International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 1
ISSN: 2088-5334

Modeling Flood Susceptible Areas Using Deep Learning Techniques with Random Subspace: A Case Study of the Mae Chan Basin in Thailand

Surachai Chantee [a], Theeraya Mayakul [a,*]

[a] *IT Management, Faculty of Engineering, Mahidol University, 999 Phuttamonthon 4 Road, Salaya, Nakhon Pathom, 73170, Thailand*
Corresponding author: [*]theeraya.may@mahidol.edu

***Abstract*— Flooding is a recurring global issue that leads to substantial loss of life and property damage. A crucial tool in managing and mitigating the impact of flooding is using flood hazard maps, which help identify high-risk areas and enable effective planning and management. This study presents a study on developing a predictive model to identify flood-prone areas in the Mae Chan Basin of Thailand using machine learning techniques, precisely the random sub-space ensemble method combined with a deep neural network (RS-DNN) and Nadam optimizer. The model was trained using 11 geographic information system (GIS) layers, including rainfall, elevation, slope, distance from the river, soil group, NDVI, road density, curvature, land use, flow accumulation, geology, and flood inventory data. Feature selection was carried out using the Gain Ratio method. The model was validated using accuracy, precision, ROC, and AUC metrics. Using the Wilcoxon signed-rank test, the effectiveness was compared to other machine learning algorithms, including random tree and support vector machines. The results showed that the RS-DNN model achieved a higher classification accuracy of 97% in both the training and testing datasets, compared to random tree (93%) and SVM (82%). The model's performance was also validated by its high AUC value of (0.99), compared to a random tree (0.93) and SVM (0.82) at a significance level of 0.05. In conclusion, the RS-DNN model is a highly accurate tool for identifying flood-prone areas, aiding in effective flood management and planning.**

***Keywords*— Deep neural networks; random subspace; flood forecasting; natural disaster; GIS.**

Manuscript received 1 Nov. 2023; revised 13 Jan. 2024; accepted 5 Feb. 2024. Date of publication 29 Feb. 2024.

I. INTRODUCTION

Flooding is a natural disaster that causes significant damage to human life and property. According to data collected by the UN Office for Disaster Risk Reduction (UNDRR), flooding was the most frequent natural disaster between 2000 and 2019, accounting for 44% of all-natural disasters and affecting up to 1.6 billion people worldwide. The economic cost of flood damage is 651 billion USD [1]. Although many countries have prepared to prevent and manage flooding to minimize losses, the latest data from 2020 indicates that the problem is still increasing and remains the most frequent natural disaster, accounting for 62% of all natural disasters and causing 40.92% of all disaster-related deaths [2]. This is mainly due to the significant increase in the frequency and severity of flooding caused by climate change [3].

Thailand is frequently affected by flooding, as reported by the Department of Disaster Prevention and Mitigation's statistics on natural disasters and damage figures from 2009-2018. The data reveals that the floods have resulted in 1,727 fatalities, impacted over 55 million people, and caused economic losses of over 50,000 million baht [4]. Mae Chan, an area in Thailand, experiences flooding almost every year. An example of flood damage in Mae Chan is shown in Fig. 1. This includes flash floods and river floods, which are attributed to various factors such as forest encroachment for agriculture. This encroachment intensifies the occurrence of flooding. The rise in average rainfall is a consequence of climate change and the encroachment on natural waterways designated for drainage. Additionally, constructing structures like residences and roads are barriers to these water channels [5].

Fig. 1 An example of flood damages in the Mae Chan basin

The impact of these floods has hindered the nation's progress, especially in regions that are not part of irrigation systems and lack facilities like dams for water management. The absence of such infrastructure exacerbates the devastation caused by flooding in these areas. While preventing floods altogether is challenging, strategic planning and prevention can mitigate the damage caused. Thus, reliable tools for flood prevention planning are essential, requiring accurate and easily understandable data to enable stakeholders to plan and respond to flood events effectively.

The utility of Flood Susceptibility Maps is paramount in the governance of flood-related challenges, as they facilitate the identification of vulnerable regions and the quantification of flood probabilities. This helps government authorities with the requisite knowledge to formulate evidence-based strategies for damage mitigation and future flood prevention [6]. Furthermore, businesses can also benefit from this tool for insurance purposes. Currently, analyzing flood-prone areas often involves handling large amounts of spatial data. Many researchers have utilized Geographic Information System (GIS) technology to convert and manage this data efficiently [7][8]. Additionally, incorporating Remote Sensing data, offering real-time, expansive aerial imagery, augments the precision of flood susceptibility forecasts, yielding superior outcomes [9], [10].

Many researchers have created flood model predictions to accurately determine flood susceptibility areas using different methods. These methods can be divided into three main types [11], as follows:

1) Hydrological methods are models that use software related to hydrological analysis, such as HEC-HMS, WetSpa, and SWAT. This software requires detailed data for processing, and extensive study areas require additional data collection, which increases the cost [12],[13]. Additionally, hydrological processing takes more time and resources than machine learning methods [14], [15].

2) Statistical and Expert-based methods use statistics and probabilities to predict flood susceptibility areas, such as frequency ratio, weights of evidence, and logistic regression. Traditional statistics are based on linear models, which do not correspond to flood data, a natural phenomenon with a complex non-linear nature, leading to low model accuracy [16] [17]. The accuracy of expert-based models depends on the knowledge and experience of the experts who weigh the scores [18].

3) Machine learning methods are data-driven learning techniques that can process non-linear data, resulting in higher accuracy than other methods. Additionally, they can be integrated with traditional models, which helps overcome their limitations[19]. Some examples of machine learning models are Support Vector Machines (SVM) [20], [21], Artificial Neural Networks (ANN) [22], Classification and Regression Tree (CART) [23], Multilayer Perceptron (MLP) [18], [24], Decision Tree (DT) [20], [25], Random Forest (RF), and Naïve Bayes (NB) [26], [27]. Ensemble and hybrid models, such as ensemble of bagging and Logistic Model Tree (LMT) [28], ensemble of Dagging and Classifier-M5P [29], Reduced Error Pruning Trees with Bagging (Bag-REPTree) [30] are also popular. However, no clear conclusion exists on which model best predicts flood risk areas in all situations.

The emergence of rapid data processing technologies has led to the proposition of Deep Learning Neural Networks as a methodology that offers enhanced learning depth and adaptability relative to conventional machine learning frameworks. However, a study has not been conducted on combining random subspace methods with deep learning algorithms applied to GIS data to enhance the accuracy of flood risk prediction models. Therefore, the aim of this study

is threefold: 1) to propose a Deep Neural Network model combined with the random subspace technique for predicting flood susceptible areas using GIS and remote sensing data, and to compare its performance with other algorithms, namely DNN, Random Tree, and SVM; 2) to rank the factors that influence flooding in the Mae Chan watershed, Thailand; and 3) to create a Flood Susceptibility Map that shows the level of risk as high, moderate, or low, to assist stakeholders in planning future flood prevention measures.

II. Materials and Method

A. Description of the Study Area

The Mae Chan Basin is in the northern part of Thailand, between longitude 99° 27' 20" E to 100° 06' 58" E and latitude 20° 01' 50" N to 20° 23' 39" N, illustrated in Fig. 2. The basin covers an area of approximately 1,193.17 square kilometers. It has a population of over 200,000 people. Most of the area is in Chiang Rai Province, with a small portion in Chiang Mai Province. The basin's topography is generally characterized by high mountains and complex terrain, with the elevation varying from 360 meters above sea level at the river mouth to 1,640 meters above sea level at the highest peak. The slope ranges from 0 to 52 degrees, with steep slopes primarily found in the headwaters in the northern and western parts of the basin. The slope gradually decreases in the middle and lower parts of the basin. The main rivers in the basin are the Mae Chan River and the Kham River, which have a length of 118 kilometers. The flow direction of the rivers is from west to east, where they join the Mekong River. Most of the area is covered by forest, accounting for 42% of the total area. Another 15% is occupied by paddy fields in the plains, while 14% is used for upland agriculture. The Mae Chan Basin is in the monsoon tropical zone, with an average annual rainfall of 1,632.66 millimeters. About 88.75% of the yearly rainfall occurs during the rainy season from May to October, which is influenced by the southwest monsoon from the Indian Ocean. The rainfall is highest in August and September, averaging 400-500 millimeters monthly. During this period, some areas may experience flooding. The flood problems in this area can be divided into two types: 1) Flash Floods, which occur suddenly due to heavy rainfall and are caused by the topography of the area being flat between mountains and the destruction of forest areas that are the sources of water, resulting in a combination of flash floods and torrential rivers. 2) River Floods occur when heavy and continuous rainfalls in areas vulnerable to water invasion cause rivers to overflow, inundating agricultural and residential areas [5]. According to the Department of Disaster Prevention and Mitigation, 2005 was the year that suffered the most damage from flooding. The total damage was valued at 112,560,400 billion baht, affecting 45,620 households.

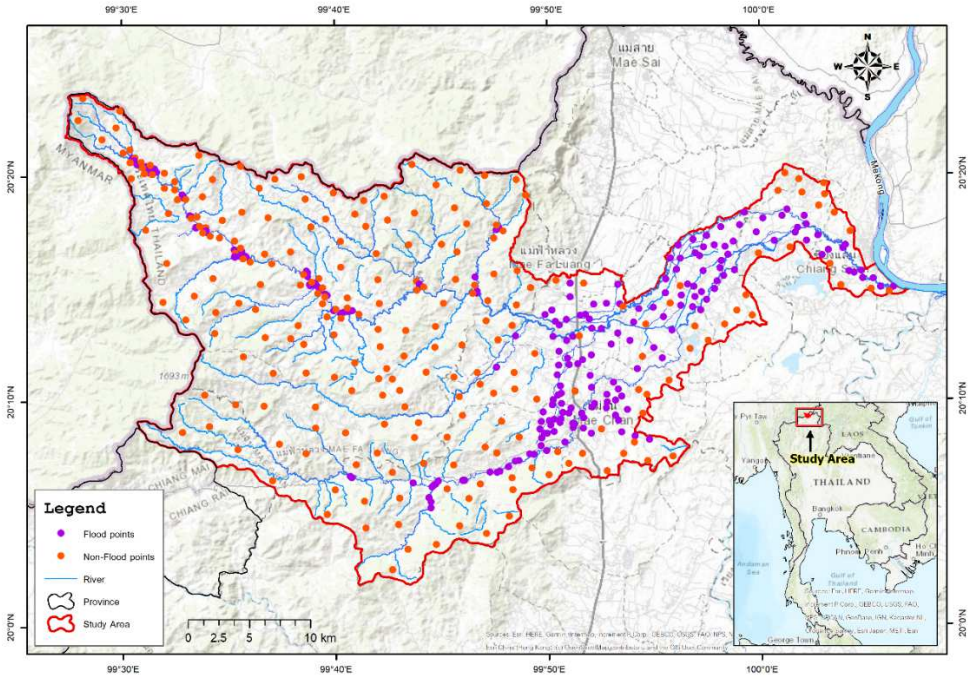


Fig. 2 Location of the Mae Chan Basin and the flood inventories

B. Data Used

1) Flood inventory: The Flood Inventory serves as an essential foundational dataset for identifying regions susceptible to flooding, given its significant impact on the accuracy of flood susceptibility maps [31]. In this study, data on flood locations spanning the years 2011-2021 were collected from the Geo-Informatics and Space Technology Development Agency (GISTDA), which included satellite images from THEOS, Sentinel1-2, Radarsat-2, field surveys conducted by the Department of Disaster Prevention and Mitigation, and news reports. A random selection of 240 flood locations was made, complemented by an equal number of non-flood locations from areas with no historical flooding, culminating in 480 data points. This dataset was subsequently partitioned into 70% for training (168 points) and 30% for testing (72 points). Binary values were assigned, with '1' denoting flood locations and '0' indicating non-flood locations, for utilization in subsequent predictive modeling.

2) Flood conditioning factors: The determinants of flood occurrences are not uniform across diverse geographical regions, with a range of topographic, meteorological, and hydrological variables contributing to flood risk. Such variables encompass factors like slope, DEM, rainfall, and proximity to rivers. However, selecting the most impactful flood conditioning factors for each study area remains controversial. Previous studies have explored different methodologies for selecting flood conditioning factors, but there is still an unclear conclusion on the best approach [30]. This study opted for flood conditioning factors frequently cited in related scholarly works to tackle this ambiguity. These include altitude, slope, distance from the river, rainfall, land use, NDVI, road density, soil group, curvature, flow accumulation, and geology. These factors were meticulously sourced from multiple authoritative agencies to ensure their accuracy and reliability. A comprehensive breakdown of these flood conditioning factors is presented in Table I.

Altitude indicates an area's elevation and is positively correlated with flood risk, given that water naturally flows from elevated to lower regions [32]. The study area exhibits an altitude range of 360-1640 meters above mean sea level. The slope is another pivotal factor influencing flood occurrences, as it dictates water flow directionality. The slope is another crucial factor influencing flood occurrences, as it dictates water flow directionality. Areas with steeper slopes are more susceptible to severe flooding as water flows from elevated to flat terrains [33]. The Mae Chan basin has a slope range between 0-54 degrees. Distance from the river is also directly related to the risk of flooding because, during rainfall, water flows towards the river and may overflow if it cannot drain away quickly enough. Areas farther from the river are less likely to flood [17]. The distance from the river is calculated using the buffer tool in ArcGIS and is measured in meters. The data range for distance from the river is 0-3,984 meters. Rainfall is the primary cause of flooding as it directly increases an area's water [34]. When rainfall exceeds the area's drainage capacity, water overflows and causes damage. Rainfall is positively correlated with flooding.

TABLE I
DATA COLLECTION

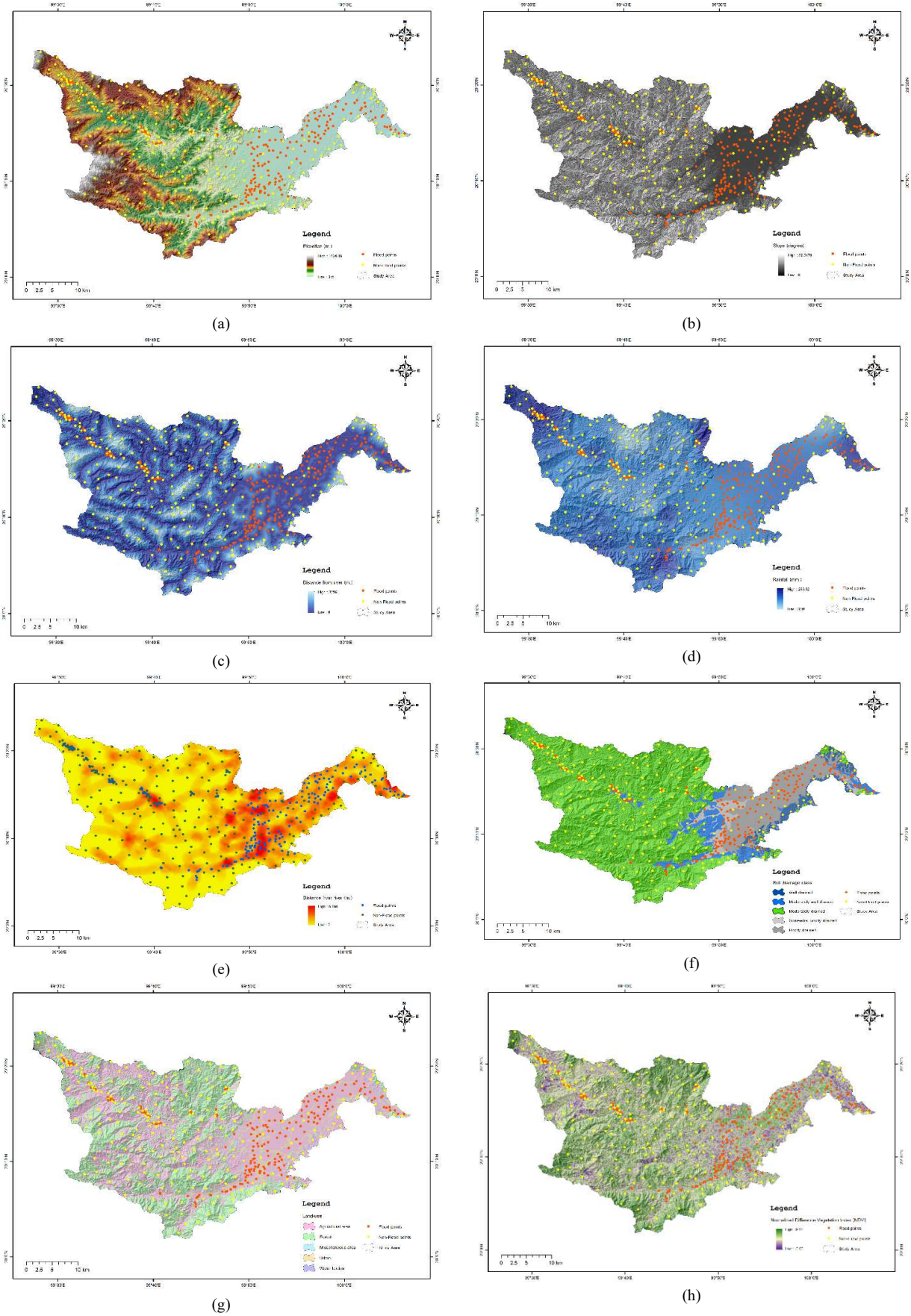
Dataset	Source	Resolution (m)	Format	Year
Altitude	Topographic Map 1:50000	30	.tif	2005
Slope	Topographic Map 1:50000	30	.tif	2005
Distance From River	Hydro-Informatics Institute	30	.tif	2020
Rainfall	Department of Water Resource	-	.shp	2011-2021
Land use	Land Develop Department	-	.shp	2018
NDVI	Landsat8 satellites	30	.tif	2021
Road density	Topographic Map 1:50000	30	.tif	2005
Soil Group	Land Develop Department	-	.shp	2018
Curvature	DEM	30	.tif	2005
Flow Accumulation	DEM	30	.tif	2005
Geology	Land Develop Department	-	.shp	2018
Flood	GISTDA	-	.shp	2011-2021

In this study, the rainfall data from the Department of Water Resources was collected and averaged for August, which is the month with the highest rainfall and flood incidents. The data were interpolated from 21 rain gauge stations using Inverse Distance Weighted (IDW) interpolation. The rainfall data range is between 3.59 - 283.82 mm. Land use is a type of land utilization that affects the occurrence of flooding. Rural areas with open spaces and no cover are more prone to frequent and severe flooding compared to areas with vegetation cover such as forests. Meanwhile, urban areas have a higher risk of flooding due to the impervious surfaces of structures, such as concrete, which make it difficult for water to seep through [32]. This information was obtained from the Land Development Department and converted into a raster format. NDVI is a dataset that displays vegetation cover on land, with values ranging between [-1,1]. It has an inverse relationship with the occurrence of flooding, where areas with more vegetation cover experience less flooding compared to open areas [13]ages acquired on March 7, 2021, were used to create the NDVI dataset, with less than 5% cloud cover (Land Cloud Cover). The NDVI values were calculated using the following formula [35]:

$$NDVI=(NIR-R)/(NIR+R) \quad (1)$$

where Red and NIR are the spectral reflectance measurements acquired in the red (visible) and near-infrared regions, respectively. Soil groups with different characteristics affect the drainage properties of the soil, which directly affects the occurrence of flooding in an area [36]. Soils that have good drainage properties are less prone to flooding. This information was obtained from the Department of Land Development and categorized according to the type of soil group. Geology is a dataset that displays information about the properties of rock layers that affect water retention. The type of rock and its properties can affect the rate of infiltration or surface runoff on the land [28]. The physical curvature of the Earth's surface is a factor that affects the flow of water in an area [37]. Concave areas can store more water, reducing the risk of flooding, compared to flat areas where water can accumulate and cause flooding. The data for curvature is generated from DEM using ArcMap software and is divided into 3 classes: (-3.99) - (-0.1), (-0.09) - (0.1), and (0.1) - (3.7), with a resolution of 30 meters. The direct relationship between flow accumulation and flooding makes areas with high flow accumulation more susceptible to flooding [38]. This is also generated from DEM data. Road Density is also a risk factor for flooding, as roads act as a barrier to the water flow [39]. In cases where water cannot drain quickly enough, flooding can occur. Therefore, areas with a high density of

roads are more likely to experience flooding. Road density is calculated within a 1 square kilometer area using ArcGIS software. Once all the data is collected, it is rasterized using the same coordinate system, UTM Zone 47N, with a pixel resolution of 30 meters for efficient processing using ArcGIS software. Display in the form of a map, as shown in Fig. 3.



(a) (b) (c) (d) (e) (f) (g) (h)

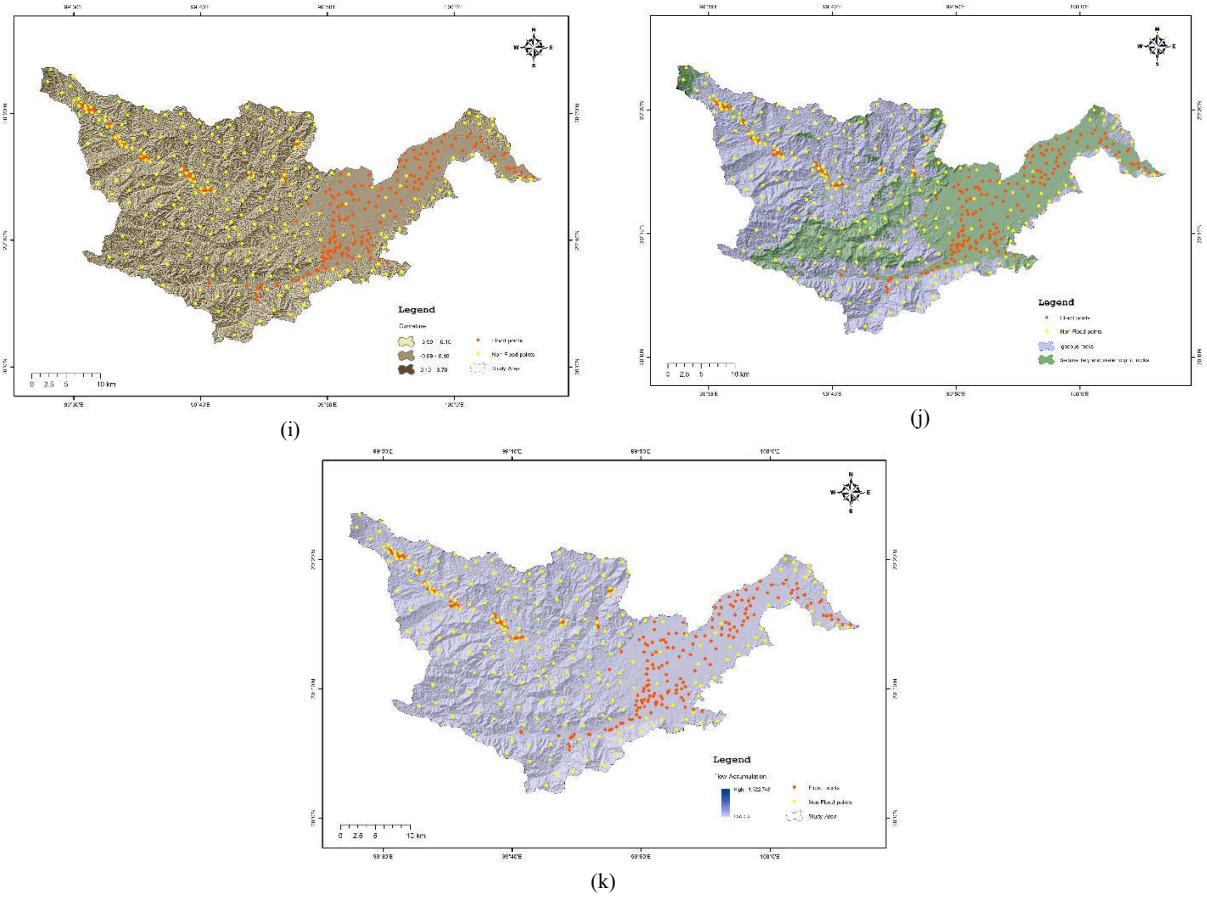


Fig. 3 Flood influencing factor: (a) Altitude; (b) Slope; (c) Distance from River; (d) Rainfall; (e) Road density; (f) Soil group; (g) Land use; (h) NDVI; (i) Curvature; (j) Geology and (k) Flow Accumulation

C. Theoretical Background of Methods Used

Mapping flood-prone areas involves several processes, from data collection, data preparation, factor selection, model building, model validation, and generating a flood susceptibility map. A process flowchart was created to provide an overview of the entire project, as shown in Fig. 4.

1) Standardization (z-score): Standardization (z-score) is a feature scaling technique used to adjust the range of data to be within the same scale. It adjusts the values to have a mean of 0 and a standard deviation of 1 to reduce bias in the machine learning algorithm computation and improve processing speed. In this study, the data was standardized to be within the range of [-1, 1] using the equation:

$$z = (x - \mu) / \sigma \quad (2)$$

where z is the z-score, x is the normalized value, μ is the population mean, and σ is the population standard deviation.

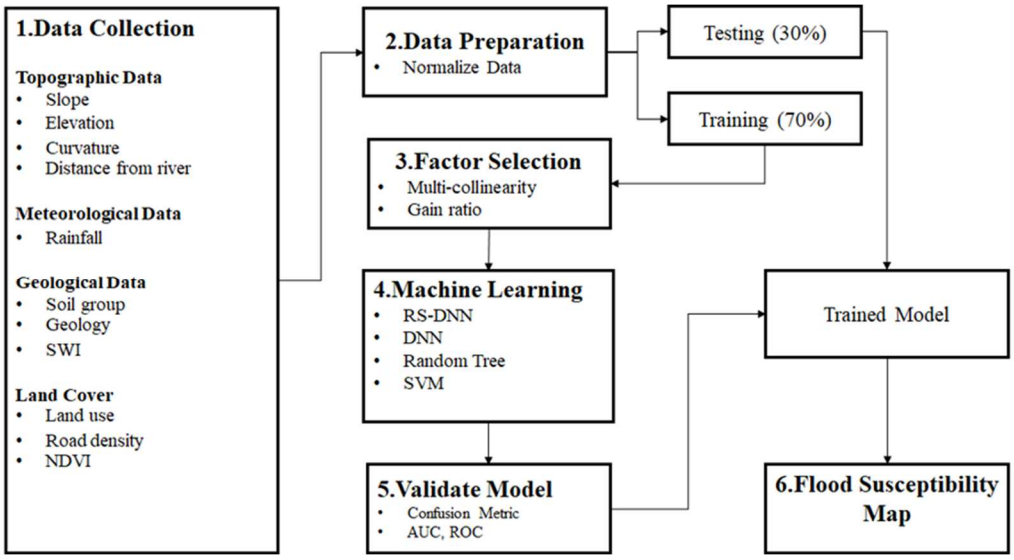


Fig. 4 Research Methodology

2) Factor selection: The selection of features for machine learning processing is crucial for the accuracy of the model and the efficiency of the processing. Therefore, the following criteria are considered for feature selection:

- Multicollinearity assessment

Multicollinearity poses a challenge in predictive modeling when two or more independent variables exhibit a high degree of correlation, potentially leading to inaccuracies in predictive equations [40]. This can lead to inaccuracies in the prediction equations [41]. Therefore, it is necessary to conduct a multicollinearity analysis of the variables that affect the occurrence of floods. This study used the Tolerance (TOL) and Variance Inflation Factor (VIF) to detect multicollinearity. A Tolerance value of less than 0.20 and a VIF below 5 were used as cutoffs to avoid collinearity problems [42]. Therefore, variables with values exceeding these cutoffs were not included in the model. The Variance Inflation Factor (VIF) and Tolerance are important in determining errors.

$$\text{Tolerance} = 1 - r^2 \quad (3)$$

$$\text{VIF} = \frac{1}{\text{Tolerance}} \quad (4)$$

- The Gain Ratio

In the model, each processed factor had a different impact on the occurrence of flooding, which can be determined by calculating the gain ratio using the following equation [29].

$$\text{Gain Ratio} = \text{Information Gain} / \text{Split Information} \quad (5)$$

where Information Gain is the difference between the entropy of the target variable before and after the split. Split Information measures the amount of information required to represent the distribution of the feature being split.

This study aims to demonstrate the influence of flood conditioning factors on flood occurrences. Flood conditioning factors with high GR values are more effective for flood prediction [33]. Therefore, only those factors with GR values exceeding zero will be selected for integration into the flood prediction model.

3) Machine Learning Algorithms:

- Deep Neural Network

Deep Neural Network (DNN) is a subset of machine learning that employs artificial neural networks to model and solve complex problems. DNN is characterized by multiple interconnected layers of nodes, also known as artificial neurons, which sequentially process input data to generate output. Each subsequent layer refines the output of its predecessor, enabling the model to learn more complex representations of the input data. In this study, we meticulously designed a neural network architecture with a total of six layers. The input layer encompasses 11 crucial factors contributing to flooding, namely altitude, slope, distance from river, rainfall, road density, soil group, land use, NDVI, curvature, geology, and flow accumulation. Subsequently, we established Hidden Layer 1 with 22 nodes and Hidden Layer 2 with 16 nodes, strategically optimizing the network's depth and complexity. To enhance the model's robustness, a dropout layer was introduced with a dropout rate (p) of 0.8, preventing overfitting and promoting generalization. Furthermore, Hidden Layer 3 was configured with 4 nodes, contributing to the intricate nature of the network. Finally, the output layer was structured with 2 nodes to facilitate the extraction of meaningful insights. To identify the optimal configuration, a meticulous grid search was conducted, employing the Nadam optimization algorithm and utilizing Mean Squared Error (MSE) as the Cost Function. The model has been formulated with specifics illustrated in Fig. 5.

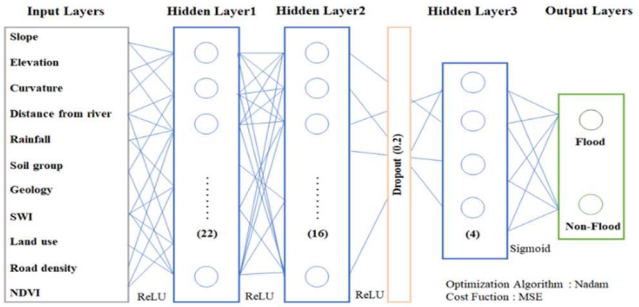


Fig. 5 The DNN Model for Flood Susceptibility

- Random Subspace

Random Subspace (RS) is a technique used in machine learning for building models by randomly selecting a subset of features from the dataset, as specified by the user-defined proportion in the parameters, instead of using the entire dataset [29]. This approach mitigates the risk of overfitting, which occurs when a model performs well on the training data but underperforms on new, unseen data. Additionally, random subspace can enhance computational efficiency by reducing the dataset's dimensionality, making it especially advantageous for handling large datasets. Moreover, this technique enhances the model's performance by allowing it to learn from a more diverse data set than the original dataset [30]. In this study, subSpaceSize was set to 0.6, and numIterations was established at 5. When combined with the DNN model, as illustrated in Figure 6.

- SVM

Support Vector Machine (SVM) is a powerful and widely used machine learning technique including GIS data analysis. It can be used for both regression and classification problems. In flood risk area analysis, SVM serves as a valuable tool for pinpointing areas susceptible to flooding [43]. The parameters for the SVM model were configured as follows: c=1.0, kernel= RBF, batchSize = 32, and numDecimalPlaces = 2

- Random Tree

Random Tree is a decision tree algorithm applicable to both classification and regression tasks. This algorithm is one of the decision tree algorithms with high accuracy [44], [45]. One of its advantages is its ability to reduce noise and overfitting. Furthermore, it is computationally efficient, requiring fewer computational resources for execution. In this Model, the following parameters were set: KValue = int(log_2(features) + 1), batchSize = 32, maxDepth = unlimited, numFolds = 0 (no backfitting), and numDecimalPlaces = 2.

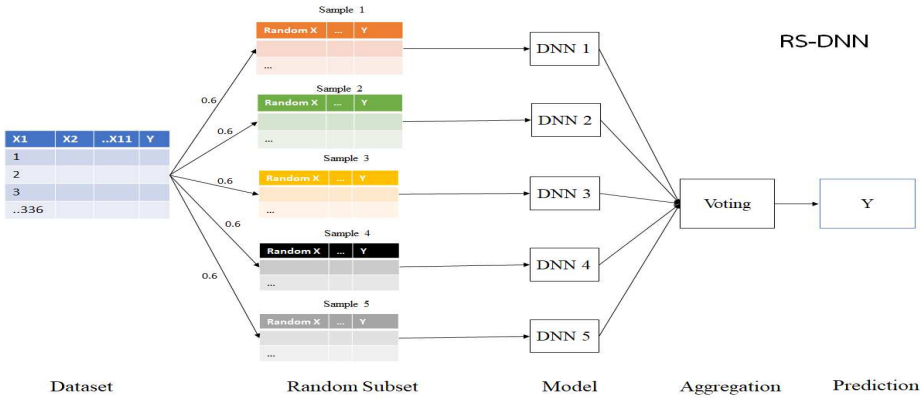


Fig. 6 The RS-DNN Model for Flood Susceptibility

4) Evaluation Method:

- Confusion matrix

A confusion matrix is a tool used in machine learning to evaluate the performance of a classification model.[46] It provides a clear visual representation of the predicted and actual values, allowing us to calculate various evaluation metrics such as accuracy, precision, recall, and F1 score. Accuracy measures the overall performance of the model in predicting the correct class, while precision measures the model's ability to correctly identify the positive class. Recall measures the model's ability to identify all positive cases correctly, and F1-score is a weighted average of precision and recall. These evaluation metrics can be computed using the following system of equations [47].

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \quad (6)$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad (7)$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (8)$$

$$\text{F1} - \text{score} = \frac{2 * (\text{Precision} * \text{Recall})}{(Precision + Recall)} \quad (9)$$

- Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC)

The ROC curve functions as an analytical instrument for identifying the most suitable cut-off threshold in the decision-making framework of the evaluated model. The graph illustrates a relationship with the X-axis representing 1-Specificity, and the Y-axis representing Sensitivity, both ranging from 0 to 100. The optimal cut-off point is identified at the juncture where Sensitivity approaches 100 and 1-Specificity approaches 0.[48] Similar to AUC, it is a metric used to estimate the area under the ROC curve, providing an evaluation of the overall classification performance of the model. Values approaching 1 indicate strong discriminatory ability with high accuracy, while values below 0.5 indicate poor discrimination. Not applicable for use.

D. Flood Susceptibility Mapping

Once the data processing stage is complete, the flood susceptibility model generates probabilities of flood occurrence, which are then utilized to create a flood risk map. The researcher employed the natural break (Jenks) technique [30], [16], a statistical method, to classify the flood risk level of the susceptible areas into four groups based on the model's predictive values. These risk categories are delineated as very low risk, low risk, moderate risk, and high risk. The advantage of using the natural break technique is that it identifies the class boundaries from data based on its inherent structure, rather than being arbitrarily divided. The classification and mapping were executed using the ArcGIS application.

III. Results and Discussion

A. Multicollinearity

Table II shows the results of a multicollinearity analysis, which examines the relationship between the predictor variables in a regression model. The analysis produces two measures of multicollinearity: Tolerance and VIF (Variance Inflation Factor). None of the factors exhibited any multicollinearity concerns, as their tolerance values exceeded 0.2, and VIF values remained below 5.

TABLE II
THE RESULT OF A MULTICOLLINEARITY ANALYSIS

Factors	Tolerance	VIF
Curvature	0.951	1.052
Flow accumulation	0.932	1.073
Rainfall	0.863	1.158
NDVI	0.836	1.196
Distance from river	0.758	1.319
Road Density	0.725	1.378
Soil	0.625	1.6
Geology	0.588	1.7
Land use	0.533	1.875
Altitude	0.445	2.245
Slope	0.352	2.843

The tolerance values for all considered factors span a range from 0.352 to 0.951. Among them, the variable "Curvature" boasts the highest tolerance value at 0.951, and is followed by "Flow Accumulation" and "Rainfall." Conversely, "Land Use," "Altitude," and "Slope" exhibit the lowest tolerance values, with the minimum recorded at 0.352. Thus, multicollinearity does not appear to be a concern in this model. Similarly, the VIF values for all factors range between 1.052 and 2.843. "Slope" registers the highest VIF value at 2.843, while this suggests a relatively strong correlation with other independent variables in the model. It does not reach the threshold commonly associated with multicollinearity, which is generally considered to occur when VIF values exceed 5.

Consequently, all factors were deemed suitable for inclusion in this research.

B. Feature selection

To improve the accuracy of the model and detect any irrelevant factors that may reduce its predictive power of the model. The Gain Ratio (GR) took into account the inherent information content of a feature, based on the entropy of the class distribution. It quantifies the relationship between the information gain and the intrinsic value of the feature. The gain ratio was used in this study to examine the influence of all factors, and we employed a K-fold cross-validation technique (K=10) to ensure a balanced distribution of data and reduce bias in data splitting. Factors with high GR values were identified as significant influencers of flooding in the model, while those with a value of zero had no impact on flooding [29]. The detailed results are presented in Table 3.

TABLE III
PREDICTIVE CAPABILITIES OF FLOOD INFLUENCING FACTORS

Factors	Average Merit	Std. Dev.
Distance from river	0.321	0.059
Slope	0.319	0.022
Soil Group	0.278	0.031
Land use	0.271	0.006
Geology	0.259	0.012
Curvature	0.243	0.243
Elevation	0.217	0.005
Road Density	0.17	0.014
Flow Accumulation	0.122	0.01
Rainfall	0.113	0.011
NDVI	0.064	0.006

We analyzed the factors influencing flooding in the Mae Chan Basin. Distance from the river emerged as the most significant, with the highest Gain Ratio (GR) of 0.321, corroborating previous studies [16], [30] that have identified the proximity to water bodies as a major risk factor for flooding. Slope followed closely with a GR of 0.319, indicating that areas with low slopes are more prone to flooding due to poor water drainage, while steep slopes facilitate rapid water flow. This finding corresponds to the location of past floods that often occurred on the eastern side of the flat area. Soil Group was the third most influential factor with a gain ratio of 0.278. Soil series such as (Hd), (Lgu), and (Ph), which are poorly drained clay series, directly affect flooding. Land use (0.271) is also a significant factor that affects flooding, with flooding occurring mostly in open areas. Geology (0.259) and curvature (0.243) are factors related to flow and water retention, while Road Density (0.17) density often affects the occurrence of flooding. Flow Accumulation (0.122) is the point of water consolidation, which is directly related to flooding, and Rainfall (0.113) is a significant factor that adds water to the area. It is worth noting that in the Mae Chan Basin, most of the precipitation falls in the forest area, which is the watershed area, and with the steepness of the area, the water flows rapidly into the river, making the two factors, NDVI (0.064) and Rainfall (0.113), less correlated. As all factors had a Gain Ratio value greater than zero, it can be inferred that each factor has an impact on flooding [31]. Hence, all factors were used to train the model.

C. Model validation and comparison

1) Confusion matrix: The performance of the training data was evaluated using a confusion matrix, with the corresponding results presented in Table IV. Likewise, the performance of the testing data was assessed using a confusion matrix, and the corresponding results are depicted in Table V.

TABLE IV
PERFORMANCE OF MODELS ON THE TRAINING DATASET (K-FOLD, K=10)

Algorithms	TP	TN	FN	FP	accuracy	precision	recall	F1
Random-DNN	160	160	8	8	0.952	0.952	0.952	0.952
DNN	155	163	13	5	0.946	0.969	0.923	0.945
Random Tree	155	160	13	8	0.937	0.951	0.923	0.937
SVM	137	150	31	18	0.854	0.884	0.815	0.848

TABLE V
PERFORMANCE OF MODELS ON THE TESTING DATASET (K-FOLD, K=10)

Algorithms	TP	TN	FN	FP	accuracy	precision	recall	F1
Random-DNN	69	71	3	1	0.972	0.986	0.958	0.972
DNN	66	69	6	3	0.937	0.957	0.917	0.936
Random Tree	65	69	7	3	0.930	0.956	0.903	0.929
SVM	60	59	12	13	0.826	0.822	0.833	0.828

Table 4 displays the results of accuracy measurement using training data, indicating that the random-DNN model achieved the highest accuracy score at 0.952, followed by the DNN at 0.946, the random tree model at 0.937, and the SVM model at 0.854. Corresponds to the F1 score, representing the harmonic mean of precision and recall, were also calculated. The random-DNN model achieved the highest F1 score of 0.952, followed by the DNN at 0.945, the random tree model at 0.937, and the SVM model at 0.848. It is important to note that these measures have predefined criteria, with a perfect accuracy score being 1.0 and a worst-case scenario scoring of 0.0.[49]

The outcomes of the testing data are illustrated in Table V, where the random-DNN model displayed the highest accuracy score at 0.972, followed by the DNN at 0.937, the random tree model at 0.930, and the SVM model at 0.826. Precision, which reflects the predictive accuracy of the random-DNN model, stood out as the highest at 0.986, trailed by the DNN at 0.957, the random tree model at 0.956, and the SVM model at 0.822. This aligns with the prior training data.

The RS-DNN model outperformed other models, registering the highest True Negative (TN) and True Positive (TP) scores while minimizing False Negative (Type II error) and False Positive (Type I error). This led to superior accuracy, precision, recall, and F1 scores. The model's efficacy is largely attributable to the random subspace technique, which enhances learning by randomly selecting feature subsets. However, this approach did increase computational time, a trade-off for the improved predictive performance achieved through ensemble learning.

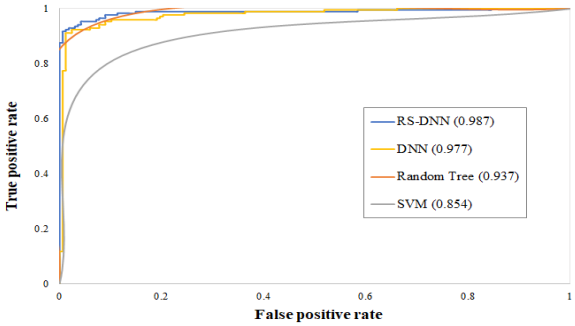


Fig. 7 The ROC and AUC on a training dataset.

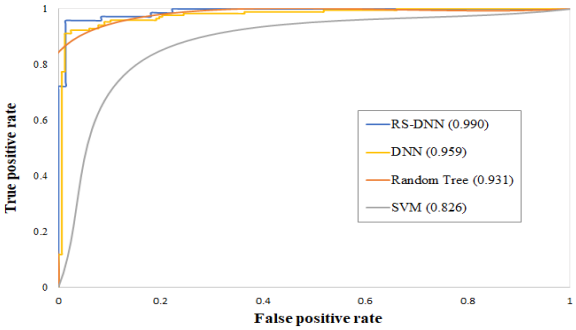


Fig. 8 The ROC and AUC on the testing dataset.

Based on the ROC curve for the training data shown in Figure 7, the Random-DNN model (blue) demonstrated the closest prediction performance to the cut-off point (upper left), followed by the DNN model (yellow), the random tree model (orange), with the gray SVM model performing the least accurately. The area under the curve (AUC) for the RS-DNN, DNN, random tree, and SVM models were 0.987, 0.977, 0.937, and 0.854, respectively.

Similarly, the results for the testing data presented in Figure 8 revealed that the RS-DNN model again showed the closest predictive performance to the cut-off point, followed by the random tree model (orange), with the SVM model (gray) exhibiting the least accurate performance. The AUC for the RS-DNN, DNN, random tree, and SVM models were 0.990, 0.959, 0.931, and 0.826, respectively. All the models performed well above the threshold value of 0.5.

After evaluating the performance of the models using the ROC and AUC graphs, the RS-DNN model demonstrated superior performance compared to other models.[50]

3) Statistical Test: In this study, we performed a statistical analysis using a Wilcoxon signed-rank test to compare the differences between the machine learning models [35]. The analysis results were calculated using the Weka application and are presented in Table VI.

TABLE VI
RESULTS OF COMPARISON OF THE PREDICTION PERFORMANCE USING THE WILCOXON SIGNED-RANK TESTS AT THE SIGNIFICANT LEVEL $\alpha = 5\%$.

Pairwise Comparison	z-value	p-value	Effect Size	Sig.
RS-DNN vs. DNN	-2.8908	0.0004	0.37	yes
RS-DNN vs. Tree	-3.1975	0.0001	0.41	yes
RS-DNN vs. SVM	-4.7872	< 0.0001	0.62	yes

We used the Wilcoxon signed-rank test on the training data to compare the performance differences between the RS-DNN and DNN models at a 95% significance level. The p-value of 0.0004 was less than 0.05, and the corresponding z-value of -2.8908 fell outside the range (-1.96, 1.96). Therefore, we conclude that the RS-DNN and the DNN model show significant performance differences at the 0.05 level. The calculated effect size of 0.37, between 0.3 to 0.5, indicates a medium effect [51]. Similarly, comparing the RS-DNN and Tree models yielded a p-value of 0.0001, less than 0.05, and a z-value of -3.1975, which fell outside the range (-1.96, 1.96). Thus, we conclude that the RS-DNN and Tree models exhibit significant performance differences at the 0.05 level. The calculated effect size of 0.41, between 0.3 to 0.5, indicates a medium effect. Finally, the comparison between the RS-DNN and SVM models using the same test resulted in a p-value of < 0.0001, less than 0.05, and a corresponding z-value of -4.7872, falling outside the range (-1.96, 1.96). Thus, we conclude that the RS-DNN and SVM model also display significantly different performance at the same 0.05 level. The calculated effect size of 0.62, which surpasses the threshold of 0.5, indicates a significant effect [51].

Upon comparing the model's performance using the confusion matrix table, ROC-AUC, and statistical tests, all three methods consistently indicate that the Ensemble model, precisely the RS-DNN combination, is more effective in classifying flood areas than using a single model. Employing diverse datasets for each model mitigates the overfitting issues often associated with DNN models and enhances the model's learning capabilities by exposing it to various data. Using multiple processing models also provides robustness against noisy data through aggregation, minimizing errors from data noise. However, it is essential to acknowledge that this advantage comes at the cost of increased time and computational resource requirements[52].

D. Generating flood Susceptibility Maps

After completing the model, the researcher employed all the available data from the Mae Chan Basin to predict and produce a flood risk map. The map is classified into four levels of risk: high risk, moderate risk, low risk, and shallow risk. It is visually represented in Fig 9.

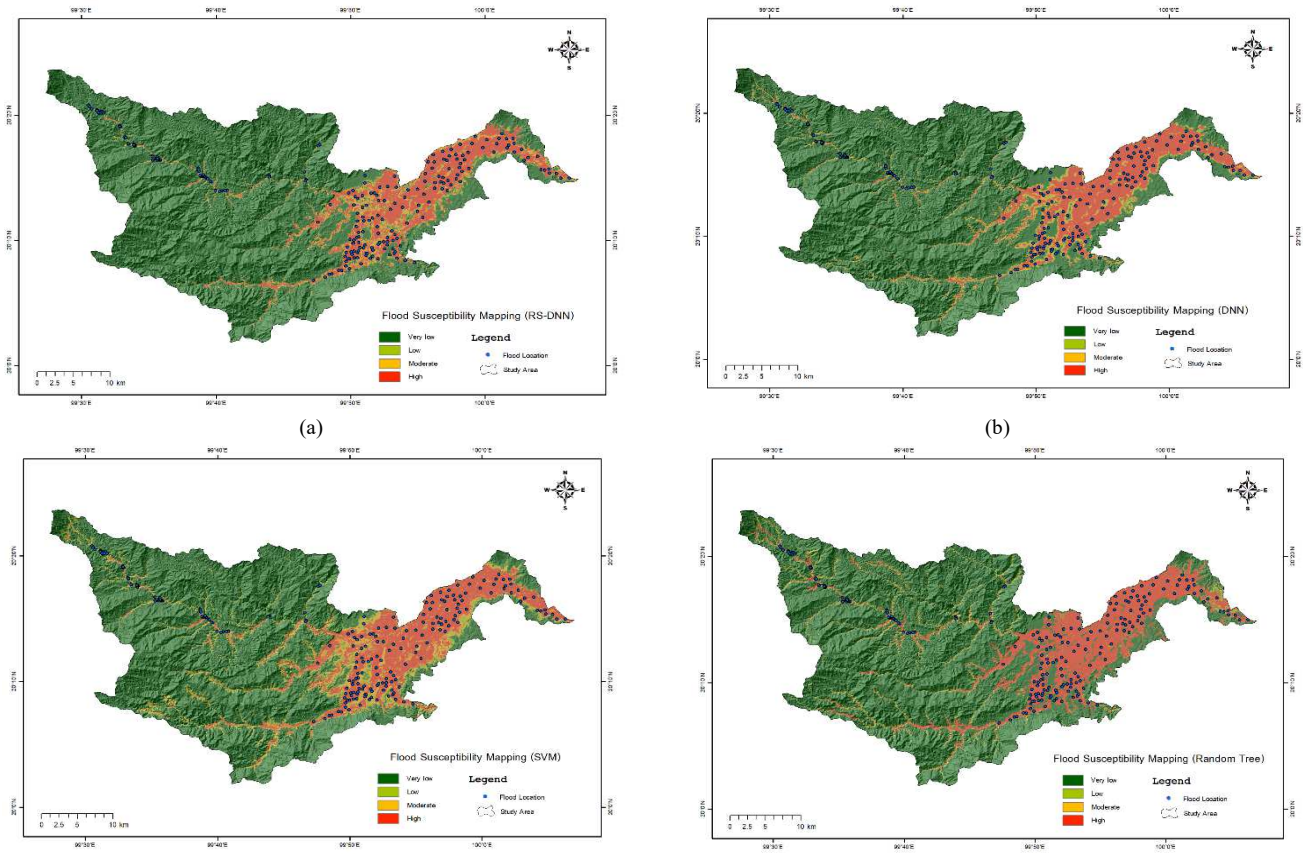


Fig. 9 Flood Susceptibility Mapping: (a) RS-DNN; (b) DNN; (c) SVM; (d) Random Tree

The low-risk level, which all four models predicted, had the most significant predicted area, at approximately 70-80% of the total area. This corresponds to the topographical conditions in the northern and western mountainous regions. The high-risk level was predicted by the RS-DNN model with the lowest rate of 12.95% and by the random tree model with the highest rate of 18.78%.

Most high-risk areas were in the eastern part of the watershed, near the mouth of the river, where it joins the Mekong River. This is consistent with the topographical conditions in that area, where there are many waterways, including the main rivers and numerous canals. The water flowing down from the watershed's upper and eastern parts passes through this area. In addition, the soil in this area is clay, which makes drainage difficult, and the relatively flat slope exacerbates the problem.

We validated it against historical flood points from field surveys. The results from the model correspond to the observed flood points. Figure 10 shows risky areas in valleys where flash floods occur. The model predicts high-risk areas primarily located in areas that receive water from streams and have steep slopes. Figure 11 shows risky regions of plain areas caused mainly by river floods. There is an excess of water in the main rivers and streams due to the inflow from upper areas and rainfall in the region. This excess water prevents proper drainage, leading it to flow into houses. The risk is reduced when farther from waterways.

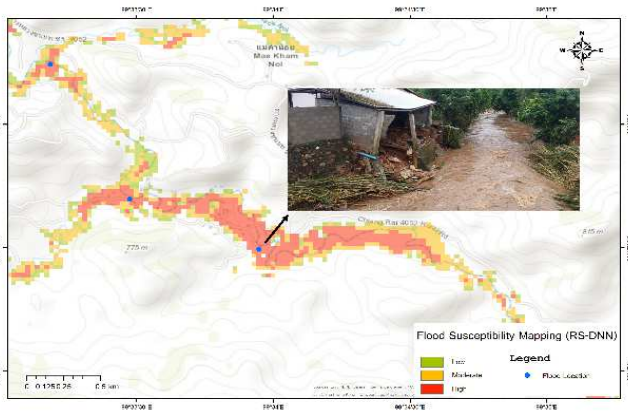


Fig. 10 Flooding locations in mountain areas.

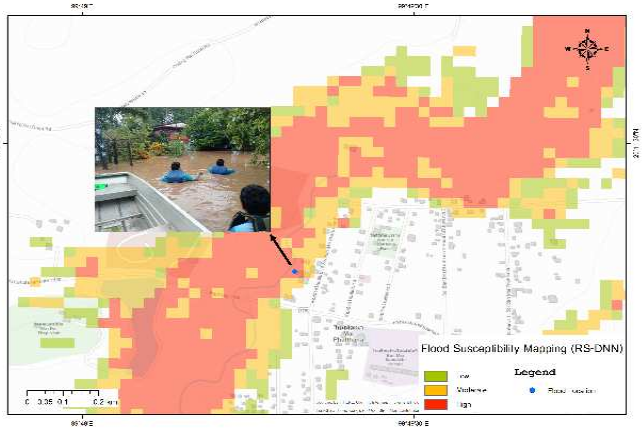


Fig. 11 Location of flooding in flat areas.

We focused on the high-risk areas in the mountainous forest, which often experience flash floods predicted by all 4 models in Fig 12. Notably, the RS-DNN model's area has a clustered and non-dispersed pattern that closely aligns with the historical data of actual flood locations. Similarly, the DNN also showed a clustered high-risk pattern, although the regions were smaller and more dispersed. In contrast, the Random Tree and SVM models have small and widely scattered high-risk areas, which deviate from the actual historical flood data and may be noise resulting from inaccurate calculations. These discrepancies could be attributed to computational noise and suggest that corrections are needed before relying on these models for accurate flood prediction.

A flood susceptibility evaluation was conducted for the study area using the RS-DNN model. The findings indicate that the lower part of the watershed, which serves as the water intake point, is at the highest risk of flooding. In contrast, the upper areas are considered a slightly lower risk. To quantify the extent of the risk, the researcher calculated the size of each flood-prone area. The results are presented in Table VII.

TABLE VII
THE SIZE OF FLOOD SUSCEPTIBILITY AREAS FOR EACH RISK LEVEL

Flood Susceptibility	RS_DNN	DNN	Tree	SVM
Very low	968.74	950.35	910.34	858.21
Low	19.45	33.47	52.29	69.13
Moderate	58.61	49.81	14.08	78.75
High	155.75	168.91	225.84	196.45
Total (km^2)	1202.55	1202.55	1202.55	1202.55

The model displays the locations of areas at risk of flooding, with a notable concentration of risk near waterways. The northern part of the area is mountainous, and the community is dispersed. Hazard maps can assist community leaders in implementing appropriate tools for citizen alerts and delineating zones where construction is prohibited to avert potential disasters.

In the lower part of the area, which is a lowland region receiving water from the north, risk maps prove valuable for water resource managers. They can analyze and establish water catchment areas to prevent overflow from insufficient drainage, storing water for use in the dry season. This includes determining the optimal distance for digging canals to facilitate proper water drainage.

IV. CONCLUSION

The utilization of machine learning techniques in the model has shown good results in providing accurate and reliable predictions of flood susceptible areas. This study aimed to explore the potential of several machine learning algorithms, including RS-DNN, deep neural network, support vector machine, and decision tree, to identify flood-susceptible areas using various physical factors such as topography, land use, and hydrological data.

The results of this study showed that the RS-DNN algorithm outperformed the other models in terms of prediction accuracy and reliability, achieving an overall accuracy of > 90%. This indicates that the RS_DNN algorithm is an effective and robust tool for mapping flood susceptible areas.

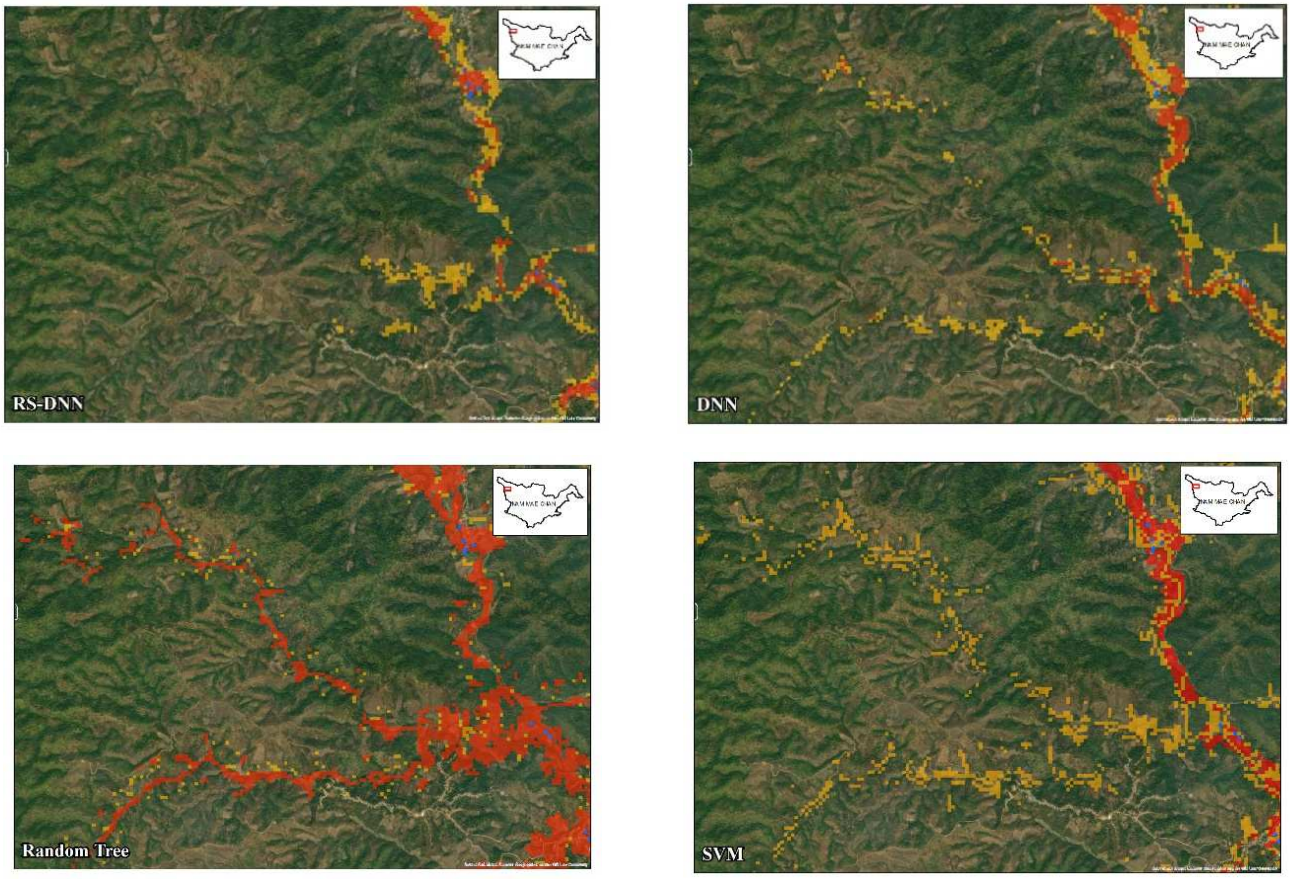


Fig. 12 Comparing the results of each model's high-risk area

Furthermore, the random subspace technique notably enhances the model's accuracy while reducing Type I and Type II errors. This lays the groundwork for developing a more effective flood-prone area prediction model. Based on the analysis using gain ratio, the top three factors that significantly affect the occurrence of flooding in the Mae Chan watershed area are the distance from the river, slope, and soil group. The proximity to the river poses a higher risk of river flooding, while the hill directly affects waterlogging. Flat areas are more susceptible to flooding than sloping areas. Additionally, soil groups with poor drainage properties can contribute to the increased risk of flooding. Therefore, these key factors should be prioritized when developing flood risk maps and management strategies for the Mae Chan watershed.

Finally, the flood risk map derived from the model's processing will serve as a valuable resource for officials and stakeholders in formulating response plans to mitigate potential impacts within the area. However, it is crucial to note that the model's applicability to other regions may depend on environmental and geographical conditions, necessitating parameter adjustments.

This research thoroughly assesses flood risk using GIS data and machine learning algorithms. The RS-DNN model has demonstrated its efficacy in accurately predicting flood-prone areas, emphasizing the role of advanced technologies in mitigating the adverse effects of natural disasters. The findings also suggest that this approach could be adapted for regions with similar geographical and environmental traits.

ACKNOWLEDGMENT

We extend our sincere gratitude to the Hydro Informatics Institute for generously providing the invaluable data that significantly contributed to the success of this research.

REFERENCES

[1] UNDRR, "The human cost of disasters: an overview of the last 20 years (2000-2019)," 2020. [Online]. Available: https://www.undrr.org/publication/human-cost-disasters-overview-last-20-years-2000-2019

[2] National Disaster Reduction Center of China, "2020 Global Natural Disaster Assessment Report," 2021. [Online]. Available: https://www.preventionweb.net/publication/2021-global-disaster-assessment-report

[3] Y.-J. Chen, H.-J. Lin, J.-J. Liou, C.-T. Cheng, and Y.-M. Chen, "Assessment of Flood Risk Map under Climate Change RCP8.5 Scenarios in Taiwan," *Water (Basel)*, vol. 14, no. 2, 2022, doi:10.3390/w14020207.

[4] NSO, "Statistics of flood in Thailand : 2009-2020," 2021. [Online]. Available: http://statbbi.nso.go.th/staticreport/page/sector/th/21.aspx

[5] HII, "The operations regarding data collection and analysis for the development of a data repository system for 25 river basins and a simulation model for flood and drought events.," Hydro – Informatics Institute, 2012. [Online]. Available: https://tiwrm.hii.or.th/web/attachments/25basins/02-khong.pdf.

[6] M. Ahmadlou *et al.*, "Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder neural networks," *J Flood Risk Manag*, vol. 14, no. 1, p. e12683, 2021, doi: 10.1111/jfr3.12683.

[7] Y. Berkat and H. Fitriana, "Flood Prone Analysis Using GIS and Remote Sensing Data; Case Study in Semarang, Central Java," *IOP Conf Ser Earth Environ Sci*, vol. 874, p. 12004, Dec. 2021, doi:10.1088/1755-1315/874/1/012004.

[8] A. R. M Amen *et al.*, "Mapping of Flood-Prone Areas Utilizing GIS Techniques and Remote Sensing: A Case Study of Duhok, Kurdistan Region of Iraq," *Remote Sens (Basel)*, vol. 15, no. 4, 2023, doi:10.3390/rs15041102.

[9] P. Hansana, X. Guo, S. Zhang, X. Kang, and S. Li, "Flood Analysis Using Multi-Scale Remote Sensing Observations in Laos," *Remote Sens (Basel)*, vol. 15, no. 12, 2023, doi: 10.3390/rs15123166.

[10] H. S. Munawar, A. W. A. Hammad, and S. T. Waller, "Remote Sensing Methods for Flood Prediction: A Review," *Sensors*, vol. 22, no. 3, p. 960, Jan. 2022, doi: 10.3390/s22030960.

[11] M. M. Msabi and M. Makonyo, "Flood susceptibility mapping using GIS and multi-criteria decision analysis: A case of Dodoma region, central Tanzania," *Remote Sens Appl*, vol. 21, p. 100445, 2021, doi:10.1016/j.rsase.2020.100445.

[12] L. S. Bruno, T. S. Mattos, P. T. S. Oliveira, A. Almagro, and D. B. B. Rodrigues, "Hydrological and Hydraulic Modeling Applied to Flash Flood Events in a Small Urban Stream," *Hydrology*, vol. 9, no. 12, 2022, doi: 10.3390/hydrology9120223.

[13] V. Kumar, K. V. Sharma, T. Caloiero, D. J. Mehta, and K. Singh, "Comprehensive Overview of Flood Modeling Approaches: A Review of Recent Advances," *Hydrology*, vol. 10, no. 7, 2023, doi:10.3390/hydrology10070141.

[14] P. Jimeno-Sáez, R. Martínez-España, J. Casalí, J. Pérez-Sánchez, and J. Senent-Aparicio, "A comparison of performance of SWAT and machine learning models for predicting sediment load in a forested Basin, Northern Spain," *Catena (Amst)*, vol. 212, p. 105953, 2022, doi:10.1016/j.catena.2021.105953.

[15] P. Kumar *et al.*, "Nature-based solutions efficiency evaluation against natural hazards: Modelling methods, advantages and limitations," *Science of The Total Environment*, vol. 784, p. 147058, 2021, doi:10.1016/j.scitotenv.2021.147058.

[16] R. Costache and D. Tien Bui, "Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania," *Science of The Total Environment*, vol. 691, pp. 1098–1118, 2019, doi:10.1016/j.scitotenv.2019.07.197.

[17] S. V. Razavi Termeh, A. Kornejady, H. R. Pourghasemi, and S. Keesstra, "Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms," *Science of The Total Environment*, vol. 615, pp. 438–451, 2018, doi:10.1016/j.scitotenv.2017.09.262.

[18] T. K. Saha *et al.*, "How far spatial resolution affects the ensemble machine learning based flood susceptibility prediction in data sparse region," *J Environ Manage*, vol. 297, p. 113344, 2021, doi:10.1016/j.jenvman.2021.113344.

[19] B. T. Pham *et al.*, "Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling?," *J Hydrol (Amst)*, vol. 592, p. 125615, 2021, doi:10.1016/j.jhydrol.2020.125615.

[20] M. Ganjirad and M. R. Delavar, "Flood Risk Mapping Using Random Forest and Support Vector Machine," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. X-4/W1-2022, pp. 201–208, 2023, doi: 10.5194/isprs-annals-X-4-W1-2022-201-2023.

[21] A. Salvati *et al.*, "Flood susceptibility mapping using support vector regression and hyper-parameter optimization," *J Flood Risk Manag*, vol. 16, no. 4, p. e12920, 2023, doi: 10.1111/jfr3.12920.

[22] G. Wang, J. Yang, Y. Hu, J. Li, and Z. Yin, "Application of a novel artificial neural network model in flood forecasting," *Environ Monit Assess*, vol. 194, Dec. 2022, doi: 10.1007/s10661-022-09752-9.

[23] M. Ahmadlou, Y. Ebrahimian Ghajari, and M. Karimi, "Enhanced classification and regression tree (CART) by genetic algorithm (GA) and grid search (GS) for flood susceptibility mapping and assessment," *Geocarto Int*, vol. 37, no. 26, pp. 13638–13657, Dec. 2022, doi:10.1080/10106049.2022.2082550.

[24] S. Shadkani, A. Abbaspour, S. Samadianfard, S. Hashemi, A. Mosavi, and S. S. Band, "Comparative study of multilayer perceptron-stochastic gradient descent and gradient boosted trees for predicting daily suspended sediment load: The case study of the Mississippi River, U.S," *International Journal of Sediment Research*, vol. 36, no. 4, pp. 512–523, 2021, doi: 10.1016/j.ijsrc.2020.10.001.

[25] F. M. Aswad, A. N. Kareem, A. M. Khudhur, B. A. Khalaf, and S. A. Mostafa, "Tree-based machine learning algorithms in the Internet of Things environment for multivariate flood status prediction," vol. 31, no. 1, pp. 1–14, 2022, doi: 10.1515/jisys-2021-0179.

[26] A. Habibi, M. Delavar, M. Sadeghian, and B. Nazari, "Flood Susceptibility Mapping and Assessment Using Regularized Random Forest and Naïve Bayes Algorithms," *ISPRS Annals of the Photogrammetry Remote Sensing and Spatial Information Sciences*, vol. X-4/W1-2022, pp. 241–248, Dec. 2023, doi:10.5194/isprs-annals-X-4-W1-2022-241-2023.

[27] H. Tang, H. Xu, X. Rui, X. Heng, and Y. Song, "The Identification and Analysis of the Centers of Geographical Public Opinions in Flood Disasters Based on Improved Naïve Bayes Network," *Int J Environ Res Public Health*, vol. 19, no. 17, 2022, doi:10.3390/ijerph191710809.

[28] C. Luu *et al.*, "Flood-prone area mapping using machine learning techniques: a case study of Quang Binh province, Vietnam," *Natural Hazards*, vol. 108, Dec. 2021, doi: 10.1007/s11069-021-04821-7.

[29] A. R. M. Towfiqul Islam *et al.*, "Flood susceptibility modelling using advanced ensemble machine learning models," *Geoscience Frontiers*, vol. 12, no. 3, p. 101075, 2021, doi: 10.1016/j.gsf.2020.09.006.

[30] W. Chen *et al.*, "Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles," *J Hydrol (Amst)*, vol. 575, pp. 864–873, 2019, doi: 10.1016/j.jhydrol.2019.05.089.

[31] A. Shirzadi *et al.*, "A novel ensemble learning based on Bayesian Belief Network coupled with an extreme learning machine for flash flood susceptibility mapping," *Eng Appl Artif Intell*, vol. 96, p. 103971, 2020, doi: 10.1016/j.engappai.2020.103971.

[32] M. Rahman *et al.*, "Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis," *Earth Systems and Environment*, vol. 3, no. 3, pp. 585–601, 2019, doi:10.1007/s41748-019-00123-y.

[33] D. Tien Bui *et al.*, "A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area," *Science of The Total Environment*, vol. 701, p. 134413, 2020, doi: 10.1016/j.scitotenv.2019.134413.

[34] M. Diakakis, G. Deligiannakis, A. Pallikarakis, and M. Skordoulis, "Factors controlling the spatial distribution of flash flooding in the complex environment of a metropolitan urban area. The case of Athens 2013 flash flood event," *International Journal of Disaster Risk Reduction*, vol. 18, pp. 171–180, 2016, doi:10.1016/j.ijdrr.2016.06.010.

[35] S. Huang, J. Xia, G. Wang, and J. Lei, "The impact of flood regime on river floodplain vegetation coverage: Insights from a 30-year Landsat record," *J Hydrol (Amst)*, vol. 626, p. 130355, 2023, doi:10.1016/j.jhydrol.2023.130355.

[36] S. Sugianto, A. Deli, E. Miswar, M. Rusdi, and M. Irham, "The Effect of Land Use and Land Cover Changes on Flood Occurrence in Teunom Watershed, Aceh Jaya," *Land (Basel)*, vol. 11, no. 8, 2022, doi:10.3390/land11081271.

[37] M. Panahi *et al.*, "Deep learning neural networks for spatially explicit prediction of flash flood probability," *Geoscience Frontiers*, vol. 12, no. 3, p. 101076, 2021, doi: 10.1016/j.gsf.2020.09.007.

[38] A. Hamlat, C. B. Kadri, A. Guidoum, and H. Bekkaye, "Flood hazard areas assessment at a regional scale in M'zi wadi basin, Algeria," *Journal of African Earth Sciences*, vol. 182, p. 104281, 2021, doi:10.1016/j.jafrearsci.2021.104281.

[39] M. Rahman *et al.*, "Flooding and its relationship with land cover change, population growth, and road density," *Geoscience Frontiers*, vol. 12, no. 6, p. 101224, Nov. 2021, doi:10.1016/J.GSF.2021.101224.

[40] M. A. Baig *et al.*, "Regression analysis of hydro-meteorological variables for climate change prediction: A case study of Chitral Basin, Hindukush region," *Science of The Total Environment*, vol. 793, p. 148595, 2021, doi: 10.1016/j.scitotenv.2021.148595.

[41] K. I. Sundus, B. H. Hammo, M. B. Al-Zoubi, and A. Al-Omari, "Solving the multicollinearity problem to improve the stability of machine learning algorithms applied to a fully annotated breast cancer dataset," *Inform Med Unlocked*, vol. 33, p. 101088, 2022, doi:10.1016/j.imu.2022.101088.

[42] J. Y.-L. Chan *et al.*, "Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review," *Mathematics*, vol. 10, no. 8, 2022, doi: 10.3390/math10081283.

[43] H. Farhadi, A. Esmaeily, and M. Najafzadeh, "Flood monitoring by integration of Remote Sensing technique and Multi-Criteria Decision Making method," *Comput Geosci*, vol. 160, p. 105045, Mar. 2022, doi:10.1016/J.CAGEO.2022.105045.

[44] G. Seeja, A. S. A. Doss, and V. B. Hency, "A Novel Approach for Disaster Victim Detection Under Debris Environments Using Decision Tree Algorithms With Deep Learning Features," *IEEE Access*, vol. 11, pp. 54760–54772, 2023, doi:10.1109/access.2023.3281461.

[45] H.-M. Lyu and Z.-Y. Yin, "Flood susceptibility prediction using tree-based machine learning models in the GBA," *Sustain Cities Soc*, vol. 97, p. 104744, 2023, doi: 10.1016/j.scs.2023.104744.

[46] M. Rasheed, SuhaShihab, O. Alabdali, and H. H. Hassan, "Parameters Extraction of a Single-Diode Model of Photovoltaic Cell Using False Position Iterative Method," *J Phys Conf Ser*, vol. 1879, no. 3, p. 032113, 2021, doi: 10.1088/1742-6596/1879/3/032113.

[47] M. Motta, M. de Castro Neto, and P. Sarmento, "A mixed approach for urban flood prediction using Machine Learning and GIS," *International Journal of Disaster Risk Reduction*, vol. 56, p. 102154, 2021, doi: 10.1016/j.ijdrr.2021.102154.

[48] M. Rasheed, O. Y. Mohammed, S. Shihab, and A. Al-Adili, "Explicit Numerical Model of Solar Cells to Determine Current and Voltage," *J Phys Conf Ser*, vol. 1795, no. 1, p. 012043, 2021, doi:10.1088/1742-6596/1795/1/012043.

[49] R. Jalal, S. Shihab, M. A. Alhadi, and M. Rasheed, "Spectral Numerical Algorithm for Solving Optimal Control Using Boubaker-Turki Operational Matrices," *J Phys Conf Ser*, vol. 1660, no. 1, p. 012090, 2020, doi: 10.1088/1742-6596/1660/1/012090.

[50] M. Rasheed, O. Alabdali, and S. Shihab, "A New Technique for Solar Cell Parameters Estimation of The Single-Diode Model," *J Phys Conf Ser*, vol. 1879, no. 3, p. 032120, 2021, doi: 10.1088/1742-6596/1879/3/032120.

[51] A. Lovakov and E. R. Agadullina, "Empirically derived guidelines for effect size interpretation in social psychology," *Eur J Soc Psychol*, vol. 51, no. 3, pp. 485–504, 2021, doi: 10.1002/ejsp.2752.

[52] M. Enneffatia, M. Rasheed, B. Louatia, K. Guidaraa, S. Shihab, and R. Barillé, "Investigation of structural, morphology, optical properties and electrical transport conduction of Li0.25Na0.75CdVO4 compound," *J Phys Conf Ser*, vol. 1795, no. 1, p. 012050, 2021, doi:10.1088/1742-6596/1795/1/012050.