International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 17, Issue 1, 2023

An Integrated Predictive Model for Analysis and Flood Disaster Warning Using Artificial Neural Networks

Tepridht Phratep[a], Surachai Thongkhaew[b] and Prasong Praneetpolgrang[c], [a,b]School of Information Technology, Sripatum University, Bangkok, Thailand, [c]Academic Faculty, Navaminda Kasatriyadhiraj Royal Air Force Academy, Thailand. *E-mail: [a]hara.german@gmail.com, [b] Surachai.th@spu.ac.th, [c] prasong_pr@rtaf.mi.th

The purpose of this research is to develop a predictive model for analyzing and warning of disaster from natural hazards using artificial neural networks. This study consists of 6 steps: 1) problem definition, 2) data collection and data cleansing. In this research, the researcher used rainfall data in Thailand from 1901 to 2018, 3) splitting data into two parts, 80% for training and 20% for testing, 4) defining a multiclass neural network model, 5) training the model, and 6) evaluating the performance of the model using cross-validation test. This research demonstrates a development approach of a predictive model for analyzing and flood disaster warning from natural hazards using artificial neural networks, with high accuracy of 95.833%. It can be further developed to create a platform for connecting and using in disaster warning systems in the future.

Keywords: *Predictive Model, Flood Disaster, Neural Networks.*

1. INTRODUCTION

Currently, we can see that the impact of uncertain and unpredictable weather conditions has increased, causing significant impacts on society and the economy. Thailand has to face various types of natural disasters, including meteorological disasters, geographical disasters, and earth surface disasters, which are all natural disasters that occur naturally and are mostly seasonal, but can also occur suddenly, causing damage to life, health, property, and the environment. Disasters often cause physical and life losses, ranging from minor to fatal, depending on the severity and nature of the disaster. The impact on mental health can vary greatly. While the severity of the disaster and the impact on mental health may prevent the affected person from being able to help themselves, some may lose family assets and experience grief. They may experience stress and sometimes lose consciousness, which may lead to mental illness.

The impact of each disaster is accompanied by simultaneous loss of assets, leading to economic, social, agricultural, industrial, public transportation, and political impacts. Activities cannot be carried out as usual, and must be halted. Transportation is cut off, and transportation of goods is interrupted. Essential consumer goods, medicines, and products become more expensive. When there is price control, there will be shortages of goods, especially for non-essential products. Prices will decrease significantly, causing business owners to lose profits, and there may be inflation or currency devaluation. People with low incomes will suffer, and the economy will decline, with the industry coming to a halt and causing unemployment problems. This will result in people not having enough income to cover their expenses and may lead to an increase in crime. In Thailand, where 80% of the population are farmers, natural disasters can damage crops, destroy farmland, and lead to lower quality and quantity of agricultural products, which means farmers lose their income and incur costs such as labor and time. Additionally, animals used for farming purposes and breeding may also be lost or become endangered. As mentioned, it can be seen that the impact of disasters creates a significant loss of life and property in Thailand, and there is an increasing trend of severity due to the changing climate conditions.

In addition, the use of modern technology for warning systems in Thailand is still limited (A. Thotsaphon, et. al, 2020). Currently, technology is being utilized to promote the safety and security of the population in coping with emergency situations that may arise. The objective is to enhance the ability to prevent and mitigate the impact of disasters, as well as to inform those residing in high-risk areas of emergency situations in a timely manner, so that people have time to prepare, evacuate from risky areas, or carry out appropriate prevention and mitigation activities. Technology for disaster alert systems consists of various channels such as audible warnings, SMS messages, or notifications through smartphone applications, wireless technology for transmitting warning data, and online social media notifications.

Technology which uses for disaster alert systems is beneficial in reducing risks to the lives and property of the population and enhancing long-term disaster preparedness for society (P. Thawatchai, 2022; N, Riem et al., 2022). The use of artificial intelligence technology in disaster management is an effective tool that can improve the efficiency of disaster management. A good example of using artificial intelligence technology in disaster management is a research project that proposes a probabilistic prediction model for analyzing and warning of flood disasters using neural networks. By analyzing and processing rainfall data, disaster management personnel can manage disasters more efficiently and reduce the risk of disasters in the future.

2. THEORY AND RELATED RESEARCH

2.1 Artificial Intelligence (AI)

Artificial Intelligence (AI) consists of the words “intelligence,” which means the ability to think independently, and “artificial,” which refers to something created by humans. This results in AI being defined as a power of human-created thinking (Limna et al., 2021). AI is a computer system that simulates human work. It is highly efficient in performing specific tasks and also contributes to change in all sectors by enabling computers to make correct decisions and improving operational efficiency (Holzinger et al., 2021; Kumar et al., 2021). In addition, AI helps people work smarter, but it is necessary to develop new abilities and innovations such as technology, emotional skills, social skills, and creativity (Ivanov & Webster, 2019; Ruel & Njoku, 2021).

AI is a thinking process that is caused and influenced by humans, including natural language processing (NLP), computer vision, robotics, and expert systems (Prinya Songvasin, 2019). Nowadays, AI is used to improve work efficiency in fields such as medicine, industry, agriculture, and business. In this research, AI has been applied to disaster warning systems to alert people of natural disasters. The relationship between AI, machine learning (ML), and deep learning (DL) is shown in Figure 1.

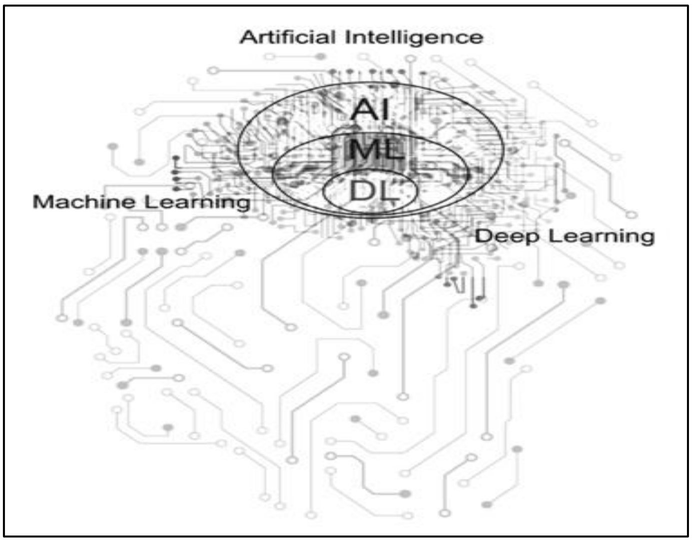


Figure 1: The Relationship between Artificial Intelligence and Machine Learning.

2.1.1 Machine Learning (ML)

Machine Learning is a process of artificial intelligence that teaches computers to understand the nature of human beings. It has both shallow and deep architectures and has the ability to learn and improve through experience. The machine learning process begins with raw data used to extract useful information that helps with decision-making. The main objective is to allow machines to learn useful information like humans do, on a moral level, by learning from experience and using machine learning algorithms to improve performance directly based on the amount of available data (C. Perez, 2019). There are four types of machine learning: 1) Supervised Learning, 2) Semi-Supervised Learning, 3) Unsupervised Learning, and 4) Reinforcement Learning (Jo, T., 2021). The types of machine learning are as follows:

1. Supervised learning is a type of machine learning that involves considering two types of results during the learning process: the target output initially provided for each training example, and the output computed by the learning algorithm. The supervised learning process is a learning paradigm that adjusts parameter values to minimize the difference between the target output and the computed output. This type of learning is commonly used for classification tasks, and the learning process is typically represented in a structure as shown in Figure 2.

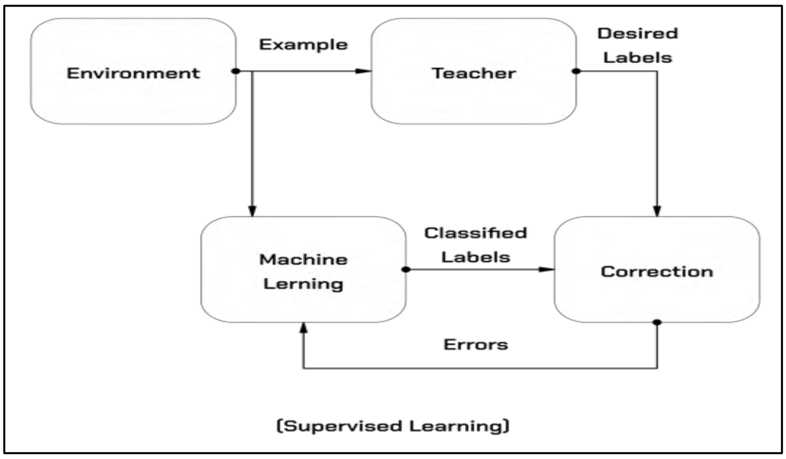


Figure 2: Machine Learning with a Supervised Learning.

2. Semi-supervised learning is represented as a block diagram with the aim of using examples without labels, as well as labeled examples for training machine learning algorithms. The machine learns from labeled examples, reducing errors between the target and calculated labels, as well as unlabeled examples. This depends on the similarity between them, which combines supervised and unsupervised learning. This is achieved by adding unlabeled examples. Semi-supervised learning is popularly used in classification and regression tasks, as shown in Figure 3.

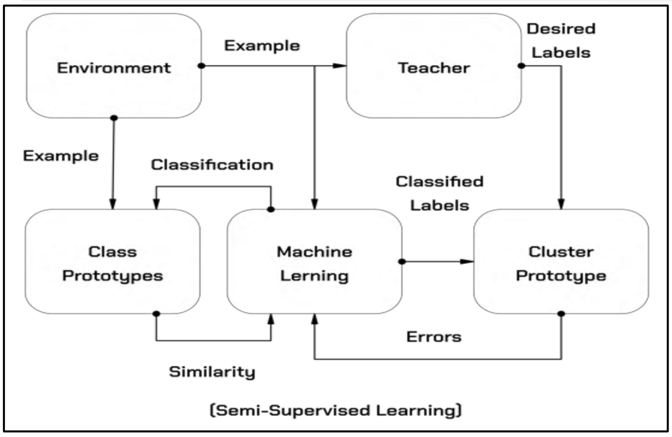


Figure 3: Machine Learning with a Semi-Supervised Learning.

3. Unsupervised Learning is a learning and training process that does not require a teacher or labeled data. It uses a set of unlabeled examples for learning, starting with random sampling. Unsupervised learning is a process of improving the efficiency of the prototype cluster, which depends on the similarity between the training examples. Algorithms for unsupervised learning are commonly used for clustering data. This is shown in Figure 4.

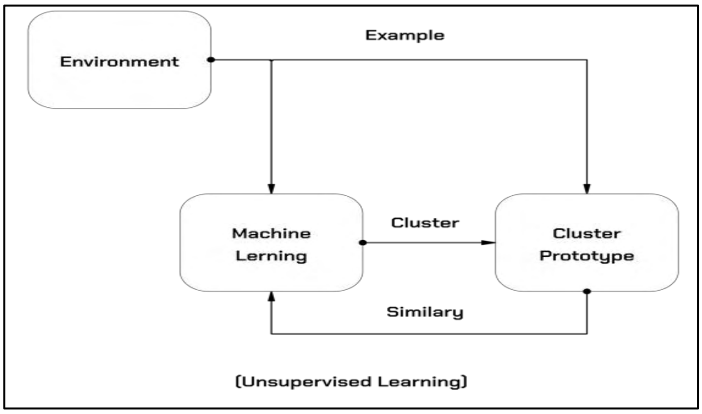


Figure 4: Machine Learning with a Unsupervised Learning.

4. Reinforcement Learning is a type of machine learning that involves updating parameters to increase rewards and minimize penalties. Inputs are received from the external environment and outputs are created based on actions taken. Rewards or penalties are received from the environment and parameters are updated to optimize rewards and avoid penalties. The principle of operation is shown in Figure 5.

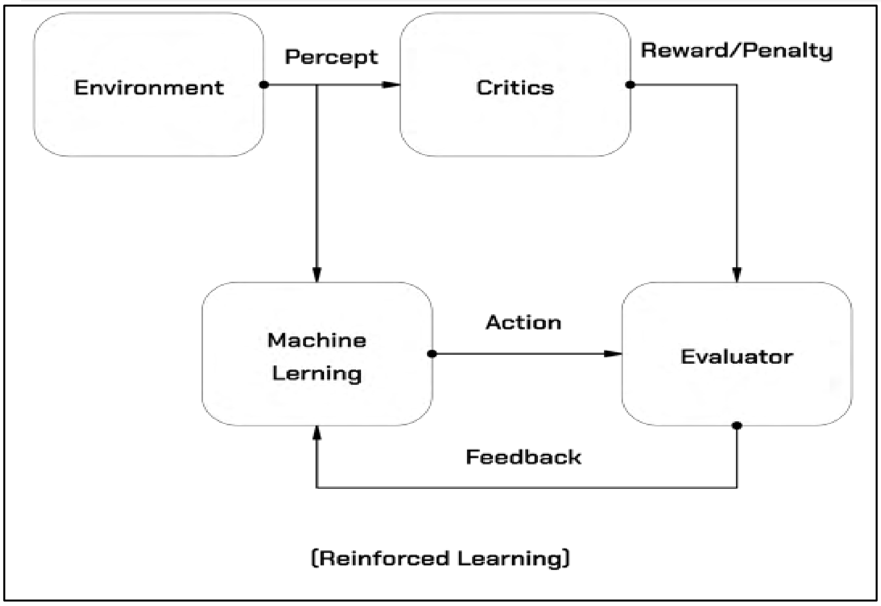


Figure 5: Machine Learning with a Reinforcement Learning.

2.1.2 Deep Learning (DL)

Deep learning is a popular form of machine learning. It refers to an architecture with multiple hidden layers (deep neural networks) to learn various features that can be applied across multiple levels. Deep learning algorithms leverage the power of unknown structures to distribute input data, in order to find the best representation, which typically involves multiple levels of learning, with increasingly higher level features. In deep learning, the problem arises in the hierarchical nature of the ideas. Each idea is built on top of previous layers of the model encoding basic representations of the problem, while the higher levels are built from lower layers to create more complex ideas.

This process can be seen as a hierarchical learning because each layer in the network uses the output from the previous layer as a "building block" to create a more complex learning pathway in higher layers. The step-by-step process of comparing the learning methods of the original model with the deep learning approach is based on the characteristics designed by hand and the deep learning approach is based on the hierarchical learning shown in Figure 6.

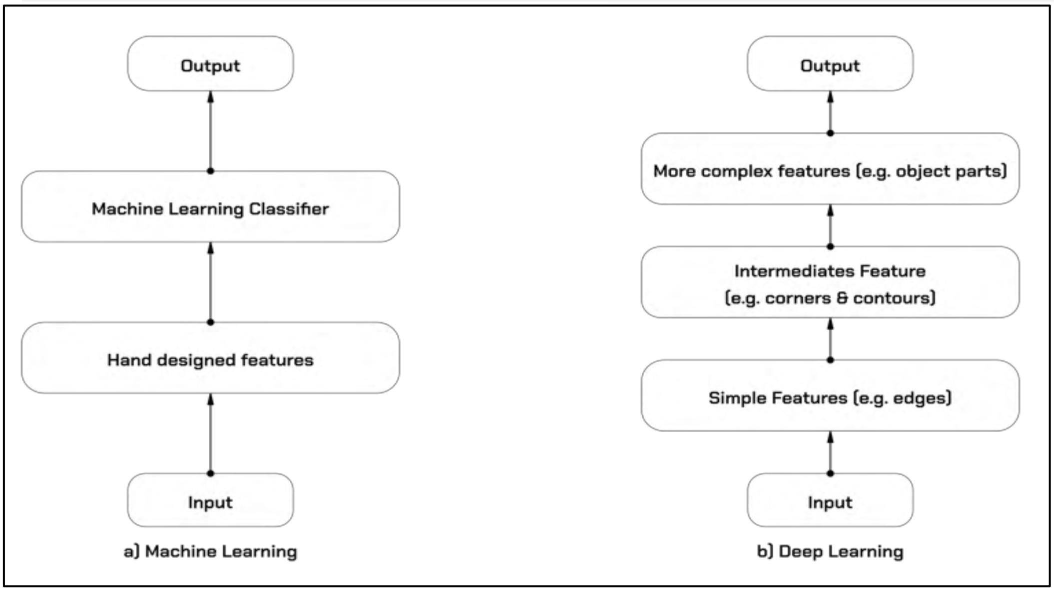


Figure 6: a) General Machine Learning Using Manually Designed Feature Extraction Algorithms
b) Deep Learning Using Automatically Learned Representations through Layered Representations.

2.2 Artificial Neural Networks (ANN)

The artificial neural network is the combination of nodes that are used for simple processing, functioning like cells of living things. The processing capabilities of the network are integrated into the points of connection between units or weights that result from the process of adaptation or learning from a set of training patterns. Artificial neural networks are widely used in statistical analysis and data modeling to serve as an alternative to linear regression or clustering techniques. Therefore, artificial neural networks are used for classification or prediction to be applied in each specific professional field. Artificial neural networks are in the form of parallel distributed computing. The necessary requirement of the artificial neural network is the ability to detect features that are important for processing data, which is consistent with reality. It can be seen that the basic unit of the artificial neural network is the Threshold Logic Unit (TLU), which is a mechanism that weights input and output data equivalent to "1" and if the result is outside of that, it will be “0”. The TLU is the integration mechanism based on the foundation of real neural cells. Details are shown in Figure 7-9.

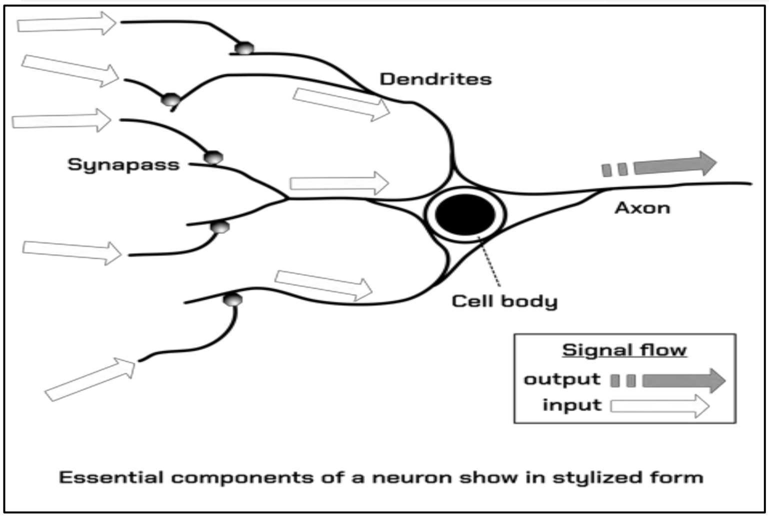


Figure 7: The Components of a Neuron.

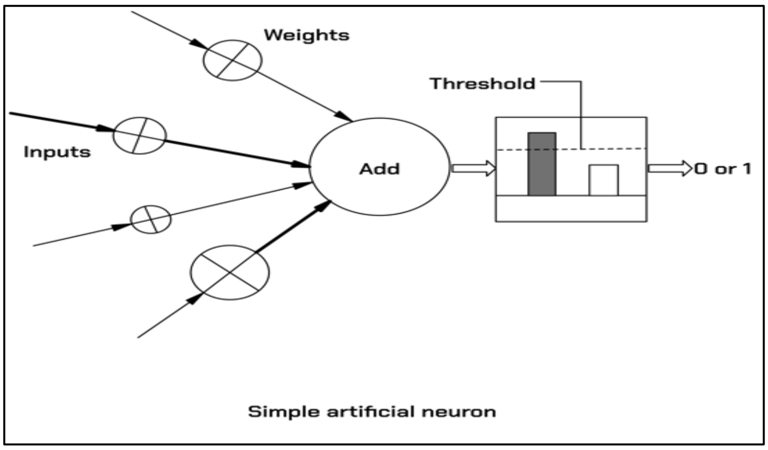


Figure 8: Simple Artificial Neural Networks and Example of Neural Networks Model.

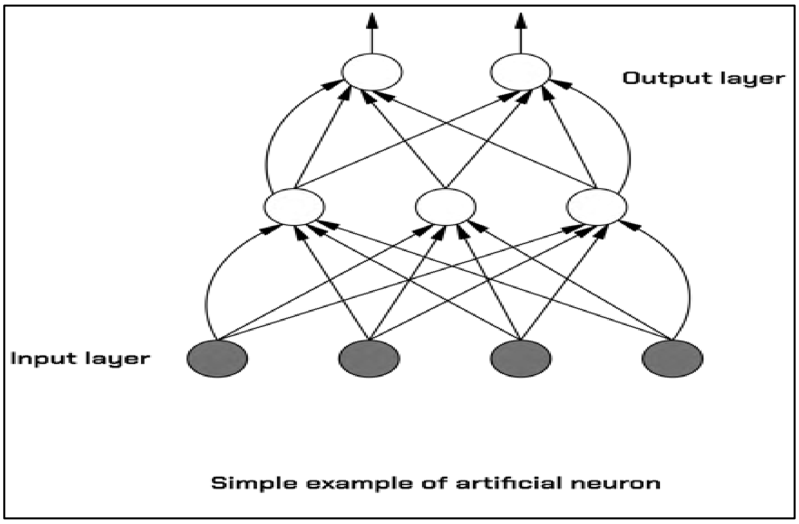


Figure 9: Simple Example of Artificial Neural Networks.

The learning process of Artificial Neural Networks (ANN) operates on the principle of examples. Generally, specific task rules are not programmed in advance. Instead, the operation of the network depends on a set of interconnected units or nodes called artificial neurons, which simulate the behavior of neurons in the brain in a simplified way. The "signals" used in the network connections are real numbers, and the output of each neuron is calculated by a non-linear function of the sum of its inputs. The connections between neurons are called edges, and these edges typically have weights that are adjusted during learning. In general, an artificial neural network may have from a few thousand to several million neurons and may have connections between multiple layers of input and output. A network with more layers creates deeper architectures, which is known as "Deep Neural Networks" or "Deep Learning".

Deep Neural Networks (DNN) or Deep Learning (DL) is a widely used machine learning technique based on artificial neural networks and learning algorithms. Deep Learning architectures, such as Deep Neural Network, Recurrent Neural Network, and Convolutional Neural Network have been applied to various fields and have produced comparable or even superior results compared to human experts in some cases.

Recurrent neural networks (RNN) are artificial neural networks designed for sequential data. They utilize the principle of incorporating internal state within the model to feed it back as new input, coupled with regular input, to help the model learn and memorize the patterns of the input sequence, leading to further advancement such as the Long Short-Term Memory (LSTM) and the Gate Recurrent Unit (GRU). The latter is particularly useful in overcoming the issues of Vanishing Gradient and Exploding Gradient, while being more efficient due to having fewer parameters than LSTM, thus making it easier and faster to train, especially with small datasets where GRU performs well.
A Convolutional Neural Network (CNN) is an artificial neural network that simulates human vision by dividing an image into small sub-regions and then combining them to recognize objects, using mathematical calculations that follow the concept of spatial convolution in image processing. The network extracts features of objects by applying multiple filters to the input image and pooling the results. This helps the network to learn and recognize patterns and objects from images.

K-Nearest Neighbor Algorithm (KNN) is a classification method that decides which class or condition a new case belongs to by checking the closest K-number of existing cases. It uses the principle of measuring the distance between each attribute of the input data to calculate the proximity between data points. KNN is suitable for numerical data with non-continuous variables that require additional special handling.

Levenberg-Marquardt (LM) algorithm is an iterative technique used to find the minimum value of a multi-variable function represented as the sum of squares of non-linear functions [4], [6]. It has become a standard technique for solving the least-squares problem that is not linear [7], and has been widely used in various fields. LM can be considered as a blend of steepest descent and Gauss-Newton methods. When the current solution is far from the correct solution, the algorithm works like the steepest descent method. As the current solution approaches the correct solution, it works like the Gauss-Newton method. The LM algorithm can update the parameter between the level set update and Gauss-Newton update, as shown in equation (1).

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

Where

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

μ is 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)

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

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

Where

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

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

is the input data which has been added to 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

Support Vector Machine (SVM) is a supervised learning technique used for classification and regression to predict continuous values. It utilizes machine learning theory to improve prediction accuracy and avoid overfitting. SVM is commonly used for face detection, image and text categorization, autonomous driving, chatbots, and more (Bansal, M., Goyal, A., & Choudhary, A., 2022). The SVM regression analysis technique is suitable for datasets with a large number of features and is effective for both linear and nonlinear classification of data.

2.3 Disaster

During times of disaster, there are losses and damages that occur within society. The United Nations Office for Disaster Risk Reduction (UNDRR) defines a disaster as a serious disruption of the functioning of a community or a society, involving widespread human, material, economic or environmental losses and impacts, which exceeds the ability of the affected community or society to cope using its own resources (Nicolas A. Valcik, 2013).

Unexpected situations can arise from natural disasters such as earthquakes, fires, floods, or from human events such as road accidents or plane crashes. Natural disasters, on the other hand, are disasters that occur due to natural causes such as storms, floods, earthquakes, and other similar events that cannot be controlled. Disaster Risk Reduction means preventing and reducing the impact of disasters by preparing in advance, such as building structures that are strong enough to withstand the force of disasters, preparing equipment for use, preparing areas for the disposal of materials generated by disasters, and using systematic approaches to analyze and manage the causal factors of disasters. This includes reducing the vulnerability of people and property, managing land and the environment wisely, and preparing for unforeseen events.

An Early Warning System (EWS) is designed to increase disaster preparedness and reduce the risk of disasters. This system works by using data from sensors and communication networks to prepare for emergency situations and reduce the impact of disasters. EWS can be applied in various fields such as environmental disasters like volcanic eruptions, earthquakes, floods, and entering the rainy season. It can also be applied to safety issues like outbreaks of infectious diseases, accidents in transportation, and economic crises such as financial crises. By using this system, those responsible for disaster management and prevention can have more accurate and timely data analysis, which helps improve disaster management and preparedness.

An effective early warning system is necessary to create and disseminate warning information at the appropriate time and with meaningful content so that individuals, communities, and organizations that are threatened by danger can prepare and take appropriate actions in a timely manner to reduce the likelihood of danger or loss. The focus should be on the ability of the early warning system to access all relevant parties. Therefore, the early warning system should not be delayed in developing both the early warning system and the warning messages (Musavi & Syed Hyder Abbas, 2020).

2.4 Related Research

Satwik P. M. & Meenatchai S. (2020) found in their research that various tasks were predicted to involve the use of wireless sensor networks and advanced neural networks. The accuracy of the predictions was evaluated by various assessment parameters such as the nature of the task, the correlation coefficient of the matrices, the mean square error, and the validation of the epoch. The research was evaluated using different machine learning algorithms to analyze the assessment from different perspectives.

The research work of Chen, J. et al. (2020) studied from the framework of driving forces, pressures, states, impacts, and responses (DPSIR) and found that the accuracy of the prediction from the training set was 0.77, the validation set was 0.75, and the testing set was 0.748. This indicates that the model has good accuracy in prediction. Finally, the risk assessment value of urban agglomeration in the Yangtze River Delta (YRD) from 2016 to 2018 was predicted, and the application of ArcGIS software and the RF-RBF model for evaluating the risk of urban communities in the YRD region from 2016 to 2018 will depend on the RF and RBF.

Ruhhee Tabbussum & Abdul Qayoom (2020) studied the development of a new flood forecasting model using artificial neural network (ANN) algorithms. The model was evaluated and verified for a case study of the river basin in the Kosi-Mahananda catchment of the Himalayan region in India-Nepal during the September 2014 flood event, which was caused by various atmospheric reactions, resulting in a maximum discharge of over 115,000 m^3/s. The study aimed to develop an ANN model for flood forecasting by developing five different neural network models: Bayesian neural network, Levenberg–Marquardt neural network, conjugate gradient descent-based GMDH neural network, size-adjustable GMDH neural network, and flexible backpropagation neural network. The Levenberg-Marquardt artificial neural network (ANN) model with a mean squared error of 0.002128 (the lowest of all models) and an R2 efficiency of 95.839% was developed, considering the hydrological, geomorphological, and hydrogeological characteristics of the Kosi-Mahananda river basin in the Himalayan region of India-Nepal during the September 2014 flood event, which was caused by various atmospheric reactions, resulting in a maximum discharge of over 115,000 m^3/s. It is crucial to develop an early warning system for floods in the long term to manage flood risk and reduce the impact of disasters in the future. This requires the integration of a well-distributed meteorological and hydrological monitoring network (including Doppler radar) and a river gauging station. The study examined the potential of using ANN techniques for flood forecasting modeling. The comparison showed that five different ANN models for flood prediction were effective, but the Levenberg-Marquardt model performed better and was more reliable.

Sung, W. T. et al. (2022) studied the early warning of impending flash floods based on AIoT by analyzing data from global natural disaster risk reduction centers and volcanic hazard centers in Indonesia. They found that floods are a recurring natural disaster every year, especially on the slopes of mountains where the risk is higher than in urban areas due to the potential for other natural disasters such as landslides and damage to walking paths. It is necessary to have sensors to detect and monitor floodwaters, and the researchers have developed a warning system using AIoT technology that analyzes real-time flood data, allowing officials to check on people in mountainous areas and issue advance warnings. The system design involves importing sensors into microcontrollers and communicating between stations using LoRa and SIM900 to send data to cloud servers via the internet. All sensor data for each station is displayed on the application, and alerts are sent via SMS and the application.

Zhou et al. (2018) studied the risk of landslides using a case study in Longju, an area of the Three Gorges Reservoir in China, by considering the characteristics of two types of landslides: rockfall landslides and debris flow landslides. In discriminating the risk of landslides, the SVM model performed better than ANN and LR with a higher AUC value of 0.881 for predicting all landslides, compared to ANN at 0.836 and LR at 0.697.

3. RESEARCH METHODOLOGY

In this research, the researcher presented a supervised algorithm of the neural network model using the rainfall data between 1901 to 2018 to predict the occurrence of floods with a model-building process shown in Figure 10. From Figure 10, the steps to create a composite model include importing data, selecting datasets, cleaning data, splitting data (for training and testing), selecting algorithms, building the model, predicting with the model, and measuring the model's performance, with details of each step as follows.

3.1 Problem Definition

In the process of defining the scope of the problem, researchers focus on building a model for analyzing and warning of disasters from natural disasters using artificial neural network techniques from rainfall data.

3.2 Data Collection

The process of collecting this data, the researcher has the following steps.

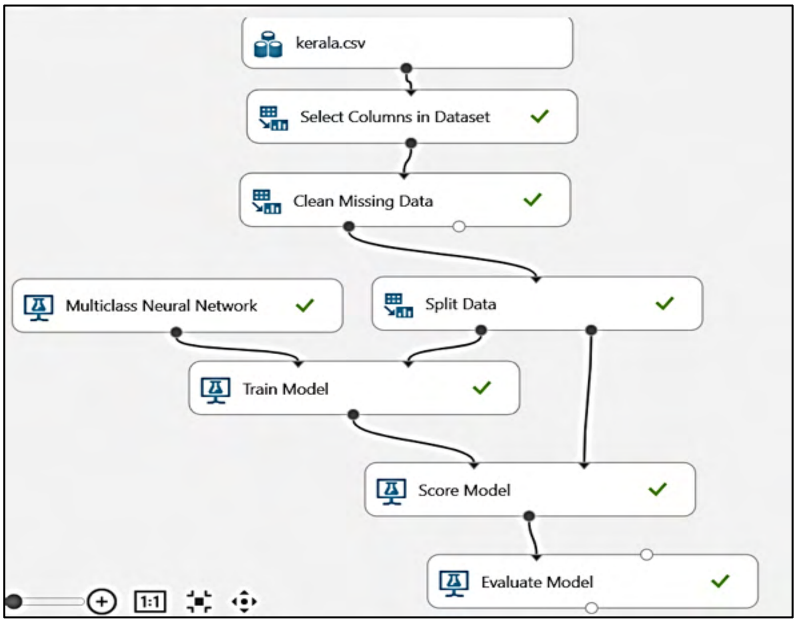


Figure 10: Steps for Creating a Model Using Microsoft Machine Learning Studio

3.2.1 Research data sources use rainfall data between 1901 to 2018 from a sample dataset from the website https://www.kaggle.com/code/mukulthakur177/flood-prediction-model/data as shown in Figure 11.

	B	C	D	E	F	G	H	I	J	K	L	M	N	P
1	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	FLOODS
2	1901	28.70	44.70	51.60	160.00	174.70	824.60	743.00	357.50	197.70	266.90	350.80	48.40	YES
3	1902	6.70	2.60	57.30	83.90	134.50	390.90	1,205.00	315.80	491.60	358.40	158.30	121.50	YES
4	1903	3.20	18.60	3.10	83.60	249.70	558.60	1,022.50	420.20	341.80	354.10	157.00	59.00	YES
5	1904	23.70	3.00	32.20	71.50	235.70	1,098.20	725.50	351.80	222.70	328.10	33.90	3.30	YES
6	1905	1.20	22.30	9.40	105.90	263.30	850.20	520.50	293.60	217.20	383.50	74.40	0.20	NO
7	1906	26.70	7.40	9.90	59.40	160.80	414.90	954.20	442.80	131.20	251.70	163.10	86.00	NO
8	1907	18.80	4.80	55.70	170.80	101.40	770.90	760.40	981.50	225.00	309.70	219.10	52.80	YES
9	1908	8.00	20.80	38.20	102.90	142.60	592.60	902.20	352.90	175.90	253.30	47.90	11.00	NO
10	1909	54.10	11.80	61.30	93.80	473.20	704.70	782.30	258.00	195.40	212.10	171.10	32.30	YES
11	1910	2.70	25.70	23.30	124.50	148.80	680.00	484.10	473.80	248.60	356.60	280.40	0.10	NO
12	1911	3.00	4.30	18.20	51.00	180.60	990.00	705.30	178.60	60.20	302.30	145.70	87.60	NO
...														
...														
...														
113	2012	7.40	11.00	21.00	171.10	95.30	430.30	362.60	501.60	241.10	187.50	112.90	9.40	NO
114	2013	3.90	40.10	49.90	49.30	119.30	1,042.70	830.20	369.70	318.60	259.90	154.90	17.00	YES
115	2014	4.60	10.30	17.90	95.70	251.00	454.40	677.80	733.90	298.80	355.50	99.50	47.20	YES
116	2015	3.10	5.80	50.10	214.10	201.80	563.60	406.00	252.20	292.90	308.10	223.60	79.40	NO
117	2016	2.40	3.80	35.90	143.00	186.40	522.20	412.30	325.50	173.20	225.90	125.40	23.60	NO
118	2017	1.90	6.80	8.90	43.60	173.50	498.50	319.60	531.80	209.50	192.40	92.50	38.10	NO
119	2018	29.10	52.10	48.60	116.40	183.80	625.40	1,048.50	1,398.90	423.60	356.10	125.40	65.10	YES
120														

Figure 11: Dataset.

3.2.2 Data Cleansing

Data cleansing is a process of managing raw data to be in a format that can be used for machine learning (ML) model creation. If the data has errors, it can affect the accuracy of the model.

3.3 Splitting Data

This step involves dividing the dataset into two parts: one for training the model and the other for testing the model's performance (Picard, R.R., & Berk, K.N., 1990). In this research, the data was split into 80% for training and 20% for testing, as shown in Figure 12.

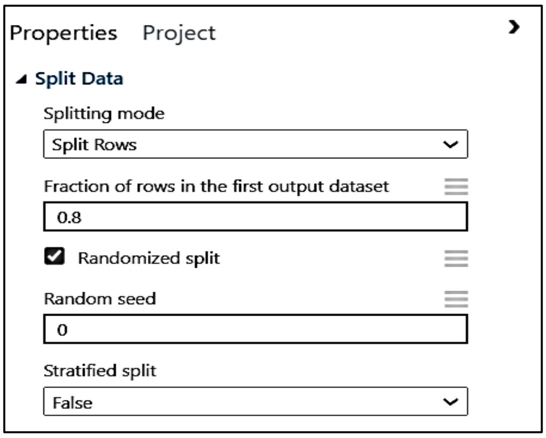


Figure 12: Split Data.

3.4 Define a Model

The researcher selected the method of creating a multi-class classification model using a neural network algorithm for training the data, as shown in Figure 13.

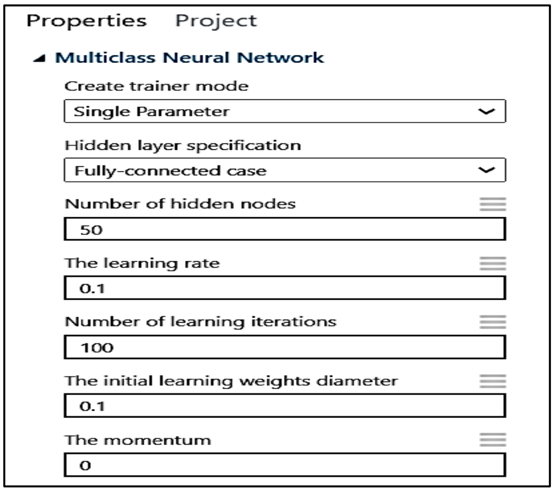


Figure 13: Model Selection.

3.5 Training the ML Model

In the model training step, the researcher determines the data label that needs to be analyzed for flooding. In this research, the researcher defines the FLOODS column to predict the likelihood of flooding. shown in Figure 14.

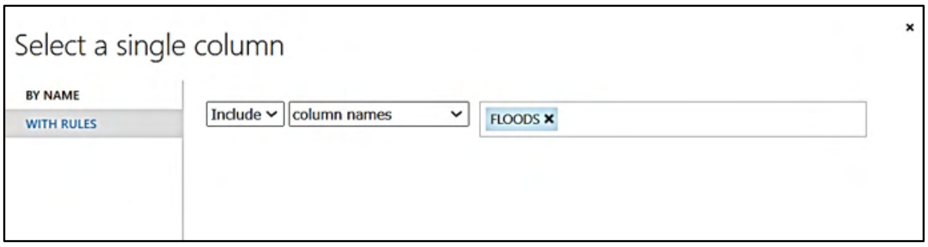


Figure 14: Step of Model Selector.

3.6 Model Evaluation

The evaluation of model performance uses the model performance metrics from a 2x2 Confusion Matrix, which is an important table for measuring the ability of Machine Learning in solving Classification problems. Details are shown in Figure 15.

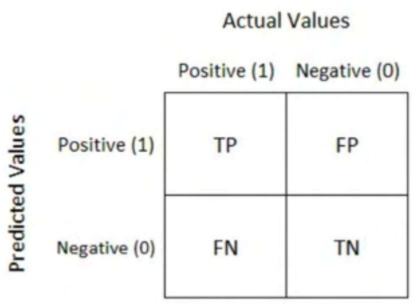


Figure 15: Confusion Matrix Dimension 2x2 (Faraji, Z., 2022)

Where:
True Positive (TP) is when the program predicts “true” and the actual value is “true”.
True Negative (TN) is when the program predicts “false” and the actual value is “false”.
False Positive (FP) is when the program predicts “true” but the actual value is “false”.
False Negative (FN) is when the program predicts “false” but the actual value is “true”.

4. RESEARCH RESULTS

In this research, the results of the study and discussion can be described as follows.

4.1 Model Overview

From this research, a diagram showing the model can be drawn as shown in Figure 16.

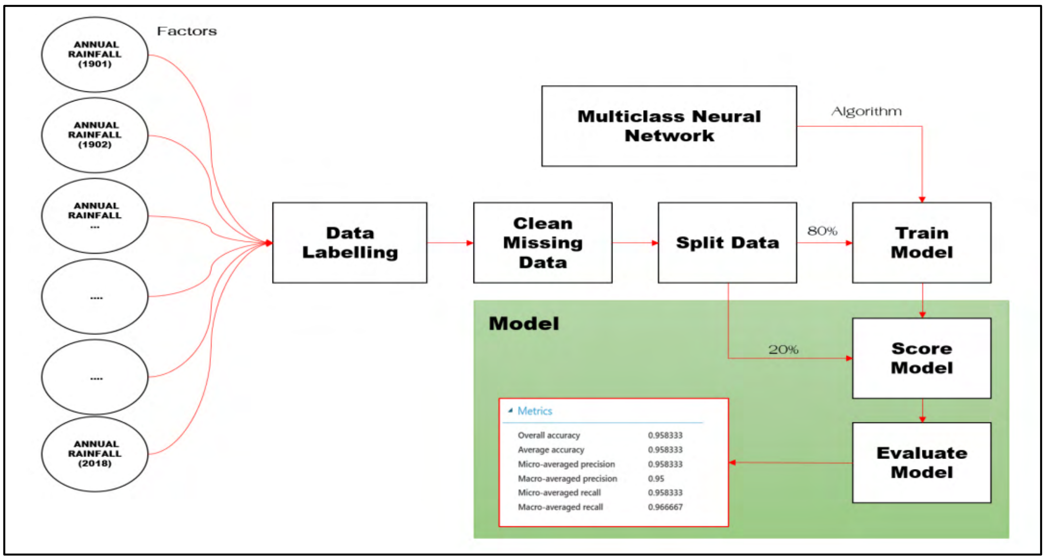


Figure 16: Overview of the Model.

4.2 Dataset

In this study, the researchers used a rainfall dataset from 1901-2018, as shown in Figure 17.

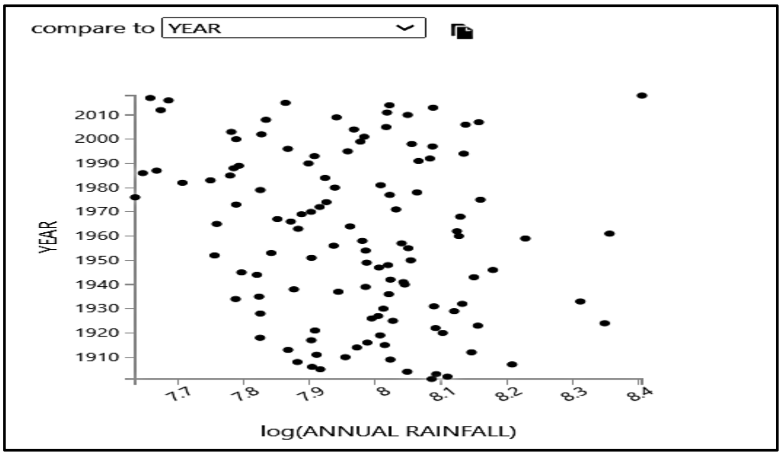


Figure 17: Graph Showing the Dataset of Rainfall Amount.

4.3 Training Data

In this research, the researcher divided dataset for model training into 80% of the total dataset, with a minimum value of 2068.8, a maximum value of 4473, and an average value of 2910.1543, as shown in Figures 18-19.

Statistics

Mean	2910.1543
Median	2906.85
Min	2068.8
Max	4473
Standard Deviation	467.5853
Unique Values	94
Missing Values	0
Feature Type	Numeric Feature

Figure 18: Statistics of the Training Dataset.

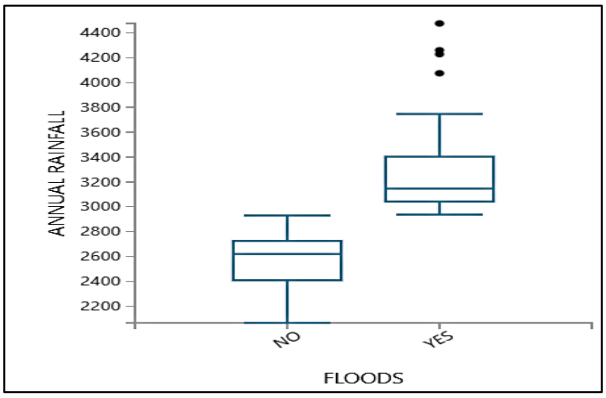


Figure 19: Statistics on Probability of Flooding in the Training Dataset.

4.4 Test Data

In this study, the researchers divided dataset for testing the model into 20% of the total dataset. The minimum value is 2223.3, the maximum value is 3671.1, and the average value is 2985.1375, as shown in Figures 20-21.

Statistics

Mean	2985.1375
Median	3013.9
Min	2223.3
Max	3671.1
Standard Deviation	389.0259
Unique Values	24
Missing Values	0
Feature Type	Numeric Feature

Figure 20: Statistics of the Test Dataset.

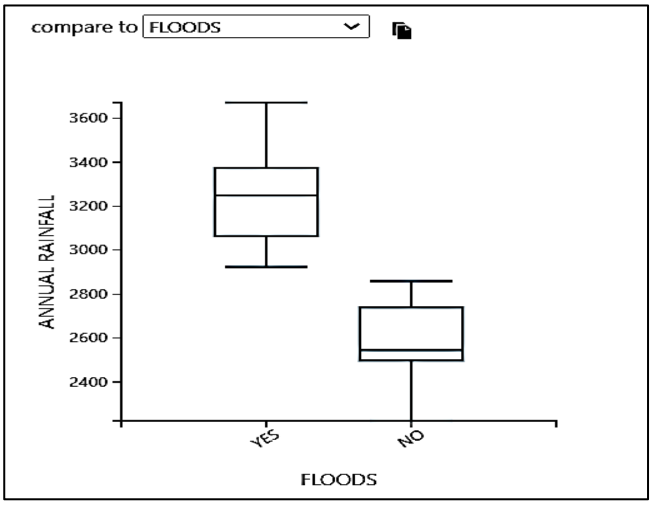


Figure 21: Statistics of Flood Probability in Test Data.

4.5 Model Comparison Results on Test Datasets

The comparison results of the model on the test datasets are shown in the FLOODS column compared to the Scored Labels. The accuracy and precision of the prediction, the statistics of the probability of no floods (Scored Probabilities for Class "NO"), and the statistics of the probability of floods (Scored Probabilities for Class "YES") are shown in Figure 22.

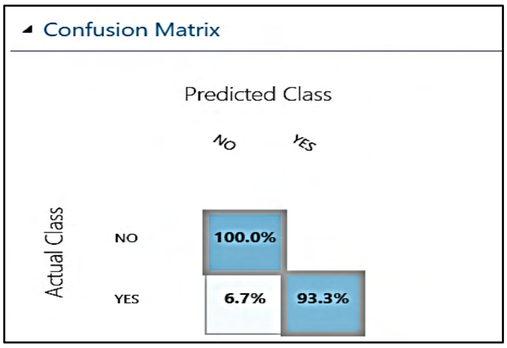
rows 24 columns 18

JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL RAINFALL	FLOODS	Scored Probabilities for Class "NO"	Scored Probabilities for Class "YES"	Scored Labels
648	484.2	255.9	249.2	280.1	53	3003.3	YES	0.003684	0.99631	YES
702.5	457.3	280	198.3	182.6	0.1	2858.8	NO	0.735445	0.265152	NO
749.6	459.9	134.1	339.8	298.1	10.2	2937.5	YES	0.27702	0.72271	YES
743	357.5	197.7	266.9	350.8	48.4	3248.6	YES	0.000129	0.99987	YES
1205	315.8	491.6	358.4	158.3	121.5	3326.6	YES	0.000019	0.999981	YES
531.3	675.9	457.7	368.9	204.3	19.9	3498.4	YES	0.000003	0.999997	YES
686.7	516.8	119.4	171	365.6	39	3176.7	YES	0.001446	0.998554	YES
662.2	383.7	211.7	163.8	261.7	23.3	2503	NO	0.999997	0.000003	NO
511.5	495	70.6	164.4	127.5	10.8	2223.3	NO	1	0	NO
392.8	236	438.5	378.2	178.1	19.8	3134.7	YES	0.001366	0.998637	YES

Figure 22: Comparison of Model Predictions with Test Datasets.

4.6 Result of Model Evaluation

The performance of the model was evaluated to assess the prediction results generated from the model that created in machine learning. The results are shown in Figure 23.



Metrics

Overall accuracy	0.958333
Average accuracy	0.958333
Micro-averaged precision	0.958333
Macro-averaged precision	0.95
Micro-averaged recall	0.958333
Macro-averaged recall	0.966667

Figure 23: Model Performance Evaluation Result.

From Figure 23, the performance evaluation of the model shows that:

The accuracy and precision have a value of 95.8333%. The recall has a value of 96.66% for correctly predicted data.

5. CONCLUSION

In this research, the predictive model for analyzing and warning of natural disasters caused by floods using artificial neural networks is developed from a dataset of monthly rainfall quantities from the years 1901-2018. In addition, the data was divided into 80% for training and 20% for testing using the Multiclass Neural Networks Algorithm. The performance measurement of the model found an accuracy of up to 95.833%. The researchers expect that this model can be used in predicting, analyzing, and warning of flood disasters in the future.

REFERENCES

Abdullahi, S. I., Habaebi, M. H., & Abd Malik, N. (2019). Intelligent flood disaster warning on the fly. developing IoT-based management platform and using 2-class neural network to predict flood status. Bulletin of Electrical Engineering and Informatics, 8(2), 706-717.

Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. Journal of pharmaceutical and biomedical analysis, 22(5), 717-727.

Ahmed, A. S., Abood, M. S., & Hamdi, M. M. (2021). Advancement of deep learning in big data and distributed systems. In 2021 3rd international congress on human-computer interaction optimization and robotic applications (HORA), (pp. 1-7). IEEE.

Bansal, M., Goyal, A., & Choudhary, A. (2022). 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, June.

Benevides, P., Catalao, J., Nico, G. (2019). Neural Network Approach to Forecast Hourly Intense Rainfall Using GNSS Precipitable Water Vapor and Meteorological Sensors. Remote Sens. Remote Sens, 11(8), 966.

Brenot H., Rohm W., Kaˇcmaˇrík, M., Möller, G., Sá, A., Tonda´s, D., Rapant, L., Biondi, R., Manning, T., & Champollion, C. (2020). Cross-Comparison, and methodological improvement in GPS tomography. Remote Sens, 12(1), 30.

Cai, H., Jia, X., Feng, J., Yang, Q., Hsu, Y. M., Chen, Y., & Lee, J. (2019). A combined filtering strategy for short term and long-term wind speed prediction with improved accuracy. Renewable energy, 136, 1082-1090.

Chen, J., Li, Q., Wang, H., & Deng, M. (2020). A machine learning ensemble approach based on random forest and radial basis function neural network for risk evaluation of regional flood disaster: a case study of the Yangtze River Delta, China. International journal of environmental research and public health, 17(1), 49.

dLimna, P., Siripipatthanakul, S., & Phayaphrom, B. (2021). The Role of Big Data Analytics in Influencing Artificial Intelligence (AI) Adoption for Coffee Shops in Krabi, Thailand. International Journal of Behavioral Analytics, 1(2), 1-17.

Dong, S., Yu, T., Farahmand, H., & Mostafavi, A. (2021). A hybrid deep learning model for predictive flood warning and situation awareness using channel network sensors data. Computer-Aided Civil and Infrastructure Engineering, 36(4), 402-420.

Faraji, Z. (2022). A Review of Machine Learning Applications for Credit Card Fraud Detection with A Case study. SEISENSE Journal of Management, 5(1), 49-59.

Gabrielle, S. I. M. M. (2018). Disaster response in Southeast Asia: The ASEAN agreement on disaster response and emergency management. Asian Journal of International Law, 8(1), 116-142.

Holzinger, A., Malle, B., Saranti, A., & Pfeifer, B. (2021). Towards Multi-Modal Causability with Graph Neural Networks Enabling Information Fusion for Explainable AI. Information Fusion, 71, 28-37.

Ivanov, S., & Webster, C. (2019). Conceptual Framework of the Use of Robots, Artificial Intelligence and Service Automation in Travel, Tourism, and Hospitality Companies. Robots, Artificial Intelligence, and Service Automation in Travel, Tourism, and Hospitality, Emerald Publishing Limited, Bingley, 7-37.

Jo, T. (2021). Machine Learning Foundations. Supervised, Unsupervised, and Advanced Learning. Cham: Springer International Publishing.

Junfei Chen, Qian Li, Huimin Wang, and Menghua Deng. (2020). A Machine Learning Ensemble Approach Based on Random Forest and Radial Basis Function Neural Network for Risk Evaluation of Regional Flood Disaster: Case Study of the Yangtze River Delta China. International Journal of Environmental Research and Public Health.

Kumar, S., Kumar, V., & Attri, K. (2021). Impact of Artificial Intelligence and Service Robots in Tourism and Hospitality Sector: Current Use & Future Trends. Administrative Development: A Journal of HIPA, Shimla, 8(SI-1), 59-83.

Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2023). A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. Annals of Data Science, 10(1), 183-208.

Limna, P. (2022). Artificial Intelligence (AI) in the hospitality industry: A review article. Int. J. Comput. Sci. Res, 6, 1-12.

Le, X. H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of long short-term memory (LSTM) neural network for flood forecasting. Water, 11(7), 1387.

Li, H., & Sriver, R. L. (2019). Impact of air–sea coupling on the simulated global tropical cyclone activity in the high-resolution Community Earth System Model (CESM). Climate Dynamics, 53(7-8), 3731-3750.

Munawar, H. S., Ullah, F., Qayyum, S., Khan, S. I., & Mojtahedi, M. (2021). Uavs in disaster management. Application of integrated aerial imagery and convolutional neural network for flood detection. Sustainability, 13(14), 7547.

Musavi, Syed Hyder Abbas, (2020). Early warning-based multihazard and disaster management systems. CRC Press.

Pham Quang, M., & Tallam, K. (2022). Predicting Flood Hazards in the Vietnam Central Region: An Artificial Neural Network Approach. Sustainability, 14(19), 11861.

Roy, S. S., Mohapatra, M., Tyagi, A., & BHOWMIK, S. R. (2019). A review of Nowcasting of convective weather over the Indian region. Mausam, 70(3), 465-484.

Ruhhee Tabbussum, Abdul Qayoom. (2020). Comparative analysis of neural network training algorithms for the flood forecast modeling of an alluvial Himalayan River. Journal of Flood Risk Management.

Ruel, H., & Njoku, E. (2021). AI Redefining the Hospitality Industry. Journal of Tourism Futures, 7, 1, 53-66.

Satwik P. M., Meenatchai Sundram. (2020). Evaluation on Predictive analysis of rain disaster using Adaptive Neural Turing Network. International Journal of Advance Research, Ideas, and Innovations in Technology, pp.870-873.

Schyns, P. G., Snoek, L., & Daube, C. (2022). Degrees of algorithmic equivalence between the brain and its DNN models. Trends in Cognitive Sciences.

Sung, W. T., Devi, I. V., & Hsiao, S. J. (2022). Early warning of impending flash flood based on AIoT. EURASIP Journal on Wireless Communications and Networking, 2022(1), 15.

Tabbussum, R., & Dar, A. Q. (2020). Comparative analysis of neural network training algorithms for the flood forecast modelling of an alluvial Himalayan River. Journal of Flood Risk Management, 13(4), e12656.

Uddin, S., Haque, I., Lu, H., Moni, M. A., & Gide, E. (2022). Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Scientific Reports, 12(1), 1-11.

V. Seal, A. Raha, S. Maity, S. Mitra, A. Mukherjee, and M. Kanti Naskar. (2012). A Simple Flood Forecasting Scheme Using Wireless Sensor Networks. International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC), vol. 3, no. 1.

Wang, Y. V., Kim, S. H., Lyu, G., Lee, C. L., Lee, G., Min, K. H., & Kafatos, M. C. (2022). Relative Importance of Radar Variables for Nowcasting Heavy Rainfall: A Machine

Zhou, C.; Yin, K.; Cao, Y.; Ahmed, B.; Li, Y.; Catani, F.; Pourghasemi, H.R. (2018). Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China.