International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V71I7P211

Volume 71 Issue 7, 105-116, July 2023

Original Article

The Local Analysis and Prediction System for Warning of Flash Floods Disaster in Thailand

Tepridht Phratep[1], Surachai Thongkhaew[2], Prasong Praneetpolgrang[3]

[1,2]School of Information Technology, Sripatum University, Bangkok, Thailand
[3]Academic Faculty, Navaminda Kasatriyadhiraj Royal Air Force Academy, Thailand

[1]Corresponding Author : hara.german@gmail.com

Received: 03 May 2023 Revised: 20 June 2023 Accepted: 24 June 2023 Published: 21 July 2023

Abstract - *The purposes of this research are 1) to develop a predictive model for alerting against flash floods caused by forest runoff in Thailand using Artificial Intelligence and 2) to present an analysis and prediction system for alerting against flash floods caused by forest runoff in Thailand using Artificial Intelligence. This study developed a model from a dataset of monthly rainfall amounts from 2018-2021 using the Neural Networks Regression algorithm. The statistical analysis found that the highest value was achieved when the Learning Weight was set to 0.1, and the Learning Rate was 0.009, with an accuracy of 99.09%. This research demonstrates the potential for developing predictive models for alerting flash floods caused by forest runoff, using neural networks that can be applied in designing and developing alert systems on various platforms to help alert and reduce losses of life and property of the people.*

Keywords - *Disaster alert system, Artificial intelligence, Prediction model, Neural network regression*.

1. Introduction

Floods are natural disasters that occur due to water and have significant impacts on society and the economy. In some cases, floods can be caused by natural factors such as prolonged heavy rainfall, which affects humans, for example, encroachment in forested areas, dam construction, and inappropriate environmental systems.

Floods caused by flash floods from forest runoff are another cause of flooding that occurs due to water flowing down from high areas to low-lying areas. The result is that the water flows faster and stronger than usual, which frequently happens during the rainy season when there is a significant volume of water in the forest. In Thailand, forest water flow disasters frequently occur in various areas, causing damage to the environment, as well as the lives and properties of people, from small losses to the loss of life, depending on their severity and nature. Additionally, it also has varying degrees of psychological impact. Each disaster that occurs simultaneously results in the loss of both property and various activities such as the economy, society, agriculture, industry, public utilities, transportation, and politics, which cannot be carried out normally. It must come to a halt, transportation is disrupted, and transportation of goods is interrupted. The price of medical supplies and consumer goods increases. In addition, 80% of Thailand's population is engaged in agriculture. Agricultural areas have been destroyed, leading to damaged crops and low productivity for farmers. This results in a lack of quality and income and increased costs, time, and labor for farmers.

The impact of disasters in Thailand has resulted in significant damage to life and property, with the severity increasing due to unpredictable weather conditions. However, the current warning system utilizing modern technology remains inadequate. Various techniques and methods are being developed to address these challenges in managing natural disasters. Technology plays a crucial role in creating a secure environment and facilitating emergency preparedness. Through different channels such as loud warnings, SMS messages, smartphone applications, wireless technology, and social media, disaster warning technology aims to promptly alert individuals in high-risk areas. This allows them to prepare, take preventive measures, or evacuate from dangerous zones.

There are several research studies, such as [1] studying water level measurement in rivers to propose a water level notification system and the speed of water using wireless networks. Research [2] uses a calculation method from the amount of rainfall from infrared satellite images and rainfall data that exceed 50 millimeters per day to analyze and create a model in hydrology for monitoring and alerting. Research[3] presents a methodology that utilizes meteorological and hydrological data to classify flooded areas and provides alerts through a web application. Research[4] introduces the development of a prototype water level notification system, incorporating geographic maps and the Internet of Things (IoT) technology for monitoring water sources. In some research studies [5], geographic information systems are applied to analyze the development of early warning systems for landslides. In research [6], neural networks are used to predict water levels, and in research [7], artificial intelligence is applied to create a flood warning system in combination with sensor work. Research [8] develops a predictive water-level model using various artificial intelligence models. Therefore, it can be seen that disaster warning prediction for hydrological disasters has various methods, depending on the environment and other factors, and also has different levels of accuracy.

From the background and problems mentioned previously, utilizing advanced techniques, particularly artificial intelligence technology, for disaster management is an efficient and accurate tool. Therefore, this research focuses on developing a predictive model for alerting forest flood disasters in Thailand using artificial intelligence, and presenting an analysis and prediction system for alerting forest flood disasters in Thailand using artificial intelligence as a guideline for utilizing artificial intelligence technology in disaster management, by analyzing and processing rainfall data, disaster management personnel can efficiently manage and reduce the risk of disasters, thus increasing their ability to prevent or mitigate future disasters.

2. Literature Review

2.1. Artificial Intelligence (AI)

AI denotes the utilization of cognitive machine capabilities to execute tasks according to predefined conditions. The word "intelligence" means the ability to think independently, and "artificial" refers to things created by humans. Therefore, AI is defined as an intellectual capability created by humans. AI is an extremely effective tool for performing specialized tasks, and it can also help computers make correct decisions and improve their operational efficiency [11]. Furthermore, AI helps people work more efficiently and become more proficient, but it is necessary to develop its abilities and create new things, such as technology, emotional skills, social skills, and creativity [13].

AI encompasses various domains such as Natural Language Processing, Computer Vision, Robotics, and Expert Systems, which involve logical and sequential thinking processes involving both computers and humans. At present, these fields are rapidly advancing and evolving. AI is being used to increase efficiency in many professions, such as medicine, industry, agriculture, and business. Recent studies have utilized AI for predicting natural disasters, specifically exploring the correlation between AI, ML, and DL, as depicted in Figure 1.

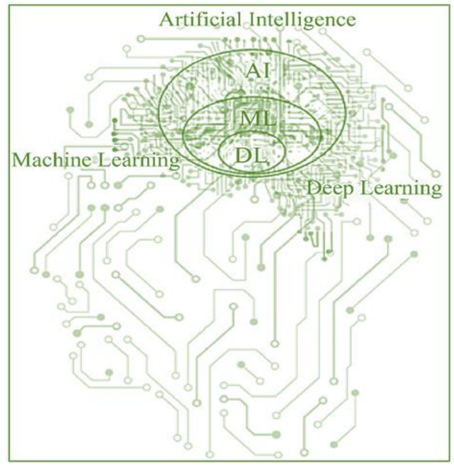

Fig. 1 The Relationship between Artificial Intelligence and Machine Learning

2.2. Machine Learning (ML)

ML is the use of AI that has the ability to learn automatically and provide accurate and appropriate results from learning experiences. It uses calculations and grouping techniques that exist in data sets and development programs to provide appropriate answers to problems and to improve efficiency in data separation. The most well-known learning method begins with basic data, learning from data and decision-making developed based on models. Computers can learn without the need for experts to help or exchange with humans, providing more accurate and faster results by analyzing data through classification techniques such as supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [15].

2.2.1. Supervised Learning

It is an ML technique that utilizes a dataset containing labeled examples to build a predictive model capable of predicting outcomes for new data. Each training example in the dataset is associated with a desired outcome, and during the model creation process, learning algorithms are employed to determine these outcomes. The model's parameters are adjusted through a learning process, aiming to minimize the discrepancy between the predicted and actual outcomes. Supervised learning is widely used in tasks involving classification [15].

2.2.2. Semi-Supervised Learning

It is an iterative learning process that leverages both labeled and unlabeled data to enhance the performance of supervised learning. By incorporating the additional unlabeled data, the model can gain a better understanding of the underlying patterns and structure in the dataset, leading to improved predictive capabilities. This approach takes advantage of the availability of large amounts of unlabeled data, which may be easier to obtain compared to labeled data. It effectively utilizes it to enhance the overall learning process.

2.2.3. Unsupervised Learning (UL)

UL is a training and learning process that operates without labeled data, relying on a set of examples and starting from an initial random state. It aims to improve the performance of clustering algorithms by identifying patterns and relationships based on the similarity between training examples. Unsupervised learning algorithms are commonly employed to group data and uncover hidden structures or clusters within the dataset. By autonomously discovering patterns, unsupervised learning provides valuable insights and can assist in various data analysis tasks.

2.2.4. Reinforcement Learning (RL)

It is a type of ML involving a feedback system to adjust a machine's behavior to fit the environment. It receives input from the external environment and must choose an action to generate the best result. The environment evaluates these actions, leading to rewards or penalties that serve as performance indicators for the machine. The machine learns by experimenting and adjusting parameters to minimize the reward or punishment and determine the best course of action in that particular environment.

2.3 Artificial Neural Networks (ANNs)

ANNs are a combination of nodes used for simple processing, serving as the cells of living organisms. The processing capability of the network is embedded within the connections or weights derived from the adaptation or learning process using a set of training patterns. Artificial neural networks are extensively utilized for statistical analysis and data modeling, serving as an alternative to linear regression or classification techniques. Therefore, artificial neural networks are used for classification or prediction and are applied in each professional field according to their expertise.

The artificial neural network is in a parallel distributed computing format. The necessary requirement for an artificial neural network is the ability to detect important data processing properties of the network, which are actual properties that correspond to each other. It can be seen that the foundation of the artificial neural network is the Threshold Logic Unit (TLU), which is a mechanism for weighing input and output data that equals “1”, and if the result is otherwise, it equals “0”. The TLU is the mechanism integrated from the basics of real neural cells.

The learning of ANNs operates on the principle of considering examples. In general, programs are not rigidly defined by rules. The operation depends on a set of interconnected units or nodes called artificial neurons, which simulate neurons in the biological brain in a crude manner. The utilization of "signals" in the connections involves real numbers, and the output of each neuron is determined through a non-linear function applied to the sum of inputs. Connections are called edges, and the weights of the edges and neurons are usually adjusted through learning. An artificial neural network typically has a few thousand to several million units and is connected between input and output layers through several layers. More layers create deeper technology, referred to as “Deep Neural Networks (DNN)”. DNNs are a part of a wide range of ML methods based on ANNs with learning capabilities. Deep Learning technology such as Deep Neural Networks, Recurrent Neural Networks, and Convolutional Neural Networks have been used in various fields, achieving comparable and sometimes superior performance to human experts.

Recurrent neural networks (RNN) are artificial neural networks designed for sequential data, using the principle of the internal state within the model to provide feedback as new input paired with normal input to help the model learn the pattern of the input sequence and make further predictions. Gated Recurrent Unit (GRU) and Long Short Term Memory (LSTM) were designed to overcome gradient-related problems. GRU demonstrated superior performance in small datasets due to its fewer parameters, resulting in faster and easier training.

The Convolutional Neural Network (CNN) ANN emulates human visual perception by partitioning an image into smaller regions, then merging them and analyzing them using mathematical computations that align with the concept of Spatial Convolution in image processing. It extracts features of objects by using multiple filters combined together.

The K-Nearest Neighbor Algorithm (KNN) is a classification technique that decides which class represents certain conditions or new cases by checking a certain number of the nearest points or conditions most similar to given attributes. It calculates the distance between each variable in the data and outputs a value (suitable for numerical data and non-continuous variables but requires special handling).

The Levenberg-Marquardt (LM) algorithm is an iterative method used to minimize the value of a multi-variable function represented as the sum of squares of non-linear functions [4, 6]. It has emerged as a widely adopted method for solving non-linear least squares problems, becoming a standard technique in the field [7]; in various domains, the LM algorithm is commonly encountered. This algorithm can be seen as a combination of the steepest descent method and the Gauss-Newton method.

When the current solution deviates significantly from the correct solution, the algorithm behaves similarly to the steepest descent method. However, as the current solution approaches the correct solution, the LM algorithm switches from updating parameters by gradient descent to updating them by the Gauss-Newton method, as shown in equation (1).

$$(J^T J + \mu I) h_{lm} = -g \tag{1}$$

Where

$g = J^T f \; and \; \mu \geq 0$

μ is the damping parameter (Learning Factor)

h_{lm} is descent direction

I is Identity Matrix

In the case μ with a large value, h_{lm} is obtained as Equation (2)

$$h_{lm} \cong -\frac{1}{\mu} g = -\frac{1}{\mu} F'(x) \tag{2}$$

Therefore μ is small, $h_{lm} \cong h_{gn}$

Where

h_{gn} is a Gauss-Newton step as in Equation (3)

$$h_{gn} = \frac{-J^T f(x)}{J^T J} \tag{3}$$

x is the input data combined with the deviation (Bias)

$F'(x)$ is the rate of change function with respect to the variable x

$f(x)$ is a function of the variable

SVM is used for classification and regression tasks in supervised learning. It leverages ML principles to enhance prediction accuracy and mitigate the risk of overfitting. It is commonly used in facial recognition, image and text classification, automatic driving, chatbots, and more [16]. SVM regression analysis is suitable for datasets with a large number of properties and has the ability to classify both linear and nonlinear data.

2.4. Microsoft Azure Machine Learning (MAML)

MAML is a cloud-based service that enables developers to create, deploy, and manage machine learning applications using the Microsoft data center network. It is a flexible service for changing various conditions of operations, supporting various regression, grouping, and categorization algorithms. In addition, users can customize the model with Python and R languages. With Azure ML, users can drag and drop modules and datasets, link them to create models, and easily and quickly adjust other parameter values [17].

2.5. Regression Analysis (RA)

RA is a statistical method utilized to examine the relationship among multiple variables. It consists of variables called predictors or independent variables (X) and variables that want to know, called response or dependent variables (Y), to determine whether they are factors or causes of each other or not. RA used in data analysis [24] can be divided into 3 types, namely linear regression analysis, polynomial regression analysis, and logistic regression analysis.

2.6. Linear Regression Analysis (LRA)

It explores the linear relationship between variables [24]. If it involves one independent variable (X) and one dependent variable (Y), it is called simple linear regression analysis. For two independent variables (X) and one dependent variable (Y), it is referred to as multiple linear regression analysis.

2.6.1. Simple Linear Regression Analysis (SLRA)

SLRA explores the connection between a single independent variable and a dependent variable, similar to correlation analysis. However, unlike correlation analysis, SLRA determines the independent variable (the cause) and the dependent variable (the effect). In addition to understanding their relationship, SLRA can use the independent variable's value to predict or estimate the dependent variable, indicating its predictive or explanatory capability.

This method involves linear regression analysis with a single predictor variable (X) and a single response variable (Y), representing their relationship through a mathematical equation (4).

$$Y = \beta_0 + \beta_1 X \tag{4}$$

Where

Y represents the output data.

β_0 represents the constant of the regression equation, which is the intercept value of the Y-axis of the equation

$\beta_1 X$ Regression Coefficient of Respondent X

2.6.2. Multiple Linear Regression Analysis (MLRA)

MLRA is a statistical method that examines the linear relationship between multiple independent variables and a dependent variable. It extends the concept of Simple Linear Regression Analysis, which focuses on the relationship between a single independent variable and a dependent variable. In Multiple Linear Regression Analysis, the relationship between the independent variables (X) and the dependent variable (Y) remains linear, but the analysis incorporates more than one independent variable. This relationship is mathematically expressed by equation (5).

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k + \in \tag{5}$$

Where

Y is the number of independent variables in the regression equation.

k is the value of each independent variable (where *i = 1, 2....k*).

x_i is the value of each independent variable (where *i = 1, 2....k*).

b_0 is the constant of the regression equation.

b_i is the regression coefficient (where *i = 1, 2,…,k*) where b_i represents the rate of change of x_i to *Y*.

$\in$ is the deviation between the x and then *Y*

2.7. Model Performance Measurement

ML techniques utilize various metrics to evaluate the accuracy of different models. When it comes to regression models, one widely used metric is the Mean Squared Error (MSE). It measures the model's accuracy by calculating the average squared errors. A lower MSE value indicates a more accurate model.

To compute the MSE, the squared difference between predicted and actual values is averaged. Every prediction model has some degree of error, and the accuracy of the predicted value depends on the degree of error. The degree of error can be calculated using the following formula [25]. Mean Square Error (MSE) is the average of squared errors, as shown in equation (6).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (6)$$

Where

y_i is the actual data value
$\hat{y}_i$ is the predicted outcome value
n is the total number of data

Root Mean Square Error (RMSE), or the root of the mean square error, is a standard error measurement. Widely used, a small value indicates that the model is very accurate. The calculation method takes the calculated MSE value to the square root, as shown in equation (7).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (7)$$

Where

n is the amount of data used
y_i is the actual value at any t
$\hat{y}_i$ is the predicted value at any t

The Coefficient of Determination, or R-square (R2), is a value used to indicate whether a model is appropriate or not. It ranges from 0 to 1, with a value closer to 1 indicating higher accuracy. Generally, a value greater than 0.6 is considered a good model.

2.8. Literature Review

In [18], a new flood forecasting model utilizing an Artificial Neural Network (ANN) was developed and evaluated using a case study of the Kosi and Jhulum rivers in the Koshi basin. These rivers experienced a major flood in September 2014, with a discharge exceeding 115,000 cubic meters per second due to atmospheric conditions. The study compared five different ANN models, including the Bayesian Neural Network, Levenberg-Marquardt Neural Network, Conjugate Gradient Descent Neural Network, Resilient Backpropagation Neural Network, and Flexible Distributed Neural Network. Results showed that the Levenberg-Marquardt Neural Network achieved the lowest average squared error.

In 2022, research [19] proposed a framework that uses the multimodal approach of Deep Learning (DL) to analyze disaster situations, consisting of VGG-19, CNN, and LSTM models, using an automated data weighting method instead of manually adjusting the weight of data. The study found that linking different data and ideas efficiently can be achieved using a large dataset from Twitter to test and detect the level of damage caused by disasters. The model was able to learn multiple tasks simultaneously, and the performance of using the standard data weighting method was better than the model using the task-specific data weighting method.

Research [12] develops a natural disaster risk map in Thailand using knowledge and information from the fields of engineering, information technology, and geography. The research analyzes the hierarchy of risks using the Analytical Hierarchy Process (AHP) to create a map of natural disaster risks such as floods, landslides, wildfires, and cyclones. This map serves as a prototype for a natural disaster warning system based on risk maps at the regional level. The study identifies the factors that cause each type of natural disaster and assigns weights to the data based on risk analysis principles.

Research [13] develops an early warning system for natural disasters using machine learning technology. The study investigates heavy rainfall patterns over short periods of time using the Logistic Regression (LR) classifier technique to build a predictive model. The results show that the LR technique is more effective than other classifiers.

In research [14], a natural disaster risk map is developed in Thailand using knowledge and information from the fields of engineering, information technology, and geography; research analyzes the hierarchy of risks using the Analytical Hierarchy Process (AHP) to create a map of natural disaster risks such as floods, landslides, wildfires, and cyclones. This map serves as a prototype for a natural disaster warning system based on risk maps at the regional level. The study identifies the factors that cause each type of natural disaster and assigns weights to the data based on risk analysis principles.

3. Materials and Methods

In this research, the researchers presented a supervised algorithm of the neural network model with backpropagation from the rainfall data during the period from 2018 to 2021 to predict the occurrence of floods. The process of building the research model consists of 8 steps, as shown in Figure 2.

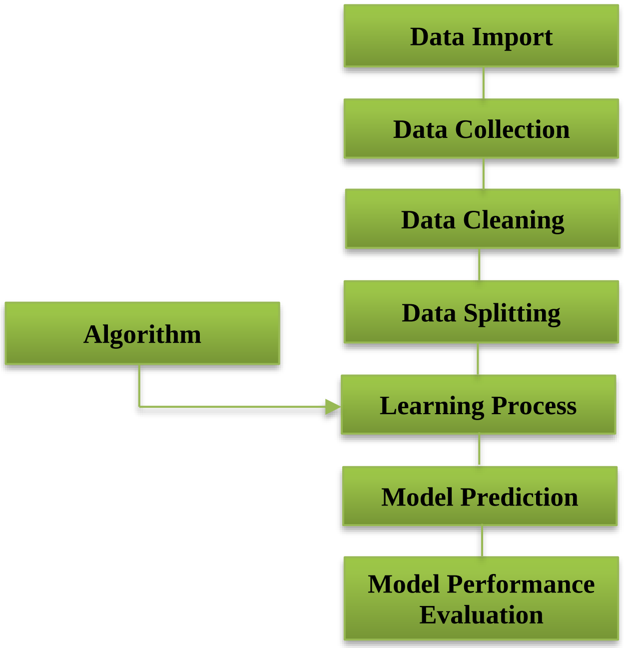

Fig. 2 Research process steps

From Figure 2, it can be explained that the steps of this research are as follows:

1. Data Import
This step is importing the Dataset used for machine learning in this research, which uses data in the form of a CSV format.

2. Data Understanding
This is the selection of factors that affect flooding. In this research, the average amount of rainfall from 50 cubic millimeters was used.

3. Data Cleaning
This is checking the data to ensure that it is in a usable format, such as verifying the correctness and completeness of the data.

4. Data Splitting
This is dividing the data for use in learning (model creation) and testing the model. This research divides the data into 80% for model creation and 20% for testing the model.

5. Model Selection
This is the selection of the algorithm for teaching ML. The Neural Network Regression algorithm is used in this research, and the various parameter settings are adjusted.

6. Model Learning
This is teaching the machine from the dataset that has been split into 80% using the learning techniques with the algorithm selected in step 5.

7. Model Prediction
This step evaluates the model learned by the machine by using the model obtained to evaluate the dataset split in the 20% section.

8. Model Performance Measurement
This measures the model's performance using statistical measures, including Mean Absolute Error (MAE).

In this research, the tool used to build the model is Microsoft Machine Learning Studio, as shown in Figure 3.

3.1. Data Importing Process

The data importing process in this research involves using a .csv file to import rainfall data from the months of 2018-2021. Microsoft Machine Learning Studio is utilized for model creation, and the .csv file comprises columns like year, month, minimum rainfall, maximum rainfall, and average rainfall. This is done so the machine can learn from the imported data and create a predictive model for forest flood warnings.

3.2. Data Collection

For the data collection process, the researcher followed these steps. The data source used for this research is rainfall data from the years 2018-2021, obtained from a sample dataset from the website: https://doc-0k-7g-docs.googleusercontent.com/docs/securesc/.../8dfcftnt483, as shown in the example in Figure 4.

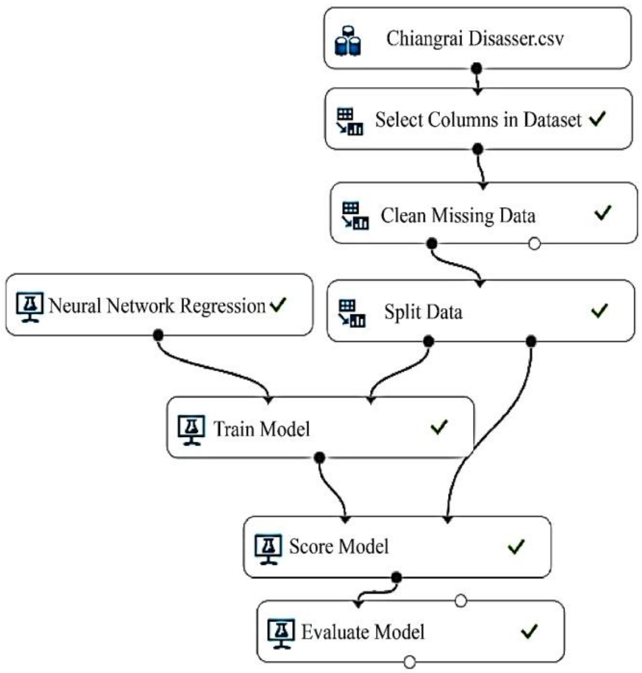

Fig. 3 shows the diagram of building a model using Microsoft Machine Learning Studio.

Year	2018			2019			2020			2021		
Month	MinRain	MaxRain	AvgRain	MinRain	MaxRain	AvgRain	MinRain	MaxRain	AvgRain	MinRain	MaxRain	AvgRain
January	4.769999981	16.86000061	8.776122468	34.25999832	67.79000092	49.65644778	0	0.01	0.000115822	3.799999952	19.38999939	9.796027796
February	3.789999962	20.5	10.91876121	0	3.720000029	0.657472051	0	0.02	0.002810958	20.52000046	68.19999695	59.01568839
March	13.93000031	35.59000015	26.05649612	0	18.43000031	4.218928392	0.800000012	11.02000046	3.938145837	0.200000003	10.43000031	2.8434082
April	120.2900009	195.3800049	162.1420708	13.5	32.74000168	18.50689798	60.79999924	123.4800034	95.25486555	83.23999786	167.6000061	126.4359462
May	267.7000122	515.5300293	409.6562403	89	162.9900055	122.6746611	60.24000168	124.75	90.86005336	64.27999878	146.3600006	120.0839349
June	112.5299988	246.9900055	202.5148927	40.18999863	96.91999817	57.65492295	75.41000366	219.0899963	170.0714916	71.48000336	219.7899933	192.9061668
July	166.8399963	274.0400085	226.2426045	103.8499985	210.2899933	188.9306375	70.83000183	219.9799957	170.4094067	163.4299927	295.2799988	243.8856108
August	156.3999939	301.8999939	274.5151788	277.6900024	439.8900146	336.7843588	197.3200073	397.2900085	354.4223849	120.1200027	415.5400085	312.8321422
September	104.6500015	245.8699951	194.6696787	63.20000076	175.0800018	86.31027397	101.7799988	242.4900055	200.7220143	169.5599976	311.6400146	247.3365607
October	133.3099976	269.9899902	194.0708188	25.31999969	96.01999664	59.55241714	25.11000061	86.25	37.71434082	107.9700012	230.8699951	195.0206345
November	8.180000305	54.88999939	37.94283411	9.229999542	39.77999878	22.70900997	13.85999966	50.58000183	26.31717393	11.69999981	77.79000092	54.5141283
December	19.78000069	45.20000076	34.87195488	10	19.20000076	16.28437909	0	0	0	0	4.75	0.934543257

Fig. 4 Dataset of rainfall amounts in Chiang Rai province

3.3. Data Cleansing

Data cleansing is the process of managing raw data to make it usable for machine learning (ML) modeling. If the data contains errors, it can affect the model's accuracy. This process is shown in Figure 5.

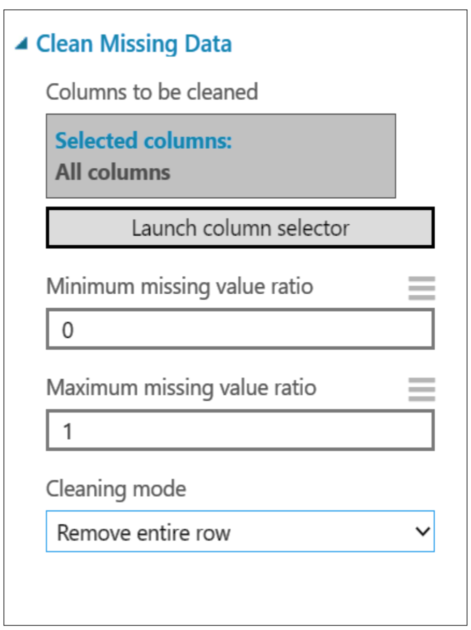

Fig. 5 Data cleansing condition setting

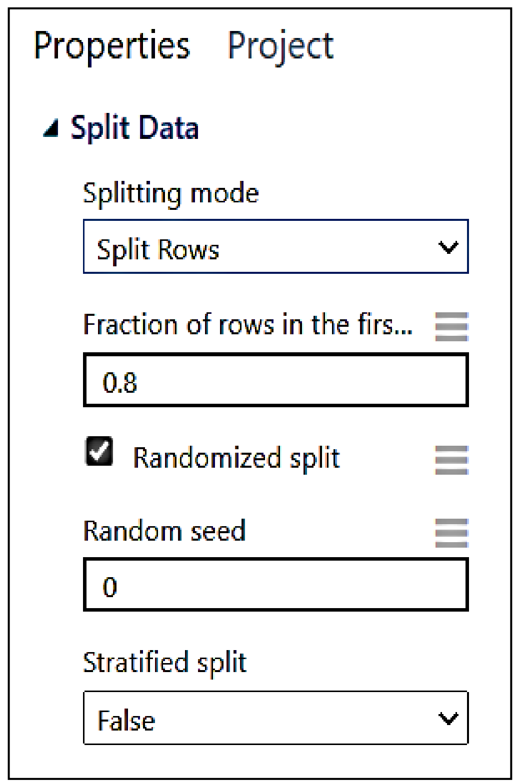

Fig. 6 Data splitting for model training

3.4. Data Splitting

This step involves splitting the dataset into two parts: one for training the model and the other for testing its performance. In this study, the researcher split the data into 80% for training and 20% for testing, as shown in Figure 6.

3.5. Model Selection

To define a model, the researcher selected a method for creating a multi-class classification model using neural network algorithms for training the data, as shown in Figure 7.

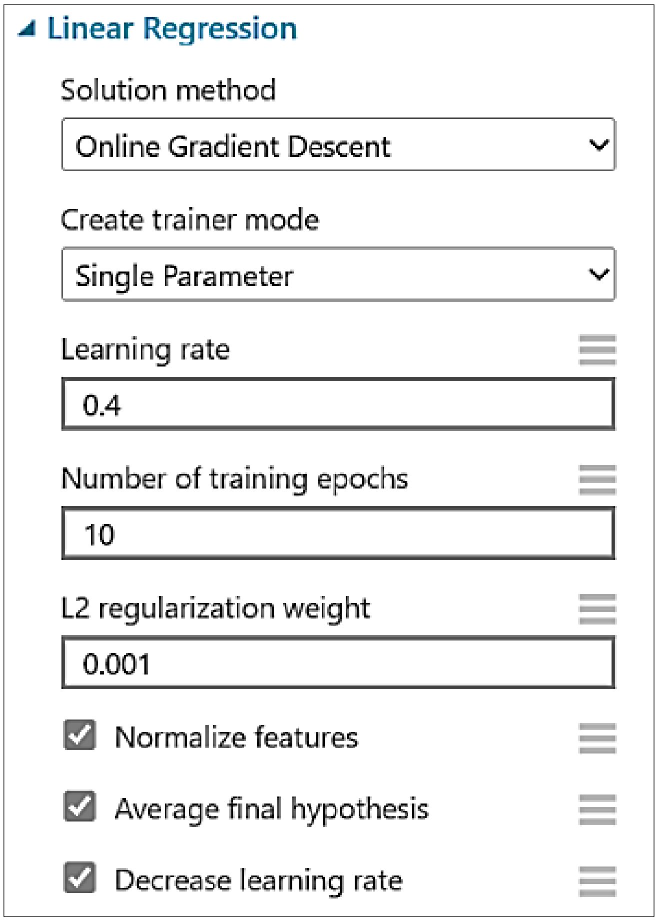

Fig. 7 Algorithm selection for model building

3.6. Training the Model

In the process of training the model, the researcher sets the data label that he wants to analyze for the occurrence of floods. In this research, the researcher sets the STATUS_WARNING column to predict the likelihood of flooding, as shown in Figure 8.

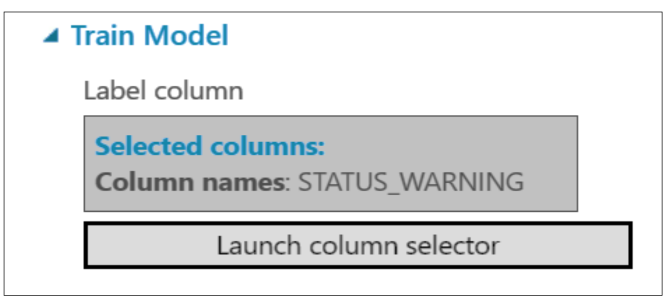

Fig. 8 Model selection

This step involves teaching the model, which consists of 2 parts: 1) selecting an algorithm to create the model and 2) importing data. In this study, the researcher chose the Neural Network Regression algorithm. In the process of creating the model, the researcher selected the STATUS_WARNING column to predict flood occurrence, with 3 steps as follows:

1. Algorithm selection for model training

During this phase, the researcher chose Neural Network Regression as the algorithm and specified the model creation parameters as follows: Trainer Mode: Single Parameter, Hidden Layer: Fully Connected, with 100 Hidden Nodes, Learning Rate: 0.001, and Learning Weights: 0.1.

2. Selecting data for teaching and learning

This step is a part of importing data, where 80% of the data was used to create the model.

3. Adjusting the Learning Rate and Learning Weights.

3.7. Model Prediction Process

This process is a step to test the model's accuracy created from the Training Model process in step 3.5. The model is compared with test data to find the model's accuracy created for predicting results or answers to the data. It uses existing data for learning and creating the model. Then, the created model is used to predict the results. The model will calculate and predict the price of a house per square meter based on given data and return the result or answer to the user.

3.8. Model Performance Evaluation Process

This process is a step for evaluating the model's performance by finding statistical values, including Mean Absolute Error, Root Mean Squared Error, Relative Absolute Error, Relative Squared Error, and Coefficient of Determination. Once a model with sufficient performance has been obtained, it is used to predict results with new data that has not been learned before.

4. Results and Discussion

In this research, the results and findings can be explained as follows:

4.1. Overview of the Model

From this research, it is possible to create a diagram illustrating the model, as shown in Figure 9.

4.2. Dataset

In this research, the researchers used a rainfall dataset from 2018 to 2021, as shown in Figure 10.

4.3. The Results of Comparing the Model with the Test Dataset

The comparison of the model with the test dataset is shown in the STATUS_WARNING column compared to the Scored Labels, which have high accuracy and precision. The statistics for predicting the chance of no flood occurrence, Scored Probabilities for Class "NO," and Scored Probabilities for Class "YES," are shown in Figure 11.

4.4. The results of measuring the performance of the model

The accuracy of the model's predictions in machine learning was assessed to evaluate its performance.

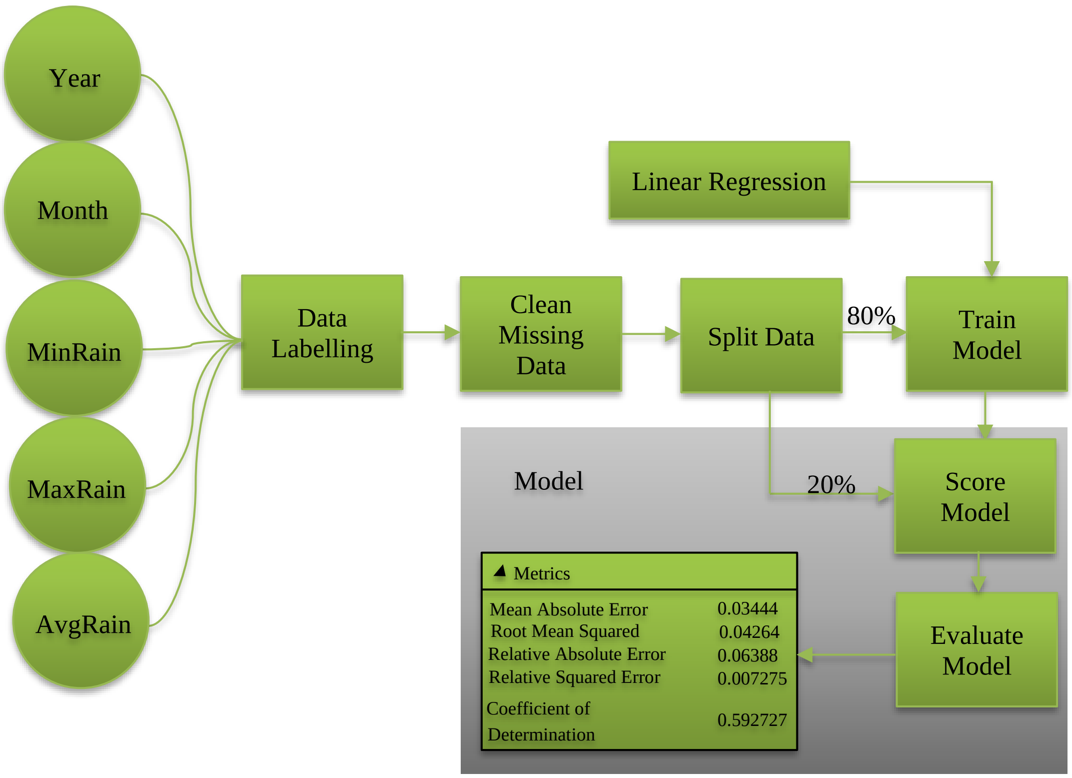

Fig. 9 Overview of the model

	YEAR	MONTH	MinRain	MaxRain	AvgRain	STATUS	STATUS_WARNING
view as							
	2018	JANUARY	4.77	16.860001	8.776122	NORMAL	0
	2018	FEBRUARY	3.79	20.5	10.918761	NORMAL	0
	2018	MARCH	13.93	35.59	26.056496	NORMAL	0
	2018	APRIL	120.290001	195.380005	162.142071	WARNING	1
	2018	MAY	267.700012	515.530029	409.65624	WARNING	1
	2018	JUNE	112.529999	246.990005	202.514893	WARNING	1
	2018	JULY	166.839996	274.040008	226.242604	WARNING	1
	2018	AUGUST	156.399994	301.899994	274.515179	WARNING	1
	2018	SEPTEMBER	104.650002	245.869995	194.669679	WARNING	1
	2018	OCTOBER	133.309998	269.98999	194.070819	WARNING	1
	2018	NOVEMBER	8.18	54.889999	37.942834	NORMAL	0
	2018	DECEMBER	19.780001	45.200001	34.871955	NORMAL	0

Fig. 10 Graph showing the dataset of rainfall volume

YEAR	MONTH	MinRain	MaxRain	AvgRain	STATUS	STATUS_WARNING	Scored Labels
2018	JULY	166.839996	274.040008	226.242604	WARNING	1	0.931103
2018	MARCH	13.93	35.59	26.056496	NORMAL	0	-0.034397
2020	MARCH	0.8	11.02	3.938146	NORMAL	0	0.013356
2021	MAY	64.279999	146.360001	120.083935	WARNING	1	0.981241
2019	DECEMBER	10	19.200001	16.284379	NORMAL	0	-0.032371
2019	FEBRUARY	0	3.72	0.657472	NORMAL	0	0.007714
2019	AUGUST	277.690002	439.890015	336.784359	WARNING	1	0.991141
2021	FEBRUARY	20.52	68.199997	59.015688	WARNING	1	0.983013
2021	JANUARY	3.8	19.389999	9.796028	NORMAL	0	0.074122
2018	JUNE	112.529999	246.990005	202.514893	WARNING	1	0.93106

Fig. 11 Comparison of the model prediction with the test dataset

Table 1. Shows the results of measuring the performance when the learning weight was Set to 0.1, and the learning rate was set between 0.001 to 0.010

LR	MAE	RMSE	RAE	RSE	Coef.
0.001	0.451536	0.46719	0.903071	0.873064	0.126936
0.002	0.368134	0.387092	0.736268	0.599361	0.400639
0.003	0.160041	0.174658	0.320082	0.122022	0.877978
0.004	0.071047	0.083607	0.083607	0.027961	0.972039
0.005	0.047553	0.075402	0.095106	0.022742	0.977258
0.006	0.041348	0.065596	0.082696	0.017211	0.982789
0.007	0.039589	0.062936	0.079177	0.015844	0.984156
0.008	0.038102	0.052398	0.076205	0.010982	0.989018
0.009	0.032125	0.047513	0.064251	0.00903	0.99097
0.010	0.039204	0.051663	0.078409	0.010676	0.989324

Table 2. Shows the performance measurement results when setting the learning weight at 0.2 and learning rate between 0.001-0.010.

LR	MAE	RMSE	RAE	RSE	Coef.
0.001	0.381085	0.394416	0.762169	0.622255	0.377745
0.002	0.189008	0.206141	0.378017	0.169976	0.830024
0.003	0.068205	0.084874	0.136411	0.028814	0.971186
0.004	0.055149	0.076454	0.110299	0.023381	0.976619
0.005	0.051675	0.07371	0.10335	0.021733	0.978267
0.006	0.041787	0.062475	0.083574	0.015613	0.984387
0.007	0.03289	0.051137	0.06578	0.01046	0.98954
0.008	0.042657	0.056386	0.085314	0.012717	0.987283
0.009	0.035718	0.054579	0.071437	0.011915	0.988085
0.010	0.035065	0.047965	0.07013	0.009203	0.990797

From Table 1, it shows the performance of the statistical analysis model when applied to real data, demonstrating the relationship between the independent and dependent variables in the model. It was found that the highest value was obtained when setting the Learning Weight to 0.1 and the Learning Rate to 0.009, with an accuracy of 99.09%.

From Table 2, it shows the performance measurement of the statistical analysis model when used with real data by demonstrating the relationship between the independent and dependent variables in the statistical analysis model. It was found that the highest value was obtained when adjusting the Learning Weight to 0.2 and the Learning Rate to 0.010, with an accuracy of 99.07%

4.5. System Design and Prediction Results for Alerting Forest Flood Disasters in Thailand using Artificial Intelligence

From this research, the study results and the high-accuracy model generated was utilized to design a system for analyzing and predicting forest flood disasters in Thailand using artificial intelligence for alerting purposes, as shown in Figure 12.

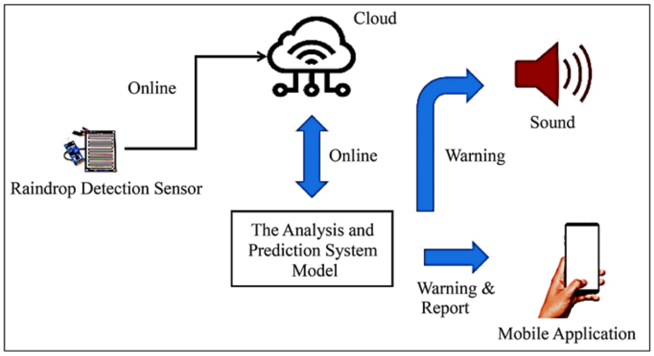

Fig. 12 The results of designing the analysis and prediction system for alerting forest flash flood disasters in Thailand using artificial intelligence

From Fig. 12, the principle of operation of the analysis and prediction system for notifying forest flash flood disasters in Thailand using artificial intelligence can be explained as follows:

Part 1 is an Internet of Things system component that checks rainfall and sends data to the cloud.

Part 2 is the cloud component that stores data, processes data, and stores the program command sets that provide services to users in the form of web services.

Part 3 is the mobile application component that monitors data and receives disaster notifications.

Part 4 is the distributed sound system component. If a disaster notification occurs, the system automatically sends a sound signal to the distributed sound system.

5. Conclusion

Based on the research results, it can be concluded that the purpose of this research is to create a predictive model for notifying forest flash flood disasters in Thailand using artificial intelligence and to present an analysis and prediction system for notifying forest flash flood disasters in Thailand using artificial intelligence. The model was created using a dataset of monthly rainfall amounts from the years 2018-2021, using a data splitting method for training (80%) and testing the model (20%) with the Neural Networks Regression algorithm. The statistical analysis model showed the highest accuracy when the Learning Weight was set to 0.1, and the Learning Rate was set to 0.009, with an accuracy of 99.09%. After obtaining a model with sufficient accuracy, it was used to design a forest flash flood warning system consisting of three main parts: 1) a rainfall measurement system, 2) a cloud computing system, and 3) an application, as well as 4) alert notification through a sound distribution system. This proposed system can continuously monitor and alert users. This system will effectively reduce the loss of life and property.

References

[1] Jong-Uk Lee et al., “RFMS: Real-Time Flood Monitoring System with Wireless Sensor Networks,” *The 5th IEEE International Conference on Mobile Ad-Hoc and Sensor Systems(MASS 2008)*, pp. 527-528, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[2] Kwabena O. Asante et al., “Developing a Flood Monitoring System From Remotely Sensed Data for the Limpopo Basin,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1709-1714, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[3] Ponthip Limlahapun, and Hiromichi Fukui, “Flood Monitoring and Early Warning System Integrating Object Extraction Tool on Web-Based,” *ICROS-SICE International Joint Conference*, pp. 2639-2643, 2009. [Google Scholar] [Publisher Link]

[4] Sathit Banmai et al., “Development a Prototype of Water Level Alert System on Geographic Map Combined with the Internet of Things,” *The 14th NPRU National Academic Conference Nakhon Pathom Rajabhat University*, 2022. [Google Scholar] [Publisher Link]

[5] Rokhmat Hidayat et al., “Development of a Landslide Early Warning System in Indonesia,” *Geosciences*, vol. 9, no. 10, pp. 1-17, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Ragil Andika Yuniawan et al., “Revised Rainfall Threshold in the Indonesian Landslide Early Warning System,” *Geosciences*, vol. 12, no. 3, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Aniessa Rinny Asnaning, and Septafiansyah Dwi Putra, “Flood Early Warning System Using Cognitive Artificial Intelligence: the Design of AWLR Sensor,” *In 2018 International Conference on Information Technology Systems and Innovation (ICITSI)*, pp. 165-170, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[8] Shuyu Zhang et al., “Short-Term Water Level Prediction Using Different Artificial Intelligent Models,” *In 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, pp. 1-6, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[9] Pongsakorn Limna, “Artificial Intelligence (AI) In the Hospitality Industry: a Review Article,” *International Journal of Computing Sciences Research*, vol. 7, no. 7, pp. 1306-1317, 2022. [Google Scholar] [Publisher Link]

[10] Cheng-Xiang Wang et al., “Artificial Intelligence Enabled Wireless Networking for 5G and Beyond: Recent Advances and Future Challenges,” *IEEE Wireless Communications*, vol. 27, no. 1, pp. 16-23, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Andreas Holzinger et al., “Towards Multi-Modal Causability With Graph Neural Networks Enabling Information Fusion for Explainable AI,” *Information Fusion*, vol. 71, pp. 28-37, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] W. Praewa., “Development of Natural Disaster Risk Map in Thailand,” *The 27th International Convention on Civil Engineering*, 2022. [Publisher Link]

[13] Seung-Hyun Moon et al., “Application of Machine Learning to an Early Warning System for Very Short-Term Heavy Rainfall,” *Journal of Hydrology*, vol. 568, pp. 1042–1054, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[14] Kavitha.R et al., "Dam Water Level Monitoring and Alerting System Using IOT," *SSRG International Journal of Electronics and Communication Engineering*, vol. 5, no. 6, pp. 19-22, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[15] Jo, T., *Machine Learning Foundations. Supervised, Unsupervised, and Advanced Learning*, Cham: Springer International Publishing, 2021. [Google Scholar] [Publisher Link]

[16] Malti Bansal, Apoorva Goyal, and Apoorva Choudhary, “A Comparative Analysis of K-Nearest Neighbour, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory Algorithms in Machine Learning,” *Decision Analytics Journal*, vol. 3, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Osama Harfoushi, Dana Hasan, and Ruba Obiedat. “Sentiment Analysis Algorithms Through Azure Machine Learning: Analysis and Comparison,” *Modern Applied Science*, vol. 12, no. 7, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[18] Ruhhee Tabbussum, and Abdul Qayoom., “Comparative Analysis of Neural Network Training Algorithms for the Flood Forecast Modeling of An Alluvial Himalayan River,” *Journal of Flood Risk Management*, vol. 13, no. 4, pp. 1-18, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[19] Tianyi Wang et al., “Multi-Task Multimodal Learning for Disaster Situation Assessment,” *In Proceedings of the 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pp. 209-212, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[20] Jean Baptiste Nsengiyumva, and Roberto Valentino, “Predicting Landslide Susceptibility and Risks Using GIS-Based Machine Learning Simulations, Case of Upper Nyabarongo Catchment,” *Geomatics, Natural Hazards and Risk*, vol. 11, pp. 1250–1277, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Binh Thai Pham et al., “Hybrid Integration of Multilayer Perceptron Neural Networks and Machine Learning Ensembles for Landslide Susceptibility Assessment at Himalayan Area (India) Using GIS,” *Catena*, vol. 149, pp. 52–63, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[22] Seung-Hyun Moon et al., “Application of Machine Learning to an Early Warning System for Very Short-Term Heavy Rainfall,” *Journal of Hydrology*, vol. 568, pp. 1042-1054, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[23] Zahra Faraji, “A Review of Machine Learning Applications for Credit Card Fraud Detection with a Case Study,” *SEISENSE Journal of Management*, vol. 5, no. 1, pp. 49-59, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[24] Joshi, J, and Saxena, S., “Regression Analysis in Data Science,” *Journal of Analysis and Computation*, vol. 14, no. 6, 2020. [Publisher Link]

[25] Chanawit Patcharacharoenwong, Kankawee Hernmek, and Warangkhana Kimpan, “Arrival Time Prediction Model to a Pier for Public Transportation Boats,” *Journal of Science Ladkrabang*, vol. 29, no. 2, pp. 31-44, 2020. [Google Scholar] [Publisher Link]