International Journal of Lakes and Rivers.
ISSN 0973-4570 Volume 15, Number 1 (2022), pp. 37-52

http://www.ripublication.com/ijlr.htm

Design and Implementation of Novel Rainfall Prediction Model Using Supervised Data Mining Techniques

Deepak Sharma[1], Dr. Priti Sharma[2]

[1,2]DCSA, Maharshi Dayanand University, Rohtak, Haryana, India.

Abstract

Supervised machine learning techniques are used to find consequential patterns using ample amount of data. By using these patterns, relationship between attributes can be established. These relationships are ultimately used to predict the outcome of the event under study. Rainfall plays a decisive role in the cultivation of crops. It is a fact that different crops appeal for different supply of water for cultivation but almost all of them relates with the amount of rainfall. Prediction of rainfall significantly contributes in the selection of crop for the season. Also, it can be very effectual in water management for the cultivation of crops. In this paper, a Novel rainfall prediction model based on supervised machine learning data mining techniques have been proposed for the prediction of rain fall. Cross validation technique is used to validate the models. This paper includes detailed implementation of the model, performance evaluation and comparison.

Keywords: Data mining, Artificial neural network, Bagging, Ensembles method, cross validation, Rapid miner, Rainfall prediction, Machine learning.

I. INTRODUCTION

Rainfall is a convoluted baroscopic phenomenon that depends upon many meteorological conditions which are bothersome to prognosticate. [1] In the total population of farmers, more than 82% have modest or medium-sized lands and confide in rainfall for the cultivation of crops. Different crops appeal for the different amplitude of water and precise Rainfall prediction can be a gigantic benefaction in the decision-making process. [2]

Accurate prediction also helps in preserving the groundwater which is used by farmers

in the truancy of rainfall for the cultivation of formidable water-demanding crops such as rice and sugarcane. [3]

Research in weather prediction specifically for dealing with natural hazards such as floods and storms has been in trend for the past two decades but predictions using complex machine learning techniques is an affair of a recent post.

In a postulated contour, there are three categories of weather forecasting models i.e. synoptic model, statistical model, computer model. Synoptic model is a combination of radial and postulate weather prediction methods that are used for short-term predictions. [4] This model is being used by skilled experts since mid of 20^{th} century. Different types of weather maps, radars, and satellite images are premeditated by professionals in order to bring out predictions.

The statistical model hinge on convoluted analytical equations which are ruled by the laws of physics and driven by many scientific atmospheric phantasms. The last in the list is computer modelling i.e. prediction using machines of high processing and storage capacity. [5]

With the recent developments, the processing power of machines has been increased exponentially. The storage capacity and accessing speed also increased which results in increased performance. [6] Supervised machine learning techniques are used for the dissection of a large quantity of data and to ascertain the relationship between different attributes of the data set for generating the aspired prediction parameter.

In this paper, Rainfall prediction has been proposed based on the artificial neural network which is broadly discussed in the implementation section of this paper. Meteorological data of the past 35 years of the district Hisar of Haryana, India have been procured from the National centres for environmental information (NCEI). [7] NCEI accumulate daily weather data of around nine thousand stations that were situated around the world. As per the World Weather Watch Program well regulated by the world meteorological Organization (WMO), the data of all the stations is available without any restrictions and can be used for research work.

There are in total 8600 instances in the data set having parameters as station code, Data, Temperature, Dew point, sea level pressure, visibility, wind speed, Maximum temperature, minimum temperature, and precipitation. A radical structure of any data mining or knowledge discovery process embraces data cleaning, Pre-processing and normalization, Data analysis, Evaluation of results.

Data cleaning pertain to removal of unwanted or incorrect data. Pre-processing and normalization dethrone various types of anomalies such as outliers or redundant data so that the model formed to analysis the data can evaluate it with commendable accuracy and diagnosticate the stencil inside the data. Results given by the model have been investigated by a validation method. In this research work, cross validation is used to analyse the performance of the proposed rainfall prediction model. [8]

Data mining techniques can conventionally be divided into supervised and unsupervised techniques. In supervised techniques the data set is divided into two parts one is training data set which is used to train the model and other is testing data set

which is used to validate the model but in unsupervised techniques, there is no training data available. Supervised Data Mining techniques are used to solve classification problems and regression problems. In classification problem, the final prediction result is a class label on the other side the final prediction results in regression problem is a numerical value.

In this work, artificial neural network supervised data mining techniques are used to predict the amount of precipitation based on the other weather parameters. Ensembles method is used to build the Rainfall prediction model having artificial neural network as base classifier and a total of ten ANNs are amalgamated to form the framework for high accuracy prediction.

II. LITERATURE REVIEW

The study and analysis of existing research work related to the topic under study is very importunate and must be pursued as essential part of research. Studying existing work not only point out the current advancements in that field but highlight the problems faced by the researchers and scientists in that type of research work.

K. C. Gouda et.al. [9] have proposed an approach for predictions of rainfall using artificial neural network back propagation algorithm. The value of root mean squared error (RMSE) of the proposed model is 0.300. In this research work, the data for Bangalore city of India is under study and the value of rainfall predicted by the model and value of actual rainfall of the region has been compared.

Moulana Mohammed et.al. [10] have proposed three rainfall prediction models using multiple linear regression, Support vector regression, Lasso regression respectively. The mean absolute error for multiple linear regression, support vector regression and lasso regression are 10.95, 4.35 and 11.71 respectively. The SVM is stand out to be the best in case of mean absolute error.

Deepali Patil et.al. [11] have been proposed prediction of rainfall using three supervised machine learning techniques namely multiple linear regression, artificial neural network and decision tree. All the three techniques have been implemented using Python 3.7. For this research work "Austin weather data set" has been collected from kaggle. Data set contains attributes such as temperature, humidity and rainfall etc.

III. DATA SET OVERVIEW

According to the world weather watch program structured by world meteorological organization (WMO). According to WMO Resolution 40(CG-XII), daily weather data for around 9000 stations worldwide has been collected and shared among all countries. This global data for all the station, is a product of different countries and is made freely available for the intension of research and development and other non-commercial and educational purposes. [12] The National Centers for Environmental Information (NCEI) formally known as National climate Data Centre (NCDC) provides global surface summary of day data (GSOD) [15]. It usually provides the current daily data in

24 hours after the current day. There are in total 18 meteorological elements whose values are collected and stored on daily basis which were present in the collected data set.

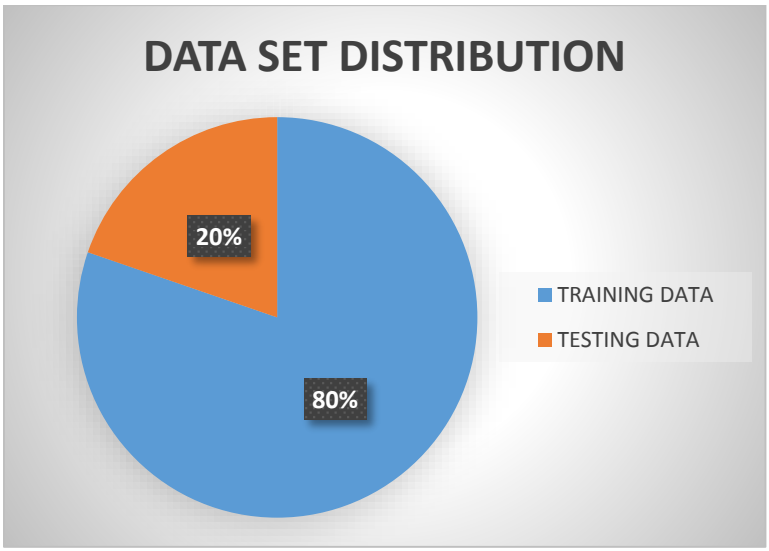

Figure 1: Data Set Distribution

Table 1: Represents data utilisation for model

S. No.	Data Set	Description	Instances
1	Training Data Set	(1988-2016) 31 years	7000
2	Testing Data Set	(2017-2021) 4 years	1720

Attribute which are present is the original collected data set are as follows:

Table 2: Attributes description of Data set

S. No.	Attribute Name	Data Type
1	Station code (STN)	Integer
2	Date (DATE)	Integer
3	Temperature (TEMP)	Numeric
4	Dew point (DEWP)	Numeric
5	Sea level pressure (SLP)	Numeric
6	Visibility (VISIB)	Real

S. No.	Attribute Name	Data Type
7	Wind speed (WDSP)	Numeric
8	Maximum sustained wind speed (MXSPD)	Numeric
9	Maximum Temperature (MAXT)	Numeric
10	Minimum Temperature (MINT)	Numeric
11	Precipitation Amount (PRCP)	Real

IV. PROPOSED MODEL

In general, there are three broad categories of predicting weather attributes i.e. synoptic modeling, statistical modelling and computer modelling which are already being discussed in the introduction section of this paper. In this research work, computer modelling approach is used. A novel rainfall prediction model is designed and implemented with the help of Rapid miner data analytical tool. [13]

In the process of fabricating this model, many supervised machine learning Data mining techniques have been critically analysed and compared on the basis of accuracy, recall, precision, RMSE (Root mean square error). [14] The comparison of different supervised machine learning data mining techniques have been shown in table.

These techniques have been validated on the data set using cross validation technique. After comparison and critical analysis of the technique, it is ascertained that artificial neural network (ANN) gives better accuracy as compared to the other techniques such as Decision tree, K-nearest neighbour, Logistic regression, linear regression. The other performance analysis parameter of ANN is better as compared to the other DM techniques.

The rainfall prediction model proposed in this research work is based on artificial neural network (ANN) and the accuracy of ANN is further improved by using ensembles method. [15] In ensemble method, more than one prediction model is used to further intensify the accuracy of the prediction model. Bagging and boosting are a type of ensemble method.

Dataset D is divided into k-set of D1, D2, D3… Dk Where used for training of classifier models M1, M2, M3… Mk respectively. [16] In ensemble method, new tuple is classified based on combine votes of all the classified based on combine votes of all the classifier models as shown in the given figure.

Ensemble method is far more superior to any individual classifier because it based on voting algorithm. When a tuple is to be classified, then every classifier predict the class label and which class label gets maximum vote is assigned to the tuple this process majority voting same as what happens in elections. Base classifiers may misclassify the tuple but ensemble will misclassify only when majority of the classifier misclassifies which is very rare. [17] In this way ensembles produce better results as compared to base classifiers. Bagging is a type of ensemble method which is used to increase the

prediction accuracy. It is also called as bootstrap aggregation because bootstrap method is used for sampling. Bagging can be understand in a simple way using example of a patient and doctor. Suppose there is a patient who want to go for a diagnosis based on his symptoms. So, instead of consulting to a single doctor he approached multiple doctors and noted their diagnosis. After consulting a number of doctor, he takes his treatment based on diagnosis given by majority of doctors. Also noted that value of diagnosis of each doctor has equal value.

Bagging also works similar, in which the Data set D having d tuples sampled using bootstrap method of "sampling with replacement". K training sets each of having d tuples are sampled from original set D and supplied to k classifiers such that set D1 is supplied to M1 and D2 is supplied to M2 and so on. Bagging technique is used for the proposed novel rainfall prediction model having artificial neural network as base classifier. After building of each classifier (M1, M2…Mk) wherever a new tuple X is present, each classifier predicts counts as one vote. After prediction by all classifier models, which class label gets majority votes will be assigned to the tuple 'X' this called as majority voting where vote of each classifier get equal value.

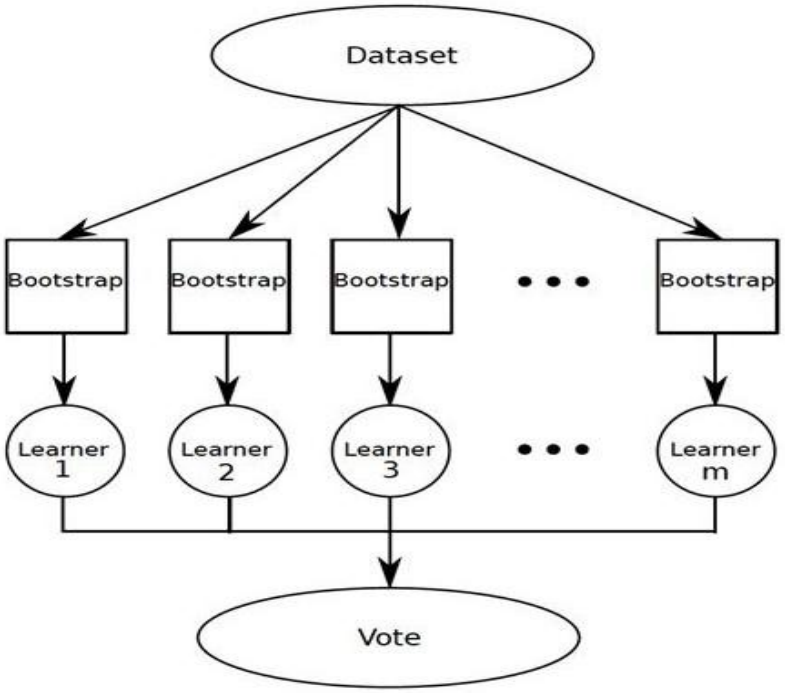

Figure 2: Bagging Ensembles Method

For this research work, Data set has been acquired from National centers for environmental information (NCEI) for district Hisar of Haryana state of India having station code: 421310. In the acquired data set, the daily summary of weather parameter such as Temperature, Dew point, Sea level pressure, Visibility, Wind Speed, Maximum sustained wind speed, Maximum temperature, Minimum temperature, Precipitation is provided.

The collected data is from the year 1988 to 2021 (35 years). It has been observed that whenever data is collected from a secondary data source embrace absurdity due to human interventions or some other factor. Also Data set contains missing values for some attributes which can be happened in case of lapse in reporting of data or incorrect measurement of attribute value.

Due to the above cited reasons it is an accustomed practice of pre-processing of collected data in order to annihilate any possible enigma in the acquired data set. Preprocessing and elimination of ambiguities leads to a better prediction model and accurate results. The process used for preprocessing of the acquired data set is shown in figure 3. In the initial phase of preprocessing the elimination of tuples or instances having outlier or missing value for any attribute is to be found. [18]

After that identification of superfluous attribute is to be aimed. Superfluous attributes are those attributes whose values have no significant relation with the attribute whose value is to be prediction i.e. Precipitation in this case. For example in this data set, station and date are two attributes which are superfluous and should be eliminated from the data set to make prediction more accurate and significant. [19] After removal of superfluous attribute, the attribute Precipitation (PRCP) is to be labeled as attribute whose value is to be predicted.

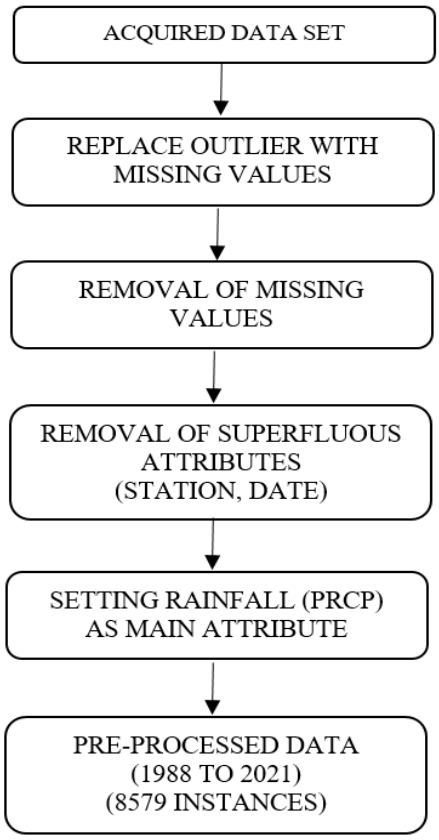

Figure 3: Pre-processing of Acquired Data Set

Data from the year 1988 to 2016 (7000 instance) is used for conformation or training of Novel Rainfall Prediction Model (NRPM) and data from year 2017 to 2021 (1720 instance) is used for analysis of the generated model by predicting the value of precipitation and comparing it with the actual value of precipitation. [21]

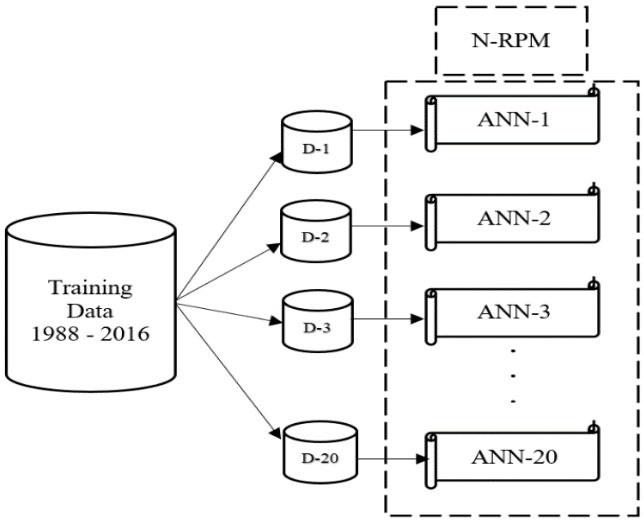

Figure 4: Training of Novel - Rainfall Prediction Model (N-RPM)

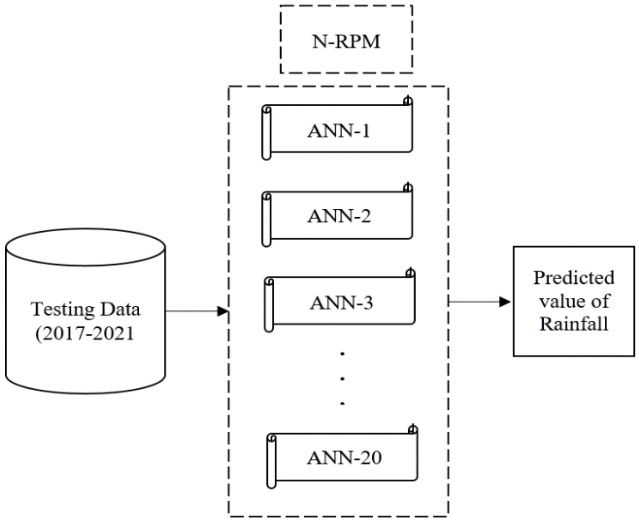

Figure 5: Testing of Novel - Rainfall Prediction Model (N-RPM)

In Novel Rainfall prediction model, Bagging ensemble technique is used which is also called as bootstrap aggregation and it is already discussed in detail previously in this section. [22]

Artificial neural network is used as base classifier after being comparing with many

classification and regression supervised Data mining techniques such as Decision tree, K-nearest neighbor, Linear regression, Logistic regression, Random forest etc.

The proposed model is composed of 20 artificial neural networks which are to be trained in parallel using 20 Data sets. These data sets are generated from the training data using sampling with replacement technique. The sample ratio for generating each new data set is 0.5. The sample ratio specifies the fraction of instances used to create new data set from the original supplied data set.

Every neural network is trained parallel by a unique data set with 500 training cycles. Each neural network is trained using back propagation algorithm having learning rate of 0.3.

Cross validation is a statistics based validation method which is used to evaluate the performance of a machine learning model. Cross validation effectively deals with the problem of overfitting which is generally found in predictive models. [23]

In problem of overfitting, the model during its training process acquired the knowledge not only from the correct data but from the ambiguity and errors present in the data which will ultimately impact the predictive performance of the model in an abrogating manner.

All these things will barricade the prediction model and block the way towards a more generalized model that can be applied on anew data in order to predict information with accuracy.

In cross validation, the complete data set is divided into 'n' number of partitions also called as folds. The performance of model is analyzed on each fold and the average is considered as the final outcome of the performance analysis. In this research, the proposed N-RPM is validated using cross validation using 10 folds. The implementation of the cross validation is done in Rapid Miner 8.1.0 as shown in figure 8 and 9. [24]

V. RESULTS AND EXEMPLIFICATION

In this paper, a Novel-Rainfall prediction model has been proposed which is immensely discussed in the previous section. The acquired data set contains data from 1988 to 2021(May). The data set is divided into two parts i.e. Training data (1988 to 2016) and testing data (2017 to 2021). The Novel- Rainfall prediction model is trained using training data. After training of the model, testing data is supplied to the model and prediction value of precipitation is generated which is shown in figure 11. [25] The N-RPM is validated using cross validation method shown in figure 8 and 9. The confusion matrix formulated after the process of validation showing true positive, true negative, false positive and false negative is shown in table 2 [26].The performance of proposed N-RPM is evaluated on SIX parameters i.e. Accuracy, Recall, Precision, F-measure, Specificity and RMSE [27]. The formulas for calculation of these performance parameters are given below after the confusion matrix and the calculated values are shown in table 3 [28].

Table 2: Confusion matrix

Confusion Matrix	**Predicted Class**		
Actual Class		**False**	**True**
	False	5765 (TN)	545 (FP)
	True	213 (FN)	459 (TP)

Formulas for computing performance of the Proposed Model:

$$Accuracy = \frac{TP + TN}{P + N}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall}$$

$$Specificity, True\ negative\ rate = \frac{TN}{N}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Predicted_i - Actual_i)^2}{N}}$$

Table 3: Performance analysis of N-RPM with existing prediction models

S. No.	**Performance Parameters**	**RPM-DPS (Proposed Model)**	**RPM- KC [9]**	**RPM-MM [10]**	**RPM-DP [11]**
1	Accuracy	89.14	85.4	84.3	81.2
2	Precision	79.87	76.2	74.3	70.1
3	Recall	71.08	69.2	68.3	65.2
4	F-measure	75.21	71.9	69.4	65.8
5	Specificity	0.964	0.897	0.854	0.836
6	RMSE	0.284	0.312	0.325	0.340

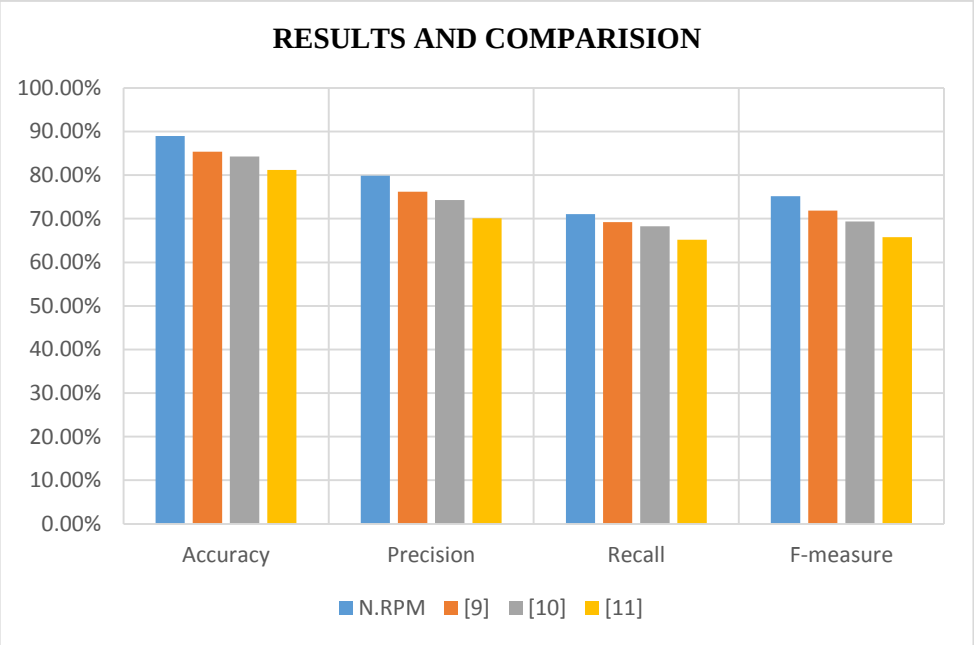

Figure 6: Results and comparison

The comparison of actual value of rainfall (in mm) and predicted value of rainfall (in mm) predicted using Novel Rainfall prediction model along with percentage error is shown in table 2. The average error percentage for the years (2017 to 2021) is **4.6%**.

$$\text{Error}(\%) = \frac{\text{Actual value} - \text{Predicted value}}{\text{Actual value}} \times 100$$

Table 4: Comparison of Actual and predicted Rainfall of HISAR DISTRICT

Year	**Predicted Rainfall (in mm)**	**Actual Rainfall (in mm)**	Error (%)
2017	484	502	3.5 %
2018	370	357	3.6 %
2019	395	415	4.8 %
2020	512	548	6.5 %
2021	681	651	4.6 %
ERROR (%)			**4.6 %**

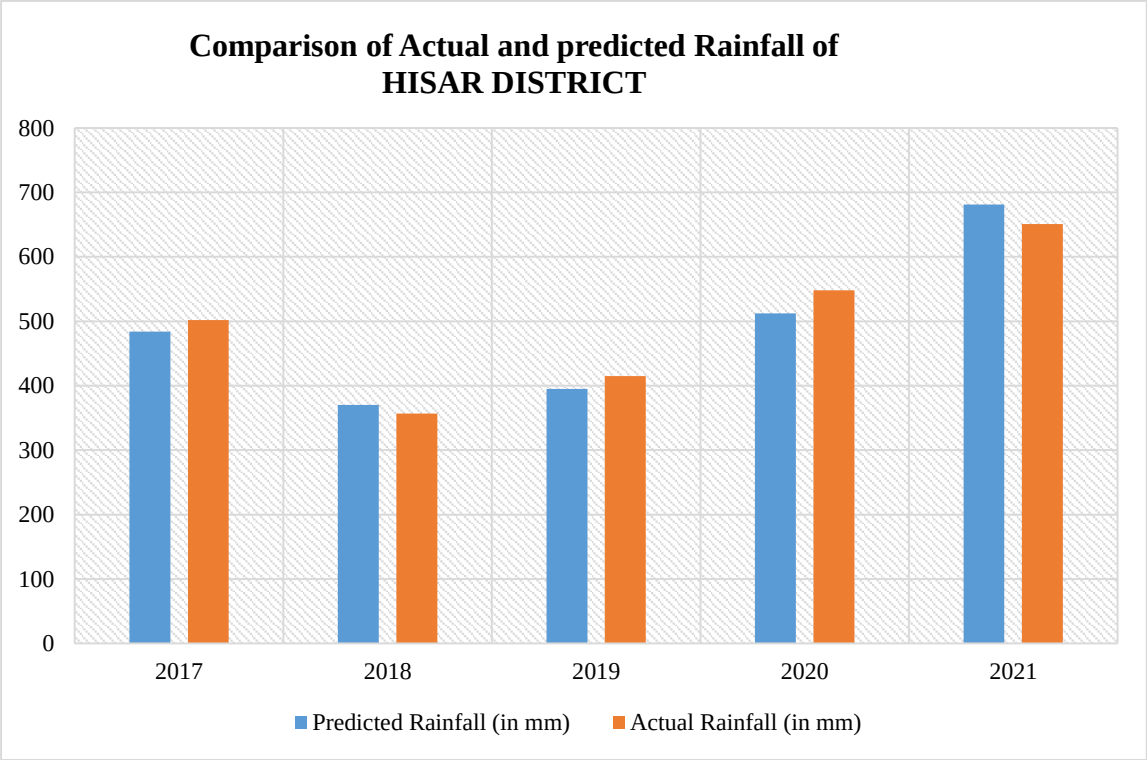

Figure 7: Comparison of Actual and predicted Rainfall

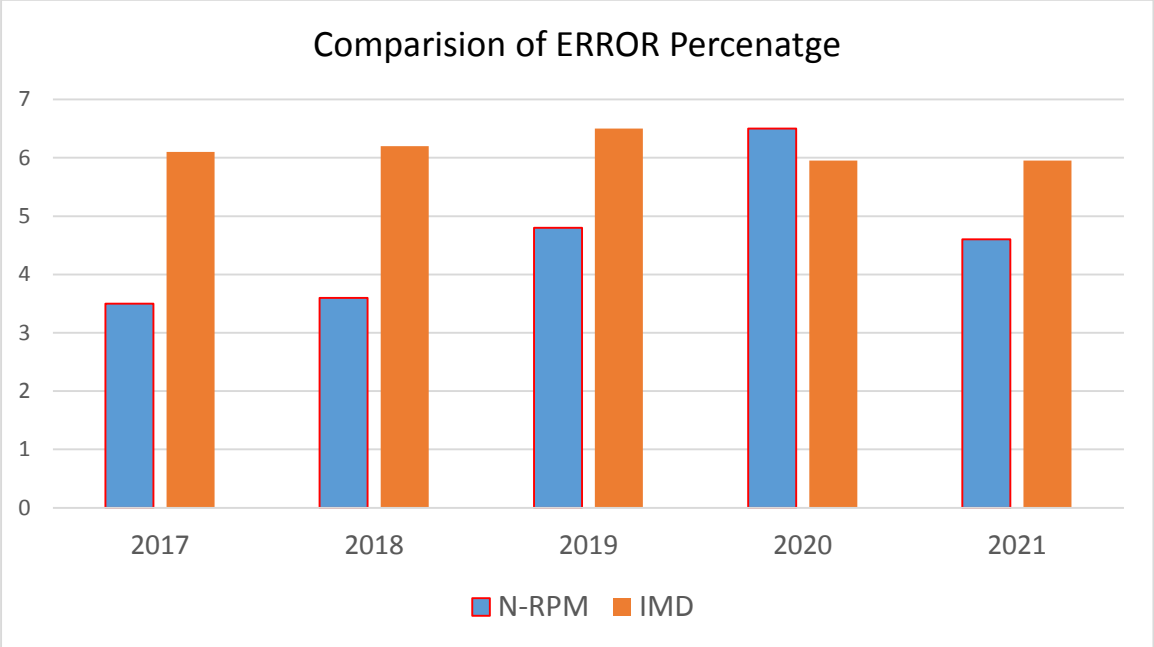

Figure 8: Comparison of Error Percentage

As per the information given by IMD i.e. Indian meteorological department in public domain which was published in many national level eminently reputed newspapers, the error (%) of predictions by IMD was 7.94% for the 1995 to 2006 and after that the error (%) of predictions by IMD declined to 5.95%. Also, from last 20 years the IMD i.e. Indian meteorological department has predicted the southwest monsoon accurately only once in every five years. The error percentage in predictions by Novel Rainfall prediction model is better as compared to Indian meteorological department (IMD).

VI. CONCLUSION AND FUTURE CONFINES

In this research work, a novel- rainfall prediction model has been proposed which is based on bagging ensemble supervised machine learning technique. Also, artificial neural network (ANN) is used as base classifier in the model. The N-RPM broadly discussed in the proposed model section of this paper. [29] A total of 20 artificial neural networks are used as base classifier and trained parallel on the training data set. The proposed model is evaluated using cross validation technique and the results are summarized in table 3.

The N-RPM has performed rainfall prediction for the year 2017,2018,2019,2020 and 2021 with an accuracy of 89.14%. [30] Also the average error rate for five years is 4.6% which is less than the error rate of prediction by Indian meteorological department (IMD). The error rate of IMD is around 5.95 % as per information given in public domain. Also from this research work it can be concluded that the N-RPM predicted the rainfall more accurately with 1.5% less error rate then IMD. Also, artificial neural network is selected as base classifier after comparing with many other supervised machine learning techniques so it also concluded that ANN with bagging shows best results for rainfall prediction as compared to other supervised machine learning techniques.

The accuracy of N-RPM model can be improved by studying the effect of climate change with rainfall patterns. This can be done by including the factors such as emission of greenhouse gases, levels of carbon dioxide, carbon mono oxide and other harmful gases in the environment. The proposed N-RPM model can be extended for the prediction of other weather attributes such as storm prediction etc.

REFERENCES

[1] Tharun V.P, Ramya Prakash, S. Renuga Devi, “Prediction of Rainfall Using Data Mining Techniques”, 2nd International Conference on Inventive Communication and Computational Technologies, IEEE-2018.

[2] Abishek.B, R.Priyatharshini, Akash Eswar M, P.Deepika, “Prediction of Effective Rainfall and Crop Water Needs using Data Mining Techniques”, International Conference on Technological Innovations in ICT For Agriculture and Rural Development, IEEE-2017.

[3] Fahad Sheikh, S. Karthick, D. Malathi, J. S. Sudarsan, C. Arun, “Analysis of Data

Mining Techniques for Weather Prediction", Indian Journal of Science and Technology, Vol 9(38), ISSN (Print): 0974-6846, IJST-2016.

[4] Ramsundram N, Sathya S, Karthikeyan S, "Comparison of Decision Tree Based Rainfall Prediction Model with Data Driven Model Considering Climatic Variables", Irrigation Drainage Sys Eng, an open access journal ISSN: 2168-9768, 2016.

[5] Bhaskar Pratap Singh, Pravendra Kumar, Tripti Srivastava, Vijay Kumar Singh, "Estimation of Monsoon Season Rainfall and Sensitivity Analysis Using Artificial Neural Networks", Indian Journal of Ecology (2017) 44 (Special Issue-5): 317-322.

[6] Nikhil Sethi, Dr.Kanwal Garg, "Exploiting Data Mining Technique for Rainfall Prediction", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3), 2014, 3982-3984.

[7] Chandrasegar Thirumalai, M Lakshmi Deepak, K Sri Harsha, K Chaitanya Krishna, "Heuristic Prediction of Rainfall Using Machine Learning Techniques", International Conference on Trends in Electronics and Informatics - ICEI 2017.

[8] Niketa Gandhi, Owaiz Petkar, Leisa J. Armstrong, "Predicting Rice Crop Yield Using Bayesian Networks", Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India.

[9] K C Gouda, Libujashree R,Priyanka Kumari,Manisha Sharma, Ambili D Nair, "An Approach for Rainfall Prediction using Soft Computing" International Journal of Engineering Trends and Technology (IJETT) – Volume 67 Issue 3 - March 2019 ISSN: 2231-5381.

[10] Moulana Mohammed, Roshitha Kolapalli, Niharika Golla, Siva Sai Maturi, "Prediction Of Rainfall Using Machine Learning Techniques" International Journal of Scientific & Technology Research Volume 9, Issue 01, January 2020 ISSN: 2277-8616.

[11] Deepali Patil, Shree L.R. Tiwari, Abhishek Jain, Shree L.R. Tiwari, Aniket Gupta, "Rainfall Prediction using Linear approach & Neural Networks and Crop Recommendation based on Decision Tree" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 9 Issue 04, April-2020.

[12] Suvidha Jambekar, Shikha Nema, Zia Saquib, "Prediction of Crop Production in India Using Data Mining Techniques", 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA).

[13] Wassamon Phusakulkajorn, Chidchanok Lursinsap, Jack Asavanant, "Wavelet-Transform Based Artificial Neural Network for Daily Rainfall Prediction in Southern Thailand", ISCIT 978-1-4244-4522-6/09 2009 IEEE.

[14] N. Tyagi and A. Kumar, "Comparative analysis of backpropagation and RBF neural network on monthly rainfall prediction," Proc. Int. Conf. Inven. Comput. Technol. ICICT 2016, vol. 1, 2017.

[15] N. Mishra, H. K. Soni, S. Sharma, and A. K. Upadhyay, "A Comprehensive Survey of Data Mining Techniques on Time Series Data for Rainfall Prediction," J. ICT Res. Appl., vol. 11, no. 2, p. 168, 2017.

[16] Deepak Sharma, Dr. Priti Sharma, "Rain Fall Prediction using Data Mining Techniques with Modernistic Schemes and Well-formed Ideas", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN (Online): 2278-3075, Volume-9 Issue-1, 2019, Page no. 258-263.

[17] Deepak Sharma, Dr. Priti Sharma. "Rainfall Prediction Using Classification and Clustering Complex Data Science Models with Geological Significance". International Journal of Computer Science Trends and Technology (IJCST) V8 (5): Page (39-44) Sep - Oct 2020. ISSN: 2347-8578. www.ijcstjournal.org.Published by Eighth Sense Research Group

[18] Kalyankar MA, Alaspurkar SJ, "Data Mining Technique to Analyze the Meteorological Data", IJARCSSE, vol. 3(2), pp. 114–118, 2013.

[19] Browning. K. A., "The mesoscale data base and its use in mesoscale forecasting", Quarterly Journal of the Royal Meteorological Society, vol. 115:488, pp. 717-762, 1989.

[20] Sharma Sanjay, Dutta Devajyoti, Das J, and Gariola R.M., "Nowcasting of severe storms at a station by using the Soft Computing Techniques to the Radar Imagery", 5thEuropean Conference on Severe Storms, Landshut-Germany, 2009.

[21] B.Renuka Devi, K.Nageswara Rao, S.Pallam Setty, M.Nagabhushana Rao, "Disaster Prediction System Using IBM SPSS Data Mining Tool", International Journal of Engineering Trends and Technology (IJETT), vol. 4(8), August 2013.

[22] Gao, S., and L. S. Chiu, "Development of statistical typhoon intensity prediction: Application to satellite-observed surface evaporation and rain rate (STIPER)", Weather Forecasting, vol. 27, pp. 240–250, doi: 10.1175/WAF-D-11-00034.1, 2012.

[23] Ganesh P. Gaikwad, Prof. V. B. Nikam, "Different Rainfall Prediction Models and General Data Mining Rainfall Prediction Model", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278- 0181, vol. 2(7), July – 2013.

[24] Dharminder Kumar, Deepak Bharadwaj, "Rise of Data Mining: Current and Future Application Areas", International Journal of Computer Science Issues, vol. 8(5), No 1, ISSN (Online): 1694-0814, 2011.

[25] Venkatadari M., Dr. Lokanataha C. Reddy, "A Review on Data Mining From Past to Future", International Journal of Computer Applications, pp. 19-22, vol. 15, No. 7, Feb 2011.

[26] Agnihotri, Dutta RK, Bhushan R, Somayajulu BLK. Evidence of solar forcing of southwest monsoon during last millennium. Earth Planetary Science Letters 2002; 198: p 521.

[27] Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. In Proc. of the ACM SIGMOD Conference on Management of Data, Washington, D.C. 1993; 207-216.

[28] Agrawal R, Srikant R. Fast Algorithms for Mining Association Rules. Proc. of the 20th International Conference on Very Large Databases, Santiago, Chile, Sept. 1994; IBM Research Report RJ9839.

[29] Agrawal R, Mannila H, Srikant R, Toivonen H, Verkamo AI. Fast Discovery of Association Rules. Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1995; Chapter 12.

[30] Akaike H. A new look at the statistical model identification. IEEE transaction on automatic control, 1977; 19: 716-723.

[31] Alexander G, Kesavamurthy RN, De VS, Chellappa P. Fluctuation of monsoon activity. Indian Journal of Meteorology, Hydrology and Geo Physics, 1978; 29: p 76.

[32] Allen, J., Christie, A., Fithen, W., McHugh, J., Pickel, J., Stoner,E. State of the Practice of Intrusion Detection Technologies. Technical report, Carnegie Mellon University.2000; http://www.cert.org/archive/pdf/99tr028.pdf.

[33] Ananthakrishnan R, Soman MK. Onset of southwest monsoon over Kerala 1901-1980. J. Climatology, 1988; 8: 283-296.

[34] Ananthakrishnan R, Pathan JM, Arakkatti SS.On the northward advantage of ITCZ and onset of southwest monsoon over southeast Bay of Bengal. J. Climatology, 1981; p 153.118

[35] Ananthakrishnan R, Ramakrishnan AR, Srinivasan V, Jambunathan R. Synoptic feature associated with onset of southwest monsoon over Kerala. Forecasting Manual, 1968; Vol.IV-18.2, India Meteorological Department, Pune.

[36] Ancell R, Sordo CM, Gutierrez JM. Applications of Bayesian Networks in Meteorology. Advances in Bayesian Networks.2004; 309-327.

[37] Asnani GC. Tropical meteorology. Sindhi Society, Pune, 1993; p 1202.

[38] Attri SD, Ajit Thyagi. Climate profile of India – Chapter I, Meteorology monograph – environmental meteorology. 01/2010, IMD, New Delhi.

[39] Balachandran S, Asokan R, Sridaran S. Global surface temperature in relation to northeast monsoon rainfall over Tamil Nadu. Journal of Earth System Science, 2006; 115(3): 349-362.

[40] Basu BK. Diurnal variation of precipitation during summer monsoon. Monthly Weather Review, 2007; 135(6): p2155.