



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 3, March 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.512

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Hybrid Prediction Model for Rainfall Forecasting using Machine Learning

Dr.V.P. Gladis Pushparathi¹, Keerthana C K², Renuga V³, Shree Gayathiri K⁴

Associate Professor, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai,
Tamil Nadu, India¹

U.G. Students, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai,
Tamil Nadu, India^{2,3,4}

ABSTRACT: Rainfall has been one of the most volatile aspects to analyze. Over the past few decades several machine learning algorithms have been applied to predict the rainfalls beforehand. In recent years countries like United States of America, Japan, China, Taiwan etc., were suffering from extreme and dangerous natural disasters due to impact of climate. The flood is one of the main reasons for the damages caused in the Asian countries like, India, Bangladesh, Sri Lanka, China etc. These floods increase the risk of death by 75%. In this project, several hybrid forecasting models are used that are combinations of two feature selection techniques, Gradient boosting and Random forest with various machine learning techniques, viz Support Vector Machine (SVM), adaboost, Neural Network (NN) and K-Nearest Neighbour (KNN). These model have been applied to the past 11 years (2007 - 2017) weather data to predict rainfall. The performances of these algorithms have been computed on different metrics F-score, precision, recall, accuracy. Adaptive boosting algorithm comes up with F-score of 0.9726 on testing data whereas K-nearest, SVM and neural network shows Fscore of 0.8754, 0.8045 and 0.7946 respectively. The hybrid predictive models have performed better than all other remaining models. The empirical results shown by the Adaboost algorithm is superior to remaining ones.

KEYWORDS: Rainfall forecasting, Random forest, Prediction duration, Gradient boosting, Neural Network, Support vectormachine, K- nearest neighbor.

I. INTRODUCTION

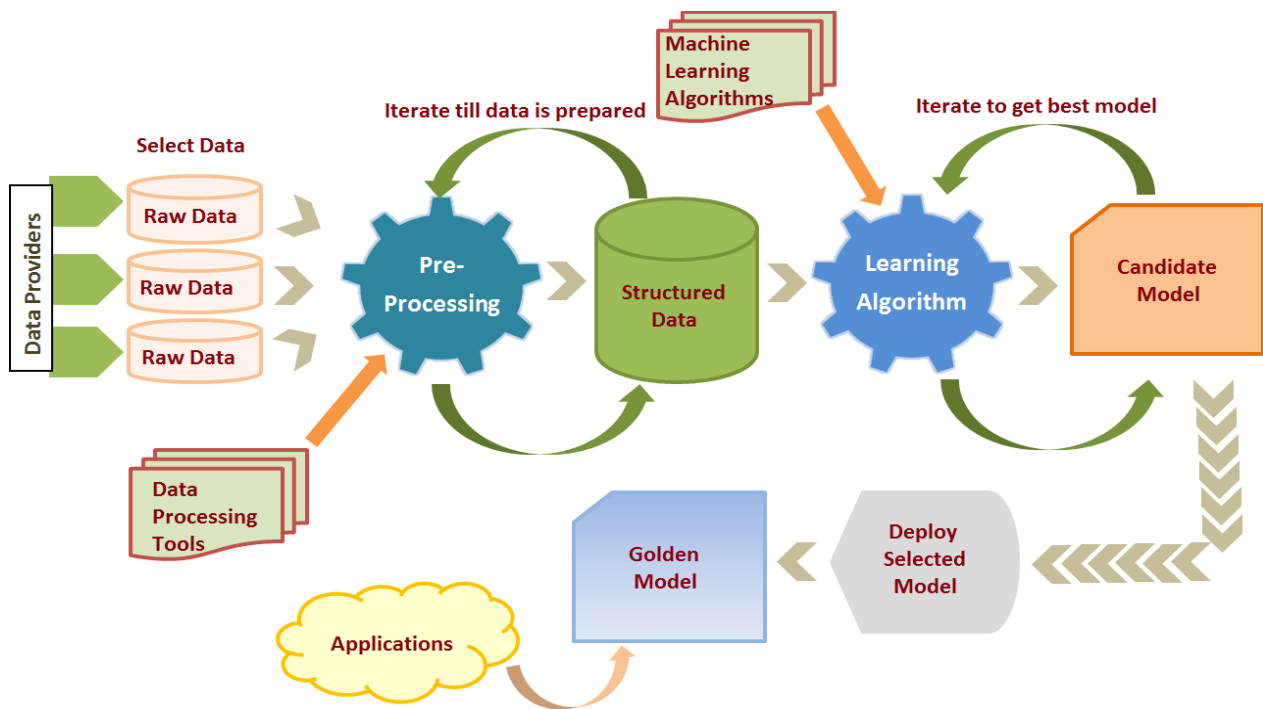
Machine learning, by its definition, is a field of computer science that evolved from studying pattern recognition and computational learning theory in artificial intelligence. It is the learning and building of algorithms that can learn from and make predictions on data sets.

In recent years countries like United States of America, Japan, China, Taiwan etc., were suffering from extreme and dangerous natural disasters due to impact of climate. The flood is one of the main reasons for the damages caused in the Asian countries like, India, Bangladesh, Sri Lanka, China etc. These floods increase the risk of death by 75%. The advancement in the information technology, the need for easy accessibility of large cloud storage and processing power is available. Data mining technologies helps us to provide reference for decision makers as summarized information even from the large amount of data. Among many data mining techniques, classification is a widely used one. Past studies proposed many techniques that could be applied to classification, such as decision trees, neural networks, Bayesian classifiers, support vector machines.

Predicting rainfall is a challenging task and especially when environment factors are in variant state due to global warming. It is multiplex phenomenon challenging problem and require expertise in multiple domain. Aftermath of this invariance cause hazardadous problems in various sectors, agri-culture sector gets effected the most. Some related problems is mentioned in . Precipitation expectation is an imperative part of climate forecasting. Exact and opportune precipitation expectation is significant for the arranging and administration of water assets, surge admonitions, development exercises and flight tasks and so on. This aims to predict rainfall few days prior and to find the cardinal attributes that play key role in rainfall. With the advancement of technology and science, specifically, the intelligent computing technology in the previous couple of decades, many rising strategies, for example, Support vector machines (SVMs) and Artificial

Neural Network (ANNs) have been broadly utilized in the rainfall prediction. The hybrid predictive models have performed better than all other remaining models. Hence, feature selection not only reduces the computational time but also augments the accuracy. All the hybrid models have chosen almost fifteen common features that have relevant contribution in rainfall prediction. Empirical findings have shown that the proposed model i.e. GB-Adaboost is superior when compared with others without feature selection.

The two feature selection techniques, viz. random forest and gradient booster conjoin with four different machine learning models: adaboost, K-Nearest Neighbour (KNN), Neural Network and Support Vector Machine (SVM) are implemented. The performance of above mentioned hybrid models is evaluated on the basis of two metrics - F1 score and testing accuracy. Results show that adaboost has better performance than remaining hybrid models. This model has the potential of approximating any non-linear function up to the arbitrary precision, without any preparatory hypothesis about data and input space.



II. RELATED WORK

In the year 2017, Natita Wangsoh, Wiboonsak waththayu and Dusadee sukawat [1] proposed a system a hybrid climate model (HCM) is a novel proposed model based on the combination of self-organizing map and analog method (AM). The main purpose was to improve the accuracy in rainfall forecasting using HCM. In combination process of HCM, SOM algorithm classifies high dimensional input data to low dimensional of several disjointed clusters in which similar input is grouped. AM searches the future day that has similar property with the day in the past.

In the year 2018, Jeongwoo Lee, Chul-Gyum Kim, Jeong Eun Lee and Nam [2] proposed a system ANN model, incorporated with lagged global climate indices, could allow for more timely and flexible management of water resources and better preparation against potential droughts in the study region.

In the year 2018, D. Khandekahakr Sachin, R. Venkata Ramana [3] proposed a system the wavelet neural network (WNN) has been widely used for modeling different kinds of nonlinear systems including rainfall forecasting. Wavelet combines



with Artificial Neural Networks (ANN) to solve different kinds of problems, especially efficient in rainfall prediction. In this paper, a model based on wavelet neural networks (WNN) decomposition is proposed as a learning tool to predict consecutivedaily rainfalls on accounts of the preceding events of rainfall data.

In the year 2019, Duong Tran Anh, Thanh Duc Dang and Song Pham Van [4] proposed a system Simulated Annealing algorithm (GA-SA) supported by Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA). Results showed both the wavelet transform and seasonal decomposition methods combined with the SANN model could satisfactorily simulate non-stationary and non-linear time series-related problems such as rainfall prediction, but wavelet transform along with SANN provided the most accurately predicted monthly rainfall.

In the year 2019, Kaveri Kar1, Neelima Thakur, Prerika Sanghvi [5] proposed a system for rainfall prediction which is one of the demanding operational responsibility carried out by meteorological services worldwide. Planning is referred to as the roadmap to success as failing to plan implies planning for failure. Information about future happenings instrumental to efficient and effective planning. Effects of natural disaster such as flooding, drought an only be prevented with effective planning. The different techniques that can be applied are as- decision tree, clustering, K-mean, fuzzy logic.

III. PROPOSED METHODOLOGY DISCUSSION

EXISTING SYSTEM

Artificial Neural Network (ANNs) has been broadly utilized in the rainfall prediction. Artificial neural network is basically a non-parametric model that can minimize the empirical risk. This model has the potential of approximating any non-linear function up to the arbitrary precision, without any preparatory hypothesis about data and input space. Neural network has the ability to model the attributes to rainfall prediction and that too without the deep knowledge of natural phenomenon. But it faces several pitfalls like nonconvex objective function and toil to find out the number of hidden layers. Moreover overfitting causes inconsistency in performance of neural networks. Major pitfall of empirical risk minimization is overfitting.

PROPOSED SYSTEM

To overcome the fallback in the existing system we propose a machine learning based system to increase the efficiency and accuracy. This project have applied several machine learning algorithms KNN, adaboost, support vector machines and Neural network to examain rainfall predictions methodically. Result analysis from this project finds that neural network has performed less efficiently among other mentioned algorithms. Performance shown by adaboost is better among others. Moreover, each machine learning algorithm has performed differently, their results are comparable but adaboost performance is found to be consistently better

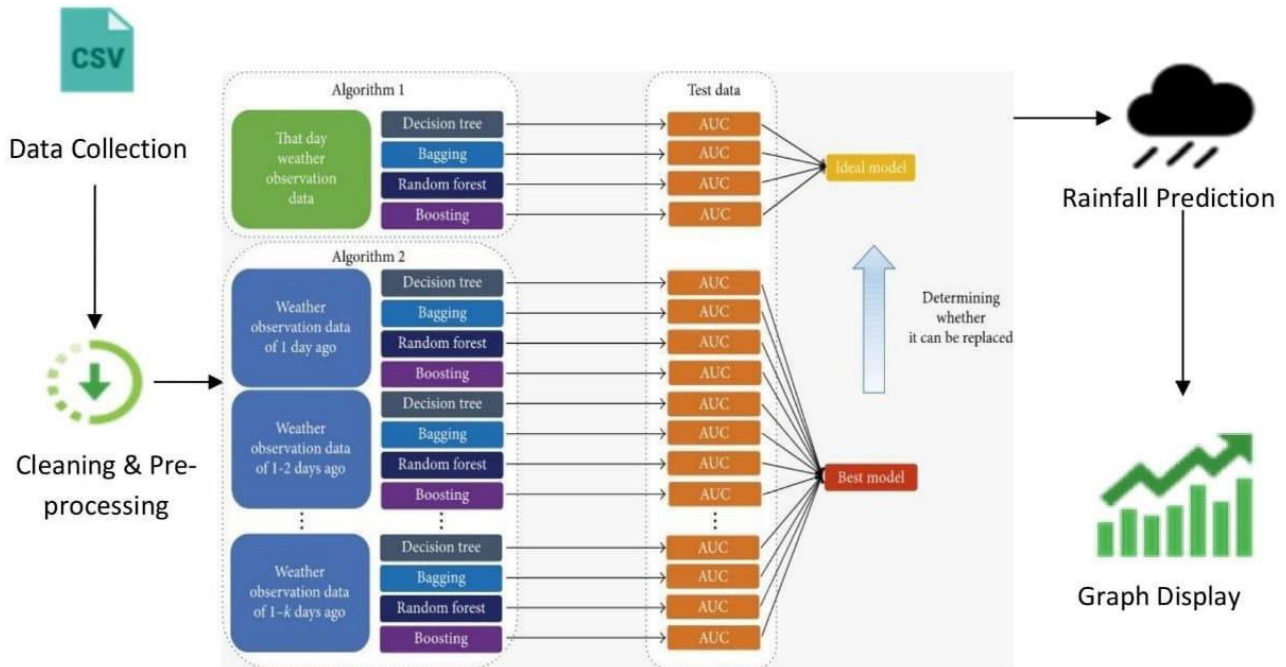


Figure 1 Architecture Diagram

Modules in the system

- **Data pre-processing**
- **Random Forest model generation**
- **Prediction, result presentation**

Data pre-processing

Weather dataset is collected in a comma separated values (CSV) file for the last three or four years of rainfall information. The dataset contains the month wise aggregation. The dataset might contain empty values, negative values or error. Dataset is cleaned in the pre-processing. The preprocessing methods involve of removing records which is not complete. Once the clean dataset is available we have to prepare it to feed to the machine learning algorithm

Random Forest model generation

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction. One big advantage of random forest is, that it can be used for both classification and regression problems, which form the majority of current machine learning systems. We train our system with dataset and create the model for future prediction

Prediction, result presentation

Advantage of random forest is that it can be used for both regression and classification tasks and that it's easy to view the relative importance it assigns to the input features. Random Forest is also considered as a very handy and easy to use algorithm, because its default hyper parameters often produce a good prediction result. The number of hyper



parameters is also not that high and they are straightforward to understand. The Random forest trained model is used to predict the rainfall for a particular month. The prediction duration is for a couple of month. Python matplotlib library can be used to produce the graphical representation of the data in a visual format. The predicted data is also plotted in the graph along with the existing history data.

IV. EXPERIMENTAL RESULTS

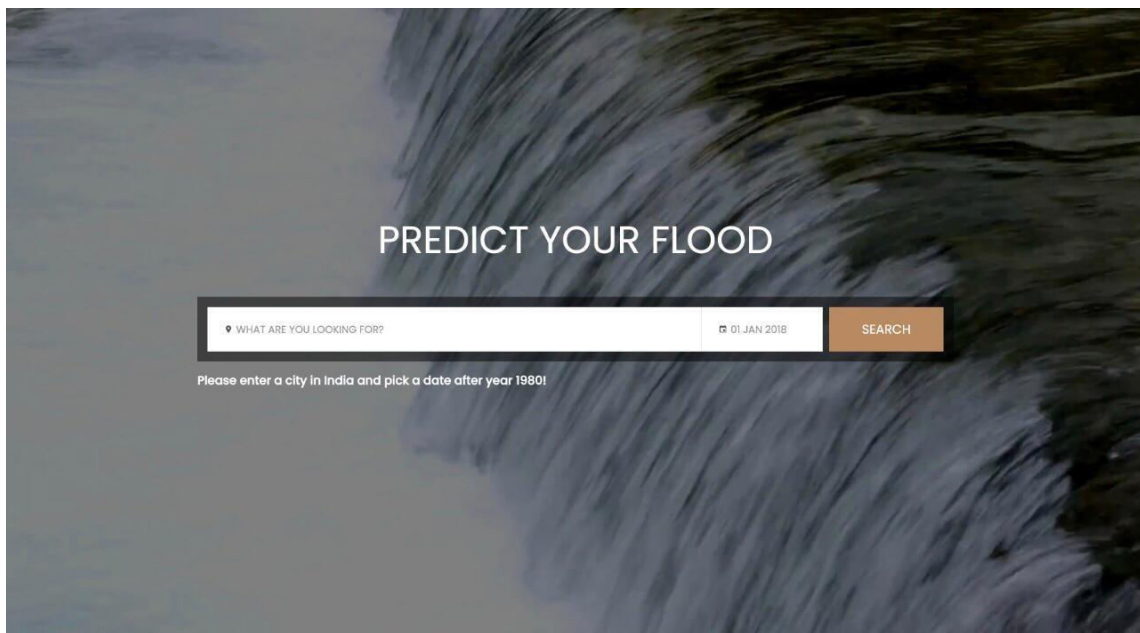


Figure 2 Home Page

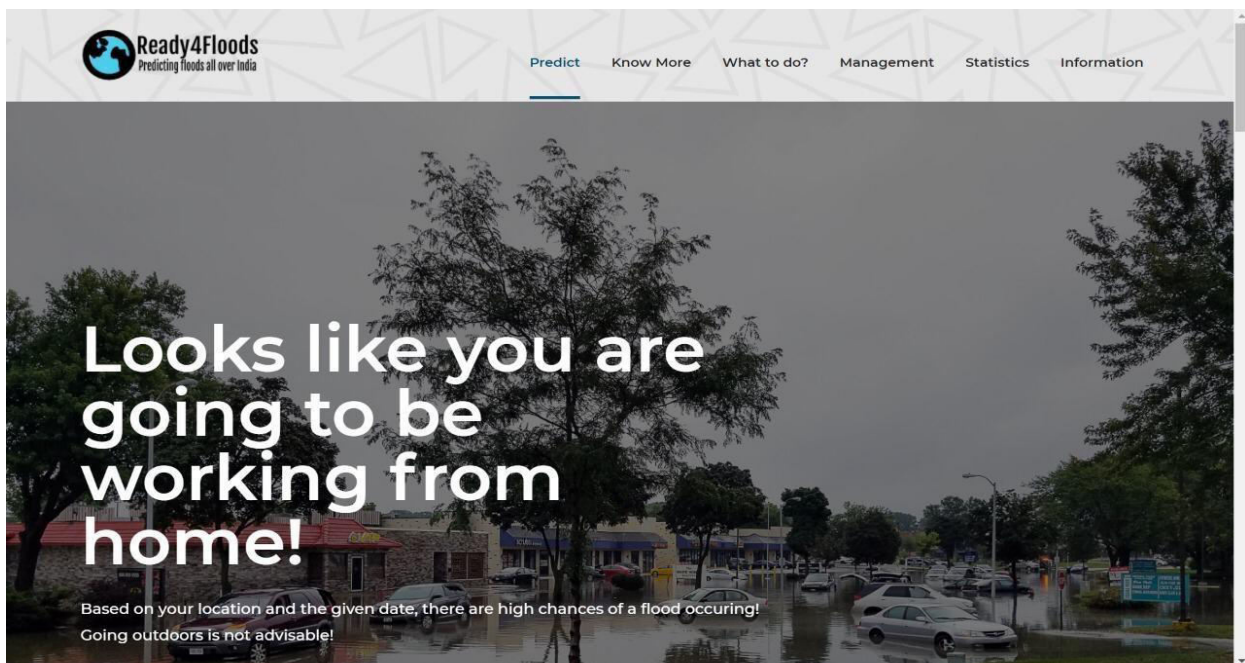


Figure 3 Prediction Result



V. CONCLUSION

In this project, various hybrid models are used, GB-SVM, GB-KNN, GB-NN, RFSVM, RF-KNN, RF-NN. For performance measurement of the above mention models, various empirical studies have been feature importance of all the features given by random forest. performed and it is measured on the basis of two metrics, viz. F-score and accuracy, here f-score has taken into account as calculating performance on the basis of accuracy can be misleading. The results have shown the classifiers (SVM, KNN) used with gradient boosting feature selection technique gave better F-score metric than the individual used classifiers. Moreover, among the two feature selection techniques i.e Random forest (RF) and Gradient boosting (GB), RF has shown better results. Empirical findings have shown that the proposed model i.e GB-Random Forest is superior when compared with others without feature selection.

REFERENCES

- [1] Q. Zhao, Y. Liu, X. Ma, W. Yao, Y. Yao and X. Li, "An Improved Rainfall Forecasting Model Based on GNSS Observations," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 7, pp. 4891-4900, July 2020, doi: 10.1109/TGRS.2020.2968124.
- [2] Schmitz, G. H., and J. Cullmann. "PAI-OFF: A new proposal for online flood forecasting in flash flood prone catchments." *Journal of hydrology* 360.1-4 (2008): 1-14.
- [3] Riordan, Denis, and Bjarne K. Hansen. "A fuzzy casebased system for weather prediction." *Engineering Intelligent Systems for Electrical Engineering and Communications* 10.3 (2002): 139-146.
- [4] Guhathakurta, P. "Long-range monsoon rainfall prediction of 2005 for the districts and sub-division Kerala with artificial neural network." *Current Science* 90.6 (2006): 773-779.
- [5] Pilgrim, D. H., T. G. Chapman, and D. G. Doran. Problems of rainfall-runoff modelling in arid and semiarid regions." *Hydrological Sciences Journal* 33.4 (1988): 379-400. *9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* 395
- [6] Lee, Sunyoung, Sungzoon Cho, and Patrick M. Wong. "Rainfall prediction using artificial neural networks." *Journal of geographic information and Decision Analysis* 2.2 (1998): 233-242..
- [7] French, Mark N., Witold F. Krajewski, and Robert R. Cuykendall. "Rainfall forecasting in space and time using a neural network." *Journal of hydrology* 137.1-4 (1992): 1-31.
- [8] Charaniya, Nizar Ali, and Sanjay V. Dudul. "Committee of artificial neural networks for monthly rainfall prediction using wavelet transform." *Business, Engineering and Industrial Applications (ICBEIA), 2011 International Conference on. IEEE, 2011.*
- [9] Noone, David, and Harvey Stern. "Verification of rainfall forecasts from the Australian Bureau of Meteorology's Global Assimilation and Prognosis (GASP) system." *Australian Meteorological Magazine* 44.4 (1995): 275-286.
- [10] Hornik, Kurt, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators." *Neural networks* 2.5 (1989): 359-366.
- [11] Haykin, Simon. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.
- [12] Rajeevan, M., Pulak Guhathakurta, and V. Thapliyal. "New models for long range forecasts of summer monsoon rainfall over North West and Peninsular India." *Meteorology and Atmospheric Physics* 73.3-4 (2000): 211-225.
- [13] Luk, K. C., James E. Ball, and Ashish Sharma. "A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting." *Journal of Hydrology* 227.1-4 (2000): 56-65.
- [14] French, Mark N., Witold F. Krajewski, and Robert R. Cuykendall. "Rainfall forecasting in space and time using a neural network." *Journal of hydrology* 137.1-4 (1992): 1-31.
- [15] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (3) (1995), pp. 273-297.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor

**Impact Factor:
7.512**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details