Rainfall Forecast Based Predictive Analytics Model Using Machine Learning

L. Monish[1], S. G. Shaila[1(✉)], A. Vadivel[2], D. Shivamma[1], and S. G. Sumana[1]

[1] Dayananda Sagar University, Bangalore, India
{monishl-cse,shaila-cse,shivammad-cse,sumana-bca}@dsu.edu.in
[2] Gitam University, Bengaluru, India

Abstract. Rainfall irrigates over half of India's land. The average annual rainfall in India is 300-600 mm. India is endowed by the southwestern monsoon. But it is unreliable as rainfall will fickle along the year. This makes prediction very challenging and important. Rainfall prediction can be achieved by using advanced computer models and simulation tools. To understand and compute the complex patterns from the data, there is a need of efficient algorithms. Artificial Neural Network (ANN) easily fixes the problem concerned with the non-linear data. The feed-forward network and the backpropagation network can be used to overcome the downsides of the traditional methods used for rainfall prediction. Support Vector Machine (SVM) is used for rainfall prediction as SVM model is used for numerical value prediction for the observed non-linear data and the performance of both the algorithms are compared .

Keywords: Rainfall · Prediction · Classification · Artificial Neural Network · Support Vector Machine · Non-linear data · Regression

1 Introduction

Rainfall benefits the mankind by supplying water to satisfy their needs. Rainfall naturally balances the environment by humidifying the air, replenishing the water table, etc. Too much rain or too less of rain leads to serious condition like flood or drought. Rainfall prediction is important in the country as India's economy is largely dependent on agriculture. From saving the lives of the people to planning a farmer's crop, rainfall prediction will enhance the growth of the country's economy. Seasonal fluctuation in rainfall makes it challenging for meteorological scientists to predict the rainfall. Predicting the rainfall accurately is a solution to many serious problems. Rain is one of the causes of natural disasters like flood and drought which are encountered by people across the country. Amongst all weather happenings in India, rainfall has most impact on farmers. Indian farmers are widely dependent on monsoon, they plan their crop accordingly. Monsoon rainfall in India has also affected the lives in the township and cities. Flooding in major cities like Mumbai, Bangalore and Chennai are common in India due to monsoon cyclones. The living civilization greatly depends on the amount of rainfall

S. Tamane et al. (Eds.): ICAMIDA 2022, ACSR 105, pp. 622–631, 2023.
https://doi.org/10.2991/978-94-6463-136-4_53

for their well-being. Forecasting the rainfall must be done to improve the way of living of the people.

The ancient Indians mentioned the date of onsets of the rainfalls and each seasoned rainfall was named. They were successful in tracing the path of the monsoon clouds. In the recent times, the rainfall forecasting is done using statistical models namely ARMA (Auto Regressive Moving Average), ARIMA (Auto Regressive Integrated Moving Average) models etc. Due to the vigorous nature of the environment, statistical techniques fail to forecast rainfall. There are numerous other models that are based on mathematical or statistical computations. These real-life scenarios are better accomplished using the artificial neurons, which can learn from experience. In the operational hydrology rainfall forecasting is a great challenge. Different researches have approached in different ways to predict rainfall, but due to the considerable non-linearity of the data it is burdensome to predict rainfall accurately. Stochastic models must be used to attempt the forecast of rainfall. Rainfall has the most composite patterns that make it difficult for the model to understand. Thus, various neural networks are used such as Multilayer Feed Forward Neural Network (MLFN), Recurrent Neural Network (RNN), Time Delay Neural Network (TDNN) and Back Propagation Neural Network (BPNN). Among these, Artificial Neural Network (ANN) is flexible and a non-linear data driven model used for rainfall prediction. ANN is a parallel distributed model that has the tendency to store the experimental data. ANN algorithm does not make any assumptions on the data as seen in the previous statistical models. Another supervised learning algorithm is Support Vector Machine (SVM). It performs both regression and pattern classification. SVM will represent the points in the space. Here, examples are represented as points and are divided among the categories. When the new example is available, it will be represented as a point in the free space and is mapped into the same address space and the category for the new example is predicted. SVM forms hyperplanes between the points, when the data is linearly separable. The best hyperplane will create a margin between the two classes. Support Vector Regression (SVR) model is used to predict numerical output. The main idea of SVR is to compute a linear regression function in a high dimensional space and a non-linear function is used to map the input data.

The paper is organized as follows: Literature review is discussed in Sect. 2 and proposed methodology in Sect. 3. Section 4 will discuss about the experimental results and conclusion is presented in last Section.

2 Related Work

Meteorologists at National Oceanic and Atmospheric Administration (NOAA) were focused on monitoring the changes in the environment using the different equipment like doppler radar, weather satellites, radiosondes, etc. Observational data collected by doppler radar, radiosondes, weather satellites, buoys and other instruments were fed into computerized Numerical Weather Prediction System (NWS) and these models will compute the rainfall. This system uses previous history and a differential equation to provide the forecast. The drawback of this system is that the partial differential equation will not work with non -linear data and irregular patterns. Koizumi et al., (1999) [1] applied a method to modify numerical model forecasts using Artificial Neural Network

to predict forecasts. The neural network was used in a lame manner, the number of hidden layers were more than the volume of the data set used. The predictors were given to the network without the selection procedure. Abraham et al., (2001) [2] have applied built Soft Computing Models and Multivariate Adaptive Regression Splines for Rainfall Forecasting. The methods used Evolving Fuzzy Neural Network (EFuNN), Artificial Neural Network using Scaled Conjugate Gradient Algorithm (ANNSCGA), Adaptive Basis Function Neural Network (ABFNN) and General Regression Neural Network (GRNN). Multivariate Adaptive Regression Splines (MARS) is a regression technique that outperforms other models, in terms of performance and Root Mean Square Error (RMSE). Among these models, EFuNN (neuro-fuzzy system) performed better with low RMSE and performance time. Boser et al., (1996) [3] have created a nonlinear classifier using the kernel trick. The different kernel function used were polynomial function, radial basis function and hyperbolic function. This transformation will create a marginal Hyperplane in the data. This model was giving a satisfactory result. Abhishek et al., (2012) [4] have used backpropagation ANN model and observed least mean squared error in predicting the rainfall. ANN with Back-Propagation is a very complex network with large number of hidden layers. This model gave the best performance than the other models. But the performance time is observed to be more. Nayak et al., (2013) [5] have used SVM classification model to predict the rainfall and observed satisfactory results. Rainfall prediction used divide and conquer approach where the regions were divided into smaller areas. Pre-modelled tools were not used. RBF was used to predict the rainfall. Rainfall prediction was accurate in some regions. But some smaller regions gave low accuracy. This method can be used in emergency conditions only.

Narvekar et al., (2015) [6] have proposed a daily based forecast system. The ANN back propagation model is used. he outputs of the models were the minimum and maximum temperatures, relative humidity, and the rainfall. Good accuracy was observed. Nagahamulla et al., (2012) [7] have proposed a model for daily rainfall prediction that uses ensemble neural network (ENN). A finite number of ANNs networks are trained on the same task. All these finite models are combined to form an ensemble neural network (ENN). Weighted average method is used to combine the weights to the model. 41 years of rainfall data with 21 different parameters across Columbo is used. The data is classified into 4 climatic seasons. This model gives higher accuracy compared to individual back propagation method. Haq, et al., (2021) [8] have used Long Short-Term Memory (LSTM) for forecasting the rainfall based on rainfall parameters: El-Nino and Indian Ocean Dipole (IOD). The Authors have also used rainfall time series pattern to predict the rainfall. The authors have predicted the rainfall across Sidoarjo, East Java. The rainfall parameters of 5 weeks across Sidoarjo were used to predict the rainfall for the 6th week. Fayaz, et al., (2022) [9] have used the geographical data to predict the rainfall using the different models of neural network like BPNN, Feed Forward Neural Network (FFNN), Grey Wolf Levenberg-Marquardt-based non-linear neural network (GWLM-NARX), RNN, and TDNN. Barrera-Animas, et al., (2022) [10] have used various Machine Learning and Deep Learning models to predict the rainfall across five major cities in United Kingdom. They have used the climate data from year 2000 to 2020. The authors have obtained the best prediction results using LSTM with minimal hidden layers. The major disadvantage of all these statistical models is that they cannot

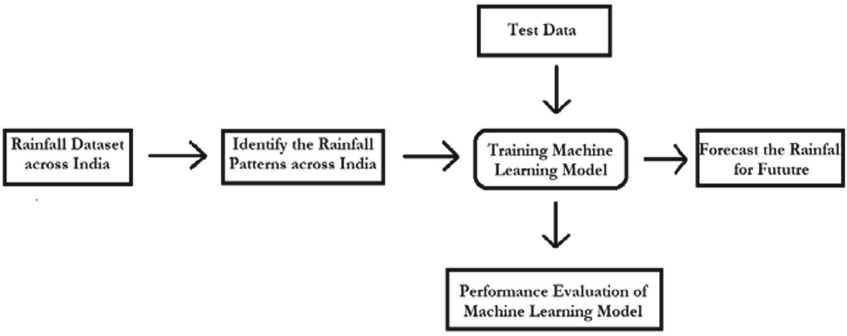

Fig. 1. Architecture Diagram for the rainfall prediction model

work with time varying non-linear data. Though, these models perform well in calculations, however, not in prediction as these models cannot adapt to the irregular and nonlinear patterns of data. This is because the data format cannot fit any function or a formula.

3 Proposed Methodology

Initially the collected input data is used to train the learning algorithm with the rainfall patterns. Data Collection has been a great challenge, rainfall data from the year 1905 to 2015 across India is available in the data.gov.in portal. Rainfall data from the year 1905 to 2015 across India is merged with the Rainfall impacts data to analyze the importance of the rainfall prediction in the country. The data contains the information about the houses lost, cattle lost and lives lost. Affects data is analyzed against the rainfall data. There are several factors that control the rainfall such as Temperature, Pressure, Wind speed, Wind Direction, Cloud cover and Humidity. Several machine learning models like logistic regression, Gaussian Naive Bayes, Decision Tree, Regression, SVM and ANN have been trained on the data. The accuracy and the F1 score of the models have been compared to choose the best algorithm for rainfall prediction. Out of all these models ANN and SVM gave the best accuracy compared to other models. Since the rainfall patterns are non-linear, ANN and SVR algorithm is used to handle this non-linear pattern of rainfall. Using this a predictive model is derived that will consider the rainfall forecast data along with the various parameters as input and the model predicts a rain value. The architecture to predict the rainfall is shown in Fig. 1. The algorithm inputs the parameters for which the rainfall needs to be predicted. After the successful intake of the input, the program is trained to predict the rain value.

ANN is a Feed forward Neural Network without a cyclic connection between the nodes. A python library Keras gives inbuilt implementation by specifying the number of nodes as input, hidden and output layers.. The Keras model has 3 dense layers each with 20 input layers, 10 hidden layers and 1 output layer. A bridge between the input and output layers is maintained by the activation function. The different activation can be used such as Sigmoid, Tanh, Sine, Cosine, and linear. For the input layer with x features, with a hidden layer with i neurons will produce the prediction y. Sigmoid function σ is

a S-shaped curve. The output value will lie between 0 and 1. Sigmoid function is used to predict the probability. However, we have more than 2 categories in the output. So, we use Softmax activation function. This is represented in Eq. (1).

$$softmax(Z_i) = \frac{\exp(Z_i)}{\sum \exp(Z_i)} \tag{1}$$

where Z_i is the vector for the features x.

We have also employed SVM model with different kernel functions that can be used as polynomial function, radial basis function and hyperbolic function, radial basis kernel function is used in the regression model. SVR will focus on identifying a function on mapping input domain to real number on the basis of training examples. The points within the decision line will form the best fit hyperplane. The decision lines are at a distance € from the hyperplanes. The hyperplane Y is represented by Eq. (2).

$$Y = wx + \mathrm{b} \tag{2}$$

w and b are constants, x represents the different features of rainfall such as temperature, humidity, etc. and y represents the rain value in millimeter(mm). The decision boundaries are represented in Eq. (3) and (4).

$$wx + b = +€ \tag{3}$$

$$wx + b = -€ \tag{4}$$

Figure 2 represents the ANN feed forward network and Fig. 3 represents the SVR graph. Dividing the dataset into training and testing set plays vital role for any machine learning task. The dataset is cross-validated into 80:20 where 80% of data is used to train the machine and 20% for validating the trained model. Similarly, the proposed approach cross-validated the dataset into 70:30 for prediction in next iteration. Both ratios were resulting in improved accuracy. Models are trained for 100 epochs and their performance is measured. The Accuracy and Root Mean Square Error are the different performance measures used to evaluate the proposed approach.

4 Results and Discussion

The rainfall dataset collected from the year 1905 to 2015 is used to predict the rainfall across India. The dataset contains details about the various features used for prediction. The data is obtained from the base Open Government Data Platform Data.gov.in. Table 1 depicts the statistics of various features in the dataset.

Figure 4 and 5 clearly depicts the maximum affect due to rainfall for the year 2007 whereas the rainfall has maintained the similar pattern from 2005 to 2013. After the disaster in 2007 due to the heavy rainfall, the precautionary steps are taken by the government to keep the people and animals in protection and to decrease the disasters in the succeeding years.

Predicted values of rainfall from both ANN and SVM are available. The mean squared errors of the models are also presented. The model with the least mean square error is

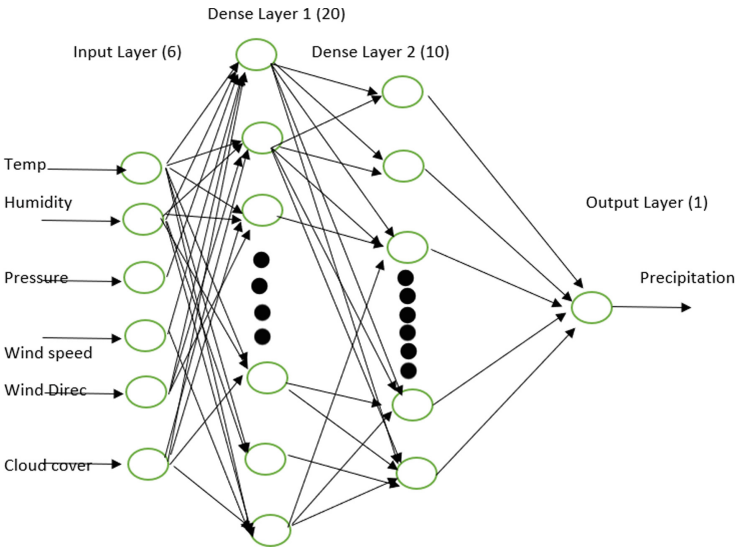

Fig. 2. ANN model for Rainfall Prediction

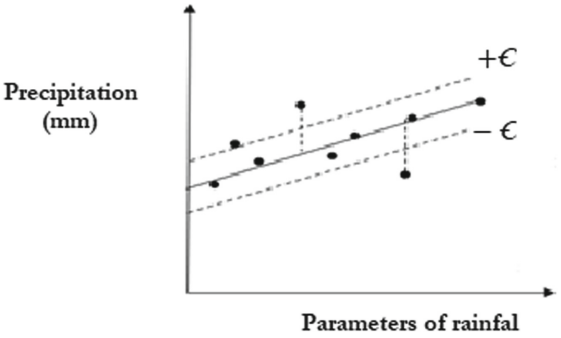

Fig. 3. SVR Model for Rainfall Prediction

Table 1. Statistics of the Dataset

	Temp (c)	Pressure (mb)	Humidity	Cloud cover	Wind Speed (mph)	direction	Rain (mm)
Count	3956	3956	3956	3956	3956	3956	3956
Mean	30.02	1004.69	56.10	26.03	6.97	8.87	0.66
Std	5.78	6.75	18.66	27.59	4.00	4.01	1.93
Min	9.00	993.00	4.00	0.00	0.00	1.0	0.00
Max	45.00	1022.00	93.00	100.00	32.00	16.00	12.53

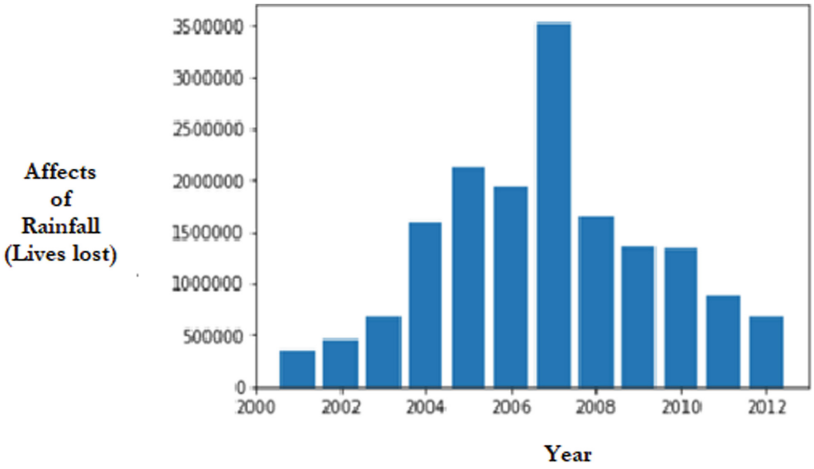

Fig. 4. Rainfall Affected for the year 2000–2013

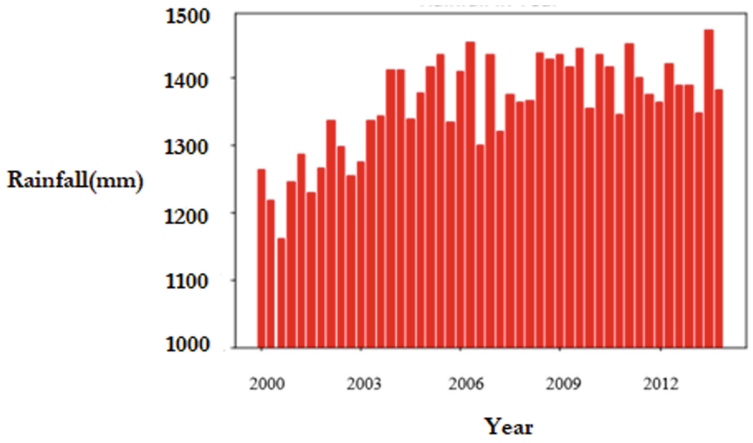

Fig. 5. Rainfall for the year 2000–2013.

the best accurate model in predicting the rainfall. From the observed result, for a given input values of temperature-35 °C, pressure 1000 mb, humidity-68%, cloud cover-27%, wind speed-2 mph and wind direction-South, expected value of rainfall is 2.56 mm and hence ANN is predicting the rainfall value 2.5 mm, which is accurate, as shown in the Table 2. It is observed that, ANN has the least mean squared error value 2.799 compared with SVM 3.298. The mean squared error of the algorithms are plotted in Fig. 6.

Table 2. Performance Comparison of ANN and SVM models.

Rainfall values Vs Features					
Temp.	**Pressure**	**Humidity**	**Cloud cover**	**Wind Speed**	**Direction**
35	1000mb	68%	27%	2mph	9
ANN Predicted value			**SVM Predicted value**		
2.5 mm			1.4 mm		
ANN mean squared error			**SVM mean squared error**		
2.799			3.298		

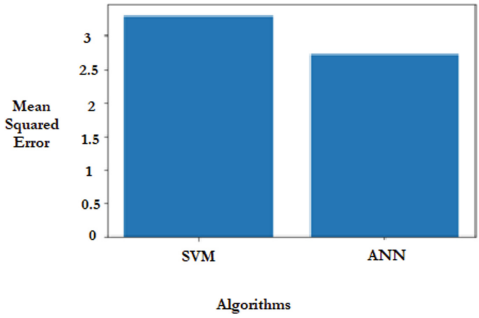

Fig. 6. Mean Squared Error of ANN and SVM models.

Table 3. Sample Test set to compare the output of ANN and SVM.

Test Set	**Observed values**	**SVM**	**ANN**
[30 996 80 42 7 7]	0.3	0.697	0.400
[29 997 81 20 5 7]	0.1	0.697	0.199
[33 994 64 45 3 7]	0.3	0.697	0.400
[31 996 50 42 9 7]	0.1	0.697	0.237

The comparison results of the testing set with respect to ANN and SVM models are depicted in the below Table 3. ANN is predicting the rainfall closely to the same values of the observed values as compared to SVM.

Figure 7 depicts the GUI designed for rainfall prediction system. It is designed for the ease of the users to predict the rainfall pattern by inputting the features such as temperature, pressure, wind speed, wind direction, cloud cover and humidity.

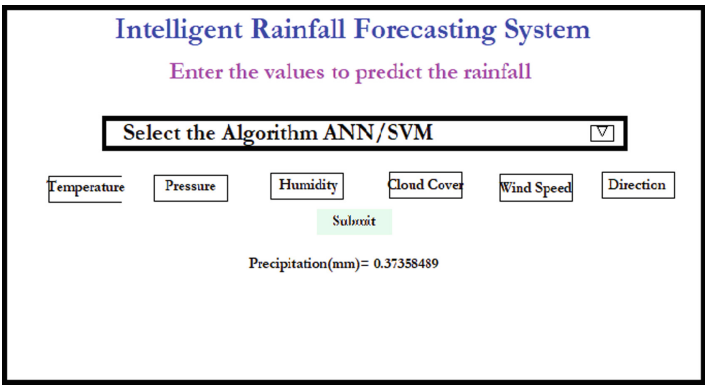

Fig. 7. GUI for Rainfall Prediction

5 Conclusion

Rainfall Prediction is an important aspect for famers in rural India, as it is concerned with the people and animal life. These days, once a year rainfall drain most of the cities in the country due to the poor infrastructure or planning. Predicting the rainfall will definitely help the government and the people living in the remoted areas. A detailed study of rainfall predictions using various models over twenty-five years is done. In this paper, the proposed approach used ANN to fix the problem concerned with the non-linear data. The feed-forward network and the backpropagation network are used to overcome the downsides of the traditional methods used for rainfall prediction. Support Vector Machine (SVM) is also employed for rainfall prediction it is used for numerical prediction for the observed non-linear data and the performance of both the algorithms are compared.

References

1. Koizumi, K. (1999). An Objective Method to Modify Numerical Model Forecasts with Newly Given Weather Data Using Artificial Neural Network, Weather and Forecasting, 14(1), 109-118.
2. Abraham, A., Steinberg, D., & Philip, N. S. (2001). Rainfall forecasting using soft computing models and multivariate adaptive regression splines. *IEEE SMC Transactions, Special Issue on Fusion of Soft Computing and Hard Computing in Industrial Applications*, *1*(xx), 1–6.
3. Boser, B. E. (1992). isabelle M. *Guyon, and Vladimir N. Vapnik."A Training Algorithm for Optimal Margin Classifiers." COLT*, *92*.
4. Abhishek, Kumar & Kumar, Abhay & Ranjan, Rahul & Kumar, Sarthak. (2012). A rainfall prediction model using artificial neural network. Proceedings - IEEE Control and System Graduate Research Colloquium, ICSGRC 2012. 82-87. https://doi.org/10.1109/ICSGRC.2012.6287140..
5. Nayak, Deepak & Mahapatra, Amitav & Mishra, Pranati. (2013). A Survey on Rainfall Prediction using Artificial Neural Network. International Journal of Computer Applications. 72. 32-40. https://doi.org/10.5120/12580-9217.

6. Narvekar, Meera & Fargose, Priyanca. (2015). Daily Weather Forecasting using Artificial Neural Network. International Journal of Computer Applications. 121. 9-13. https://doi.org/10.5120/21830-5088.
7. Nagahamulla, H. R., Ratnayake, U. R., & Ratnaweera, A. (2012, December). An ensemble of artificial neural networks in rainfall forecasting. In *International Conference on Advances in ICT for Emerging Regions (ICTer2012)* (pp. 176–181). IEEE.
8. Haq, D. Z., Novitasari, D. C. R., Hamid, A., Ulinnuha, N., Farida, Y., Nugraheni, R. D., ... & Widjayanto, A. (2021). Long short-term memory algorithm for rainfall prediction based on El-Nino and IOD data. *Procedia Computer Science*, *179*, 829-837.
9. Fayaz, S. A., Zaman, M., & Butt, M. A. (2022). Knowledge discovery in geographical sciences—a systematic survey of various machine learning algorithms for rainfall prediction. In *International Conference on Innovative Computing and Communications* (pp. 593–608). Springer, Singapore.
10. Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, *7*, 100204.
11. https://scikit-learn.org/stable/modules/svm.html#svm-regression.
12. https://www.noaa.gov/stories/6-tools-our-meteorologists-use-to-forecast-weather