Global J. Environ. Sci. Manage. 10(4): 1501-1518, Autumn 2024, Serial #40

Global Journal of Environmental Science and Management
(GJESM)

Homepage: https://www.gjesm.net/

Podcasts

ORIGINAL RESEARCH PAPER

Analysis of sensitive urban form indicators of flood susceptible prediction based on machine learning models

M. Srivanit[1], S. Pattanasri[1], N. Phichetkunbodee[2], S. Manokeaw[3], S. Sitthikankun[4], D. Rinchumphu[5,*]

[1] *Urban Planning Programs, Thammasat University, Thailand*

[2] *Department of Civil Engineering, National Taiwan University, Taiwan*

[3] *Office of Research Administration, Chiang Mai University, Thailand.*

[4] *Department of Industrial Technology, Chiang Mai Rajabhat University, Thailand*

[5] *Department of Civil Engineering, Chiang Mai University, Thailand*

ARTICLE INFO

Article History:
Received 03 Feruary 2024
Revised 09 April 2024
Accepted 11 June 2024

Keywords:
Flood susceptible
Machine learning
Resilient City
Urban form

DOI: 10.22034/gjesm.2024.04.02

ABSTRACT

BACKGROUND AND OBJECTIVES: Flooding is one of the biggest challenges affecting the economy and people's well-being. Various approaches have been employed in prior research to examine spatial data and enhance flood response effectiveness, such as the utilization of machine learning methodologies to aid in decision-making within the realm of urban planning. However, different machine learning models serve different purposes depending on their learning processes and computation techniques. The objective of this study is to create a machine learning algorithm that can evaluate areas prone to flooding. The study purpose is to offer valuable insights to city authorities and aid in urban planning, ultimately enhancing the safety and adaptability of the city's inhabitants

METHODS: In order to evaluate flood risk areas and offer valuable insights for urban management and planning, a technique was devised. A total of eleven multiclass classification algorithms were employed to analyze eight urban factors, enabling the assessment of flood risk. The outcomes were then visualized on a map using geographic information systems.

FINDINGS: The study discovered that the Bagging Decision Tree Algorithm model produced the best flood risk assessment model, with an accuracy of 88.58 percent compared to the government's flood simulation model results. Furthermore, precipitation, building coverage ratio, and floor area ratio emerged as the top three crucial factors influencing the likelihood of flooding in the city.

CONCLUSION: The Bagging Decision Tree Algorithm model provides a thorough evaluation of flood risk, presenting crucial information for urban governance and development. Integrating key variables such as rainfall, building coverage ratio, and floor area ratio into flood risk management strategies is crucial for mitigating the impact of floods in economically significant urban areas.

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
48	8	9

*Corresponding Author:
Email: *damrongsak.r@cmu.ac.th*
Phone: +669 5995 9519
ORCID: 0000-0002-6037-4974

Note: Discussion period for this manuscript open until January 1, 2025 on GJESM website at the "Show Article".

INTRODUCTION

Flooding is a natural phenomenon that can be attributed to both climate change and human actions. The consequences of flooding are far-reaching, impacting various aspects of society such as the economy, infrastructure, and the daily routines of individuals. The infrastructure uses a substantial budget for repair and is also an obstacle to the transportation network, hindering the movement of goods and people across the city. In terms of society, individuals encounter obstacles pertaining to income, employment, and job displacement. As a result, adept urban planning and risk mitigation are essential in addressing these concerns (Glago, 2021). Due to these challenges, managing water in the city collaboratively is important, involving both the community and the government (Woldesenbet, 2020). Urban flooding can be effectively addressed through a comprehensive policy framework consisting of three key components. Firstly, the government should implement a flood management policy that incorporates a water management strategy specifically tailored for urban areas, with active involvement from the local population to ensure its effectiveness and relevance. Secondly, there should be a policy in place that guarantees equitable distribution of benefits to all members of the community. Lastly, it is crucial to acknowledge and address the intricate nature of urban environments to develop sustainable solutions for managing flood risks (Li *et al.,* 2020). Flood hazard maps (FHM) were developed in the past decade for various essential purposes. They aid in assessing and mitigating flood risks, informing emergency preparedness, and guiding effective response efforts. The role of understanding, mitigating, and responding to the risks associated with flooding holds great significance for a range of stakeholders, communities, and decision-makers (Atanga, 2020). Establishing a successful FHM is difficult, and several obstacles remain. It is crucial to highlight that, due to climate change and changes in relative sea level, FHM will require frequent updates to represent the growing risk of floods. If there is a lack of easily accessible high-quality data, the creation of advanced and precise maps may necessitate the use of intricate models and costly data-collection techniques (Diehl *et al.,* 2021). FHM depends on topography, long-term extreme event data, and advanced numerical modeling approaches. To address these issues, geographic information systems (GIS) technology was integrated with FHM to facilitate spatial analysis, allowing insightful decisions based on geographical data (da Silva *et al.,* 2020). A prior investigation has utilized technology and scientific methods to address flood management, exemplified by the utilization of GIS to analyze spatial data. This approach has proven to be highly effective in enhancing the efficiency of flood management strategies (Petty *et al.,* 2016). GIS systems can use remote sensing data (RS) to create spatial modeling by defining spatial variables influencing flooding and categorizing them into hazard and vulnerability (Sar *et al.,* 2015; Kokila *et al.,* 2021). Calculating the weight of each variable using the analytic hierarchy process (AHP) that takes data from many factors and stacks them together can be done quickly. Finally, a flood susceptibility map can be created by using risk scores (Burayu *et al.,* 2023). Flood risk management services now utilize advanced technology such as artificial intelligence (AI) in addition to GIS for analyzing, planning, and decision-making processes (Pham *et al.,* 2021). The intelligence of AI comes from machine learning (ML), in which output data is specified so that the model can predict results according to the conditions of the data entered by creating prediction equations from statistics (Ramkumar *et al.,* 2024). In the process of decision-making through the utilization of Machine Learning (ML), Classification Algorithms are employed to assign responses to distinct classes. These classes can be categorized into two types: the first type assigns a value of 1, indicating the occurrence of flooding, while the second type assigns a value of 0, indicating the absence of flooding. The result from the prediction emerges as a probability within the range of 0 to 1 according to the equation of the sigmoid function. Examples of commonly used classification algorithms such as logistic regression (LR), naive bayes (NB), k-nearest neighbors (KNN), support vector machine (SVM), decision trees (DTs), random forest (RF), and deep learning neural network (DLNN). The performance of these models will be evaluated through different techniques including Confusion Matrix and Receiver operating characteristic (ROC). (De Diego *et al.,* 2022). Pattaya, in Chonburi Province, Thailand, is another economically important area affected by annual flooding due to its geography. The region consists of undulating hills interspersed with flat plains, with certain sections forming a basin.

This particular locale is renowned as a significant tourist spot in Thailand, advocating for eco-tourism and cultural tourism to bolster the tourism industry. There are 14.68 million tourists and tourism income of 231,313 million baht (Sangpikul, 2020). For this reason, Pattaya City is a repeatedly flooded area that is flooded for a long time when there is heavy rain, causing damage to the well-being of the people and the economy. This study endeavors to identify an appropriate machine learning model to evaluate the areas at risk of flooding in Pattaya, stemming from the aforementioned challenges. The approach integrates city flood risk assessment tools with GIS, contributing to resilient city management planning. The study aims to analyze urban indicators influencing flood susceptibility predictions through machine learning models, thereby developing prototype models. The prototypes can help urban designers or government officials make efficient and sustainable decisions for flooding management in Pattaya City, Chon Buri Province, Thailand, in 2022 - 2023.

MATERIALS AND METHODS

In order to accurately forecast areas at risk of flooding, this study utilizes a blend of Geographic Information Systems (GIS) and Machine Learning (ML) techniques to improve the reliability and exactness of flood risk evaluations. By utilizing spatial data, this method offers in-depth understanding of flood patterns. The following sections will explain the application of GIS in flood risk prediction, the integration of ML algorithms, and the data collection and preparation processes involved in this study.

Application of GIS to predict flood risk areas

The application of GIS for assessing flood risk areas has gained plenty of academic interest. FHM has improved with advances in predicting and identifying regions in danger of floods. These advancements strive to provide policymakers with more accurate data regarding regions susceptible to flooding or affected by it. Three primary methods exist for constructing the FHM, encompassing physically-based, empirical, and physical modeling approaches (Teng *et al.,* 2017). The physical modeling approach involves experiments to validate the model's predicting abilities. Numerical models, conversely, retain their significance as long as they effectively simulate or portray the real dynamics of a flow or flood occurrence. Empirical models, on the other hand, offer numerous scientific and data-driven approaches that can be utilized. Based on statistical and data-driven methods, these methods use data such as hydrological, topographic, digital elevation model (DEM), and geomorphology. Remote sensing is sometimes utilized to acquire this data, which is subsequently analyzed through the implementation of GIS (Wang *et al.,* 2018). Physically-based models are excellent for flood prediction and early warning systems that require a large amount of input data, including hydrological parameters, river network geometry, topography features, and, on occasion, remote sensing data processing within GIS (Praskievicz *et al.,* 2020). Flood risk assessment in GIS commonly involves the use of topographic information and data on yearly precipitation. In the realm of GIS, there has been a clear focus on incorporating urban variables into flood risk analysis. Numerous research trends emphasize the significance of incorporating factors like floor area ratio (FAR) (Park and Won, 2019), building density (BD), building coverage ratio (BCR), and other urban parameters that represent urban characteristics (Lin *et al.,* 2021), into GIS to improve flood-prone region assessments. The assessment of flood risk can be enhanced by considering urban factors, which unveil the spatial distribution and intensity of floods. Due to the complexity and volume of data in assessing flood-prone areas, machine learning in flood risk analysis within the GIS framework has become increasingly crucial (Pham *et al.,* 2021). While traditional methods such as physically-based, empirical, and physical modeling approaches have significantly contributed, machine learning techniques can enhance predictive accuracy and efficiency (Shen *et al.,* 2023). Machine learning algorithms have the capability to manage a wide range of large datasets, extracting intricate patterns and relationships that could pose difficulties for traditional methods. This capability becomes particularly important when dealing with urban parameters, such as FAR, BD, and BCR, which contribute to a more comprehensive understanding of flood risk in metropolitan areas. Researchers and decision-makers can enhance their flood risk analysis by integrating machine learning techniques into GIS, leading to more accurate predictions and informed decisions when developing flood risk management plans (Rezvani *et al.,* 2023). This research aims to create a tool that assesses flood risk areas in

Pattaya City, offering valuable information for city administration and planning. The study combines GIS with multiclass classification machine learning models. The flood risk levels are classified by these models according to the eight spatial factors that impact flood susceptibility predictions. The objective is to provide a comprehensive and efficient approach to support decision-making and planning in the context of flood risk management for Pattaya City.

Machine learning-based method for assessing flood risk

Classification algorithms play a crucial role in the field of machine learning as they streamline the decision-making process by sorting items into different categories using patterns that have been learned beforehand (Shrestha *et al.,* 2021). They play a crucial role in identifying spam, recognizing images, and diagnosing medical conditions. Additionally, assessing the effectiveness of classification models is of utmost importance. According to Vujović (2021), accuracy, precision, recall, F1-score, ROC curves, and confusion matrices provide information about a model's effectiveness. Accurate performance measurement ensures reliable predictions and aids in understanding a model's strengths and weaknesses, contributing to model refinement and real-world applicability. Various classification algorithms, including LR (Vakili *et al.,* 2020), Genetic Algorithm (GA) (Saleh *et al.,* 2022), NB (Liu *et al.,* 2017), KNN (Shahabi *et al.,* 2020), SVM (Chang et al., 2018), DTs (Khosravi *et al.,* 2018), and RF (Wang *et al.,* 2015), have found use in FHM. Another significant tool in engineering risk analysis is the DLNN (Tien-Bui *et al.,* 2020). It plays a significant role in FHM work by leveraging its pattern recognition, feature extraction, and complex modeling capabilities. The following section will explain the details of each classification algorithm. An indispensable component in spam detection, image recognition, and medical diagnosis applications, they play a crucial role in ensuring accurate results. Additionally, assessing the effectiveness of classification models is of utmost importance. Validation is crucial to comprehending, enhancing, and guaranteeing the reliability of a classification model in practical situations.

Classification algorithms

LR is one of the machine learning models used for binary classification problems. According to the sigmoid function equation, the model predicts probabilities ranging from 0 to 1. In order to enable the segregation of classes, a decision boundary is established to categorize each class. This is particularly applicable in scenarios involving binary classification, where the classes are denoted as 1 (representing "yes") and 0 (representing "no"). The decision boundary is set at 0.5; if the predicted probability is more significant than 0.5, the prediction is assigned as class 1; otherwise, if it is less than 0.5, it is class 0. This process is mathematically using Eq. 1 (Zaidi and Al Luhayb, 2023)

$$y=\frac{1}{\left(1+e^{-x}\right)} \tag{1}$$

NB is a model that predicts which variable or instance belongs to the category. It is a supervised model that requires prior data labeling (LR). The fundamental principle of NB involves utilizing probability concepts to predict the category. The equation for Naive Bayes Classification, using Eq. 2 (Balayla, 2022)

$$P(c\,|\,x)=\frac{P(x\,|\,c)}{P(x)} \tag{2}$$

Where C is class, x is attribute, P is probability, posterior probability P(c|x) is the probability of the class (C) given the observed data with attributes (x). Likelihood P(x|c) is the probability of observing the attributes (x) given that the class is C and naïve attributes. Prior probability P(c) is The prior probability of a specific class (C) occurring before observing the data. Predictor Prior probability P(x) is The prior probability of observing the given set of attributes (x) across all classes.

KNN is a classification method that partitions classes by evaluating the similarity between the target data and other data points. By measuring the proximity of the target data to a specific dataset, the system can determine the most similar data and provide it as the answer. This method operates without using training data to create a model but relies on the available data directly as the model. SVM is an algorithm that addresses data classification problems commonly used for analysis and classification. This algorithm relies on finding the coefficient of the equation to create a separation line

for the input data during the training process. The primary emphasis lies on the boundary line, which serves as the most effective means of categorizing data. Decision Trees (DTs) represent a classification model based on rules; these rules are generated from the values of individual Features without the need for equations to govern the association between the feature and target variable, a relationship that can be illustrated in a tree diagram. RF is one of the models belonging to the ensemble learning group. The principle of training the same model multiple times (multiple instances) on the same data set. In each training instance, a different subset of the data is selected, and the decision-making of these models is then combined through voting to determine the most frequently chosen class. DLNN is an automated learning technique that mimics the operations of the human brain's neural network. This method consists of arranging numerous layers of neural networks and instructing them using example data. Subsequently, the data is employed for tasks such as recognizing patterns or classifying data.

Performance evaluation for classification algorithms

The evaluation and accuracy assessment of machine learning models for predicting flood-prone areas involves using metrics such as the confusion matrix or the area under the receiver operating characteristic curve (AUC-ROC). The metrics play a crucial role in evaluating the accuracy and suitability of the generated model.The confusion matrix provides insights into the model's performance by displaying true positive, true negative, false positive, and false negative values. The AUC-ROC curve provides a thorough evaluation of the model's capacity to differentiate between positive and negative cases.

Confusion matrix

Evaluating the performance of a model is imperative before deploying it for practical purposes, in order to gauge its effectiveness for potential development or application in different scenarios. Performance evaluation is typically conducted using metrics derived from the confusion matrix, a crucial table in assessing the capabilities of machine learning models in addressing classification problems. The values of the confusion matrix include:

1) Accuracy is calculated using the formula Accuracy = (True positive (TP)+ True negative (TN))/(TP+TN+ False positive (FP)+ False negative (FN), representing the proportion of correctly predicted instances to the total instances.

2) Precision (Positive predictive value) is calculated as Precision = TP/(TP+FP), representing the accuracy of the model when predicting the positive class.

3) Sensitivity or Recall is calculated as Sensitivity/ Recall = TP/(TP+FN), measuring the model's ability to identify positive instances correctly.

4) Specificity is calculated as Specificity = TN/ (TN+FP), indicating the model's ability to identify negative instances correctly.

5) The F1-score is a single metric that combines Precision and Recall, calculated as F1 = 2 * [(Precision * Recall) / (Precision + Recall)]. It offers a balanced measure of a model's overall performance without choosing between Precision and Recall.

Receiver operating characteristic

AUC stands for "Area Under <ROC> Curve," another popular metric used for model evaluation. The AUC value, ranging from 0 to 1, serves as a measure of the model's predictive accuracy for the outcome variable (y). A higher AUC value indicates a greater effectiveness in predicting the outcome variable. The AUC score is often interpreted on a scale between 0 and 1, where a value closer to 1 suggests better overall predictive performance. Researchers have suggested interpreting the results from the AUC based on score ranges (Jeong *et al.,* 2016), as outlined in Table 1.

Table 1: Criteria of ROC score

ROC score	Rating
0.9 < AUC < 1.0	excellent
0.8 < AUC < 0.9	good
0.7 < AUC < 0.8	fair
0.6 < AUC < 0.7	poor
0.5 < AUC < 0.6	fail

Factors influencing flooding

The development of FHM is vital in reducing the negative effects of flooding, thoroughly examining relevant variables using a combination of GIS and machine learning approaches. Flooding during the risk assessment process can be influenced by a multitude of factors, such as topology, precipitation, modifications in urban form, and other pertinent elements. In the field of topology variables, DEM emerges as a critical component, providing insights about terrain elevation that are crucial for identifying low-lying areas prone to flooding (McClean *et al.,* 2020). The analysis of slope and elevation plays a crucial role in identifying and characterizing various landscape features that have a significant impact flood dynamics (Cabrera and Lee, 2020). The Ogato *et al.* (2020) study also mentioned that soil type data refines the assessment even more, as changes in soil features play an essential role in water absorption and runoff, determining flood susceptibility. The precipitation variable, specifically annual rainfall data, is a cornerstone in understanding the hydrological cycle and predicting potential flooding events (O'Donnell *et al.,* 2020). Elevated levels of precipitation may result in higher runoff, underscoring the importance of this factor in flood risk evaluations. Alterations in city blocks have a notable impact on drainage patterns and flow pathways (Mohtat and Khirfan, 2022). Lin *et al.* (2021) deployed metrics like FAR and BCR to analyze flood risk levels in the context of urban parameters. FAR serves as a metric for building density, indicating the degree of urbanization, while BCR provides insights into the density of built-up areas. Land use data enhances the analysis by capturing the diverse effects of various land uses on water runoff. Urban areas characterized by extensive pavement may contribute to heightened runoff, underscoring the importance of considering land use patterns in flood hazard assessments. Another urban parameter, such as the alignment of road networks (Zaoui *et al.,* 2020), becomes a crucial variable as it influences drainage patterns. Inadequate or impeded drainage systems have the potential to worsen the likelihood of flooding occurrences. Beyond these variables, incorporating flood risk information derived from government simulation models enhances the accuracy and reliability of FHM. Machine learning algorithms, including ensemble models or neural networks, prove instrumental in synthesizing the complex interplay of these variables. The utilization of this analytical method enables the development of predictive models that can generate precise flood susceptibility maps, providing a reliable decision-making tool for individuals engaged in urban planning, emergency response, and risk mitigation efforts. In essence, the synergistic integration of GIS and machine learning techniques, coupled with a nuanced consideration of diverse variables, emerges as a pivotal methodology for advancing the development and precision of FHM. These can be concluded for the input indicators to predict the flooding model in eight indicators**: 1)** rainfall is a key variable in flood research as it directly impacts flood occurrence and severity. It measures the amount of precipitation over time, affecting surface runoff, river levels, and potential flooding. Precise precipitation data plays a vital role in forecasting potential flood hazards, simulating various scenarios, and establishing timely alert mechanisms. These aspects are of utmost importance in ensuring effective disaster preparedness and response strategies (Wu *et al.,* 2024). 2) Land use is crucial in flood research as it affects water runoff and absorption. It refers to how land is utilized, including agriculture, urban areas, forests, and wetlands. Different land uses impact how water flows and is absorbed into the ground. Urban areas, characterized by impervious surfaces such as roads and buildings, frequently result in elevated runoff levels and heightened flood vulnerability. In contrast, forests and wetlands possess the ability to absorb greater amounts of water, thereby diminishing runoff and the likelihood of flooding. An in-depth comprehension of land use is essential for effective flood risk evaluation and the development of appropriate management tactics (Marchioni *et al.,* 2022). **3)** FAR is a key urban planning metric affecting flood risk. It measures the ratio of a building's total floor area to the size of the land plot. High FAR values indicate denser developments, often leading to more impervious surfaces like concrete, reducing rainwater absorption, increasing runoff, and raising flood risks. Incorporating green spaces and sustainable drainage systems into urban planning can effectively mitigate these risks **(**Park and Lee, 2019). 4) BCR measures the percentage of a plot covered by buildings. It is calculated by dividing the area covered by buildings by the total plot area.

Table 2: All input parameters used for creating the risk prediction model

Dataset	Resolution	Unit	Source
Digital elevation model	30m x 30m	-	DPT Rama IX
Slope	30m x 30m	Percent (%)	DEM
Elevation	30m x 30m	Meter (m)	DEM
Annual rainfall	-	Milli meter (mm)	https://www.tmd.go.th/service/tmdData
Soil type	30m x 30m	-	DPT Rama IX
Gross building area	30m x 30m	Square meter (m^2)	Generated from GIS
Land ownership	30m x 30m	m^2	Generated from GIS
Building coverage ratio	-	%	Generated from GIS
Floor area ratio	-	-	Generated from GIS
Road Alignment	-	-	DPT Rama IX
Land use	30m x 30m	-	DPT Rama IX
Flood risk from Flood Simulation Model	30m x 30m	-	DPT Rama IX

BCR is crucial in urban planning, influencing flood risk management. High BCR means more impervious surfaces, increasing runoff, and higher flood risks. Thus, urban planners utilize the benefit-cost ratio (BCR) as a tool to achieve a harmonious equilibrium between development and flood resilience. This is frequently accomplished by incorporating green spaces into urban designs, which serve the dual purpose of absorbing rainwater and minimizing the impact of flood (Chen *et al.,* 2020). 5) Road density measures the length of roads in a given area. High road density can increase surface runoff and flooding risks because impervious surfaces prevent water absorption (Lu and Wu, 2023). 6) Slope refers to the incline or steepness of a land surface, playing a crucial role in flood dynamics. Steeper slopes can cause water to run off quickly, increasing flood risks downstream. Gentler inclines facilitate more water infiltration into the soil, thereby diminishing the amount of runoff (Robertson *et al.,* 2019). 7) Elevation above sea level influences flood risk, with lower elevations being more prone to flooding due to water accumulation from higher areas. This point makes elevation data vital for designing drainage systems and flood defenses (Robertson *et al.,* 2019). 8) Soil drainage ability refers to the capacity of soil to absorb and transmit water. This characteristic significantly impacts flood risk, as soils with poor drainage can pool water on the surface, thereby heightening the probability of flooding when intense rainfall occurs. The drainage capacity is affected by various factors including soil texture, structure, and the amount of organic matter present (Sandink and Binns, 2021).

Methods

The objective of this study is to develop a tool that can be utilized for the analysis of flood risk areas within the Pattaya City region of Chonburi Province. The secondary data were used, such as rainfall, land use, FAR, BCR, road density, slope, elevation, soil drainage ability, and flood risk levels from the Rama IX Department of Public Works and Town and Country Planning's flood simulation model. It will be utilized in developing a machine learning model which can be employed to assess the utmost critical flood hazard for residents, urban planning experts, or interested parties. This assessment will aid in formulating effective city planning strategies aimed at mitigating and minimizing the occurrence of floods within the urban area. The methodology section is divided into three sections: data collection, analysis, and validation, which will be discussed below.

Data collection

The data acquisition process necessitates close examination because of its potential use in machine learning. The data used for ML model training consists of precipitation, DEM, slope, elevation, soil type, gross building area, land ownership, FAR, BCR road alignment, land use, and flood simulation model. The Meteorological Department provided monthly precipitation data for Chonburi Province spanning from 2015 to 2022. The research team gathered geographic information regarding the Pattaya city area from the Department of Public Works and Town and Country Planning (DPE). This geographic data contains DEM, road alignment, land use, and flood risk from the flood simulation model. The last set of

data is geographic information about the Pattaya city area from the research team, obtained using various quantum geographic information system (QGIS) tools to generate based on the polygons' characteristics in the DPE shapefile. This dataset contains gross building area, land ownership, BCR, and FAR. All data used for creating a risk prediction model in ML can be shown in Table 2.

Data analysis

A suitable GIS dataset is required for the preparation of a research study on urban flood risk assessment using Machine Learning integrated with GIS. Prior to the development of a machine-learning model, thorough analysis of this dataset is necessary. The subsequent section will provide an outline of the methods to be employed.

Dataset preparation on QGIS

Preparing GIS data using hazard and vulnerability information and flood risk assessment results is integrated with Grid H3 to create a data set that can be utilized to build machine learning models. The required data types and preparation methods are illustrated in Fig. 1.

To prepare flood risk area analysis data in GIS, the H3 Grid size must be suitably selected to draw each data layer together. Fig. 2 illustrates the process of creating a dataset containing hazard variables, while Fig. 3 demonstrates the steps involved in preparing a

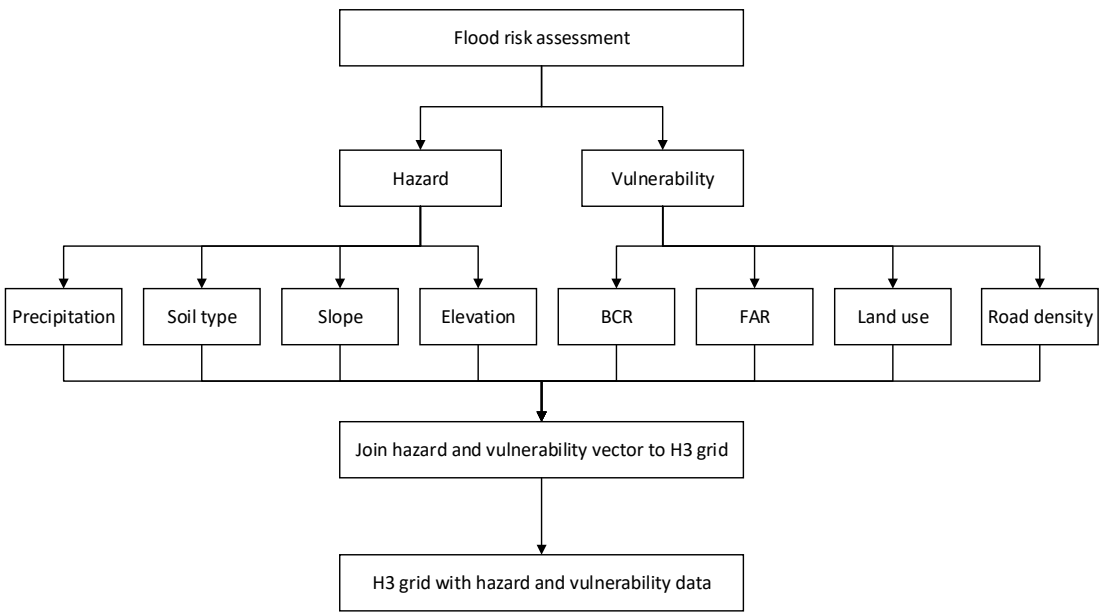


Fig. 1: GIS data preparation, Hazard, and Vulnerability factors

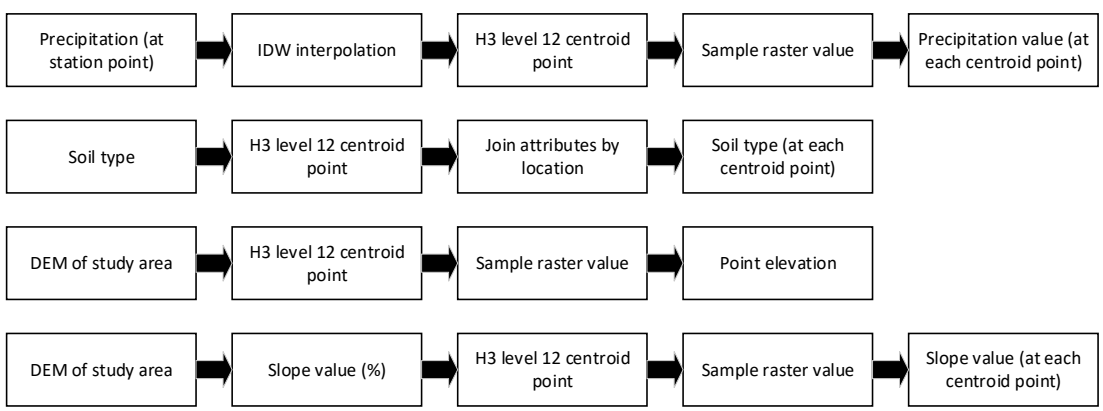


Fig. 2: Hazard-type data preparation using QGIS.

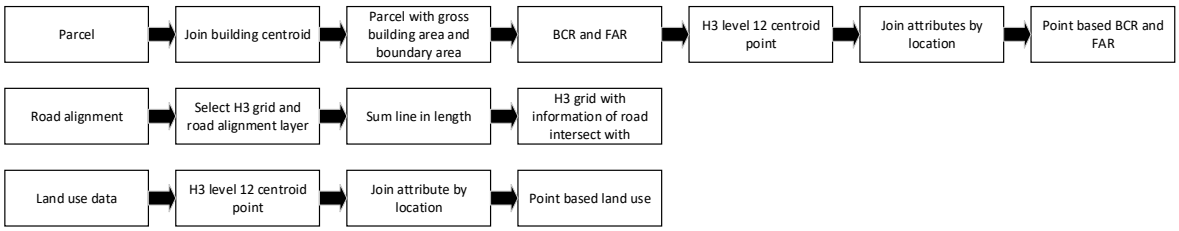


Fig. 3: Vulnerability-type data preparation



Fig. 4: An example of an H3 grid in the QGIS platform containing the necessary data for flood risk assessment.

dataset focused on vulnerability factors.

After the preparation process, the next step is integrating all data into the H3 grid. Fig. 4 illustrates the input of essential data, such as precipitation, land use, FAR, BCR, road density, slope, elevation, and soil drainage, into each grid centroid in QGIS. This comprehensive display of information ensures that all necessary data is accurately represented and easily accessible for analysis and interpretation.

Data analysis using a machine learning model

When the dataset containing factors from both sides is combined with Grid H3, it produces a tabular format that can be utilized for the creation of a machine-learning model, as depicted in Fig. 5.

Developing a machine learning model encompasses six primary stages, which can be delineated as follows:

1. Using QGIS, combine spatial data on hazard and vulnerability factors and export the results to a tabular data format.

2. Prepare data for analysis by cleaning, transforming, and removing unusable data.

3. Split Data into train data divided into 70% (3A) and test data into 30% (3B).

4. Choose an algorithm to develop the model to make predictions by using train Data (5A)

5. Use the roc_auc_score value to evaluate the model and determine whether the accuracy is appropriate (5B). If the accuracy is unsuitable, choose a different algorithm (5C).

6. Machine learning model prediction will obtain a Flood risk assessment score.

RESULTS AND DISCUSSION

This study aims to build a tool for flood risk assessment in the Pattaya City region, Chonburi

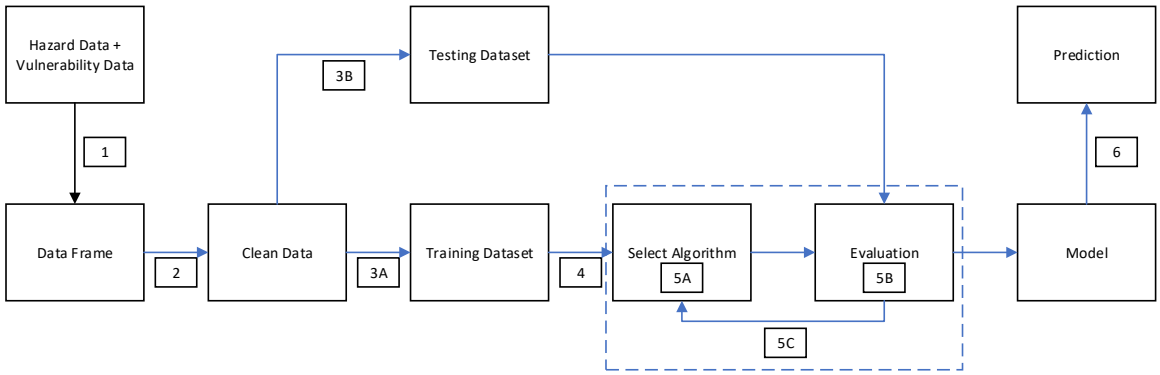


Fig. 5: Analysis process using machine learning to assess flood risk.

Province, utilizing machine learning and 11 multiclass classification algorithms, as follows: LR, NB, KNN, SVM, DTs, AdaBoost decision tree, bagging decision tree, extra tree, RF, and neural network (deep learning) to identify the best model for assessing flood risk in Pattaya. The flood vulnerability predictions were determined by eight parameters in the construction of the model. These parameters include precipitation, land use, FAR, BCR, road density, slope, elevation, and soil drainage capacity. The efficiency of the model was then investigated comparatively, leading to the following results.

Hazard factors

The hazards of Pattaya include the accumulated rainfall unit, which is mm, soil drainage capacity, slope (%), and elevation unit, which is (m). The details of each variable can be described below.

1) Accumulated rainfall (mm): In order to maintain consistency with the findings of the Department of Public Works Flood Simulation Model in 2016, this research endeavor has developed a learning model by utilizing cumulative rainfall data (measured in millimeters) from the corresponding year. The rain measuring station's proximity will result in higher levels of accumulated rainfall in central and southern Pattaya compared to the northern region.

2) The Pattaya City soil series is classified into three types of soil drainage ability: poorly drained, moderately drained, and well-drained. It depicts the Pattaya community area and highways (Pattaya Beach Road, Third Road, South Pattaya Road, Sukhumvit Road, South Pattaya, Sukhumvit Road, and Road along the Railway) in soil with poor drainage.

3) The Pattaya city area exhibits a slope (%) derived from the Digital Elevation Model (DEM) with a resolution of 30m x 30m. The topography of Pattaya predominantly comprises undulating plains, coastal plains, and gently sloped hills. Notably, the Pratumnak Hill area stands out as the location with the most pronounced slope, appearing as a white patch on the western side.

4) Elevation (m) of the Pattaya city area based on the DEM with a resolution of 30m x 30m. It depicts Pattaya's undulating lowlands, coastal plains, and hills, with Pratumnak Hill being the highest location.

Vulnerability factors

Vulnerability factors include FAR, BCR, land use, and road density. Below are details of each variable.

1) In 2016, Pattaya's FAR, or ratio of total building area to land, ranged from 0.29 to 7.26.

2) BCR represents Pattaya City's percentage of building coverage per land area in 2016. The area's BCR ranges from 0.67% to 100%.

3) Land use in 2016 is classified as agricultural area (A), forest area (F), miscellaneous area (M), community and building area (U), and water area (W). Land use in most areas of Pattaya city is U.

4) The road density of Pattaya city is determined by the number of roads intersecting per grid unit. Areas with high road density are typically found in close proximity to major thoroughfares like Pattaya Beach Road, Third Road, South Pattaya Road, Sukhumvit Road, South Pattaya, Sukhumvit Road, and the railway road.

Comparative results of model performance

It is crucial to establish the parameters for each mode during the development of machine learning

Table 3: Training accuracy, testing accuracy, the difference between training and testing accuracy, and ROC score

Model	Train Accuracy	Test Accuracy	Difference between Training Testing Accuracy	ROC Score
Decision tree	0.912	0.809	0.103	0.878
Bagging decision tree	0.913	0.822	0.091	0.941
Adaboost decision tree	0.905	0.793	0.112	0.870
Extra tree	0.912	0.802	0.110	0.908
Random forest	0.916	0.810	0.106	0.926
K-Nearest neighbors	0.835	0.772	0.063	0.866
Gaussian naive baye	0.611	0.608	0.003	0.597
Logistic regression	0.634	0.631	0.003	0.593
Support vector machine	0.771	0.735	0.036	0.565
Neural network	0.635	0.633	0.002	0.763
Gradient boosting	0.775	0.756	0.018	0.870

Table 4: Precision, recall, and F1-score of each model

Model	Flood Risk level	Train			Test		
		Precision	Recall	F1-score	Precision	Recall	F1-score
Decision tree	Low risk	0.802	0.875	0.837	0.64	0.644	0.642
	Moderate risk	0.778	0.82	0.799	0.581	0.589	0.585
	High risk	0.791	0.916	0.849	0.657	0.734	0.693
	No risk	0.991	0.936	0.963	0.919	0.902	0.91
Bagging Decision tree	Low risk	0.815	0.859	0.836	0.654	0.668	0.661
	Moderate risk	0.767	0.833	0.799	0.585	0.62	0.602
	High risk	0.792	0.911	0.847	0.677	0.76	0.716
	No risk	0.988	0.941	0.964	0.936	0.908	0.922
Adaboost decision tree	Low risk	0.792	0.854	0.822	0.592	0.667	0.627
	Moderate risk	0.745	0.845	0.792	0.538	0.609	0.571
	High risk	0.774	0.925	0.843	0.634	0.742	0.684
	No risk	0.994	0.926	0.959	0.94	0.865	0.901
Extra tree	Low risk	0.802	0.875	0.837	0.623	0.638	0.63
	Moderate risk	0.778	0.82	0.799	0.577	0.576	0.576
	High risk	0.791	0.916	0.849	0.66	0.72	0.688
	No risk	0.991	0.936	0.963	0.913	0.896	0.905
Random forest	Low risk	0.815	0.875	0.844	0.638	0.64	0.639
	Moderate risk	0.789	0.813	0.8	0.594	0.58	0.587
	High risk	0.812	0.894	0.851	0.688	0.705	0.696
	No risk	0.986	0.947	0.966	0.91	0.91	0.91
K-nearest neighbors	Low risk	0.755	0.622	0.682	0.624	0.504	0.558
	Moderate risk	0.686	0.599	0.64	0.565	0.471	0.514
	High risk	0.717	0.705	0.711	0.622	0.616	0.619
	No risk	0.884	0.951	0.916	0.846	0.919	0.881
Gaussian naive bayes	Low risk	0.292	0.118	0.168	0.284	0.111	0.159
	Moderate risk	0.252	0.069	0.109	0.219	0.067	0.103
	High risk	0.199	0.097	0.13	0.207	0.105	0.139
	No risk	0.669	0.905	0.769	0.668	0.905	0.769
Logistic regression	Low risk	0.32	0.001	0.002	0.107	0	0.001
	Moderate risk	0.313	0.012	0.023	0.2	0.007	0.013
	High risk	0.833	0.002	0.004	0.889	0.003	0.005
	No risk	0.635	0.999	0.777	0.633	0.998	0.775
Support vector machine	Low risk	0.738	0.408	0.525	0.644	0.334	0.44
	Moderate risk	0.688	0.426	0.526	0.581	0.34	0.429
	High risk	0.724	0.486	0.581	0.647	0.424	0.512
	No risk	0.785	0.969	0.867	0.762	0.957	0.849
Neural network	Low risk	0.484	0.121	0.194	0.46	0.118	0.187
	Moderate risk	0.514	0.015	0.03	0.522	0.013	0.026
	High risk	0.431	0.085	0.142	0.412	0.084	0.14
	No risk	0.668	0.984	0.796	0.666	0.981	0.793
Gradient boosting	Low risk	0.711	0.405	0.516	0.663	0.369	0.474
	Moderate risk	0.719	0.415	0.526	0.663	0.363	0.469
	High risk	0.83	0.476	0.605	0.792	0.453	0.577
	No risk	0.784	0.978	0.871	0.773	0.973	0.862

algorithms aimed at classifying flood risk levels. Each model's results include training accuracy, testing accuracy, the difference between training and testing accuracy, and ROC score, which can be seen in Table 3. The Precision, Recall, and F1 scores of each model are depicted in Table 4.

The study categorized flood risk levels into four groups - high risk, moderate risk, low risk, and no risk - in order to assess the model's precision. The ratio of correct and incorrect predictions is then used to draw a ROC graph to find the AUC or ROC Score of each model, which is used as a criterion for selecting the best model.

Selecting the most appropriate model for assessing urban flood risk

From the model performance test, some models with ROC Score according to the Excellent criteria with AUC values in the range of 0.9 < AUC < 1.0 (Jeong *et al.,* 2016), shown in Table 5.

Based on the evaluation of training accuracy, testing accuracy, and ROC score, it was determined that the bagging decision tree model exhibited the highest AUC value. Furthermore, the discrepancies between the values were within a 10% range. The model's did not experience overfitting as a result of its performance, with the training accuracy surpassing the testing accuracy, thereby avoiding any instances of underfitting (Hasan *et al.,* 2023). Therefore, it is considered that the bagging decision tree model is most suitable for use in assessing flood risk in Pattaya City. It was found that the top three variables that most influence the city's flood risk are rainfall, BCR, and FAR, respectively, as shown in Table 6.

Table 7 demonstrates that the exclusion of the three most significant factors, namely rainfall, BCR, and FAR, will result in a decrease in the training accuracy, testing accuracy, and ROC score of the model. This issue demonstrates that the model's efficiency is sensitive to these three variables.

Studied by the Department of Public Works and Town and Country Planning

The inclusion of eight indicators pertaining to flood

Table 5: Models with an excellent ROC score

Model	Training accuracy	Testing accuracy	ROC score
Bagging decision tree	0.913	0.822	0.941
Random forest	0.916	0.810	0.926
Extra trees	0.912	0.802	0.908

Table 6: The feature importance for each factor of the bagging decision tree model

Feature	Importance
Rainfall	0.521
BCR	0.141
FAR	0.111
Slope	0.064
Road density	0.053
Elevation	0.049
Soil drainage capacity	0.037
Land use	0.024

Table 7: Model performance when removing the top 3 factors that most affect the model

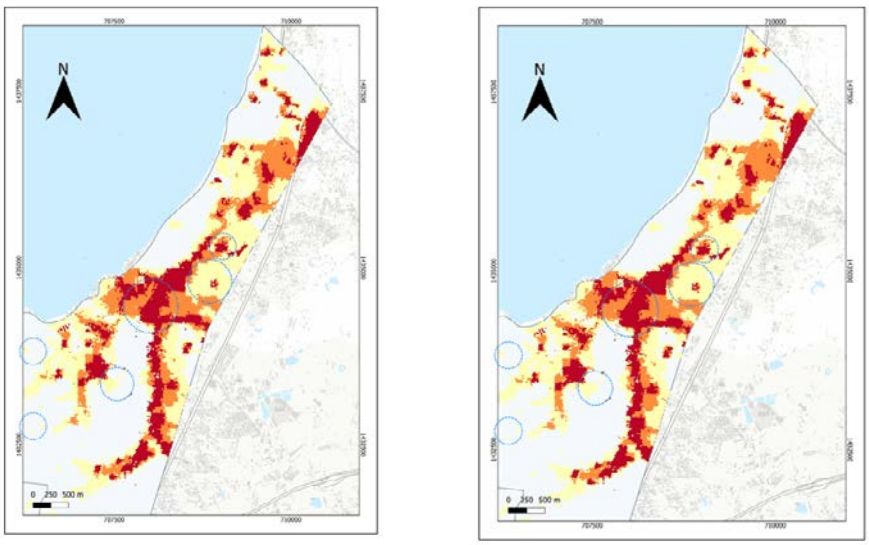
Model	Training accuracy	Testing accuracy	ROC score
Bagging decision tree	0.680	0.669	0.717

Table 8: Accuracy, precision, recall, and F1-score values compared to the flood simulation model

Class	Precision	Recall	F1-score
No risk	0.972	0.932	0.952
Low risk	0.768	0.801	0.784
Moderate risk	0.714	0.769	0.740
High risk	0.760	0.865	0.809
Accuracy		88.58%	

susceptibility forecasts resulted in an overall accuracy of 88.58%, as compared to the flood simulation model. Precision, recall, and F1-score values can be used to rank the accuracy of different classes. The bagging decision tree model was the most effective at classifying flood risk levels in non-risk areas, with precision values of 0.972 because the dataset included more non-risk areas than danger areas, as shown in Table 8.

Upon comparing the risk analysis outcomes of the model, it is evident that they bear resemblance to the flood simulation model studied by the Department of Public Works and Town and Country Planning. Fig. 6 depicts an example comparison in the North Pattaya area (enlarged picture), while (Fig. 7) depicts flood risk susceptibility for the entire Pattaya area using the



(a) (b)

Fig. 6: Comparison from (a) Flood simulation model and (b) Bagging decision tree

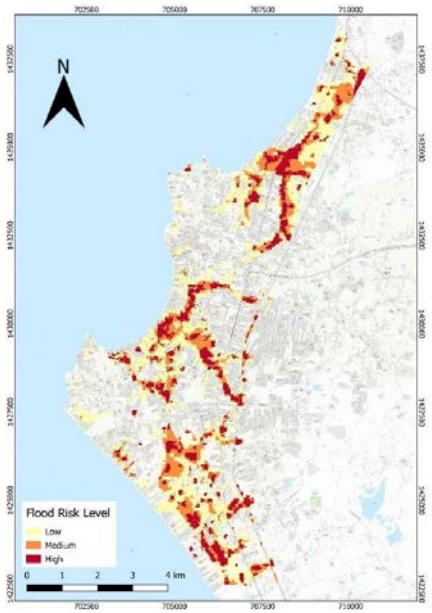


Fig. 7: Flood risk susceptibility for Pattaya area using bagging decision tree model

Table 9: Risk-affected areas information in Pattaya City

Risk level	Risk-affected area (points)	Area (m^2)	Proportion
Low risk	27,784	8.53	21.34
Moderate risk	12,397	3.81	9.52
High risk	11,009	3.38	8.46



(a) (b)

Fig. 8: High-risk area predicted from bagging decision tree model (a), and actual flood situation occur in 2022 (b)

bagging decision tree model.

Fig. 7 exclusively depicts the regions that are prone to flooding, with classifications denoting varying levels of vulnerability, namely low, moderate, and high risk. Table 9 provides comprehensive details about the areas that are impacted by these risks. It furnishes specific information such as the number of risk points, the corresponding sizes of the affected areas, and their proportion in relation to the total area, expressed as a percentage.

It was discovered that the roads along the road are sensitive to high-risk flooding, which matches news reports concerning flooding in Pattaya City, Chonburi Province, in 2022, such as flooding detected along Sukhumvit Road and the road along the railway, as shown in Fig. 8.

CONCLUSION

Flooding, which arises from the combined effects of climate change and human activity, gives rise to substantial challenges in terms of economics, infrastructure, and everyday existence. The resulting destruction necessitates meticulous urban planning and the implementation of risk management strategies. FHM plays an essential role in risk mitigation, but its effectiveness depends on frequent updates, high-quality data, and innovative technology such as GIS. Prior research has used GIS technology and a machine learning algorithm to estimate flood risk and make decisions more efficiently. The annual flooding in Pattaya, Thailand, creates a problem for the economically crucial locality, leading to an investigation aimed at discovering an effective machine learning framework for assessing flood hazard. This research aims to combine GIS tools and machine learning models to examine urban indicators, resulting in prototypes for resilient city management planning in Pattaya. This study extracted from GIS and imported eight variables that affect flood susceptibility predictions, including rainfall, land use, FAR, BCR, road density, slope, elevation, and soil drainage capacity, into 11 multiclass classification algorithms to assess the flood risk of the Pattaya city area. According to the study

findings, the implementation of the bagging decision tree algorithm model proved to be the optimal approach for flood risk assessment. It achieved an impressive accuracy rate of 88.58%, surpassing the flood simulation model results analyzed by the Department of Public Works and Town and Country Planning. Furthermore, rainfall, BCR, and FAR were identified as Pattaya's top three variables influencing flood risk. Using machine learning models to study and evaluate flood-prone city areas is a useful urban planning tool. Enhancing the accuracy of machine learning models can be achieved through the provision of robust and structured data sets, databases, and methodologies tailored for urban environments. The absence of dedicated organizations responsible for managing urban data in Thailand has led to cities having inadequate data sets, thereby hindering the development of highly precise machine learning models. For example, no historical record of flood incidents in the city makes it challenging to locate the precise sites where flooding occurred. Consequently, the conclusions drawn from the study are constrained by the conditional probability, which serves as an indicator of whether the risk is categorized as high or low. Moreover, it should be noted that the rain measuring stations in Pattaya only provide coverage for certain areas. If a city data management agency is established, designing a resilient city system will be possible. This model can be easily adapted for other provinces in Thailand by considering each area's specific local data. By incorporating region-specific information, the model is able to generate accurate flood risk predictions that are tailored to different locations. The adaptability of this model ensures that it can serve as a valuable tool for urban planners and policymakers across the country. The model's accuracy will be significantly enhanced with the inclusion of comprehensive historical data. The more historical flood data available, the better the model can learn from past events and predict future risks. This approach enables efficient flood risk management in Pattaya and provides a flexible solution that can be implemented in different areas, thereby enhancing the overall urban resilience across the country in the face of flooding.

AUTHOR CONTRIBUTIONS

M. Srivanit, the corresponding author, contributed to supervising the study and manuscript preparation. S. Pattanasri performed the literature review, conducted research experimentation, interpreted the results, and wrote the manuscript. N. Phichetkunbodee contributed to the manuscript's conceptualization, methodology, and revising. S. Manokeaw contributed to revising the manuscript and performed validation on the analysis. S. Sitthikankun contributed to the manuscript's conceptualization, methodology, and revising. D. Rinchumphu supervised the work, concept formulation, data evaluation, and manuscript edition.

ACKNOWLEDGEMENT

The authors thank the City Research and Development Center support staff and Frontier Research in Urban Intelligence, Faculty of Engineering, Chiang Mai University, Thailand. This research project is supported by Fundamental Fund 2024 [Number R66EX01091], Chiang Mai University

CONFLICT OF INTEREST

The authors declare that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancy have been completely observed by the authors.

OPEN ACCESS

©2024 The author(s). This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit: http://creativecommons.org/licenses/by/4.0/

PUBLISHER'S NOTE

GJESM Publisher remains neutral with regard

to jurisdictional claims in published maps and institutional afflictions.

ABBREVIATIONS

%	Percent
A	Agricultural area
AI	Artificial intelligence
AHP	Analytic hierarchy process
AUC-ROC	Area under the receiver operating characteristic curve
BCR	Building coverage ratio
BD	Building density
DEM	Digital elevation model
DLNN	Deep learning neural network
DTs	Decision trees
Eq.	Equation
F	Forest area
F1-score	Performance metric commonly used in binary classification problems
FN	False negative
FP	False positive
FAR	Floor area ratio
FHM	Flood hazard map
Fig.	Figure
GA	Genetic algorithm
GIS	Geographic information systems
KNN	K-nearest neighbors
LR	Logistic regression
M	Miscellaneous area
ML	Machine learning
mm	Millimeter
m	Meter
m^2	Square meter
NB	Naive bayes classification
RF	Random forest
RS	Remote sensing data
ROC curve	Receiver operating characteristic
QGIS	quantum geographic information system
SVM	Support vector machine
TP	True positive
TN	True negative
QGIS	Quantum geographic information system
U	Community and building area
W	Water area

REFERENCES

Atanga, R.A., (2020). The role of local community leaders in flood disaster risk management strategy making in Accra. Int. J. Disaster Risk Reduct. 43: 101358 **(12 pages)**.

Balayla, J., (2022). Bayesian updating and sequential testing: overcoming inferential limitations of screening tests. BMC Med. Inf. Decis. Making. 22(1) **(18 pages)**.

Burayu, D.G.; Karuppannan, S.; Shuniye, G., (2023). Identifying flood vulnerable and risk areas using the integration of analytical hierarchy process (AHP), GIS, and remote sensing: A case study of southern Oromia region. Urban Clim., 51: 101640 **(17 pages)**.

Cabrera, J.S.; Lee, H.S., (2020). Flood risk assessment for Davao Oriental in the Philippines using geographic information system-based multi-criteria analysis and the maximum entropy model. J. Flood Risk Manage., 13(2): e12607 **(17 pages)**.

Chang, M.J.; Chang, H.K.; Chen, Y.C.; Lin, G.F.; Chen, P.A.; Lai, J.S.; Tan, Y.C., (2018). A Support Vector Machine Forecasting Model for Typhoon Flood Inundation Mapping and Early Flood Warning Systems. Water,.10(12): 1734 **(19 pages)**.

Chen, Y.; Liu, T.; Chen, R.; Zhao, M., (2020). Influence of the Built Environment on Community Flood Resilience: Evidence from Nanjing City, China. Sustainability, 12(6): 2401 **(16 pages).**

da Silva, L.B.L.; Humberto, J.S.; Alencar, M.H.; Ferreira, R.J.P.; de Almeida, A.T., (2020). GIS-based multidimensional decision model for enhancing flood risk prioritization in urban areas. Int. J. Disaster Risk Reduct. 48: 101582 **(15 pages)**.

De Diego, I.M.; Redondo, A.R.; Fernández, R.R.; Navarro, J.; Moguerza, J.M., (2022). General Performance Score for classification problems. Appl. Intell., 52(10): 12049-12063 **(15 pages)**.

Diehl, R.M.; Gourevitch, J.D.; Drago, S.; Wemple, B.C., (2021). Improving flood hazard datasets using a low-complexity, probabilistic floodplain mapping approach. Plos One. 16(3): e0248683 **(20 pages)**.

Glago, F.J., (2021). Flood disaster hazards; causes, impacts and management: a state-of-the-art review. Nat. Hazard., 29-37 **(19 pages)**.

Khosravi, K.; Pham, B.T.; Chapi, K.; Shirzadi, A.; Shahabi, H.; Revhaug, I.; Prakash, I.; Tien Bui, D., (2018). A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. Sci. Total Environ., 627: 744-755 **(12 pages)**.

Kokila, R.; Balasubramanian, B.; Meenambigai, K.; Alagirisamy, K.; Kumar, A.; Kamyab, H.; Yadav, Krishna K.; Nareshkumar, A., (2021). A GIS-based tool for the analysis of the distribution and abundance of Chilo sacchariphagus indicus under the influence of biotic and abiotic factors. Environ. Technol. Innovation, 21: 101357 **(14 pages)**.

Li, L., Collins, A.M.; Cheshmehzangi, A.; Chan, F.K.S., (2020).

Identifying enablers and barriers to the implementation of the Green Infrastructure for urban flood management: A comparative analysis of the UK and China. Urban For. Urban Greening, 54: 126770 **(11 pages)**.

Lin, J.; He, X.; Lu, S.; Liu, D.; He, P., (2021). Investigating the influence of three-dimensional building configuration on urban pluvial flooding using random forest algorithm. Environ. Res., 196: 110438 **(33 pages)**.

Liu, R.; Chen, Y.; Wu, J.; Gao, L.; Barrett, D.; Xu, T.; Li, X.; Li, L.; Huang, C.; Yu, J., (2017). Integrating Entropy-Based Naïve Bayes and GIS for Spatial Evaluation of Flood Hazard. Risk Anal., 37(4): 756-773 **(18 pages)**.

Lu, S.; Huang, J.; and Wu, J., (2023). Multi-Dimensional Urban Flooding Impact Assessment Leveraging Social Media Data: A Case Study of the 2020 Guangzhou Rainstorm. Water, 15(24): 4296 **(20 pages)**.

McClean, F.; Dawson, R.; Kilsby, C., (2020). Implications of using global digital elevation models for flood risk analysis in cities. Water Resour. Res., 56(10): e2020WR028241 **(15 pages)**.

Jeong, M.S.; Kim, J.S.; Jang, H.W.; Lee, J.H., (2016). ROC evaluation for MLP ANN drought forecasting model. J. Korea Water Resour. Assoc., 49(10) **(14 pages)**.

Marchioni, M.; Fedele, R.; Raimondi, A.; Sansalone, J.; Becciu, G., (2022). Permeable Asphalt Hydraulic Conductivity and Particulate Matter Separation With XRT. Water Resour. Manage., 36(6): 1879-1895 **(17 pages)**.

Mohtat, N.; Khirfan, L., (2022). Distributive Justice and Urban Form Adaptation to Flooding Risks: Spatial Analysis to Identify Toronto's Priority Neighborhoods [Original Research]. Front. Sustainable Cities. 4 **(20 pages)**.

O'Donnell, E.C.; Thorne, C.R.; Yeakley, J.A.; Chan, F.K.S., (2020). Sustainable Flood Risk and Stormwater Management in Blue-Green Cities; an Interdisciplinary Case Study in Portland, Oregon. JAWRA J. Am. Water Resour. Assoc., 56(5): 757-775 **(19 pages)**.

Ogato, G.S.; Bantider, A.; Abebe, K.; Geneletti, D., (2020). Geographic information system (GIS)-Based multicriteria analysis of flooding hazard and risk in Ambo Town and its watershed, West shoa zone, oromia regional State, Ethiop. J. Hydrol. Reg. Stud., 27: 100659 **(18 pages)**.

Park, K.; Won, J.h., (2019). Analysis on distribution characteristics of building use with risk zone classification based on urban flood risk assessment. Int. J. Disaster Risk Reduct., 38: 101192 **(10 pages)**.

Park, K.; Lee, M.H. (2019). The Development and Application of the Urban Flood Risk Assessment Model for Reflecting upon Urban Planning Elements. Water. 11(5): 920 **(17 pages)**.

Petty, T.R.; Noman, N.; Ding, D.; Gongwer, J.B., (2016). Flood Forecasting GIS Water-Flow Visualization Enhancement (WaVE): A Case Study. J. Geogr. Inf. Syst., 37: 72935 **(38 pages)**.

Pham, B.T.; Luu, C.; Dao, D.V.; Phong, T.V.; Nguyen, H.D.; Le, H.V.; von Meding, J.; Prakash, I., (2021). Flood risk assessment using deep learning integrated with multi-criteria decision analysis. Knowledge-Based Syst., 219: 106899 **(15 pages)**.

Pham, B.T.; Luu, C.; Phong, T.V.; Nguyen, H.D.; Le, H.V.; Tran, T.Q.; Ta, H.T.; Prakash, I., (2021). Flood risk assessment using hybrid artificial intelligence models integrated with multi-criteria decision analysis in Quang Nam Province, Vietnam. J. Hydrol., 592: 125815 **(15 pages)**.

Praskievicz, S.; Carter, S.; Dhondia, J.; Follum, M., (2020). Flood-inundation modeling in an operational context: sensitivity to topographic resolution and Manning's. J. Hydroinf., 22(5): 1338-1350 **(13 pages)**.

Ramkumar, G.; Tamilselvi, M.; D Sundarsingh Jebaseelan, S.; Mohanavel, V.; Kamyab, H.; Anitha, G.; Thandaiah P., R.; Rajasimman, M., (2024). Enhanced machine learning for nanomaterial identification of photo thermal hydrogen production. Int. J. Hydrogen Energy. 52: 696-708. **(12 pages)**.

Rezvani, S.M.; Falcão, M.J.; Komljenovic, D.; de Almeida, N.M., (2023). A systematic literature review on urban resilience enabled with asset and disaster risk management approaches and GIS-based decision support tools. Appl. Sci., 13(4): 2223 **(41 pages)**.

Robertson, W.M.; Allen, J.T.; Wolaver, B.D.; Sharp, J.M., (2019). Aridland spring response to mesoscale precipitation: Implications for groundwater-dependent ecosystem sustainability. J. Hydrol., 570: 850-862 **(13 pages)**.

Saleh, A.; Yuzir, A.; Sabtu, N.; Abujayyab, S.K.M.; Bunmi, M.R.; Pham, Q.B., (2022). Flash flood susceptibility mapping in urban area using genetic algorithm and ensemble method. Geocarto Int., 37(25): 10199-10228 **(30 pages)**.

Sandink, D.; Binns, A.D., (2021). Reducing Urban Flood Risk Through Building- and Lot-Scale Flood Mitigation Approaches: Challenges and Opportunities. Front. Water. 3 **(20 pages)**.

Sangpikul, A., (2020). Learning about the real-world in MICE education: The case of exhibition learning from Thailand. J. Conv. Event Tourism. 21(3): 225-253 **(29 pages)**.

Sar, N.; Chatterjee, S.; Das Adhikari, M., (2015). Integrated remote sensing and GIS based spatial modelling through analytical hierarchy process (AHP) for water logging hazard, vulnerability and risk assessment in Keleghai river basin, India. Model. Earth Syst. Environ., 1(4): 31 **(21 pages)**.

Shahabi, H.; Shirzadi, A.; Ghaderi, K.; Omidvar, E.; Al-Ansari, N.; Clague, J. J.; Geertsema, M.; Khosravi, K.; Amini, A.; Bahrami, S.; Rahmati, O.; Habibi, K.; Mohammadi, A.; Nguyen, H.; Melesse, A.M.; Ahmad, B.B.; Ahmad, A., (2020). Flood Detection and Susceptibility Mapping Using Sentinel-1 Remote Sensing Data and a Machine Learning Approach: Hybrid Intelligence of Bagging Ensemble Based on K-Nearest Neighbor Classifier. Remote Sens., 12(2): 266 **(30 pages)**.

Shen, C.; Appling, A.P.; Gentine, P.; Bandai, T.; Gupta, H.; Tartakovsky, A.; Baity-Jesi, M.; Fenicia, F.; Kifer, D.; Li, L.; Liu, X.; Ren, W.; Zheng, Y.; Harman, C. J.; Clark, M.; Farthing, M.; Feng, D.; Kumar, P.; Aboelyazeed, D.; Lawson, K., (2023). Differentiable modelling to unify machine learning and physical models for geosciences. Nat. Rev. Earth Environ., 4(8): 552-567 **(16 pages)**.

Shrestha, Y.R.; Krishna, V.; von Krogh, G., (2021). Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges. J. Bus. Res., 123: 588-603 **(16 pages)**.

Teng, J.; Jakeman, A. J.; Vaze, J.; Croke, B.F.W.; Dutta, D.; Kim, S., (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. Environ. Model. Software, 90: 201-216 **(16 pages)**.

Tien Bui, D.; Hoang, N.D.; Martínez-Álvarez, F.; Ngo, P.T.T.; Hoa, P.V.; Pham, T.D.; Samui, P.; Costache, R., (2020). A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm

area. Sci. Total Environ., 701: 134413 **(38 pages)**.
Vakili, M., Ghamsari, M., Rezaei, M., (2020). Performance analysis and comparison of machine and deep learning algorithms for IoT data classification. arXiv., 2001: 09636 **(19 pages)**.
Vujović, Ž., (2021). Classification model evaluation metrics. Int. J. Adv. Comput. Sci. Appl., 12(6): 599-606 **(8 pages)**.
Wang, Z.; Lai, C.; Chen, X.; Yang, B.; Zhao, S.; Bai, X., (2015). Flood hazard risk assessment model based on random forest. J. Hydrol., 527: 1130-1141 **(12 pages)**.
Wang, Z.; Wu, J.; Cheng, L.; Liu, K.; Wei, Y.M., (2018). Regional flood risk assessment via coupled fuzzy c-means clustering methods: an empirical analysis from China's Huaihe River Basin. Nat. Hazard., 93(2): 803-822 **(20 pages)**.
Woldesenbet, W.G., (2020). Analyzing multi-stakeholder collaborative governance practices in urban water projects in Addis Ababa City: procedures, priorities, and structures. Appl. Water Sci., 10(1): 44 **(19 pages)**.
Wu, Y.; Tang, M.; Zhou, Z.; Chu, J.; Zeng, Y.; Zhan, M.; Xu, W., (2024). Rainfall Pattern Construction Method Based on DTW-HCA and Urban Flood Simulation: A Case Study of Nanchang City, China. Water. 16(1): 65. **(20 pages)**.
Zaoui, M.; Himouri, S.; Kadri, T.; Benaouina, C., (2020). Selection of the best alternative for a road project to replace a section in a flood-prone area using GIS and AMC tools. J. Mater. Eng. Struct., 7(2): 18 **(18 pages)**.
Zaidi, A.; and Al Luhayb, A.S.M., (2023). Two Statistical Approaches to Justify the Use of the Logistic Function in Binary Logistic Regression. Math. Probl. Eng., 2023: 5525675 **(11 pages).**

AUTHOR (S) BIOSKETCHES

Srivanit, M., Ph.D., Associate Professor, Urban Planning Programs, Thammasat University, Thailand.
- Email: *manat-s@tu.ac.th*
- ORCID: 0000-0002-4660-4518
- Web of Science ResearcherID: AFR-4557-2022
- Scopus Author ID: 55553537700
- Homepage: https://expertise-grad.tds.tu.ac.th/professors/4

Pattanasri, S., M.Eng. Student, Urban Planning Programs, Thammasat University, Thailand.
- Email: *atomsiwawut@gmail.com*
- ORCID: 0009-0005-6870-3559
- Web of Science ResearcherID: NA
- Scopus Author ID: NA
- Homepage: https://www.tds.tu.ac.th/programs-up/

Phichetkunbodee, N., Ph.D. Candidate, Department of Civil Engineering, National Taiwan University, Taiwan.
- Email: *d11521023@ntu.edu.tw*
- ORCID: 0000-0003-4569-669X
- Web of Science ResearcherID: NA
- Scopus Author ID: 57222813126
- Homepage: https://www.ce.ntu.edu.tw/en/home

Manokeaw, S., M.D., Proactive researcher, Office of Research Administration, Chiang Mai University, Thailand.
- Email: *sattaya.manok@cmu.ac.th*
- ORCID: 0000-0001-8355-2211
- Web of Science ResearcherID: NA
- Scopus Author ID: NA
- Homepage: https://ora.oou.cmu.ac.th/home/

Sitthikankun, S., Ph.D. Candidate, Lecturer, Department of Industrial Technology, Chiang Mai Rajabhat University, Thailand.
- Email: *sitthikorn_sit@cmru.ac.th*
- ORCID: 0000-0001-6107-8010
- Web of Science ResearcherID: NA
- Scopus Author ID: 57437092300
- Homepage: https://www.science.cmru.ac.th/courses/architectural-technology/

Rinchumphu, D., Ph.D., Associate Professor, Department of Civil Engineering, Chiang Mai University, Thailand.
- Email: *damrongsak.r@cmu.ac.th*
- ORCID: 0000-0002-6037-4974
- Web of Science ResearcherID: NA
- Scopus Author ID: 57219602276
- Homepage: https://civil.eng.cmu.ac.th/people/damrongsak-rinchumphu

HOW TO CITE THIS ARTICLE

Srivanit, M.; Pattanasri, S.; Phichetkunbodee, N.; Manokeaw, S.; Sitthikankun, S.; Rinchumphu, D., (2024). Analysis of sensitive urban form indicators of flood susceptible prediction based on machine learning models. Global J. Environ. Sci. Manage., 10(4): 1501-1518.

DOI: 10.22034/gjesm.2024.04.02

URL: https://www.gjesm.net/article_713741.html