ANVESAK
ISSN: 0378–4568
UGC Care Group1 Journal

ASSESSMENT OF THE RELIABILITY OF A COMPOSITE LEARNING SYSTEM FOR PREDICTION USING PRECISION, RECALL, ACCURACY, AND F1-SCORE

[1]G. Sudarsanam, [2]M. Kanchana, [3]V. Mahesh Kumar, [4]V.Praveen Kumar
[1,2,3]Dept of Computer Science and Engineering, Sree Venkateswara College Of Engineering, Nellore (Dt), Andhra Pradesh, India.
[4]Dept of Electronics and Communication Engineering, Sree Venkateswara College Of Engineering, Nellore (Dt), Andhra Pradesh, India.

ABSTRACT
The whole economy of several nations depends on rainfall. Early rainfall forecasting reveals the effects on several fields. Accurate and timely rainfall predictions are particularly beneficial to the building industry, transportation activities, agricultural duties, aviation operations, and flood scenario. Every year, there is a lot of damage caused by heavy rains to both the infrastructure and human lives. Many studies are being conducted to forecast rainfall based on geographic location and meteorological factors. Each year, disasters brought on by rain cause considerable harm and loss to both infrastructure and human lives. This study suggests utilizing the Indian dataset and the Composite Learning Algorithm (CLA) to estimate rainfall using global characteristics. The technique is used to estimate rainfall using a dataset that is collected from UCI repository. The dataset consists of multiple meteorological parameters to predict the rain fall more precisely. Performance is calculated based on precision, recall, accuracy, F1-score are shown.
Keywords: Composite Learning, UCI repository, rainfall prediction.

1. INTRODUCTION

Rainfall is important in many nations across the world. The countries' average rainfall demonstrates the enormous influence on the economy of the nation. Rainfall and dams in some areas are the only sources of agriculture. Apart from agriculture, other locations, such as coastal regions worldwide, need to know the precise rainfall in order to avoid areas from flooding due to severe rainfall.

Prediction of rainfall has a bigger significance in sustaining water conservation systems around the nation. One of the biggest difficulties in predicting rainfall is complexity. The system's non-linear patterns and hidden layers are both extremely difficult to anticipate. The suggested approach is effective in locating hidden layers and also with non-linear patterns [1]. To predict the rainfall in any specific region there is a need to develop the application which can be used globally. The application can be developed as two types such as first and second. The first present with physical law which effects with rainfall and second one is used to develop the hidden patterns which can show the impact on physical factors and also the process involved in achieving it. Among these two second one works very efficiently with the integration of mathematical calculations [2]. The main objective of the proposed system is to find and resolve the issues such as identifying the hidden layers that are present in the system. This will provide the accurate and actual prediction that impacts the countries agriculture and economy [3].

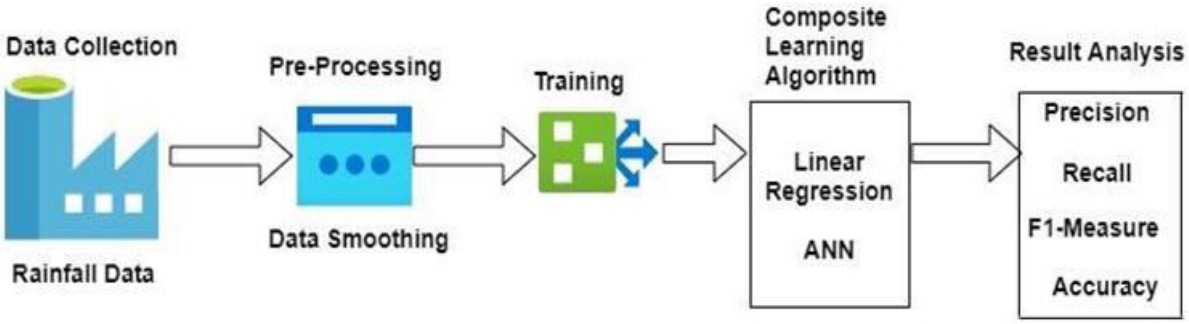

Figure1: System Architecture

2. LITERATURE REVIEW

The author [4] proposed the new rainfall prediction model which utilized the neural networks.This model is applied on Udupi region from the state of Karnataka in India. This model isintegrated with the BPNN and FFNN which gives the more accuracy at 90.12%. The training isappliedon 70%and testingis donefor 30%.

The authors in [5] have presented precipitation expectation utilizing momentary strategy due toits difficulties in the forecast. To address this CNN model was utilized to foresee the momentaryprecipitation by gathering set of climate highlights from numerous encompassing perceptions. Itwascontrasted andpublicclimategaugemodel and demonstratedessentiallybetter.

The author in [6] proposed the integrated model with the merging of Deep Learning Algorithmssuch as LSTM and ConvNet. These two models are applied on rainfall prediction which consistsof 10368 Geographic Locations from all the world for 468 months. The performance is existingmodel RMSE for LSTM is 2.55 and for ConvNet with 2.44. By increasing the secret layers, theerrors are reduced.

Theauthor[7]focusedonpredictingtherainfallwithadoptionofneuralnetworkalgorithms.Thismodelisappliedondatasetwhichiscollectedfrom1971to2000whichconsistsofmeteorological borders. In the existing system, the network objective error is set to 0.01 and thelearning rate is 0.01. The implementation is carried out with the MATLAB which uses theaccuracyof 100%.

The author [8] introduced the unique model that predicts the rainfall prediction using geneticalgorithm (GA). GA and MCRP merged and applied on 21 different datasets that are collectedfrom urban areas of Europe. Every day precipitation information for a very long time were takenaspreparinginformationandone-yearprecipitationinformationweretakenastestinginformation. GP defeats the shortcoming of MCRP by anticipating the different environmentsbetterthan MCRP.

Kala et al., [9] proposed the new method which is developed to predict the accurate rainfall.Artificial neural networks (ANN) are most widely used and soft computing prediction methodthat learns the past data and make the predictions on present data. This prediction is totally basedon the water resources. Theaccuracyis calculatedbyusingconfusion matrixand RMSE.

Santhosh et al., [10] introduced the new model that predicts the temperature. This is the NeuralNetwork libraries that support the various types training. This paper used the Back PropagationNeural Network (BPN) approach. The main advantage of this approach is to work efficiently onhuge classes.It is moreeffectivethan mathematical models.

Wu et al., [11] proposed the new model with preprocessing technique that includes the movingaverage (MA) and singular spectrum analysis (SSA). This shows the huge accuracy of 89.98%.Haviluddin et al., [12] proposed the new combined model which is combination of ANN andBPNN. Thisalgorithm shows the hugeperformancebyshowingMSE.

Aakashetal.,[13]discussedthevariousapproachesthatareusedtoanalyzetherainfallprediction which is shown in the table format. The author wants to show the performance ofseveralalgorithms to predict rainfall.

Tomoakietal.,[14]introducedthenewimprovedrainfallpredictionmodelthatisbasedonANN.Thisisautomatedalgorit

hmthatretrievesthemeteorologicalinformationthatpredictsthe rainfall by using internet. This paper mainly focused on using the big data on internet to analyzetheaccuraterainfall.
Grace et al., [15] proposed a new model that predicts rainfall that uses the Multiple LinearRegression (MLR). This is applied to the Indian dataset. This input dataset consists of variousparameters to predict accurate rainfall. The parameters are such as Mean Square Error (MSE),accuracy,correlation to analyzetheperformanceof theproposedmodel.

3. DATA SET DESCRIPTION

The dataset is UCI repository dataset consists small data which consists of 10000 records andlarge data 25551 records with 5 attributes such as DATE, PRCP, TMAX, TMIN and RAIN thedataabout theclimatic conditions ofvarious places are collected.

DATE	PRCP	TMAX	TMIN	RAIN
1948-01-01	0.47	51	42	TRUE
1948-01-02	0.59	45	36	TRUE
1948-01-03	0.42	45	35	TRUE
1948-01-04	0.31	45	34	TRUE
1948-01-05	0.17	45	32	TRUE
1948-01-06	0.44	48	39	TRUE

Figure 2:Data set Columns

4. DATA PRE PROCESSING

In rainfall dataset the noise should be removed by using binning data pre-processing technique.Binningwillsortthedataanddividethedataintoequalfrequencybins.Thisprocessisoneofthe data cleaning process that cleans the noisy data from the dataset. This technique is also reduces the impact of small errors that are observed. The actual values are divided into smallintervals which are called as bins. The smoothing effect is also reducing the chances of overfittingin caseof small datasets. The dividingof bins canbedonebyusing

$$[\min + w],\ [\min + 2w] \ldots\ldots [\min + nw],\ \text{where } w = \frac{\max - \min}{\text{no of bins}} \quad - - - \quad (1)$$

Especially in rainfall prediction dataset the TMAX and TMIN are two attributes considered forbinning.Afterthis step thetotal dataforprocessingis 6890 tuples.

TRAINING

Training data is one of the important step to analyze the attributes in the rainfall dataset. Aproposed pre-trained model applied on rainfall dataset. In this model, the features are identifiedbasedonthegivendataset.TheattributessuchasPRCP,TMAX,TMINarethethreeindependent variables and RAIN is one attribute which is dependent variable. From small dataset7000tuples areconsideredas trainingand 3000 areconsidered astesting.

COMPOSITE LEARNING ALGORITHM

The algorithm is combination of linear regression with ANN
Linear regression

- The real values are estimated such as (climate conditions, season, locations, and averagerainfall) based on continuous variable(s). By fitting the best line, the relation betweendependentand independent variables areused.
- Theregression lineisrepresented byalinearequation Y=a*X +b.
- X, Y are considered as two variables. Where Y is the dependent variable (that‟s thevariable that goes on the Y axis), X is the independent variable (i.e., it is plotted on the Xaxis), b is the slope of the line and „a‟ is the y-intercept. Here X is considered as thechances of high rainfall predictions and Y is considered as the less chances of rainfall i.e.,thedata classifiedas X and Yvariableand thesearetakenas inputs forANN.

Table1: Total values after applying the liner regression

	Total number of days for huge rain fall	Total number of days for less rain fall	Total number of days for medium rain fall
6890	6812	31	40

Artificial Neural Network

Artificial Neural Networks (ANN) using Linear Regression are most frequently employed in deep learning to handle a wide range of issues with complicated datasets. Particularly, a dataset like the prediction of rainfall calls for more efficiency to address the accuracy problems. A copied structure and learning process are mimicked by ANN.

Artificial neural networks, or ANNs for short, are adaptive systems that change their structure in response to input data. The network follows the flow throughout the learning phase. The neural network's neurons are interconnected, so each neuron's output feeds into the input of the next one until the final output is achieved. The weighting factors are modified or by a programmed method, and these connection weights hold the information required to get the final output closer to the known result. The network is trained by presenting it with a collection of input data with known results/output.

Initialize the input data X values to

Step 1: The data provided by the X values is transmitted in the first phase along with certain weights associated to the hidden layer as input values. The letter "h" stands for how many levels are buried. Examples of data inputs are x1, x2, x3,... xn. Each input is given a weight before being transferred to the hidden levels. The properties indicate that there are five input layers, including DATE, PRCP, TMAX, TMIN, and RAIN.

Step2:Each hidden layer neuron receives data from the input layer via the linear sum operation, and the hidden layer neuron's output is the result of feeding the linear sum to the activation function. From the concealed layer to the output layer, the same process is used. A linear combination of the transferred input values may be used to mathematically represent a three-layer neural network as follows:

$$\hat{y}_k = f_0\left|\sum_{j=1}^{n} w_{kj} f_h\left(\sum_{j=1}^{m} w_{ji} x_i + w_{jb}\right) + w_{kb}\right| - - - (2)$$

where y is the anticipated kth output value, 0 represents the activation function for the output neuron, n represents the number of output neurons, wkj represents the weight connecting the jth neuron in the hidden layer and the kth neuron in the output layer, h represents the activation function for the hidden neuron, m represents the number of hidden neurons, wji represents the weight connecting the ith neuron in the input layer and the jth neuron in the hidden

While wjb stands for the bias for the jth hidden neuron, wkb stands for the bias for the kth output neuron:

Step 3: A phase in the training process for mastering the ANN model is determining the optimal weight

vector to use in Equation (1). The back-propagation approach was utilised in this study to identify the weights that reduce the total network faults in Equation 2.

$$E = \sum_{p=1}^{P} E_p = \sum_{p=1}^{p} \sum_{k=1}^{n} (y_{pk} - \hat{y}_{pk})^2 - - - \ (3)$$

Where Ep is the error based on the squared difference between the genuine outputs ypk and the projected outputs ypk for pattern p, and E is the error for all input patterns.

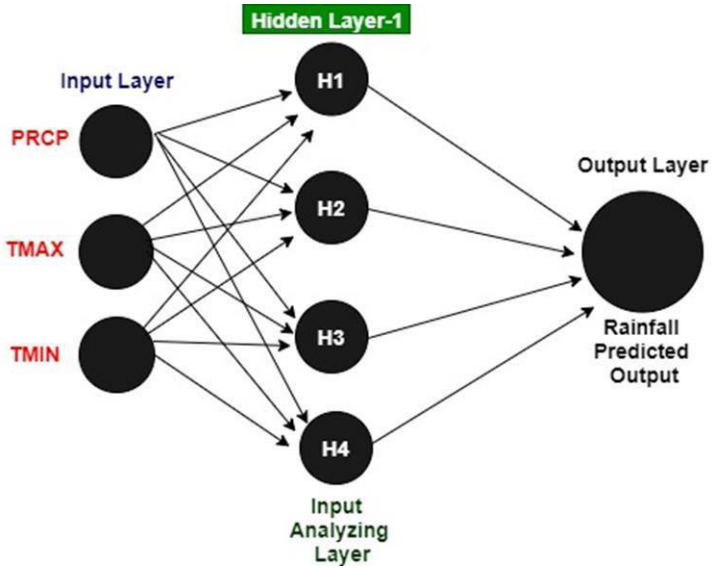

Figure 2: Architecture of Artificial Neural Network

EVALUATION METRICS

Performance Metrics:

Utilizing the following metrics—False Positive Rate (FPR), False Negative Rate (FNR), accuracy, recall, and F1 score—the performance of the rainfall prediction system is measured. These metrics make use of the fundamental count values True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

Precision

The accuracy is measured by the percentage of real positives that are accurately detected. It has to do with the test's capacity to recognise favourable outcomes.

$$\textbf{Precision} = \frac{\text{No. of TP}}{\text{No. of TP} + \text{No. of TN}} - - - (4)$$

1 Score: This evaluates the precision of a model on a dataset. It is employed to assess binary classification schemes that categorise examples as "positive" or "negative."

$$\textbf{F1 Score} = 2\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} - - - - (5)$$

Accuracy:Theoverallaccuracyof theoutputis calculated byusingthebelow equation:

$$\textbf{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} - - - (6)$$

Recall: Appropriate when minimizing false negatives is the focus.

$$\text{Recall} = \frac{TP}{TP + FN} - - - - - (7)$$

Experimental Results

The experiments were performed using win python and the programming language python. The accuracy, F1-Score, and precision are the parameters that are used in this stage to analyse the results of the proposed algorithm. The section's approximate results are displayed.

Table2: Comparison of Proposed Approach with Existing Approach

Algorithms	ANN	LR	CLA
Precision(%)	88.76	89.1	98.9
Recall(%)	87.88	89.7	99.2
F1-Score(%)	87.87	91.21	97.12
Accuracy(%)	89.12	92.34	98.89

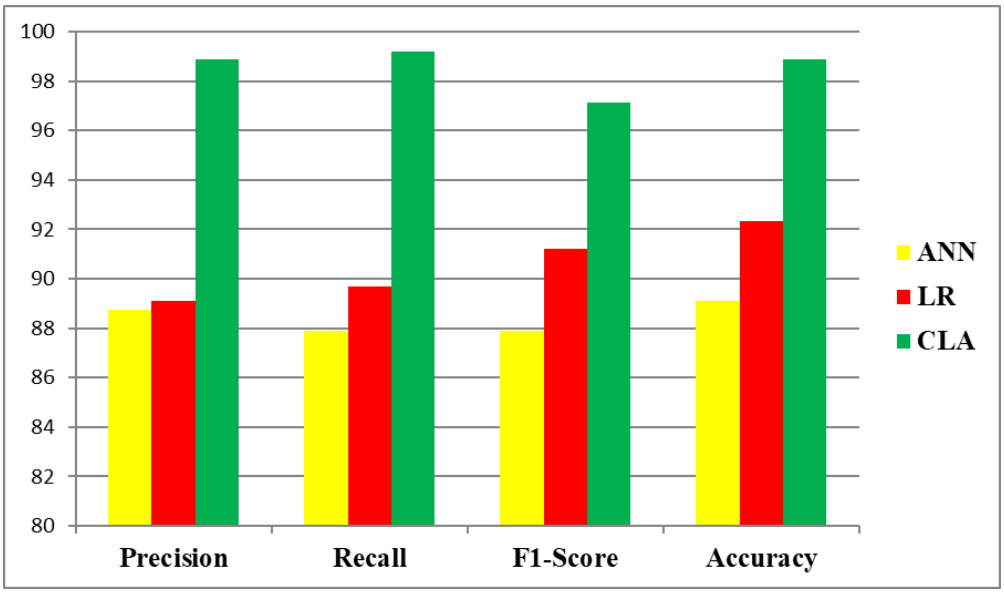

Figure 3: performance analysis

CONCLUSION

The difficult task is predicting rainfall. The ability of machine learning algorithms to extract the features and predict the rainfall more effectively is increasing.ML algorithms will own the secret information about the previous weather circumstances. Numerous scholars are experimenting with the idea of predicting rainfall using an improved ML system. The proposed system was primarily concerned with obtaining a more precise and improved rainfall prediction. The approach is a combination of linear regression and ANN for learning. The suggested system employs a multi-layered architecture to provide reliable predictions. Precision (98.9%), Recall (99.2%), F1-Score (97.12%), and Accuracy (98.89%) are used to calculate performance for the CLA, which has significantly improved performance when compared to existing methods.

REFERENCES

1. Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. Water (Switzerland),10(11).

2. Janani,B;Sebastian,P.(2014).Analysisontheweatherforecastingandtechniques.International Journal of Advanced Research in Computer Engineering & Technology, 3(1), 59–61.
3. Chaudhari, M. S., & Choudhari, N. K. (2017). Open Access Study of Various RainfallEstimation&PredictionTechniquesUsingDataMining.AmericanJournalofEngineeringResearch(AJER), 7, 137–139.
4. Kumar Abhishek, Abhay Kumar, Rajeev Ranjan, Sarthak Kumar," A Rainfall PredictionModel using Artificial Neural Network", 2012 IEEE Control and System Graduate ResearchColloquium(ICSGRC-2012),pp. 82-87, 2012.
5. Minghui Qiu, Peilin Zhao, Ke Zhang, Jun Huang, Xing Shi, Xiaoguang Wang, Wei Chu, "AShort-Term Rainfall Prediction Model using Multi-Task Convolutional Neural Networks", IEEEInternationalConferenceonData Mining,pp. 395-400, 2017,DOI10.1109/ICDM.2017.49.
6. Aswin S, Geetha P and Vinayakumar R, "Deep Learning Modelsfor the Prediction ofRainfall", International Conference on Communication and Signal Processing, April 3-5, 2018,India,pp.0657-0661.
7. XianggenGan,LihongChen,DongbaoYang,GuangLiu,"TheResearchofRainfallPrediction Models Based on Matlab Neural Network", Proceedings of IEEE CCIS2011, pp. 45-48.
8. Sam Cramer, Michael Kampouridis, Alex A. Freitas and Antonis Alexandridis, "PredictingRainfallintheContextofRainfallDerivativesUsingGeneticProgramming",2015IEEESymposium Series on ComputationalIntelligence,pp. 711– 718.
9. Kala and S. G. Vaidyanathan, "Prediction of Rainfall Using Artificial Neural Network,"2018 International Conference on Inventive Research in Computing Applications (ICIRCA),2018,pp. 339-342.
10. S Baboo Santhosh and Kadar Shereef, "An Efficient Weather Forecasting System usingArtificial Neural Network", International Journal of Environmental Science and Development,vol. 1, 2010.
11. C.L. Wu and K.W. Chau, "Prediction of rainfall time series using modular soft computingmethods", Engineering Applications of Artificial Intelligence, vol. 26, no. 3, pp. 997-1007, 2013,ISSN0952-1976.
12. Haviluddin Mislan, Sumaryono Sigit Hardwinarto and Marlon Aipassa, "Rainfall MonthlyPrediction Based on Artificial Neural Network: A Case Study in Tenggarong Station", EastKalimantan - Indonesia Procedia Computer Science, vol. 59, no. 2015, pp. 142-151, ISSN 1877-0509.
13. Aakash Parmar, Kinjal Mistree and Mithila Sompura, "Machine Learning Techniques forRainfall Prediction: A Review", 2017 International Conference on Innovations in informationEmbeddedand Communication Systems.
14. TomoakiKashiwao,KoichiNakayama,ShinAndo,KenjiIkeda,MoonyongLeeandAlireza Bahadori, "A neural network-based local rainfall prediction system using meteorologicaldata on the Internet: A case study using data from the Japan Meteorological Agency", AppliedSoftComputing, vol. 56, pp. 317-330, 2017,ISSN1568-4946.
15. R. K. Grace and B. Suganya, "Machine Learning based Rainfall Prediction," 2020 6thInternationalConferenceonAdvancedComputing andCommunicationSystems(ICACCS),2020,pp. 227-229.