iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VII **Month of publication:** July 2023

DOI: https://doi.org/10.22214/ijraset.2023.55101

Advancing Flood Prediction: Leveraging Machine Learning for Accurate Prediction

Harshal Gawas
Information Technology, Thakur College of Engineering and Technology

Abstract: Floods, a force of nature that cannot be escaped, relentlessly strike without discrimination. Every nation, to shield itself from this calamity, strives to make adequate preparations. Countless countries endure the devastation of floods annually, suffering immeasurable losses in various aspects such as finance, environment, and human lives. Regrettably, these losses are irreparable, no matter the efforts made. Amongst the different types of floods, flash floods pose the gravest danger as they possess both the destructive force of a flood and an astonishing velocity. These flash floods occur when intense rainfall surpasses the soil's ability to absorb water. They also occur when water fills dry creeks or streams or when sufficient water accumulates to breach riverbanks, resulting in rapid surges in a short period. Flash floods can materialize within minutes after triggering incidents, leaving little time for public warning and protection. As one facet of artificial intelligence, machine learning greatly contributes to the progression of predictive systems, offering enhanced performance and cost-effective solutions. Flood prediction must be accurate and fast to mitigate its damage. Machine learning algorithms have gained popularity for flood prediction recently because of their capacity to analyse massive volumes of data and find subtle patterns. This article presents an overview of flood prediction technology utilizing machine learning, encompassing the diverse algorithms employed, the benefits and drawbacks of this technology, and the challenges that lie ahead along with potential future developments.
Keywords: Flood, Machine Learning, Flood Prediction, Support Vector Machine (SVM)

I. INTRODUCTION

Floods are among nature's most devastating natural disasters, wreaking havoc on infrastructure, crops, socioeconomic systems, and human lives. Consequently, authorities are under more responsibility to provide accurate and trustworthy flood risk globes, as well as plan for long-term flood risk management with a focus on mitigation, safeguarding, and preparedness [1]. Techniques for forecasting floods are essential for risk control and disaster preparedness.

Robust and accurate estimates can benefit the governance of water resources strategies, suggestions for policy and research, and future evacuation models [2]. Reliable approaches for long-term and short-term prediction of flooding and other hydrological events are required to avoid damage [3]. However, because climatic conditions are changeable, forecasting flood lead periods and places are intrinsically difficult.

Because of this, current large-scale flood forecasting models rely heavily on particular data and make several naive assumptions [4]. Machine learning (ML) models for flood forecasting aid in flood mitigation or prevention, as well as flood alerts. Machine learning (ML) approaches have become more popular for this use due to their minimal computing needs and dependence mostly on observational data.

Accurate projections reduce the effects of flooding. Because of their capacity to analyse enormous volumes of data and detect complicated patterns, predictive machine learning algorithms have emerged as a viable technique for flood prediction in recent years. The use of machine learning algorithms is increasingly being utilised for flood prediction because they produce more accurate and faster forecasts than previous methods. There are several benefits of the use of machine learning that can aid in flood prediction, some of which are listed below:

1) *Improved Accuracy:* The use of Machine learning algorithms may evaluate enormous volumes of data and identify complicated patterns, resulting in more precise predictions.
2) *Faster Predictions:* The use of Machine learning algorithms can swiftly process enormous volumes of data, resulting in speedier predictions.
3) *Integration of Multiple Factors:* Machine learning algorithms can consider multiple factors that contribute to floods, including rainfall measurements, river levels, and land use patterns.

Machine learning has great benefits as mentioned above but there are also some dependencies or requirements which need to be fulfilled or can affect the accuracy of the flood prediction model, few of the requirements are as follows:

a) *Need for high-quality Data:* It might be difficult to get complete, high-quality data that machine learning algorithms can use to produce accurate predictions.
b) *Risk of over fitting:* There is a chance that the algorithms may become too closely matched to the data, which could produce inaccurate predictions.
c) *Need for Regular Updates:* Machine learning algorithms need to be updated regularly as new data becomes available, which can be time-consuming and costly.

A good and accurate model can be a great help to the government to plan in planning further plans for tackling the flood situations which will help large masses and countries which have to suffer through this disaster every year.

II. RELATED WORK

The following are some research papers on predicting floods using machine-learning techniques:

1) Amir Mosavi, Pinar Ozturk, Kwok Wing Chau (2018) work review

Their paper outlines the significant advancements in machine learning (ML) over the past two decades as well as the numerous benefits of applying ML to flood prediction. By incorporating new ML approaches and hybridising current ones, researchers hope to find more precise and effective prediction models.

This paper's primary contribution is to provide cutting-edge ML models for predicting floods and recommendations for the best models.

This article explicitly discusses the literature evaluating ML models through in-depth evaluations of robustness, accuracy, efficacy, and speed in order to provide a broad overview of the various ML algorithms used in the field. This work discusses the most effective forecasting techniques to predict short-term and long-term floods. Additionally, the most significant developments for enhancing the accuracy of flood forecasting models are taken into account. The best approaches to enhance ML techniques are said to include hybridization, data decomposition, algorithm ensemble, and model optimisation.[5]

2) Nur-Adib Maspo, Aizul Nahar Bin Harun, Masafumi Goto, Faizah Cheros, Nuzul Azam Haron and Mohd Nasrun Mohd Nawi (2019) work review

The purpose of the study is to assess current machine learning (ML) methods for flood forecasting and flood forecasting parameters. Based on an analysis of prior research publications, the rating was determined. The two sections of this study are separated to accomplish this objective. Finding flood forecasting tactics that specifically use ML techniques is the first step. The next is identifying the flood forecasting parameters that were used as the model's input parameters. The major feature of this paper is to evaluate cutting-edge ML methods for flood forecasting and to list significant parameters that are used as model inputs.

This enables scientists and flood managers to investigate ML approaches for early flood forecasts using the prediction findings as a reference.[6]

3) Jannatul Tanjil, Mr. Md. Ajwad Anwar, Nazmus Salehin Asif, Sagor Ghosh (2022) work review

The study tries to reflect how Machine Learning approaches outperform other approaches in predicting flood efficiently and it also talks a lot about the different requirements for different ML models such as machine learning-based forecasting systems including precipitation, humidity, temperature, water flow, and water level. This review article examines flood forecasting techniques.[7]

4) Ioanna Bouri, Manu Lahariya, Omer Nivron, Enrique Portales Julia, Dietmar Backes, Piotr Bilinski, Guy Schumann. (2022)work review

The study is regarding providing a cost-efficient ML framework that is affordable to poor countries. Custom data analysis splits are included to support generalizability claims. Furthermore, we propose a new bidirectional LSTM architecture (2P-LSTM) and evaluate it on three basic models. Finally, we analyse models built in Africa and Asia that were not included in the caravan dataset.[8]

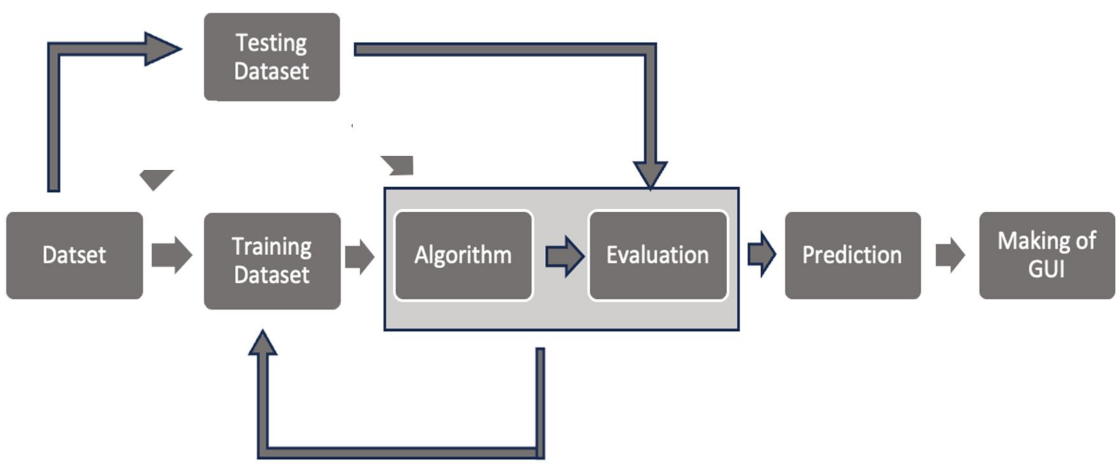
5) *Sella Nevo , Efrat Morin , Adi Gerzi Rosenthal , Asher Metzger, Chen Barshai , Dana Weitzner1, Dafi Voloshin…….[9] (2022) work review*

The study talks about how Long Short- Term Memory (LSTM) are better than available linear models. The study also covers the currently operational flood prediction systems such as one of the operational models developed by Google. In India and Bangladesh, flood warning systems are activated during the rainy season, covering flood-prone areas near rivers that cover an area of more than 470 000 sq km and are home to more than 350 million people. To the affected communities, the government, and emergency response organisations, more than 100 million flood notifications were sent.

6) *Chenkai Cai , Jianqun Wang , Zhijia Li , Xinyi Shen , Jinhua Wen , Helong Wang and Changhuai Wu(2022) work review*

To reduce the uncertainty associated with the meteorological system, the research presents a novel hybrid machine learning framework that improves quantitative precipitate forecasting by combining multi-model group and probabilistic post-processing approaches. Three distinct postprocessing strategies were used in this study to eliminate mistakes in NPFs.[10]

III. METHODOLOGY



The methodology used for predicting the flood can be understood in the above image. This is the basic methodology used in the process of making any predictive model.

The parts of the methodology are explained as follows:

1) *Data:* A dataset, which consists of several discrete data points, may be used to train an algorithm to identify predictable patterns within the entire dataset.
2) *Training Dataset:* The subset of the initial data set that is employed to train or fit the predictive model developed using machine learning is the largest (in terms of size). The ML algorithms are first fed training data so they can learn how to predict the task at hand.
3) *Testing Dataset:* A separate from the training dataset subset of the original data is the test dataset. However, it utilizes them as a benchmark to assess the model once model training is finished since they have certain comparable feature types and class probability distributions.
4) *Algorithm:* Programmes known as machine learning algorithms can predict results, find hidden patterns in data, and improve performance based solely on prior knowledge. Machine learning algorithms may be used for a variety of reasons.
5) *Evaluation:* Model evaluation is an integral part of the ML lifecycle. Data scientists can measure, interpret, and explain model performance. Speeds up model development timelines by providing insight into how and why models work the way they do.

As per the methodology used by us, the evaluation phase keeps on going till the best-suited algorithm is not founded, if the used model is not good enough then the model is changed and re-evaluation is done only after getting a satisfactory result, the predictive system is made for the well-suited algorithmic model and the GUI works proceeded.

IV. EXISTING WORK AND ALGORITHM

The existing literature on machine learning algorithms for flood prediction includes decision trees, random forests, support vector machines (SVM), and artificial neural networks (ANN). The algorithms were trained on historical data, including rainfall measurements, river levels, and land use patterns, to make predictions about future flood events. Few of this machine-learning algorithms are explained below:

A. Support Vector Machine (SVM)

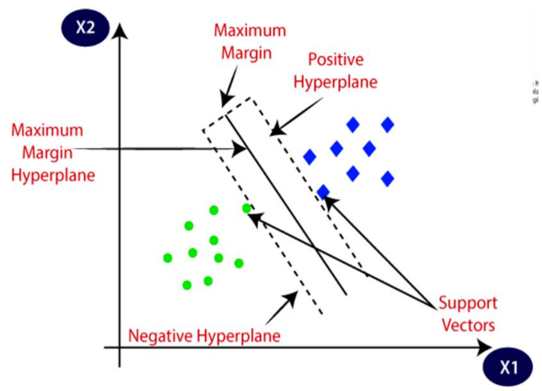


Fig 1: SVM Model

SVM is an algorithm for supervised machine learning that can be used to perform regression and classification. They are most suitable for categorization even though they are thought of as regression difficulties. The SVM method looks for a hyperplane in N-dimensional space that correctly classifies the input points. The quantity of components affects the hyperplane's size. When just two input items are present, the hyperplane is simply a line. When the total amount of input components approaches three, the hyperplane changes into a 2D plane. Once the total amount of elements reaches three, it becomes hard to visualise.[11]

B. Random Forest Classifier

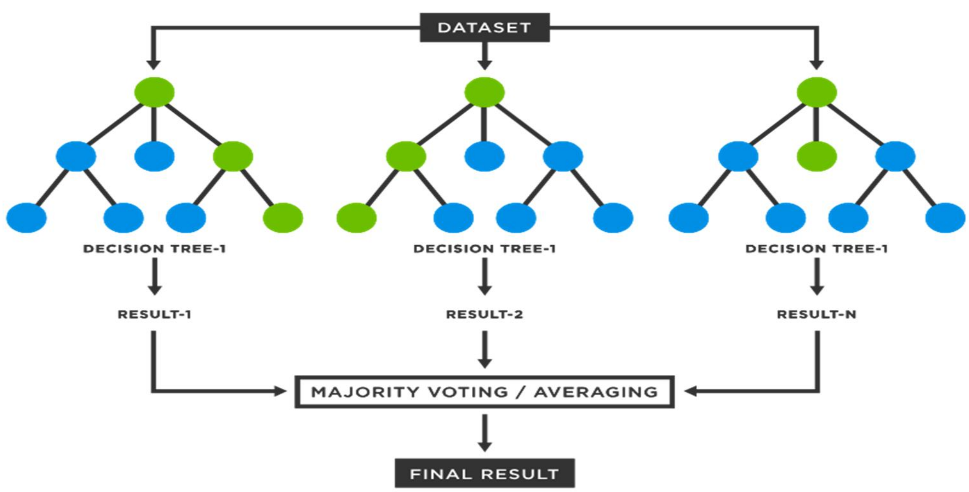


Fig 2: Random Forest Classifier

The variance of regression trees is reduced by using the random sampling approach used to select the optimum division by lowering correlation. enhances the ability of each tree in the forest to forecast. Bootstrap sampling is another method for strengthening individual tree independence.

A random forest structure is built using N decision trees combined in the first stage of the Random Forest's operation, and predictions are then made for each tree that was built in the first stage.

The following steps can be used to explain how things work:

1) *Step 1:* Take K random data points from the training set.
2) *Step 2:* Establish decision trees based on subsets of the chosen data points.
3) *Step 3:* Choose the decision tree N that you want to build.
4) *Step 4:* Repeat steps 1 and 2.
5) *Step 5:* Assign new data points to the category with the most support by checking up the predictions for the new data points in each decision tree.

Some of the existing

a) *Machine Learning Techniques:* Machine learning approaches that are often used to anticipate floods include artificial neural networks (ANN), support vector machines (SVM), decision trees, and random forests. These algorithms can examine vast volumes of data to find trends and generate precise forecasts.
b) *Remote Sensing:* Remote sensing technology such as satellites and aerial drones can be used to monitor flood-prone areas and provide real-time information on water levels, land use patterns, and meteorological conditions.
c) *Geographic Information Systems (GIS):* GIS technology can be used to create digital maps of flood-prone areas and integrate data from various sources to provide a comprehensive view of the situation.
d) *Early Warning Systems:* Early warning systems can use machine learning algorithms to predict floods based on real-time data from sensors and provide alerts to communities in advance of the flood.

V. FUTURE SCOPES

With the emergence of new technologies, the waves of new solutions have also increased that can be implemented in future to increase the accuracy of existing models or can even create better models. The following are some of the new technologies that can be used in future:

1) *Internet of Things (IoT):* IoT devices such as sensors, cameras, and drones can be used to gather real-time data on water levels, weather conditions, and land use patterns. This data can then be fed into machine learning algorithms to make more accurate predictions.
2) *Artificial Intelligence (AI):* Deep learning algorithms, for instance, may be used to increase flood forecast accuracy by analysing enormous volumes of data and discovering trends in real time.
3) *5G Technology:* With the help of 5G technology, real-time data transfer and quicker, more accurate predictions are both possible.
4) *Big Data Analytics:* Big data analytics may be utilised for analysing massive amounts of data from many sources, such as satellite images, meteorological models, and social media, to improve the precision of floods. Communities may better plan for and respond to floods by adopting this technology, which lessens their effect and helps save lives and property. However, it is crucial to remember that implementing these technologies demands substantial infrastructure investments and risks, including those related to data privacy and security.

VI. CONCLUSION

A promising technology that has the potential to revolutionise flood prediction and management is machine learning for flood prediction. The bulk of the research employed land use patterns, meteorological data, and measurements of rainfall and river levels as the data sources for training the algorithms. Accuracy, precision, and recall were the three evaluation metrics that were most frequently utilised. According to the research findings, machine learning algorithms can make predictions more quickly and accurately than conventional approaches. Compared to conventional approaches, the algorithms can anticipate events more accurately, more quickly, and with the ability to take into account a variety of flood-related parameters. However, there are several drawbacks to using machine learning for flood prediction, such as the necessity for high-quality data and the potential of overfitting. More research is needed to enhance and perfect the technology, as well as to guarantee that it is used efficiently to mitigate flood damage.

REFERENCES

[1] Danso-Amoako, E.; Scholz, M.; Kalimeris, N.; Yang, Q.; Shao, J. “Predicting dam failure risk for sustainable flood retention basins: A generic case study for the wider greater manchester area. Comput. Environ. Urban Syst.” 2012, 36, 423–433.

[2] Xie, K.; Ozbay, K.; Zhu, Y.; Yang, H, “Evacuation zone modelling under climate change: A data-driven Method”. J. Infrastruct. Syst. 2017, 23, 04017013.

[3] Pitt, M. Learning Lessons from the 2007 Floods; Cabinet Office: London, UK, 2008.

[4] Lohani, A.K.; Goel, N.; Bhatia, K. “Improving real-time flood forecasting using fuzzy inference system. J. Hydrol.” The 7th AUN/SEED-Net Regional Conference on Natural Disaster (RCND 2019), 25-26 November 2019

[5] Amir Mosavi, Pinar Ozturk, Kwok Wing Chau “Flood Prediction Using Machine Learning, Literature Review” October 2018

[6] Nur-Adib Maspo, Aizul Nahar Bin Harun, Masafumi Goto, Faizah Cheros, Nuzul Azam Haron and Mohd Nasrun Mohd Nawi ” Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review”, The 7th AUN/SEED-Net Regional Conference on Natural Disaster (RCND 2019)November 2019

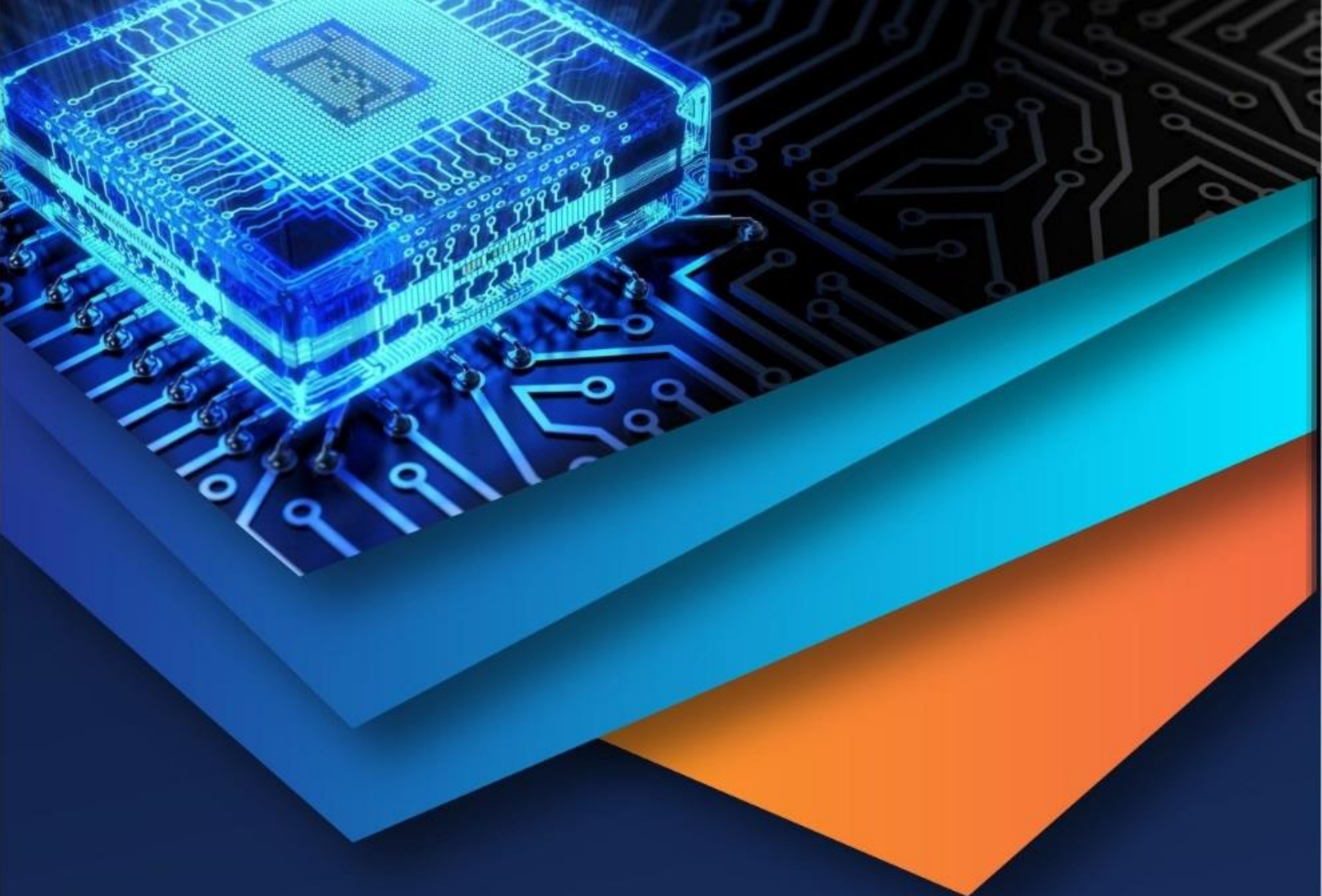
[7] Jannatul Tanjil, Mr. Md. Ajwad Anwar, Nazmus Salehin Asif, Sagor Ghosh “FLOOD PREDICTION USING MODERN TECHNOLOGICAL APPROACHES: A REVIEW”, 6th International Conference on Civil Engineering for Sustainable Development (ICCESD 2022), Feb 2022

[8] Ioanna Bouri, Manu Lahariya, Omer Nivron, Enrique Portales Julia, Dietmar Backes, Piotr Bilinski, Guy Schumann ” ML framework for global river flood predictions based on the Caravan dataset” December 2, 2022

[9] Sella Nevo, Efrat Morin , Adi Gerzi Rosenthal , Asher Metzger , Chen Barshai , Dana, Dafi Voloshin , Frederik Kratzert , Gal Elidan, Gideon Dror , Gregory Begelman , Grey Nearing1 , Guy Shalev, Hila Noga, Ira Shavitt, Liora Yuklea, Moriah Royz, Niv Giladi, Nofar Peled Levi, Ofir Reich , Oren Gilon , Ronnie Maor , Shahar Timnat , Tal Shechter , Vladimir Anisimov , Yotam Gigi, Yuval Levin, Zach Moshe, Zvika Ben-Haim, Avinatan Hassidim, and Yossi Matias “Flood forecasting with machine learning models in an operational framework” 5 August 2022

[10] Chenkai Cai , Jianqun Wang , Zhijia Li , Xinyi Shen , Jinhua Wen , Helong Wang and Changhuai Wu “A New Hybrid Framework for Error Correction and Uncertainty Analysis of Precipitation Forecasts with Combined Postprocessors” 29 September 2022

[11] https://www.geeksforgeeks.org/support-vector-machine-algorithm/



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY