2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) | 979-8-3503-1590-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ACCAI58221.2023.10199905

# Rainfall Prediction Using Deep Learning and Machine Learning Techniques

**Kothai G, Y Sushmitha, K Lakshmi Saranya, P Naga Ramya Sri and P Amulya**

*Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, Virudhunagar, India*

*E-mail : emailtokothaiganesan@gmail.com 9920004658@klu.ac.in 9920004649@klu.ac.in 9920004655@klu.ac.in 9920004628@klu.ac.in*

**Abstract- Predicting rainfall is a crucial component of contemporary weather forecasting, with consequences for agriculture, hydrology, water management, and disaster preparedness and response. The conventional method of predicting rainfall has depended on historical information and meteorological observations. Yet, more precise and timely predictions of rainfall patterns are now possible thanks to the development of sophisticated computer models, remote sensing technologies, machine learning, and deep learning algorithms. An overview of the methods for predicting rainfall is given in this work, including the analysis of data from satellite imaging, atmospheric conditions, ocean temperatures, and other climate variables. We also look at how machine learning and deep learning algorithms are evolving to build more sophisticated models that can evaluate vast quantities of data and predict rainfall patterns with greater accuracy.**

***KEYWORDS:*** ***Predicting Rainfall, Weather Forecasting, MLP, Temperature, Machine Learning, Deep Learning.***

## I.INTRODUCTION

The process of predicting the time and amount of precipitation in a certain area is known as rainfall prediction. For many industries, including agriculture, hydrology, and water management, as well as for disaster preparedness and response, the accuracy of rainfall prediction is crucial. In the past, basic meteorological observations and historical data have been used to predict rainfall. Yet, more precise and timely predictions of rainfall patterns may now be made thanks to technological advancements including sophisticated computer models, remote sensing techniques, and machine learning and deep learning algorithms.

Data from satellite photography, atmospheric conditions, ocean temperatures, and other climatic parameters are analysed using a variety of approaches by researchers and practitioners. Then, taking into consideration a variety of environmental and meteorological elements that affect precipitation, these data are utilised to develop models that can forecast rainfall patterns and probabilities for a specific time period. More complicated models may be built to examine vast quantities of data and produce more accurate predictions of rainfall patterns thanks to the advancement of machine learning and deep learning techniques.

Several sectors and activities can be greatly impacted by accurate rainfall forecasts. Using it, for instance, can help water management authorities more effectively manage their water supplies and assist farmers in planning their planting and harvesting dates. It can also assist in early warning and disaster response by advising authorities of imminent flooding or landslides.

## II.REVIEW OF LITERATURERESEARCH:

Rashad Ahmad et.al [1] has used machine learning techniques for predicting the rainfall in smart cities. They have used ANN, KNN, Decision tree, SVM techniques in their paper for prediction of rainfall. They got ANN-based classification as highest accuracy. Compared to machine learning techniques; deep learning technique got the highest accuracy as 82.1 %.

Muhammad Bilal et.al [2] predicted the rainfall using some machine learning algorithms for easily forecasting the time-series. They took some popular cities and predicted the rainfall based on that. Models they have used are LSTM, XGBoost, Linear support vector Regression, Extra tress Regressor. In these model's LSTM got highest accuracy and has

been predicted very well.

Feng Liang et.al [3] done the research article on Rainfall for Short-Time Prediction using deep learning. They have collected some echo images and prepared the dataset for predicting the rainfall which is very low. Conv3D-GRU has the highest accuracy. The publishers said that there are no conflicts in the paper. It is showing the correct prediction.

DemekeEndalie et.al [4] case study for Model of Deep Learning for predicting rainfall. He took the dataset based on agriculture and he gave the results that how the rainfall was going day by day. He proposed the Long-short term memory and gave the correct percentage. Decision tree, K-Nearest Neighbor also used but LSTM has the highest accuracy.

Nikhil Oswal et.al [5] Basically the rainfall is very hard to predict and it is difficult task. They showed the different accuracies for different methods. The data was collected from different cities of Australia. Different methods, different inputs and techniques. Real world data has been taken. Under sampling and Oversampling methods had been used to rainfall. Linear Regression and Random Forest also used.

Peter Watson et.al [6] prepared the models based on change of climate which causes the floods and heavy rainfall. Deep learning and machine learning both techniques for the rainfall. They plotted some graphs which can show the values and easily identified. TRU-NET has the highest accuracy compared to others. CNN and other neural networks also used.

Selva Balan et.al [7] used the data which is not linear and got the results for rainfall. Artificial neural network is used to train the dataset. Calculated the bias and changed the input weights of the layer. Because the artificial neural network will definitely work with the regression and classification techniques. Difference for without normalization and with normalization has done.

Nagulla Bhavana et.al [8] by using the crop dataset predicted the rainfall; it is useful for the farmers in the agriculture. ARIMA(Auto-Regressive Integrated Moving Average), ANN, SVM, Logistic Regression are used. Deep learning techniques has the highest accuracy compared to machine learning algorithms which they have taken.

Aswin et.al [9] different-different deep learning models for rainfall prediction. LSTM, CNN are used in these models. The three distinct gates that make up the LSTM act as its controllers.Artificial Neural Network also used. The graph shows the monthly rainfall distribution and the average rainfall distribution.

Suganya et.al [10] developed a machine learning algorithm to predict rain. We employ the LSTM and ConvNet deep learning architectures.Clustering, Classification and Regression are used. MATLAB, Nntool and Pandas, scikit learn, handoop tools are also used for the best accuracy.

Chalachew et.al [11] using the dataset with amount of daily rainfall they had predicted the machine learningtechniques. Finally, rainfall is predicted based on the atmosphere. MLR, RF and XGBOOST algorithms are used forpredicting rainfall. XGBOOST is the better algorithm for predicting the weather among these three algorithms.

AfanGalih Salman et al. [12] predicted weather have done two experiments with two datasets for training and testing data namely ENSO(El Nino Southern Oscillation) dataset, Weather dataset and the datasets are collected from Kaggle. The first experiment is done using ENSO dataset in which the dataset is split into 75% training and 25% testing it performed with accuracy of 84.8% of Recurrent neural network. The second experiment is done using Weather dataset and the dataset is split into 50% training and 50% testing it performed with accuracy of 59.9%

Mihir Bhawsaretal.[13] have done weather forecasting using information gatheredatthe meteorologist's center. And have done research on howthemachine learning techniques like Classification,Clustering, Decision tree and deep learning techniques like Recurrent Neural network, (CRBM), Convolution network (CN) works for forecasting weather.

Xiaoli Ren et.al [14] based on the weather prediction using the deep learning techniques the weather forecasting hasdone. They have used different kinds of algorithms like DNN, RNN and LSTM. They used the convolution LSTMfor the prediction of weather is done. The dataset is trained and tested according to the algotithms. RNN Has thehighest accuracy compared to the other models.

Tomah Sogabe et.al [15] did research on weather forecasting using deep learning models and predicted the resultsby using two combination of algorithms and they are RNN and LSTM. With the help of these models, they predictedthe accurate values for the models and the LSTM Acquires the highest accuracy among the algorithms.

## III.Problem Statement:

Design machine learning and deep learning models that predict rainfall in a certain area using previous weather data and other relevant information.Predicting rainfall is crucial for many industries, including agriculture, hydrology, water management, and disaster relief. Planning agricultural cycles, irrigation systems, and flood control can benefit from accurate rainfall forecasts.To construct a sustainable deep learning model capable of accurately forecasting rainfall in a specific area

**DATASET DESCRIPTION:**
The "weatherAUS" dataset is a collection of daily weather observations recorded by the Australian Bureau of Meteorology from various weather stations dispersed around Australia. Information from 49 weather stations is included in the dataset, which covers the years 2007 to 2017.The dataset includes a wide range of weather-related parameters, such as temperature, humidity, rainfall, wind speed, and air pressure. The goal variable is the total rainfall, expressed in millimetres, for each day's 24-hour period ending. Total number of records in the dataset is 142,193, and each one corresponds to one day's worth of weather observations. During data preprocessing, it is necessary to deal with the missing values in a few of the dataset's features.With this dataset, machine learning models may be created to forecast how much rain will fall in a specific area on a given day. Several applications, including agriculture, water resource management, and flood predictions, can benefit from this.

**DATASET DETAILS:**
The dataset is collected from Kaggle. The dataset consists of Date, Location, Minimum Temperature, Maximum Temperature, Rainfall, Evaporation, Sunshine, Wind Gust direction, Wind gust Speed, Wind direction 9AM, Wind direction 3PM, Wind speed 9AM, Wind speed 3PM, Humidity at 9AM, Humidity at 3PM, Pressure at 9AM, Pressure at 3PM, Cloud, Temperature at 9AM, Temperature at 3PM, Rain Today, Rain Tomorrow. 142,193 of records in dataset.

## IV.Proposed System:

The system has climate data from the past, such as heat, humid, wind speed, and altitudeand geographic position, as well as other relevant elements including topography, land use, and vegetation cover. Prior to being input into both standard machine learning (logistic regression, decision tree, and random forest) and deep learning algorithms, the data is preprocessed and cleaned (MLP).The output from each algorithm is then integrated using an ensemble learning technique, which helps to increase the accuracy of the system as a whole. The model's accuracy is then assessed by comparing the projected rainfall levels to actual rainfall data.

The design offers a comprehensive and adaptable system for predicting rainfall by using both conventional machine learning techniques and deep learning algorithms. Deep learning algorithms are very good at dealing with complicated, nonlinear interactions between input components, which can increase the accuracy of the system as a whole.

Overall, this architecture offers a reliable and flexible method for rainfall forecasting, It has applications in a number of areas, such as agro, hydrology, water management, and disaster response.

**DATA EXPLORATION:**
Normally data exploration is the crucial stage in building a model. Some steps are taken to explore the data:

**Gathering the Data**: The rainfall data was collected from different sources like weather prediction centers, and from the satellites.

**Data Cleaning**: Missed values and the errors are checked and will rectify the data

**FEATURE ENGINEERING:**
While constructing a model to predict rainfall, feature engineering is a vital phase. Some factors that can be considered for rainfall prediction also included the following:

**Temporal Features:**Create components that record the dynamic behavior of rainfall, such as the hour of the day, the day of the week, the month of the year, and the season. These features can help to detect any recurrent trends in the data.

**Weather - Related Features:** Establish features that track weather-related data that can affect precipitation, such as gravity, air velocity, humid, and climate.These characteristics can aid in identifying any weather-related patterns in the data.

## V.METHODOLOGY:

### MULTILAYER PERCEPTRON:

The multilayer perceptron (MLP) is a fully linked feed-forward artificial neural network.A linear activation function compensates this arrangement. To imitate the frequency of biological neurons' action potentials, or firing, some MLP neurons contain a nonlinear activation function. When each piece of data is examined, the perceptron learns by changing the connection weights in accordance with the degree of error in the output relative to the predicted outcome. The supervised learning model includes it. The Representation of ROC Curve is in fig(2).

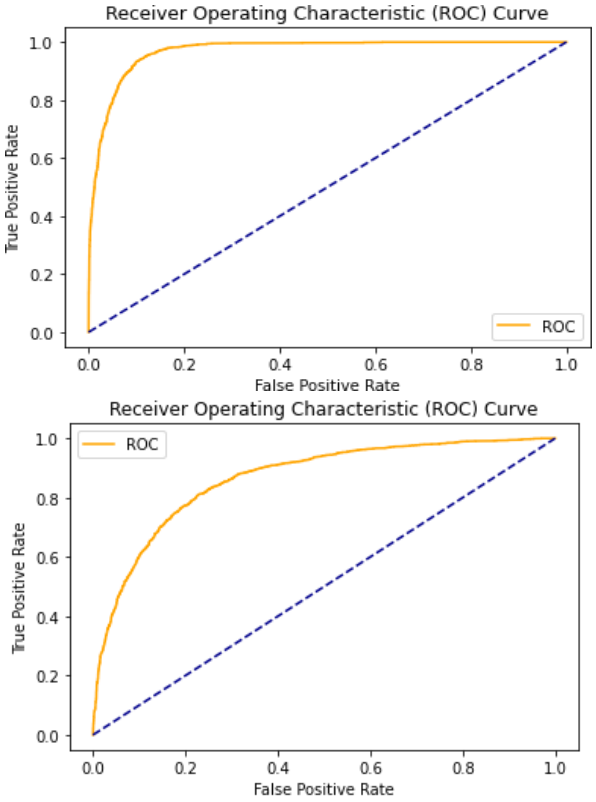


Fig.2.ROC Curve of MLP

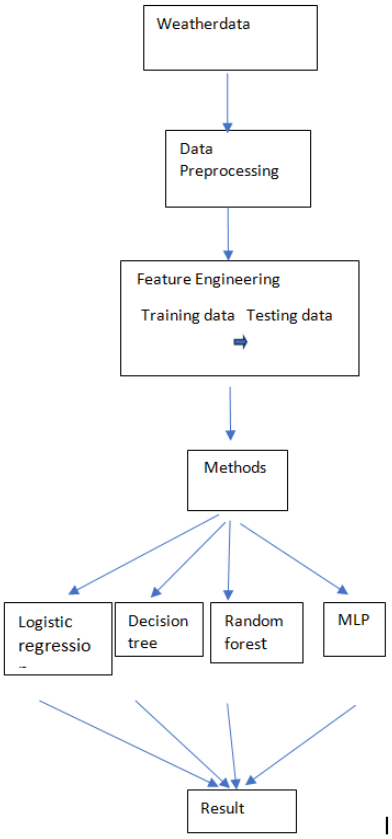


Fig.3.Correlation matrix for MLP

From the above fig(3) represents the correlation matrix and the values are True positive rate is 0.56, False positive rate is 0.037, False negative rate is 0.054, True negative rate is 0.35.

Accuracy obtained by using this model is 0.979.
ROC Area under Curve using multilayer perceptron is 0.908.

### LOGISTIC REGRESSION:

The statistical method of logistic regression can be used to determine the probability that it willrain.The process involves creating a prediction model that uses historical weather data to anticipate the likelihood of precipitation on a certain day. Using the acquired weather data, the logistic regression model determines the likelihood of rainfall. The model calculates the likelihood that it will rain based on the values of the weather variables. The result of the model is a probability score with a range of 0 to 1. Then, by establishing a threshold value, the status of a certain day is decided. The Logistic Regression ROC Curve is Represented in fig(4).

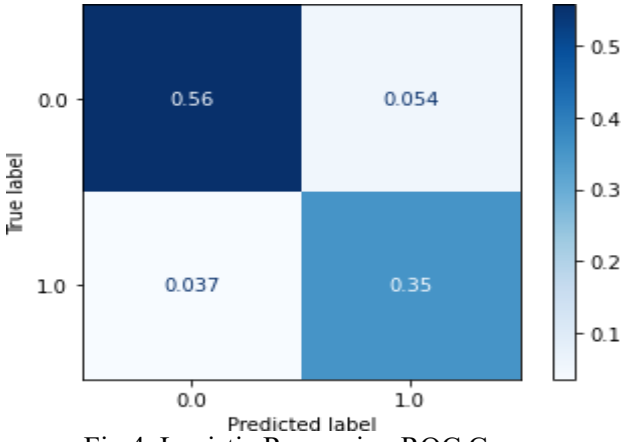


Fig.4. Logistic Regression ROC Curve

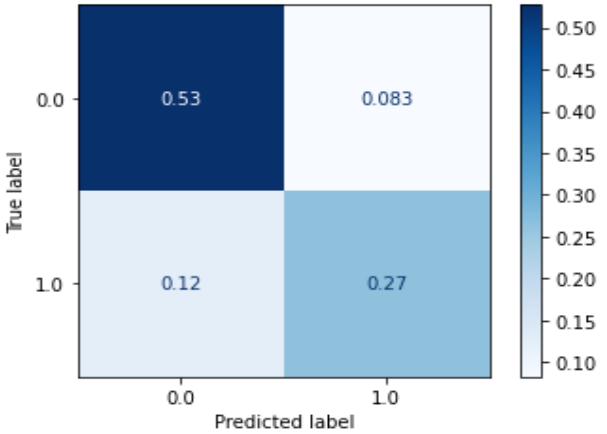


Fig.5. Logistic Regression correlation matrix

From the above fig(5) shows the correlation matrix values. True positive rate is 0.53, True Negative rate is 0.27, False positive rate is 0.12, False negative rate is 0.083.
ROC Area under Curve = 0.7723646519867347
Accuracy = 0.7925747453729055

**DECISION TREE:**
Decision trees can be trained to forecast whether it will rain or not using input features like climate, humid,wind speed, and pressure. The decision tree approach produces a tree structure by repeatedly breaking the data into subgroups depending on the input's value features until a halting condition is reached.Each node of the decision tree runs a test on one of the input features to determine which branch of the tree should be taken. The test separates the data into two or more subsets based on a threshold value of the input characteristic. Up until a leaf node with the projected output label is reached, the process is repeated. The ROC Curve for Decision tree is placed in fig(6).

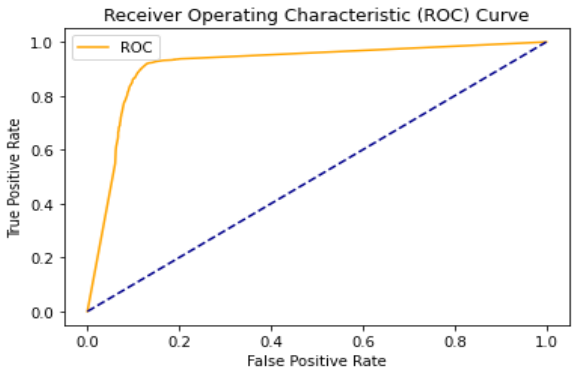


Fig.6. ROC Curve of Decision tree

ROC Area under Curve = 0.8903422709778986
Accuracy = 0.8894973168327674

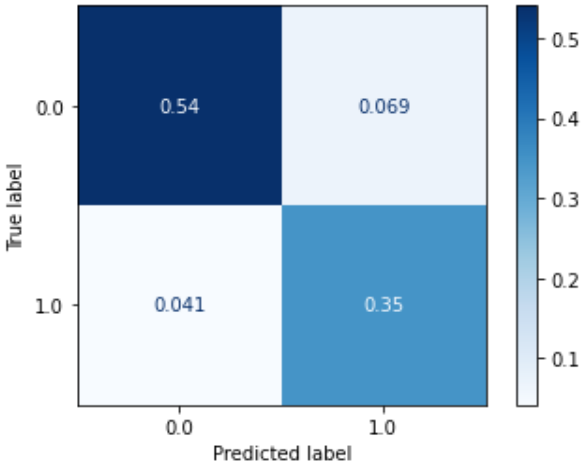


Fig.7. Correlation matrix decision tree

In the above fig(7) represents the decision tree correlation matrix and the matrix consists of the values. The value of True positive rate is 0.54, The value for false positive rate is 0.041, The value for True negative rate is 0.35, The value for False negative rate is 0.069.

**RANDOM FOREST CLASSIFIER:**
A random forest classifier can be trained on historical meteorological data using input factors such as humid, gravity, air velocity, and altitude to forecast whether it will rain or not. The method creates a collection of decision trees by randomly selecting a fraction of the data and characteristics at each split.

After being trained on a separate subset of the data and features, each decision tree in the random forest makes a vote for the final forecast. The output of the random forest classifier is the majority vote of each individual decision tree.

The random forest classifier, once trained, can be used to predict the likelihood of rainfall given new input data. The method evaluates the input features and combines the output of each tree to provide the final prediction. The ROC Curve for random forest classifier is represented in fig(8).

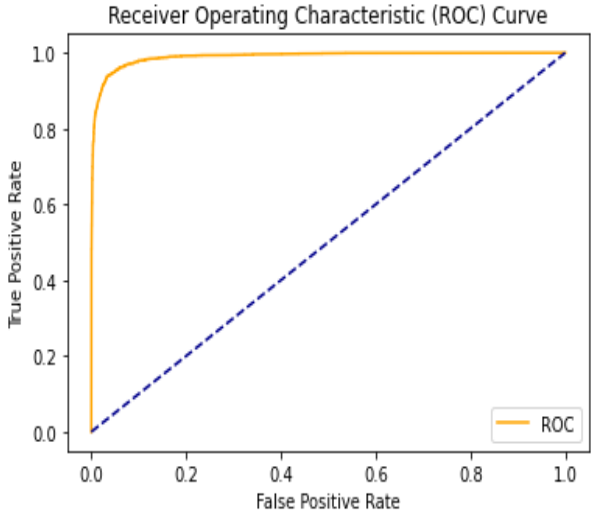


Fig.8. Random Forest Classifier ROC Curve

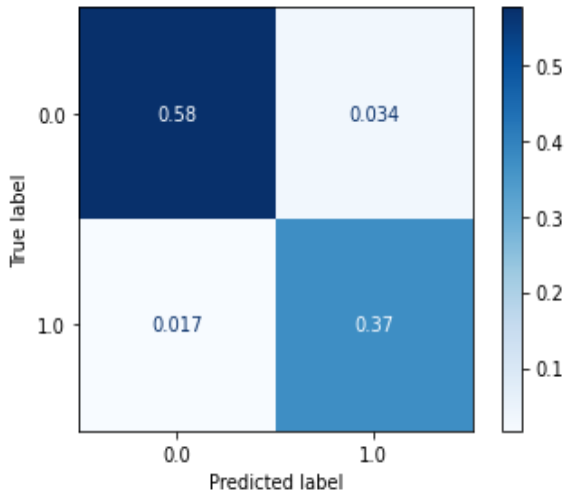


Fig.9.Correlation matrix for random forest classifier

The fig(9) represents the correlation matrix. The values of correlation matrix are as follows: True positive rate is 0.58, False positive rate is 0.017, False negative rate is 0.034, True negative rate is 0.37. The accuracies are represented below.
ROC Area under Curve = 0.9500405619618622
Accuracy = 0.9487460300076662

## VI. Experimental Results:

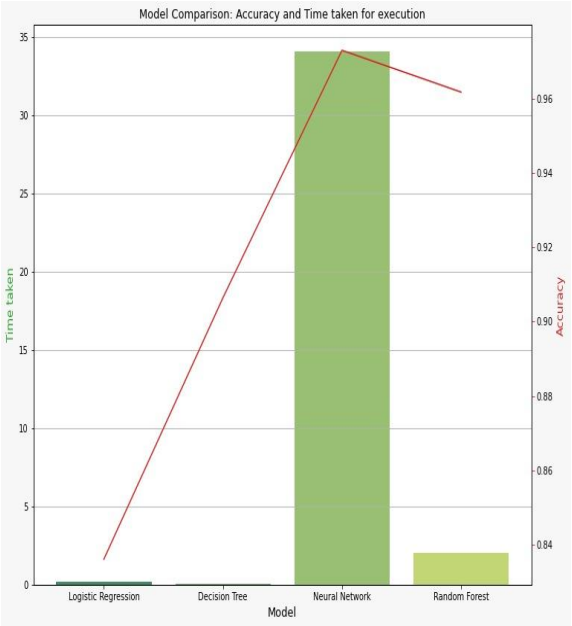


Fig.10.Comparision of Deep learning and Machine learning Algorithms histogram

Multilayer Perceptron(MLP) is a type of neural network that can learn complex non-linear relationships between input features and output labels, and performed well with high accuracy 97%. MLP has several advantagescompared to other machine learning techniques like random forest, decision tree, and logistic regression, including its ability to handle non-linear relationships and its high flexibility in modeling complex systems.

Table.1. Predicting the Algorithms and their accuracies.

| Algorithms | Accuracy |
|---|---|
| Multi layer Perceptron | 0.97 |
| Logistic Regression | 0.79 |
| Decision Tree | 0.88 |
| Random Forest Classifier | 0.94 |

Table(1) shows the algorithms with their accuracies. Multilayer perceptron has the highest accuracy of 97%, Logistic Regression has the accuracy of 79%, Decision tree has the accuracy of 88%, Random forest has the accuracy of 94%.

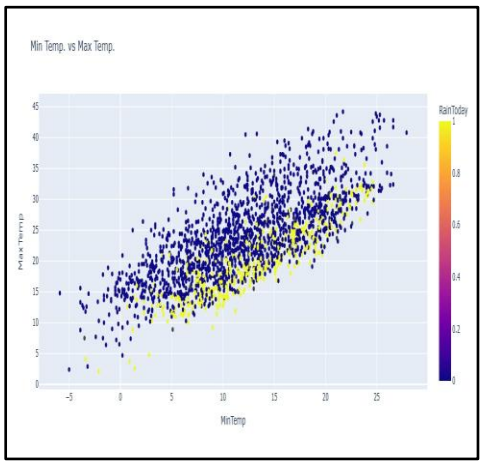

Fig.11.Scatter plot For Minimum temperature and maximum temperature

## V.CONCLUSION:

As a result, rainfall prediction is a critical tool for many different businesses and endeavours, and advances in technology and data processing methods have substantially increased its precision and dependability. Rainfall prediction using the algorithms of logistic regression, decision tree, random forest, and MLP (multi-layer perceptron) were compared, and MLP(multi-layer percepton) was found to have the highest accuracy, it means that the MLP(multi-layer percepton) algorithm performed better in predicting rainfall based on the dataset and evaluation metrics used.

## REFERENCES:

[1] Rahman, A. U., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., ... &Mosavi, A. (2022). Rainfall prediction system using machine learning fusion for smart cities. Sensors, 22(9), 3504.

[2] 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.

[3] Sun, D., Wu, J., Huang, H., Wang, R., Liang, F., & Xinhua, H. (2021). Prediction of short-time rainfall based on deep learning. Mathematical Problems in Engineering, 2021, 1-8.

[4] Endalie, D., Haile, G., & Taye, W. (2022). Deep learning model for daily rainfall prediction: case study of Jimma, Ethiopia. Water Supply, 22(3), 3448-3461.

[5] Oswal, N. (2019). Predicting rainfall using machine learning techniques. arXiv preprint arXiv:1910.13827.

[6] Adewoyin, R. A., Dueben, P., Watson, P., He, Y., & Dutta, R. (2021). TRU-NET: a deep learning approach to high resolution prediction of rainfall. Machine Learning, 110, 2035-2062.

[7] Balan, M. S., Selvan, J. P., Bisht, H. R., Gadgil, Y. A., Khaladkar, I. R., &Lomte, V. M. (2019). Rainfall prediction using deep learning on highly non-linear data. International Journal of Research in Engineering, Science and Management, 2(3), 590-592.

[8] Basha, C. Z., Bhavana, N., Bhavya, P., & Sowmya, V. (2020, July). Rainfall prediction using machine learning & deep learning techniques. In 2020 international conference on electronics and sustainable communication systems (ICESC) (pp. 92-97). IEEE.

[9] Aswin, S., Geetha, P., &Vinayakumar, R. (2018, April). Deep learning models for the prediction of rainfall. In 2018 International Conference on Communication and Signal Processing (ICCSP) (pp. 0657-0661). IEEE.

[10] Grace, R. K., &Suganya, B. (2020, March). Machine learning based rainfall prediction. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 227-229). IEEE.

[11] Liyew, C. M., & Melese, H. A. (2021). Machine learning techniques to predict daily rainfall amount. Journal of Big Data, 8, 1-11.

[12] Salman, A. G., Kanigoro, B., &Heryadi, Y. (2015, October). Weather forecasting using deep learning techniques. In 2015 international conference on advanced computer science and information systems (ICACSIS) (pp. 281-285). Ieee.

[13] Bhawsar, M., Tewari, V., &Khare, P. (2021). A survey of weather forecasting based on machine learning and deep learning techniques. International Journal of Emerging Trends in Engineering Research, 9(7).

[14] Ren, X., Li, X., Ren, K., Song, J., Xu, Z., Deng, K., & Wang, X. (2021). Deep learning-based weather prediction: a survey. Big DataResearch, 23, 100178

[15] Sogabe, T., Ichikawa, H., Sakamoto, K., Yamaguchi, K., Sogabe, M., Sato, T., & Suwa, Y. (2016, November). Optimization ofdecentralized renewable energy system by weather forecasting and deep machine learning techniques. In 2016 IEEE Innovative Smart GridTechnologies-Asia (ISGT-Asia) (pp. 1014-1018). IEEE.