# TO PREDICT THE FLOOD DURING A HEAVY DOWNPOUR WITH THE HIGHEST ACCURACY USING LR FLOOD PREDICTION MODEL.

[1] A.Jyoshna, [2] Yaidyapu Pranathi , [3] Mohammed Shujauddin, [4]Laxmapuram Harish, [5]Deshapaga Vignesh

[1] Assistant Professor, Department of Information Technology, Teegala Krishna Reddy  Engineering College Hyderabad, Telangana, India.

[1] allenkijyoshna@gmail.com

[2, 3,4]UG Scholars Students Department of Information Technology, Teegala Krishna Reddy  Engineering College , Hyderabad, Telangana, India.

[2] ypranathi07@gmail.com  , [3] shuja.dmm@gmail.com ,[4] harishreddy1318@gmail.com , [5]vigneshdeshapaga7262@gmail.com

## Abstract

Flood is one of the unfavorable natural disasters. This advancement of the flood prediction system provides cost-effective solutions and better performance. In this project, a prediction model is constructed using rainfall data to predict the occurrence of floods due to rainfall. The model predicts whether "flood may happen or not" based on the rainfall range for particular locations. Indian district rainfall data is used to build the prediction model. Machine learning methods are widely used in building an efficient prediction model for weather forecasting. This advancement of the prediction system provides cost-effective solutions and better performance. The dataset is trained with various algorithms like Random Forest Regression, Linear Regression, Lasso Regression, Support Vector Machine and Multilayer Perceptron. Among this, Lasso Regression algorithm performed efficiently with the highest accuracy of 89.40%. The LR flood prediction model can be useful for the climate scientist to predict the flood during a heavy downpour with the highest accuracy.

## 1. INTRODUCTION

The environment includes other factors such as pollutants. When the ratio of the pollutants become higher it becomes harder for the human to determine the type of weather. Environmental monitoring is important as it helps us to monitor the weather conditions. Thus, we need a model which can predict climate  change and can monitor the climate. For this reason, we need to create flood level sensing devices which will detect the water level. This system is integrated to the microcontroller board which will help to send the data each time the water reaches the threshold value. Ultrasonic sensor is used to detect the water level. In India the flood prone areas cover almost 12.5% of the country. Therefore, we need a model or a prototype that can predict the flood that might occur.

Flood forecasting techniques and Flood warning system are required different type of architectures of flood. Flood forecasting techniques able to mitigate the hazards for population and environment in real time with an early warning. Flood forecasting has been approached through rainfall – runoff and flood routing model. Flood forecast predict inflow at selected location with High Flood Level value at selected locations of river with time depends on watershed or catchment area. The dataset sources are rainfall-runoff, water levels using automatic rain gauges with satellites technology, infiltration rate etc.

Flood forecasting technique organized method suitably based on available data and an appraisal of rating criteria with an inspired performance. Flood forecasting using real time estimation gives chances of flood value in GUI. Flood estimation using Machine Learning in real time can calculate large data instantly. Comparison between flood modelling by machine learning and stochastic method gives machine learning is accurate, easy and can be applied for numbers of calculation. By using the forecasted inflow, rate of inflow in reservoir can decide the time of operation i.e., opening and closing of gate.

## 2. LITERATURE SURVEY

Muskingham equation is used for flood forecasting using combination of single and multi-inflow model form on Tar River basin. Real time flood forecast is used for Godavari River basin with improved time 12 hr as compared to conventional method with model accuracy improved by computing flood plain inundations. ANN model tool used to forecast flood. Stream flow estimation using multiple linear regression to generate Rainfall – Runoff data. Overview on different methods for Flood Forecast and Early Warning with advanced development in England, Scotland. A comparison between statistical and ANN method for flood forecasting in river Mahanadi, ANN methods are better beyond the calibration and efficient Kim and Barros modified an ANN model to improve flood forecasting short-term lead time through consideration of atmospheric conditions. They used satellite data from the ISCCP-B3 dataset. This dataset includes hourly rainfall from 160 rain gauges within the region. The ANN was reported to be considerably more accurate than the statistical models. In another similar work, Reference developed an ANN forecast model for hourly lead time. In their study, various datasets were used, consisting of meteorological and hydrodynamic parameters of three typhoons. Testing of the ANN forecast models showed promising results for 5-h lead time. In another attempt, Danso-moako provided a rapid system for predicting floods with an ANN. They provided a reliable forecasting tool for rapidly assessing floods. An $R^2$ value of 0.70 for the ANN model proved that the tool was suitable for predicting flood variables with a high generalization ability. The results of provides similar conclusions.

## 3. PROBLEM STATEMENT

There was much research for prediction of flood ahead, but not many methods give the estimate with high accuracy. The coastal cities flood loss prediction by AHP-GM-

ANN model would reduce the predictive error to get reliable results significantly the model solved the problem in nonlinear relationship variables and improved the quantitative system accuracy in the predictive method. Kartika et al. implemented the system to predict the flood using Radial Basis Function (RBF). It determines the next month's water level and daily rainfall. But it may be impossible to predict the rainfall due to climate change. As per the literature survey, authors have not combined the classification algorithm and unfocused on the accuracy, which reduces the execution time for optimization.

## 4. PROPOSED SYSTEM

The fundamental facts in the development of water logging are surface runoff and insufficient passage of water to drainage. Thus, it is essential for flood forecasting. The dataset from different sources is to form a generalized dataset  on which ML techniques are applied for extracting patterns, and to obtain results with higher accuracy. The dataset is trained along with algorithm Support Vector Machine, Logistic regression, K-nearest neighbour, and Multi-Layer Perceptron. Parameter used for calculation. The proposed work is a way to evaluate the rainfall dataset to predict the flood using machine learning techniques with higher accuracy.

Step1: The collected dataset of rainfall is preprocessed.

Step2: The dataset of rainfall is randomly partitioned into testing and training.

Step3: The dataset trained with LR, SVM, KNN and MLP algorithm.

Step4: The model is constructed using the MLP algorithm with the highest accuracy and validated with the parameters.

Step5: Input test data to the prediction model and validate the results.

## 5. DESIGN

System design is transition from a user-oriented document to programmers or data base Personnel. The design is a solution, how to approach to the creation of a new system. This is composed of several steps. It provides the understanding and procedural details necessary for implementing the system recommended in the feasibility study. Designing goes through logical and physical stages of development, logical design reviews the present physical system, prepare input and output specification, details of implementation plan and prepare a logical design walkthrough.
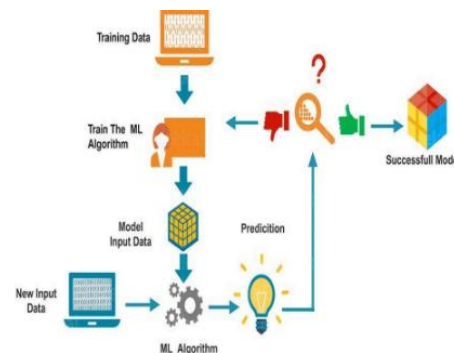


Fig 5.1 System Architecture

## 6. IMPLEMENTATION

### 6.1 SVM Algorithm

"Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well Support Vectors are simply the co- ordinates of individual observation. The SVM classifier is a frontier which best segregates the two classes.
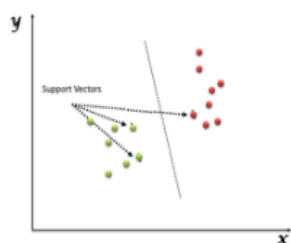


### 6.2 Lasso Regression

Lasso Regression is a machine learning algorithm that can be used to perform linear regression while also reducing the number of features used in the model. Lasso stands for least absolute shrinkage and selection operator. Pay attention to the words, "least absolute shrinkage" and "selection". We will refer to it shortly. Lasso regression is used in machine learning to prevent overfitting. It is also used to select features by setting coefficients to zero. Lasso regression is also called "L1-norm regularization". In L1 regularization, a penalty term is added to the cost function that is proportional to the sum of the absolute values of the coefficients. This encourages the model to select only the most important features and set the coefficients of less important features to zero.

### 6.3. Random Forest Regression

Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model. A Random Forest Regression model is powerful and accurate. It usually performs great on many problems, including features with non-linear relationships. Disadvantages, however, include the following: there is no interpretability, overfitting may easily occur, we must choose the number of trees to include in the model.

### 6.4 Multi-layer Perceptron

Multi-layer perception is also known as MLP. It is fully connected dense layers, which transform any input dimension to the desired dimension. A multi layer perception is a neural network that has multiple layers. To create a neural network, we combine neurons together so that the outputs of some neurons are inputs of other neurons.
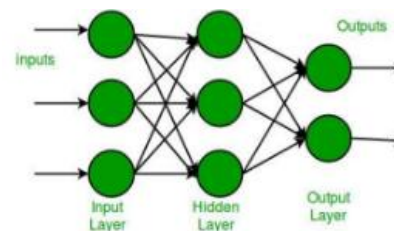
## 7. LIBRARIES

## 8. OUTPUT SCREENS

### 7.1 NumPy

NumPy is a Python package which stands for 'Numerical Python'. It is the core library for scientific computing, which contains a powerful n-dimensional array object, provide tools for integrating C, C++ etc. It is also useful in linear algebra, random number capability etc. NumPy array can also be used as an efficient multi-dimensional container for generic data. Now, let me tell you what exactly is a python NumPy array. To install Python NumPy, go to your command prompt and type "pip install NumPy". Once the installation is completed, go to your IDE (For example: PyCharm) and simply import it by typing: "import NumPy as np".

### 7.1 Pandas

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data. In 2008, developer Wes McKinney started developing pandas when in need of high performance, flexible tool for analysis of data. Prior to Pandas, Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data — load, prepare, manipulate, model, and analyze.
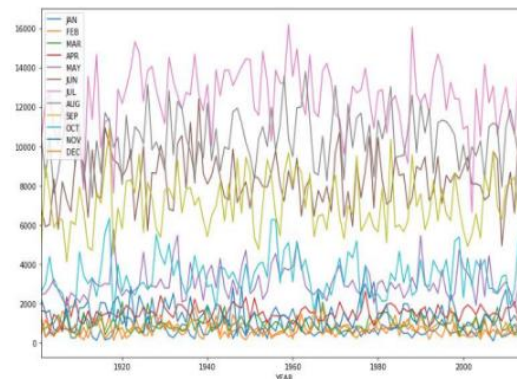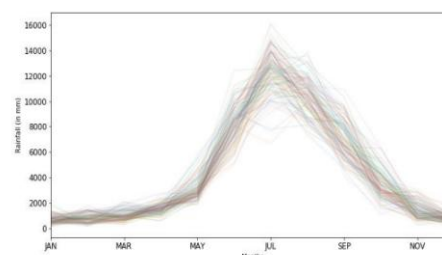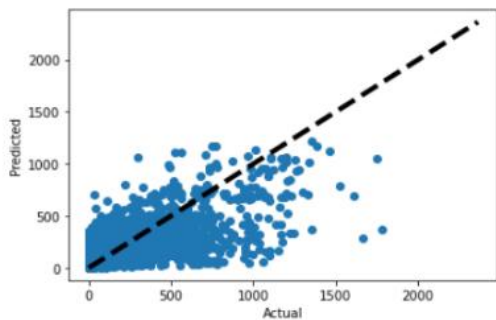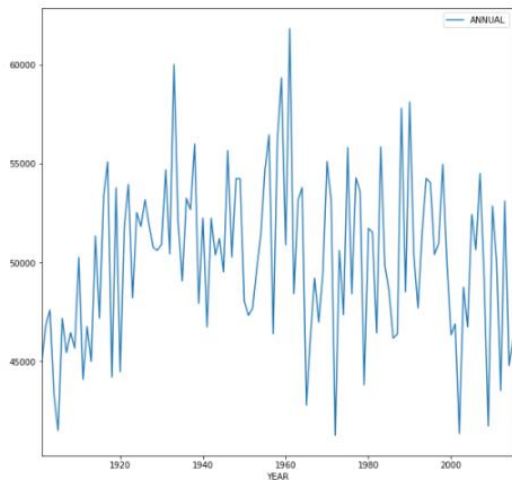


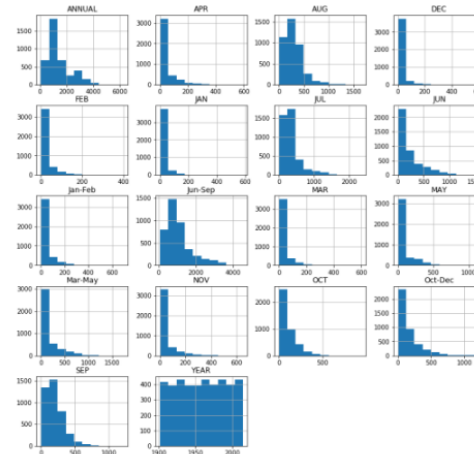6.6.1 Monthly Rainfall (1920-2000)



6.6.2 Seasonal Variation (1920-2000)
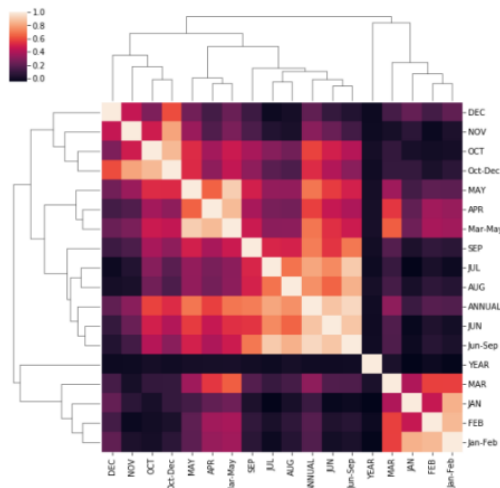


6.6.3 Seasonal Variation (in mm)

6.6.7 Loss Range of Algorithms

6.6.5 Prediction vs Actual Plot Graph



6.6.8 Histogram Graphs



6.6.6 Cluster Map (three months)

## 9. CONCLUSION

Damages that occur due to flood to living and non-living are very large. In this paper, flood prediction model is built. District wise Indian rainfall data collected between the periods of 1901 to 2015 is used for analysis. The pre-processed rainfall data was split into 70% training data and 30% testing data. The dataset is trained with Support Vector Machine, Linear regression, Random Forest Regression, Lasso Regression and Multi-Layer Perceptron. The performance factors like precision, recall, F1 score, sensitivity, specificity was calculated for each technique. The classification accuracy achieved by LR is 89.5%, SVM is 82.5%, Random Forest Regression is 80.1%, and

1807

MLP is 86.7%. Among the four techniques LR performed with highest accuracy. The LR flood prediction model predicts whether "flood may happen or not" based on the rainfall range for particular locations. This prediction model can be used by disaster management department to forecast flood.

## 10. FUTURE SCOPE

In future, we aim to use other artificial intelligence techniques to improve the prediction accuracy. The process can be automated by displaying the result of prediction in webpage or desktop application.

## 11. REFERENCES

[1] G. Corani and G. Guariso, "Coupling fuzzy modelling and neural networks for river flood prediction," IEEE Trans. on Systems Man and Cybernetics, vol.35, no.3, pp.382-390, Aug.2005.

[2] A. Luchetta and S. Manetti, "A real time hydrological forecasting system  using a fuzzy clustering approach," Computers & Geosciences, vol.29, no.9, pp.1111-1117, Nov.2003.

[3] Todini, E. (2004). Role and treatment of uncertainty in real- time flood forecasting. Hydrological Processes, 18(14), 2743-2746.

[4] Tienfuan Kerh and C.S. Lee, "Neural networks forecasting of flood discharge at an unmeasured station using river upstream information," Adv. In Eng. Software, vol.37, no.8, pp.533-543, Aug.2006.

[5] Antony, T., Raju, C. S., Mathew, N., & Moorthy, K. K. (2015). Flood Extent Analysis Over the Major River Basins in the Indian Subcontinent Using Satellite Microwave Radiometric Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(9), 4373-4378.

[6] Iqbal, N., Hossain, F., Lee, H., & Akhter, G. (2016). Satellite gravimetric estimation of groundwater storage variations over Indus Basin in Pakistan. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(8), 3524-3534.

[7] Khan, F., Memon, S., Jokhio, I. A., & Jokhio, S. H. (2015, November). Wireless sensor network based flood/drought forecasting system. In 2015 IEEE SENSORS (pp. 1-4). IEEE.

[8] D.P. Lettenmaier and E.F. Wood, 1993, Hydrological Forecasting, Chapter 26 in Handbook of Hydrology. (D. Maidment, ed.), McGraw-Hill. D.P.

[9] Adeli, H. (2001). Neural networks in civil engineering: 1989–2000. Computer-Aided Civil and Infrastructure Engineering, 16(2), 126-142.48

[10] Benoudjit, A., & Guida, R. (2019). A Novel Fully Automated Mapping of the Flood Extent on SAR Images Using a Supervised Classifier. Remote Sensing, 11(7), 779.

[11] Zhao, G., Pang, B., Xu, Z., Peng, D., & Xu, L. (2019). Assessment of urban flood susceptibility using semi-supervised machine learning model. Science of The Total Environment, 659, 940-949.

[12] Bauer, B., & Kohler, M. (2019). On deep learning as a remedy for the curse of dimensionality in nonparametric regression. The Annals of Statistics, 47(4), 2261-2285.

[13] Engle, R. F., D. F. Hendry, and J. F. Richard. 1983. Exogeneity.Econometrica 51:277-304

[14] Ruslan, F. A., Samad, A. M., Zain, Z. M., & Adnan, R. (2013, August). Flood prediction using NARX neural network and EKF prediction technique: A comparative study. In 2013 IEEE 3rd International Conference on System Engineering and Technology (pp. 203-208). IEEE. International Journal of Sciences: Basic and Applied Research (IJSBAR) (2020) Volume 49, No 2, pp 24-34 34

[15] H. M. Noor, D. Ndzi, G. Yang and N. Z. M. Safar, "Rainfall-based river flow prediction using NARX in Malaysia," 2017 IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA), Batu Ferringhi, 2017, pp. 67-72.

[16] R. Adnan, A. M. Samad, Z. M. Zain and F. A. Ruslan, "5 hours flood prediction modeling using improved NNARX structure: case study Kuala Lumpur," 2014 IEEE 4th International Conference on System Engineering and Technology (ICSET), Bandung, 2014, pp. 1-5.

[17] Q. A. Lukman, F. A. Ruslan and R. Adnan, "5 Hours ahead of time flood water level prediction modelling using NNARX technique: Case study terengganu," 2016 7th IEEE Control and System Graduate Research Colloquium (ICSGRC), Shah Alam, 2016, pp. 104-108.