FLOOD PREDICTION USING MACHINE LEARNING MODELS AND FORECASTINMETHODS

[1]Dr. NASINA KRISHNA KUMA, [2]ATTLA VINAY KUMAR

[1]Professor, Department of Master of Computer Applications in Audisankara College of Engineering and Technology, Gudur.

[2]PG Scholar, Dept of MCA, Audisankara College of Engineering and Technology (Autonomous), Gudur.

Received:
Accepted:

Keywords:
flood forecasting, support vector machine learning, artificial neural network.

ABSTRACT

The most well-known and deadly natural disasters of our century have been floods. The lack of an effective framework for flood forecasting has resulted in severe losses to infrastructure and human life. This has served as a reminder of how crucial it is to have a mechanism in place for predicting floods. The goal of this study is to create the best flood prediction model. All the vital aid and assistance required by the citizens and the government will be provided by AI calculations and a robust, effective, and precise flood expectation framework. The Decision Tree Model is being created as a result. With a wide range of precision, this model actualizes numerous calculations on datasets. The approach makes use of an AI computation that forecasts floods and uses an Android application to notify local and governmental agencies. Decision Tree, Random Forest, and Gradient Boost are the three machine learning algorithms that were put to the test. By utilizing a sophisticated algorithm and dealing with more complex information, this model focuses on increasing the rate of prediction.

1 INTRODUCTION

There is a plethora of ways in which water can enter typically dry terrain and cause a flood. A burst dam, rapid melting of ice, or unusually heavy rains can all overflow a river and cause it to flood the surrounding area. When an enormous storm or tsunami causes the ocean to surge inland, ocean front flooding occurs. The second-most frequent natural calamity on Earth after forest fires is thought to be flooding. Due to its role in extreme weather conditions and the rising sea level, environmental change is increasing the risk of floods globally, especially in coastal and low-lying areas. Tropical storms may travel even more slowly and drop more rain as a result of the increase in temperatures that comes with an unnatural weather change, channelings moisture into climatic streams. Therefore, it is absolutely crucial to develop frameworks that can predict floods before the rainfall ever reaches the ground. Since the turn of the century, the frequency and severity of floods have increased, making them one of the worst natural disasters to ever occur. The frequency at which the floods are growing is

anticipated to increase tremendously due to global warming.

Thus, flood prediction systems have tended to be developed in this new era of invention and technology. In every event, the information and algorithm used determine the capability and accuracy of these models.

The Organization for Economic Cooperation and Development estimates that every year, floods cost the global economy more than $40 billion in losses. In reality, the majority of countries lack effective flood warning systems. Twenty percent of flood fatalities, according to the Central Water Commission, occur in India. Bihar is the state that has been most clearly negatively impacted, with over 73 percent of its entire surface area being inundated annually. According to reports, the cost of harm to public utilities, crops, and infrastructure across all of India in 2018 could have reached 3% of the country's GDP.

Even though there are numerous techniques that might be Attempts made to prevent floods are probably the most effective and the most basic early warning system [1] uses AI algorithms. Due to heavy rains and flooding, there will be floods. When the sensor approaches diverse qualities have been linked to creativity Prepare for floods. There are currently a variety of datasets available. It can be used to build frames for expectations. Machine Learning would ensure that one is active, competent, and exact predictions.

2 LITERATURE SURVEY

In addition to T. A flood has been generated by Sasi Praba [2]. warning system for which information was gathered from distant sources satellites for sensing and ground use. The specifics amount of rainfall and the neighboring bodies of water are taken into swimming pools. Using the Gradient Boost Algorithm [3], a relationship that is nonlinear between the overall amount of rainfall and runoff, lowering the mean-squared error in the process. the datasets the Decision Tree used models that weren't included in the training dataset an algorithm is used to forecast floods using datasets that weren't utilized in the practice dataset. The structure is separated into three layers that are linked together. Three layers make up the Physical, network, and application layers are the three levels. Data is gathered by the Physical Layer from sensors and the User Interface is the Application Layer, which is the Desktop and mobile application. The historic data collection included gathered from the Flood and Meteorological Departments Commission for Forecasting. The real-time data is gathered and transmitted to the network layer every hour using the GSM Module and information stored on a cloud server are transferred to the model for machine learning.

Floods were anticipated in the Lech River basin region of Europe by Ding et al. Floods were predicted using a Spatig-Temporal Attention Long Short-Term Memory (STA-LSTM) model. This model, which is data-driven, establishes a link between the runoff and observable hydrological features. The Recurrent Neural Organization is adjusted by LTSM, which also Moulds the basic component of the neural network framework (RNN). Using the attention model eliminates the concern over how similar hydrological highlights affect various floods. The Support Vector Machine (SVM), Fully Connected Network (FCN), and the original LSTM are also used to examine the STA-precision LSTM's [5]. For the STA-LSTM Model, the test dataset's skewness is 0.0651 while the training dataset's skewness is 0.58. The FCN model performed better than the SVM, LSTM, and STA-LSTM network models when the correlation at T+3 was completed. However, the STALSTM Model performs comparably the best at T+6 and T+9, whereas FCN performs the least well.

Yuri Wu, Yukie Ding, and Jun Feng [6] conducted flood prediction tests in Changhua River using the Sparse Bayes Model. By employing the SMOTE [7] technique to solve the problem of unbalanced sample distribution and the AdaBoost Methodology to train a Sparse Bayesian model, over-fitting performance is enhanced. When compared to using a single Sparse Bayesian Model, using a group of Sparse Bayesian models achieves a high level of accuracy. It was observed that the model outperforms the single model. Additionally, it was assumed that the testing limit was a key factor in determining how well the model performed. The Changhua River Basin's annual summer flood data from 1998 to 2010 make up the dataset for the "Sparse Bayesian Flood Forecasting Model based on SMOTE Boost. "Each hour, the real-time data is recorded. The Changhua stream and rainfall, as well as the rainfall of the stations in the upper ranges of Changhua, make up the data credits.

A mixed Deep Learning (DL) and Fuzzy Logic (FL) based algorithm is used in an urban flood estimating and checking platform developed by

Karyotins et al. [8] as a part of a UK Newton Fund project in Malaysia. This model collects real-time data using inexpensive sensors. Artificial Neural Organizations (ANN) are a reliable solution for time forecasting problems, and DL [9] uses them for both supervised and unsupervised training. FL, which is based on the concept of fuzzy sets, is capable of handling "fractional truth." For the Deep Learning Model, a data window of 200 data points is the most suitable size. It was also discovered that there is a higher risk of flooding if the strength of the rain, the length of the storm, and the soil absorption are all high.

Dolo and. The technique created by al. [10] to forecast the with the use of machine learning models, flood occurrence. The Predictions of rainfall are made using data from previously collected information. the month's expected precipitation. Forecasting is possible for rain, both short-term and long-term. Data have been collected. originating with the Indian Meteorological Department. There are two independent datasets that include average rainfall information for each month and district from 1951 to 2000; the information after that is from 1901 to 2015 and includes average rainfall information for each state. IoT is used by this Low Cost IoT based Flood Monitoring System to calculate how long it would take for the flood to reach land. ML algorithms are used to estimate the rainfall's severity. Linear regression, support vector machines, and artificial neural networks are the algorithms used for the same. The several IoT devices employed include IoT Gecko, water-float sensors, and rain-drop sensors. A buzzer beeps and sends out a warning of an impending flood as the water level increases. The dataset from the previous three months is used in the linear regression model to forecast the amount of precipitation for the following month. The same is true for SVM. The CNN 1-D [11] technique is used for ANN. The linear regression algorithm's mean absolute error is 40.2467874. The mean absolute error reported by the SVM model was 90.606787. For ANN, a mean absolute error of 21.8097545 was found.

3 EXISTING SYSTEM

Detection and Warning System for Floods (Flows) [12] is the steps the Malaysian government has taking to assist halt the actual harm done to homes, streets, organizations, government agencies, and people according to the yearly floods. It aids in observing and controlling this important by providing important information like flood conditions, a strategy to the general public and the local community, including preparation authorities in the region impacted. The apparatus can determine the water level to alert both the general public and the local authorities by delivering an SMS and MMS warning in reference to the flooding circumstances. The system furthermore enables the general public and local government to observe the using an Android app, real-time water level graph data. It measures the water level using an ultrasonic sensor. The Raspberry Pi 3 acts as a server to handle and store data. utilize each microcontroller output and this information to the Raspberry Pi camera to take a picture of the flooding circumstance The microcontroller collects data, such as distance from the sea, temperature, humidity, and flood stage. Then, the responsibility for sending the data from the GSM SIM 900/900A Using the AT command, connect the microcontroller to the server. It's GSM is also in charge of sending flood alerts and warning messages. levels that have been determined from the image and sensor to the target phone's cellphone.

A computer technology called Global Flood Monitoring System (GFMS) This may be used to map the flood conditions in the area world. Robert Adler and Huan Wu from the Zooming in on a location, it is used at University of Maryland. visit the system's worldwide interactive tour to see anything of interest if the water is increasing, falling, or at flood stage. It can can also be used to determine whether an upstream rain event is occurring, whether the rain has stopped, and how. Water is flowing downhill. GFMS is available 24/7, in any when there is cloud cover or another impedance, for example. It is based on data on precipitation collected by NASA's Earth observation satellites. The quantity of water that is absorbing and what amount is feeding the streamflow are determined by combining precipitation data from GFMS with a land surface model that integrates plant cover, soil type, and terrain. Every three hours, visitors can view information for flooding, streamflow, rainfall, and water depth. Users can also enlarge the view to show inundation maps with a 1 km goal resolution.

4 PROPOSED SYSTEM ARCHITECTURE

The suggested system design is displayed in Figure 1. Labeled Training makes use of a data set. The characteristics are taken from as input, the collection of features and the training data set are provided. to the algorithm for machine learning. Figure 2 depicts the actions. contributing to the prediction model.

4.1 Report:

A report is a particular way to describe, pinpoint, and analyse problems that might occur in the wake of an incident. High rainfall with a change in weather patterns could be considered such an event, which could enhance the likelihood of flooding.

4.2 Analysis:

An ML report describes the data's accuracy. By separating the variables in the data, their dependencies are discovered and the data are analysed. These aid in a better understanding of the facts.

4.3 monitor:

To ensure the accuracy of the output from the given dataset, the data are continuously measured and the performance is tracked.

4.4 prediction:

The acquired historical dataset is used to train an algorithm, which is then used to produce an output. The prediction analysis employs the forecast methodology.

4.5 simulate:

Machine learning simulations assist us in predicting changes that have never occurred before and in obtaining possibilities that are outside the realm of possibility.

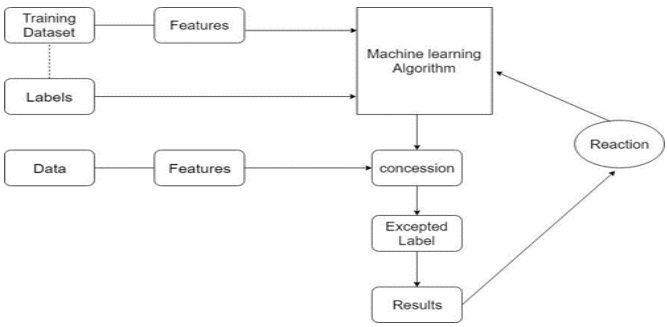

Figure1: Proposed rainfall prediction model architecture

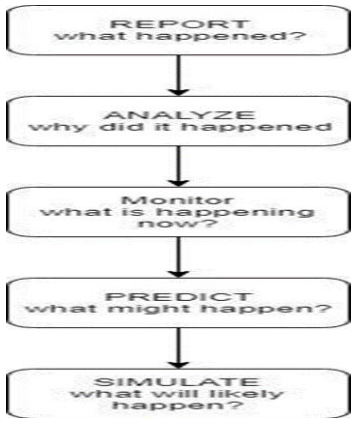

Figure 2: Prediction Methodology

5 DATASETS

In the Indian districts of Bihar and Orissa, this Decision Tree model aids in the flood forecasting process. Bihar's Patna, Shekhar, Darbhanga, Kishanganj, East Champa ran, West Chiampa ran, Gopalganj, Sitamarhi, Muzaffarpur, and Saran districts are taken into consideration. Bhadrak, Cuttack, Kalahandi, Kendra Para, Baleshwar, Koraput, Puri, Jaipur, and Sambalpur are the districts in Orissa that are similar to these. The period covered by the data collection is 1992–2002. There are 2640 data points that were gathered during this time. The India Water Portal, a reputable source, provided the dataset [13]. District-by-district downloads of the dataset in CSV format were required. Monthly data collection has been done for the dataset. Then, all of the districts had to be combined.

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUA RAINFAL
YEAR	1.000000	-0.225531	0.003879	-0.012842	0.086865	-0.059661	-0.174938	-0.223403	0.044173	0.107655	-0.030223	-0.130129	-0.123643	-0.19804
JAN	-0.225531	1.000000	0.019613	0.078626	0.034807	0.071420	0.189375	0.034423	0.008677	-0.113502	-0.035044	-0.011034	-0.089809	0.11864
FEB	0.003879	0.019613	1.000000	0.245375	0.123706	-0.083500	0.054114	0.005789	0.023259	0.066317	0.053133	-0.162880	-0.127025	0.06145
MAR	-0.012842	0.078626	0.245375	1.000000	0.074014	-0.102961	0.019000	0.018330	0.042411	0.143850	-0.023066	-0.032612	0.026292	0.11610
APR	0.086865	0.034807	0.123706	0.074014	1.000000	-0.114566	0.072990	0.014977	-0.047842	0.012928	0.113172	0.022206	-0.110392	0.11235
MAY	-0.059661	0.071420	-0.083500	-0.102961	-0.114566	1.000000	0.001235	-0.046518	-0.124412	0.116860	0.197102	0.094934	-0.118077	0.31472
JUN	-0.174938	0.189375	0.054114	0.019000	0.072990	0.001235	1.000000	0.094939	-0.014549	-0.052634	0.001156	0.015967	-0.085188	0.45340
JUL	-0.223403	0.034423	0.005789	0.018330	0.014977	-0.046518	0.094939	1.000000	0.154467	0.209441	0.025223	-0.028526	-0.013573	0.65199
AUG	0.044173	0.008677	0.023259	0.042411	-0.047842	-0.124412	-0.014549	0.154467	1.000000	0.098215	-0.181496	-0.112729	0.142090	0.41303
SEP	0.107655	-0.113502	0.066317	0.143850	0.012928	0.116860	-0.052634	0.209441	0.098215	1.000000	-0.032348	-0.027615	-0.011007	0.42834
OCT	-0.030223	-0.035044	0.053133	-0.023066	0.113172	0.197102	0.001156	0.025223	-0.181496	-0.032348	1.000000	-0.024060	-0.039067	0.20586
NOV	-0.130129	-0.011034	-0.162880	-0.032612	0.022206	0.094934	0.015967	-0.028526	-0.112729	-0.027615	-0.024060	1.000000	0.070720	0.14878
DEC	-0.123643	-0.089809	-0.127025	0.026292	-0.110392	-0.118077	-0.085188	-0.013573	0.142090	-0.011007	-0.039067	0.070720	1.000000	0.04296
ANNUAL RAINFALL	-0.198048	0.118648	0.061457	0.116103	0.112358	0.314723	0.453407	0.651990	0.413036	0.428344	0.205861	0.148783	0.042967	1.00000

Figure 3:snapshot of dataset

6 SCREEN SHOTS AND RESULTS

```
: data.isnull().sum()  # cheaking if any colomns is Left empty or not.

: SUBDIVISION          0
  YEAR                 0
  JAN                  0
  FEB                  0
  MAR                  0
  APR                  0
  MAY                  0
  JUN                  0
  JUL                  0
  AUG                  0
  SEP                  0
  OCT                  0
  NOV                  0
  DEC                  0
   ANNUAL RAINFALL     0
  FLOODS               0
  dtype: int64
```

Figure 4: finding number of missing values

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL RAINFALL	FLOODS
0	KERALA	1901	28.7	44.7	51.6	160.0	174.7	824.6	743.0	357.5	197.7	266.9	350.8	48.4	3248.6	1
1	KERALA	1902	6.7	2.6	57.3	83.9	134.5	390.9	1205.0	315.8	491.6	358.4	158.3	121.5	3326.6	1
2	KERALA	1903	3.2	18.6	3.1	83.6	249.7	558.6	1022.5	420.2	341.8	354.1	157.0	59.0	3271.2	1
3	KERALA	1904	23.7	3.0	32.2	71.5	235.7	1098.2	725.5	351.8	222.7	328.1	33.9	3.3	3129.7	1
4	KERALA	1905	1.2	22.3	9.4	105.9	263.3	850.2	520.5	293.6	217.2	383.5	74.4	0.2	2741.6	0

Figure 5: Sample Datasets

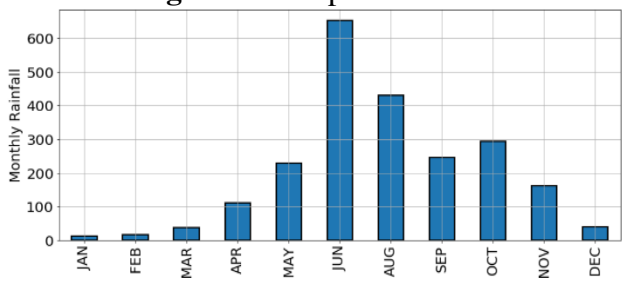

Figure 6: rainfall in Kerala all months

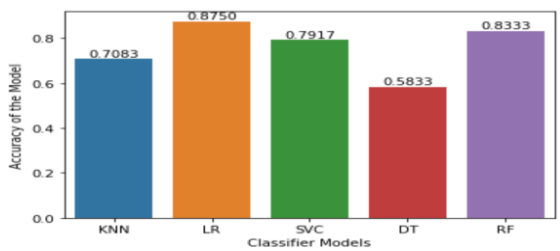

Figure 7: Comparing all the prediction models

```
In [148]: svc_scores=svc_proba[:,1]
          svc_scores

Out[148]: array([0.87402279, 0.31680621, 0.03027299, 0.46791439, 0.39919284,
                 0.1146196 , 0.76251407, 0.77067404, 0.9659059 , 0.82489191,
                 0.84081485, 0.85039042, 0.95608681, 0.91659179, 0.73091789,
                 0.04533431, 0.00547164, 0.0144526 , 0.46617238, 0.5982857 ,
                 0.17931983, 0.22843243, 0.85521272, 0.58594676, 0.9444819 ,
                 0.01667973, 0.47109385, 0.51987668, 0.88530034, 0.88004345,
                 0.89619659, 0.78253182, 0.00192442, 0.20397255, 0.9788859 ,
                 0.91245668, 0.58715733, 0.95437448, 0.00104091, 0.6431293 ,
                 0.1274714 , 0.95012418, 0.8767119 , 0.33462178, 0.01769035,
                 0.00865127, 0.78777968, 0.00619317, 0.97385357, 0.74955443,
                 0.98757871, 0.00675502, 0.69124623, 0.00805598, 0.06748037,
                 0.70144754, 0.0261892 , 0.96104268, 0.99220093, 0.39210633,
                 0.2258943 , 0.16668526, 0.80993758, 0.12662222, 0.31527738,
                 0.09222602, 0.79722201, 0.46090118, 0.98242613, 0.93323197,
                 0.0662195 , 0.58546203, 0.01522909, 0.22218462, 0.07907861,
                 0.87769627, 0.23319018, 0.41903765, 0.89772917, 0.69034705,
                 0.39503935, 0.64196435, 0.26475539, 0.85978593, 0.99384842,
                 0.39155917, 0.92837038, 0.3331483 , 0.97764931, 0.45980528,
                 0.7276532 , 0.23657413, 0.80860683, 0.05835456])
```

Figure 8: Support Vector Classification to prediction

7 CONCLUSIONS

Figure 4 displays the rainfall information for the three months of June through September in 2016. The sample rainfall dataset is shown in Figure 5. The dataset's histogram is displayed in Figure 6. The data's distance plot is displayed in Figure 7. Figure 8 depicts the binary logistic regression pattern for the three months of data, and Figure 9 depicts the linear regression model's coefficients. Figure 10 displays the decision tree prediction model's accuracy. In this study, a straightforward machine learning system is used to forecast the likelihood of a flood with accuracy. The proposed method displays the outcomes of a flood occurring in the following year. The decision tree method produces findings that are more accurate and easier to grasp when compared to the other algorithms. For nonlinear datasets, the decision tree also generates a model. This nonlinear can be used to determine the precision of a dataset that is logistic or linear. The decision tree provides greater accuracy than other straightforward machine learning algorithms, according to the findings of the comparison.

The collected dataset contains a large number of variables, making it impossible to include them into a straightforward machine learning technique. It is possible to use a neural network to process a large amount of data, which will increase the output's accuracy. The neural network's usage of fuzzy state machines allows it to generate a variety of results with various probabilities. It can offer a more flexible and adaptive form for historical datasets.

ACKNOWLEDGMENT

The writers would like to express their gratitude to everyone who has supported and guided us along the path. They would especially like to thankDr.N.Krishna Kumar**,** their project guide, for her kind assistance, direction, and support during the project's preparation and progress. She has consistently encouraged them to put up their best effort and assisted them in promptly achieving deadlines and criteria.

REFERENCES

[1] Flood Prediction Using Machine Learning Models: Literature Review by A. Moosavi, P. Ozturk, and K. Chau. 1–41 in Water, 10 (2018) Flood Detection Using Gradient Boost Machine Learning Approach.

[2] Felix, A., Sasi Praba, T. Pages 779–783 of the book International Conference on Computational Intelligence and Knowledge Economy (2019).

[3] Diffusion gradient boosting for networked learning by Ying, B. and Sayed, A. Pages 2512–2516 of IEEE International Conference on Acoustics, Speech, and Signal Processing (2017).

[4] Spatig-Temporal Attention LSTM Model for Flood Forecasting. Ding, Y., Zhu, Y., Wu, Y., Feng, J., Cheng, Z. IEEE Smart Data (Smart Data),

IEEE Cyber, Physical and Social Computing (Spescom), and IEEE Green Computing and Communications (Green Com) together with the International Conference on Internet of Things (things), pp. 458-465 (2019).
[5] Long short-term memory, Hochreiter, S. and Schmidhuber, J. 1–32 in Neural Computation, 9(8).
[6] Sparse Bayesian Flood Forecasting Model Based on SMOTE Boost by Wu, Y., Ding, and Feng. IEEE Green Computing and Communications (Green Com), IEEE Cyber, Physical and Social Computing (Spescom), and IEEE Smart Data (Smart Data), International Conference on Internet of Things (things), pp. 279–284. (2019).
[7] Synthetic minority over-sampling technique (SMOTE) by Chawla, Bowyer, Hall, and Kegel Meyer. 321-357 in J. Arif. Intell. Res (2002).
[8] Deep Learning for Flood Forecasting and Monitoring in Urban Environments: Kariotis, C., Maniak, T., Doctor, F., Iqbal, R., Palade, V., Tang, R. Pages 1392–1397 of the 18th IEEE International Conference on Machine Learning and Applications (2019).
[9] Deep learning Lacuna, Y., Bengi, Y., and Hinton, G. 521(7553), pages of Nature (2015).
[10] Low Cost IoT based Flood Monitoring System Using Machine Learning and Neural Networks: Rani, D., Dr. Jayalakshmi G. N., and Baligad, V. In: Innovative Mechanisms for Industry Applications: Proceedings of the Second International Conference, pp. 261-267 (2020).
[11] Alawi, S., Mohammed, T., and Al-Sawi, S.: Convolutional neural network understanding. Pages 1–6 of the International Conference on Engineering and Technology (2017).
[12] Flood Detection and Warning System: Marzuki, S., Sidik, M., Nasir, H., Zainol, Z., and M. Ismail (Flows). Pages 1–4 of the 12th International Conference on Ubiquitous Information Management and Communication (2018).
[13] Data Finder on the India Water Portal. (2005) India Water Portal Datafinder.
[14] Ad boosting neural networks: Application to online character recognition. Schwenk, H., and Bengie, Y. [14]. Pages 967-972 of the 7th International Conference (1997).
[15] Flood Forecasting System Based on Integrated Big and Crowdsource Data by Using Machine Learning Techniques Puttinaovarat, S., and Horkey, P. Access IEEE 8, 5885-5905 (2020).
[16] Ranit, A., and Durga, P.: Machine Learning-Based Flood Forecasting. 166–169 are included in the publication Proceedings of the Fourth International Conference on Communication and Electronics Systems (2019)
[17] Smote boost for Regression: Improving the Prediction of Extreme Values by Moniz, Ribeiro, Cerqueira, and Chawla Data Science and Advanced Analytics, IEEE 5th International Conference, pp (2018).
[18] Abdullahi, S., Habeeba, M., and Malik, N.: On-the-go Flood Disaster Warning System. Pages 258–263 of the 7th International Conference on Computer and Communication Engineering (2018).
[19] A Flood Forecasting Model based on Deep Learning Algorithm by Integrating Stacked Autoencoders with BP Neural Network Liu, F., Xu, and Yang, S. Pages 58–61 of IEEE's Third International Conference on Multimedia Big Data (2017).
[20] Urban Pluvial Flood Forecasting Using Open Data and Machine Learning Techniques in the Pattani Basin by Nominee, Nikitin, and Kaluzny. Pages 288–297 of 6th International Young Scientists Conference on HPC and Simulation (2017).
[21] Menon, K., and Kala, L.: Mobile app for flood alarm and video surveillance system for real-time flood detection. pp. 515–519 in International Conference on Computing Methodologies and Communication (2017).
[22] Bayesian flood forecasting methods: A review, Han, S., Coulibaly, P. Hydrology Journal, 551, 340-351 (2017).
[23] 15. Modeling for flash floods using evolutionary predictive theory, Segre tier, Collard, and Clergue. 844–851 in IEEE Congress on Evolutionary Computation (2013).

NOMENCLATURE

Subscripts	
Ff	Flood forecasting
Svm	Support vector machine
Ann	Artificial neural network