Proceedings of the 6th International Conference on Civil Engineering for Sustainable Development (ICCESD 2022), 10~12 February 2022, KUET, Khulna, Bangladesh (ISBN-978-984-35-1972-6)

FLOOD PREDICTION USING MODERN TECHNOLOGICAL APPROACHES: A REVIEW

Jannatul Tanjil[1], Mr. Md. Ajwad Anwar[2], Nazmus Salehin Asif*[3], Sagor Ghosh[4]

[1]Student, Department of Civil Engineering, Ahsanullah University of Science and Technology, e-mail: tanjil.tapty@gmail.com
[2]Assistant Professor, Department of Civil Engineering, Ahsanullah University of Science and Technology, e-mail: ajwad.anwar19@gmail.com
[3]Student, Department of Civil Engineering, Ahsanullah University of Science and Technology, e-mail: asif143.bd@gmail.com
[4]Student, Department of Civil Engineering, Ahsanullah University of Science and Technology, e-mail: sagorghosh10@gmail.com

***Corresponding Author**

ABSTRACT

In the field of hydrology, floods are a topic of study, and they pose a serious threat to agriculture as well as civil engineering and the public health sector. When used in the meaning of "flowing water," it can also refer to the tide's inflow. Flooding can be caused by the overflow of water from rivers, lakes, and oceans. Floods have wreaked havoc on people's lives and property in recent decades. Escaping quickly depends on an early warning system that can foresee floods. As a result of global flood predictions, a variety of technical methodologies have been used (such as AI machine learning, GIS, remote sensing). So, different methodologies necessitate distinct data sets. For example, a machine learning-based prediction system involves rainfall, humidity, temperature, water flow, and water level. Flow rates, typical temperature ranges, cloud visibility data, soil type, land cover categorization, and other variables are needed for remote sensing and GIS-based prediction. Flood prediction technology is examined in this review paper. The performance comparison of technological approaches provides a detailed grasp of the various strategies within the context of a thorough review and discussion. Using high-precision technology tools, it is hoped that the communities living near wetlands would be better served. Also, it's important to come up with concrete advice in the Bangladesh's context on how to improve the precaution by using flood prediction. When picking the appropriate technological procedure for a given assignment forecast, hydrologists and civil engineers can utilize the results of this study as a reference. This study has also, by comparative analysis tried to show that why machine learning (ML) is better to use rather than other approaches. Initially stating that this model is suitable for any region around the world.

Keywords: Hydrology, Flood-Prediction, ML, AI, Quantum computing.

1. INTRODUCTION

Flooding is a common occurrence in flood-prone wetlands. Over the last century, it has become one of the greatest major causes of suffering for the residents of flood-prone regions. Floods have wreaked havoc on human lives and properties. Timing is crucial when it comes to evasion, which is ascertained by early warning systems (Al Qundus et al., 2020). As a result, it is necessary to use innovative approach, mostly technological, to forecast the flood, which will help improve human existence on a massive scale. Furthermore, machine learning, remote sensing, and geological data management may aid in the understanding of hydrology and the application of its features to detect early warning signs. Flood risk prospective mapping is necessary for flood mitigation and management (Rahmati et al., 2016).

There has been significant research conducted across the world to assess the possible outcomes and potentials of these advanced methods. A study conducted in Brisbane, Australia evaluated the influence of individual predictor variables on flood vulnerability mapping and their significance in the development of accurate modeling of potential flood areas using machine learning. However, it was concluded that the inclusion of additional factors in the modelling work does not necessarily require better precision (Tehrany et al., 2019). Another research performed in Iran found that the analytical hierarchical process (AHP) and geographic information system (GIS) techniques are capable of making consistent and precise predictions for flood hazard, particularly in no-data areas (Rahmati et al., 2016). Flood forecast models demonstrate a large correlation between both the processing variables and flood outcomes (Mitra et al., 2016). The findings demonstrate that the deep convolutional neural network can be used effectively and quickly for flood prediction based on monsoon variables only prior to flood occurrence (Sankaranarayanan et al., 2020). Another research in Kuwait used a wireless sensor network decision technique to identify flood damage by comparing weather patterns conditions at a given place to historical data using compressed air, wind direction, water level, temperature, and humidity. In making binary decisions (flood or no flood), the precision was 98 percent (Al Qundus et al., 2020). As a result, the assessment of flood prediction varies from place to place, model to model, and, most importantly, variable to variable.

The World Resources Institute (WRI) examined which nations have the maximum percentage of the total GDP impacted by flood events on average annually, and Bangladesh comes in second after India, with 5.4 million USD impacted per year. Bangladesh must take the appropriate measures to preclude this economic threat, as well as the haphazard situation, from severely affecting people's lives. In 2021, the Ministry of Water Resources will implement a digital weather prediction and warning system to improve disaster risk management in Bangladesh. Bangladesh Water Development Board (BWDB) and a2i, with aid from Google, devised this enhanced flood forecasting and warning system under the governance of the Ministry of Water Resources. The prediction is provided by Google via push notifications on mobiles. In 2020, one million smart alerts were sent to flood-affected residents via 300,000 Android phones. In this study, modern strategies will be assessed and a viable method for developing and underdeveloped Countries will be recommended in a more convenient manner.

2. BACKGROUND STUDIES

2.1 Machine Learning and Deep Neural Network

Machine learning is a data science technology that has an emerging future in every aspect. If the skeleton of machine learning is observed it is a statistical solution to a concerning problem. In the machine learning process a certain amount of data is usually given to train a model. The model is then evaluated by statistical approaches and then machine learning can provide a solution by observing the evaluation of data sets. Deep neural network (DNN) is an extended form of machine learning that employs multilayer machine learning approach. DNN mainly deals with unstructured data to provide a useful output. To predict flood multiple algorithms of ML and DNN are used in various researches. Those approaches will be briefly discussed here.

2.1.1 SVM

SVM stands for “Support Vector Machine”. It's a machine learning-based data analysis algorithm. The major application of SVM are for regression and classification of specific data sets (Sankaranarayanan et al., 2020). It is, after all, a supervised learning method based on the notion of structural risk minimization (Bhaduri et al., 2008; Jebur et al., 2014; Yao et al., 2008, Tehrany et al., 2019). The classification hyperplane is constructed in the middle of the biggest margin by SVM, which identifies the widest separation between two classes (Pradhan, 2013; Tehrany et al., 2019). A hyperplane or a spatial line is used to design the binary SVM method while classifying (Sankaranarayanan et al., 2020). A point is classed as +1 if it is above the hyperplane, and as -1 if it is below the hyperplane. The training ints that are closest to the ideal hyperplane are known as support vectors. SVM aims to discover the optimal hyperplane separation between flood and non-flood data in the training set (Tehrany et al., 2019). The process of SVM modeling is described step-by-step in Tehrany et al., 2014.

2.1.2 Naïve Bayes Algorithm

The Naive Bayesian classifier is based on Bayes' theorem, with the predictors being assumed to be independent. Because it does not involve iterative parameter estimation, a Naive Bayesian model is straightforward to build and is especially helpful for huge datasets. The Bayes' theorem can be used to calculate the likelihood of an event occurring, assuming the probability of another event that has already occurred. In simple terms, the algorithm seeks to predict the probability of another event occurring given the occurrence of one. In this instance, the first occurrence might be considered evidence. The event's prior probability is the probability of the second event prior to the finding of evidence. The data suggests that the first case involved an attribute value of an unknown instance. Now is the time to make a naive assumption about the Bayes' theorem, namely that the features are all independent of one another. After that, the evidence will be broken down into its constituent parts. After that, the classifier will be given a model to determine the likelihood off a certain input set that contains all possible values for the class variable, and the output with the highest probability will be picked. Finally, the conditional probability and class probability will be calculated (Sankaranarayanan et al., 2020).

2.1.3 K-Nearest Neighbor

KNN is a non-parametric algorithm used for classification and regression. The input is the closest significant workspace, and the output is decided by the classification or regression of the KNN algorithm (Sankaranarayanan et al., 2020). The K-NN approach assumes that the new case/data and old cases are similar and places the new case in the most similar category to the existing categories. It saves all previous data and groups new data points into categories based on their similarity. As a result, using the K-NN approach, new data can be quickly sorted into a suitable category. It's also known as a lazy learner algorithm since it doesn't learn from the training set straight away; instead, it saves the dataset and uses it to classify it later. During the training phase, it merely saves the dataset, and when it receives new data, it categorizes it into a category that is extremely similar to the new data. It is used for the Classification problems.

2.1.4 Decision Tree

DT is a predictive modeling technique that finds and describes structural trends in data using tree topologies. It is not necessary to have a pre-existing relationship between the input variables and the goal variable (Saito et al., 2009; Tehrany et al., 2019). Data may be explained and used in a predicted manner using DT (Witten et al., 2002; Tehrany et al., 2019), it may also accept data generated at different sizes using only non-linear relationships and no frequency distribution assumptions (Kheir et al., 2010; Tehrany et al., 2019). DT is used as a rule-based technique in two areas: data classification and predictive modeling (Bhaduri et al., 2008; Murthy, 1998; Tehrany et al., 2019). Based on susceptibility levels, DT organizes and categorizes conditioning factors into hierarchical and homogeneous categories. The purpose of building a tree is to come up with a set of decision rules that can be used to predict the outcome based on a set of input factors (Debeljak and Džeroski, 2011; Tehrany et al., 2019). As a result, the rules are developed by examining a set of factors in order to predict an event from a similar set of data (Myles et al., 2004; Tehrany et al., 2019).

2.2 Artificial Neural Networks

An artificial neural network (ANN) is a computer program that simulates the human brain by simulating neurons, that are similar to their biological counterparts in the brain. This has application forms in pattern learning, such as having trained the automated system with a relevant information set and using it to predict future events (Wang, 2003; Haga et al.,1996; Mitra et al., 2016). An ANN is trained to provide accurate results for a particular problem. The feedback data and produce morals are nourished into the ANN, and the weight values for the interconnection are designated at random. The ANN weights are adjusted of the neurons until they generates the appropriate result for the set of input data given to it. As a result, the number of data points affects an ANN's precision more than the set of variables. The connectivity weights are the process by which the ANN learns the solution to these problems (Haga et al., 1996; Mitra et al., 2016). An ANN's basic implementation consists of three layers: an input nodes, a hidden units, and an output vector. The weight of the interconnection between the 3 levels determines how the layers work properly. A neural network can be trained either under supervision or unmonitored. Another concern for an ANN model is lowering the error measure, which is typically given by the mean squared error and is a reliable way of verifying the training phase (Mitra et al., 2016).

2.3 Remote Sensing

To quantify the increase in the water level, a wireless sensor network (WSNL) premised on LoRaWAN is used. WSNL employs the recipient modulation feature principle known as Ask, wherein the senders are Raspberry Pis equipped with sensors (referred to as endpoints) and the receiver is an Arduino (referred to as the coordinator). A group is made up of each end - point and its co-ordinator. Because each End - point is treated as a separate terminal, the circumstances and environment conditions of the each terminal are entirely independent of any other station, which indicates that each sensor must always be calibrated using statistical readings. Readings are taken out of each sensor for initializing, and the mean, variability, and standard deviations are determined by calculating and applying. These initializing values have been used to establish a threshold level for the sensor terminal and completely remove any severe readings that each sensor can produce. Then, each endpoint produces readings from five sensors and reproduces them for a short period of time to form a data packet. The produced data packet is sent through LoRaWAN client machine, where LoRa modulates and encodes the data using its own instrumentation function (ASK). LoRa transmits data over an RF stream at a particular frequency. After every Endpoint sends its data packets to the LoRaWAN server station, they are filtered further on the LoRaWAN server station. To maintain the rec The Raspberry Pi sends Python code to the Google Weather API to seek two more values, which is then added to the endpoint measurements. These are all the pressure difference at sea level and the rainfall valuation for the sector where the existing endpoint is positioned. Some measurements are overlooked or corrupted during interaction phases over the radio wave channel and/or serial channel, likely to result in NULLs in the data set. These are rectified by using data pre-processing and a Z-score design for standardization, which enables the sequence of flooding to be predicted or recognized (Al Qundus et al., 2020). Cognition of data packets from every endpoint, the LoRaWAN server co-ordinator needs to perform demodulation and decrypting.

2.4 GIS

The function of geographic information systems (GIS) in disasters is wide support in critical life-saving measures introduced by rapidly modernizing countries all around the world. Numerous GIS-based disaster management concepts and infrastructures were proposed by researchers, scientific experts and engineers from around the world. In the event of a flood, the implementation of GIS emergency aid authority can rescue the lives of thousands of disaster-affected before or after the occurrence of this severe situation.

2.4.1 MCDM

Every conditioning component in the Multiple-Criteria Decision-Making (MCDM) framework was first transformed to a stretched raster format with the needed pixel resolution. In ArcGIS 10.2, the spacing measure can be used to calculate the distance from a river and other variables. Following that, every factor has a few rows and columns, resulting in a single matrix. For the purpose of providing data for flood vulnerability surveying, all necessary actions for each framework are followed. Finally, in ArcGIS 10.2, all of the factors are normalized between 0 and 1 using the appropriate fuzzy membership function. This strategy eliminates the possibility of the evaluation process becoming a source of bias (Khosravi et al.,2019). The Vise Kriterijumska Optimizacijaik Ompromisno Resenje (VIKOR) process, developed by Duckstein and Opricovic (1980) to enhance complex systems, is among the most widely known MCDM methods. It determines the negotiated settlement ranking list and solution, as well as the weighting of stability interims. This method utilizes a multicriteria ranking index to find the way to solve a problem that is closest to the perfect (Opricovic and Tzeng, 2004), and the alternative solutions are assessed using all standardized guidelines (Chitsaz and Banihabib, 2015; Opricovic and Tzeng, 2004; Khosravi et al.,2019). The TOPSIS method, developed by Ching and Yoon (1981), is predicated on the distance function between decision-making choices (Ameri et al., 2018; Chitsaz and Banihabib, 2015). It was designed to solve decision-making issues with competing and incommensurable requirements. The ranking of alternative solutions in this technique is based on the smallest distance from the ideal solution and the greatest distance from the negative ideal way to solve (Jozaghi et al., 2018; Opricovic and Tzeng, 2004; Khosravi et al., 2019). Fishburn (1967) developed the SAW method, which seems based on the weighting factor (Jain and Raj, 2013; Khosravi et al., 2019). An assessment rating is computed for every substitute by multiplying the weighted value assigned to that attribute's potential substitute by the weight training of relative value immediately delegated by the decision making process, then summarizing the products for all requirements (Pourkhabbaz et al., 2014; Khosravi et al., 2019). This technique has the advantage of being a commensurate change in specific of the raw data, which indicates that the comparative order of magnitude of the measured values remains constant (Jain and Raj, 2013; Khosravi et al.,2019).

3. DISCUSSION

In terms of machine learning, it appears that each research strategy was constrained by some commonly applied analyses in the reviews. However, each strategy was distinct in its ability to select designs and variables. As a result, a common modeling path emerged from all of the findings. In a flood probability study, two parameters were determined: temperature and rainfall intensity. To anticipate the flood, SVM, KNN, and Naive Bayes designs were used in this study. Relying on monsoon characteristics prior to flooding, the precision of this strategy was impressive and effective. In another study, two datasets with various parameters were used, but only two ML designs were used: SVM and DT. There were 2 types of variables in these data sets: meteorological parameters and geological parameters. The percentage of prediction was substantial. However, this method has drawbacks because it necessitates a massive volumes of information, which is challenging to provide to a developing country like Bangladesh. It also has a possibility although this method is cost efficient.

Researchers are also using remote sensing strategies to obtain access to the elegance of flood predictability. All strategies, more or less, involve the similar types of meteorological conditions, such as air density, wind speed, volume of water, temperature, and humidity. Wireless sensor nodes were used in these studies for modeling, which instantaneously analyzes the data. SVM was used in the procedure, which was carried out in MATLAB. Because it is a wireless technology that requires a perfectly functioning communications system and internet access, as well as being IoT (Internet of Things) based, it would be difficult to implement in some developing countries. It would necessitate a working communication infrastructure. Another study substituted an artificial neural network for SVM, but the parameters remained the very same. The findings were similar to and substantial in comparison to previous research.

The Geographic Information System introduced a new dimension to flood prediction. The strategies of GIS concepts used in the studies were almost identical. The input variables in those research were the same, such as the vegetation indices, topographic wetness index, stream power index, soil composition, slope, and rainfall data. Multiple criteria decision making methodologies and machine learning techniques such as NBT (Naive Bayes Tree) and NB were used to analyze the results (Naive Bayes). The results of this assessment were meaningful and demonstrated a high level of flood prediction potential. Flood vulnerability maps created as a result of these studies could be useful for preplanning of flood-prone areas not only in particular regions, but also globally.

While the first machine learning model demands few parameters and has a massive effect on flood prediction, it is suitable for developing countries such as Bangladesh. True, Google has created GIS flood prediction technology in Bangladesh in recent history. However, it places the government in a position of reliance on other organizations. That Machine learning model would aid in the development of an independent organization, as well as assisting the government in micromanaging the scheme at the root level and taking the appropriate precautions to protect people's lives from extreme risks.

4. CONCLUSION

Floods have caused widespread concern over the years because they are a natural calamity. For that reason, a proactive approach was required to save people's lives from ruination by notifying them even before flood occurred. In this regard, the most promising strategies are machine learning, deep learning, geographic information systems, and remote sensing. The use of such methods significantly improved people's lives. Our review demonstrated a significant influence on flood prediction and efficient functioning using those strategies across the world. Briefly the findings can be stated as-

- Machine learning models can predict flood with high precision using metereological data.
- SVM, KNN and NB are the ML models which are easy to generate for forecasting using limited parameters.
- Blending ML and GIS models together can not only forcast the flood but also can be used for mapping the flood-prone areas precisely.
- IoT based remote sensing techniques can be used to aware people about upcoming flood priorly.

FUTURE WORK

Flood warning necessitates the analysis of massive amounts of data usually contains several vibrant variables, such as air temperature, pressure, and density, which interact in non-trivial ways. Nevertheless, using traditional computers–even supercomputers–to develop numerical climate and weather forecast models has restrictions. Furthermore, the process of analysis weather information by traditional computers may be too slow to catch up with constantly varying atmospheric conditions. Quantum computing is a technology to improve traditional numerical methods for monitoring and forecasting weather patterns by quickly and efficiently handling large amounts of data with many factors, harnessing the computing power of qubits, and employing quantum-inspired optimization techniques. Furthermore, quantum machine training can enhance pattern recognition, which is important for understanding the flood. We think as future expansion quantum computing has massive potential in the aspect of flood forecasting.

REFERENCES:

Sankaranarayanan, S., Prabhakar, M., Satish, S., Jain, P., Ramprasad, A., & Krishnan, A. (2020). Flood prediction based on weather parameters using deep learning. Journal of Water and Climate Change, 11(4), 1766-1783.

Tehrany, M. S., Jones, S., & Shabani, F. (2019). Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. Catena, 175, 174-192.

Bhaduri, K., Wolff, R., Giannella, C., & Kargupta, H. (2008). Distributed decision-tree induction in peer-to-peer systems. Statistical Analysis and Data Mining: The ASA Data Science Journal, 1(2), 85-103.

Jebur, M. N., Mohd Shafri, H. Z., Pradhan, B., & Tehrany, M. S. (2014). Per-pixel and object-oriented classification methods for mapping urban land cover extraction using SPOT 5 imagery. Geocarto International, 29(7), 792-806.

Yao, X., Tham, L. G., & Dai, F. C. (2008). Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. Geomorphology, 101(4), 572-582.

Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Computers & Geosciences, 51, 350-365.

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. Journal of hydrology, 512, 332-343.

Saito, H., Nakayama, D., & Matsuyama, H. (2009). Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: the Akaishi Mountains, Japan. Geomorphology, 109(3-4), 108-121.

Witten, I. H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with Java implementations. Acm Sigmod Record, 31(1), 76-77.

Kheir, R. B., Greve, M. H., Abdallah, C., & Dalgaard, T. (2010). Spatial soil zinc content distribution from terrain parameters: A GIS-based decision-tree model in Lebanon. Environmental Pollution, 158(2), 520-528.

Bhaduri, K., Wolff, R., Giannella, C., & Kargupta, H. (2008). Distributed decision-tree induction in peer-to-peer systems. Statistical Analysis and Data Mining: The ASA Data Science Journal, 1(2), 85-103.

Murthy, S. K. (1998). Automatic construction of decision trees from data: A multi-disciplinary survey. Data mining and knowledge discovery, 2(4), 345-389.

Debeljak, M., & Džeroski, S. (2011). Decision trees in ecological modelling. In Modelling complex ecological dynamics (pp. 197-209). Springer, Berlin, Heidelberg.

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. Journal of Chemometrics: A Journal of the Chemometrics Society, 18(6), 275-285.

Al Qundus, J., Dabbour, K., Gupta, S., Meissonier, R., & Paschke, A. (2020). Wireless sensor network for AI-based flood disaster detection. Annals of Operations Research, 1-23.

Wang, S. C. (2003). Artificial neural network. In Interdisciplinary computing in java programming (pp. 81-100). Springer, Boston, MA.

Haga, M., Demuth, H., Beale, M., & De Jesús, O. (1996). Neural Network Design (Vol 20).

Khosravi, K., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., ... & Prakash, I. (2019). A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. Journal of Hydrology, 573, 311-323.

Chitsaz, N., Banihabib, M.E., 2015. Comparison of different multi criteria decision-making models in prioritizing flood management alternatives. Water Resources Management, 29(8): 2503-2525.

Ameri, A.A., Pourghasemi, H.R., Cerda, A., 2018. Erodibility prioritization of sub-watersheds using morphometric parameters analysis and its mapping: A comparison among TOPSIS, VIKOR, SAW, and CF multi-criteria decision making models. Science of The Total Environment, 613: 1385-1400.

Jozaghi, A. et al., 2018. A Comparative Study of the AHP and TOPSIS Techniques for Dam Site Selection Using GIS: A Case Study of Sistan and Baluchestan Province, Iran. Geosciences, 8(12): 494.

Opricovic, S., Tzeng, G.-H., 2004. Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. European journal of operational research, 156(2): 445-455.

Pourkhabbaz, H., Javanmardi, S., Faraji Sabokbar, H., 2014. Suitability analysis for determining potential agricultural land use by the multi-criteria decision making models SAW and VIKORAHP (Case study: Takestan-Qazvin Plain). Journal of Agricultural Science and Technology, 16(5): 1005 1016.

Jain, V., Raj, T., 2013. Evaluation of flexibility in FMS using SAW and WPM. Decision Science Letters, 2(4): 223-230.

Rahmati, O., Zeinivand, H., & Besharat, M. (2016). Flood hazard zoning in Yasooj region, Iran, using GIS and multi-criteria decision analysis. Geomatics, Natural Hazards and Risk, 7(3), 1000-1017.

Mitra, P., Ray, R., Chatterjee, R., Basu, R., Saha, P., Raha, S., ... & Saha, S. (2016, October). Flood forecasting using Internet of things and artificial neural networks. In 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 1-5). Ieee.