International Journal of Data Science and Analysis
2024; 10(6): 109-128
http://www.sciencepublishinggroup.com/j/ijdsa
doi: 10.11648/j.ijdsa.20241006.11
ISSN: 2575-1883 (Print); ISSN: 2575-1891 (Online)

ANN-based and DT-based Classification Approaches to Predict the Rainfall Level of the Grid ($90°E - 92°E, 23°N - 25°N$) in Bangladesh

Md. Habibur Rahman

Department of Statistics and Data Science, Jahangirnagar University, Dhaka, Bangladesh

Email address:

habib.drj@juniv.edu

To cite this article:

Md. Habibur Rahman. (2024). ANN-based and DT-based Classification Approaches to Predict the Rainfall Level of the Grid ($90°E - 92°E, 23°N - 25°N$) in Bangladesh. *International Journal of Data Science and Analysis, 10*(6), 109-128.
https://doi.org/10.11648/j.ijdsa.20241006.11

Received: 29 October 2024; **Accepted:** 10 December 2024; **Published:** 18 December, 2024

Abstract: The study area defined by the coordinates ($90°E - 92°E, 23°N - 25°N$) is a significant region in Bangladesh, where accurate rainfall predictions are crucial for both the local population and policymakers. Understanding rainfall patterns in this area is vital for effective planning and resource management. Data on atmospheric variables, including temperature, rainfall, humidity, sea level pressure, and wind speed were collected from the Bangladesh Meteorological Department for various locations across the study grids for the period of 1964 to 2015. The descriptive statistics revealed that the pattern of the data of climate parameters is not normal. This dataset serves as the foundation for analyzing climate parameters and forecasting rainfall levels within the specified regions of Bangladesh. This study evaluates machine learning techniques, focusing on artificial neural networks (ANN) and classification and regression trees, C5.0, Random Forest, and Gradient Boosting as alternatives to traditional statistical models for predicting atmospheric phenomena. It reveals that conventional models often rely on assumptions unsuitable for chaotic systems like the atmosphere. Among the assessed models ANN, CART, C5.0, Random Forest (RF), and Gradient Boosting Machines (GBM) the ANN demonstrated the highest predictive capabilities for rainfall forecasting in Bangladesh, achieving superior training accuracy and Kappa values while also being recognized as the best overall performer based on ranking metrics.

Keywords: ANN, Decision Tree, GBM, Cross-validation, Rainfall, Bangladesh

1. Introduction

Rainfall is crucial for Bangladesh, significantly impacting various life aspects. It plays an essential role in the agricultural sector, which is a major economic contributor. Sufficient rainfall promotes crop growth, thereby ensuring food security for the population. It affects planting, irrigation, and harvesting schedules, directly influencing crop yields and farmers' livelihoods. Additionally, rainfall is vital for replenishing water sources like rivers, lakes, and groundwater, which are necessary for drinking water, sanitation, and domestic and industrial uses. While excessive rainfall can cause floods, regulated and predictable rainfall is essential for managing river and reservoir water levels to prevent droughts and severe flooding. Bangladesh depends on hydropower as a renewable energy source; consistent rainfall ensures adequate water in reservoirs for electricity generation, thus supporting the country's energy supply. Rainfall also sustains ecosystem health by nourishing forests, wetlands, and wildlife habitats, maintaining biodiversity and environmental balance. Various economic activities such as fishing, transportation, and industry rely on sufficient rainfall. Furthermore, rainfall affects public health by influencing water quality and availability, ensuring clean water supplies, and controlling airborne dust and pollutants. Lastly, rainfall is integral to climate balance by regulating temperatures and supporting weather patterns essential for diverse ecological and human activities. In summary, rainfall is vital for agriculture, water resources management, flood control, hydropower generation, ecosystem health, economic activities, public health, and maintaining climate balance in Bangladesh.

Predicting rainfall in Bangladesh is essential for various reasons. As an agricultural nation, a large segment of the population relies on farming for their livelihoods. Accurate rainfall forecasts empower farmers to make informed choices regarding planting, irrigation, and harvesting, which enhances crop yields and reduces losses. Due to its low-lying terrain and extensive river networks, Bangladesh is particularly vulnerable to flooding. Timely and accurate rainfall predictions are crucial for effective flood management and mitigation, enabling authorities to issue warnings and execute evacuation plans to safeguard lives and property. In addition to flooding, the country faces other weather-related challenges such as cyclones and landslides. Rainfall forecasting is a critical component of early warning systems, enhancing preparedness and response strategies to lessen the impact of these disasters on communities. Effective management of water resources depends on precise rainfall predictions, which facilitate the planning and operation of reservoirs, dams, and irrigation systems, ensuring water availability during dry spells while controlling excess during the rainy season. Heavy rainfall can result in waterlogging and the spread of waterborne diseases; reliable rainfall forecasts allow health authorities to prepare for and mitigate risks associated with potential disease outbreaks.

Accurate rainfall forecasting is vital in urban settings for the design and maintenance of effective drainage systems to avert urban flooding, an issue that is increasingly pressing due to ongoing urbanization in Bangladesh. Weather-related disruptions can significantly affect the economy. Reliable rainfall predictions help various sectors, including agriculture, infrastructure, and public services, to minimize disruptions and plan more effectively, thereby enhancing economic stability. As one of the countries most susceptible to climate change effects, dependable rainfall forecasts are essential for formulating long-term strategies to adapt to shifting weather patterns and mitigate adverse impacts. Accurate rainfall forecasting is crucial for ensuring food security, protecting lives and property, managing water resources, maintaining public health, supporting economic stability, and adapting to climate change in Bangladesh. Rainfall significantly influences reservoir water levels, with climate change causing unpredictable variations that can lead to either overflow or drought. This study utilized several machine learning models to predict rainfall in Tasik Kenyir, and Terengganu, employing Bayesian Linear Regression, Boosted Decision Tree Regression, Decision Forest Regression, and Neural Network Regression across various scenarios and timeframes. The results indicated that Boosted Decision Tree Regression achieved the highest accuracy [1].

Bangladesh, located in South Asia, receives significant rainfall each year, particularly during the monsoon season. Accurate rainfall predictions for specific regions are essential to prevent flooding and agricultural losses in the country. This study focuses on forecasting rainfall levels in the grid area defined by $(90°E - 92°E, 23°N - 25°N)$, which has been identified as vulnerable to flood risks due to heavy precipitation. Rainfall is a crucial climatic factor that profoundly impacts Bangladesh's economy, society, and environment. The country is situated in a region highly prone to climate variability and change, leading to frequent natural disasters such as floods, landslides, and droughts. These disasters adversely affect the economy, agriculture, infrastructure, and human lives. Consequently, precise and reliable rainfall predictions are vital for effective decision-making across various sectors, including agriculture, water resource management, disaster management, and urban planning.

Several researchers have explored the uncertainties related to weather systems [2-8]. A variety of studies have concentrated on data mining, particularly through the use of classification algorithms [9-12]. One of the most effective and straightforward methods for extracting information from large datasets is the decision tree construction technique. This paper reviews and discusses various decision tree algorithms applied to different datasets. Numerous studies confirm the effectiveness of data mining techniques for predictive purposes. Weather forecasting, including the prediction of precipitation labels, is a chaotic system influenced by both temporal and spatial factors. It represents one of the most critical and challenging tasks undertaken by meteorological services globally, requiring expertise from multiple specialized fields. The complexity in meteorology arises from making decisions amid uncertainty. A review highlighted the use of artificial neural networks in weather prediction and examined their advantages [13]. The chaotic characteristics associated with atmospheric phenomena have also garnered interest from contemporary scientists [9-12, 14]. Classification and Regression Trees (CART) utilize the GINI Index as a splitting criterion within a binary tree structure [15, 16]. CART can handle both nominal and continuous attributes while constructing decision trees. It also accommodates missing values through surrogate tests to achieve fairly accurate results and employs cross-validation in its pruning method. The researcher presents a small application of the CART algorithm for weather prediction in Hong Kong based on factors such as year, month, average pressure, relative humidity, cloud cover, precipitation, and average temperature [17]. The authors in the study applied data mining techniques to analyze weather data and uncover hidden patterns within large datasets [18]. This approach transformed the collected information into actionable knowledge for classifying and predicting weather conditions in Gaza City, providing valuable forecasts that aided decision-making processes. Kumar compared various classification methods including Decision Trees, Rule-based Methods, Neural Networks, Naive Bayes, Bayesian Belief Networks, and Support Vector Machines [19]. Effective weather prediction also requires advanced statistical models that do not rely on assumptions about the underlying system.

Numerous global studies have developed stochastic weather models, which are statistical models functioning as random number generators. These models produce outputs that closely mimic the weather data they are designed to replicate [3]. The artificial neural network (ANN), first introduced in 1964 [20], has become a prominent soft computing technique in weather forecasting. Chattopadhyay et al. specifically applied soft computing methods for rainfall prediction [21]. Kalogirou et. al. utilized ANN for forecasting rainfall by dividing data into homogeneous sub-populations [22]. Other soft computing techniques, such as self-organizing maps, backpropagation neural networks, and fuzzy rule systems, have also been used for rainfall prediction [23]. Michaelides et al. also conducted a comparative analysis of ANN and multiple linear regression for estimating missing rainfall data in Cyprus [24]. Badr et. al. developed ten distinct statistical models, incorporating both linear and nonlinear methods, to forecast seasonal rainfall anomalies in the Sahel region [25]. Generalized linear models (GLMs) and generalized additive models (GAMs) are regression models based on likelihood principles. GLMs extend the standard ordinary least squares (OLS) model [26] by integrating a link function that connects the mean of the response variable to the predictors [27]. GAMs replace the link function in GLMs with a nonparametric smoothing function, allowing for nonlinear relationships between predictors and responses [28]. Given that Sahel rainfall anomalies follow a normal Gaussian distribution, a normal identity link function was employed in constructing the GLM, while a cubic regression spline was used in the GAM. In addition to GLM, GAM, and MARS (Multivariate Adaptive Regression Splines), four tree-based modeling techniques were applied: classification and regression trees [16], Bayesian additive regression trees (BART) [29], bagged categorical and regression trees [30], and a random forest (RF) model [31]. The tree models underwent cost-complexity pruning, and suitable prior information for error variance was utilized in the BART model [25].

Meteorological phenomena have historically been forecasted using statistical models such as simple linear regression, multiple linear regression, and the Markov Chain Model. However, these models rely on assumptions that may not hold in the chaotic nature of the atmosphere. Numerical models, which are based on nonlinear operator equations, provide an alternative but are sensitive to initial conditions, complicating the solutions. In contrast, machine learning methods do not depend on such assumptions, allowing for rapid information processing and effective mapping of input to output variables. This capability makes them particularly suitable for modeling complex nonlinear relationships. In this research, the following steps will be undertaken: data collection and exploratory data analysis, which will include visualization, outlier detection, and data trimming to refine and summarize atmospheric data. Various machine learning models, including the MLP model and CART model, will be fitted to the data. The fit quality of these models will be assessed using performance metrics, and the best-performing model will be selected.

In recent years, machine learning (ML) algorithms have been utilized to develop precise rainfall prediction models, surpassing traditional statistical methods. Among the most commonly used ML models in rainfall prediction studies are artificial neural networks (ANNs) and decision trees (DTs). ANNs are nonlinear models that can learn from large datasets and make accurate predictions based on input data. The researchers are inspired by the human brain's structure, where numerous interconnected neurons work together to process and analyze information. Conversely, DTs are a machine learning algorithm that predicts outcomes by recursively splitting the input data into subsets based on the characteristics. This study aims to predict rainfall levels in the grid area defined by $(90°E - 92°E, 23°N - 25°N)$ in Bangladesh, a region prone to flood risks due to heavy rainfall during the monsoon season. The primary goal of this research is to compare the performance of ANN-based and DT-based classification methods in predicting rainfall levels in this area. We will train and test both models using a comprehensive dataset of historical rainfall data for the region and evaluate their performance using various metrics. This study seeks to contribute to the development of accurate and reliable rainfall prediction models that can aid effective flood and disaster management decision-making processes. Machine learning techniques have increasingly been applied in weather prediction and forecasting, with ANNs and DTs being among the most favored methods. ANNs are artificial intelligence models that learn from extensive datasets to make accurate predictions based on input data, while DTs make predictions by recursively partitioning the input data into subsets according to their features. This research is significant as it supports the creation of accurate and dependable rainfall prediction models that can improve decision-making processes for flood and disaster management. Reliable rainfall prediction models can help mitigate or reduce the impact of natural disasters by providing early warning systems to stakeholders, including farmers, water managers, and policymakers. Additionally, these models can facilitate efficient resource allocation, such as relief supplies, during and after a disaster. Furthermore, comparing the performance of ANN-based and DT-based classification methods in predicting rainfall levels can provide insights into the relative strengths and weaknesses of both models. This information can be valuable for researchers and practitioners in the field of machine learning, guiding future research and the development of rainfall prediction models. Overall, the findings of this study may influence the application of machine learning in climate and weather prediction. Moreover, this research adds to the growing body of literature on rainfall prediction in Bangladesh, which is vital for sustainable development and disaster risk reduction initiatives in the country. The main objectives of this research are as follows: to analyze the characteristics of atmospheric data, to apply machine learning models for predicting precipitation labels in the selected study areas, to evaluate the quality of fit of the machine learning models using performance metrics, to assess the cross-validation results of the models fitted to the study areas, and to identify the best-fitting model among those used for the study regions. These studies highlight the potential of machine learning algorithms for accurate rainfall prediction while suggesting that the choice of algorithm may depend on specific regional characteristics and dataset attributes. In Bangladesh, further research is needed to determine the most effective machine learning algorithms for rainfall prediction, particularly in specific regions or localities.

2. Method and Materials

Atmospheric precipitation forecasting will be analyzed for environmental phenomena. The previous section covered earlier research, along with the background and rationale for this study. In the upcoming section, the methodologies employed to forecast rainfall levels will be detailed. Various machine learning techniques, including Artificial Neural Networks (ANN), Classification and Regression Trees (CART), Random Forest (RF), and Gradient Boosting Machines (GBM), were utilized to make rainfall predictions. To identify the most effective predictive model, the performance of these machine learning methods was assessed using specific evaluation metrics.

2.1. Artificial Neural Network

Artificial Neural Network (ANN) is a fundamental model that includes at least three distinct layers: an input layer, multiple hidden layers, and an output layer.

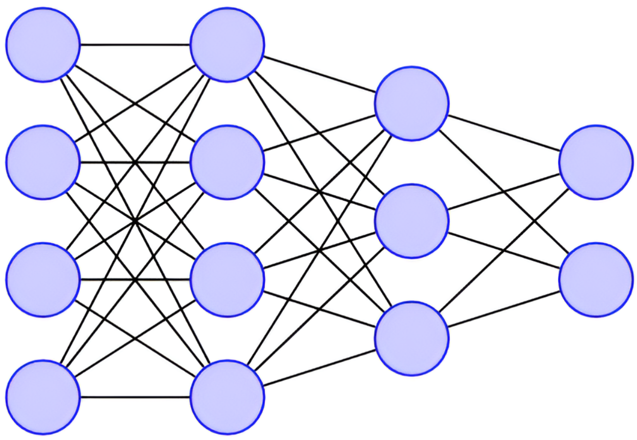

Figure 1. An Artificial Neural Network (ANN) with four nodes at the input layer, four nodes at the first hidden layer, three nodes at the second hidden layer, and finally two nodes at the output layer.

It necessitates that neurons in adjacent layers are fully interconnected, which results in a substantial number of weight parameters that must be optimized during training. The ANN architecture applies to various types of learning, including supervised learning, unsupervised learning, and reinforcement learning (RL). While this model enjoyed widespread use in earlier applications, its popularity has diminished due to its complexity, slow convergence rates, and average performance. Nevertheless, Multi-Layer Perceptron (MLP) serves as a foundational model for more sophisticated neural network architectures. For instance, the advanced adaptive learning neural network (AdaNet) allows MLPs to adjust their structures dynamically in response to the training dataset, making it suitable for optimizing mobile networks that are continuously evolving. The Multi-layer Perceptron (MLP) was created to address this limitation. It is a type of neural network that enables a non-linear relationship between inputs and outputs. An MLP consists of input and output layers, along with one or more hidden layers that contain numerous interconnected neurons. In contrast to the original Perceptron, where each neuron is required to have a specific activation function that sets a threshold such as ReLU or sigmoid neurons in a Multilayer Perceptron can utilize any chosen activation function.

Backpropagation is the learning algorithm that enables the Multilayer Perceptron to iteratively update the weights in the network, to minimize the cost function. For backpropagation to function correctly, there is a crucial requirement. The function that combines inputs and weights in a neuron, such as the weighted sum, and the activation function, like ReLU, must be differentiable. These functions need to have a bounded derivative because Gradient Descent is commonly employed as the optimization algorithm in the Multilayer Perceptron.

2.2. Decision Tree Induction

Decision tree (DT) learning is a type of supervised learning widely used in data mining, statistics, and machine learning. Originating in the late 1970s and early 1980s, J. Ross Quinlan, a prominent figure in machine learning, developed the decision tree method known as Iterative Dichotomiser (ID3) [32]. ID3 was succeeded by C4.5 and Classification and Regression Tree (CART), which further refined the decision tree approach. According to Han *et al.* ID3, C4.5, and CART utilize an optimization technique involving a non-backtracking approach, constructing decision tree algorithms in a top-down recursive divide-and-conquer manner [32]. The methodology for decision trees is outlined below.

i) The algorithm is invoked by passing three parameters: D indicates the dataset, the attribute list indicates the list of attribute descriptions, and the Attribute selection method selects the best attribute according to class.
ii) The tree begins with a single leaf, N, which represents the tuples of training in D.
iii) If all of the tuples in D belong to the same class, leaf N will be categorized with the class.
iv) Else the algorithm uses the Attribute selection approach to establish the splitting criterion. The splitting criterion instructs which characteristic to check at node N by selecting the ”optimal” method for separating or partitioning the tuples in D into discrete classes.
v) The splitting criterion is designated at node N, and it acts as a test at the node. For each conclusion of the splitting criterion, a branch is formed from node N. The data items in D are split in this manner.
vi) The method recursively applies the same technique to create a decision tree for the data items for every resulting partition.
vii) The recursive partitioning process is terminated immediately if all of the itemsets in subdivision D are members of the same class, or there are not any additional features that allow the tuples can be further

divided up, or there are also no tuples for a certain branch, hence partition D_j is empty.

viii) The eventual results decision tree is restored.

Classification and Regression Tree: In 1984, statisticians L. Breiman, J. Friedman, R. Olshen, and C. Stone published the book "Classification and Regression Trees" (CART) [33]. CART represents a specific approach within decision tree induction. Similar to other decision tree methods, CART employs the Gini index as a criterion for effectively splitting a given dataset into partitions. The Gini index quantifies the impurity of dataset D, whether it's a partition of data or a set of training instances.

$$Gini(D) = 1 - \sum_{i=1}^{n} p_i^2 \tag{1}$$

Here, p is the probability of tuple D. $Gini(D)$ gives the partitioning as,

$$Gini_A(D) = \frac{|D_1|}{|D|} \times Gini(D_1) + \frac{|D_2|}{|D|} \times Gini(D_2) \tag{2}$$

And finally,

$$\Delta Gini(A) = Gini(D) - Gini_A(D) \tag{3}$$

The splitting property is chosen to maximize impurity reduction. The procedure continues till the data can no longer be split any further. CART is capable of handling both data types, numerical as well as categorical.

C5.0: Computer scientist J. Ross Quinlan introduced the C5.0 algorithm as an enhancement to his earlier C4.5 algorithm [34]. C5.0 is a well-known decision tree method that partitions data based on the attribute that provides the highest information gain. This splitting criterion selects attributes that maximize the information gain at each step. Each subset resulting from the initial partition is recursively divided based on different attributes until further partitioning is not feasible. Compared to other methods, C5.0 decision trees operate similarly but are notably easier to interpret and implement. C5.0 is capable of handling both quantitative and qualitative criteria efficiently. The anticipated information required to categorize a tuple in D is expressed as follows: $Info(D) = -\sum_{i=1}^{m} p_i log_2 p_i$. In this context, p_i represents the nonzero probability that a randomly selected tuple from D is part of class C_i, which is estimated using the formula $\frac{|C_{i,D}|}{|D|}$. Additionally, $Info(D)$ is referred to as the entropy of D.

If attribute A takes on discrete values, these values directly relate to the v outcomes of a test performed on A. Attribute A can divide D into v distinct partitions. The measure $Info_A(D)$ quantifies the expected information needed to classify a tuple from D based on the partitioning created by A. A lower expected information requirement indicates higher purity among the partitions. Information gain is defined as the difference between the initial information requirement and the new requirement. In other words, the information gain is defined as $Gain(A) = Info(D) - Info_A(D)$, indicating how much is gained by making a decision based on attribute A. The attribute A that provides the highest information gain, $Gain(A)$, is selected as the splitting attribute at node N. The measure of information gain tends to favor tests with numerous outcomes, meaning it tends to choose attributes that have a larger number of values. C4.5, which is an advancement of ID3, utilizes a modified version of information gain called the gain ratio to address this bias. It implements a form of normalization for information gain by utilizing a "split information" value that is defined $Info(D)$ as

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right)$$

The Gain ratio (GR) is defined as $GR(A) = \frac{Gain(A)}{SplitInfo_A(D)}$.

2.3. Random Forest

Random Forest, developed by L. Breiman in 2001, stands as one of the most influential algorithms in machine learning [31]. It is a classification method that operates by constructing a collection of tree-structured classifiers. The Random Forest (RF) algorithm selects a random subset of features at each iteration and constructs decision trees using the CART procedure [32]. As described by Hastie et al. (2009), the operation of Random Forest can be summarized as follows [35]:

i) First, the algorithm creates bootstrapping samples Z of the total size of the training data N from a training set with a replacement which is also known as bagging.

ii) Expand the random-forest tree towards the bootstrap sample by recursively repeating the procedures below for each tree terminal node until the minimum node size is attained.
   (a) Among the p variables, choose m variables at random.
   (b) Choose the best variable or split-point from the m options.
   (c) Divide each node into two daughter nodes.

iii) Return the tree ensemble.

This approach effectively builds a robust model capable of handling both qualitative and quantitative data. The resulting Random Forest model can then be employed for classifying new datasets or test datasets.

2.4. Gradient Boosting Machine

Gradient Boosting Machine (GBM) is a well-known boosting method pioneered by Jerome Friedman [36]. GBM is considered comparable in performance to advanced techniques like random forests. Unlike random forests, GBM constructs trees sequentially, where each tree learns from the mistakes (pseudo residuals) of its predecessors [37]. GBM begins with an initial weak learner and iteratively builds trees based on the residuals of the previous models. Each tree is trained on a modified version of the original dataset. GBM operates as a boosting method by sequentially creating trees, aiming to improve predictive accuracy with each iteration. All the individual trees collectively form a single predictive model used for making predictions on new data.

2.5. Assessment of Performance of Algorithm

In this study, these algorithms are assessed using evaluation metrics including accuracy (ACY), Recall (RCL) or sensitivity, specificity (SPE), positive predictive value (PPV), negative predictive value (NPV), Cohen's Kappa (KPA), F_1 score (F_1), Detection Rate (DER), balanced accuracy (BAC), and ROC curve with AUC. The k-fold cross-validation techniques are applied to determine these performance metrics. For each classifier, a confusion matrix is generated, which is a 2×2 table used to evaluate the performance of a classification algorithm. This matrix includes true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The confusion matrix facilitates the calculation of these evaluation metrics [38]. *Sensitivity* denotes the model's capacity to accurately detect true positive cases. It is alternatively known as recall or the true positive rate. Sensitivity is calculated by dividing the total number of correctly identified positive instances by the sum of true positives, which also encompasses false positives. Mathematically, sensitivity can be computed as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

Specificity pertains to the model's ability to correctly identify true negative cases. It is also known as the true negative rate. Specificity is calculated by dividing the total number of correctly identified negative instances by the sum of true negatives, which also includes false negatives. Specificity can be defined as:

$$Specificity = \frac{TN}{FP + TN} \tag{5}$$

The *accuracy* of any predictive method is fundamental in evaluating its performance. It measures the ratio of correctly predicted data points across the entire evaluation. The highest achievable accuracy is 1.0, and the lowest is 0.0. It is easily calculated by dividing the number of correctly predicted points by the total number of projections. Additionally, it can be stated as,

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

This study gathered the highest levels of precision attained using various ML algorithms.

Negative predictive value refers to the model's ability to predict the presence of negative cases among those predicted as negative. NPV is a metric used to evaluate the accuracy of a specific model. Mathematically, NPV can be computed as follows:

$$Negative\ Predictive\ Value = \frac{TN}{FN + TN} \tag{7}$$

Positive predictive value indicates the model's capability to predict the presence of positive cases among those predicted as positive. It is also known as precision. This characteristic assesses the likelihood of someone being a true positive case given a positive test result. PPV can be quantified by

$$Positive\ Predictive\ Value = \frac{TP}{TP + FP} \tag{8}$$

The F_1 score evaluates the accuracy of a test by combining its precision and recall. Precision is also known as positive predictive value, and recall as sensitivity. The F_1 score assesses how accurately a model predicts outcomes across all data. It is calculated using the harmonic mean of a model's precision and recall. The equation below can be used to compute it:

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{9}$$

The *ROC curve*, also called the receiver operating characteristic curve, illustrates how well a classification model performs across various thresholds. It plots two parameters: true positive rate and false positive rate. AUC stands for the area under the ROC curve, which measures the area beneath the entire ROC curve in two dimensions. AUC compares two models and evaluates how effectively a single model performs across different thresholds.

Cohen's kappa statistic provides a more effective method for addressing multi-class and misaligned class issues. It quantifies agreement between expected and actual classifications within a dataset. The kappa statistic can evaluate not only a single classifier but also multiple classifiers together. Its values indicate different levels of agreement: 0 indicates no agreement, 0 to 0.20 suggests slight agreement, 0.21 to 0.40 indicates fair agreement, 0.41 to 0.60 signifies moderate agreement, 0.61 to 0.80 represents substantial agreement, and 0.81 to 1 denotes almost perfect agreement [39]. It can be calculated by

$$Kappa = \frac{Accuracy - Random\ Accuracy}{1 - Random\ Accuracy} \tag{10}$$

Random Accuracy is the ration between {(TN+FP)(TN+FN)+(FN+TP) (FP+TP)} and {(TP+FP+TN+FN)(TP+FP+TN+FN)}. Predicting Precipitation is crucial for both humans and the environment, and various models are employed for this purpose. Each model operates based on its specific algorithms. Evaluating the performance of these models helps identify the most effective one for forecasting Precipitation levels. The next section will delve into the dataset's characteristics used in this study. Understanding the dataset will provide insights into how well these models can be applied to achieve accurate predictions.

3. Data Description

The data on the atmospheric variables- temperature, dew point temperature, maximum temperature, minimum temperature, humidity, rainfall, sea level pressure, and wind speed of different locations of the study grid were collected from the Bangladesh Meteorological Department (BMD) for the period of 1964 to 2015.

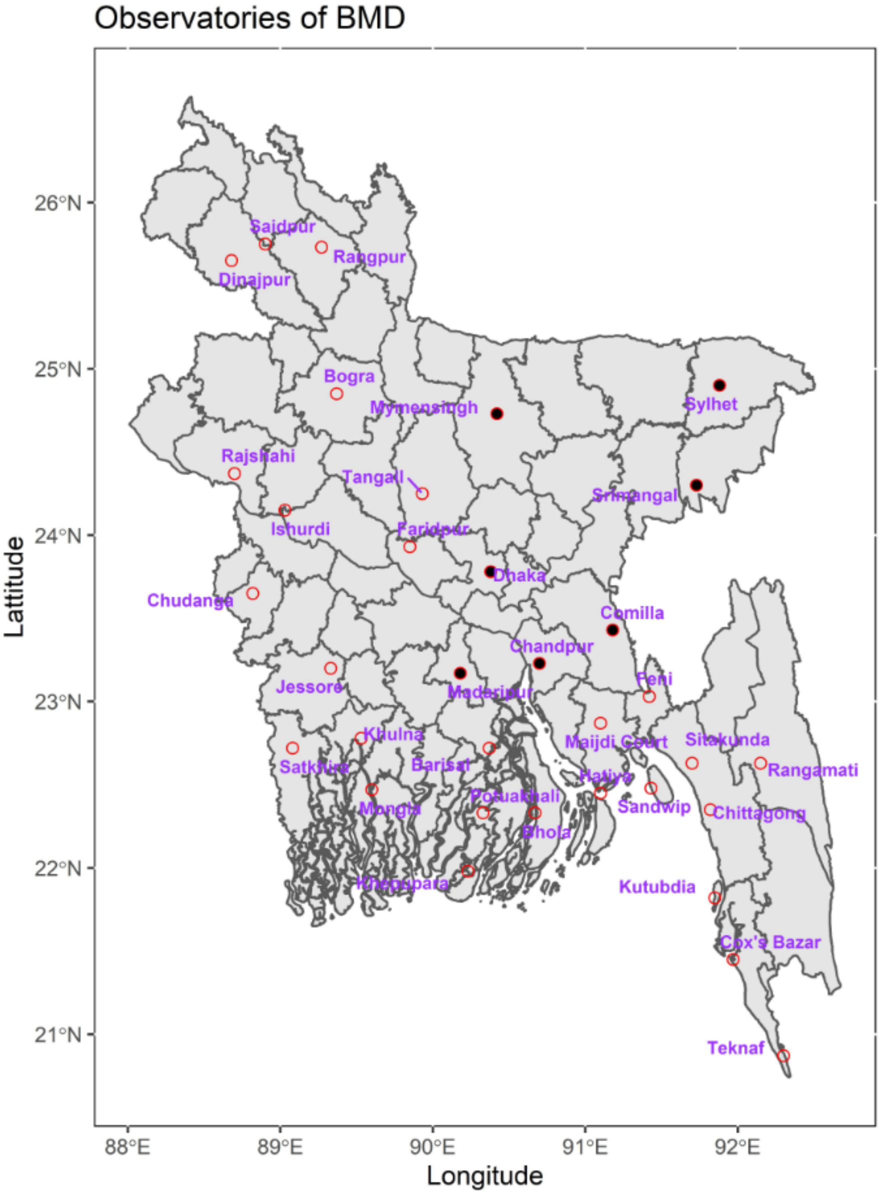

Figure 2. *The 34 observatories of Bangladesh Meteorological Department (BMD) and the locations considered for this study are black-filled circled.*

The whole country is divided into five grids which are Grid $(88^\circ E - 90^\circ E, 25^\circ N - 27^\circ N)$ or North-West Region: Dinajpur, Rangpur, Sydpur; Grid $(88^\circ E - 90^\circ E, 23^\circ N - 25^\circ N)$ or West-Middle Region: Jessore, Chudanga, Faridpur, Ishurdi, Tangail, Rajshahi, Bogra; Grid $(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ or North-East Region: Madaripur, Chandpur, Comilla, Dhaka, Srimangal, Mymensingh, Sylhet; Grid $(88^\circ E - 90^\circ E, 21^\circ N - 23^\circ N)$ or South-West Region: Khepupara, Bhola, Potuakhali, Mongla, Satkhira, Barisal, Khulna; and Grid $(91^\circ E - 93^\circ E, 21^\circ N - 23^\circ N)$ or South-East Region: Teknaf, Cox's Bazar, Kutubdia, Chittagong, Hatiya, Sandwip, Rangamati, Sitakunda, Maijdi Court, Feni. The series for the climate parameters are generated from the available data from the locations of the grid $(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$. The available data for the different stations are presented in Table 1. The available locations (Latitude, Longitude) in the grid $(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ in Bangladesh to collect data are Chandpur (23.23, 90.7), Comilla (23.43, 91.18), Dhaka (23.78, 90.38), Madaripur (23.17, 90.18), Mymensingh (24.73, 90.42), Srimangal (24.3, 91.73), Sylhet (24.9, 91.88), also presented in Table 1 and in the Figure 2. The table shows the available data for different locations in Bangladesh within the longitude range of $90^\circ E$ to $92^\circ E$ and latitude range of $23^\circ N$ to $25^\circ N$ (Table 1). The locations included are Chandpur, Comilla, Dhaka, Madaripur, Mymensing, Srimangal, and Sylhet. The table provides the coordinates (latitude and longitude) for each location, the duration of the data available, and the total number of months of data. The data durations vary, with the earliest starting in 1960 and the latest ending in 2015. The locations with the longest data duration are Comilla, Dhaka, Mymensing, Srimangal, and Sylhet, each with 672 months of data, while Madaripur has the shortest duration with 468 months of data (Table 1). The data for this study are derived from the available location data by calculating the average and the summary results are presented in Table 2 and the correlations between the parameters are illustrated in Table 3.

Table 1. *Available data of the different locations of the Grid $(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ in Bangladesh considered in this study.*

Location (Latitude, Longitude)	Duration	No. of Month
Chandpur (23.23, 90.7)	1964-2015	624
Comilla (23.43, 91.18)	1960-2015	672
Dhaka (23.78, 90.38)	1960-2015	672
Madaripur (23.17, 90.18)	1977-2015	468
Mymensing (24.73, 90.42)	1964-2015	672
Srimangal (24.3, 91.73)	1964-2015	672
Sylhet (24.9, 91.88)	1964-2015	672

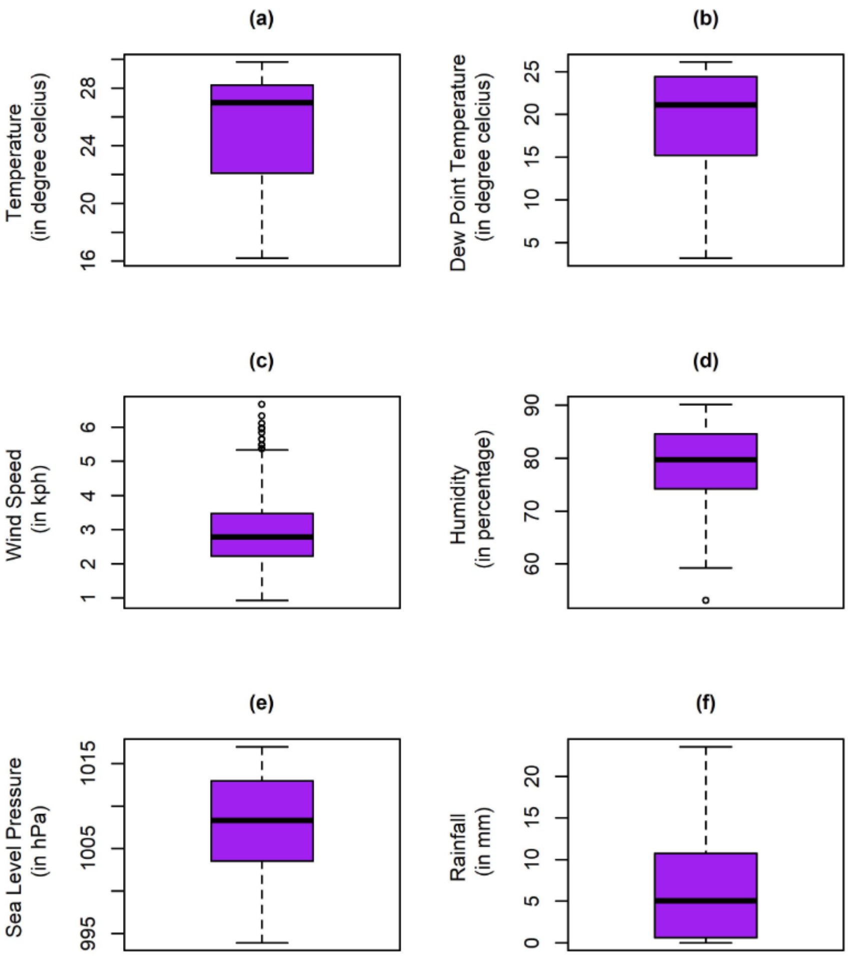

Figure 3. *Box-and-whisker plot for the atmospheric parameters (a) temperature (b) dew point temperature (c) wind speed (d) humidity (e) sea level pressure and (f) rainfall of the Grid $(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ in Bangladesh considered in this study.*

The box-and-whisker plots for the atmospheric parameters are presented in Figure 3. Most of the distribution of the atmospheric parameters is skewed (Figure 3, Table 2). The atmospheric variables temperature, dew point temperature, humidity, and sea level pressure are negatively skewed with skewness -0.85,-0.52, -0.61, -0.10 respectively (Table 2). Besides, the parameter wind speed is positively skewed with skewness 0.81 (Figure 3 and Table 2). The correlations between atmospheric parameters are highly positive and highly negative (Table 3). The sea level pressure and temperature are negatively related with the correlation -0.849 (Table 3).

***Table 2.** Summary Statistics for the atmospheric parameters temperature (TEM), dew point temperature (DPT), wind speed (WIS), humidity (HUM), and sea level pressure (SLP) of the Grid $(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ in Bangladesh considered in this study.*

Statistics	TEM	DPT	WIS	HUM	SLP	RAN
Mean	25.19	19.89	2.91	78.94	1008.14	6.40
Median	27.00	21.10	2.79	79.72	1008.30	5.04
Mode	28.40	24.40	2.34	84.51	1014.10	0.00
Std Dev.	3.77	5.03	0.93	6.56	5.18	6.11
Kurtosis	-0.74	-1.01	0.81	-0.26	-1.24	-0.63
Skewness	-0.85	-0.52	0.83	-0.61	-0.10	0.68
Range	13.60	22.90	5.73	37.03	23.10	23.55
Minimum	16.20	3.20	0.93	53.11	993.90	0.00
Maximum	29.80	26.10	6.66	90.14	1017.00	23.55
Count	672.00	672.00	672.00	672.00	672.00	672.00

***Table 3.** Correlation between the atmospheric parameters - temperature, dew point temperature, wind speed, humidity, sea level pressure, and rainfall of the Grid $(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ in Bangladesh considered in this study.*

Name	TEM	DPT	WIS	HUM	SLP	RAN
TEM	1.000	0.895	0.502	0.484	-0.849	0.697
DPT	0.895	1.000	0.376	0.728	-0.821	0.783
WIS	0.502	0.376	1.000	0.119	-0.563	0.468
HUM	0.484	0.728	0.119	1.000	-0.617	0.728
SLP	-0.849	-0.821	-0.563	-0.617	1.000	-0.832
RAN	0.697	0.783	0.468	0.728	-0.832	1.000

The table presents summary statistics for various atmospheric parameters collected from the grid area in Bangladesh defined by the coordinates $90^\circ E$ to $92^\circ E$ and $23^\circ N$ to $25^\circ N$ (Table 2). The parameters include temperature (TEM), dew point temperature (DPT), wind speed (WIS), humidity (HUM), sea level pressure (SLP), and rainfall (RAN). The statistics cover measures such as mean, standard error, median, mode, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum, and count for each parameter. For instance, the mean temperature is 25.19Â°C with a standard deviation of $3.77^\circ C$, while the dew point temperature averages $19.89^\circ C$ with a standard deviation of $5.03^\circ C$. The wind speed has a mean of 2.91 Knots and a maximum recorded speed of 6.66 Knots. Humidity averages 78.94% with a range of 37.03%, and sea level pressure averages 1008.14 hPa (Table 2). Each parameter has a consistent count of 672 data points, indicating a robust dataset for analysis. The rainfall is categorized as (i) No rain (dry day) and Trace (NRT), (ii) Light rain (1-10 mm) (LTR), (iii) moderate rain (11-22 mm), moderately heavy rain (23-43 mm), heavy rain (44-88 mm), and very heavy rain with greater than 88 mm (MHR) in this study and the percentage of these three categories are presented in Figure 4. The different rainfall categories in percentage are 31 percent, 41 percent, and 28 percent for no rain and trace (NRT), light rain, moderate rain (LTR), and moderately heavy rain and more (MHR), respectively (Figure 4). The table displays the correlation coefficients among various atmospheric parameters, including temperature (TEM), dew point temperature (DPT), wind speed (WIS), humidity (HUM), sea level pressure (SLP), and rainfall (RAN) for the specified grid area in Bangladesh.

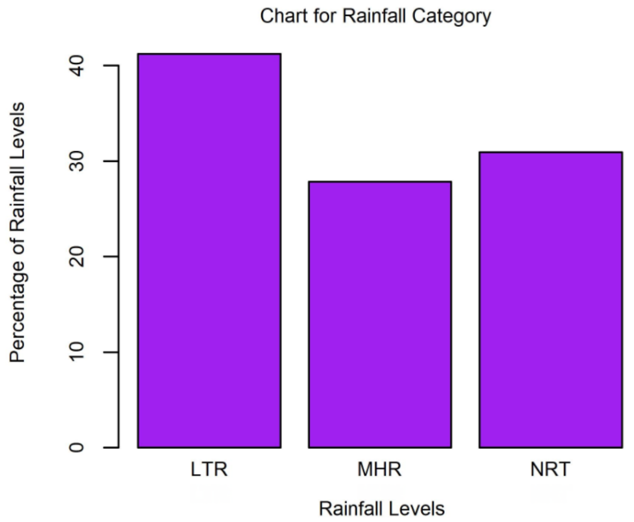

***Figure 4.** Schematic plot for the portion of the different rainfall categories of the Grid $(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ in Bangladesh considered in this study.*

Each parameter is correlated with the others, providing insights into their interrelationships. For instance, there is a strong positive correlation between temperature and dew point temperature (0.895), indicating that as the temperature rises, the dew point also tends to increase (Table 3). Conversely, there is a significant negative correlation between sea level pressure and both temperature (-0.849) and dew point temperature (-0.821), suggesting that higher temperatures are associated with lower sea level pressures. Wind speed shows moderate positive correlations with temperature (0.502) and rainfall (0.468), while humidity has a strong positive correlation with dew point temperature (0.728) and rainfall (0.728). The correlation matrix thus highlights the complex interactions between these atmospheric parameters within the studied region (Table 3).

4. Results and Discussions

Rainfall prediction plays a vital role in climate modeling and resource management, especially in areas susceptible to extreme weather. Accurate precipitation forecasts can greatly impact agricultural strategies, water resource allocation, and disaster preparedness. This section evaluates various machine learning models utilized for predicting rainfall. The study employs advanced machine learning techniques, including Classification and Regression Trees (CART), C5.0, Random Forest (RF), and Gradient Boosting Machine (GBM). By assessing the performance of these models through metrics such as accuracy, precision, recall, and F_1-score, the research aims to determine the most effective method for rainfall prediction in the specified regions. Additionally, incorporating relevant atmospheric features such as temperature, humidity, wind speed, and pressure into the models enhances their predictive accuracy. The results of this study are anticipated to improve rainfall forecasting techniques, thereby facilitating better decision-making for water management and agricultural practices in Bangladesh.

4.1. ANN Model

Table 4 provides a detailed confusion matrix for an Artificial Neural Network (ANN) model designed to predict precipitation levels in the specified grid area of Bangladesh. It includes data from both training and test datasets, categorized into three rainfall levels: Low Rain (LTR), Medium Rain (MHR), and No Rain (NRT). The confusion matrix indicates the number of correct and incorrect predictions for each category, showing that the model correctly classified 181 instances of LTR and 115 instances of MHR during training, while the test data revealed 43 correct classifications for LTR and 44 for MHR (Table 4). Additionally, the table presents various performance metrics for the model, including Recall (RCL), Specificity (SPE), Precision (PRN), Negative Predictive value (NPV), Detection Rate (DER), Detection Prevalence (DEP), Balanced Accuracy (BAC) for both training and test datasets. For instance, the training recall for LTR is 0.883, indicating a high sensitivity in predicting low rainfall events, while the test recall is lower at 0.662, suggesting some challenges in generalization. The results are presented for various values of k, ranging from 2 to 19 folds, highlighting metrics such as accuracy, Kappa statistic, F_1 score, recall, specificity, precision, negative predictive value (NPV), detection rate (DER), and balanced accuracy (BAC) (Table 5).

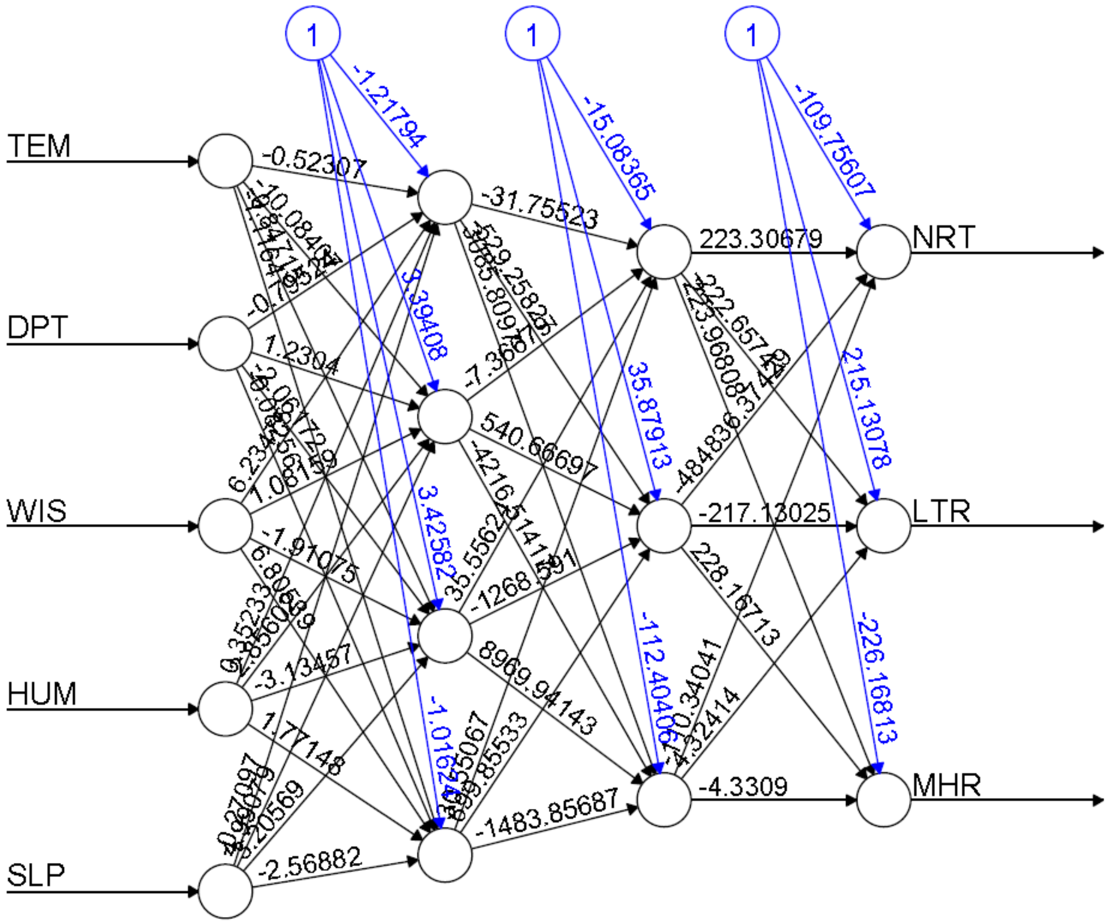

Figure 5. *The ANN Model to predict the rainfall level of the Grid* $(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ *in Bangladesh.*

Overall, the training accuracy is notably high at 88.89%, while the test accuracy is lower at 74.4%, reflecting the model's performance across different datasets (Table 4). The fitted ANN model with estimated parameters to predict the rainfall level of the study grid is presented in Figure 5.

***Table 4.** The Confusion matrix found for the ANN model for training and test data to predict the rainfall category of the Grid $(90°E - 92°E, 23°N - 25°N)$ in Bangladesh considered in this study.*

Confusion Matrix and performance metrics for ANN							
		Training			**Test**		
	Category	**LTR**	**MHR**	**NRT**	**LTR**	**MHR**	**NRT**
	LTR	181	15	17	43	12	9
Predicted	MHR	16	115	0	12	44	0
	NRT	8	0	152	10	0	38
RCL		0.883	0.885	0.899	0.662	0.786	0.809
SPE		0.893	0.957	0.976	0.796	0.893	0.917
PRN		0.850	0.878	0.950	0.672	0.786	0.792
NPV		0.918	0.960	0.951	0.789	0.893	0.925
PRE		0.407	0.258	0.335	0.387	0.333	0.280
DER		0.359	0.228	0.302	0.256	0.262	0.226
DEP		0.423	0.260	0.318	0.381	0.333	0.286
BAC		0.888	0.921	0.938	0.729	0.839	0.863
Overall Statistics		Train Accuracy: 0.8889			Test Accuracy: 0.744		
		Train Kappa : 0.8303			Test Kappa : 0.6131		

The k-fold cross-validation results for an Artificial Neural Network (ANN) model aimed at predicting rainfall categories in the grid area of Bangladesh defined by the coordinates $90°E$ to $92°E$ and $23°N$ to $25°N$ are presented in Table 5.

***Table 5.** The k-fold cross-validation results Accuracy (ACY), Kappa (KPA), F_1 Score, Recall (RCL), Specificity (SPE), Precision (PRN), Negative Predictive value (NPV), Detection Rate (DER), Balanced Accuracy (BAC) of ANN model considering various numbers of the fold to predict the rainfall category of the Grid $(90°E - 92°E, 23°N - 25°N)$ in Bangladesh considered in this study.*

k-Fold cross-validation results for ANN model								
k	**ACY**	**KPA**	F_1	**RCL**	**SPE**	**PRN**	**NPV**	**BAC**
2	0.781	0.669	0.787	0.792	0.888	0.784	0.888	0.840
3	0.798	0.694	0.802	0.809	0.897	0.802	0.897	0.853
5	0.804	0.702	0.806	0.810	0.899	0.808	0.900	0.854
7	0.805	0.704	0.809	0.812	0.899	0.815	0.901	0.856
10	0.802	0.701	0.806	0.813	0.899	0.808	0.901	0.856
11	0.811	0.714	0.816	0.820	0.903	0.817	0.904	0.861
13	0.804	0.702	0.806	0.811	0.899	0.812	0.901	0.855
17	0.805	0.704	0.807	0.813	0.900	0.812	0.901	0.856
19	0.805	0.704	0.809	0.815	0.900	0.810	0.901	0.858

As the number of folds increases, the accuracy of the model shows a general upward trend, peaking at 0.811 with 11 folds. The Kappa statistic, which measures agreement between predicted and observed classifications, also improves with higher k, reaching a maximum of 0.714 at 11 folds. The F_1 score, which balances precision and recall, exhibits a similar pattern, indicating the model's effectiveness in predicting rainfall categories. Notably, recall and specificity values remain high across all folds, suggesting that the model consistently identifies both positive and negative instances of rainfall. Overall, the table demonstrates the robustness of the ANN model across different k-fold configurations, with the best performance metrics observed at 11 folds (Table 5).

4.2. CART Model

Table 6 presents the confusion matrix results for a Classification and Regression Tree (CART) model used to predict precipitation levels in the specified grid area of Bangladesh. It includes both training and test datasets, categorizing the rainfall into three classes: Low Rain (LTR), Medium Rain (MHR), and No Rain (NRT). The confusion matrix indicates the number of correct and incorrect predictions for each category, with the training data showing that the model accurately classified 167 instances of LTR and 116 instances of MHR while incorrectly classifying 32 and 15 instances, respectively. For the test data, the model achieved 41 correct classifications for LTR and 44 for MHR, with some misclassifications noted. Key performance metrics are also provided, such as recall, specificity, precision, negative predictive value (NPV), detection rate (DER), and balanced accuracy (BAC) (Table 6). The recall for LTR in the training set is 0.861, indicating a high sensitivity to low rainfall events, while the test recall drops to 0.631, reflecting some challenges in generalization. Overall, the training accuracy is recorded at 85.52%, while the test accuracy is lower at 72.02%, suggesting that while the model performs well on training data, it faces difficulties in accurately predicting rainfall categories in unseen data.

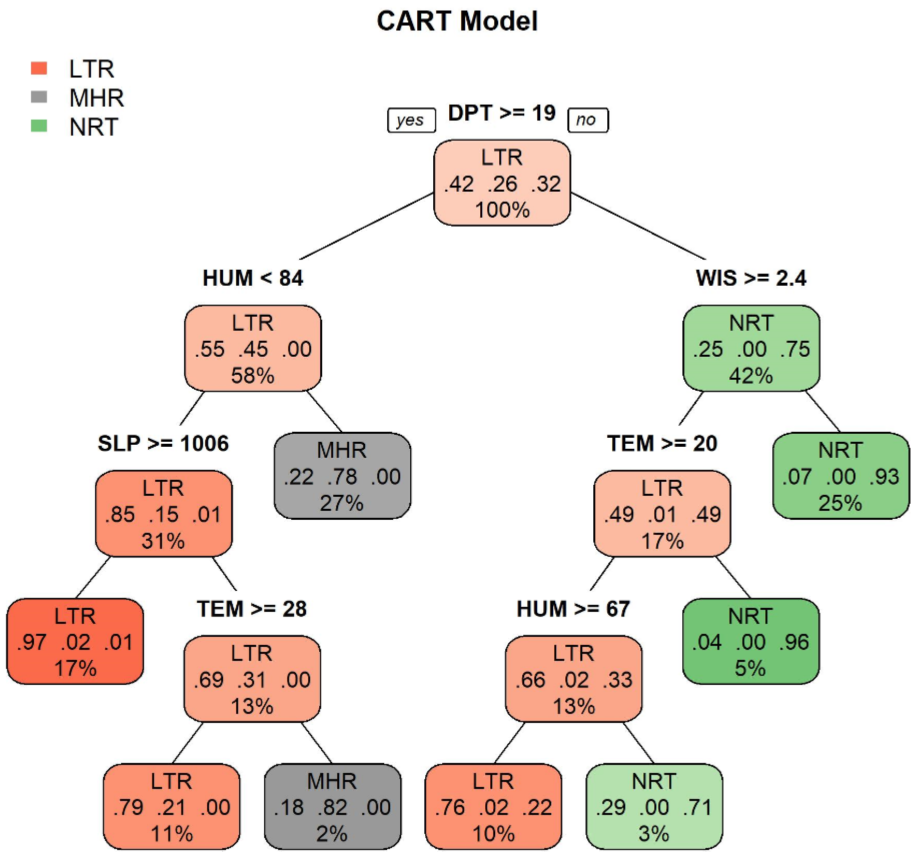

Figure 6. *The CART model to predict the rainfall level of the Grid $(90^{\circ}E - 92^{\circ}E, 23^{\circ}N - 25^{\circ}N)$ in Bangladesh.*

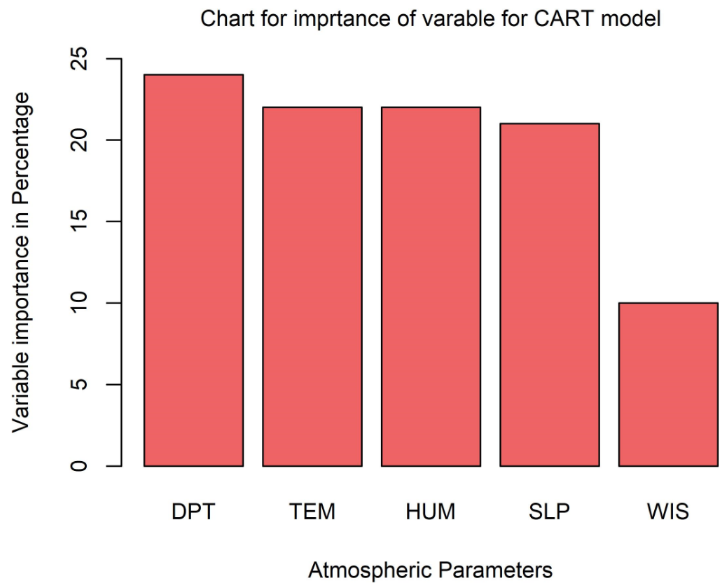

Figure 7. *Schematic Chart for the importance in the percentage of atmospheric variables Dew Point Temperature (DPT), Sea Level Pressure (SLP), Temperature (TEM), Humidity (HUM), and Wind Speed (WIN) to predict the rainfall level for CART model.*

The Kappa statistic, which assesses agreement between predicted and actual classifications, is also reported, with values of 0.7802 for training and 0.5766 for testing, further illustrating the model's performance (Table 6). The CART model obtained for the data set is presented in Figure 6. The significance of atmospheric variables Dew Point Temperature (DPT), Sea Level Pressure (SLP), Temperature (TEM), Humidity (HUM), and Wind Speed (WIS) for predicting rainfall levels using the CART model is illustrated in the figure 7.

Table 6. *The Confusion matrix found for the CART model for training and test data to predict the rainfall category of the Grid*$(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ *in Bangladesh considered in this study.*

Confusion Matrix and performance metrics for CART							
		Training			**Test**		
	Category	**LTR**	**MHR**	**NRT**	**LTR**	**MHR**	**NRT**
	LTR	167	32	14	41	15	8
Predicted	MHR	15	116	0	12	44	0
	NRT	12	0	148	12	0	36
RCL		0.861	0.784	0.914	0.631	0.746	0.818
SPE		0.852	0.958	0.965	0.777	0.890	0.903
PRN		0.784	0.886	0.925	0.641	0.786	0.750
NPV		0.907	0.914	0.959	0.769	0.866	0.933
PRE		0.385	0.294	0.321	0.387	0.351	0.262
DER		0.331	0.230	0.294	0.244	0.262	0.214
DEP		0.423	0.260	0.318	0.381	0.333	0.286
BAC		0.856	0.871	0.939	0.704	0.818	0.861
Overall Statistics		Train Accuracy: 0.8552			Test Accuracy: 0.7202		
		Train Kappa : 0.7802			Test Kappa : 0.5766		

Table 7. *The k-fold cross-validation results-Accuracy, Kappa, F_1 Score, Recall, Specificity, Precision, Negative Predictive value (NPV), Detection Rate (DER), Balanced Accuracy (BAC) of CART model considering various numbers of the fold considering various numbers of the fold to predict the rainfall category of the Grid*$(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ *in Bangladesh considered in this study.*

k-Fold cross-validation results for CART model								
k	**ACY**	**KPA**	F_1	**RCL**	**SPE**	**PRN**	**NPV**	**BAC**
2	0.771	0.652	0.776	0.779	0.882	0.777	0.883	0.831
3	0.798	0.692	0.801	0.802	0.895	0.805	0.896	0.849
5	0.786	0.673	0.790	0.788	0.888	0.794	0.889	0.838
7	0.784	0.672	0.788	0.791	0.889	0.797	0.891	0.840
10	0.779	0.665	0.783	0.787	0.886	0.787	0.887	0.837
11	0.774	0.655	0.779	0.778	0.882	0.785	0.884	0.830
13	0.774	0.657	0.776	0.783	0.884	0.783	0.887	0.833
17	0.778	0.664	0.779	0.786	0.886	0.802	0.893	0.836
19	0.789	0.678	0.792	0.793	0.890	0.805	0.894	0.842

The table provides the k-fold cross-validation results for a Classification and Regression Tree (CART) model used to predict rainfall categories in the grid area of Bangladesh defined by the coordinates $90^\circ E$ to $92^\circ E$ and $23^\circ N$ to $25^\circ N$ (Table 7). The results are organized by varying the number of folds, denoted as k, ranging from 2 to 19. For each k, several performance metrics are presented, including accuracy, Kappa statistic, F_1 score, recall, specificity, precision, negative predictive value (NPV), detection rate (DER), and balanced accuracy (BAC). The accuracy of the model ranges from 0.771 with 2 folds to a peak of 0.798 with 3 folds, indicating that the model's performance improves slightly with more folds (Table 7). The Kappa values, which assess the agreement between predicted and actual classifications, also show a positive trend, reaching a maximum of 0.692 at 3 folds. The F_1 score, which balances precision and recall, is highest at 0.801 for 3 folds, while recall values hover around 0.779 to 0.793 across different k values, indicating consistent sensitivity in detecting rainfall categories. Specificity remains high, particularly for 3 folds (0.895) and 5 folds (0.888), suggesting that the model effectively identifies non-rainfall instances. Overall, the table illustrates the CART model's robust performance across various configurations of k-fold cross-validation, with the best results generally observed at lower fold counts (Table 7).

4.3. C5.0 Model

The table presents the confusion matrix results for a C5.0 decision tree model used to predict rainfall categories in the specified grid area of Bangladesh (Table 8). The model classifies precipitation into three levels: Low Rain (LTR), Medium Rain (MHR), and No Rain (NRT).

The confusion matrix shows the number of correct and incorrect predictions for each category in both the training and test datasets. For the training data, the model accurately classified 180 instances of LTR and 107 instances of MHR while incorrectly classifying 28 and 24 instances, respectively. In the test data, the model achieved 46 correct classifications for LTR and 38 for MHR, with some misclassifications. Various performance metrics are also reported, such as recall, specificity, precision, negative predictive value (NPV), detection rate (DER), and balanced accuracy (BAC) (Table 8). The recall for LTR in the training set is 0.769, indicating a high sensitivity to low rainfall events, while the test recall drops to 0.541, suggesting some challenges in generalizing the model to unseen data (Table 8). The overall training accuracy is 82.74%, while the test accuracy is lower at 66.07%, indicating that the model performs better on the training data compared to the test data. The Kappa statistic, which measures the agreement between predicted and actual classifications, is

0.7341 for training and 0.4791 for testing, further highlighting the model's performance differences between the two datasets (Table 8). The fitted C5.0 model is presented in the figure 8. The table presents the k-fold cross-validation results for a C5.0 decision tree model used to predict rainfall categories in the grid area of Bangladesh defined by the coordinates $90^\circ E$ to $92^\circ E$ and $23^\circ N$ to $25^\circ N$. The results are organized by varying the number of folds, denoted as k, ranging from 2 to 19. For each k, several performance metrics are reported, including accuracy, Kappa statistic, F_1 score, recall, specificity, precision, negative predictive value (NPV), detection rate (DER), and balanced accuracy (BAC) (Table 9).

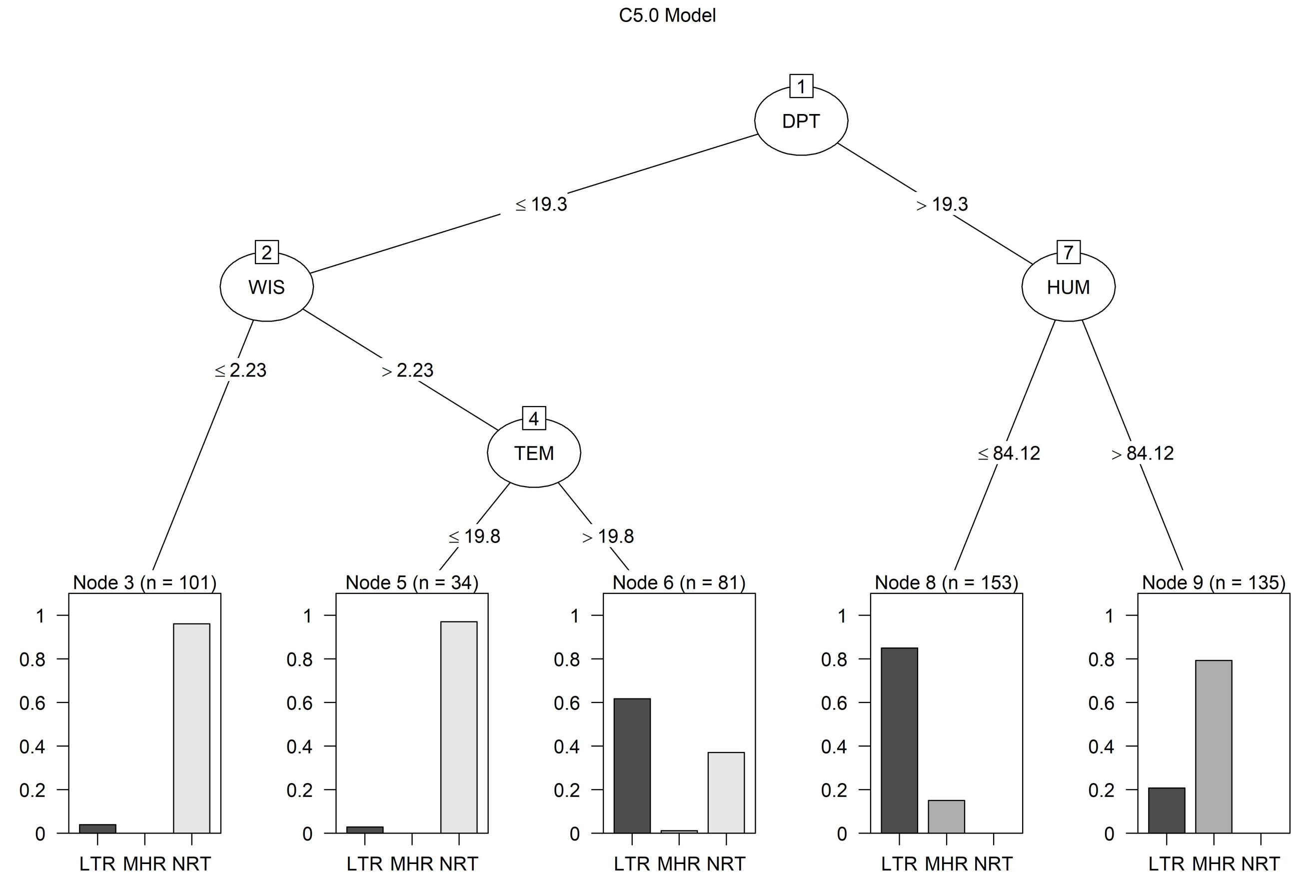

Figure 8. *The C5.0 Model to predict the rainfall level of the Grid $(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ in Bangladesh.*

Table 8. *The Confusion matrix found for the C5.0 model for training and test data to predict the rainfall category of the Grid$(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ in Bangladesh considered in this study.*

Confusion Matrix and performance metrics for C5.0							
		Training			**Test**		
	Category	**LTR**	**MHR**	**NRT**	**LTR**	**MHR**	**NRT**
Predicted	LTR	180	28	5	46	14	4
	MHR	24	107	0	18	38	0
	NRT	30	0	130	21	0	27
RCL		0.769	0.793	0.963	0.541	0.731	0.871
SPE		0.878	0.935	0.919	0.783	0.845	0.847
PRN		0.845	0.817	0.813	0.719	0.679	0.563
NPV		0.814	0.925	0.986	0.625	0.875	0.967
PRE		0.464	0.268	0.268	0.506	0.310	0.185
DER		0.357	0.212	0.258	0.274	0.226	0.161
DEP		0.423	0.260	0.318	0.381	0.333	0.286
BAC		0.824	0.864	0.941	0.662	0.788	0.859
Overall Statistics		Train Accuracy: 0.8274			Test Accuracy: 0.6607		
		Train Kappa : 0.7341			Test Kappa : 0.4791		

Table 9. The k-fold cross-validation results-Accuracy, Kappa, F_1 Score, Recall, Specificity, Precision, Negative Predictive value (NPV), Detection Rate (DER), Balanced Accuracy (BAC) of C5.0 model considering various numbers of the fold to predict the rainfall category of the Grid$(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ in Bangladesh considered in this study.

k-Fold cross-validation results for C5.0 model								
k	ACY	KPA	F_1	RCL	SPE	PRN	NPV	BAC
2	0.799	0.694	0.803	0.804	0.896	0.805	0.897	0.850
3	0.808	0.710	0.810	0.816	0.902	0.812	0.904	0.859
5	0.816	0.719	0.819	0.818	0.904	0.823	0.905	0.861
7	0.804	0.701	0.806	0.808	0.898	0.810	0.900	0.853
10	0.797	0.693	0.800	0.805	0.896	0.804	0.898	0.850
11	0.809	0.711	0.812	0.817	0.902	0.816	0.904	0.860
13	0.803	0.701	0.806	0.808	0.898	0.812	0.901	0.853
17	0.807	0.707	0.810	0.815	0.901	0.818	0.903	0.858
19	0.817	0.720	0.819	0.818	0.904	0.836	0.908	0.861

The accuracy of the model peaks at 0.817 with 19 folds, indicating that the model's performance improves with more folds. The Kappa values, which assess the agreement between predicted and actual classifications, also show a positive trend, reaching a maximum of 0.720 at 19 folds. The F_1 score, which balances precision and recall, is highest at 0.819 for both 5 and 19 folds, while recall values range from 0.804 to 0.818 across different k values, indicating consistent sensitivity in detecting rainfall categories. Specificity remains high, particularly for 5 folds (0.904) and 11 folds (0.902), suggesting that the model effectively identifies non-rainfall instances. Overall, the table illustrates the C5.0 model's robust performance across various configurations of k-fold cross-validation, with the best results generally observed at higher fold counts, particularly 19 folds (Table 9).

4.4. Random Forest

The table presents the confusion matrix results for a Random Forest (RF) model used to predict rainfall categories in the specified grid area of Bangladesh (Table 10). The model classifies precipitation into three levels: Low Rain (LTR), Medium Rain (MHR), and No Rain (NRT).

The confusion matrix shows the number of correct and incorrect predictions for each category in both the training and test datasets. For the training data, the model accurately classified 165 instances of LTR and 108 instances of MHR while incorrectly classifying 28 and 23 instances, respectively. In the test data, the model achieved 46 correct classifications for LTR and 43 for MHR, with some misclassifications noted (Table 10).

Various performance metrics are also reported, such as recall, specificity, precision, negative predictive value (NPV), F_1 score, detection rate (DER), and balanced accuracy (BAC). The recall for LTR in the training set is 0.793, indicating a high sensitivity to low rainfall events, while the test recall drops to 0.667, reflecting some challenges in generalization. The overall training accuracy is 81.94%, while the test accuracy is lower at 75.60%, suggesting that while the model performs well on training data, it faces difficulties in accurately predicting rainfall categories in unseen data. The Kappa statistic, which assesses agreement between predicted and actual classifications, is 0.7242 for training and 0.6302 for testing, further illustrating the model's performance differences between the two datasets (Table 10). The error rate to predict the rainfall levels including out-of-bag (OOB) against the number of trees of Random forest are presented in the figure 9.

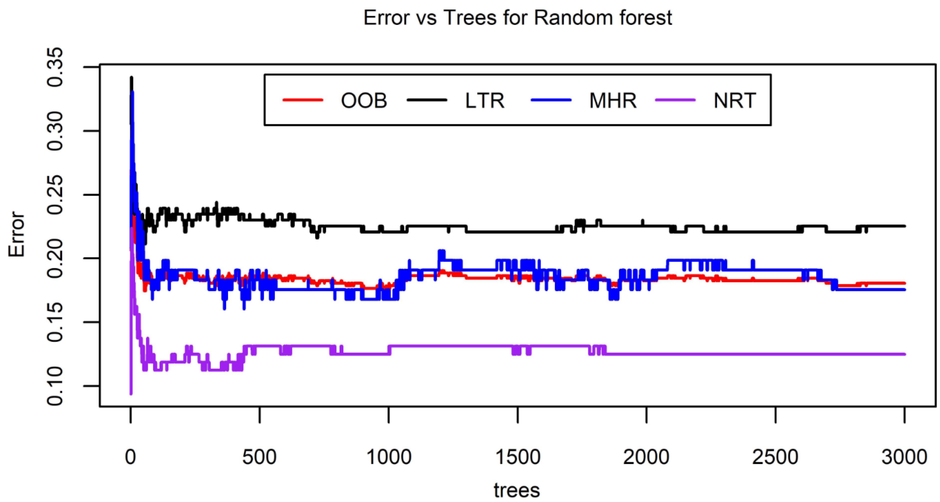

Figure 9. Schematic plot for Error rate to predict the rainfall level and trees of Random Forest.

Table 10. *The Confusion matrix found for the Random Forest model for training and test data to predict the rainfall category of the Grid*($90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N$) *in Bangladesh considered in this study.*

Confusion Matrix and performance metrics for Random Forest							
		Training			**Test**		
	Category	**LTR**	**MHR**	**NRT**	**LTR**	**MHR**	**NRT**
	LTR	165	28	20	46	11	7
Predicted	MHR	23	108	0	13	43	0
	NRT	20	0	140	10	0	38
RCL		0.793	0.794	0.875	0.667	0.796	0.844
SPE		0.838	0.938	0.942	0.818	0.886	0.919
PRN		0.775	0.824	0.875	0.719	0.768	0.792
NPV		0.852	0.925	0.942	0.779	0.902	0.942
F_1		0.784	0.809	0.875	0.692	0.782	0.817
PRE		0.413	0.270	0.317	0.411	0.321	0.268
DER		0.327	0.214	0.278	0.274	0.256	0.226
DEP		0.423	0.260	0.317	0.381	0.333	0.286
BAC		0.816	0.866	0.908	0.742	0.841	0.882
Overall Statistics		Train Accuracy: 0.8194			Test Accuracy: 0.7560		
		Train Kappa : 0.7242			Test Kappa : 0.6302		

Table 11. *The k-fold cross-validation results-Accuracy, Kappa, F_1 Score, Recall, Specificity, Precision, Negative Predictive value (NPV), Detection Rate (DER), Balanced Accuracy (BAC) of Random Forest model considering various numbers of the fold to predict the rainfall category of the Grid*($90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N$) *in Bangladesh considered in this study.*

k-Fold cross-validation results for Random forest								
k	**ACY**	**KAP**	F_1	**RCL**	**SPE**	**PRN**	**NPV**	**BAC**
2	0.793	0.686	0.797	0.799	0.893	0.797	0.894	0.846
3	0.798	0.692	0.802	0.804	0.895	0.804	0.896	0.850
5	0.818	0.724	0.821	0.824	0.906	0.824	0.908	0.865
7	0.820	0.726	0.823	0.824	0.907	0.825	0.907	0.865
10	0.810	0.711	0.813	0.816	0.902	0.817	0.903	0.859
11	0.814	0.717	0.817	0.819	0.904	0.819	0.904	0.862
13	0.809	0.710	0.813	0.814	0.901	0.822	0.903	0.858
17	0.811	0.712	0.814	0.815	0.902	0.817	0.903	0.859
19	0.812	0.714	0.814	0.816	0.903	0.823	0.905	0.859

The table presents the k-fold cross-validation results for a Random Forest (RF) model used to predict rainfall categories in the specified grid area of Bangladesh. The results are organized by varying the number of folds, denoted as k, ranging from 2 to 19. For each k, several performance metrics are reported, including accuracy, Kappa statistic, F_1 score, recall, specificity, precision, negative predictive value (NPV), detection rate (DER), and balanced accuracy (BAC). The accuracy of the model peaks at 0.820 with 7 folds, indicating that the model's performance improves with more folds (Table 11). The Kappa values, which assess the agreement between predicted and actual classifications, also show a positive trend, reaching a maximum of 0.726 at 7 folds. The F_1 score, which balances precision and recall, is highest at 0.823 for 7 folds, while recall values range from 0.799 to 0.824 across different k values, indicating consistent sensitivity in detecting rainfall categories (Table 11). Specificity remains high, particularly for 5 folds (0.906) and 7 folds (0.907), suggesting that the model effectively identifies non-rainfall instances. Overall, the table illustrates the Random Forest model's robust performance across various configurations of k-fold cross-validation, with the best results generally observed at higher fold counts, particularly 7 folds (Table 11).

4.5. Gradient Boosting Machine

The table presents the confusion matrix results for a Gradient Boosting Machines (GBM) model used to predict rainfall categories in the specified grid area of Bangladesh. The model classifies precipitation into three levels: Low Rain (LTR), Medium Rain (MHR), and No Rain (NRT). The confusion matrix shows the number of correct and incorrect predictions for each category in both the training and test datasets (Table 12).

For the training data, the model accurately classified 166 instances of LTR and 104 instances of MHR while incorrectly classifying 26 and 27 instances, respectively. In the test data, the model achieved 42 correct classifications for LTR and 39 for MHR, with some misclassifications noted (Table 12).

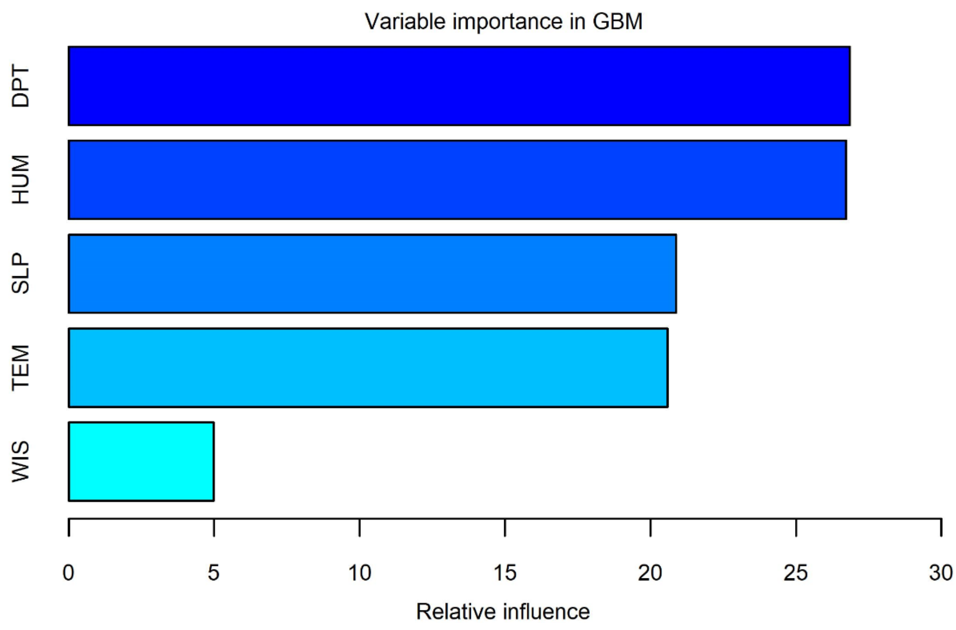

Figure 10. *Schematic plot for the importance of variables in Gradient Boosting Machine.*

Table 12. *The Confusion matrix found for Gradient Boosting Machines for training and test data to predict the rainfall category of the Grid* $(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ *in Bangladesh considered in this study.*

Confusion Matrix and performance metrics for GBM							
		Training			**Test**		
	Category	**LTR**	**MHR**	**NRT**	**LTR**	**MHR**	**NRT**
	LTR	166	26	21	42	11	11
Predicted	MHR	27	104	0	17	39	0
	NRT	12	0	148	11	0	37
RCL		0.810	0.800	0.876	0.600	0.780	0.771
SPE		0.843	0.928	0.964	0.776	0.856	0.908
PRN		0.779	0.794	0.925	0.656	0.696	0.771
NPV		0.866	0.930	0.939	0.731	0.902	0.908
F_1		0.794	0.797	0.900	0.627	0.736	0.771
PRE		0.407	0.258	0.335	0.417	0.298	0.286
DER		0.329	0.206	0.294	0.250	0.232	0.220
DEP		0.423	0.260	0.317	0.381	0.333	0.286
BAC		0.826	0.864	0.920	0.688	0.818	0.840
Overall Statistics		Train Accuracy: 0.8294			Test Accuracy: 0.7024		
		Train Kappa : 0.7393			Test Kappa : 0.5494		

Table 13. *The k-fold cross-validation results-Accuracy, Kappa,* F_1 *Score, Recall, Specificity, Precision, Negative Predictive value (NPV), Detection Rate (DER), Balanced Accuracy (BAC) of Gradient Boosting Machines (GBM) considering various numbers of the fold to predict the rainfall category of the Grid*$(90^\circ E - 92^\circ E, 23^\circ N - 25^\circ N)$ *in Bangladesh considered in this study.*

k-Fold cross-validation results for Gradient Boosting Machines (GBM)								
k	**ACY**	**KPA**	F_1	**RCL**	**SPE**	**PRN**	**NPV**	**BAC**
2	0.784	0.671	0.788	0.787	0.888	0.789	0.889	0.837
3	0.801	0.698	0.805	0.808	0.897	0.803	0.897	0.853
5	0.799	0.695	0.802	0.806	0.897	0.802	0.897	0.851
7	0.807	0.706	0.810	0.810	0.900	0.812	0.901	0.855
10	0.796	0.691	0.800	0.804	0.895	0.801	0.896	0.849
11	0.810	0.710	0.812	0.814	0.901	0.822	0.904	0.857
13	0.802	0.700	0.804	0.810	0.898	0.812	0.901	0.854
17	0.795	0.689	0.797	0.803	0.895	0.806	0.898	0.849
19	0.807	0.706	0.810	0.813	0.900	0.815	0.903	0.857

Table 14. *The table for the values of accuracy, and Cohen's Kappa with rank produced from different machine learning models to predict the rainfall category of the Grid($90°E - 92°E, 23°N - 25°N$) in Bangladesh considered in this study with the overall rank to determine the best model compared to other models.*

Model	Accuracy (Rank)		Kappa (Rank)		Sum of rank
	Training	Test	Training	Test	(Rank)
ANN	88.89 (1)	74.40 (2)	83.03 (1)	61.31 (2)	6 (1)
CART	85.52 (2)	72.02 (3)	78.02 (2)	57.66 (3)	10 (2)
C5.0	82.74 (4)	66.07 (5)	73.41 (4)	47.91 (5)	18 (5)
RF	81.94 (5)	75.60 (1)	72.42 (5)	63.02 (1)	12 (3)
GBM	82.94 (3)	70.24 (4)	73.93 (3)	54.94 (4)	14 (4)

Various performance metrics are also reported, such as recall, specificity, precision, negative predictive value (NPV), F_1 score, detection rate (DER), and balanced accuracy (BAC). The recall for LTR in the training set is 0.810, indicating a high sensitivity to low rainfall events, while the test recall drops to 0.600, suggesting some challenges in generalizing the model to unseen data. The overall training accuracy is 82.94%, while the test accuracy is lower at 70.24%, indicating that the model performs better on the training data compared to the test data. The Kappa statistic, which measures the agreement between predicted and actual classifications, is 0.7393 for training and 0.5494 for testing, further highlighting the model's performance differences between the two datasets (Table 12). The highest importance is found for the dew point temperature to predict the rainfall level using the gradient boosting machine (Figure 10).

The table presents the k-fold cross-validation results for a Gradient Boosting Machines (GBM) model used to predict rainfall categories in the grid area of Bangladesh defined by the coordinates $90°E$ to $92°E$ and $23°N$ to $25°N$ (Table 13). The results are organized by varying the number of folds, denoted as k, ranging from 2 to 19. For each k, several performance metrics are reported, including accuracy, Kappa statistic, F_1 score, recall, specificity, precision, negative predictive value (NPV), detection rate (DER), and balanced accuracy (BAC). The accuracy of the model peaks at 0.810 with 11 folds, indicating that the model's performance improves with more folds (Table 13). The Kappa values, which assess the agreement between predicted and actual classifications, also show a positive trend, reaching a maximum of 0.710 at 11 folds. The F_1 score, which balances precision and recall, is highest at 0.812 for 11 folds, while recall values range from 0.787 to 0.814 across different k values, indicating consistent sensitivity in detecting rainfall categories. Specificity remains high, particularly for 7 folds (0.900) and 11 folds (0.901), suggesting that the model effectively identifies non-rainfall instances. Overall, the table illustrates the Gradient Boosting Machines model's robust performance across various configurations of k-fold cross-validation, with the best results generally observed at higher fold counts, particularly 11 folds (Table 13).

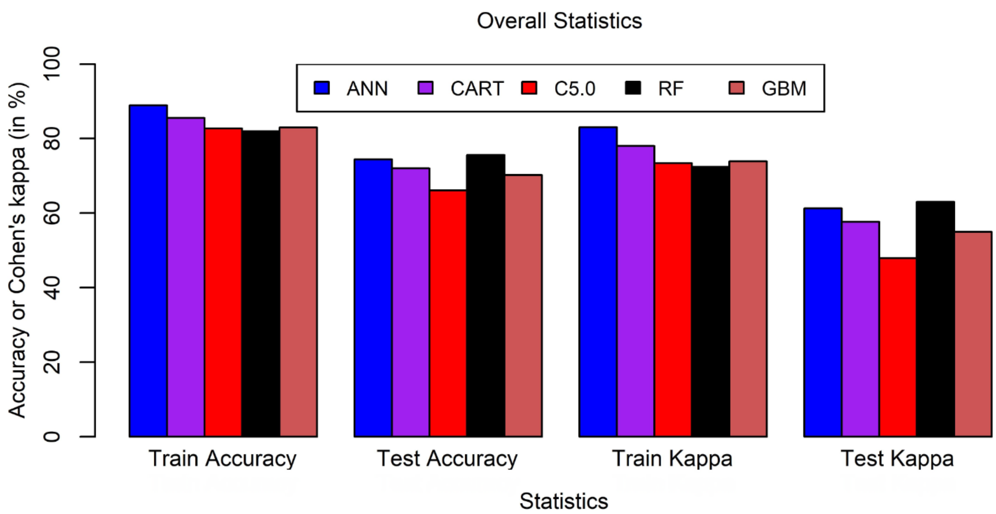

Figure 11. *Schematic plot for the accuracy and Cohen Kappa of different five machine learning models to predict the rainfall levels.*

The table presents a comparison of various machine learning models in terms of their accuracy and Cohen's Kappa statistic for predicting rainfall categories in the grid area of Bangladesh defined by the coordinates $90°E$ to $92°E$ and $23°N$ to $25°N$ (Table 14). The models included are Artificial Neural Network (ANN), Classification and Regression Tree (CART), C5.0 decision tree, Random Forest (RF), and Gradient Boosting Machines (GBM) (Table 14). For each model, the training and test accuracy, as well as the training and test Kappa values, are provided along with their respective ranks in parentheses. The model with the highest value for a particular metric is ranked 1, followed by the next highest, and so on. The ANN model achieves the highest training accuracy of 88.89% and training Kappa of 0.8303, ranking 1st in both categories. However, its test accuracy of 74.40% and test Kappa of 0.6131 rank 2nd. The CART model ranks 2nd in training accuracy and Kappa and 3rd in test accuracy and Kappa. The C5.0 model has the lowest test accuracy of 66.07% and test Kappa of 0.4791, ranking 5th in both categories. The RF model excels in test accuracy and Kappa, ranking 1st with 75.60% and 0.6302, respectively, but lags in training metrics. The GBM model falls in the middle, ranking 3rd in training accuracy and Kappa and 4th in test accuracy and Kappa (Table 14). The sum of ranks for each model is calculated to determine the overall best performer. The ANN model has the lowest sum of 6, making it the best overall, followed by CART (10), RF (12), GBM (14), and C5.0 (18) in descending order (Table 14 and Figure 11).

5. Conclusion

This study explores the efficacy of machine learning techniques, specifically artificial neural networks (ANN) and classification and regression trees, as alternatives to traditional statistical models for predicting atmospheric phenomena. The findings indicate that conventional models often rely on tacit assumptions that may not apply to chaotic systems like the atmosphere. Among the evaluated machine learning models, the ANN emerged as the most effective for forecasting rainfall levels in Bangladesh, demonstrating superior predictive capabilities. This study compares the performance of various machine learning models Artificial Neural Network (ANN), Classification and Regression Tree (CART), C5.0 decision tree, Random Forest (RF), and Gradient Boosting Machines (GBM) in predicting rainfall categories across a specified region in Bangladesh. The analysis includes metrics such as training and test accuracy, Cohen's Kappa statistic, and overall rankings based on these metrics. Results indicate that the ANN model achieves the highest training accuracy and Kappa, while RF excels in test metrics, with ANN emerging as the best overall performer based on the sum of ranks.

Acknowledgments

The author expresses gratitude to Jahangirnagar University for its support in the research's development. Special acknowledgment is given to the Bangladesh Meteorological Department for providing essential data for this study. The author also appreciates the constructive feedback from the Editor and anonymous reviewers, which significantly improved the paper's quality. Additionally, the author thanks Jahangirnagar University for its financial backing, which has been crucial in advancing the research and development efforts of this study. Such support plays a vital role in enhancing academic work and contributes meaningfully to knowledge advancement in the field.

Conflicts of Interest

The author states that they do not have any conflicts of interest.

References

[1] Ridwan, W. M., Sapitang, M., Aziz, A., Kushiar, K. F., Ahmed, A. N. and El-Shafie, A., 2021. Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. Ain Shams Engineering Journal, 12(2), pp. 1651-1663.

[2] Brown, B. G. and Murphy, A. H., 1988. On the economic value of weather forecasts in wildfire suppression mobilization decisions. Canadian Journal of Forest Research, 18(12), pp. 1641-1649.

[3] Wilks, D. S., 1991. Representing serial correlation of meteorological events and forecasts in dynamic decision-analytic models. Monthly Weather Review, 119(7), pp. 1640-1662.

[4] Elsner, J. B. and Tsonis, A. A., 1992. Nonlinear prediction, chaos, and noise. Bulletin of the American Meteorological Society, 73(1), pp. 49-60.

[5] Cartalis, C. and Varotsos, C., 1994. Surface ozone in Athens, Greece, at the beginning and at the end of the twentieth century. Atmospheric Environment, 28(1), pp. 3-8.

[6] Jacovides, C. P., Varotsos, C., Kaltsounides, N. A., Petrakis, M. and Lalas, D. P., 1994. Atmospheric turbidity parameters in the highly polluted site of Athens basin. Renewable Energy, 4(5), pp. 465-470.

[7] Kondratyev, K. Y. and Varotsos, C. A., 2001. Global tropospheric ozone dynamics. Environmental Science and Pollution Research, 8(2), p. 113.

[8] Varotsos, C., Kondratyev, K. Y. and Efstathiou, M., 2001. On the seasonal variation of the surface ozone in Athens, Greece. Atmospheric Environment, 35(2), pp. 315-320.

[9] Sivakumar, B., Liong, S. Y., Liaw, C. Y. and Phoon, K. K., 1999. Singapore rainfall behavior: chaotic? Journal of Hydrologic Engineering, 4(1), pp. 38-48.

[10] Sivakumar, B., 2000. Chaos theory in hydrology: important issues and interpretations. Journal of Hydrology, 227(1-4), pp. 1-20.

[11] Men, B., Xiejing, Z. and Liang, C., 2004. Chaotic analysis on monthly precipitation on Hills Region in Middle Sichuan of China. Nature and Science, 2(2), pp. 45-51.

[12] Varotsos, C., 2005. Power-law correlations in column ozone over Antarctica. International Journal of Remote Sensing, 26(16), pp. 3333-3342.

[13] Saxena, A., Verma, N. and Tripathi, K.C., 2013. A review study of weather forecasting using artificial neural network approach. Int. J. Eng. Res. Technol, 2(11), pp. 2029-2036.

[14] Varotsos, C. and Kirk-Davidoff, D., 2006. Long-memory processes in ozone and temperature variations at the region 60 S-60 N. Atmospheric Chemistry and Physics, 6(12), pp. 4093-4100.

[15] Breiman, L., Friedman, J., Olshen, R., Stone, C. Steinberg, D. and Colla, P. 1983. CART: Classification and regression trees wadsworth: Belmont.

[16] Breiman, L., Friedman, J., Olshen, R. and Stone, C. 1984. Classification and regression trees. monterey, Calif., USA: Wadsworth.

[17] Petre, E. G., 2009. A decision tree for weather prediction. Bul. Univ. Pet.-Gaze din Ploiesti, 61(1), pp. 77-82.

[18] Kalyankar, M. A. and Alaspurkar, S. J., 2013. Data mining technique to analyse the metrological data. International Journal of Advanced Research in Computer Science and Software Engineering, 3(2).

[19] Kumar, R., 2013. Decision tree for the weather forecasting. International Journal of Computer Applications, 76(2), pp. 31-34.

[20] Hu, M. J. C., 1964. Application of the adaline system to weather forecasting (Doctoral dissertation, Department of Electrical Engineering, Stanford University).

[21] Chattopadhyay, S. and Chattopadhyay, M., 2007. A soft computing technique in rainfall forecasting. arXiv preprint nlin/0703042.

[22] Kalogirou, S.A., Neocleous, C. and Michaelides, S., 1997. A time series reconstruction of precipitation records using artificial neural networks. In European Congress on Intelligent Techniques and Soft Computing.

[23] Wong, K. W., Wong, P. M., Gedeon, T. D. and Fung, C. C., 2003. Rainfall prediction model using soft computing technique. Soft Computing, 7, pp. 434-438.

[24] Michaelides, S. C., Neocleous, C. C. and Schizas, C. N., 1995. Artificial neural networks and multiple linear regression in estimating missing rainfall data. In Proceedings of the DSP95 International Conference on Digital Signal Processing, Limassol, Cyprus (pp. 668-673).

[25] Badr, H. S., Zaitchik, B. F. and Guikema, S. D., 2014. Application of statistical models to the prediction of seasonal rainfall anomalies over the Sahel. Journal of Applied Meteorology and Climatology, 53(3), pp. 614-636.

[26] Nelder, J. A. and Wedderburn, R. W., 1972. Generalized linear models. Journal of the Royal Statistical Society Series A: Statistics in Society, 135(3), pp. 370-384.

[27] Cameron, A. C. and Trivedi, P. K., 2013. Regression analysis of count data (No. 53). Cambridge University Press.

[28] Hastie, T. J., 2017. Generalized additive models. In Statistical Models in S (pp. 249-307). Routledge.

[29] Chipman, H. A., George, E. I. and McCulloch, R.E., 2010. BART: Bayesian additive regression trees.

[30] Sutton, C. D., 2005. Classification and regression trees, bagging, and boosting. Handbook of statistics, 24, pp. 303-329.

[31] Breiman, L., 2001. Random forests. Machine learning, 45, pp. 5-32.

[32] Han, J., Pei, J. and Tong, H., 2022. Data mining: concepts and techniques. Morgan Kaufmann.

[33] Breiman, L., 2017. Classification and regression trees. Routledge.

[34] Quinlan, J. R., 2004. Data mining tools See5 and C5.0. http://www.rulequest.com/see5-info.html

[35] Hastie, T., Tibshirani, R. and Friedman, J. 2009. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media,

[36] Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics, pp. 1189-1232.

[37] Gutierrez, D. D., 2015. Machine learning and data science: an introduction to statistical learning methods with R. Technics Publications.

[38] Brownlee, J., 2016. Machine learning mastery with R: Get started, build accurate models and work through projects step-by-step. Machine Learning Mastery.

[39] Landis, J. R. and Koch, G. G., 1977. The measurement of observer agreement for categorical data. biometrics, pp. 159-174.