16.40**	**10-yr (T2) / 5-yr (T1)**
2020	Sentinel-2	LDA	90.9	0.8793	89.1	Almost Perfect	50.82***	**3-yr (T1)**
2021	Sentinel-2	RF	85.6	0.8068	82.3	Substantial	60.39***	**5-yr (T2)**
2023	Sentinel-2	RF	83.4	0.7793	79.7	Substantial	26.55***	**3-yr (T2)**
2026	Sentinel-2	LDA	88.8	0.8496	86.8	Almost Perfect	61.21***	**Validation**

**** $p < 0.001$; ** $p < 0.01$. Agreement categories follow Congalton (2001).*

The higher classification accuracy obtained from Sentinel-2 imagery (OA = 83.4–90.9%) compared to Landsat data (OA = 78.1–86.1%) is consistent with finding from recent comparative studies. Madiela et al. (2024) reported that Sentinel-2 (OA = 93.8%) outperformed Landsat 9 (OA = 91.6%) in LULC classification within the Congo Basin rainforest, attributing the improvement to Sentinel-2's finer spatial resolution (10 m vs. 30 m) and additional red-edge bands that enhance vegetation discrimination. Similarly, Keyamf et al. (2023) found that Sentinel-2 achieved an accuracy of 91%, compared to 85% for Landsat-8, when mapping shifting agricultural landscapes in Cameroon, particularly in the delineation of small-size LULC units.

The finding that no single classification algorithm consistently outperformed other across all years where RF performing best in 2006, 2021, and 2023, while LDA excelled in 2020 and 2026, highlights the sensitivity of classifier performance to data characteristics and class separability. This result is aligned with Aftab et al. (2024), who demonstrated that optimal algorithm performance varies with landscape complexity, sensor type, and the spectral characteristics of target classes. The consistently lower performance of SVM further corroborates findings by Basheer et al. (2022), who reported that ensemble-based methods, such as RF, XGBoost, generally outperform kernel-based classifiers in heterogeneous tropical environments.

Table 4 summarizes the five most important predictor variables for the Random Forest classifier across all study years. Spectral bands (Red, Blue, Green, and SWIR) and vegetation indices (NDVI and NDBI) consistently ranked highest, whereas topographic variables (DEM, Slope, and Aspect) contributed relatively less to overall classification accuracy. The observed transition from optical-band dominance (Green and Red) during the Landsat era (2006–2016) to index-based dominance (NDVI, NDBI, SWIR) in the Sentinel-2 imagery (2020–2026) reflects the enhanced spectral information, available from the newer sensor platform.

Table 4.

Top five predictor variables by relative importance (%) for each study year.

Rank	2006	2016	2020	2021	2023	2026
1	Green (100.0)	Blue (100.0)	Red (99.7)	SWIR (100.0)	NDVI (100.0)	NDVI (99.5)
2	Red (96.2)	Green (72.2)	NDVI (99.6)	NDVI (89.1)	SWIR (87.9)	Red (98.9)
3	SWIR (74.9)	MNDWI (53.2)	Blue (99.4)	NDBI (83.2)	NDBI (75.7)	NDBI (98.5)
4	Blue (45.9)	DEM (35.7)	SWIR (99.4)	Green (42.9)	Red (53.4)	SWIR (97.7)
5	NIR (41.6)	EVI (31.9)	Green (98.3)	Blue (41.2)	Blue (47.8)	Blue (96.0)

Values in parentheses represent relative importance (%). Importance normalized to 0–100 scale within each year.

Table 5.

Per-class CA–Markov prediction accuracy by temporal interval (validated against 2026).

Class	3-Year			5-Year			10-Year		
	PA	UA	F1	PA	UA	F1	PA	UA	F1
Forest	0.8159	0.7653	0.7898	0.8295	0.6695	0.7410	0.6757	0.5246	0.5906
Agriculture	0.7166	0.7077	0.7121	0.5784	0.7283	0.6448	0.1497	0.5355	0.2340
Urban/Village	0.3772	0.1458	0.2103	0.2348	0.1149	0.1543	0.2344	0.0763	0.1152
Grassland	0.3632	0.6638	0.4695	0.2233	0.7385	0.3429	0.2783	0.1777	0.2169
Degraded Forest	0.1762	0.5000	0.2606	0.2834	0.4070	0.3341	0.0524	0.1429	0.0767
OA / Kappa	**0.6592 / 0.4922**			**0.6160 / 0.4134**			**0.3940 / 0.1141**		

PA = Producer's Accuracy; UA = User's Accuracy; F1 = harmonic mean of PA and UA.

The 3-year interval model achieved the highest overall accuracy (OA = 0.6592, Kappa = 0.4922), followed by the 5-year model (OA = 0.6160, Kappa = 0.4134) and 10-year model (OA = 0.3940, Kappa = 0.1141). Across all intervals, the forest class consistently exhibited the highest F1 scores. In contrast, the Urban/Village and Degraded Forest classes demonstrated lower accuracy, primarily due to spectral confusion. Per-class prediction accuracy across all intervals is detailed in **(Table 5).**

Table 6.

LULC area comparison across three temporal intervals.

3-year Interval (2020, 2023, 2026)						
LULC Class	**2020 (ha)**	**2020 (%)**	**2026 (ha)**	**2026 (%)**	**Net Δ (ha)**	**Δ/yr (ha)**
Forest	9867.4	48.95	9104.1	45.16	-763.2	-127.2
Agriculture	3102.8	15.39	6213.0	30.82	+3110.2	+518.4
Urban/Village	1869.8	9.28	990.4	4.91	-879.4	-146.6
Grassland	2846.1	14.12	1750.7	8.68	-1095.4	-182.6
Degraded Forest	2471.6	12.26	2099.5	10.42	-372.2	-62.0
5-year Interval (2016, 2021, 2026)						
LULC Class	**2016 (ha)**	**2016 (%)**	**2026 (ha)**	**2026 (%)**	**Net Δ (ha)**	**Δ/yr (ha)**
Forest	10943.6	54.29	9104.1	45.16	-1839.5	-183.9
Agriculture	2095.6	10.40	6213.0	30.82	+4117.4	+411.7
Urban/Village	2215.1	10.99	990.4	4.91	-1224.7	-122.5
Grassland	2745.0	13.62	1750.7	8.68	-994.3	-99.4
Degraded Forest	2158.4	10.71	2099.5	10.42	-58.9	-5.9
10-year Interval (2006, 2016, 2026)						
LULC Class	**2006 (ha)**	**2006 (%)**	**2026 (ha)**	**2026 (%)**	**Net Δ (ha)**	**Δ/yr (ha)**
Forest	12921.8	64.10	9104.1	45.16	-3817.6	-190.9
Agriculture	1671.0	8.29	6213.0	30.82	+4542.0	+227.1
Urban/Village	2011.4	9.98	990.4	4.91	-1021.1	-51.1
Grassland	1410.6	7.00	1750.7	8.68	+340.1	+17.0
Degraded Forest	2142.8	10.63	2099.5	10.42	-43.4	-2.2

Δ = change from first to last year of each interval. Negative values indicate area loss.

Across the 20-year observation period (2006–2026), the net forest loss reached 3,817.6 ha, corresponding to an average annual loss of approximately 190.9 ha. The majority of this lost forest area was converted to agriculture land, which expanded by 4,542.0 ha over the same period. Notably, the rate of agricultural expansion increased substantially in recent years, raising from 227.1 ha/year during the 10-year interval to 518.4 ha/year over the most recent 3-year interval (2020–2026). This continuous and accelerating transition highlights intensifying economic and agricultural pressure on forest, with significant implications for LULC changes, local climate and biodiversity (**Fig. 3**).

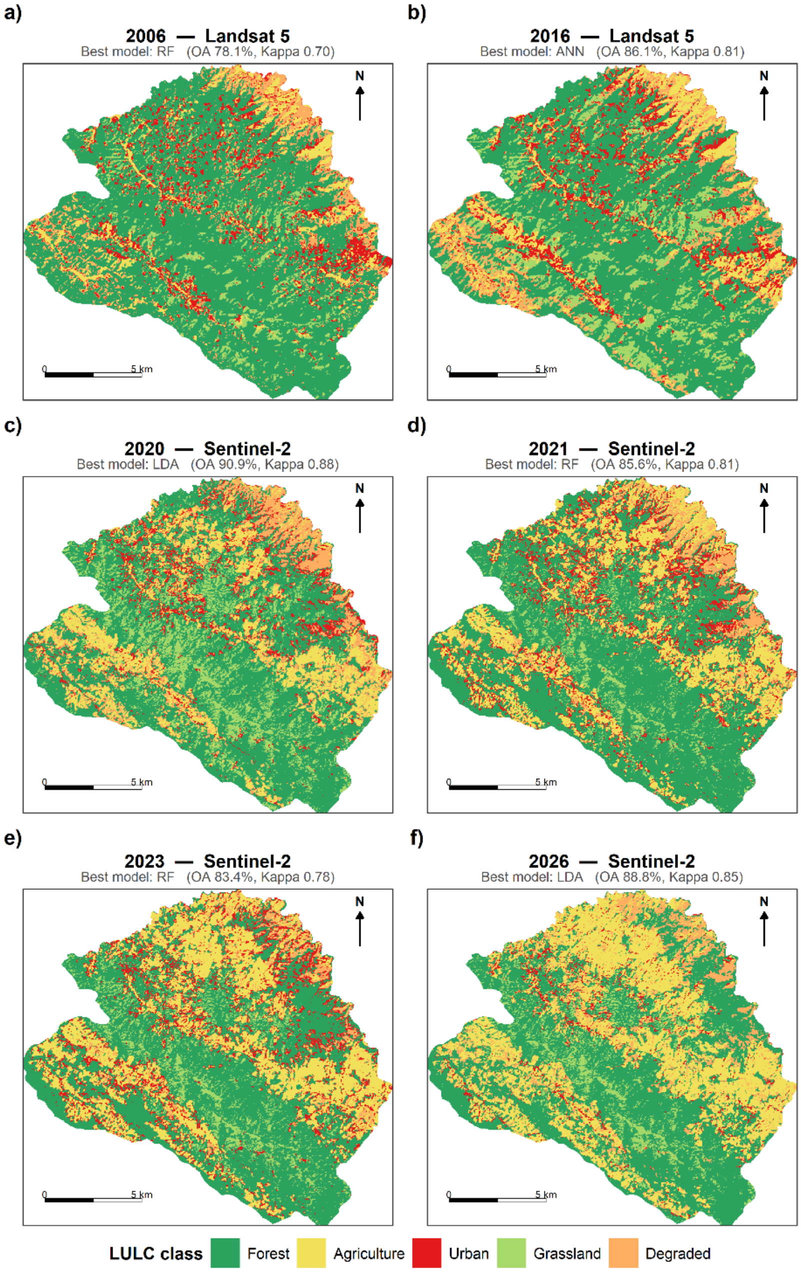


Fig. 3. Classified LULC maps for the six study years (a–f: 2006, 2016, 2020, 2021, 2023, 2026), each produced by the best-performing classifier, with five-classifier comparison insets (red box). Classes: Forest, Agriculture, Urban/Village, Grassland, Degraded Forest.

4.2. LULC Change Analysis

The analysis of land use changes in the Loe Tor Royal Project area across multiple temporal intervals (2006–2016, 2016–2021, and 2020–2023–2026) revealed that the most significant transformations occurred within the Forest and Agriculture categories. Across all intervals, forest cover exhibited consistent decline, while agricultural land expanded, reflecting a clear pattern of forest conversion to agricultural use. The 10-year interval (2006–2026) captured the largest absolute magnitude of change, with forest area decreasing by 3,817.6 ha and agricultural land increasing by 4,542.0 ha. In contrast, the 3-year interval showed the marked recent acceleration in agricultural expansion, increasing of 3,110.2 ha over the 6 years period. These findings are consistent with Bairwa et al. (2025), who identified the conversion of forested land to agriculture as a primary global driver of deforestation.

Table 6 presents area statistics for each LULC class across all study years. Across all three temporal intervals, forest cover exhibited a consistent declining trend, whereas agricultural land demonstrated continuous expansion. In the 3-year interval (2020–2026), forest area decreased from 9,867.4 ha (48.95%) to 9,104.1 ha (45.16%), representing a net loss of 763.2 ha (–127.2 ha/yr). In contrast, agricultural land increased from 3,102.8 ha (15.39%) to 6,213.0 ha (30.82%), corresponding to a net gain of 3,110.2 ha (+518.4 ha/yr), thereby confirming direct forest conversion to agricultural use over this period. Similar patterns were observed during the 5-year interval (2016–2026), in which agricultural land expanded by 4,117.4 ha (+411.7 ha/yr), primarily at the expense of forest (–1,839.5 ha) and Urban/Village areas (–1,224.7 ha). The 10-year interval (2006–2026) captured the largest absolute magnitude of change, with agriculture land increasing by 4,542.0 ha (+227.1 ha/yr) and forest cover declining by 3,817.6 ha (–190.9 ha/yr). Across all intervals, the progressive acceleration of agricultural expansion from 227.1 ha/yr in the 10-year interval to 518.4 ha/yr in the 3-year intervals, underscores the intensifying pressure on remaining forest cover in recent years.

The accelerating rate of forest-to-agriculture conversion identified in this study (from 190.9 ha/yr over 20 years to 518.4 ha/yr in the most recent 6-year period) mirrors regional deforestation trends observed across mainland Southeast Asia. Zhai et al. (2023) reported that the region experienced some of the highest deforestation rates globally, driven predominately by agricultural expansion into highland forest areas, a pattern consistent with the dynamics observed in the Loe Tor Royal Project area. The expansion of maize and cassava cultivation as a principle drivers of upland deforestation in northern Thailand has been documented by Hurni and Fox (2018). The result identified market integration and increasing road accessibility as key catalysts of this process findings that align closely with Cramér's V study, in which distance to settlements (V = 0.85) and distance to roads (V = 0.67) emerged as the most influential drivers.

The observed decline in Urban/Village area from 9.98% in 2006 to 4.91% in 2026 appears counterintuitive and is unlikely reflects an actual reduction in settlement extent. As noted by Bai et al. (2025), spectral confusion between bare agricultural soil and built-up surfaces in mountainous landscape frequently leads to overestimation of urban or village areas in Landsat-based classifications, a bias that tends to be corrected when higher-resolution Sentinel-2 imagery becomes available. This sensor-dependent reclassification effect should therefore be carefully considered when interpreting long-term LULC trends across multi-sensor datasets.

4.3. Driver Variable Analysis and Multicollinearity

Cramér's V analysis indicated that accessibility-related variables are the dominant drivers of LULC transitions. Among the examined factors, distance to settlements exhibited the highest V value (0.8452), suggested that areas located closer to villages are highly susceptible to land transformation. Distance to roads emerged as the second most influential factor (V = 0.6691), underscoring the importance of transportation infrastructure in facilitating agricultural expansion and forest conversion. Elevation (DEM) also demonstrated a strong association (V = 0.6060), particularly in governing transitions between forest and grassland in higher altitude zones. In contrast, variables slope, stream proximity, Aspect, and Hillshade exhibited only moderate to weak influence (V = 0.37–0.43).Overall, these results are consistent with the findings of Pande et al. (2018), who identified proximity to roads and settlements as primary drivers of agricultural and urban/village expansion.

The Cramér's V importance values for all driver variables across all temporal intervals were further analyzed. Overall, distance to settlements (mean importance = 174.14 for 3-year and 202.58 for 5-year intervals) and elevation (DEM; mean importance = 228.23 for 5-year and 179.87 for 10-year intervals) shown the strongest influential drivers of LULC change, consistent across all three temporal intervals. Distance to roads ranked among the top three drivers across all intervals (mean importance = 169.21 for 3-year, 202.82 for 5-year, and 124.65 for 10-year). For the 3-year interval, the ranking was distance to settlements followed by DEM and distance to roads respectively, reflecting the dominance of accessibility driven agricultural expansion near established communities. In contrast, during the 5-year interval, DEM ranked highest, followed by distance to roads and distance to settlements, suggesting that topographic constraints become more pronounced over longer calibration periods. Similarly, the 10-year interval, DEM remained the most influential variable, followed by distance to settlements and aspect ranging second and third, respectively, indicating the combined influence of physical terrain and accessibility in long-term transition dynamics. Collectively, these findings reinforce the assertion that proximity to human infrastructure acts as the primary catalyst for forest-to-agriculture conversion, while topographic factors modulate the spatial patterns and intensity of land-use change in mountainous landscape.

Prior to model implementation, multicollinearity among the seven driver variables was assessed using a pairwise Cramér's V matrix **(Table 7)**. The highest correlation was observed between Aspect and Hillshade (V = 0.46), followed by Slope and Hillshade (V = 0.23). All remaining pairwise association exhibited the values below 0.20, confirming that the selected drivers factors provide largely independent explanatory contributions and can therefore be retained in the transition suitability modeling framework without significant multicollinearity concerns.

Table 7.

Cramér's V Correlation Matrix Between Driver Variables.

	Aspect	DEM	Hillshade	Dist.Rd	Slope	Dist.Str	Dist.Set
Aspect	1.0000	0.0355	0.4628	0.0441	0.0571	0.0599	**0.0511**
DEM	0.0355	1.0000	0.0361	0.1382	0.0522	0.0889	**0.1570**
Hillshade	0.4628	0.0361	1.0000	0.0491	0.2348	0.0658	**0.0625**
Dist. Roads	0.0441	0.1382	0.0491	1.0000	0.0802	0.0658	**0.1492**
Slope	0.0571	0.0522	0.2348	0.0802	1.0000	0.0626	**0.0738**
Dist. Streams	0.0599	0.0889	0.0658	0.0658	0.0626	1.0000	**0.0720**
Dist. Settle.	0.0511	0.1570	0.0625	0.1492	0.0738	0.0720	**1.0000**

4.4. Transition Probability Matrix (TPM) and Model Validation

The transition probability matrix for 2006–2016 period indicates that forest exhibited a highest degree of stability, retaining 74.73% of its original area, with the largest conversion towards Grassland (9.19%) and Urban/Village (8.58%). Agriculture showed moderate persistence, with 46.49% remaining unchanged, accompanied by a substantial conversion to Urban/Village (26.25%). Grassland exhibited the highest stability (67.32%), while Degraded Forest showed strong reversion to Forest (38.83%) and Urban/Village (37.50%).

During 2016–2026 period, Forest retention dropped to 54.05%, with 24.59% converting to Agriculture, reflecting accelerated agricultural expansion. Agriculture maintained a retention rate of 58.70%. Grassland showed substantial reversion to Forest (54.89%), while Degraded Forest continued to convert primarily to Forest (38.04%) and Agriculture (40.53%).

Table 8.

Transition Probability Matrices for the three temporal intervals.

Interval	From \ To	Forest	Agriculture	Urban/Village	Grassland	Degraded
3-year (2020 to 2023)	**Forest**	**0.7425**	0.0869	0.0928	0.0543	0.0234
	Agriculture	0.0232	**0.8266**	0.0711	0.0021	0.0770
	Urban/Village	0.2705	0.2469	**0.3831**	0.0176	0.0819
	Grassland	0.4921	0.0593	0.0513	**0.3911**	0.0063
	Degraded	0.0696	0.4253	0.2566	0.0014	**0.2471**
5-year (2016 to 2021)	**Forest**	**0.6950**	0.0980	0.0877	0.0908	0.0286
	Agriculture	0.0624	**0.6286**	0.0785	0.0036	0.2269
	Urban/Village	0.2076	0.3639	**0.2675**	0.0161	0.1449
	Grassland	0.5909	0.1382	0.1201	**0.1032**	0.0476
	Degraded	0.3467	0.2910	0.1842	0.0080	**0.1700**
10-year (2006 to 2016)	**Forest**	**0.7473**	0.0228	0.0858	0.0919	0.0522
	Agriculture	0.0817	**0.4649**	0.2625	0.1165	0.0744
	Urban/Village	0.3296	0.1247	**0.4005**	0.0528	0.0924
	Grassland	0.1829	0.0368	0.0620	**0.6732**	0.0452
	Degraded	0.3883	0.0778	0.3750	0.0383	**0.1206**

Diagonal values (highlighted) represent class retention probability. Off-diagonal values represent transition probability to other classes.

Table 9.

Olofsson Unbiased Area Estimates (ha).

Class	Mapped (ha)	Unbiased (ha)	SE (ha)	CI95 Lower	CI95 Upper
Forest	15093	12257	337	11596	**12918**
Agriculture	10013	12070	431	11224	**12915**
Urban/Village	1698	1553	182	1196	**1909**
Grassland	2812	2067	141	1791	**2342**
Degraded Forest	3541	5210	364	4496	**5924**

The Olofsson unbiased area estimates for 5-year interval, validated against the 2026reference data, are presented in **Table 9** and reveal notable discrepancies between mapped and corrected land-use areas. Agricultural land was substantially underestimated by the classified map of 10,013 ha compared to an Unbiased estimated of 12,070 ha. Whereas, forest area was overestimated (mapped: 15,093 ha; unbiased: 12,257 ha). Degraded Forest showed the largest relative underestimation (mapped: 3,541 ha; unbiased: 5,210 ha). The Olofsson correction was applied to the 5-year interval as a representative example, however, the same methodological approach can be extended to other temporal intervals in future analyses.

The correspondence between the simulated land-use map for 2026 and the reference dataset derived from visual interpretation of Sentinel-2 satellite imagery for the same year was examined. Such comparison between the simulated output and empirical observations are essential for verifying the accuracy of the spatial transition probability model and for determining the applicability of the Transition Matrices for future projections under the BAU scenario across all three temporal intervals.

The two-period BAU comparison approach adopted in this study follows the framework proposed by Hamad et al. (2018) and provides critical insights into the temporal stationarity assumption underlying Markov-based models. The pronounce performance difference, observed for the 10-year intervals, between BAU-Early (FOM = 0.153) and BAU-Recent (FOM = 0.318) demonstrates that transition probabilities are non-stationary over extended time horizons. This finding carried important methodological implications for CA–Markov applications. Consistent with this result, Aburas et al. (2019) identified the assumption of temporal stationarity is the primary limitation of Markov-based models, particularly when calibration periods encompass the phases of rapid socio-economic change.

The projected continued decline in forest cover under all BAU scenarios carries significant implications for ecosystem service provision within the Salween and Ping watershed system. According to Wang et al. (2023), the conversion of high-integrity forests to cropland is accelerating globally, with tropical highland regions identified as particularly vulnerable. In Thailand, the Royal Project system was established precisely to encounter this trajectory by promoting sustainable highland agriculture as an alternative to shifting cultivation. However, the discrepancy between the Royal Project's conservation mandate and the continued deforestation observed in this study suggests that existing interventions may be insufficient to offset persistent market-driven pressures for agricultural expansion.

Future land-use development would benefit from moving beyond the BAU assumption by incorporating policy intervention scenarios simulation. Gebresellase et al. (2023) demonstrated that integrating alternative policy intervention into CA–Markov framework produces more realistic and decision-relevant predictions than single-scenario approaches. Such an extension would allow stakeholders to evaluate the potential effectiveness of measures cush as buffer zones, agroforestry programs, and conservation agriculture initiatives in mitigating the projected deforestation trajectory.

The validation of the 2026 land-use projection indicated that the 3-year interval model achieved the highest accuracy (OA = 0.6592, Kappa = 0.4922, FOM = 0.3254), followed by the 5-year model (OA = 0.6160, Kappa = 0.4134, FOM = 0.3432) and the 10-year model (OA = 0.3940, Kappa = 0.1141, FOM = 0.1527). At this class level, notable discrepancies were observed for the forest and agriculture categories, where the simulations projected greater forest loss and agricultural expansion than were evident in the data. This deviation likely reflects model assumptions that assigned relatively high important to accessibility of drivers and agricultural suitability. Nevertheless, the proportions for Urban/Village, Grassland, and Degraded Land closely remained with observed values, demonstrating the model's capacity to forecast overall trends, as shown in **Table 10.**

Table 10.

CA-Markov Model Validation Metrics - 2026 Reference.

Metric	3-Year	5-Year	10-Year
Overall Accuracy (OA)	0.6592	0.6160	0.3940
Kappa Coefficient	0.4922	0.4134	0.1141
Figure of Merit (FOM)	0.3254	0.3432	0.1527
Quantity Disagreement	0.1060	0.1556	0.2852
Allocation Disagreement	0.2348	0.2284	0.3208
Hit (A)	343	406	246
Miss (B)	538	596	1264
False Alarm (C)	173	181	101

The moderate Kappa values obtained in this study (0.49 for the 3-year and 0.41 for the 5-year interval) require contextualization within the broader CA–Markov modeling literature. Although these values are lower than those reported in some studies, for instance, Aburas et al. (2019) achieved Kappa as 0.80 for urban expansion prediction in Egypt using only four LULC classes, they remain comparable to results reported for complex mountainous landscapes. Gasirabo et al. (2023) reported a Figure of Merit of 0.19 for a 6-class CA–Markov model applied to the Nyabarongo River Basin in Rwanda, and similarly concluded that model performance deteriorates with increasing landscape heterogeneity and classification complexity. Several factors specific to this study likely contributed to moderate accuracy. First, the use of five LULC classes compared to three or four commonly employed in many previous studies, increases the potential for spectral confusion, particularly between Urban/Village and Degraded Forest, which exhibit similar spectral characteristic in both Landsat and Sentinel-2 imagery (Aftab et al., 2024). Second, the complex terrain of the Loe Tor area (elevations 219–1,464 m) introduces topographic shadowing and aspect-dependent reflectance variations that amplify classification uncertainty (Shandu et al., 2025). Third, the multi-sensor integration between Landsat (30 m) and Sentinel-2 (10 m resampled to 30 m) introduces spectral response inconsistencies that propagate into transition probability matrices (Claverie et al., 2018).

Nevertheless, the 3-year model's FOM (0.33) exceeds the benchmark range of 0.05–0.30 typically reported for CA–Markov simulations (Pontius et al., 2007), confirming the model ability to captures the dominant forest-to-agriculture conversion trajectory. Consistent with these finding, Kundu and Kumar (2024) similarly demonstrated that predictive accuracy generally declines as the calibration interval length increase, consistent with this finding that the 10-year model (Kappa = 0.11) performed substantially worse than the 3-year model. Detailed confusion matrices for all three intervals are provided in **(Table 11).**

Table 11.

Confusion Matrix 3-Year, 5-Year, 10-Year Interval.

Confusion Matrix 3-Year Interval.

Predicted \ Reference	Forest	Agriculture	Urban/Village	Grassland	Degraded
Forest	913	103	32	109	36
Agriculture	69	569	36	14	116
Urban/Village	93	85	43	12	62
Grassland	35	3	0	77	1
Degraded	9	34	3	0	46

Confusion Matrix 5-Year Interval.

Predicted \ Reference	Forest	Agriculture	Urban/Village	Grassland	Degraded
Forest	934	187	50	144	80
Agriculture	62	461	34	16	60
Urban/Village	74	90	27	7	37
Grassland	12	5	0	48	0
Degraded	44	54	4	0	70

Confusion Matrix 10-Year Interval.

Predicted \ Reference	Forest	Agriculture	Urban	Grassland	Degraded
Forest	769	385	81	141	90
Agriculture	18	113	10	0	70
Urban/Village	123	150	30	10	80
Grassland	191	64	5	59	13
Degraded	37	43	2	2	14

4.5. Future LULC Projections and BAU Scenarios

Future projections indicate that Forest cover is predicted to decline from 12,113.40 ha (2026*), to 10,363.15 ha by 2036*, followed by a partially recovery to 11,699.61 ha by 2056*. Agricultural land expands markedly from 1,606.61 ha (2026*) to a peak of 4,070.01 ha (2036*), before moderating to 2,347.84 ha by 2056*. In contrast, Urban/Village, Grassland, and Degraded Forest classes exhibit relatively stable extents, with only minor inflection throughout the projection period. It is important to noted that the apparent Forest cover recovery in the 10-year BAU-Early projections (2046–2056) represents an artifact rather than a genuine ecological trend. This arises from sensor inconsistency within the 10-year interval calibration interval, which incorporates data from Landsat 5 TM (2006), Landsat 8 OLI (2016), and Sentinel-2 MSI (2026), three sensors with different spectral responses and spatial resolutions. For example, Urban/Village area fluctuated from 2,010 ha (2006, Landsat 5) to 3,240 ha (2016, Landsat 8) before declining sharply to 990 ha (2026, Sentinel-2), a pattern inconsistent with field observations and attributable to sensor-dependent classification differences. Consequently, The T1 to T2 Markov transition matrix recorded an anomalously high Urban-to-Forest transition probability (0.330), which generated the spurious forest recovery signal in long-term projections. In contrast, the BAU-recent scenario (T2 to T3), which relies solely on the Landsat 8 to Sentinel-2 transition, provides greater internal consistency and is recommended as the primary projection scenario for the 10-year interval.

The transition sample analysis confirmed that sufficient training samples were available for all 20 transition pairs, with sample sizes ranging from 101 (Grassland to Agriculture) to 2,316 (Forest to Grassland). Predicted LULC areas under the 10-year BAU-Early scenario are summarized in (**Table 12**). All transitions with Markov probability exceeding 0.02 were successfully modeled.

Table 12.

Predicted LULC areas (ha) for 2026 (predicted), 2036, 2046, and 2056.

LULC Class	2026* (ha)	2036* (ha)	2046* (ha)	2056* (ha)
Forest	12113.40	10363.15	11189.17	**11699.61**
Agriculture	1606.61	4070.01	2952.30	**2347.84**
Urban/Village	3001.23	2460.85	2801.17	**2814.46**
Grassland	2676.80	2231.48	2534.91	**2733.94**
Degraded Forest	759.58	1032.13	680.07	**561.77**

* *Predicted values.*

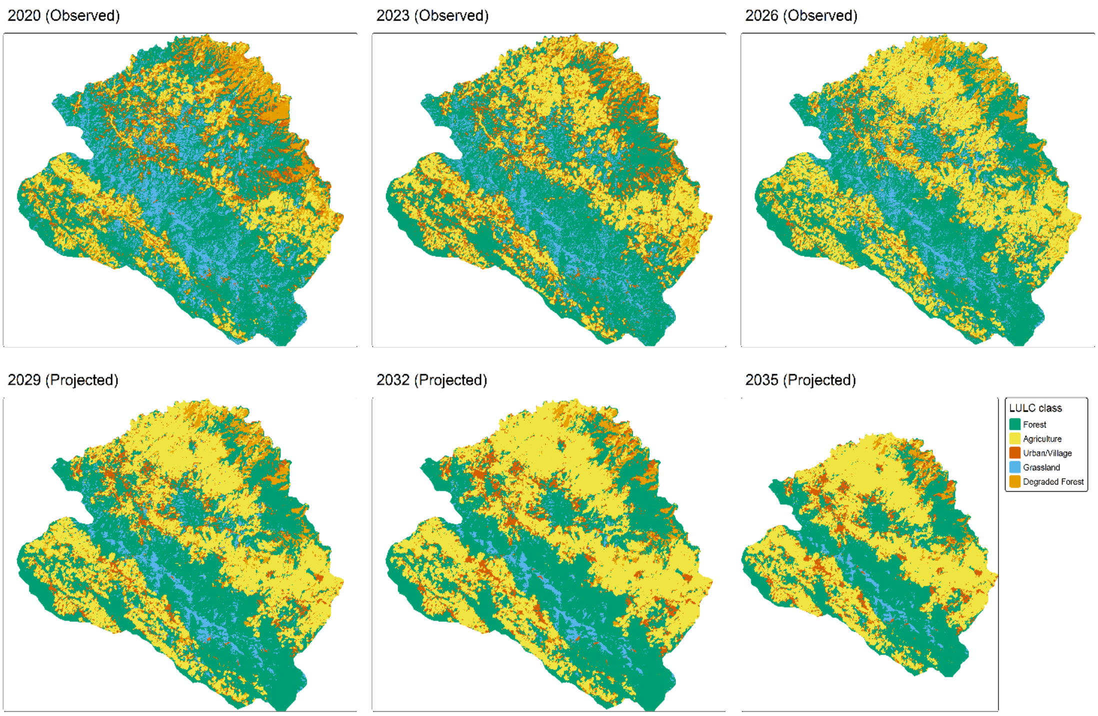


Fig. 4. Observed (2020, 2023, 2026) and 3-year BAU-Recent projected (2029, 2032, 2035) LULC maps, Loe Tor Royal Project.

Under the 10-year BAU scenario, the Forest cover is projected to decline from 12,921.8 ha (64.10%) in 2006 to 10,363.15 ha by 2036, followed by a partial recovery to 11,699.61 ha by 2056. In contrast, the 3-year BAU scenario exhibits a more severe and consistently declining trajectory, with Forest reduce from 9,867.4 ha (2020) to 8,356 ha by 2035. It is important to noted that the apparent forest recovery observed in the 10-year BAU-Early scenario does not reflect plausible ecological dynamics. As discussed in Section 3.4, this recovery arises from artifacts introduced by multi-sensor transition probability estimation and should therefore be interpreted with caution.

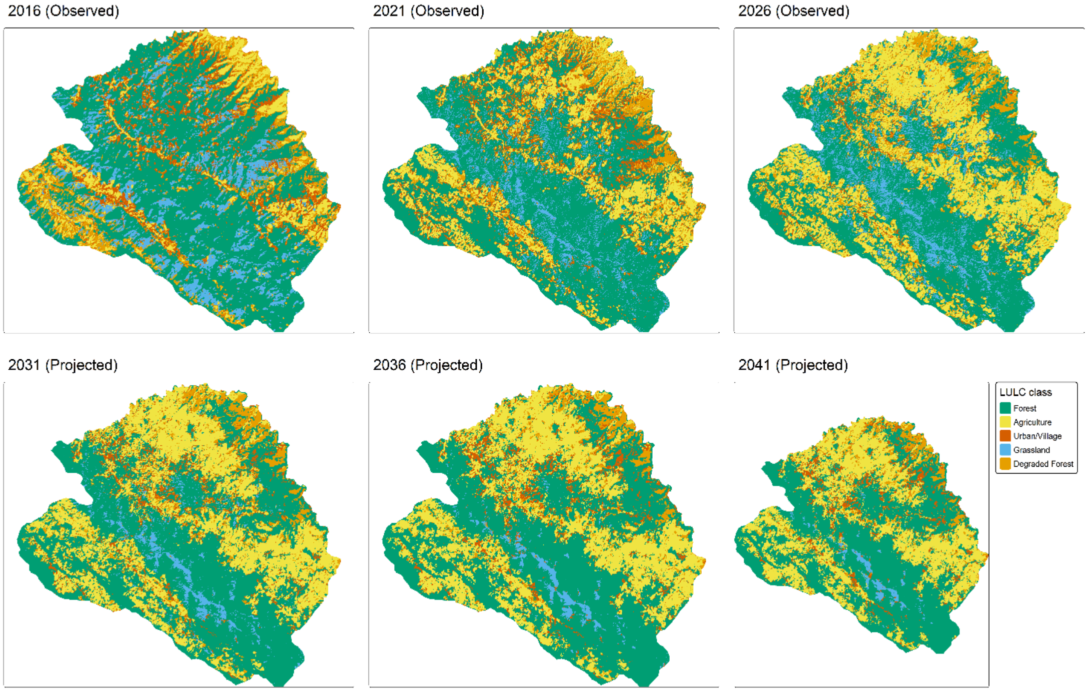


Fig. 5. Observed (2016, 2021, 2026) and 5-year BAU-Recent projected (2031, 2036, 2041) LULC maps.

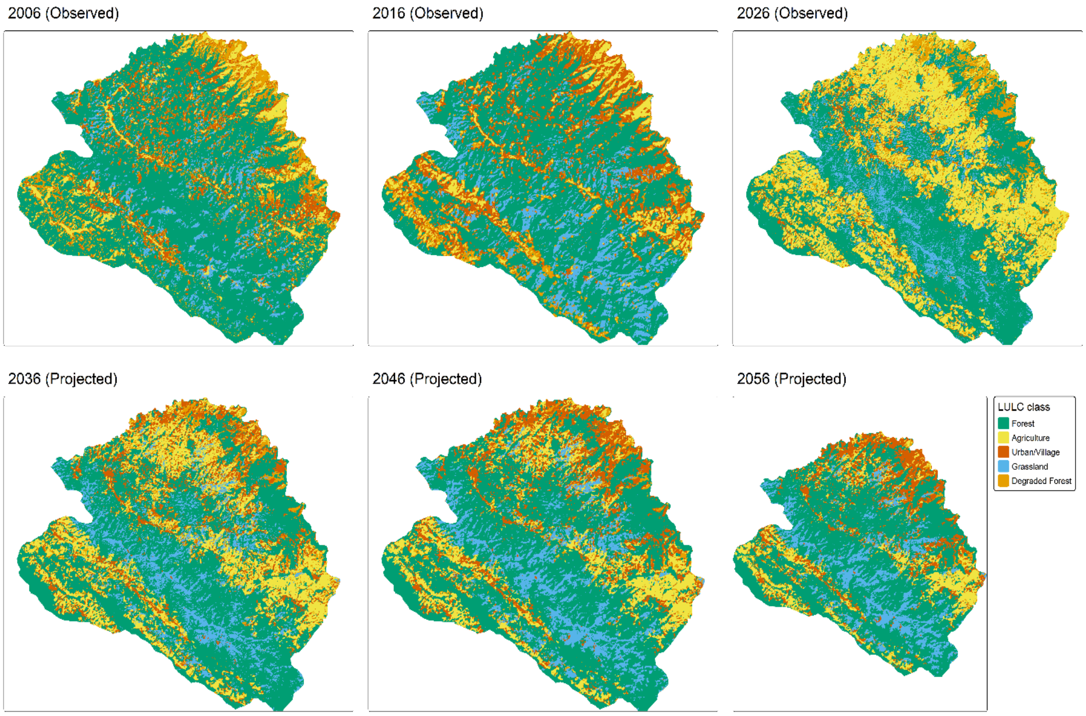


Fig. 6. Observed (2006, 2016, 2026) and 10-year BAU-Recent projected (2036, 2046, 2056) LULC maps; BAU-Recent is shown to avoid the cross-sensor artifact (see Section 3.5).

The 10-year BAU-Recent scenario, which relies exclusively on the Landsat 8 to Sentinel-2 transition matrix, provides a more ecologically realistic projection and indicates a continued decline in forest cover consistent with trends observed in the 3-year and 5-year interval. Projected LULC maps under the 3-year, 5-year, and 10-year BAU-Recent scenarios are illustrated in (**Figs. 3, 4, and 5**), respectively. Overall, this scenario reflects a cumulative forest loss across 4,400 hectares, equivalent to around 22% of the original forest area, highlights a high degree of ecosystem vulnerability and may directly impact on the ecological balance, specifically regarding biodiversity, carbon sequestration, and the maintenance of the hydrological cycle.

In contrast, the agricultural land expanded continuously from 1,671.0 ha (8.29%) in 2006 to 6,213.0 ha (30.82%) in 2026 under the 10-year interval, representing a net gain of 4,542.0 ha (+227.1 ha/yr). The expansion rate was more pronounced during the most recent 3-year interval (2020–2026) reaching +518.4 ha/yr, indicating an accelerating conversion of forested land to agriculture. This rapid expansion is largely driven by economic incentives and raising demand for agricultural production, particularly in foothill regions characterized by favorable topography and road infrastructure. Consequently, forest-to-agriculture conversion emerged as the primary mechanism driving LULC dynamics in the region.

On the contrary, Urban/Village areas exhibited a high degree of spatial stability, remaining relatively constant at approximately 990 ha (4.91% in 2026). This stability indicates that the urban expansion is not a primary driver of recent Land Use and Land Cover (LULC) change in this specific region. Similarly, Degraded forest maintained relatively stable trends across the BAU projections, accounting for approximately 10.42% of the study area in 2026. Nevertheless, under the 10-year scenario, Degraded Forest is projected to decline from 1,032 ha in 2036 to 562 ha by 2056, indicating a gradual transition toward other land-cover types.

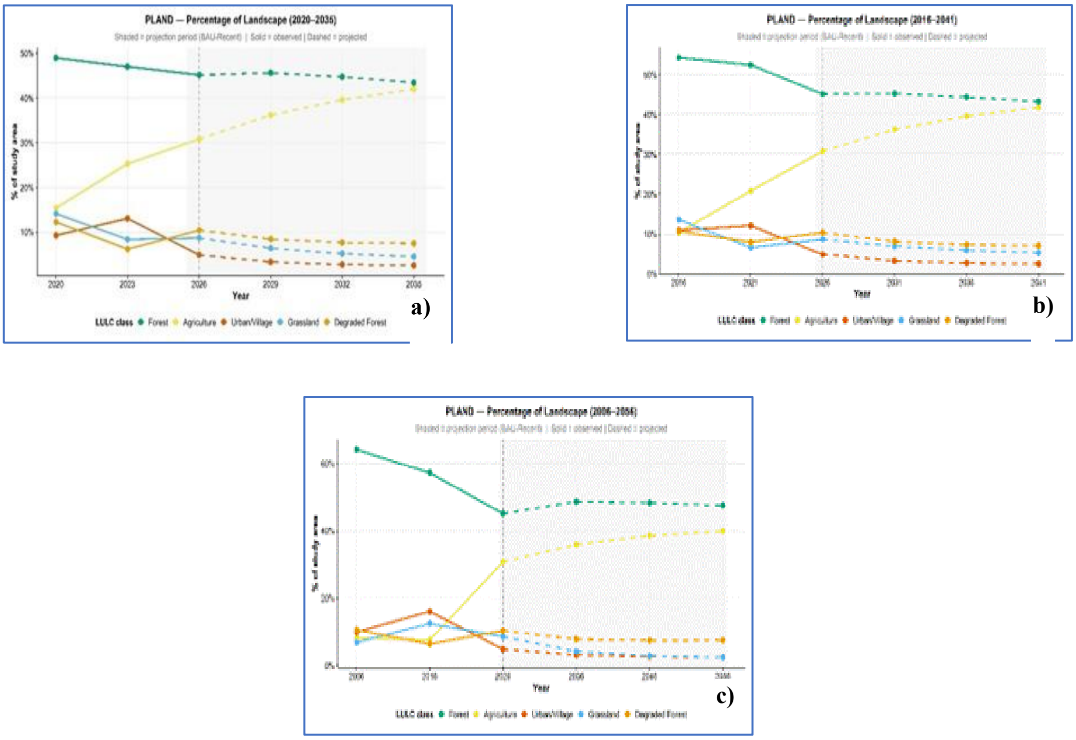


Fig. 7. Percentage-of-landscape (PLAND) trends for the five LULC classes under BAU-Recent: (a) 3-year, (b) 5-year, (c) 10-year interval. The vertical dashed line marks the observed-to-projected boundary (2026).

Grassland demonstrated instability across all BAU projections. Under the 3-year scenario, Grassland is projected to decline sharply from 1,750.7 ha in 2026 to 483 ha by 2035. In contrast, the 10-year projection indicate more moderate fluctuations ranging between 2,231 and 2,734 ha through 2056. These contrast behavior likely reflect the uncertainty of land-use dynamics at the transitional interface between agriculture and forestry. PLAND trends for all LULC classes across all BAU-Recent scenarios are shown in (**Fig. 7**).

Following Hamad et al. (2018), two Business As Usual (BAU) scenarios were defined for each temporal interval: BAU-Early, which applies the transition probability matrix derived from the T1 to T2 period, and BAU-Recent, which applies the matrix derived from the T2 to T3 period. For the 3-year interval, BAU-Early (P_2020 to 2023) yielded FOM of 0.325, whereas BAU-Recent (P_2023 to 2026) produced FOM of 0.312. Both scenarios performed similarly, with BAU-Early exhibiting a marginally higher FOM but comparable OA and Kappa values.

For the 5-year interval, BAU-Early (P_2016 to 2021) yielded an OA of 0.614, Kappa of 0.396, and FOM of 0.286, whereas BAU-Recent (P_2021 to 2026) achieved OA of 0.627, Kappa of 0.432, and FOM of 0.299. The BAU-Recent scenario consistently outperformed BAU-Early across all key metrics and presented substantially lower Quantity Disagreement (0.056 vs. 0.146).

For the 10-year interval, BAU-Early (P_2006 to 2016) yielded a FOM of 0.153, whereas BAU-Recent (P_2016 to 2026) produced a substantially higher FOM of 0.318. This large discrepancy reflects the previous discussed multi-sensor artifact, whereby BAU-Early contains an anomalous high Urban-to-Forest transition probability (0.330) resulting from cross-sensor Landsat 5 to Landsat 8 classification differences, leading to spurious Forest recovery in long-term projections. In contrast, BAU-Recent, based solely on the Landsat 8–to–Sentinel-2 matrix, avoids this artifact and is therefore recommended as the primary scenario. A visual comparison of BAU-Early and BAU-Recent projections across all intervals is presented in (**Fig. 8**).

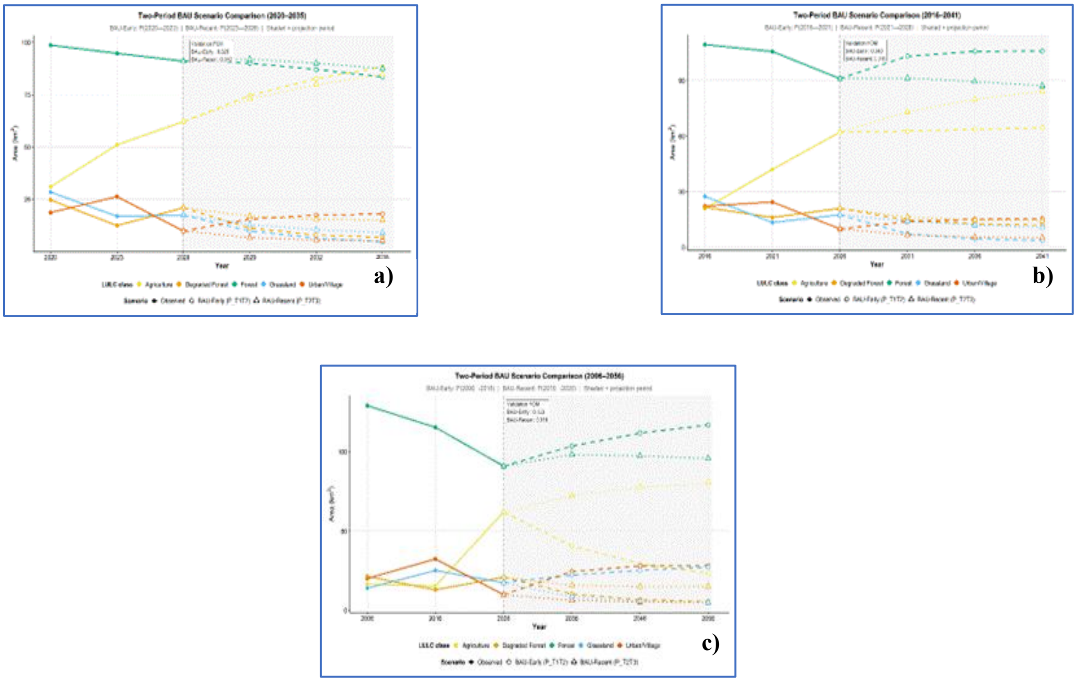


Fig. 8. Two-period BAU comparison — BAU-Early (P_T1→T2) versus BAU-Recent (P_T2→T3) — across the (a) 3-year, (b) 5-year, and (c) 10-year intervals.

In terms of class-specific accuracy, Forest consistently achieved the highest F1 scores across all intervals (F1 = 0.59–0.79), followed by Agriculture (F1 = 0.23–0.71). Urban/Village and Degraded Forest classes exhibited substantial confusion with other classes, resulting in lower F1 scores (F1 = 0.08–0.26 for Degraded Forest; F1 = 0.12–0.21 for Urban/Village). This reflects the model's limitations in distinguishing between land-uses classes with similar spectral characteristics or subtle spatial differences. However, in terms of overall precision, the model is considered sufficiently reliable for forecasting future land-use changes.

The validation analysis further indicates that while the model successfully captures a substantial number of predicted areas (Hits) across classes, False Alarms and Misses were also prevalent. These errors were particularly evident in the transitions of Forest-to-Agriculture and Forest-to-Degraded Land, which represent the dominant change process in the study area. False Alarms suggests an overestimation in certain classes, while Misses reflect underestimation of actual changes. Nevertheless, the spatial distribution of Hits indicates that the model adequately captures overall LULC trends.

4.6. Landscape Fragmentation Analysis

Landscape-level metrics were computed using the landscape metrics package in R for all observed years (2006, 2016, 2021, 2026) and projected years under the BAU-Recent scenario (5-year interval: 2031, 2036, 2041). Three complementary indices were evaluated: Shannon's Diversity Index (SHDI), Contagion Index (CONTAG), and Patch Density (PD). The results indicate a consistent decline in landscape diversity over the observed period. SHDI decreased from 1.417 (2006) to 1.294 (2016) and 1.326 (2026), reflecting the progressive landscape simplification as Agriculture replaced more diverse natural cover types. Concurrently, CONTAG increased from 38.3 (2006) to 43.8 (2016), indicating increasingly aggregation of dominant agricultural class and the expansion of large contiguous farmland.

Patch Density declined from 230.0 to 179.1 patches per 100 ha over the same period, suggesting fewer but larger and more spatially concentrated patches. Under BAU-Recent projections, SHDI is expected to continue declining while CONTAG increases, consistent with ongoing agricultural consolidation at the expense of forest patch diversity. At the class level, Forest exhibited a marked increase in the Number of Patches (NP) between 2016 and 2026 (from approximately 2,972 to 4,059 patches), indicating an active fragmentation of the forest matrix into smaller and more isolated units. A corresponding decline in the Aggregation Index (AI) confirmed reduced spatial connectivity among remaining forest patches, with implications for biodiversity persistence and ecosystem functioning.

The observed increase in forest patch number (from 2,972 to 4,059 patches between 2016 and 2026) coupled with declining Aggregation Index, represents a classic fragmentation trajectory widely documented in tropical montane forests. Chidodo et al. (2025) reported similar patterns in the East Usambara Mountains of Tanzania, where agricultural encroachment progressively fragmented contiguous forest blocks into smaller and more isolated patches. The ecological consequences of such fragmentation are well-established: reduced patch size limited habitat availability for area-sensitive species, while increased edge-to-interior ratio alters microclimate conditions and facilitate invasion by pioneer species (Haddad et al., 2015).

The declining in Shannon Diversity Index (SHDI from 1.417 to 1.326) indicates progressive landscape simplification, consistent with the dominance of forest-to-agriculture conversion as a single prevailing transition pathway. This pattern contrasts with landscapes experiencing multiple simultaneous transitions (e.g., urbanization combined with agricultural expansion), which typically exhibit stable or increasing SHDI values (Gasirabo et al., 2023). The concurrent increase in CONTAG (from 38.3 to 43.8) further confirms that agricultural expansion is creating large, contiguous farmland blocks rather than dispersed patches, a pattern with potentially irreversible implications for forest connectivity and wildlife corridor viability in the Salween and Ping watersheds.

5. CONCLUSIONS

This study integrated multi-temporal satellite imagery, five supervised classification algorithms, and CA‑Markov modelling to analyse LULC dynamics in the Loe Tor Royal Project area between 2006 and 2026. The results reveal a consistent and accelerating loss of forest to agriculture, with proximity to settlements and roads identified as the primary spatial drivers. Sentinel-2 consistently outperformed Landsat, and no single classifier was superior across all years, underscoring the sensitivity of classifier performance to sensor characteristics and landscape complexity.

The comparison of calibration intervals is the principal methodological contribution. The 3-year interval produced the most reliable CA‑Markov projections, substantially outperforming the 5- and 10-year models; the poor performance of the 10-year model is attributable to Markov non-stationarity and cross-sensor spectral inconsistency over long horizons. Across all Business-as-Usual scenarios, forest cover is projected to continue declining while agriculture expands along low-elevation margins near roads and settlements.

These findings should be read against three limitations: cross-sensor integration introduces spectral-response differences that affect transition estimation (notably for the 10-year interval); minor classes (Urban/Village, Degraded Forest) remain affected by spectral confusion in complex terrain; and the BAU projections assume stationary transition rates without explicit policy or climate drivers. Within these bounds, the study points to three priority interventions − protective buffer zones around remaining forest, conservation agriculture as an alternative to forest conversion, and GIS-based monitoring for proactive management − and provides methodological guidance for CA‑Markov applications in tropical mountainous landscapes. Future work should incorporate socio-economic drivers and compare CA‑Markov with alternative modelling frameworks.

6. DATA AVAILABILITY

The Landsat imagery is freely available from the USGS Earth Explorer (https://earthexplorer.usgs.gov). Sentinel-2 Level-2A imagery was obtained from the Copernicus Open Access Hub (https://dataspace.copernicus.eu). The R scripts used for classification and CA–Markov modeling is available from the corresponding author upon reasonable request.

ACKNOWLEDGEMENTS

This study was conducted under the auspices of the Princess Sirindhorn Center for Sustainable Development, Kasetsart University. The authors also gratefully acknowledge the Royal Project Foundation and the Land Development Department (LDD) for providing field reference data and local land-use information. Satellite imagery was obtained free of charge from the Copernicus Open Access Hub (European Space Agency) and the United States Geological Survey (USGS) Earth Explorer. This research received no specific grant from any funding agency in the public, commercial, or nonprofit sectors.

REFERENCES

Aburas, M. M., Ahamad, M. S. S. and Omar, N. Q. (2019). "Spatio-temporal simulation and prediction of land-use change using conventional and machine learning models: a review." *Environ Monit Assess* **191**(4): 205. Available from: https://doi.org/10.1007/s10661-019-7330-6 [Accessed 5 March 2026].

Aftab, B., Wang, Z., Wang, S. and Feng, Z. (2024). "Application of a Multi-Layer Perceptron and Markov Chain Analysis-Based Hybrid Approach for Predicting and Monitoring LULCC Patterns Using Random Forest Classification in Jhelum District, Punjab, Pakistan." *Sensors (Basel)* **24**(17): 5648. Available from: https://doi.org/10.3390/s24175648 [Accessed 30 August 2025].

Akoglu, H. (2018). "User's guide to correlation coefficients." *Turk J Emerg Med* **18**(3): 91-93. Available from: https://doi.org/10.1016/j.tjem.2018.08.001 [Accessed 1 September 2025].

Bai, M., Zhang, X., Ai, W., Jing, X. and Liu, L. (2025). "A high-resolution global annual city and town boundaries dataset (2000-2022) derived from GLC_FCS30D product." *Sci Data* **13**(1): 50. Available from: https://doi.org/10.1038/s41597-025-06368-9 [Accessed 4 December 2025].

Bairwa, B., Sharma, R., Kundu, A., Sammen, S. S., Alshehri, F., Pande, C. B., Orban, Z. and Salem, A. (2025). "Predicting changes in land use and land cover using remote sensing and land change modeler." *Frontiers in Environmental Science* **13**: 1540140. Available from: https://doi.org/10.3389/fenvs.2025.1540140 [Accessed 1 March 2026].

Basheer, S., Wang, X., Farooque, A. A., Nawaz, R. A., Liu, K., Adekanmbi, T. and Liu, S. (2022). "Comparison of Land Use Land Cover Classifiers Using Different Satellite Imagery and Machine Learning Techniques." *Remote Sensing* **14**(19): 4978. Available from: https://doi.org/10.3390/rs14194978 [Accessed 1 March 2026].

Butt, A., Shabbir, R., Ahmad, S. S. and Aziz, N. (2015). "Land use change mapping and analysis using Remote Sensing and GIS: A case study of Simly watershed, Islamabad, Pakistan." *The Egyptian Journal of Remote Sensing and Space Science* **18**(2): 251-259. Available from: https://doi.org/10.1016/j.ejrs.2015.07.003 [Accessed 30 December 2025].

Chidodo, S., Zhang, L., Zarei, A. and Feger, K.-H. (2025). "Remote sensing-based assessment of four decades of land use/land cover change in the Sigi River Watershed, East Usambara Mountains, Tanzania." *Frontiers in Remote Sensing* **6**: 1594331. Available from: https://doi.org/10.3389/frsen.2025.1594331 [Accessed 30 January 2026].

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V. and Justice, C. (2018). "The Harmonized Landsat and Sentinel-2 surface reflectance data set." *Remote Sensing of Environment* **219**: 145-161. Available from: https://doi.org/10.1016/j.rse.2018.09.002 [Accessed 1 March 2026].

Cochran, W. G. (1950). "The comparison of percentages in matched samples." *Biometrika* **37**(3-4): 256-266. Available from: https://doi.org/10.1093/biomet/37.3-4.256 [Accessed 20 April 2026].

Cohen, J. (1960). "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* **20**(1): 37-46. Available from: https://doi.org/10.1177/001316446002000104 [Accessed 1 March 2026].

Congalton, R. G. (2001). "Accuracy assessment and validation of remotely sensed and other spatial information." *International Journal of Wildland Fire* **10**(4): 321-328. Available from: https://doi.org/10.1071/wf01031 [Accessed 1 March 2025].

Gasirabo, A., Xi, C., Hamad, B. R. and Edovia, U. D. (2023). "A CA–Markov-Based Simulation and Prediction of LULC Changes over the Nyabarongo River Basin, Rwanda." *Land* **12**(9): 1788. Available from: https://doi.org/10.3390/land12091788 [Accessed 30 March 2026].

Gebresellase, S. H., Wu, Z., Xu, H. and Muhammad, W. I. (2023). "Scenario-Based LULC Dynamics Projection Using the CA–Markov Model on Upper Awash Basin (UAB), Ethiopia." *Sustainability* **15**(2): 1683. Available from: https://doi.org/10.3390/su15021683 [Accessed 30 March 2026].

Haddad, N. M., Brudvig, L. A., Clobert, J., Davies, K. F., Gonzalez, A., Holt, R. D., Lovejoy, T. E., Sexton, J. O., Austin, M. P., Collins, C. D., Cook, W. M., Damschen, E. I., Ewers, R. M., Foster, B. L., Jenkins, C. N., King, A. J., Laurance, W. F., Levey, D. J., Margules, C. R., Melbourne, B. A., Nicholls, A. O., Orrock, J. L., Song, D.-X. and Townshend, J. R. (2015). "Habitat fragmentation and its lasting impact on Earths ecosystems." *Science Advances* **1**(2): e1500052. Available from: https://doi.org/10.1126/sciadv.1500052 [Accessed 30 March 2026].

Hamad, R., Balzter, H. and Kolo, K. (2018). "Predicting Land Use/Land Cover Changes Using a CA-Markov Model under Two Different Scenarios." *Sustainability* **10**(10): 3421. Available from: https://doi.org/10.3390/su10103421 [Accessed 30 March 2026].

Hurni, K. and Fox, J. (2018). "The expansion of tree-based boom crops in mainland Southeast Asia: 2001 to 2014." *Journal of Land Use Science* **13**(1-2): 198-219. Available from: https://doi.org/10.1080/1747423x.2018.1499830 [Accessed 30 March 2026].

Keyamf, C., Yemefack, M., Silatsa, F. and Oben, T. (2023). "Sentinel-2 and Landsat-8 potentials for high-resolution mapping of the shifting agricultural landscape mosaic systems of Southern Cameroon." *International Journal of Applied Earth Observation and Geoinformation* **124**: 103545. Available from: https://doi.org/10.1016/j.jag.2023.103545 [Accessed 1 November 2025].

Kiriwongwattana, K. and Waiyasusri, K. (2024). "Spatial evolution of smart cities for sustainable tourism: A case study of phuket province, thailand." *GeoJournal of Tourism and Geosites* **55**: 1312-1320. Available from: https://doi.org/10.30892/gtg.55331-1303 [Accessed 30 June 2026].

Kundu, S. and Kumar, V. (2024). "Significance of Measuring the Accuracy of Cellular Automata Markov Chain for Land Use Projections in District Gurgaon in Haryana, India." *International Journal of Engineering Trends and Technology* **72**(9): 304-311. Available from: https://doi.org/10.14445/22315381/ijett-v72i9p125 [Accessed 30 April 2026].

Madiela, B., Clobite, B. and Midel, M. (2024). "Comparing Sentinel-2 and Landsat 9 for land use and land cover mapping assessment in the north of Congo Republic: a case study in Sangha region." *International Journal of Remote Sensing* **45**: 1-22. Available from: https://doi.org/10.1080/01431161.2024.2394238 [Accessed 1 October 2025].

Mansour, S., Al-Belushi, M. and Al-Awadhi, T. (2020). "Monitoring land use and land cover changes in the mountainous cities of Oman using GIS and CA-Markov modelling techniques." *Land Use Policy* **91**: 104414. Available from: https://doi.org/10.1016/j.landusepol.2019.104414 [Accessed 1 August 2025].

Mohanrajan, S. N., Loganathan, A. and Manoharan, P. (2020). "Survey on Land Use/Land Cover (LU/LC) change analysis in remote sensing and GIS environment: Techniques and Challenges." *Environ Sci Pollut Res Int* **27**(24): 29900-29926. Available from: https://doi.org/10.1007/s11356-020-09091-7 [Accessed 15 August 2025].

Nouri, J., Gharagozlou, A., Arjmandi, R., Faryadi, S. and Adl, M. (2014). "Predicting Urban Land Use Changes Using a CA–Markov Model." *Arabian Journal for Science and Engineering* **39**(7): 5565-5573. Available from: https://doi.org/10.1007/s13369-014-1119-2 [Accessed 30 November 2025].

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E. and Wulder, M. A. (2014). "Good practices for estimating area and assessing accuracy of land change." *Remote Sensing of Environment* **148**: 42-57. Available from: https://doi.org/10.1016/j.rse.2014.02.015 [Accessed 25 November 2025].

Pande, C. B., Moharir, K. N., Khadri, S. F. R. and Patil, S. (2018). "Study of land use classification in an arid region using multispectral satellite images." *Applied Water Science* **8**(5): 164. Available from: https://doi.org/10.1007/s13201-018-0764-0 [Accessed 1 April 2026].

Pontius, R. G., Thontteh, O. and Chen, H. (2007). "Components of information for multiple resolution comparison between maps that share a real variable." *Environmental and Ecological Statistics* **15**(2): 111-142. Available from: https://doi.org/10.1007/s10651-007-0043-y [Accessed 30 March 2026].

Pontius, R. G. J. and Millones, M. (2011). "Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment." *International Journal of Remote Sensing* **32**(15): 4407-4429. Available from: https://doi.org/10.1080/01431161.2011.552923 [Accessed 30 March 2026].

Royal Development Projects Board. 2025. *Sufficiency Economy Philosophy* [Online]. Available: https://www.rdpb.go.th/th/Sufficiency [Accessed 5 October 2025].

Royal Project Foundation. 2025. *History of Lertor Royal Project* [Online]. Available: https://royalproject.org/page/History_Lertor/ [Accessed 5 October 2025].

Sari, D. N., Mutashim, M. H. F., Restuvica, F. V., Putra, K. P. B., Mukasyaf, A. A., Anggani, N. L. and Nugraga, A. R. (2025). "Prediction of Spatial Land Use Land Cover Changes in Boven Digoel, South Papua in 2031 and 2041 Using LCM and CA-Markov Models." *Geographia Technica* **20**(2): 283-302. Available from: https://doi.org/10.21163/GT_2025.202.18 [Accessed 30 June 2026].

Shandu, I. D., Xulu, S. and Gebreslasie, M. (2025). "Enhancing land cover classification in the heterogeneous landscape by integrating auxiliary data with Sentinel-2 imagery using the random forest algorithm." *Frontiers in Remote Sensing* **6**: 1697897. Available from: https://doi.org/10.3389/frsen.2025.1697897 [Accessed 5 October 2025].

Sharma, G. and Sharma, M. K. (2024). "Development of multispectral post-classification change detection technique." *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*: 1-9. Available from: https://doi.org/10.1109/SCEECS61402.2024.10482012.

Thammanu, S., Marod, D., Han, H., Bhusal, N., Asanok, L., Ketdee, P., Gaewsingha, N., Lee, S. and Chung, J. (2020). "The influence of environmental factors on species composition and distribution in a community forest in Northern Thailand." *Journal of Forestry Research* **32**(2): 649-662. Available from: https://doi.org/10.1007/s11676-020-01239-y [Accessed 15 August 2025].

Waiyasusri, K. (2021). "Monitoring the Land Cover Changes in Mangrove Areas and Urbanization using Normalized Difference Vegetation Index and Normalized Difference Built-up Index in Krabi Estuary Wetland, Krabi Province, Thailand." *Applied Environmental Research* **43**(3): 1-16. Available from: https://doi.org/10.35762/AER.2021.43.3.1 [Accessed 30 June 2026].

Waiyasusri, K. and Chotpantarat, S. (2022). "Spatial Evolution of Coastal Tourist City Using the Dyna-CLUE Model in Koh Chang of Thailand during 1990-2050." *ISPRS International Journal of Geo-Information* **11**(1): 49. Available from: https://doi.org/10.3390/ijgi11010049 [Accessed 30 June 2026].

Wang, L., Wei, F. and Svenning, J.-C. (2023). "Accelerated cropland expansion into high integrity forests and protected areas globally in the 21st century." *iScience* **26**(4): 106450. Available from: https://doi.org/10.1016/j.isci.2023.106450 [Accessed 15 March 2026].

Worachairungreung, M., Thanakunwutthirot, K. and Kulpanich, N. (2023). "A study on oil palm classification for ranong province using data fusion and machine learning algorithms." *Geographia Technica* **18**(1): 161-176. Available from: https://doi.org/10.21163/GT_2023.181.12 [Accessed 30 June 2026].

Zhai, J., Xiao, C., Feng, Z. and Liu, Y. (2023). "Are there suitable global datasets for monitoring of land use and land cover in the tropics? Evidences from mainland Southeast Asia." *Global and Planetary Change* **229**: 104233. Available from: https://doi.org/10.1016/j.gloplacha.2023.104233 [Accessed 8 September 2025].

ARTIFICIAL INTELLIGENCE IN GEOGRAPHICAL RESEARCH: FROM SPATIAL OBSERVATION TO INFORMED ACTION - A SYNTHESIS OF THE AIAG STUDIES AND THE GEOGRAPHICAL FOUNDATIONS OF GEOAI

***Zsolt MAGYARI-SÁSKA*[1]**

An Afterword from the Co-editor

DOI: 10.21163/GT_2026.212.13

ABSTRACT

This paper synthesises the twelve contributions included in the AIAG special issue and considers their findings within the broader development of geospatial artificial intelligence. The studies show how AI and related computational methods can support the identification of geographical phenomena, the estimation of environmental conditions in data-sparse areas, the analysis of spatiotemporal processes, the construction of land-use and hazard scenarios, and the integration of data-driven models with physical knowledge. Rather than treating algorithms as the principal outcome, the synthesis focuses on the geographical evidence they produce, including flood boundaries, pollution estimates, rainfall forecasts, biomass inventories, burned-area maps, land-use projections and susceptibility assessments. The discussion then shifts from what AI performs in the individual studies to what geography contributes to AI. It examines how geographical research defines meaningful entities, represents scale, connectivity and temporal structure, organises heterogeneous evidence, keeps environmental and social processes visible, and connects model outputs to places, institutions and decisions. These contributions are essential because spatial coordinates alone do not make a model geographically meaningful.

The paper also considers the conditions required for reliable GeoAI, particularly spatially appropriate validation, transferability between regions, interpretability and the spatial communication of uncertainty. It argues that AI is most useful when it complements GIS, remote sensing, field observations, statistical analysis, physical modelling and expert knowledge. Its contribution lies not in replacing geographical judgement, but in extending the capacity to observe, estimate and anticipate spatial processes while keeping model outputs connected to the geographical contexts and decisions they represent.

Keywords: *GeoAI; geographical knowledge; spatial dependence; geographical modelling; decision support; spatial validation; explainable AI*

1. INTRODUCTION

Geospatial artificial intelligence (GeoAI) represents an evolving set of analytical methodologies that extend the capacity of geographers to observe, measure, model, and anticipate spatial phenomena. The twelve contributions assembled in this special issue demonstrate this analytical progression across diverse environmental and socio-economic contexts:

[1] *Babeş-Bolyai University, Faculty of Geography, 400006 Cluj-Napoca, Romania, zsolt.magyari@ubbcluj.ro co-editor of AIAG.*

1. Kritchayan INTARAT, Phatsachon KUANKHAMNUAN and Woraman JANGSAWANG / *Flood Mapping in Phra Nakhon Si Ayutthaya, Thailand, Utilizing Sentinel-1 SAR Imagery and Deep Learning Approaches*
2. Triyatno TRIYATNO, Lailatur RAHMI, Syafri ANWAR, Beni AULIA and Sumayyah Aimi Mohd NAJIB / *Projecting LULC Change (2025–2039) and Modeling Landslide Susceptibility Using Cellular Automata–Markov Chain (CA–MC) in the Kambang Watershed, West Sumatra, Indonesia*
3. Sri NURDIATI, Mochamad Tito JULIANTO, Ionel HAIDU, Muhammad Daryl FAUZAN, Hari NURDIANTO, Syukri Arif RAFHIDA and Mohamad Khoirun NAJIB / *Predicting Fire Hotspots in Kalimantan Based on Climate Factors Using Gaussian Process Regression and Long Short-Term Memory*
4. Uca SIDENG, Nurul Afdal HARIS and Mustari S. LAMADA / *An Integrated ANN–CA Model for Land-Use Change Prediction and Flood-Risk Mitigation: A Case Study of Enrekang Regency, Indonesia*
5. Ananto A. AJI, Syaiful Muflichin PURNAMA and Vina Nurul HUSNA / *Random Forest–Based Mapping of Riverine Plastic Pollution from Sentinel-2 in Google Earth Engine: LULC and Rainfall Controls in Kendal Regency, Indonesia*
6. Jumadi JUMADI, Kuswaji Dwi PRIYONO, Ali Hasan ABDULLAH, Supari SUPARI, Hamza AIT ZAMZAMI, Farha SATTAR, Muhammad NAWAZ, Hamzah HASYIM and Steve CARVER / *Performance of Machine-Learning and Deep-Learning Models for Predicting Rainfall in a Large Watershed: Case Study of the Bengawan Solo River Basin, Indonesia*
7. Imad KATIBA, Hanae BELMAJDOUB and Khalid MINAOUI / *Hybrid Deep-Learning and Reinforcement Framework for Physically Consistent Precipitation Control along the Moroccan–Iberian Coasts*
8. Maliwan THABTHONG, Teerawong LAOSUWAN, Satith SANGPRADID, Yannawut UTTARUK, Jenjira PIAMDEE, Piyatida AWICHIN, Titipong PHOOPHATHONG and Maharaja SINGHARAJ / *An AI-Based Framework for Estimating $PM_{2.5}$ Using Satellite Aerosol Optical Depth and Meteorological Data in a Coastal Industrial Region of Thailand*
9. Pimpilai WUNAPHAI, Teerawong LAOSUWAN, Satith SANGPRADID, Yannawut UTTARUK, Narueset PRASERTSRI, Thinnakon ANGKAHAD and Piyatida AWICHIN / *Geospatial Deep Learning for Biomass and Carbon-Stock Estimation Using UAV-Derived Canopy Metrics*
10. Worawit SUPPAWIMUT and Ratchaphon SAMPHUTTHANONT / *Deep Learning–Based Detection of Agricultural Burned Areas Using True-Color Satellite Imagery*
11. Thanh Long NGUYEN, Thi Ngoc Ha TRAN, Quoc Lap KIEU and Dieu Trinh NGUYEN / *A Shear-Strength-Derived Lithological Weakness Index for GIS-Based Landslide-Susceptibility Modelling in Mountainous Road Corridors of Northern Vietnam*
12. Sorawit ONPHUA, Sakhan TEEJUNTUK and Chakrit NA TAKUATHUNG / *Multi-Algorithm Classification and CA–Markov Prediction of Land-Use Change across Three Temporal Intervals: A Case Study of the Loe Tor Royal Project, Tak Province, Thailand*

These studies address diverse environmental phenomena, including riverine flooding, precipitation patterns, forest fires, atmospheric pollution, riverine plastic pollution, plantation biomass, land-use transitions, and slope instability. Study areas span alluvial floodplains, tropical river basins, industrial coastal zones, commercial plantations, and mountainous transport corridors. Collectively, these investigations examine complex spatial and temporal systems by integrating multi-source geospatial data, computational models, and domain knowledge of localized physical processes.

The papers employ a broad spectrum of AI and related computational techniques, including semantic image segmentation, supervised classification, nonlinear regression, recurrent neural networks, ensemble tree architectures, cellular automata, Markov chains, and reinforcement learning. The algorithmic development is rarely the primary objective; the principal outputs consist of empirical spatial evidence, such as mapped flood extents, aerosol concentration fields, rainfall forecasts, burned scar geometries, land-use trajectories, biomass inventories, and susceptibility zonations. The geographical value of AI lies in this conversion of heterogeneous observations into spatially explicit evidence.

In the following sections, numbers in square brackets refer to the twelve AIAG contributions listed above.

2. THEMATIC SYNTHESIS OF THE AIAG STUDIES

2.1. Seeing geographical phenomena more clearly

A first group of studies uses AI to map features and processes that are difficult to identify consistently through conventional interpretation. This is especially relevant to satellite and UAV imagery, where the growing volume of observations has made manual analysis increasingly impractical. Remote-sensing archives provide repeated coverage of large areas, but the images still need to be classified, segmented and interpreted in relation to conditions on the ground. Intarat, Kuankhamnuan and Jangsawang [1] examine this problem through flood mapping in Phra Nakhon Si Ayutthaya Province, Thailand. They combine Sentinel-1 synthetic aperture radar imagery with a U-Net segmentation model using different ResNet encoders. The model distinguishes flooded from non-flooded pixels and produces continuous maps of the inundated area. Sentinel-1 data are well suited to this task because radar observations remain available under cloud cover and at night, when optical imagery may be unavailable or difficult to interpret. The comparison of the three encoders also shows that increasing network depth does not necessarily improve the result. ResNet101 provided a better balance between segmentation accuracy, generalisation and computational demand than ResNet50 or ResNet152. Maps produced soon after a flood can support the identification of affected settlements, roads and agricultural land and can be compared with the spatial extent of earlier events. Their wider operational use, however, depends on processing speed, computing capacity and validation across several floods and geographical settings. Good performance within the original study area does not guarantee that the same model will work equally well elsewhere, since part of what it learns is specific to the landscape and the radar characteristics of the training data.

Suppawimut and Samphutthanont [10] investigate the detection of agricultural burning in rice-growing areas of northern Thailand. Reference burned-area masks are first derived from the differenced Normalized Burn Ratio, after which a U-Net model is trained to identify burned patches from true-colour satellite imagery. The spectral index is therefore used to prepare the training data, while the resulting model operates with simpler RGB inputs. A particularly relevant part of the study is the spatial separation of the training, validation and test areas. Nearby pixels and image tiles often resemble one another, so random sampling can allow almost identical spatial patterns to appear in both the training and test datasets, leading to overly optimistic estimates of model performance. The spatial hold-out design provides a more realistic assessment of how well the model performs in previously unseen areas. The resulting burn maps can also show where agricultural fires are concentrated and where they may contribute to seasonal haze and fine-particle pollution. In this way, the method provides spatial information that can support the targeting of mitigation measures without treating image classification as an end in itself.

Wunaphai and colleagues [9] work at a finer spatial scale, combining UAV imagery, canopy-height models, field measurements and deep-learning-based tree detection to estimate biomass and carbon stocks in rubber plantations. Tree height, crown dimensions and stand density are derived from the UAV data, while the field plots provide the biomass measurements needed for calibration. Deep learning is mainly used to identify individual trees and organise the canopy information into a plantation inventory. The value of the approach lies in the connection it creates between a relatively small number of field plots and a much more detailed spatial representation of the plantation. Field observations therefore remain essential; the model does not replace them, but extends their usefulness beyond the locations where measurements were collected.

2.2. Estimating what cannot be measured everywhere

A second group of studies uses AI to estimate environmental conditions in areas where direct observations are limited. This is a common problem in geographical research: monitoring stations are unevenly distributed, field measurements are costly, and conditions may vary considerably within the same administrative region. Models can partly address these gaps by combining observations from several sources and estimating values for locations where no direct measurements are available.

Thabthong and colleagues [8] apply this approach to the estimation of daily $PM_{2.5}$ concentrations in Rayong Province, a coastal and industrial area of Thailand. They combine measurements from ground stations with satellite-derived aerosol optical depth and meteorological variables obtained from reanalysis data, including temperature, relative humidity, wind speed and boundary-layer height. Random Forest performed best among the methods tested. The analysis also identified wind speed, humidity and boundary-layer conditions as relevant factors in the spatial and temporal variation of pollution concentrations. Each dataset contributes different information. Monitoring stations provide direct measurements of $PM_{2.5}$, but only for a limited number of locations. Satellite observations cover a much larger area, although aerosol optical depth is not a direct measure of particle concentrations near the ground. Reanalysis data add information about the atmospheric conditions that affect the transport, dispersion and accumulation of pollutants. By relating these sources to one another, the model produces a more spatially complete estimate of $PM_{2.5}$ than could be obtained from the station network alone.

Aji, Purnama and Husna [5] use a similar combination of data sources to examine riverine plastic pollution in Kendal Regency, Indonesia. Their analysis brings together Sentinel-2 imagery, land-use and land-cover data, precipitation records and a remotely sensed Plastic Index. Higher index values were generally associated with settlements and industrial areas, while forests and plantations tended to show lower values. The relationship with rainfall was weaker and varied across the study area. These results suggest that the distribution of the index is more closely related to the location of human activities than to precipitation alone. The study is useful in this respect because it does not treat every spatial correlation as equally informative. Instead, it allows several possible explanations for the observed pattern to be considered together, including rainfall, land use, industrial activity and settlement distribution. The same principle applies to the $PM_{2.5}$ study, where pollution levels reflect not only the location of emission sources but also the atmospheric conditions governing their dispersion. In both cases, the value of the model lies not simply in reproducing a spatial pattern, but in showing which geographical factors are most closely associated with it.

2.3. Learning from environmental time series

Geographical processes develop over time as well as across space. Several studies in the volume therefore examine whether AI methods can identify temporal relationships, predict environmental variables and improve the selection of suitable modelling approaches.

Nurdiati and colleagues [3] analyse monthly fire hotspots in Kalimantan using climatic indicators that include rainfall, the duration of dry spells, the El Niño–Southern Oscillation and the Indian Ocean Dipole. They compare Gaussian Process Regression with Long Short-Term Memory networks. The LSTM model performs better on the test data, suggesting that it captures relationships extending across successive time steps more effectively. Neither approach, however, reproduces all of the observed changes. In particular, both models have difficulty accounting for a substantial reduction in fire hotspots that appears to be associated with socioeconomic intervention rather than climatic variation. This result indicates a limitation of the predictor set rather than only of the algorithms themselves. Fire occurrence depends on rainfall and large-scale climate variability, but it is also influenced by land management, agricultural practices, enforcement, market conditions and policy measures. A model based exclusively on environmental variables can therefore represent only part of the process. The remaining prediction errors may provide an indication of where important human influences have not been included.

Jumadi and colleagues [6] compare a wider range of machine-learning and deep-learning methods for rainfall prediction in the Bengawan Solo River Basin. Nine algorithms are tested using a long CHIRPS rainfall record sampled at several hundred grid points. Extreme Gradient Boosting and a Gated Recurrent Unit model produce the most stable results, whereas some of the more elaborate deep-learning architectures show larger errors and stronger variation between months. The comparison shows that increasing model complexity does not necessarily improve rainfall prediction. Performance also depends on the length and consistency of the time series, the temporal resolution of the data, spatial differences within the basin and the way the models are trained and evaluated. For this reason, the comparison of algorithms is not simply a technical preparation for the analysis. It helps establish which modelling assumptions are best suited to the available data.

Taken together, the two studies also point to different sources of predictive uncertainty. In rainfall modelling, errors may arise because an algorithm does not capture the temporal structure of the data sufficiently well. In the fire-hotspot study, part of the uncertainty results from the absence of socioeconomic variables that could explain changes not driven by climate. A more complex temporal model may improve the representation of dependencies within a time series, but it cannot account for processes that are missing from the input data.

2.4. From prediction to geographical scenarios

Prediction in geographical research does not always refer to estimating a single value for the following day or month. It may also involve constructing spatial scenarios that show how a landscape could develop if recently observed patterns and relationships continue.

Triyatno and colleagues [2] apply this approach in the Kambang watershed of West Sumatra. They combine supervised land-cover classification with Cellular Automata–Markov Chain modelling and Random Forest estimation of landslide potential. Land-use and land-cover patterns are projected for 2029 and 2039, after which the simulated changes are used to assess possible shifts in landslide susceptibility. Particular attention is given to the expected reduction in primary forest and the expansion of other land uses. The resulting maps indicate both the projected extent of highly susceptible terrain and the areas where this increase is most likely to occur.

Sideng, Haris and Lamada [4] use a related modelling framework in Enrekang Regency, where an artificial neural network is combined with cellular automata to simulate land-use change and examine its relationship with flood risk. The neural network estimates the association between land-use transitions and their driving factors, while the cellular component accounts for the influence of neighbouring cells on the location of future change. Their results suggest that built-up areas may continue to expand into zones already exposed to flooding. The relevance of the simulation lies less in predicting the exact future state of individual pixels than in identifying where current development tendencies may create additional risk. Cellular automata are useful here because land-use change is spatially dependent. Development often follows existing settlements and roads, while slope, elevation and neighbouring land uses may either encourage or restrict expansion. The model therefore provides a conditional scenario: it shows how the landscape might develop if the relationships identified during the calibration period remain broadly similar.

Onphua, Teejuntuk and Na Takuathung [12] examine how the choice of calibration period affects this type of projection. In the Loe Tor Royal Project area, they compare several land-cover classification algorithms and test CA–Markov models based on three-, five- and ten-year intervals. The model calibrated over the shortest period produces the most reliable results, whereas the ten-year interval performs less well. The authors relate this difference to changes in transition probabilities over time and to spectral inconsistencies between images acquired by different sensors. Their findings show that a longer historical record is not necessarily more suitable for model calibration. A long interval may combine periods with different planning policies, technologies, land-management practices or data characteristics, even though the model treats them as parts of the same process. Selecting a calibration period therefore requires attention not only to the number of available years, but also to whether the observations represent a reasonably consistent geographical and institutional context.

2.5. Connecting data-driven models with physical knowledge

The volume also includes studies in which data-driven models are combined with physical constraints, process knowledge or variables designed to represent geographical conditions more directly. In these cases, AI is used as one component of the analysis rather than as an independent source of explanation.

Katiba, Belmajdoub and Minaoui [7] propose a hybrid framework in which a BiLSTM–MLP network predicts precipitation, while a reinforcement-learning controller tests small changes in selected atmospheric variables. The search is restricted by stability and energy constraints so that the simulated combinations remain physically plausible. The framework is presented as a tool for exploring the sensitivity of precipitation to changes in atmospheric conditions and for informing water-management analysis, not as a system for modifying weather in practice. The distinction is important because statistical associations between rainfall, pressure, humidity and wind do not imply that these variables can be controlled independently in the atmosphere. The main contribution of the study lies in the way it combines prediction with a constrained exploration of possible atmospheric states. Rather than allowing the controller to search without limits, the model excludes solutions that would conflict with basic physical conditions.

Nguyen and colleagues [11] incorporate physical information at an earlier stage of the modelling process. Their study develops a lithological weakness index from shear-strength data and uses it as an explanatory variable in GIS-based landslide-susceptibility modelling along mountainous road corridors in northern Vietnam. This provides a more direct representation of material resistance than a conventional classification based only on broad geological units. The study shows that model performance depends partly on how well the input variables reflect the processes being investigated. In landslide analysis, replacing a geological category with an indicator of mechanical weakness may improve the physical meaning of the model even when the underlying algorithm remains unchanged. The same consideration applies to other commonly used predictors, including slope, rainfall and land cover. These variables are not direct representations of environmental processes, but simplified measures constructed from particular datasets and assumptions. Their definition therefore requires geographical and physical judgement before they are introduced into the model.

3. WHAT GEOGRAPHY CONTRIBUTES TO AI

The studies discussed in the previous section show how AI can transform satellite images, time series and field measurements into flood maps, pollution estimates, rainfall forecasts, biomass inventories and land-use scenarios. Yet these outputs do not derive their geographical meaning from the algorithms alone. Before a model can recognise, estimate or predict anything, the research must define what constitutes the phenomenon, at what scale it should be observed, which relationships matter and how the result will be interpreted.

Geography therefore contributes more than spatial data to AI. It provides concepts for distinguishing meaningful entities, methods for organising heterogeneous evidence, an understanding of scale and spatial dependence, and knowledge of the environmental and social processes behind observed patterns. It also connects model outputs to places, institutions and decisions. Without this framework, an AI system may produce an accurate classification or prediction whose geographical significance remains uncertain (Janowicz et al., 2020; Wang et al., 2024; Mai et al., 2025).

3.1. Defining meaningful geographical entities

AI models require target categories, but geographical entities are rarely given in an unambiguous form. A flood boundary, burned area, land-use class, hazardous slope or urban neighbourhood must first be defined through observational rules and disciplinary knowledge. These definitions determine what the model learns.

Flood mapping illustrates the problem. Radar backscatter can help distinguish inundated surfaces, but low values may also represent permanent water, smooth artificial surfaces or radar shadow. The models used by Intarat, Kuankhamnuan and Jangsawang [1] can segment likely floodwater, but the resulting class becomes geographically meaningful only when it is interpreted in relation to floodplain morphology, drainage, land cover and the timing of image acquisition.

The preparation of reference data raises similar questions in burned-area mapping. Suppawimut and Samphutthanont [10] use the differenced Normalized Burn Ratio to create training masks for a deep-learning model. These masks are not neutral observations. Their meaning depends on the selected spectral threshold, the dates compared and the distinction between agricultural fire, harvest, bare soil and other forms of surface change. Geography contributes to AI here by defining the environmental process represented by the label.

Some geographical entities are multidimensional rather than directly visible. A geomorphosite, for example, is not simply a terrain form that can be recognised from elevation data. Irimuș et al.

evaluate volcanic landforms through scientific, scenic, cultural-historical, ecological and socioeconomic criteria (Irimuș et al., 2015). Computer vision might identify volcanic cones or craters, but it cannot derive the full significance of a site from morphology alone. The geographical classification combines observable form with scientific interpretation and place-based value.

Comparable issues arise in human geography. Accessibility may refer to proximity, travel time, available opportunities or the capacity of different population groups to reach them. Each definition measures something different and may lead to different planning conclusions (Geurs & van Wee, 2004). Health exposure is equally dependent on definition. A temperature value assigned to a residence, a neighbourhood average and the conditions experienced during daily mobility are not interchangeable measures (Clark et al., 2025).

The target variable is therefore already a geographical argument. AI can improve the consistency or speed of classification, but geography determines what the categories represent and which distinctions should remain visible.

3.2. Representing scale, connectivity and temporal structure

Geographical processes do not occur independently at individual pixels or observation points. They develop within neighbourhoods, catchments, transport networks, routes and regions. The spatial unit used in a model influence both the pattern detected and the conclusion drawn from it. Land-use change provides a clear example. Cellular automata are useful because the probability of transition depends partly on neighbouring cells. Development tends to follow existing settlements and roads, while slope, protected areas and surrounding land uses may constrain its direction. The studies by Triyatno et al. [2], Sideng et al. [4] and Onphua et al. [12] combine classification or machine-learning methods with CA–Markov approaches precisely because land-use transitions are spatially connected.

The choice of temporal interval matters as well. Onphua et al. [12] find that the shortest calibration period produces the most reliable projection. A longer record is not automatically better when it combines periods governed by different policies, land-management practices, technologies or sensor characteristics. Geography contributes by asking whether observations belong to the same underlying spatial and institutional process before they are treated as one training dataset.

Hydrological modelling faces a similar problem. Catchments differ in terrain, soils, vegetation, storage and river-network organisation. Models trained across many basins can identify shared rainfall–runoff relationships and transfer some information to poorly monitored regions (Kratzert et al., 2019). The catchment cannot, however, be treated as an arbitrary collection of grid cells. Flow connectivity, upstream–downstream relations and water balance remain part of the structure of the problem (Nearing et al., 2021).

Regional variation is equally important. Blöschl et al. found that river floods have increased in some parts of Europe and decreased in others, reflecting different combinations of precipitation, snow processes and catchment conditions (Blöschl et al., 2019). A model fitted to the continent as a whole could obscure these contrasting regional processes even if its overall statistical performance were acceptable.

Temporal relations also require geographical and process-based interpretation. Gasparrini et al. show that temperature may affect mortality over several subsequent days and that the exposure–response relationship differs between locations (Gasparrini et al., 2015). The relevant observation is therefore not simply the temperature and mortality recorded on the same date. Lag structure, seasonality and local adaptation form part of the phenomenon being modelled.

Route-level risk offers a smaller-scale illustration. Tourist-trail conditions depend on the sequence of terrain segments, exposure, weather, communication coverage and the characteristics of the user. A trail cannot be represented adequately by one average slope or a single hazard value (Magyari-Sáska, 2014). Its meaning comes from the connection between segments and from changing conditions along the route.

Geography thus contributes the units and relations through which AI interprets spatial data. It indicates when observations should be treated as neighbours, parts of a network, stages in a sequence or members of distinct regions.

3.3. Organising heterogeneous geographical evidence

Many geographical questions cannot be answered from one data source. The $PM_{2.5}$ study by Thabthong et al. [8], for example, combines monitoring stations, satellite aerosol optical depth and meteorological reanalysis. Each source observes a different aspect of atmospheric pollution. Ground stations measure near-surface concentrations directly but provide limited spatial coverage. Satellite data offer broader coverage but do not measure ground-level $PM_{2.5}$ directly. Reanalysis adds information on the atmospheric conditions controlling transport and dispersion.

The model can relate these sources statistically, but geographical reasoning is needed to understand why they can be combined and where their meanings differ. A continuous prediction surface should not conceal the fact that some values are measured while others are inferred.

Aji, Purnama and Husna [5] similarly combine a remotely sensed plastic index, rainfall and land-use information. The weaker relationship with precipitation and stronger association with settlements and industrial areas suggest that the spatial pattern cannot be explained by hydrology alone. Land use, waste generation and human activity offer competing or complementary interpretations. AI can rank variables or identify interactions, but the explanation requires knowledge of how pollutants enter and move through the river system.

Heterogeneity is even greater when numerical and documentary evidence are combined. Historical GIS links maps, archival texts, place names and changing administrative boundaries (Gregory & Healey, 2007). Natural-language processing may extract names and dates, but older spellings, transliterations and historical changes in place identity require contextual interpretation.

Multi-criteria evaluation brings together another mixture of evidence. In developing a geopark-attractiveness model, Nyulas et al. reviewed 52 assessment methods and 862 variables before selecting 33 attributes concerning natural and cultural heritage, accessibility, infrastructure, scientific value, safety and other dimensions (Nyulas et al., 2024). This reduction is not merely a technical feature-selection problem. It reflects the purpose of geopark evaluation, institutional requirements and decisions about which values should be represented.

Similar considerations apply to ecosystem-service assessment. Land-cover classes may provide a convenient basis for expert scoring, but the same class can support different services depending on soils, climate, management and landscape context (Burkhard et al., 2009). Machine learning can test these generalisations against observations, yet the services themselves remain defined through ecological knowledge and social priorities.

Geospatial knowledge graphs offer one way to preserve these distinctions. They can represent settlements, geosites, roads, protected areas, documents and organisations as different kinds of entities connected by explicit relationships (Bordogna & Fugazza, 2023; Mai et al., 2022). Their value lies not simply in storing more information, but in retaining the meaning and provenance of the links among sources.

Geography contributes to data integration by determining which forms of evidence are commensurable, which differences must be retained and how uncertainty should be carried from the original observations into the model output.

3.4. Keeping environmental and social processes visible

Predictive performance does not necessarily provide an explanation. A model can reproduce an observed pattern without identifying the process that generated it. Geographical knowledge is therefore needed both when predictors are selected and when the remaining errors are interpreted.

The fire-hotspot study by Nurdiati et al. [3] provides a useful example. Climate variables explain part of the temporal variation, and the LSTM model performs better than Gaussian Process Regression. Neither model fully captures a marked decline in hotspot occurrence that appears to be related to socioeconomic intervention. The error is informative because it suggests that the predictor set omits land-management, regulatory or behavioural changes. Increasing network complexity would not solve the absence of relevant variables.

Rainfall modelling leads to a related conclusion. Jumadi et al. [6] compare nine algorithms and find that some relatively compact models perform more consistently than more elaborate deep-learning architectures. The choice of method must correspond to the length, temporal resolution and spatial structure of the available record. Complexity alone does not supply missing process information.

The physical meaning of predictors also matters. Nguyen et al. [11] derive a lithological weakness index from shear-strength data rather than relying only on broad geological classes. The improvement lies partly in representing material resistance more directly. The example shows that a geographically and physically meaningful variable may contribute more than a change of algorithm.

Katiba, Belmajdoub and Minaoui [7] take this further by restricting a reinforcement-learning search through atmospheric stability and energy constraints. The controller is not allowed to explore every statistically advantageous combination of variables. Its solutions must remain compatible with basic physical conditions.

Theory-guided and hybrid approaches generalise this principle. Conservation laws, process-model outputs and known environmental relationships can be incorporated into machine-learning systems to reduce physically implausible predictions (Karpatne et al., 2017; Reichstein et al., 2019; Irrgang et al., 2021). Such models recognise that the available data may not contain enough examples to teach every relevant constraint.

Social processes require similar caution, although they cannot usually be represented through physical equations. Satellite imagery may predict economic conditions from buildings, roads and cultivated land, but it cannot directly observe tenure, household relations or informal employment (Jean et al., 2016; Yeh et al., 2020). A transport model may learn low demand in a poorly served neighbourhood without recognising that the lack of service itself suppresses travel.

Geography helps distinguish prediction from explanation. It asks whether a model has identified a process, a proxy or merely a stable correlation, and whether the relationship is likely to persist when environmental or institutional conditions change.

3.5. Connecting model outputs to geographical decisions

The output of an AI model becomes useful only when it can inform a geographical decision. A flood mask must be related to settlements, roads, agricultural land or evacuation routes. A pollution surface must be interpreted in relation to population exposure. A land-use scenario must be evaluated against planning objectives and environmental constraints.

These decisions introduce values that cannot be inferred from accuracy alone. A highly accurate accessibility model cannot determine whether the existing distribution of services is fair. A tourism recommender trained on previous visits may reproduce the concentration of visitors in already popular places. A geopark assessment must balance conservation, accessibility, education and local development. The algorithm can compare alternatives, but it cannot establish the priorities on its own.

Scenario models require similar interpretation. The land-use projections in the volume do not describe a predetermined future. They show what may occur if the relationships observed during the calibration period continue. Their planning value lies in identifying where present tendencies could increase flood or landslide exposure, not in claiming certainty about individual future pixels.

Operational warning systems add questions of responsibility. Flood forecasts, heat warnings and trail-risk assessments should indicate where observations are sparse, when the data were updated and how certain the prediction is. A model may support a decision, but formal responsibility remains with the relevant authorities and professional organisations.

Cartography also plays a role in this transition. Model results must be presented in a form that allows users to distinguish predicted values, uncertainty, observed data and areas of extrapolation. Automated mapping can improve efficiency, but map design remains an interpretive act involving classification, generalisation and visual emphasis (Usery et al., 2022).

The geographical contribution continues after the model has produced its prediction. It connects the result to affected places and populations, evaluates whether the output is suitable for its intended purpose and makes the assumptions and uncertainties visible.

3.6. Geography as part of the model, not merely its application area

The twelve studies in this volume demonstrate the breadth of AI applications in geographical research. They also show that the most successful models depend on decisions made outside the algorithm: how reference classes are defined, how spatial units are connected, which temporal period is representative, which variables reflect the process and how outputs are linked to action.

Geography contributes these decisions. It defines meaningful entities, represents scale and connectivity, organises heterogeneous evidence, supplies environmental and social interpretation, and relates predictions to institutional and public objectives. Field inventories, epidemiological models, accessibility frameworks, heritage assessments and physical process knowledge are not preliminary forms of AI. They are sources of the structure that AI needs.

This perspective also changes how model improvement should be understood. A deeper architecture or larger dataset may raise predictive accuracy, but improvement may also come from a better geographical unit, a more meaningful predictor, an appropriate temporal lag or a clearer distinction between observation and inference. In many cases, these choices have a greater effect on the usefulness of the result than the selection of the algorithm.

GeoAI is therefore strongest when geography remains present throughout the research process: before modelling, when the problem and data are defined; within the model, when spatial relations and process constraints are represented; and after modelling, when predictions are interpreted, mapped and used. Geography does not simply supply locations for artificial intelligence. It provides the concepts through which spatial predictions become knowledge.

4. CLOSING REFLECTIONS

The value of an AI model in geography depends not only on predictive accuracy, but also on whether it is tested under realistic spatial conditions. Nearby observations often share similar environmental and social characteristics, so randomly dividing them between training and test datasets may produce overly optimistic results. In such cases, the model is largely interpolating within familiar surroundings rather than predicting in genuinely new areas. Spatially informed cross-validation provides a more reliable assessment, although its design must reflect the intended application, scale and sampling structure (Wang et al., 2023).

Transferability must likewise be demonstrated rather than assumed. A model developed for one city, watershed, mountain region or tourism destination learns relationships shaped by local climate, morphology, land use, institutions and data quality. Transfer learning may reduce the amount of new training data required, but it cannot replace local calibration and independent validation (Ma et al., 2024). A transferable methodology is therefore not the same as a fitted model that can be applied unchanged in every region.

Predictions must also be understandable to their intended users. A geomorphologist may need to know which terrain properties produced a classification, a planner which criteria altered a suitability score, and a hiker whether a warning reflects steep slopes, snow or an approaching storm. General feature-importance rankings are often insufficient because they may neither explain individual predictions nor establish causal relationships. Explanations should therefore combine model diagnostics with disciplinary knowledge, uncertainty and the institutional context in which the result will be used (Gevaert, 2022).

Uncertainty should be communicated spatially. A single accuracy value may conceal poor performance in particular regions, seasons, terrain types or land-cover classes. Confidence surfaces, prediction intervals and maps showing distance from the training domain can reveal where results are well supported and where they depend mainly on extrapolation. This is particularly important in health, safety and environmental management, where the consequences of errors vary geographically.

The choice of predictors, target variables and evaluation criteria remains a geographical decision. AI can optimise a model after the research problem has been defined, but it cannot independently decide whether tourism development should prioritise accessibility or conservation, whether a route is acceptably safe, or whether a landform has cultural significance. Such questions require scientific interpretation, local knowledge, institutional objectives and public values.

The wider use of AI in geography can be understood as a progression from recognising spatial entities, through estimating and comparing their characteristics, to anticipating change and supporting decisions. Across these stages, established geographical frameworks define the entities, relationships, scales and evaluation criteria that AI systems must respect.

AI is therefore most useful not when it replaces GIS, field surveys, statistical analysis, physical modelling or expert interpretation, but when it extends them. It can process more observations, update assessments more frequently and compare alternatives that would be difficult to evaluate manually. The studies considered here also show that greater model complexity is not automatically beneficial: deeper architectures, longer calibration periods or more elaborate methods may perform less reliably than simpler alternatives.

The future of AI in geography will depend on careful integration with geographical knowledge. Models must be grounded in place, validated across meaningful spatial and temporal separations, interpreted through environmental and social processes, and accompanied by visible uncertainty. The objective is not simply to produce more predictions, but to create spatial knowledge that can be examined, challenged and used responsibly.

REFERENCES

Blöschl, G., Hall, J., Viglione, A., et al. (2019). Changing climate both increases and decreases European river floods. *Nature, 573*, 108–111. https://doi.org/10.1038/s41586-019-1495-6

Bordogna, G., & Fugazza, C. (2023). Artificial intelligence for multisource geospatial information. *ISPRS International Journal of Geo-Information, 12*(1), 10. https://doi.org/10.3390/ijgi12010010

Burkhard, B., Kroll, F., Müller, F., & Windhorst, W. (2009). Landscapes' capacities to provide ecosystem services: A concept for land-cover-based assessments. *Landscape Online, 15*, 1–22. https://doi.org/10.3097/LO.200915

Clark, L. P., Zilber, D., Schmitt, C., Fargo, D. C., Reif, D. M., Motsinger-Reif, A. A., & Messier, K. P. (2025). A review of geospatial exposure models and approaches for health data integration. *Journal of Exposure Science & Environmental Epidemiology, 35*(2), 131–148. https://doi.org/10.1038/s41370-024-00712-8

Gasparrini, A., Guo, Y., Hashizume, M., et al. (2015). Mortality risk attributable to high and low ambient temperature: A multicountry observational study. *The Lancet, 386*(9991), 369–375. https://doi.org/10.1016/S0140-6736(14)62114-0

Gevaert, C. M. (2022). Explainable AI for Earth observation: A review including societal and regulatory perspectives. *International Journal of Applied Earth Observation and Geoinformation, 112*, 102869. https://doi.org/10.1016/j.jag.2022.102869

Geurs, K. T., & van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: Review and research directions. *Journal of Transport Geography, 12*(2), 127–140. https://doi.org/10.1016/j.jtrangeo.2003.10.005

Gregory, I. N., & Healey, R. G. (2007). Historical GIS: Structuring, mapping and analysing geographies of the past. *Progress in Human Geography, 31*(5), 638–653. https://doi.org/10.1177/0309132507081495

Irrgang, C., Boers, N., Sonnewald, M., Barnes, E. A., Kadow, C., Staneva, J., & Saynisch-Wagner, J. (2021). Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nature Machine Intelligence, 3*, 667–674. https://doi.org/10.1038/s42256-021-00374-3

Irimuș, I. A., Bálint-Bálint, L., Dombay, Ș., Crișan, H. F., & Magyari-Sáska, Z. (2015). Classification and evaluation criteria for volcanic geomorphosites in Harghita Mountains. In *Proceedings of the 15th International Multidisciplinary Scientific GeoConference SGEM 2015* (Book 1, Vol. 1, pp. 77–84). https://doi.org/10.5593/SGEM2015/B11/S1.010

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). GeoAI: Spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science, 34*(4), 625–636. https://doi.org/10.1080/13658816.2019.1684500

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science, 353*(6301), 790–794. https://doi.org/10.1126/science.aaf7894

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., & Kumar, V. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering, 29*(10), 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences, 23*(12), 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

Ma, Y., Chen, S., Ermon, S., & Lobell, D. B. (2024). Transfer learning in environmental remote sensing. *Remote Sensing of Environment, 301*, 113924. https://doi.org/10.1016/j.rse.2023.113924

Magyari-Sáska, Z. (2014). Quantifying threats along tourist trails: An initial approach. *Geographia Technica, 9*(1), 78–86.

Mai, G., Hu, Y., Gao, S., Cai, L., Martins, B., Scholz, J., Gao, J., & Janowicz, K. (2022). Symbolic and subsymbolic GeoAI: Geospatial knowledge graphs and spatially explicit machine learning. *Transactions in GIS, 26*(8), 3118–3124. https://doi.org/10.1111/tgis.13012

Mai, G., Xie, Y., Jia, X., Lao, N., Rao, J., Zhu, Q., Liu, Z., Chiang, Y.-Y., & Jiao, J. (2025). Towards the next generation of geospatial artificial intelligence. *International Journal of Applied Earth Observation and Geoinformation, 136*, 104368. https://doi.org/10.1016/j.jag.2025.104368

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., & Gupta, H. V. (2021). What role does hydrological science play in the age of machine learning? *Water Resources Research, 57*(3), e2020WR028091. https://doi.org/10.1029/2020WR028091

Nyulas, J., Dezsi, Ș., Haidu, I., Magyari-Sáska, Z., & Niță, A. (2024). Attractiveness assessment model for evaluating an area for a potential geopark—Case study: Hațeg UNESCO Global Geopark (Romania). *Land, 13*(2), 148. https://doi.org/10.3390/land13020148

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature, 566*, 195–204. https://doi.org/10.1038/s41586-019-0912-1

Usery, E. L., Arundel, S. T., Shavers, E. J., Stanislawski, L. V., Thiem, P. T., & Varanka, D. E. (2022). GeoAI in the US Geological Survey for topographic mapping. *Transactions in GIS, 26*(1), 25–40. https://doi.org/10.1111/tgis.12830

Wang, S., Huang, X., Liu, P., Zhang, M., Biljecki, F., Hu, T., Fu, X., Liu, L., Liu, X., Wang, R., Huang, Y., Yan, J., Jiang, J., Chukwu, M., Reza Naghedi, S., Hemmati, M., Shao, Y., Jia, N., Xiao, Z., … Bao, S. (2024). Mapping the landscape and roadmap of geospatial artificial intelligence (GeoAI) in quantitative human geography: An extensive systematic review. *International Journal of Applied Earth Observation and Geoinformation, 128*, 103734. https://doi.org/10.1016/j.jag.2024.103734

Wang, Y., Khodadadzadeh, M., & Zurita-Milla, R. (2023). Spatial+: A new cross-validation method to evaluate geospatial machine-learning models. *International Journal of Applied Earth Observation and Geoinformation, 121*, 103364. https://doi.org/10.1016/j.jag.2023.103364

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D. B., Ermon, S., & Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications, 11*, 2583. https://doi.org/10.1038/s41467-020-16185-w

Aims and Scope

Geographia Technica is a journal devoted to the publication of all papers on all aspects of the use of technical and quantitative methods in geographical research. It aims at presenting its readers with the latest developments in G.I.S technology, mathematical methods applicable to any field of geography, territorial micro-scalar and laboratory experiments, and the latest developments induced by the measurement techniques to the geographical research.

Geographia Technica is dedicated to all those who understand that nowadays every field of geography can only be described by specific numerical values, variables both oftime and space which require the sort of numerical analysis only possible with the aid of technical and quantitative methods offered by powerful computers and dedicated software.

Our understanding of ***Geographia Technica*** expands the concept of technical methods applied to geography to its broadest sense and for that, papers of different interests such as: G.I.S, Spatial Analysis, Remote Sensing, Cartography or Geostatistics as well as papers which, by promoting the above mentioned directions bring a technical approach in the fields of hydrology, climatology, geomorphology, human geography territorial planning are more than welcomed provided they are of sufficient wide interest and relevance.

Targeted readers:

The publication intends to serve workers in academia, industry and government.
Students, teachers, researchers and practitioners should benefit from the ideas in the journal.

Guide for Authors

Submission

Articles and proposals for articles are accepted for consideration on the understanding that they are not being submitted elsewhere.

The publication proposals that satisfy the conditions for originality, relevance for the new technical geography domain and editorial requirements, will be sent by email to the address editorial-secretary@technicalgeography.org.

This page can be accessed to see the requirements for editing an article, and also the articles from the journal archive found on www.technicalgeography.org can be used as a guide.

Content

In addition to full-length research contributions, the journal also publishes Short Notes, Book reviews, Software Reviews, Letters of the Editor. However the editors wish to point out that the views expressed in the book reviews are the personal opinion of the reviewer and do not necessarily reflect the views of the publishers.

Each year two volumes are scheduled for publication. Papers in English or French are accepted. The articles are printed in full color. A part of the articles are available as full text on the www.technicalgeography.org website. The link between the author and reviewers is mediated by the Editor.

Peer Review Process

The papers submitted for publication to the Editor undergo an anonymous peer review process, necessary for assessing the quality of scientific information, the relevance to the technical geography field and the publishing requirements of our journal.

The contents are reviewed by two members of the Editorial Board or other reviewers on a simple blind review system. The reviewer's comments for the improvement of the paper will be sent to the corresponding author by the editor. After the author changes the paper according to the comments, the article is published in the next number of the journal.

Eventual paper rejections will have solid arguments, but sending the paper only to receive the comments of the reviewers is discouraged. Authors are notified by e-mail about the status of the submitted articles and the whole process takes about 3-4 months from the date of the article submission.

Indexed by: **CLARIVATE ANALYTICS**
SCOPUS
GEOBASE
EBSCO
SJR
CABELL

ISSN: 1842 - 5135 (Print)
ISSN: 2065 - 4421 (Online)