International Journal of Advances in
Soft Computing and Intelligent Systems (IJASCIS)
2023, Vol 02, Issue 02, 87-105

Original Paper

A Comprehensive Review on AI Based Model for Rainfall Prediction

Bhavini K. Shah[a], Dr. Kamlendu S. Pandey[b]

[a]KRP Darjee Indo-American Inst Of Comp. Sci.,Vyara, 394640,
bhavi583shah@yahoo.co.in
[b]Veer Narmad South Gujarat University, Surat, 395007,
kspandey@vnsgu.ac.in

ABSTRACT

Our nation's economy is heavily dependent on agriculture and industries. To generate a profit, we should rely substantially on the availability of water. But the outcome is severely hampered by the irregularity of rainfall and the depletion of available water supplies. In 2014, India earned net $8 billion from $304 billion trade in commercial services. On the other hand, agricultural trade of $56 billion fetched as much as $18 billion in trade surplus. This is because while in services trade imports account for a lion's share, in agriculture imports component is negligible since basic resources such as sunlight, land, water, labor etc. are all available in the country. Agriculture is not just important for feeding the local population but to gain foreign exchange. Sometime heavy rainfall can cause damage to farms and crops but if it is predicted earlier than early warning can help to reduce the damage of life and resources. This paper provides a systematic literature review of state of art machine learning and deep learning techniques proposed by various authors to predict the rainfall. This paper gives information about Ensemble Learning, Logistic Regression, various Linear Regression, Multiple Linear Regression, Artificial Neural Network, K-Nearest Neighbor, Support Vector Regression, Decision tree and other miscellaneous models.

KEYWORDS

Rainfall Prediction, Machine Learning, Ensemble learning

1.INTRODUCTION

Troposphere is the lowest segment of earth's atmosphere and it holds in the order of 80% of air mass of atmosphere. The vapor form of water represents a minute, but vital component of the atmosphere(Pathak & Shastri, 2016).The percentage water vapor insurface air in desert regions varies to about 4% over oceans. Roughly 99% of it is contained in the troposphere. The precipitation of water vapor to the liquid or ice phase is accountable for clouds, rainfall, snow, and other precipitation, which plays important role in climate. The latent heat of vaporization is released to the atmosphere whenever precipitation occurs. This is one of the important parameters in the energy budget of atmosphere on both local and global scales (Maghrabi & Al Dajani, 2013). There are mainly two approaches to predict rainfall. They are Empirical method and Dynamic method. The Empirical approach is implemented by analyzing the historical data of the rainfall and how it is associated with other weather parameters. The most widely used Empirical approaches used for climate prediction are Regression, Artificial Neural Network (ANN), Fuzzy Logic and Group method of data handling. In Dynamic approach, predictions are generated by physical models based on systems of equations that predict the evolution of the global climate system in response to initial atmospheric conditions(Kannan et al., 2010). The remaining paper is arranged into following parts: Literature Review contains survey of different types of machine learning models for prediction of rainfall as well other weather attributes like temperature, humidity etc., Survey Statistic gives information about statistics related to the prediction model used by different authors, Observation and Discussion section describes which machine learning techniques is used by number maximum number of authors, Research Challenges Ahead section describes the work which is required to for future research . Finally, the conclusion of our work is described. After that Acknowledge from the authors is given and last section gives the list of references.

2.LITERATURE REVIEW

The following keywords: ("machine learning" OR "deep learning") AND ("precipitation prediction" OR "rainfall prediction") were used to search Google Scholar for works published between 2001 and 2023. Nearly 326 results were found, and only the supervised rainfall prediction studies that used meteorological data from, for example, radar, satellites, and stations were chosen; papers that used data from regular cameras, for example, pictures, were not included. The techniques employed to achieve this can be expanded upon and used to other geophysical characteristics like temperature and wind, even though the focus of this analysis is on the prediction of rainfall. The results and arguments of this chapter can therefore be modified to fit different conditions. (Philip & Joseph, 2001) predicted the rainfall using Artificial Neural Network in Kerala, India. They processed data of 87 years from 1893 to 1980. They predicted rainfall data using parameters like Wind, temperature, precipitation, latitude-longitude, sea surface pressure. They used variation of back-propagation like Adaptive Basis Function Neural Network for the research. For the test dataset the got RMSE around 0.09.

(Chantasut et al., n.d.) predicted rainfall using Stuttgart Artificial neural networks. They used data from 1941-1999 of 245 rainfall monitor stations in Thailand around Chao Phraya River. They got MSE 0.004000 for training data and 0.030799 for testing data.

(Lai et al., 2004) predicted rainfall and temperature for east coast of China. They applied some necessary data pre-processing technique and predicted the outcome using dynamic weighted time-delay neural networks (DWTDNN).

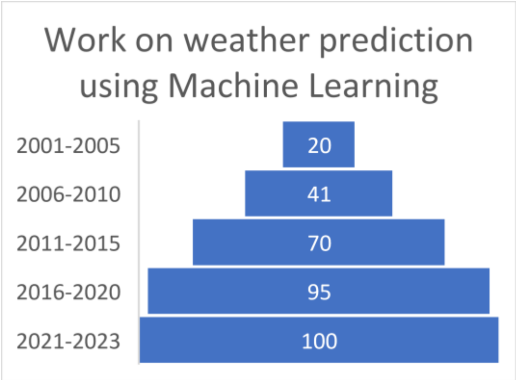
(Narayanan & Govindarajan, 2007) collected data of 102 years from 1901 to 2002 of Cuddalore district, Tamil Nadu. They built an ensemble model to predict rainfall. The dataset contains attributes for every month of a particular year. From the dataset 60% of records were used for training while 40% of records were used for testing purpose. They developed two models to forecast the rainfall. They used Ada Boost technique to develop ensemble models. The main aim of using AdaBoost was to train weak learner so that the weak learner can be trained appropriately. So, they combined Ada Boost with SVM which became AdaSVM. They combined Ada Boost with Naive Bayes algorithm which resulted in AdaNaive. Using AdaSVM and AdaNaive the data were analyzed. They implemented the classification algorithms in the Rapid Miner data analysis tool. They used “Read CSV” operator in rapid miner tool to load the data in the CSV file. They used “Select Attributes” operator to select some of data from the loaded data for classification process. Then, the selected data was given as input to the “X Validation” operator. X-validation is nested operator. The training part of Xvalidation contains Ada Boost using SVM (AdaSVM), SVM, Ada Boost using Naive Bayes (AdaNaive) and Naïve Bayes. The performance operator is used to get result of algorithm in terms of Accuracy and Classification error. The Accuracy noted for Ada Boost using SVM and Ada Boot using Naive are 98.66% and 97.62% respectively. The Classification error noted for AdaSVM and AdaNaive were 1.34% and 2.38% respectively. The error could be decreased if the correlated attributes could be selected using correlation matrix rather than directly manually selecting subset of attributes or Principle Component Analysis (PCA) should be used for the dimensionality reduction.

(Chattopadhyay & Chattopadhyay, 2008) has collected data for 128 years from 1871 to 1999 from all over the India. They predicted average summer rainfall using different Artificial Neural Network algorithms like momentum learning, conjugate gradient descent (CGD) learning, and

Levenberg–Marquardt (LM) learning and asymptotic regression. They got RMSE for these algorithms 0.26, 0.42, 0.47 and 0.15 respectively.

(B. Dutta et al., 2011) the authors anticipated humidity with usage of two different models 1. Self-Organizing Feature Map-Multilayer Perceptron (SOFM-MLP) 2. Feature Selection Multilayer Perceptron (FSMLP). In SOFM-MLP model, training examples were divided into homogeneous subgroup with the use of SOFM. After that it is pertained on the trained network. In FSMLP model, online feature selection was prepared by choosing the good features or attributes that will enhance the humidity prediction even as getting the estimation job or much less beneficial features can be removed through no longer permitting them into network. During the initial part of the training FSMLP permits only a very small "fraction" of each input feature value to go by the typical part of the MLP. As the network trains, it selectively permits simplest vital features to be vigorous by enhancing their attenuator weights as dictated by the way of gradient descent. Features with low attenuator weights were eliminated from the feature set. After feature selection they got RMSE value is 3.89. Mistakes can be reduced by applying other dimension reduction methods. Better result can be achieved if the FSMLP and SOFM-MLP can be ensemble together.

Figure 1. Count of literature review based on year



(Omary et al., 2012) has predicted precipitation using numerical weather prediction model in Jordan geographic area. They developed model based on HIRLAM and ALADIN models. The dataset contains 5 years of data containing different attributes like Min Temperature, Max Temperature, Mean Temperature, Cooling Degree Days, Growing Degree Days, Dewpoint, Avg. Humidity, Max Humidity, Min. Humidity to predict the precipitation.

(Agrawal, n.d.) predicted rainfall of Raipur. As the real time-series data has linear and nonlinear patterns within it, so it would be better to combine the linear and nonlinear model. So, they developed an ensemble model that deals with linear and nonlinear pattern of data. They developed an ensemble model combining ANFIS, ARIMA and IT2FLS algorithms. The proposed ensemble model could gain more accuracy than individual model. For the research, the authors collected weather data from meteorological center of Raipur, India for the period 2000 to 2009. The data contained important weather parameters like Year, Month, MinTemp, MaxTemp, Rainfall, Windspeed and RH. After collecting the data, they applied the preprocessing activity to replace missing value with zero. Then they used ARIMA model to reduce the set of parameters. Then these parameters were fuzzified using IT2 FLSs. After this number of fuzzy rules were reduced using Takagi-Sugeno fuzzy rule based model. Using the centroid type-reduction technique output was reduced. Then they set different parameters in simulation process of ANFIS+ARIMA+IT2FLS model. Finally, they compared the results of ANFIS+ARIMA+IT2FLS and ANFIS model using different statistical measures like RMSE and MSE. They found that the ensemble model gives more accuracy and less error than individual model. The result of the prediction can be increased if the data correlation can be there between weather parameters so that duplicasy can be removed and all the desired information in data can be considered for the analysis. Table 1 shows the literature review of rainfall prediction with its approach and result retrieved.

(Kumar, 2013) has predicted rainfall in Varanasi District of Uttarpradesh, India. They collected data from 2003 to 2007 from B.H.U rain gauge station. They got the R2 0.96 in their experiment.

(Geetha & Nasira, 2014) have considered weather data of international airport of United States. The authors applied machine learning algorithm like Feed Forward Networks or Multi-Layer Perceptrons (MLP) for predicting Minimum Temperature and Maximum Temperature. Back propagation network learns by example. Back propagation network is a supervised learning technique. It has mainly two steps 1. propagation 2. weight updates. Till the performance of the network is not improved its steps are repeated. The network is trained by giving examples and by changing the weights of network which generates the output. The authors developed ANN based model in Rapid Miner. They changed single feature of the ANN algorithm. They achieved 81.78% accuracy. ANN is better to get the strength of connections between attributes and it can learn from training and exhibit intelligence in making predictions. They considered only one property of ANN like training cycle for the prediction which may not be sufficient.

(Bushara & Abraham, n.d.) predicted rainfall using 14 base algorithms like Gaussian Processes, Linear Regression, Multilayer Perceptron, IBk, KStar, Additive Regression, Bagging, Random SubSpace, Regression by Discretization, Decision Table, M5Rules, M5P, REPTree and UserClassifier. They collected data of Sudan, South Africa for 13 years from 2000 to 2012 for 24 meteorological stations over the country. According to (Bushara & Abraham, n.d.) Kstar algorithm gave maximum correlation coefficient, minimum RMSE and MAE among 14 algorithms. On second place M5P algorithm was there then User Classifier came third among all.

(Gholami et al., 2015) predicted rainfall using the basic MLP-ANN model. They collected weather data of Kechik station in Golestan Province of northern Iran. The data were collected after 10 minutes of interval from year 2006-2007. The data had different attributes like temperature, evaporation, air pressure, humidity and wind velocity. They developed six models with different numbers of neurons in input and hidden layers. For each six models as training algorithms, they used three algorithms like Gradient Descent with momentum and adaptive learning rate (GDX), Conjugate Gradient (CG) and Levenberg-Marquart (L–M). They compared each architectures using different training algorithm. The result of analysis showed that L-M was best among three algorithms as it gave minimum RMSE and maximum R2 for each of architectures. But the RMSE can be reduced and CC can be increased by using different preprocessing techniques along with the algorithm.

(Zainudin et al., 2016)has predicted rainfall using different data mining techniques like Naïve Bayes, Support Vector Machine, Decision Tree, Neural Network and Random Forest. They collected data of multiple stations in Selangor, Malaysia. They applied cleaning and normalization on the data as part of pre-processing techniques. The authors have used humidity, pressure, temperature and wind speed. Neural Network has been used to classify the instances into low, medium and high classes based on a predefined training set.

(Aswin et al., 2018) predicted rainfall using Deep Learning Architectures like LSTM and ConvNet. They got RMSE 2.55 for LSTM as and 2.44 for ConvNet for time series dataset.

(Dash et al., 2018) predicted the rainfall of summer monsoon and post monsoon rainfall of Kerala state using K-nearest neighbour (KNN), artificial neural network (ANN), and extreme learning machine (ELM). They got that ELM technique has shown better performance in terms of MAE for summer monsoon (3.075) and post-monsoon (3.149) then other two algorithms KNN and ANN.

(Aguasca-Colomo et al., 2019) compared different rainfall prediction models like XGBoost (eXtreme Gradient Boosting), RF (Random Forest), GBM (Stochastic Gradient Boosting), LMT (Logistic Model Trees), SVM (Support Vector Machine), GLM (Generalized Linear Model) and

LDA (Linear Discriminant Analysis). Firstly, long short-term memory (LSTM) networks seem to be suitable for predictions based on time series data, since there can be lags of unknown duration between important events in these series. Secondly, Machine Learning for Streams, through monitoring streaming data from weather stations, is currently one of the most successful applications to forecast data sets in real time. This could be another research line to be explored.

Anfis and SVR were used to estimate rainfall by (Novitasari et al., 2020). The meteorological parameters (wind speed, relative humidity, and temperature) are forecasted using ANFIS, and the rainfall is projected using SVR based on these weather data. The MSE error for the projected rainfall using the Support Vector Regression (SVR) approach was 0.0928.

(Praveen et al., 2020) developed ensemble deep learning techniques like long-short-term memory (LSTM) networks to achieve very high quality for rainfall prediction. They also predicted rainfall using non-parametrical Mann-Kendal test based on rank system, to detect the trend in long-term rainfall data series.

Rainfall was forecasted by (Anwar et al., 2020) using a decision tree model created by the J48 algorithm. The J48 algorithm provided 77.8% accuracy for the training model, and the prediction model provided 86% accuracy for the 2020 testing dataset.

We propose the use of logistic regression (LR) and random forest (RF) to forecast the occurrence of landslides (Kuradusenge et al., 2020). For Rawanda, these methods make use of rainfall information from 2011 to 2018. There are 3970 records in the collection. With an AUC of 0.995 and 0.997, RF and LR fared better.

Using an interactive random forest-based model, (Asha et al., 2021) has forecast rainfall for Chennai city for the following 365 days. The dataset contained the variables needed to estimate rainfall during the last 18 years. The data processing tool of choice was Matlab. The dataset was first subjected to random forest classification, after which it was fed into back propagation neural networks, where the weights and biases were changed until the desired or appreciable result was obtained. They averaged a 100% accuracy rate.

Over South America, seasonal precipitation has been predicted by (Anochi et al., 2021). The season-specialist neural network was trained using a large dataset from 1980 to 2016 utilising a variety of techniques, including CPTEC's BAM model, NN-MPCA model, and NN-TensorFlow. Their upcoming research will examine the use of more precise seasonal forecasting using higher-resolution data.

Four alternative ML algorithms—Bayesian Linear Regression (BLR), Boosted Decision Tree Regression (BDTR), Decision Forest Regression (DFR), and Neural Network Regression (NNR)—were used by (W. M. Ridwan et al., 2021) to predict rainfall. On the other hand, multiple ML algorithms, including method 1 (M1): Forecasting Rainfall Using Autocorrelation Function (ACF) and method 2 (M2): Forecasting Rainfall Using Projected Error, were used to forecast rainfall over a range of time horizons. After adjusting the hyperparameter, BDTR has the highest coefficient of determination (R2) of any regression generated for ACF in M1. The author proposed a hybrid model for the future to increase accuracy.

In order to forecast hourly rainfall volumes using time-series data, (Barrera-Animas et al., 2022) used LSTM, Stacked-LSTM, Bidirectional-LSTM Networks, XGBoost, and an ensemble of Gradient Boosting Regressor, Linear Support Vector Regression, and an Extra-trees Regressor. They gathered data on five UK cities between the years 2000 and 2020.

Using a variety of techniques, including Decision Tree (DT), Random Forest (RF), Multilayer Perceptron (MLP), Extreme Gradient Boosting (XGB), and K-Nearest Neighbour (KNN), (Appiah-Badu et al., 2022) has predicted rainfall in several regions of Ghana. They came to the conclusion that KNN performed the least across all zones, while RF, XGB, and MLP all performed well. Decision Tree has regularly shown to be the quickest when it comes to model execution time, while MLP required the most run time.

Using four popular supervised machine learning techniques—decision tree, naive bayes, K-nearest neighbours, and support vector machines—(Rahman et al., 2022) have predicted rainfall. They combined fuzzy logic with machine learning approaches to predict rainfall more accurately. It is known as fusion. For the city of Lahore, they used 12 years' worth of historical weather data, from 2005 to 2017.

Rainfall prediction was proposed by (Gu et al., 2022) using a stacking ensemble model, which combines various machine learning model topologies.

3.SURVEY STATISTICS

We looked at 60 different papers that used a variety of algorithms to forecast weather parameters, including Artificial Neural Networks, Multi-Layer Perceptrons, Linear Regression, Multiple Linear Regression, Support Vector Regression, K-Nearest Neighbour, Decision Tree, ANFIS, ARIMA, Random Forest, Support Vector Machine, Boosting algorithms, LSTM, etc. We discovered that a maximum of 32% of these studies use artificial neural networks. Numerous artificial neural network implementations have been made, including those by (Philip & Joseph, 2001), (Chantasut et al., n.d.), (Lai et al., 2004), (Chattopadhyay & Chattopadhyay, 2008), (Wu et al., 2010), (B. Dutta et al., 2011), (Lakshmi et al.

We have found that among these papers maximum 10% papers are using ensemble model to predict the weather parameters. (Narayanan & Govindarajan, 2007), (Wu et al., 2010), (Barrera-Animas et al., 2022), (Gu et al., 2022), (Oswal, 2019), (Nong, 2010) have predicted weather parameter using ensemble models. We have found that 20% of authors have used different linear regression algorithm for predicting weather parameters.(Wu et al., 2010), (P. S. Dutta & Tahbilder, 2014), (Aksornsingchai & Srinilta, 2011), (Mathur, 2012), (Sumi et al., 2012), (Oswal, 2019), (Novitasari et al., 2020), (Kuradusenge et al., 2020), (W. Ridwan et al., 2020), (W. M. Ridwan et al., 2021), (Barrera-Animas et al., 2022) ,(Prakhar et al., n.d.) and (Kohail & Halees, 2011) have used different regression algorithms like Quantile Regression, Partial Squares Least Regression, M-regression, Multiple Linear Regression, Multivariate Adaptive regression splines, Support Vector Regression, Bayesian Linear Regression (BLR), Boosted Decision Tree Regression (BDTR), Decision Forest Regression (DFR), Neural Network Regression (NNR) and Linear Regression. . We have found that 10% of authors have used different K- Nearest Neighbour algorithm for predicting weather parameters. (Farouk et al., 2019), (Pabreja, 2012), (Sumi et al., 2012), (Dash et al., 2018), (Appiah-Badu et al., 2022) and (Rahman et al., 2022) have used KNN for predicting weather parameters. We have found that 12% of authors have used different Decision Tree algorithm for predicting weather parameters. (Farouk et al., 2019), (Prasad L V et al., 2012), (Zainudin et al., 2016), (Anwar et al., 2020), (Appiah-Badu et al., 2022) and (Rahman et al., 2022) have used Decision Tree algorithm for predicting weather parameters.

We have found that 4% of authors have used different ANFIS, ARIMA, IT2FLS algorithms for predicting weather parameters. (Agrawal, n.d.) and (Novitasari et al., 2020) have used ANFIS, ARIMA, IT2FLS algorithm for predicting weather parameters. We have found that 7% of authors have used Random Forest algorithm for predicting weather parameters. (Zainudin et al., 2016), (Aguasca-Colomo et al., 2019), (Kuradusenge et al., 2020) and (Appiah-Badu et al., 2022) have used Random Forest algorithm for predicting weather parameters. We have found that 7% of authors have use Support Vector Machine to predict the weather parameters. (Zainudin et al., 2016), (Aguasca-Colomo et al., 2019), (Rahman et al., 2022), (Aksornsingchai & Srinilta, 2011) have used Support vector machine to predict the weather parameters.

We have found that 2% of authors have used Boosting algorithm for predicting weather parameters. (Aguasca-Colomo et al., 2019) and (Appiah-Badu et al., 2022) have used Boosting algorithms to predict the weather parameters. We have found that 7% of authors have used Deep

Learning algorithm for predicting weather parameters. (Aswin et al., 2018), (Aguasca-Colomo et al., 2019), (Praveen et al., 2020), (Barrera-Animas et al., 2022) have used Deep Learning algorithms to predict the weather parameters. Table 3 shows work done by different authors by using different methods.

Table 1. Literature Review of Weather parameter prediction

Author/Citation	**Algorithm/ Approach**	**Dataset Time period**	**Region**	**Result**	**Predicted element/ Outlook**
Ali, M., Asklany, S., El-wahab, M., Hassan, M.(Farouk et al., 2019)	Naive Bayes KNN Decision tree	January 2010 to December 2015.	Cairo Airport	DT-97.45% KNN-77.34% NB- 97.45%	Haze (HZ) , Dust (Du) ,dust storm (DS) ,Cavok (clear weather case),fog (FG) and mist (BR
Pabreja, K.,(Pabreja, 2012)	KNN	ECMWF	Jammu and Kashmir	---	Humidity Temperature
Sri Lakshmi N., Ajimunnisa P., Lakshmi Prasanna V., YugaSravani T. , RaviTeja M.(Lakshmi et al., 2021)	Back Propagation Neural Network			70% accuracy	Temperature
Ashfaq Ali Shafin(Shafin, 2019)	Linear Regression, Polynomial Regression, Isotonic Regression, Support Vector Regressor	Kaggle 1900-2018=118 years Per years average	Bangladesh	R2= 0.83 MAE=0.172628 MSE=0.050341	Temperature
Hewage, P., Behera, A., Trovati, M., Pereira, E., Ghahremani, M., Palmieri, F., Liu, Y.(Hewage et al., 2020)	SR-Standard regression, LSTM -long short-term memory, TCN- Temporal convolutional network	23/08/2018 to 11/09/2018		MAE- Rain 0.000 016 Barometer 0.00042381656 Pressure 0.0005586127 Temperature 2.048237 Humidity 19.33102 Wind speed 1.8668145	Rain Barometer Pressure Temperature Humidity Wind Speed Wind Direction Dew point Heat Index

				Wind direction 3732.787 Dew point 18.028023 Heat index 6.774312	
Liu, J., Hu, Y., You , J., Chan, P. (Liu et al., n.d.)	Classical SVR DNN with SVR in Top Layer	MSLP	Hong Kong	Classical SVR - R2 0.851 DNN with SVR in Top Layer - R2 0.924	Temperature Wind Speed
Jakaria, M., Hossain, M., Rahman, M.(Jakaria et al., 2020)	Ridge Regression (Ridge), Support Vector (SVR), Multi-layer Perceptron (MLPR) Extra-Tree Regression (ETR), Random Forest Regression	Two months from 1st day of July, 2018 7 days from September 1, 2018	Tennessee n	RMSE RFR-3 Ridge-4 SVR-1 MLPR- ETR-3	Temperature
Holmstrom, M., Liu, D., Vo, C. (Holmstrom et al., n.d.)	Linear Regression, Sequential Comparison Model	2011-2015	Stanford		Temperature
Holmstrom, M., Liu, D., Vo, C. (Holmstrom et al., n.d.)	Linear Regression, functional regression, Professional	2011-2015	Melbourne, Australia	LR-5.039 FR-5.252 Professional - 2.612	Temperature
Rubhi gupta(Gupta, 2012)	Naive Bayes Bernoulli, Logistic Regression, Naive Bayes Gaussian, KNN classification			Precision NBB-1.0 LR-0.7857 NBG-0.6111 KNN-0.7500 Recall NBB-1.0 LR-1.0 NBG-1.0 KNN-0.2727	

Naik, A., Pathan, S.(Naik & Pathan, 2012)	Levenberg Marquardt BPNN				Rainy, Cloudy, Sunny, Partly Cloudy
(Balamurugan, n.d.)	MLR	01-09-2016 to 11-09-2016	Delhi	MSE-38.349 Variance score-0.491	Temperature
Singh, S., Nagrami, F., Pillai, A.(Singh et al., n.d.)	Random Forest Regression (RFR) Ridge Regression (Ridge) Support Vector Regression (SVR) Multi-layer Perceptron (MLPR) Extra-Tree Regression (ETR)		Nashville city	RMSE Ridge-4.6 SVR-3.1 MLPR-4.4 RFR-3.0 ETR-3.0	Temperature
(Hanoon et al., 2021)	Gradient Boosting Tree (G.B.T.) Random forest (R.F.) Linear regression (LR) MLP-NN RBF-NN	1985 to 2012	Terengganu, Malaysia	D factor Temperature MLP-NN 0.00039 RBF-NN 0.3037 Humidity MLP-NN-0.000647 RBF-NN-0.0842	Temperature, Humidity
(Sanders, n.d.)		2003 to 2013	Georgia	MAE RF-29.96	solar radiation, temperature, and wind speed

Table 2 shows the literature review of rainfall prediction with its approach and result retrieved.

Table 2: Literature Review of Rainfall prediction

Author/Citation	**Algorithm/ Approach**	**Dataset Time period**	**Region**	**Result**

Prakhar, A., Tripathi, A., Shrivastava, P., Gupta, U.(Prakhar et al., n.d.)	Linear Regression SVR Artificial neural net	Dataset1-1951-2000 Dataset2-1901-2015	Every state of India	MAE 57.08 116.6 42.13
Wu, J., Huang, L., Pan X. (Wu et al., 2010)	Bayesian Additive Regression Trees (BART) ensemble model	2005 to 2008	Guangxi	RMSE 4.41
Aksornsingchai, P., Srinilta, C.: (Aksornsingchai & Srinilta, 2011)	MLR SVM-POL SVM-RBF	2005 to 2007	Bangkok, Thailand	RMSE MLR-3.594 SVM-POL-3.284 SVM-RBL-3.199 MAE MLR-2.717 SVM-POL-2.257 SVM-RBF-2.158
Dutta, P., Tahbilder, H.: (P. S. Dutta & Tahbilder, 2014)	Multiple LR Algorithm	2007-2012	Assam, India	Adjusted R-squared - 0.628 R Square - 0.70
Nong, J.(Nong, 2010)	Non-parametric regression ensemble approach, K-nn-ANN model	1951 to 2002	Guangxi	MAE-1.855 CC-0.979
Kohail, S., Halees, A. (Kohail & Halees, 2011)	Least Median Squares Linear Regression (LMSL) NNs	1977 to 1985	Gaza	Correlation Coefficient LMSL-0.924 NN - 0.933 RMSE LMSL-1.69 NN -1.72
Luk, K., Ball, J., Sharma, A.(Luk et al., 2001)	Multi-Layer Feed Forward Network, Partial Recurrent NN, and Time Delay NN.		Western suburb of Sydney	NMSE-0.40

4.OBSERVATION AND DISCUSSION

We have observed that maximum 32% authors have used Neural Network algorithm to predict the weather parameters. Then 20% of authors have used Regression algorithms for the weather data prediction. After that 12% of authors have used Decision Tree for the prediction of weather parameters. We have reviewed 60 papers of weather prediction, in which we have examined data processing methods, models used, source of data, area of research, attributes used for prediction and finally the results.

5.RESEARCH CHALLENGES AHEAD

(W. Ridwan et al., 2020) predicted the rainfall data in Tasik Kenyir, Terengganu. The comparative study was conducted focusing on developing and comparing several Machine Learning (ML) models, evaluating different scenarios and time horizon, and forecasting rainfall The forecasting model uses four different ML algorithms, which are Bayesian Linear Regression (BLR), Boosted Decision Tree Regression. They predicted their result in terms of RMSE, MAE and R. They have written that more accurate rainfall prediction might be achieved by proposing hybrid machine learning algorithms and with the inclusion of different climate change scenarios.

In (Gholami et al., 2015), the authors predicted rainfall using the basic MLP-ANN model. They collected weather data of Kechik station in Golestan Province of northern Iran. The data were collected after 10 minutes of interval from year 2006-2007. The data had different attributes like temperature, evaporation, air pressure, humidity and wind velocity. They developed six models with different numbers of neurons in input and hidden layers. For each six models as training algorithms, they used three algorithms like Gradient Descent with momentum and adaptive learning rate (GDX), Conjugate Gradient (CG) and Levenberg-Marquart (L–M). They is compared each architectures using different training algorithm. The result of analysis showed that L-M was best among three algorithms as it gave minimum RMSE and maximum R2 for each of architectures. But the RMSE can be reduced and CC can be increased by using different pre-processing techniques along with the algorithm.

In (Bushara & Abraham, n.d.), considered only seven predictors for rainfall prediction, if we use some more climate factors such as atmosphere pressure, sea surface temperature, etc, so we may obtain more accurate prediction, also if Ensemble methods have been applied the results may be improved.

In paper (W. Ridwan et al., 2020) , we found the gap that they have used individual machine learning algorithm to predict the Rainfall and said that ensemble or hybrid machine learning algorithm should be developed for more accuracy. In paper (Gholami et al., 2015), they told to perform the pre-processing techniques to improve the result. In (Bushara & Abraham, n.d.), we found that more attribute should be included to improve the result of rainfall prediction.

Table 3: Work done by different authors by using different methods

ANN	**Ensemble**	**Regression**	**KNN**	**DT**	**NB**	**ANFIS+ ARIMA+ IT2FLS**	**RF**	**SVM**	**Boosting**	**LSTM**
(Philip & Joseph, 2001)	(Narayanan & Govindarajan, 2007)	Quantile Regression, Partial Squares Least Regression and the M-regression (Wu et al., 2010)	(Farouk et al., 2019)	(Farouk et al., 2019)	(Farouk et al., 2019)	(Agrawal, n.d.)	(Zainudin et al., 2016)	(Zainudin et al., 2016)	(Aguasca-Colomo et al., 2019) c	(Aswin et al., 2018)
(Chantasut et al., n.d.)	(Wu et al., 2010)		(Pabreja, 2012)	(Prasad L V et al., 2012)	(Zainudin et al., 2016)	(Novitasari et al., 2020)	(Aguasca-Colomo et al., 2019) c	(Aguasca-Colomo et al., 2019) c	(Appiah-Badu et al., 2022)	(Aguasca-Colomo et al., 2019)
(Lai et al., 2004)	(Barrera-Animas et al., 2022)	MLR (Mathur, 2012)	(Sumi et al., 2012)	Supervised Learning In Quest (SLIQ)	(Rahman et al., 2022)		(Kuradusenge et al., 2020)			(Praveen et al., 2020)
(Chattopadhyay & Chattopadhyay, 2008)	(Gu et al., 2022)		(Dash et al., 2018)	(Zainudin et al., 2016)			(Appiah-Badu et al., 2022)	(Rahman et al., 2022)		(Barrera-Animas et al., 2022)
(Wu et al., 2010)	(Oswal, 2019)	multivariate adaptive regression splines (Sumi et al., 2012)	(Appiah-Badu et al., 2022)	(Anwar et al., 2020)						
(B. Dutta et al., 2011)			(Rahman et al., 2022)	(Appiah-Badu et al., 2022)						

(Lakshmi et al., 2021)		Linear classifier, Tree-based, Distance-based, Rule-based and Ensemble (Oswal, 2019)		(Rahman et al., 2022)						
(Sumi et al., 2012)										
(Geetha & Nasira, 2014)		SVR (Novitasari et al., 2020)								
(Gholami et al., 2015)										
(Zainudin et al., 2016)		(Kuradusenge et al., 2020)								
(Aswin et al., 2018)		(W. Ridwan et al., 2020)								
(Dash et al., 2018)		Bayesian Linear Regression (BLR), Boosted Decision Tree Regression (BDTR), Decision Forest Regression (DFR) and Neural								
(Asha et al., 2021)										

		Network Regression (NNR) (W. M. Ridwan et al., 2021)								
(Anochi et al., 2021)		SVR (Barrera-Animas et al., 2022)								
(Appiah-Badu et al., 2022)										

6.CONCLUSION

This research presents a review showing that utilizing an artificial neural network (ANN) technique to anticipate the weather produces good results and can be thought of as an alternative to conventional metrological approaches. The study discusses the capabilities of artificial neural networks (ANN) in predicting a variety of weather phenomena, including temperature, thunderstorms, and rainfall, and came to the conclusion that ensemble model can be developed using more attributes like pressure, wind speed, atmosphere pressure and sea surface temperature.

ACKNOWLEDGEMENTS

We really appreciate the anonymous reviewers who provided their insightful feedback to improve the article's quality. The writers would like to thank as a result of Mrs. Sapna Gamit and Mr. Pratap Vasava over the course of this investigation.

REFERENCES

[1] Aguasca-Colomo, Castellanos-Nieves, & Méndez. (2019). Comparative Analysis of Rainfall Prediction Models Using Machine Learning in Islands with Complex Orography: Tenerife Island. Applied Sciences, 9(22), 4931. https://doi.org/10.3390/app9224931

[2] Aksornsingchai, P., & Srinilta, C. (2011). Statistical Downscaling for Rainfall and Temperature Prediction in Thailand. Hong Kong.

[3] Anochi, J. A., de Almeida, V. A., & de Campos Velho, H. F. (2021). Machine Learning for Climate Precipitation Prediction Modeling over South America. Remote Sensing, 13(13), 2468. https://doi.org/10.3390/rs13132468

[4] Appiah-Badu, N. K. A., Missah, Y. M., Amekudzi, L. K., Ussiph, N., Frimpong, T., & Ahene, E. (2022). Rainfall Prediction Using Machine Learning Algorithms for the Various Ecological Zones of Ghana. IEEE Access, 10, 5069–5082. https://doi.org/10.1109/ACCESS.2021.3139312

[5] Asha, P., Jesudoss, A., Mary, S. P., Sandeep, K. V. S., & Vardhan, K. H. (2021). An Efficient Hybrid Machine Learning Classifier for Rainfall Prediction. Journal of Physics: Conference Series, 1770(1), 012012. https://doi.org/10.1088/1742-6596/1770/1/012012

[6] Aswin, S., Geetha, P., & Vinayakumar, R. (2018). Deep Learning Models for the Prediction of Rainfall. 2018 International Conference on Communication and Signal Processing (ICCSP), 0657–0661. https://doi.org/10.1109/ICCSP.2018.8523829

[7] Balamurugan, M. B. (n.d.). SCHOOL OF COMPUTING SCIENCE AND ENGINEERING.

[8] 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. https://doi.org/10.1016/j.mlwa.2021.100204

[9] Chattopadhyay, S., & Chattopadhyay, G. (2008). Comparative study among different neural net learning algorithms applied to rainfall time series. Meteorological Applications, 15(2), 273–280. https://doi.org/10.1002/met.71

[10] Dash, Y., Mishra, S. K., & Panigrahi, B. K. (2018). Rainfall prediction for the Kerala state of India using artificial intelligence approaches. Computers & Electrical Engineering, 70, 66–73. https://doi.org/10.1016/j.compeleceng.2018.06.004

[11] Dutta, B., Mitra, S., Road, V., & Hooghly, P. O. (2011). A Connectionist Model for Prediction of Humidity Using Artificial Neural Network. 2(2).

[12] Dutta, P. S., & Tahbilder, H. (2014). PREDICTION OF RAINFALL USING DATAMINING TECHNIQUE OVER ASSAM. 5.

[13] Farouk, M., Ali, M., Asklany, S., Abd, M., & Hassan, M. (2019). Data Mining Algorithms for Weather Forecast Phenomena: Comparative Study. 76.

[14] Geetha, A., & Nasira, G. M. (2014). Artificial Neural Networks' Application in Weather Forecasting – Using RapidMiner. 4(3).

[15] Hanoon, M. S., Ahmed, A. N., Zaini, N., Razzaq, A., Kumar, P., Sherif, M., Sefelnasr, A., & El-Shafie, A. (2021). Developing machine learning algorithms for meteorological temperature and humidity forecasting at Terengganu state in Malaysia. Scientific Reports, 11(1), 18935. https://doi.org/10.1038/s41598-021-96872-w

[16] Hewage, P., Behera, A., Trovati, M., Pereira, E., Ghahremani, M., Palmieri, F., & Liu, Y. (2020). Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. Soft Computing, 24(21), 16453–16482. https://doi.org/10.1007/s00500-020-04954-0

[17] Holmstrom, M., Liu, D., & Vo, C. (n.d.). Machine Learning Applied to Weather Forecasting.

[18] Jakaria, A. H. M., Hossain, M. M., & Rahman, M. A. (2020). Smart Weather Forecasting Using Machine Learning:A Case Study in Tennessee (arXiv:2008.10789). arXiv. http://arxiv.org/abs/2008.10789

[19] Kannan, M., Prabhakaran, S., & Ramachandran, P. (2010). Rainfall Forecasting Using Data Mining Technique. .

[20] Kohail, S., & Halees, A. (2011). Implementation of Data Mining Techniques for Meteorological Data Analysis (A case study for Gaza Strip). International Journal of Information and Communication Technology Research. http://www.esjournals.org/

[21] Kumar, R. (2013). Forecasting of Rain Fall in Varanasi District, Utta r Pradesh Using Artificial Neural Network. https://www.semanticscholar.org/paper/Forecasting-of-Rain-Fall-in-Varanasi-District%2C-Utta-Kumar/c7d432c292d593c37b5f0a8005875b2321bdda86

[22] Lakshmi, N. S., Ajimunnisa, P., Prasanna, V. L., YugaSravani, T., & RaviTeja, M. (2021). PREDICTION OF WEATHER FORECASTING BY USING MACHINE LEARNING. International Journal of Innovative Research in Computer Science & Technology, 9(4).

[23] Liu, J. N. K., Hu, Y., You, J. J., & Chan, P. W. (n.d.). Deep Neural Network Based Feature Representation for Weather Forecasting.

[24] Luk, K. C., Ball, J. E., & Sharma, A. (2001). An application of artificial neural networks for rainfall forecasting. Mathematical and Computer Modelling, 33(6–7), 683–693. https://doi.org/10.1016/S0895-7177(00)00272-7

[25] Mathur, S. (2012). A Simple Weather Forecasting Model Using Mathematical Regression.

[26] Naik, A. R., & Pathan, S. K. (2012). Weather Classification and Forecasting using Back Propagation Feed-forward Neural Network. 2(12).

[27] Nong, J. (2010). A novel artificial neural network ensemble model based on K-nn nonparametric estimation for rainfall forecasting. Third International Workshop on Advanced Computational Intelligence.

[28] Omary, A., Wedyan, A., Zghoul, A., Banihani, A., & Alsmadi, I. (2012). An interactive predictive system for weather forecasting. 2012 International Conference on Computer, Information and Telecommunication Systems (CITS), 1–4. https://doi.org/10.1109/CITS.2012.6220375

[29] Oswal, N. (2019). Predicting Rainfall using Machine Learning Techniques (arXiv:1910.13827). arXiv. http://arxiv.org/abs/1910.13827

[30] Pabreja, K. (2012). Clustering technique to interpret Numerical Weather Prediction output products for forecast of Cloudburst. 3.

[31] Pathak, K., & Shastri, N. (2016). Estimation of the Effective Thickness and its Seasonal Variation of the Moist Layer over the Indian Subcontinent and its Comparative Study. Estimation of the Effective Thickness and its Seasonal Variation of the Moist Layer over the Indian Subcontinent and its Comparative Study.

[32] Philip, N. S., & Joseph, K. B. (2001). On the predictability of Rainfall in Kerala- An application of ABF Neural Network (arXiv:cs/0102014). arXiv. http://arxiv.org/abs/cs/0102014

[33] Prakhar, A., Tripathi, A., Srivastava, P., & Gupta, U. (n.d.). Rainfall Prediction Using Machine Learning.

[34] Prasad L V, N., Kumar Reddy K, P., & M m, N. (2012). A Novel Decision Tree Approach for the Prediction of Precipitation Using Entropy in SLIQ. International Journal of Simulation Systems Science & Technology. https://doi.org/10.5013/IJSSST.a.13.01.07

[35] Ridwan, W. M., Sapitang, M., Aziz, A., Kushiar, K. F., Ahmed, A. N., & El-Shafie, A. (2021). Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. Ain Shams Engineering Journal, 12(2), 1651–1663. https://doi.org/10.1016/j.asej.2020.09.011

[37] Ridwan, W., Sapitang, M., Aziz, A., Kushiar, K., Najah, A.-M., & El-Shafie, A. (2020). Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. Ain Shams Engineering Journal, 12. https://doi.org/10.1016/j.asej.2020.09.011

[38] Shafin, A. A. (2019). Machine Learning Approach to Forecast Average Weather Temperature of Bangladesh. 11.

[39] Singh, S., Nagrami, F. A., & Pillai, A. P. (n.d.). WEATHER PREDICTION BY MACHINE LEARNING.

[40] Sumi, S. M., Zaman, M. F., & Hirose, H. (2012). A rainfall forecasting method using machine learning models and its application to the Fukuoka city case. International Journal of Applied Mathematics and Computer Science, 22(4), 841–854. https://doi.org/10.2478/v10006-012-0062-1

[41] Wu, J., Huang, L., & Pan, X. (2010). A Novel Bayesian Additive Regression Trees Ensemble Model Based on Linear Regression and Nonlinear Regression for Torrential Rain Forecasting. 2010 Third International Joint Conference on Computational Science and Optimization.

Authors Biography

Bhavini Shah is an Assistant professor at KRP Darjee Indo-American Inst of Computer Science, Vyara, Gujarat, India. She has 17 years of experience in academics. She has taught a number of important papers of Computer Science at BCA. Her current research interests are Artificial Neural Network, Machine Learning and Data Mining. She also acted as Head of the Department, paper setter and examiner for different university examinations.

Dr. Kamlendu Kumar Pandey is Assistant Professor in Dept. Of Information and Communication Veer Narmad South Gujarat University, Surat, Gujarat in India. The interest areas are Wireless Sensor Networks, Cloud Computing, and Service Oriented Architecture. He is having 20 years of rich experience in academics and industry in the various fields of Computer Science. He has more than 50 publications in reputed national and international journals. He is author of 3 books related to Computer Science.

,