

Article

Flood Forecasting by Using Machine Learning: A Study Leveraging Historic Climatic Records of Bangladesh

Adel Rajab ¹, Hira Farman ^{2,3}, Noman Islam ^{2,3}, Darakhshan Syed ⁴, M. A. Elmagzoub ⁵, Asadullah Shaikh ⁶, Muhammad Akram ¹ and Mesfer Alrizq ^{6,*}

- ¹ Department of Computer Science, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia; 77adel.rajab@gmail.com (A.R.); maakram@nu.edu.sa (M.A.)
- ² Computer Science Department, Iqra University, Karachi 75300, Pakistan; hira.farman@iqra.edu.pk (H.F.); noman.islam@kiet.edu.pk (N.I.)
- ³ Department Computer Science, Karachi Institute of Economics and Technology, Karachi 74600, Pakistan
- ⁴ Computer Science Department, Bahria University Karachi Campus, Karachi 75300, Pakistan; darakhshansyed.bukc@bahria.edu.pk
- ⁵ Department of Network and Communication Engineering, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia; meabdullah@nu.edu.sa
- ⁶ Department of Information Systems, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia; asshaikh@nu.edu.sa
- * Correspondence: msalrizq@nu.edu.sa

Abstract: Forecasting rainfall is crucial to the well-being of individuals and is significant everywhere in the world. It contributes to reducing the disastrous effects of floods on agriculture, human life, and socioeconomic systems. This study discusses the challenges of effectively forecasting rainfall and floods and the necessity of combining data with flood channel mathematical modelling to forecast floodwater levels and velocities. This research focuses on leveraging historical meteorological data to find trends using machine learning and deep learning approaches to estimate rainfall. The Bangladesh Meteorological Department provided the data for the study, which also uses eight machine learning algorithms. The performance of the machine learning models is examined using evaluation measures like the R^2 score, root mean squared error and validation loss. According to this research's findings, polynomial regression, random forest regression, and long short-term memory (LSTM) had the highest performance levels. Random forest and polynomial regression have an R^2 value of 0.76, while LSTM has a loss value of 0.09, respectively.

Keywords: forecasting rainfall; RMSE; ANN; regression



Citation: Rajab, A.; Farman, H.; Islam, N.; Syed, D.; Elmagzoub, M.A.; Shaikh, A.; Akram, M.; Alrizq, M. Flood Forecasting by Using Machine Learning: A Study Leveraging Historic Climatic Records of Bangladesh. *Water* **2023**, *15*, 3970. <https://doi.org/10.3390/w15223970>

Academic Editors: Marco Franchini and Gwo-Fong Lin

Received: 23 September 2023
Revised: 10 November 2023
Accepted: 12 November 2023
Published: 15 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Natural calamities like hurricanes, earthquakes, floods, wildfires, and tsunamis are caused by the forces of nature and can happen suddenly. Environmental variables, such as climate change, deforestation, and urbanisation, frequently feed these occurrences and increase their frequency and intensity. Natural catastrophes can have severe effects, leading to extensive destruction and fatalities. One of the most critical weather factors that affects many parts of our everyday lives is rainfall [1,2]. Floods, one of the planet's most common catastrophes, seriously negatively impact the economy and agribusiness. They are regularly observed when there is inadequate drainage and a lot of rain.

Various kinds of rainfall exist, and unique mechanisms and climatic factors distinguish each. A few typical types of precipitation are mentioned in Figure 1. Water Supply [3], Plant Growth and Agriculture, Erosion and Soil Moisture, Flooding [4–10], Water Quality and Pollution and Weather and Climate Patterns [11] are some notable effects of rainfall.

The potential of using machine learning algorithms for flood/rainfall prediction and the importance of the problem cannot be denied. The severity and frequency of flood occurrences are expected to rise due to the dual threats of a fast-warming environment

and increasing urbanisation, endangering more people's lives, ecological systems, and economic systems. For many years, flood management strategies have been built on the foundation of conventional flood forecasting techniques based on hydrological and meteorological models and have produced significant insights. A more sophisticated, adaptive, and data-driven methodology must be investigated, though, due to the complex interaction of components that cause floods and the ever-increasing amount of data that are accessible [12]. The field of artificial intelligence concentrates on creating machines that can process data, learn from it, and make judgements. The use of machine learning is an appealing approach for flood forecasting because it holds the promise of revealing intricate, complicated correlations within huge datasets. Its capacity to incorporate data from numerous sources, including satellite images, river gauge data, and climate models, offers chances to improve floods' precision, predictability, and lead time.

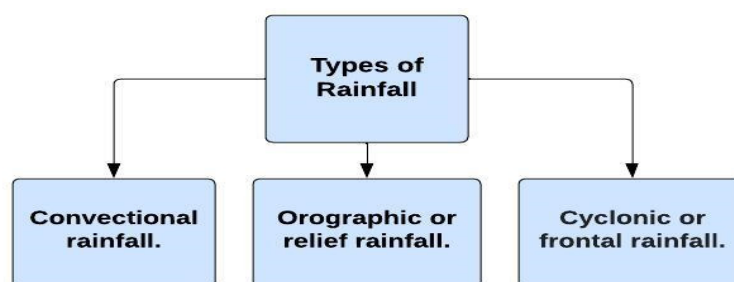


Figure 1. Various types of rainfall according to literature.

In this direction, a number of studies have been presented, such as those of Asif et al. [1], and Luo et al. [13]. Similarly, effective machine learning has been utilised to construct a rain forecast model in numerous research [1,6,14]. Osmani et al. [15] suggest a novel approach for predicting monthly dry days at six target points using several machine learning (ML) techniques. Various other studies, such as those of Manandhar et al. [16], Gude et al. [3], and Nguyen and Chen [4], have reported utilising the concepts of fuzzy logic, support vector and k-nearest neighbours approach. Aswad et al. [10] proposed the use of an Internet-of-Things-driven flood status forecast framework in order to make it easier to forecast when rivers would flood. A similar approach has been presented in Cihan and Elif's work [17].

Based on a critical study of the literature, it has been observed that work on rainfall prediction is at infancy. Bangladesh is a region which is highly affected by rainfall and lots of lives are lost yearly because of floods. The work on flood prediction for Bangladesh has not been extensive. This study performs a comparative analysis of different ML and deep learning (DL) methods for rainfall prediction. Then, we identify the best models for predicting the rainfall in Bangladesh. Finally, this paper offers a thorough investigation of the suggested model, for which a lengthy experiment was used. To summarise, the significant contributions of the discussed work are as follows:

1. Highlight the serious and long-lasting effects that floods have on the socioeconomic system, agriculture, and human life while acknowledging the growing challenge in accurately estimating rainfall because of climatic changes, non-linear qualities, and variable attributes.
2. Suggest a combined approach for data with computationally intensive flood channel mathematical models to predict flooding levels and velocities across a wide area.
3. Identify undetected trends in historical meteorological data to identify machine learning and deep learning approaches as valuable tools for precisely estimating rainfall with quantitative results to prove the usefulness of the machine learning models.
4. Implement evaluation measures to assess the efficiency and progress made by the machine learning models, such as the R^2 score and root mean squared error.

5. Observation that the random forest regressor and k-nearest neighbours algorithms achieved high accuracy, 96% and 99%, respectively.

The remaining sections of this work are organised as follows: Section 2 presents a detailed literature review regarding the materials and methods, and partly examines the theoretical background on flood early prediction utilising deep learning and various machine learning techniques. Section 3 discusses the proposed methodology. Sections 4–6 present the results and discussion for comparing various machine learning and deep learning techniques for flood earlier forecasting based on a dataset. Section 7 provides a conclusion and future recommendations.

2. Literature Review

We structured this section as follows. First, the past studies are highlighted. A discussion on the justification of this research follows this.

2.1. Related Work

Scholars have employed both qualitative and quantitative techniques to determine flood exposure. One of the qualitative approaches utilised for identifying vulnerable areas to flooding is the Analytic Hierarchy Process. Nevertheless, statistics-based and machine-learning-related quantitative techniques can be divided into two groups. Over the years, numerous mathematical and probabilistic algorithms have been employed for predicting floods [18,19]. Various researchers [20,21] propose AI-based metaheuristics algorithms to solve problems that have high complexity like security, load balancing, resource optimisation and forecasting. The capability of ML-based approaches to manage huge volumes of data has boosted interest in these algorithms for forecasting flooding over the past few years. ML makes it possible to learn from historical data [22–25]. It also creates models for forecasts based on historical data. The ability to predict floods will be assisted by this method. We must first tell the system how to generate the output and outcomes. But now, with the assistance of machine learning, it creates models and provides results on its own. The majority of flood-related machine learning research either forecasts future floods or aids in developing safety measures. Floods can be devastating in some years, though, when there has been a lot of rain and water flowing upstream [26]. Kerala, an Indian state in the south, saw a once-in-a-century flood. The cost of damage to both life and property was considerable. This inspired us to conduct research on the rainfall pattern in Kerala. Bangladesh has flooded every year, taking lives, livelihoods, crops, and property.

Flooding happens when a lake, river, or water overflows and engulfs neighbouring land. Each year, floods affect over 4.84 million people in India, 3.84 million in Bangladesh, and 3.28 million in China [27]. India is now one of the nations that has suffered the worst floods, with the most recent calamity in Kerala in August 2018 being an exceptional instance [11,28,29]. Over the years, much effort has been made to forecast the possibility of flooding based on precipitation, humidity, temperature, water velocity, and other characteristics using Internet of Things (IoT) and ML approaches. Nothing has attempted to predict the likelihood of a flood depending on the temperature and severity of the rainfall, which is the research's main flaw. In contradiction to a model developed using machine learning, the results show that a deep neural network may be employed effectively for forecasting floods with the maximum accuracy based simply on monsoon characteristics before flood occurrences. According to Adnan et al. [30], flood prediction has been a significant area of study for scholars worldwide. This is because efficient and real-time prediction of floods is essential for giving individuals who live close to flood zones the warning they need to flee. Consequently, in this study, a 5 h flood forecast model for Kuala Lumpur's rainfall area was provided using an enhanced Neural Network Autoregressive Model methodology. The 5 h NNARX flood water level forecast framework was created using MATLAB Neural Network Toolkit. The findings showed that the NNARX model effectively accurately estimated the flood water level five hours early. Mosavi et al. [31] combine new ML techniques with traditional methods to create more precise and effective

prediction models. This publication's essential contribution is its discussion of the state of ML models for flood prediction and recommendations for the most effective models to use. To give a comprehensive picture of the many ML methodologies utilised in the area, this study predominantly looks at the studies where ML models were assessed through a qualitative review of reliability, accuracy, effectiveness, and performance. According to Chen et al. [32], the area significance is divided into grids based on longitude and latitude, and the data on precipitation and drainage collected at stations are combined into tensors depending on station coordinates. Instead of a one-dimensional time sequence, the input characteristic is a two-dimensional time series with spatial data.

According to Motta et al. [33], this effort will integrate machine learning classifiers with Geographic Information Systems (GIS) methods to provide a flood prediction platform that can be helpful for resiliency management. With the help of this approach, it is possible to create realistic variables and risk indicators for the likelihood of flooding at the municipality level, which may be used to create long-term plans for smart cities. According to a review of past research articles, Maspo et al. [7] review the ML methods currently being utilised for flood forecasting. This research tries to identify the most helpful flood forecast methods. This research aims to list the critical variables and the most recent ML methods for flood prediction. According to Sankaranarayanan et al. [13], the public and the government may be able to plan both short- and long-term mitigation strategies, be ready for evacuation and rescue operations, and provide relief for flood victims if they receive early warning of a flood calamity. In this case, the location of the impacted areas and their respective seriousness are two of the critical considerations in most flood mitigation methods. There is still no reliable method for predicting floods in advance. Previous technologies usually relied on prepared and manually entered data. Because the processes were time consuming, making early and real-time projections was impossible. Innovative operational approaches have been examined by Parag et al. [34]. The researchers observe and examine the current trend regarding data-driven solutions for flood forecasting. For algorithms based on machine learning that have been developed using historical data for climatic variables, predicting jobs is growing more and more significant. A review offered by Furquim et al. [5] presents the use of data gathered from urban rivers to forecast floods in an effort to decrease the damage caused by floods. After the involvement hypothesis had shown the mutual dependence of the information, the artificial neural networks were reviewed to see how accurate their forecasting algorithms were. WSNs have been set up whenever serious flooding-related issues have arisen. Adnan et al.'s [35] approach was suggested to develop risk-based plans for growth and enhance current warning systems for emergencies. To forecast and determine potential flooding locations or flood-sensitive regions in the Teesta River basin, Talukdar et al. [6] applied cutting-edge novel ensemble machine learning strategies. A rainfall forecasting algorithm employing the effective machine learning random forest was proposed by Adnan et al. [30] by giving a flood prediction utilising a range of machine learning techniques. Gauhar et al. [36] employed a variety of coefficients of association for feature selection and the k-NN technique to forecast a flood. It is widely established that quantifying and lowering the uncertainty associated with hydrologic prediction is crucial for predicting the risk of flooding and making educated decisions [37]. The current paper thoroughly examines the Bayesian forecasting methods used in flood forecasting. According to Haque et al. [38], 180 individual models were produced using five different machine learning techniques based on multiple combinations of temporal lags for input data and lead times in prediction. The 5772 km² Someshwari-Kangsa sub-watershed in Bangladesh's North Central hydrological region was the subject of modelling. Using conventional machine learning approaches, it is challenging to predict when it will rain [39]. However, several studies have been presented that forecast rainfall using various computer algorithms. Osmani et al.'s [15] innovative method for predicting monthly dry days (MDD) at six target stations in Bangladesh makes use of a variety of ML techniques. The datasets for monthly days without precipitation and monthly days with rainfall were produced using a range of rainfall limitations. Manandhar et al. [16] recommend employing

machine learning approaches to look into the long-term implications of flood prevention in Bangladesh. Data from socioeconomic surveys and historical events (such as migration and mortality) are available from 1983 to 2014. Yaseen et al.'s [40] emphasis is on the greater necessity and duty of handling human-caused catastrophes. Humans invented an area of science called artificial intelligence (AI) that could be applied in this situation.

Aakash et al. [41] thoroughly analyse and compare the many approaches and algorithms that scholars have used to estimate precipitation. The core objective is to present non-experts with access to the techniques and approaches used in rainfall forecasting. In the present research, a flood vulnerability map for Iran is produced by utilising the concept of the convolutional neural networks (CNN) method [42], one of the more current and effective techniques in enormous datasets. In their discussion of several case studies, Sergey et al. [43] use the example of ensemble-based storm surge simulation for forecasting floods in St. Petersburg, Russia, to look at the opportunities presented by the established methodology. Mosavi et al.'s [31] main contribution is to show the current state of ML models for flood prediction and to provide insight into the most appropriate models. India [11] has some of the worst flood damage in the world right now, with the most recent disaster in Kerala in August 2018 as a prime example. The problem is that no one has attempted to forecast the likelihood of a flood using rainfall volume and temperature. Therefore, the Neural Network has been utilised to forecast the likelihood of floods based on temperature and rainfall intensity. Therefore, Gude et al. [3] propose flood prediction as one of the primary topics to be researched in hydrology. Although many academics have studied this problem using various approaches, such as physical models and image processing, the accuracy and time steps still fall short for all applications. This study examines deep learning techniques for gauging height and evaluates the associated uncertainty. Current ML techniques for flood prediction are evaluated by Maspo et al. [7], and the parameters utilised to predict floods are based on an analysis of past research publications. According to Nevo et al. [44], the multidimensional model is a machine learning replacement for the hydraulic modelling of flooding flows. Compared to past information, all models meet expectations of performance that are high enough for use in operational situations. Mitra et al. [8] offer an embedded system that utilises IoT-based machine learning to forecast the possibility of floods in a river basin. The device uses a ZigBee connection to link the WSN to a customisable mesh network, and then it uses a GPRS module to send data over the internet.

Jeerana et al. [9] examine the possibility of using machine learning methods to forecast flood occurrences in the Pattani River using open data. A probabilistic prediction framework was created by Chen et al. [32] using several machine learning approaches. Three techniques utilising multiple scenarios for decision-making were used along with the ML based approaches to evaluate how well they were able to model the risk of flooding in the Ningdu Catchment, which Khosravi et al. [45] address as one of China's foremost flood-prone geographic areas. El-Magd et al. [46] employed the extreme gradient boost and KNN methods to produce flash flood prediction maps for the River El-Laquita in Egypt's eastern centre area. To predict floods on the shores of the rivers Daya and Bhargavi, which flow across the Indian state of Odisha, Nayak et al. [47] use the Deep Belief Network (DBN). In a comparative investigation, additional machine learning techniques are applied to fully depict the effects of dam construction. Tayfur et al.'s [48] main concern is to present the application of swarm-based optimisation, ant-colony optimiser, artificial neural network (ANN) and genetic-algorithm-based approaches to flood hydrograph prediction.

To predict river flooding in the Barak River, Sahoo et al. [49] examined the corresponding precision of radial basis functions. As reported by Qian et al. [50], there have been considerable financial and human damages due to an increase in flash floods in metropolitan areas. To precisely describe the specifics of flood improvement, the current flood prediction techniques are either too sluggish or excessively straightforward. This research uses deep neural networks to accelerate the mathematical calculation of a 2D urban flood forecasting technique based on thermodynamics and controlled by the Shal-

low Water Equation (SWE). Researchers retrieve flood patterns from data generated via a partial differential equation (PDE) generator using convolutional neural networks (CNN) and conditionally generative adversarial networks (cGANs). The four ML-based FSMs mentioned by Adnan et al. [35] are random forest (RF), KNN, multilayer perceptron (MLP), and hybridised genetic algorithm–gaussian radial basis function–support vector regression (GA; RBF; SVR). Scott Miao and Wei-Hsi Hung [51] also utilise a deep learning framework.

Hossain et al. [52] describe an effort to create a system for analysing long-term seasonal rainfall trends in Western Australia using Lavenberg–Marquardt multiple linear regression and artificial-based methodologies. RBFs, including linear and nonlinear kernel parameters, perform better in the same catchment under different circumstances. The response to lighter precipitation would be very different from that to heavier one, which is a handy way to reveal the dynamics of an SVM classifier. The study also shows an unexpected outcome in the SVM response to diverse rainstorm-related inputs.

According to Aswad et al. [10], forecasting flood status is challenging and calls for in-depth research into the underlying causes of floods. This study recommends a TpoT-based model to help anticipate when rivers may flood. The IoT-FSP concept uses the Internet of Things framework to facilitate flood data collection and three approaches for ML for flood forecasts. As Ighile et al. [53] reported, this study forecasted the flood-prone locations in Nigeria using historical flood records from 1985 to 2020 and plenty of conditioned variables. An exact flood prediction model can be made using various machine learning techniques, according to Kunvergi et al.'s [54] investigation. The generalised additive model (GAM), the boosted regression tree (BTR), and the multivariate adaptive regression splines (MARS) are three novel machine learning methods that Dodangeh et al. [55] suggested. These models were built using random subsampling (RS), bootstrapping (BT), and multi-time resampling techniques. The province of Ardabil, to which this approach was employed, is situated near the Caspian Sea's coast and frequently endures severe flooding.

The study by Khairudin et al. [56] aims to investigate the impact of various time-series scales of rainfall information from eight rainfall stations along the Kelantan River on the accuracy of the water level forecasts at Kuala Krai station. To create a flood forecasting framework, Dtissibe, Francis Yongwa et al. [57] employed the multiple-layer perceptron and flow as input–output parameters. For this, a set of data corresponding to the estimates of rainfall recorded in Australia's major cities throughout the previous ten years was provided to the primary machine learning methods (kNN, decision tree, random forest, and neural networks). Sarasa-Cabezuelo's study [58] outlines a qualitative investigation of using machine learning to predict the probability of rain. The outcomes demonstrate that neural networks are the best approach. This work compares the efficacy of rainfall forecasting techniques based on modern machine learning techniques for forecasting hourly volumes of rainfall using weather time-series data from cities across the United Kingdom.

Liyew and Melese [59] analyse the performance of these algorithms. The analytical hierarchical technique, a multi-parameter modelling tool, is used by MC Aydin and Birinciolu [17] to analyse the risk of flooding assessments in the Turkish province of Bitlis. Table 1 contrasts the pros and cons of various methods, datasets, and resources used in the literature for rainfall or weather forecasting.

Table 1. A comparison of literature on rainfall prediction using machine learning.

Reference	Dataset	Model	Pros	Cons
[33]	Urban datasets from January 2013 and December 2018	Machine Learning (RF) and GIS	A flood risk score was produced by combining the results of the random forest model and the Hot Spot research.	Hourly dataset; not a consistent forecast for the entire year.

Table 1. Cont.

Reference	Dataset	Model	Pros	Cons
[5]	Rainfall in April 2014 in Brazil	Chaos theory MLP, E-RNN	The outcomes demonstrate that the MLP outperforms the ERNN.	Not a consistent forecast for the entire year.
[4]	Yilan River basin and Taiwan 2012 to 2018	(SVR), fuzzy inference model (FIM), (k-NN)	Statistical parameters are used to analyse time series data.	A less extensive training dataset, a smaller feature set, along with lower precision and recall.
[25]	Dumdum weather station	Hybrid neural framework	Selection of features: hybrid neural system.	Less precision just for a tiny area.
[29]	India's annual rainfall is included in the data collection.	Linear regression	Assist farmers in making the best decision for harvesting a particular crop.	Based on only one characteristic, no experimental results were discovered.
[14]	Used the dataset of 2016 to 2019 of Bangladesh.	DT, KNN, LR, NB, RF	According to the results, random forest can make reliable forecasts for daily rainfall estimates.	Small dataset utilised for experiment.
[57]	Dataset reported of France. 2002 to 2018 events	Multiple linear regression (MLR) and non-linear modelling technique, (ANN)	The created model was tested extensively, and the results demonstrated the usefulness of forecasting.	Event-wise data analysis.
[60]	Data on Indian summer monsoon rainfall	Expert system based on fuzzy entropy	Statistical parameters are used to analyse time series data.	Less precision; covers a much smaller area.
[61]	The Indian Meteorological Institute in Pune collected data on North India's monthly rainfall.	Artificial Neural Network (ANN)	Dataset with a long-time series.	Only 1- and 2-month forward forecast; minimal feature set.

2.2. Discussion on Past Studies

Before proceeding further, let us discuss key findings obtained from the literature, as follows:

- Classical meteorological systems: In the past, physical-based models like the HEC-HMS, SWAT, and MIKE SHE predominated flood forecasting. These representations incorporate physical equations that represent the flow and accumulation of water [2].
- Mathematical Models based on statistics: The exponential smoothing and ARIMA statistical analysis of time series techniques were also employed to anticipate river flows and flood levels. These techniques rely on the data's statistical patterns [62].
- Incorporation of machine learning techniques: Machine learning techniques have become more prevalent in recent years. Investigations have demonstrated that when there are a lot of data available, machine learning can frequently match or even surpass conventional hydrological projections [63].

Focusing on machine learning, LSTM and ANN are now favoured options for flood forecasting.

Besides those discussed above, there have been several other developments and breakthroughs when using machine learning (ML) for flood forecasting. However, there are several gaps and restrictions in prior work [13,33,64,65]:

- Numerous studies rely on small or sparse datasets, which might not fully account for all possible flood situations. Records for extreme occurrences or isolated flooding disasters are frequently lacking. Climatic datasets are often incomplete and inaccurate, particularly in nations with limited resources.

- Several models are inefficient when used differently because they are overfit to datasets or areas. Model portability between several geographic areas is still rugged.
- Numerous ML models, intense learning models, behave as “black boxes”, making it challenging to comprehend how they make decisions. This makes it difficult to win over the trust of stakeholders and end users.
- When ML models are combined with conventional meteorological models, which have been widely used and relied upon for years, there is frequently a gap. The physical processes that play a role in flood generation and propagation are not always adequately taken into consideration in studies.
- For real-time prediction, the computational burden of some sophisticated ML models may be too high. Specific models could be useless in time-sensitive situations due to the time required for data collection, initial processing, and forecasting.
- Certain approaches might be practical in local watersheds or urban areas, but they might have difficulties when scaled up to substantial river valleys.

Regardless of these challenges, there is a lot of scope for machine learning in flood forecasting. To utilise machine learning’s maximum potential, it will be essential to carry out ongoing research, collect data, engage stakeholders, and integrate ML with conventional modelling techniques. This paper employs several machine learning and deep learning models for rainfall prediction. We briefly discuss the rationale behind our choice.

For choosing ANN, the reason is that dynamic non-linear correlations in data, which are frequently present in the hydrological system, can be captured using ANNs. A sufficiently large ANN can hypothetically approximate any function. Because of this, they are adaptable for various jobs, including flood forecasting. ANNs can change their architecture (depth, width) to accommodate various datasets and prediction timeframes [57,66].

Similarly, the choice of LSMT is obvious. Flood forecasting is a time-series challenge by nature. By preserving a “stored memory” of previous inputs in the internal neuron states, LSTMs, a form of recurrent neural network (RNN), are created to handle sequential information. The issue of vanishing gradients affects traditional RNNs, making it difficult for them to learn dependence over time. LSTMs are less prone to this issue because of their latching mechanisms, which enable them to learn and remember over lengthy periods [67]. Additionally, convolutional neural networks (CNNs) and other neural network architectures can be integrated with LSTMs to capture spatial and temporal patterns [68,69].

The proposed study can forecast rainfall for every season and region in Bangladesh. From Table 1, it is clear that most efforts are for various restricted regions and have some significant flaws, like a relatively tiny dataset, a small feature set, and lower precision. On the other hand, in this study, machine learning methods and a two-layer long short-term memory (LSTM) method have been utilised [1,7,21] and an artificial neural network (ANN) [33,58] for predicting rainfall in Bangladesh has been developed. It solves the backflow problem found in other works. It uses a wider dataset with 18 features, the 16 most important of which are used. The proposed study can forecast rainfall in Bangladesh for any season.

3. Proposed Methodology

This proposed research aims to determine whether, by utilising machine and deep learning algorithms, a higher accuracy rate can be attained while also reducing error. The dataset includes information on Bangladesh’s monthly and yearly rainfall (1949 to 2013) index as well as information on the number of times a year that floods occur close to 35 stations: Khulna, Dinajpur, Bogra, Srimangal, Satkhira, Mymensingh, Jessor, Comilla, Cox’sBaza, Faridpur, Barisal, Chittagong (IAP-Patenga), Maijdee, Court, Dhaka, Rangpu, Sylhet, Rangamati, Ishurd, Rajshahi, Chandpur, Hatiya, Bhola, Sandwip, Patuakhali, Feni, Khepupara, Madaripur, Kutubdia, Sitakunda, Teknaf, Tangail, Mongla, Chuadanga, Syedpur and Chittagong (City-Ambagan). Furthermore, the data were preprocessed using feature engineering, data normalisation, and feature encoding. After splitting the dataset into training and testing portions in an 80:20 ratio, applying the machine learning model

is essential. For comparison, it is necessary to use models like the k-nearest neighbour, support vector machine, decision tree regressor, random forest mode, AdaBoostRegressor, Stacking Regressor, and artificial neural network. Finally, the model that is best at predicting floods can be identified based on the RMSE and R^2 scores of the models used. Figure 2 shows the workflow of the methodology in detail.

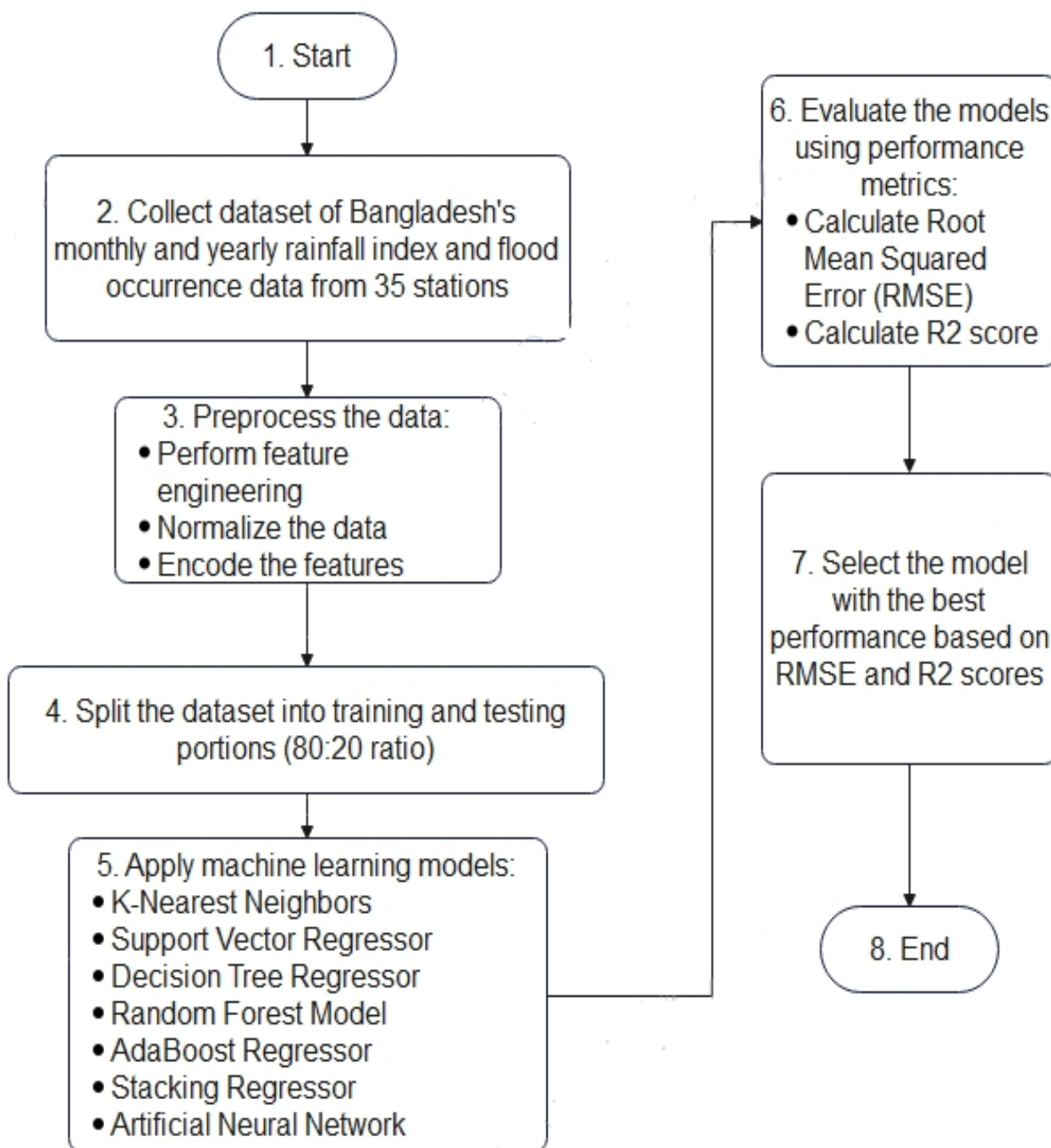


Figure 2. Machine learning pipeline describing the proposed methodology.

Figure 3 represents the complete architecture of the proposed work. The only goal of this research is to use deep learning and supervised learning to achieve maximum accuracy. The points in the validation set are used to determine the accuracy of the regressor following learning with the training data.

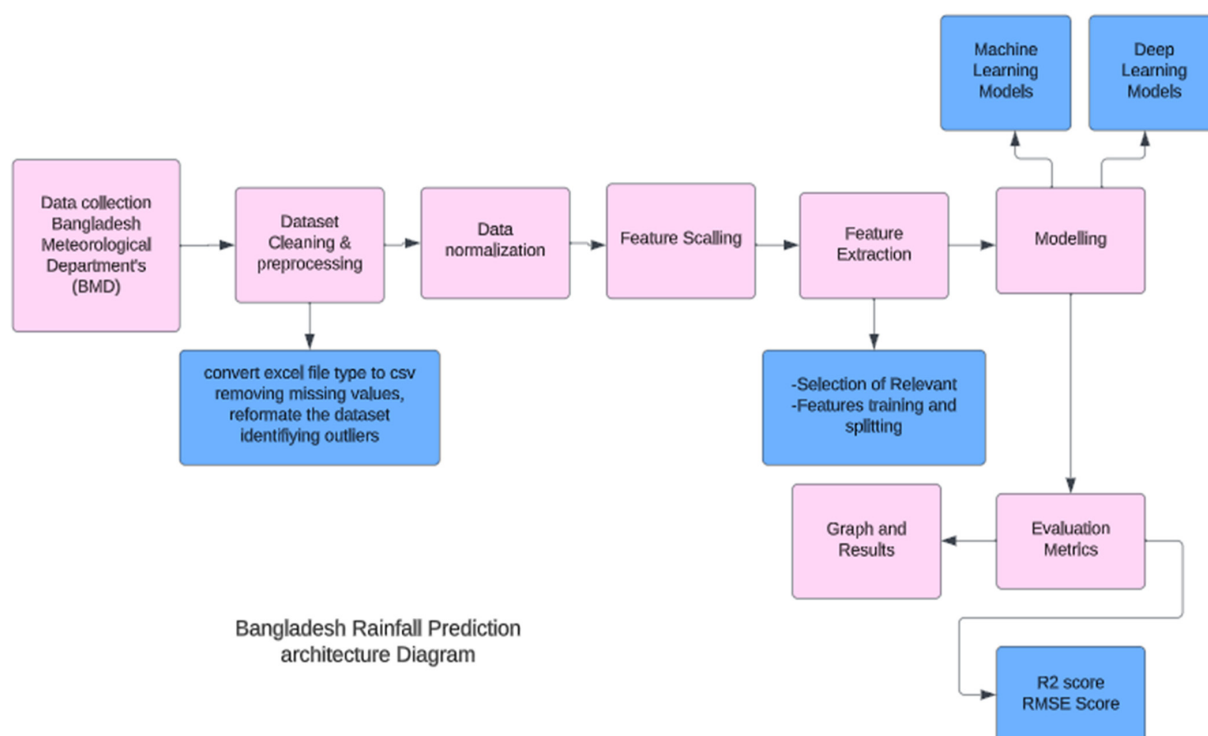


Figure 3. Block diagram describing the proposed system.

3.1. Dataset Description

The data are acquired from Bangladesh's Weather Department in Dhaka, the main authority for tracking and making predictions for every natural catastrophe to reduce mortality. To deliver accurate forecasts for the weather, we want to learn what we can deduce about previous times and how it connects to present-day climate change and the general trends of the planet's weather by using this dataset. From 1948 to 2013, the dataset includes comprehensive monthly averages for Bangladesh that are area-specific for maximum temperature, minimum temperature, rainfall, relative humidity, wind speed, cloud cover, and brilliant sunshine. Also included are the weather station numbers, X and Y coordinates, latitude, longitude, and altitude. This research develops and evaluates the top eight deep learning and machine learning models using our dataset. Figure 4 shows the snapshot of the dataset. The data for Bangladesh's maximum and minimum monthly temperatures and annual rainfall are shown in Figures 5 and 6, respectively, throughout the entire 65-year period. As can be seen, the maximum temperature in Bangladesh is in April, March or May, whereas the lowest temperature occurs in January or December.

	Unnamed: 0	Station Names	YEAR	Month	Max Temp ° C	Min Temp ° C	Rainfall (mm)	Relative Humidity %	Wind Speed (m/s)	Cloud Coverage %	Bright Sunshine (Hours)	Station Number	X_COR	Y_COR	LATITUDE	LONGITUDE	ALT	Period
0	0	Barisal	1949	1	29.4	12.3	0.0	68.0	0.453704	0.6	7.831915	41950	536809.8	510151.9	22.70	90.36	4	1949.01
1	1	Barisal	1950	1	30.0	14.1	0.0	77.0	0.453704	0.8	7.831915	41950	536809.8	510151.9	22.70	90.36	4	1950.01
2	2	Barisal	1951	1	28.2	12.3	0.0	77.0	0.453704	0.6	7.831915	41950	536809.8	510151.9	22.70	90.36	4	1951.01
3	3	Barisal	1952	1	26.6	12.3	2.0	77.0	0.453704	1.0	7.831915	41950	536809.8	510151.9	22.70	90.36	4	1952.01
4	4	Barisal	1953	1	30.0	13.3	10.0	75.0	0.453704	1.6	7.831915	41950	536809.8	510151.9	22.70	90.36	4	1953.01
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
21115	21115	Teknaf	2009	12	30.3	16.5	0.0	72.0	2.800000	0.9	8.700000	41998	734765.4	308914.1	20.87	92.26	4	2009.12
21116	21116	Teknaf	2010	12	31.7	16.7	193.0	79.0	2.400000	1.4	5.500000	41998	734765.4	308914.1	20.87	92.26	4	2010.12

Figure 4. Snapshot of the dataset [70].

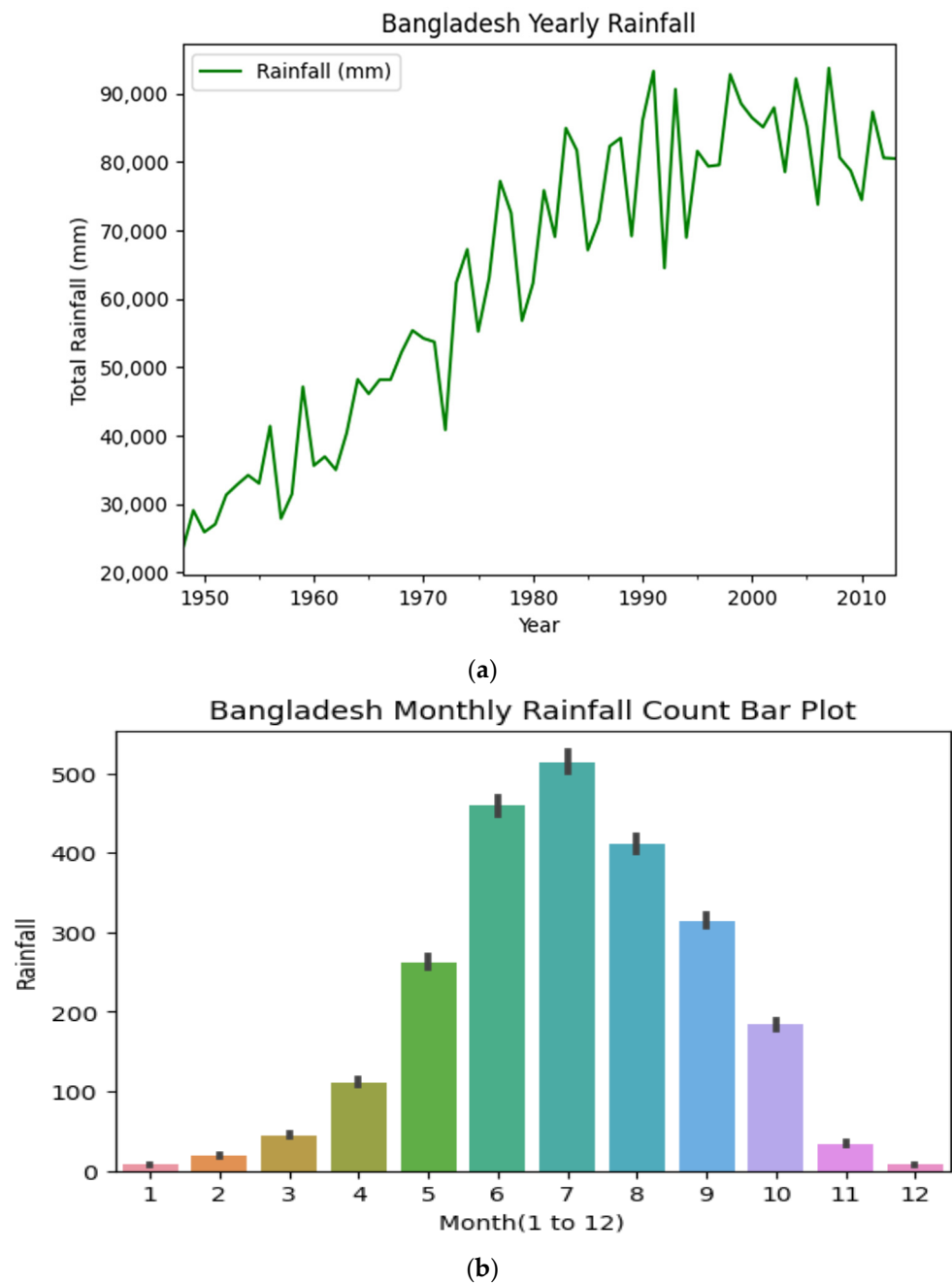


Figure 5. Yearly and monthly rainfall in Bangladesh: (a) yearly rainfall in Bangladesh, (b) monthly rainfall in Bangladesh.

These figures show the bar plot, monthly precipitation, and the time series of weather forecasting. According to the figure, the middle months experience the highest frequency of rainfall during a 12-month period, which steadily declines afterwards. Projecting future values using past data is known as time series forecasting. The graphs indicate that the peak of rainfall progressively rises after a few years. The highest temperature in Bangladesh is depicted in Figure 6. The bar plot shows the variation in the peak temperature over time. The varying patterns of the bar indicate periods when the maximum temperature was abnormally high or low. The right graph displays the variations in the minimum temperature over time. The variation in the count plot indicates the seasonal variations in the minimum temperature.

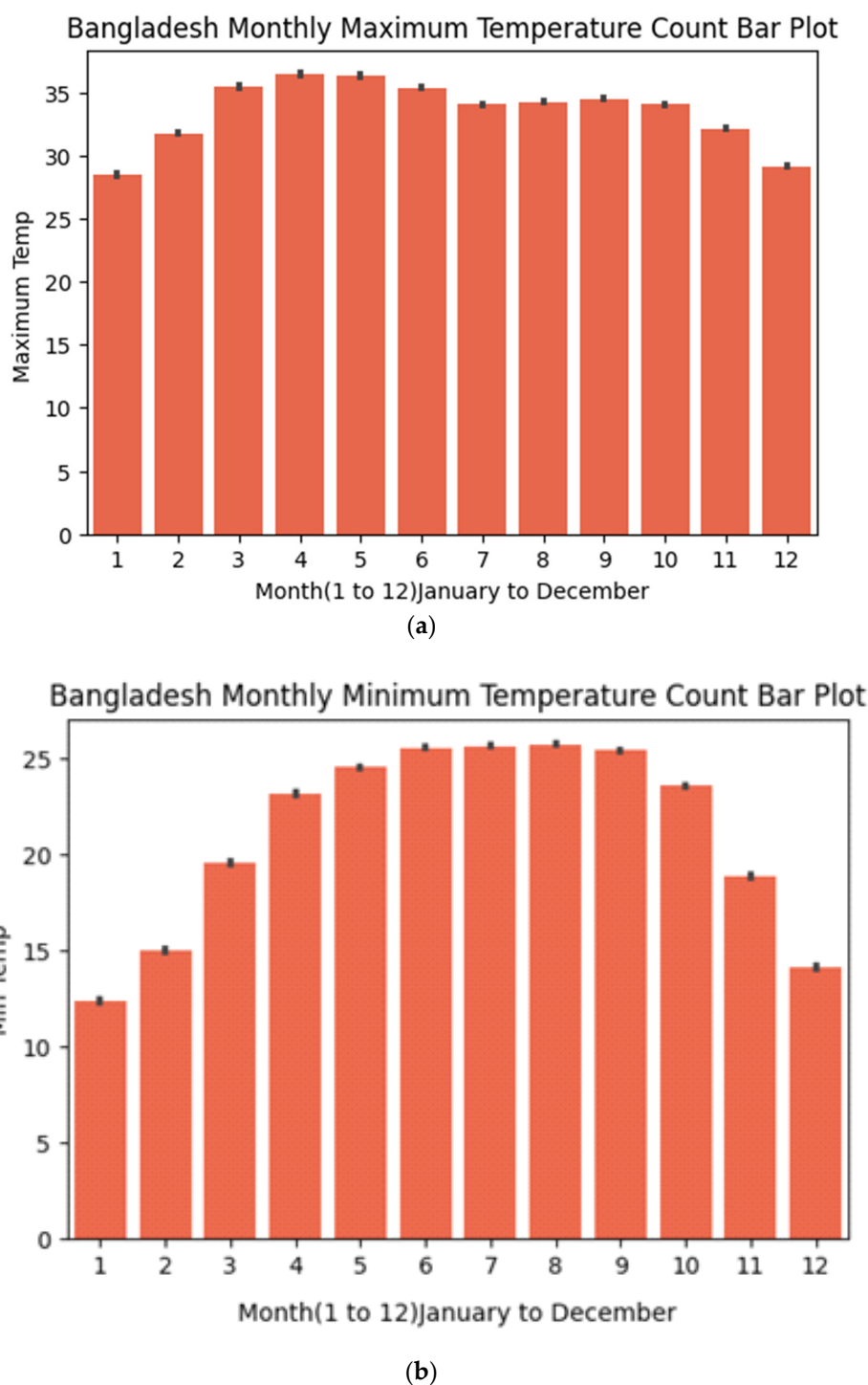


Figure 6. Bar diagram for monthly minimum and maximum temperature in Bangladesh: (a) maximum temperature, (b) minimum temperature.

Designing a deep-learning- and machine-learning-based strategy for predicting rain is the goal of this endeavour. To lay the foundation for building this model, the Bangladesh Meteorological Department's (BMD) dataset with 21,120 records was employed.

Figure 7's bar plot shows that the city in Bangladesh with the highest rainfall amount is Teknaf, followed by Sylhet. Teknaf is a city located in the southeastern part of Bangladesh. The heavy rainfall in Teknaf is caused by various things, including topography, geographic features, and orographic effects. These elements work together to make Teknaf experience more rainfall than other areas of Bangladesh. Additionally, there is the potential for yearly

variation in weather conditions and rainfall amounts, which a variety of weather-related variables and climate change may impact. Details of the eighteen features are provided in Table 2 below.

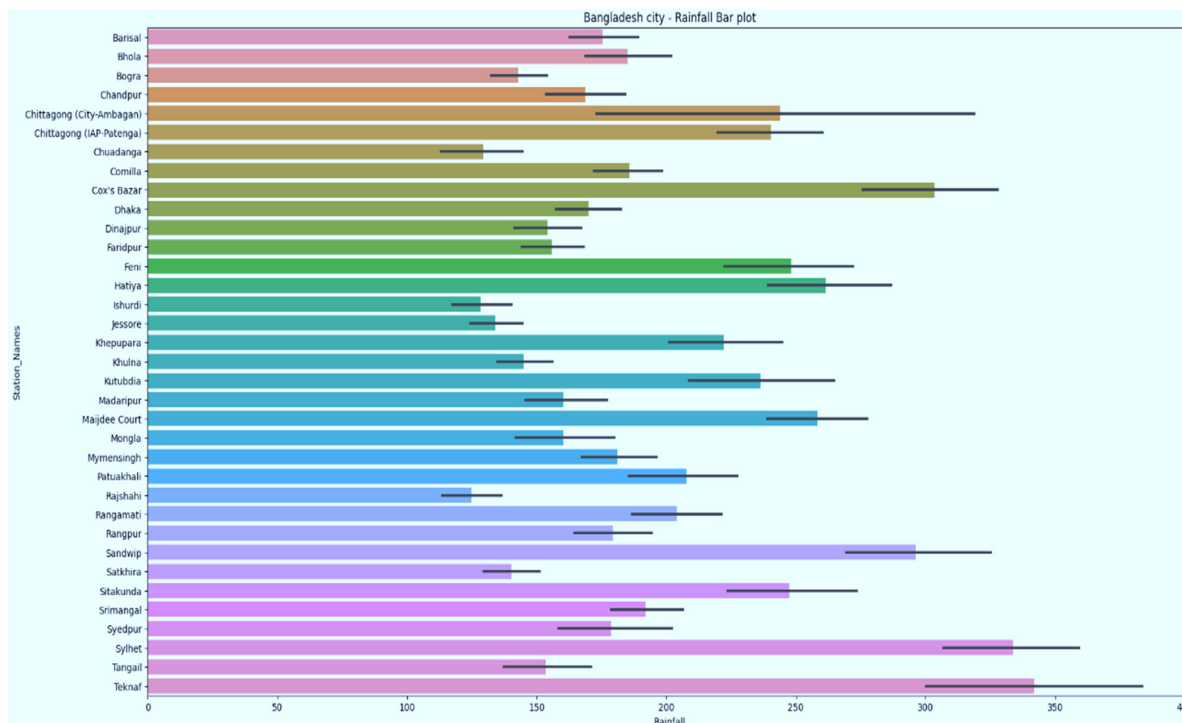


Figure 7. Bar plot for rainfall in Bangladesh.

Table 2. Features used for training machine learning models.

Sr. No.	Attribute	Attributes Description	Type	Measurement
1.	'Unnamed: 0'	This is likely an index or identifier column for the dataset.	integer	serial no
2.	'Station Names'	The name of the city or station where the flood occurred	string	categorical
3.	'Year'	This column represents the year for which the weather data are recorded.	integer	numerical
4.	'Month'	The month of the recorded data.	integer	numerical
5.	'Max Temp 0C'	The maximum recorded temperature of a day.	float	degrees Celsius
6.	'Min Temp 0C'	The minimum temperature experienced on a specific day (degrees Celsius).	float	degrees Celsius
7.	'Rainfall (mm)'	The amount of rainfall recorded (millimetres).	float	millimetres
8.	relative humidity	This column represents the relative humidity recorded for a specific month and year.	float	percentage
9.	'Wind Speed'	Wind speed in a particular direction at a given location and time	float	metres per second.
10.	Cloud Coverage	This column represents the cloud coverage or cloudiness level recorded for a specific month and year.	float	percentage
11.	'Bright Sunshine'	This column contains the duration of bright sunshine recorded for a specific month and year. It represents when the sun is visible, or the sky is clear.	float	measured in hours
12.	'Station Number'	In meteorology and weather monitoring, a station number is a unique identification declared to an individual weather station or monitoring place.	integer	numerical identifier

Table 2. Cont.

Sr. No.	Attribute	Attributes Description	Type	Measurement
13	'X_COR',	This column could represent the X-coordinate or longitude values associated with the location of each weather station.	float	coordinates of the station
14	Y_COR	This column could represent the Y-coordinate or latitude values associated with the location of each weather station.	float	coordinates of the station
15	LATITUDE	The latitude of rainfall on specific locations and weather conditions.	float	latitude coordinate of the station.
16	LONGITUDE	The longitude of rainfall in specific locations and weather conditions.	float	longitude coordinate of the station.
17	'ALT',	This column likely represents the altitude or elevation of each weather station.	num	metres
18	'Period'	Rainfall measurements are gathered or recorded at a specific time step or period (year and month combined).	float	numeric

To perform automated regression evaluations, this work uses machine learning and deep learning techniques. The information from the meteorological department is used to forecast rainfall in Bangladesh. The collection includes several characteristics. However, the output class for the forecast uses an attribute called “Rainfall”. The data collection’s overall histogram illustration is shown in Figure 8. Following the data collection, several preprocessing procedures are carried out, including verifying values, handling, scaling, and transforming some characteristics, like station names, etc. This will permit the supervised learning regression algorithms to provide more accurate predictions. Figure 8 represents the bar plot for rainfall in Bangladesh’s cities.

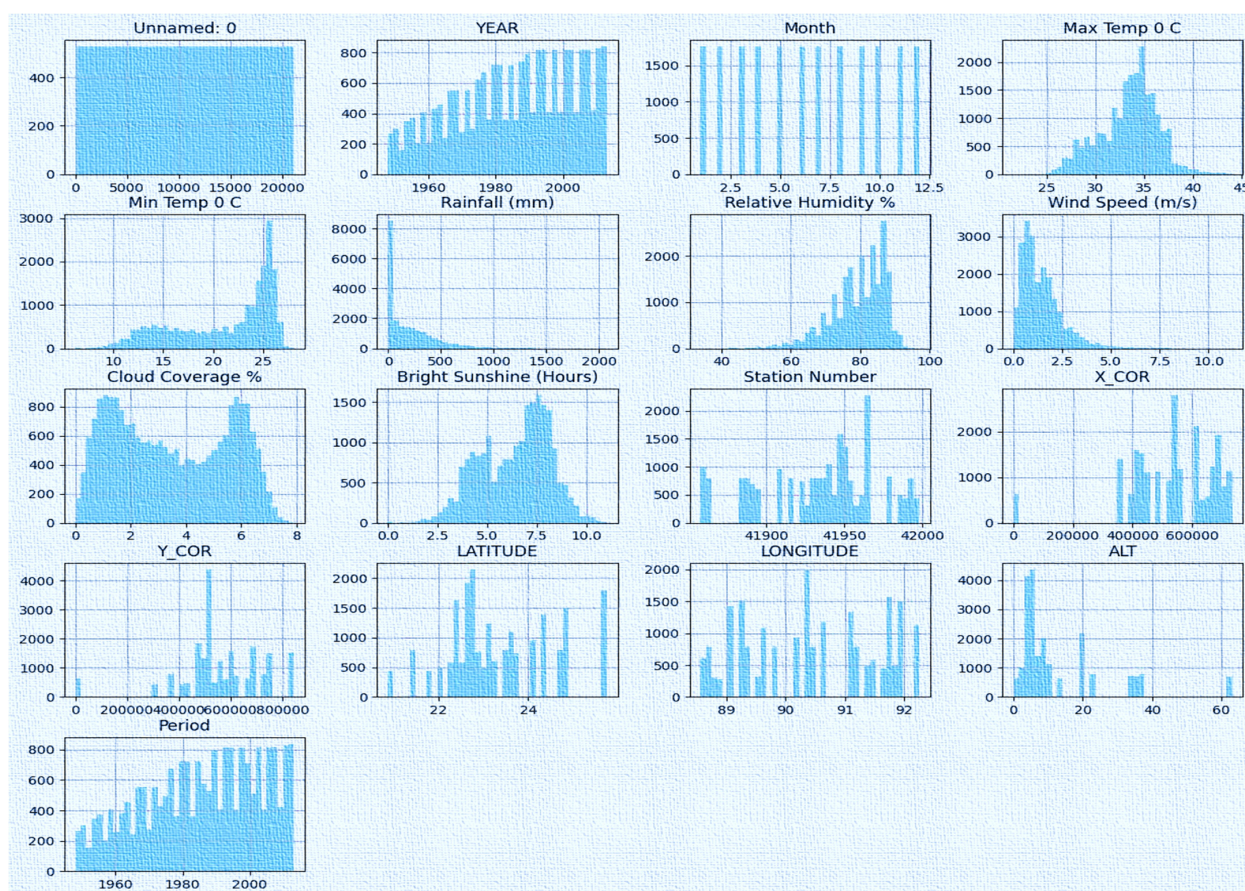


Figure 8. Histogram depicting the feature values of the dataset.

3.2. Dataset Preprocessing or Cleaning

A preliminary processing data mining technique turns unstructured, incorrect input into a format that the model can use and understand. Raw data are uneven, lacking many key aspects, and full of errors. According to data exploration and estimation, there are no redundant, invalid, or null values in the raw data for the used model. It is necessary to choose only the features pertinent to our model for forecasting rainfall during the preprocessing phase of feature selection. This reduces training time and raises the accuracy of the model. Table 3 and Figure 9 show the correlation coefficients of feature rainfall with various variables. Then, the work employs the dropping of features. Correlation is calculated between dependent and independent variables that are further used for modelling. The following columns were dropped. These features have the lowest correlation with the rainfall variable, as seen in Table 3.

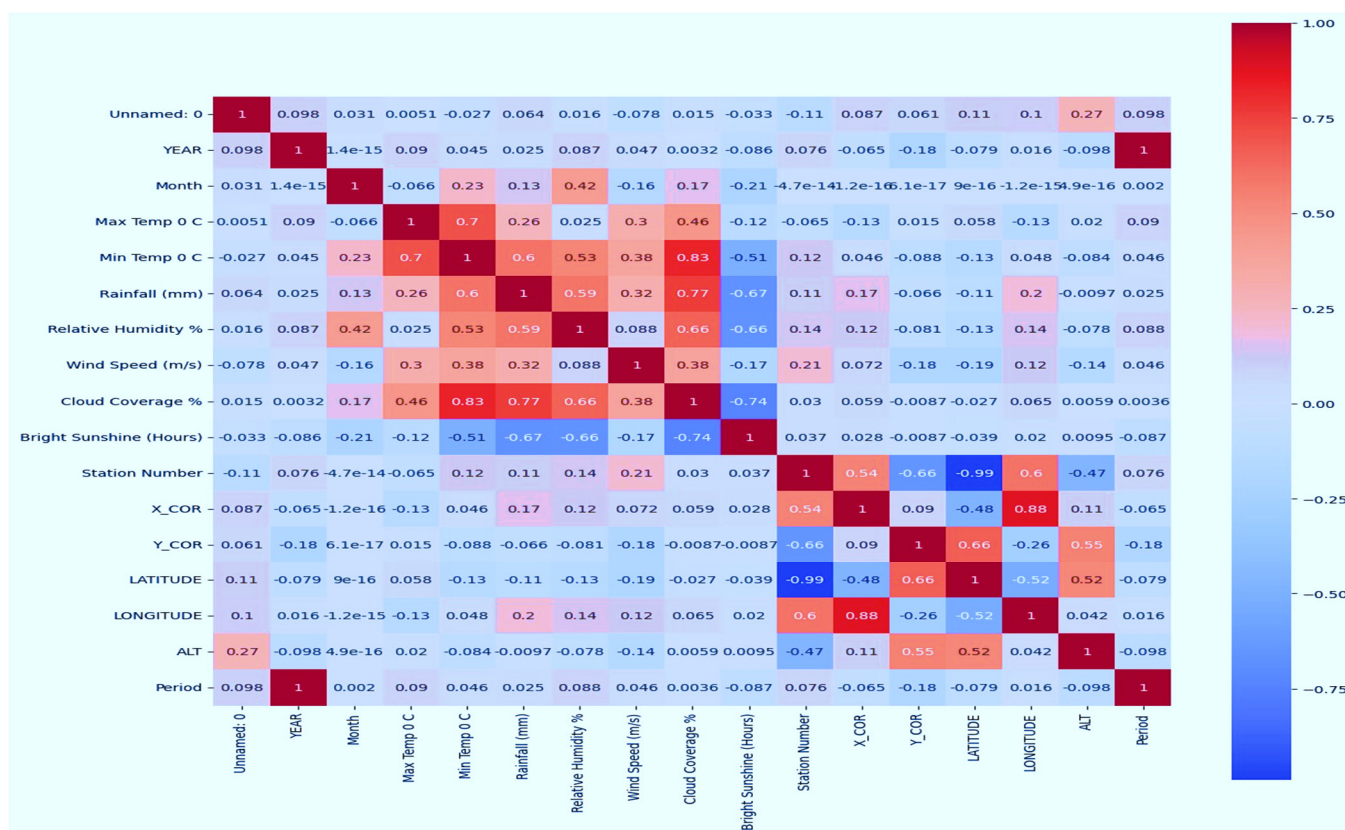


Figure 9. Pictorial representation of correlation between features.

- YEAR: This represents the year;
- X_COR: This represents the x-coordinate where rainfall is happening;
- Y_COR: This represents the y-coordinate;
- LATITUDE: The latitude of the location;
- LONGITUDE: The longitude of the location;
- ALT: The altitude of the location.

Table 3. Correlation coefficients of rainfall with various variables.

Unnam	YEAR	Month	Max Temp °C	Min Temp °C	Rainfall	Humidity	Wind Speed	Cloud Cover- age	Bright Sun- shine	Station Num	X_ COR	Y_ COR	LATI- TUDE	LONGI- TUDE	ALT	Period
0.064153	0.025109	0.132680	0.256821	0.596625	1.000000	1.0000	0.316366	0.766821	−0.673333	0.113804	0.167625	−0.066154	−0.105569	0.197805	−0.009696	0.02536

3.3. Data Normalisation

Data normalisation is scaling the values of variables so that they have the same quantitative weight and lie in an identical interval or scale. They rescale the entire dataset to a standard range or distribution.

The equation for data normalisation using the min–max scaling technique is as follows:

$$x_{normalized} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

where x represents the original value in the dataset, $\min(x)$ is the minimum value of the dataset, $\max(x)$ is the maximum value of the dataset and, $x_{normalized}$ is the normalised value of x within the range $[0, 1]$. Equation (1) scales the values of the entire dataset to the range $[0, 1]$,

3.4. Feature Encoding

Feature encoding is an essential stage in data preprocessing, specifically for machine learning and deep learning. It entails converting textual or category information into a numerical representation that algorithms can utilise. The dataset's primary feature, "Station Names", which provides string-type data, needs attention. Although machine learning algorithms perform better with numbers, textual data are encoded. The attribute "Station" contains a list of all the station names from which the daily rainfall statistics have been gathered. Without adding any further fields, the category values of the "Station" variable were then translated using label encoding into numerical values. There are numerous varieties of feature-encoding methods. In our experiment, we employ label encoding, which gives each category a distinct integer and is generally applied to data with ordinal values. Label encoding assigns a unique integer to each station name.

3.5. Feature Scaling

Feature scaling is the scaling of individual features to have similar magnitudes within the dataset. To ensure that the dataset was fair and applicable to the models utilised, the dataset's features were scaled using the Standard Scaler. Equation (2) for feature scaling using the z-score scaling or standardisation technique is as follows, where feature scaling is independently performed on individual features:

$$x_{scaled} = \frac{x - \mu}{\sigma} \quad (2)$$

where x = original value of the feature, μ = mean (average) of the feature in the dataset, σ = standard deviation of the feature in the dataset, and x_{scaled} = scaled value of the feature.

3.6. Machine Learning Models

One of the most popular and effective types of algorithmic learning is supervised machine learning, and the types of machine learning algorithms are presented in Figure 10. When we have a few instances of features–label pairings and want to predict a specific outcome or label from a set of features, we apply supervised learning. Our training set, which consists of these features–label pairings, is used to create a machine learning model. Our objective is to accurately predict new, unforeseen data. The most effective machine learning and deep learning methods for analysing the daily rainfall quantity forecasts have been chosen after evaluating many articles on rainfall prediction [8–20].

The used dataset falls under the category of a regression problem since rainfall prediction is a continuous number, or what programmers refer to as a floating-point number. In this study, the regression algorithm is used to train the dataset. Eight machine learning algorithms (polynomial linear regression, multiple linear regression, k nearest neighbours' regression, regression with decision trees, support vector machine, random forest model, Ada boost regression, and stacking regression) were tested and compared using real-time

environmental data to forecast the daily intensity of the rainfall. The algorithms with the best accuracy were reported.

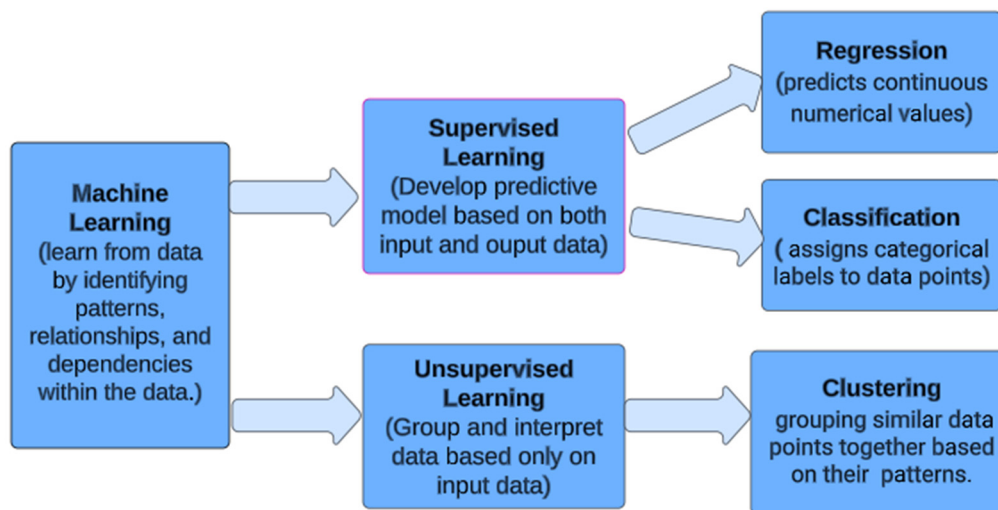


Figure 10. Various types of machine learning algorithms proposed in the literature.

3.6.1. Polynomial Linear Regression

A variation on simple linear regression, polynomial linear regression, enables more intricate connections between the predictor variables (features) and the variable of interest. The predictors are changed by being raised to different powers to create polynomial terms in polynomial linear regression. Equation (3) represents polynomial linear regression, as follows:

$$y = \beta_0 + \beta_1 x^1 + \beta_2 x^2 + \dots + \beta_n x^n + \epsilon \quad (3)$$

where y represents the target variable (dependent variable), x denotes the predictor variable (independent variable), $\beta_1 \dots \beta_n$ are coefficients to be estimated, and ϵ is the error term by introducing higher-order terms ($x^1, x^2, x^3 \dots x^n$). A polynomial linear regression can capture nonlinear relationships between the predictors and the target variable.

3.6.2. Multiple Linear Regression

With many predictor variables, multiple linear regression attempts to forecast the target variable. It requires that the predictors and the target variable have a linear relationship. Multiple linear regression's equations may be written as in Equation (4):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (4)$$

where y represents the target variable (dependent variable), x_1 and x_2 denotes the predictor variable (independent variable), $\beta_0, \beta_2 \dots \beta_n$ are coefficient to be estimated, and ϵ is the error tem. Figure 11 shows the results using multiple linear regression actual and predicted rainfall minimum temperature in terms of R^2 and RMSE score.

3.6.3. K-Nearest Neighbours Regressor

In order to deal with classification and regression forecasting problems, the KNN method is used. It falls within the category of supervised machine learning. The KNN method uses feature similarity to predict the values of new data points, which implies that the value of the new data point will depend on how closely it resembles the points in the training set. Although there are numerous distance functions, Euclidean is the most widely used one for calculating distance. The given distance equations for two- and multidimensional values are shown in Equations (5) and (6), as follows:

For two dimensions:

$$d = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]} \quad (5)$$

For multiple dimensions:

$$d = \sqrt{\sum_{i=1}^m [(x_i - y_i)^2]} \quad (6)$$

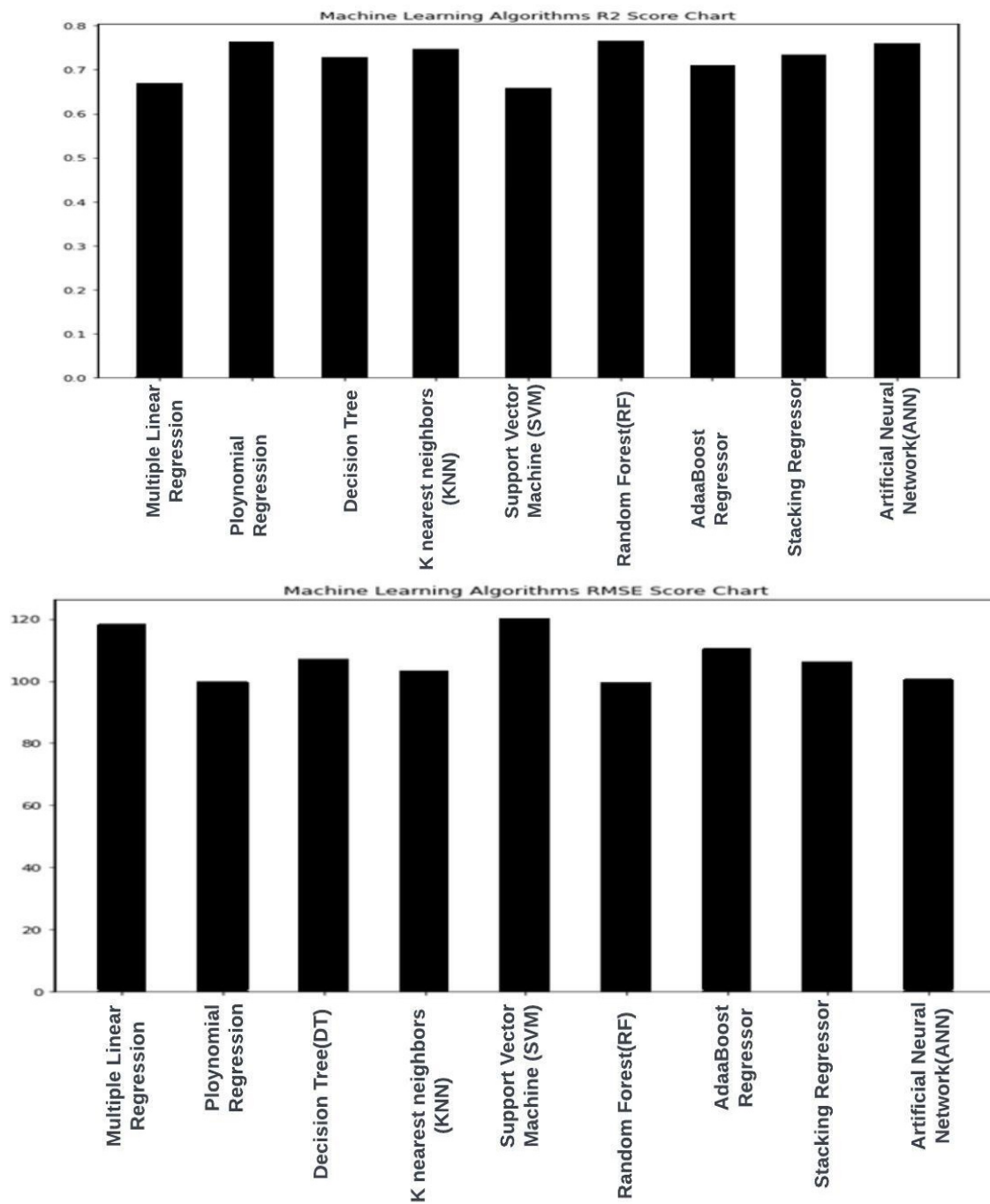


Figure 11. Results illustrating R² and RMSE values for various machine learning algorithms.

3.6.4. Decision Tree Regressor

A supervised learning approach for regression tasks is called the Decision Tree Regressor. It is a non-parametric approach that creates a model that resembles a tree in order to generate predictions using a set of decision rules deduced from training data.

3.6.5. Support Vector Machine

A supervised learning approach extends the core concept of the Support Vector Machine (SVM) method to accommodate continuous target variables. To convert the provided data characteristics into a higher dimension space, SVR employs a core implementation of kernel. The decision boundary's form and the model's flexibility are influenced by the selection of the main function. In our research, we use radial basis function for execution.

3.6.6. Random Forest Model

An adaptable, simple machine learning method called random forest typically produces outstanding results even without the use of hyper-parameter modification. It has become one of the most extensively used methods due to its simplicity and adaptability (it can be used for regression and classification).

3.6.7. AdaBoostRegressor

A supervised machine learning approach called AdaBoostRegressor is used for regression problems. It is an adaptation of the AdaBoost (Adaptive Boosting) technique, which combines a number of weak learners (regression models) to produce an ensemble model that is robust.

3.6.8. Stacking Regressor

The Stacking Regressor is a method based on machine learning that mixes various separate regressor algorithms to build a meta-regressor for regression problems. It is based on the idea of "model stacking", in which inputs from several models are combined to create the final prediction using an advanced model.

3.7. Deep Learning Model

Relying on this kind of training data and the learning goals, deep learning models can be divided into supervised and unsupervised learning, as shown in Figure 12. We used in our research, for the prediction of rainfall, supervised deep learning. Supervised deep learning, input features and their matching target labels are supplied, and the models are trained using labelled data. The objective is to learn the relationships between the provided features and the desired output labels. Several well-liked supervised deep learning architectures are as follows: artificial neural network (ANN), convolutional neural network (CNN) and recurrent neural network (RNN). Multiple-layered neural networks are used in deep learning algorithms and machine learning approaches, which enable the model to automatically extract representations and patterns from data. Supervised and unsupervised learning are the two main types of deep learning. A labelled dataset is used to train the model in supervised learning, where each data point is linked to a specific target or label. Training a model on an unlabelled dataset—one without predetermined goal labels—is known as unsupervised learning. The dataset in this study has been trained using the deep learning technique. Using real-time environmental data, three deep learning algorithms—recurrent neural network (RNN), long short-term memory (LSTM), and artificial neural network (ANN)—were evaluated and contrasted in order to predict the typical intensity of the rainfall. The algorithms with the best accuracy were reported.

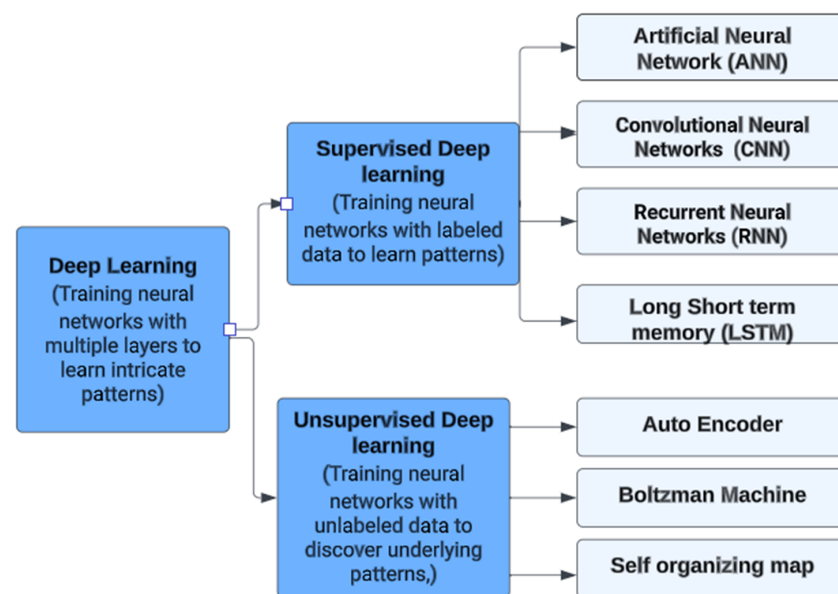


Figure 12. Various types of deep learning algorithms.

3.7.1. Artificial Neural Network (ANN)

A computational model called an artificial neural network (ANN) is motivated by the organisation and operation of the neural networks in the human brain. It is a particular kind of ML approach that can consider input data while making predictions or choices. The input layer, hidden layer(s), and output layer are the three primary types of layers in an ANN (as shown in Figure 13). The weights indicate the connections between neurons and specify the strength or significance of the information transmitted between them. During the training phase, the ANN updates the weights of the connections based on the given input data and the needed output. The error between the projected and actual output is often transmitted backwards across the network to update the weights in a procedure known as backpropagation. The ANN learns and develops its prediction or decision-making skills thanks to this iterative approach [8,57,61]. The net input for the general artificial neural network model mentioned below can be computed by using Equations (7) and (8), as follows:

$$Y_{in} = x_1 \cdot w_1 + x_2 \cdot w_2 + x_3 \cdot w_3 + \dots + x_m \cdot w_m \quad \text{i.e., } Y_{in} = \sum_i^m x_i \cdot w_i \quad (7)$$

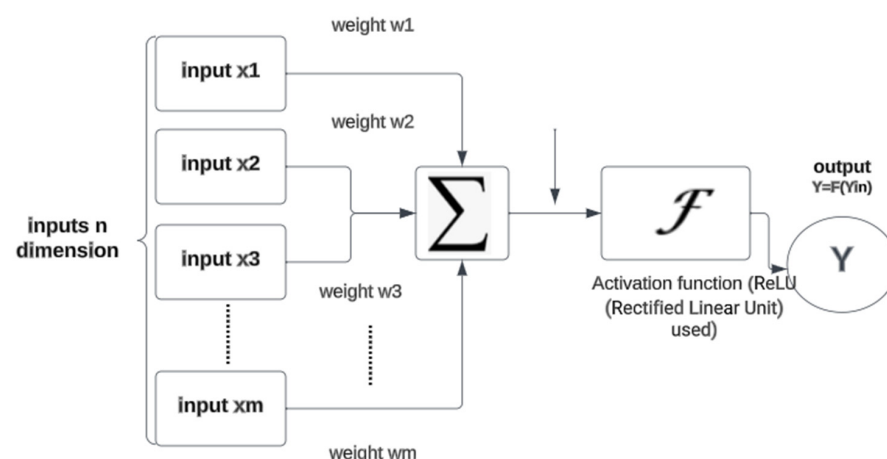


Figure 13. Artificial neural network.

Applying the activation function to the net input allows for the output estimation.

$$Y = F(Y_{in}) \quad (8)$$

3.7.2. Recurrent Neural Network (RNN)

The RNN [5] analyses time series or sequential data. Instead of processing input data in a single pass from input to output like feedforward neural networks do, RNNs have a feedback loop that enables information to persist over time and be shared between temporal stages (Figure 14). An RNN's internal capacity to maintain a state or memory enables it to recognise dependencies and patterns in sequential input. Through every time step, its internal state is changed while considering both the recent and the last input state. It can use data from earlier time steps to make predictions or generate outputs influenced by the entire input sequence. A recurrent neural network (RNN) has various parameters that govern the network's behaviour and properties. The critical RNN parameters are as follows: Input size: The dimensionality of input. It determines the number of RNN inputs. The hidden size determines the number of memory cells in the RNN. It indicates the dimensionality of the network's internal state or memory. The output size specifies the dimensionality of the output at each time step. It determines the number of RNN output nodes—Sigmoid, tanh, and ReLU are examples of standard activation functions. Weight matrix: Weight matrices manage the flow of information between the input, hidden, and output layers in RNNs. These matrices are learnt throughout the training process and regulate how each layer affects the others. The initial concealed state is the beginning point for the internal state of the RNN. It can be set to zero or learnt during training as a parameter. Recurrent weight matrix: The recurrent weight matrix relates the last time step hidden state to the current time step hidden state. It governs how the memory of the RNN is refreshed and transmitted over time. Bias Terms: At each layer, bias terms are added to the weighted inputs to induce an offset or bias in the computation. They enable the RNN to learn various intercepts for various characteristics. The following equations regulate the computations of an RNN. For hidden state computation, Equation (9) is used:

$$h(t) = f(W_{xh} * x(t) + W_{hh} * h(t-1) + b_h) \quad (9)$$

Output computation is performed by using Equation (10):

$$y(t) = f(W_{hy} * h(t) + b_y) \quad (10)$$

where t = time step, $x(t)$ = input at ' t ', and $h(t)$ = hidden state at ' t '. W_{xh} , W_{hh} , W_{hy} are weight matrices that control the flow of information, and b_h and b_y are bias vectors.

3.7.3. Long Short-Term Memory

This framework is used to prevent the issue of vanishing gradients and preserve persistent dependencies in sequential provided data. The input is represented by $x(t)$ at time t . The hidden state from the previous time step ($t-1$) is represented by $h(t-1)$. Cell state: the cell state from the prior time step ($t-1$) is represented by $c(t-1)$. Output: at time t , the hidden state or output is represented by $h(t)$. The hidden state in an LSTM (Long Short-Term Memory) represents the output or information propagated to the next time step in the sequence (as shown in Figure 14). It serves as a memory of the previous time steps and captures relevant information from the input sequence. The work for an LSTM unit is as follows: for the forget gate, we use Equation (11):

$$f(t) = \text{sigmoid}(W_f * [h(t-1), x(t)] + b_f) \quad (11)$$

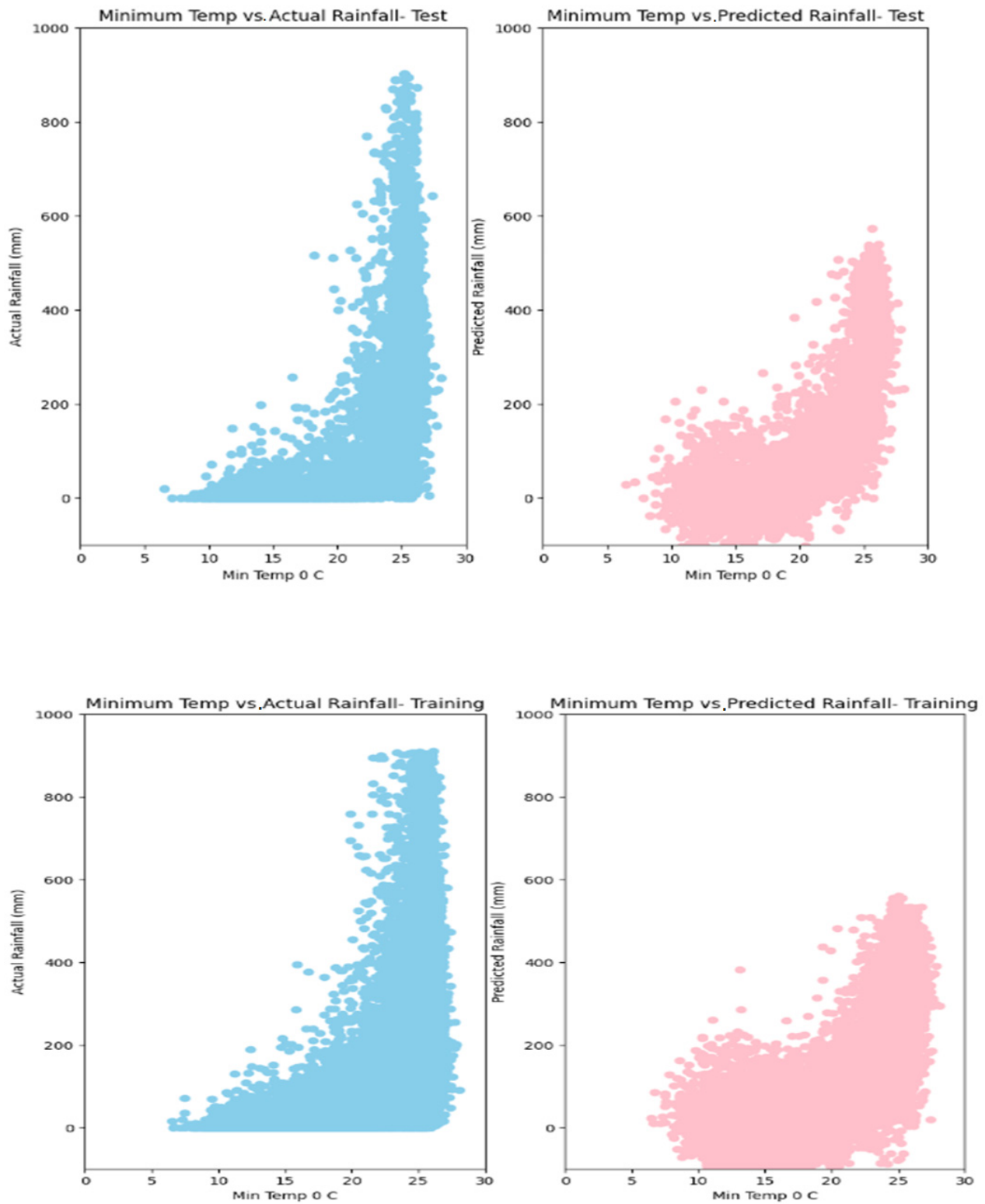


Figure 14. Illustration of actual and predicted rainfall using multiple linear regression.

For the input gate, we use Equations (12)–(16):

$$i(t) = \text{sigmoid}(W_i * [h(t-1), x(t)] + b_i) \quad (12)$$

$$\hat{C}(t) = \text{tanh}(W_c * [h(t-1), x(t)] + b_c) \quad (13)$$

Updating cell state works as follows:

$$c(t) = f(t) * c(t-1) + i(t) * \hat{C}(t) \quad (14)$$

Output gate calculations are as follows:

$$o(t) = \text{sigmoid}(W_o * [h(t-1), x(t)] + b_o) \quad (15)$$

Hidden state calculation is carried out as follows:

$$h(t) = o(t) * \tanh(c(t)) \quad (16)$$

3.8. Implementation Details

We can efficiently capture the correlations between meteorological factors and rainfall by utilising machine learning algorithms like regression models, decision trees, random forests, or support vector machines. We can estimate rainfall with a respectable degree of accuracy by taking advantage of these models' strengths to learn from previous data and make predictions based on input features. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks are examples of deep learning techniques that are particularly good at detecting complex connections and patterns in sequential or spatial data. These models are ideally suited for applications requiring rainfall prediction because they can easily handle the temporal or spatial character of rainfall patterns. Deep learning algorithms may learn complicated representations and produce precise forecasts by training on previous rainfall data. In our research work, we implemented both techniques for producing a better rainfall model. The steps for implementation are as follows. Table 4 summarises the details.

Table 4. Details of implementation.

Parameter	Values
Framework	Sk-learn, tensorflow
Training, validation, testing	60%, 20%, 20%
Number of epochs	30
Stopping criterion	Early stopping
Activation functions	ReLU
Optimiser	Adam
Validation criterion	3-fold cross validation

- The various machine learning models used in this study have been implemented using sk-learn.
- For deep learning models, tensorflow has been used.
- For optimising and finetuning of various hyperparameters, k-fold cross validation has been performed.
- In addition, earlystopping callback of tensorflow has been used.
- The models are trained for 30 epochs.
- The validation split is 20%.
- The ReLu activation function has been used along with Adam optimiser.

As mentioned in the table, a resampling method called 3-fold cross-validation is used to assess machine learning models on a small sample of data. The goal is to evaluate how effectively a model's output will transfer to a different collection of data. Preventing overfitting, a situation in which a model learns the training data too well—including its noise and outliers—and hence performs badly on unknown data, is one of the main goals of cross-validation [71].

4. Criterion for Evaluating Models

A measurement metric is a quantitative measure utilised to analyse the quality in terms of efficiency or performance accuracy of a machine learning method. It gives a standardised technique to evaluate how effectively the model functions concerning its intended task or aim. In regression tasks, numerous evaluation metrics can be used to analyse the performance of a regression model. Based on forecasting data and training data, here are some indicators we utilised for our research result.

4.1. RMSE (Root Mean Squared Error)

RMSE is calculated by taking the square root of the average of the squared differences between the predicted values (y_{pred}) and the actual values (y_{true}) (see Equations (17) and (18)).

$$RMSE = \sqrt{MSE} \quad (17)$$

$$RMSE = \frac{1}{N} \sqrt{\sum_{i=1}^N (y_i - \hat{y})^2} \quad (18)$$

where n is the number of data points and Σ represents the sum of squared differences across all data points.

4.2. R-Squared (Coefficient of Determination)

R-squared (see Equation (19)) measures the proportion of variance in the dependent variable (y_{true}) that is explained by the independent variables (y_{pred}). It ranges from 0 to 1, with 1 indicating a perfect fit.

$$R^2 = \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \quad (19)$$

where Σ represents the sum of squared differences between the predicted and actual values and y_{mean} is the mean of the actual values.

5. Results

Figures 14–22 compare the actual rainfall with predicted rainfall based on temperature. The graphs are plotted for both training data and testing data. The graph shows the data distribution for both actual values during training and testing. The best results are obtained using polynomial regression and random forest with an R^2 value of 0.76. The RMSE values are also shallow for these machine learning models. Also, the graphs for these models (Figures 14 and 18) have similar distributions of actual and predicted values during testing, whereas for other models, the distribution varies a lot between actual and predicted values.

The results describe the actual vs. predicted results of the models. The precipitation recorded or witnessed at particular locations in Bangladesh over a specified period is called actual rainfall. It is founded on information from satellite imaging, weather and precipitation monitoring stations, and other reliable data sources.

The ground truth, or actual rainfall, is used to compare predictions and is essential for evaluating the model's accuracy. The estimations or forecasts of future rainfall produced using a hydrological model, weather forecasting system, or predictive model are the predicted rainfall. These forecasts are often based on historical weather data, atmospheric conditions, and mathematical models that simulate precipitation patterns. In this work, the variation in rainfall based on feature minimum or maximum temperature is measured using multiple linear regression, polynomial linear regression, decision tree regression, k-nearest neighbours, support vector machine, random forest, AdaBoostRegressor, Stacking Regressor, and artificial neural network. Table 5 shows the results of model implementation.

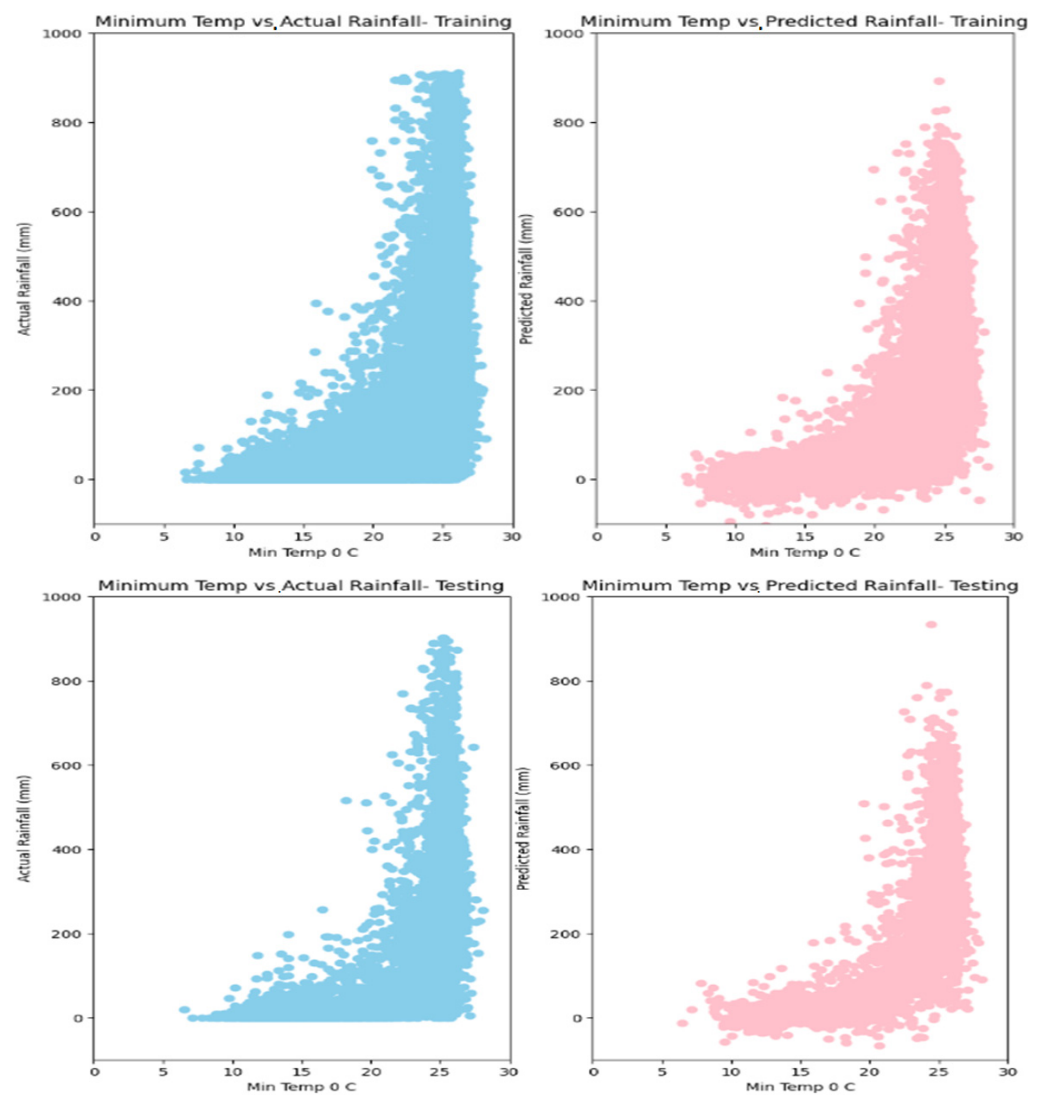


Figure 15. Illustration of actual and predicted rainfall using polynomial regression.

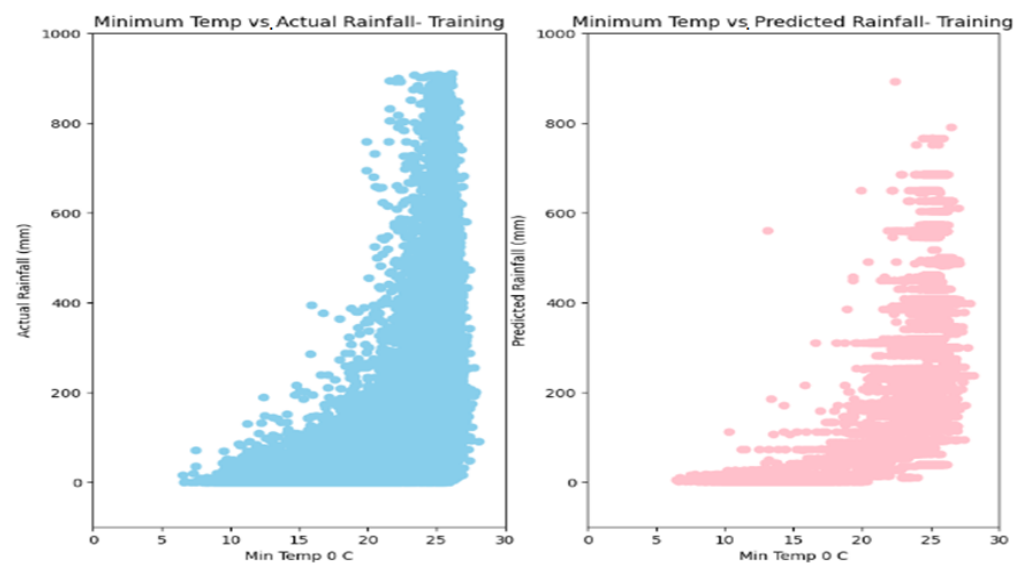


Figure 16. Cont.

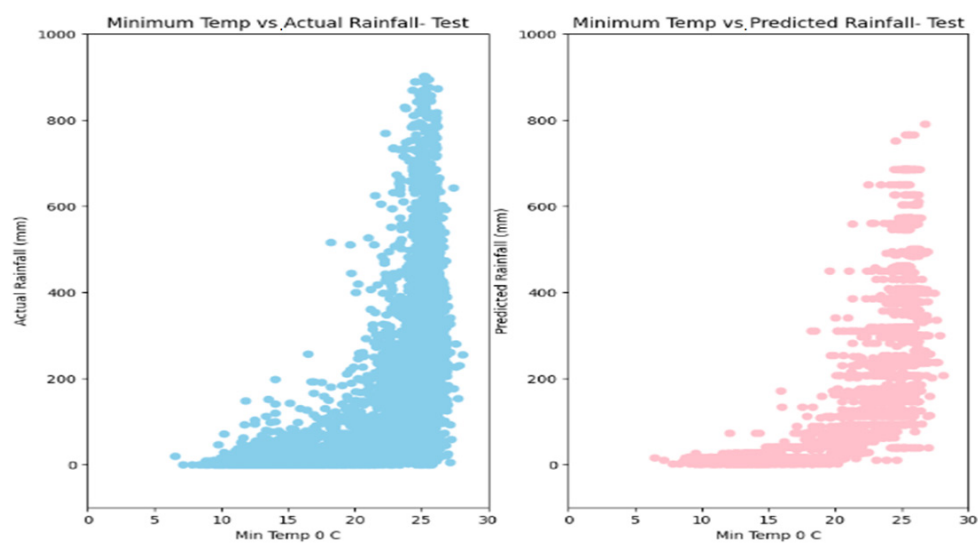


Figure 16. Illustration of actual and predicted rainfall using decision tree.

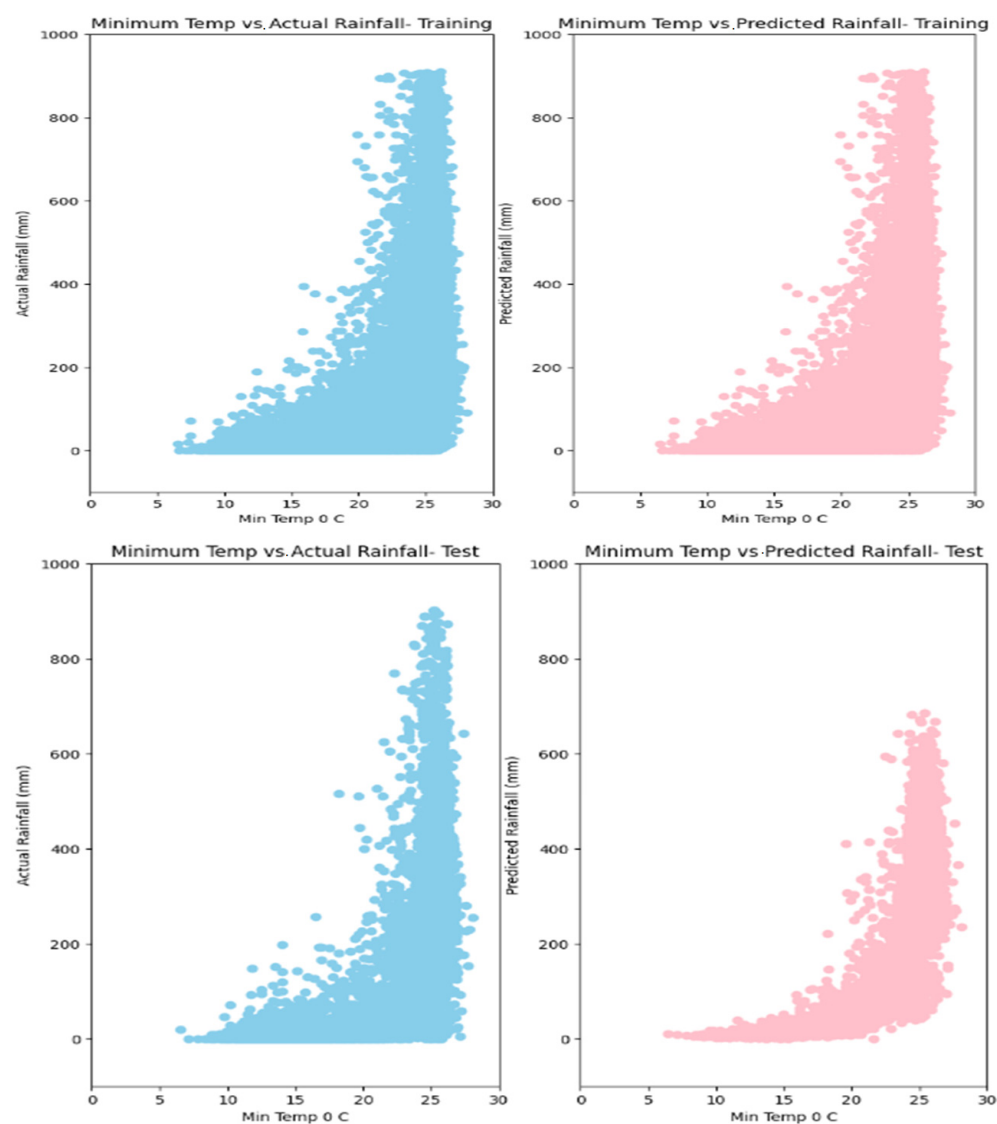


Figure 17. Illustration of actual and predicted rainfall using k-nearest neighbour.

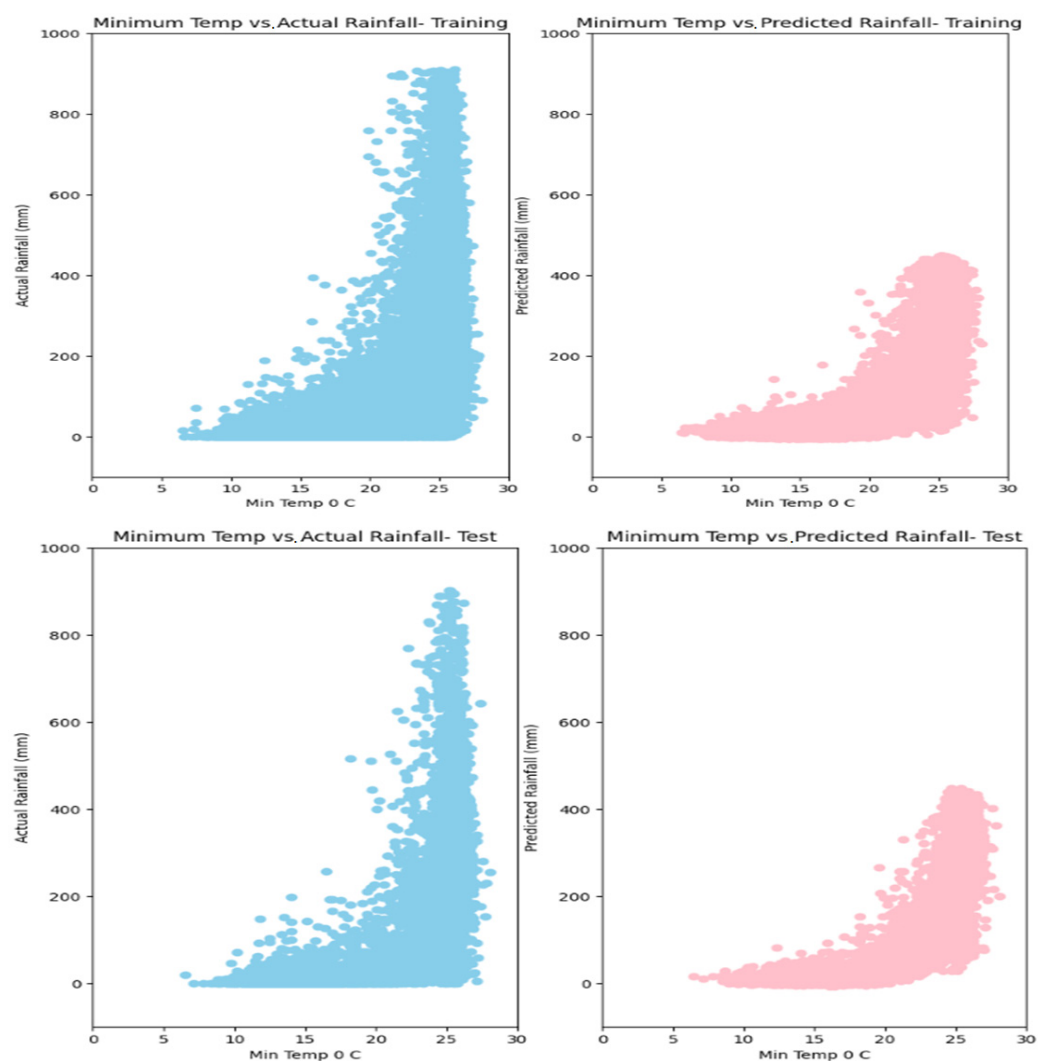


Figure 18. Illustration of actual and predicted rainfall using support vector machine.

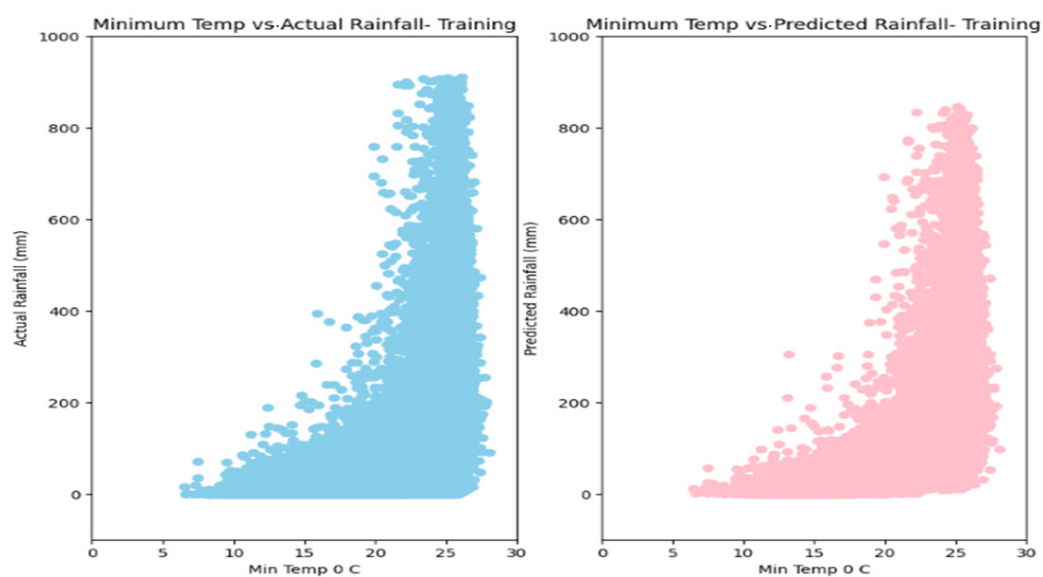


Figure 19. Cont.

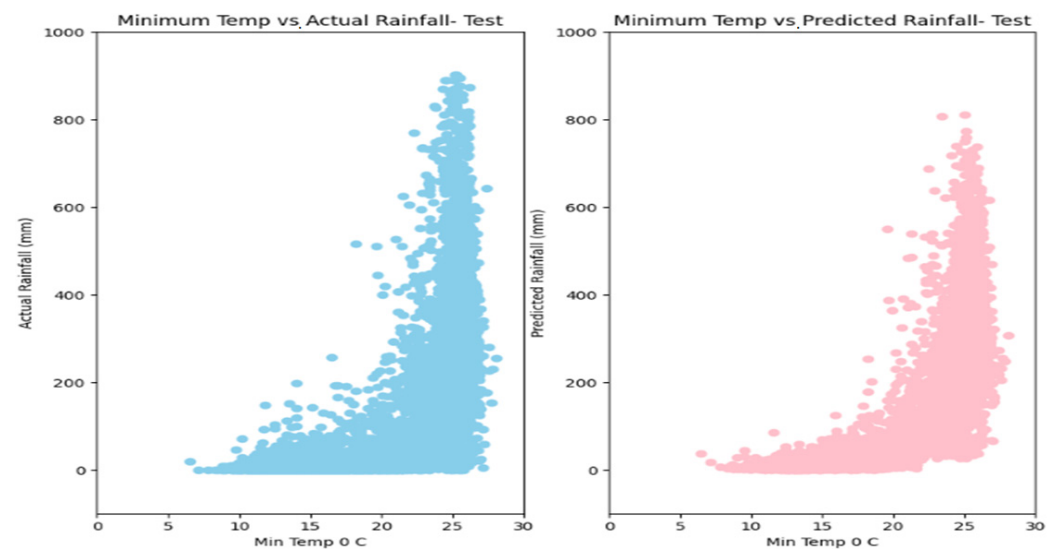


Figure 19. Illustration of actual and predicted rainfall using random forest.

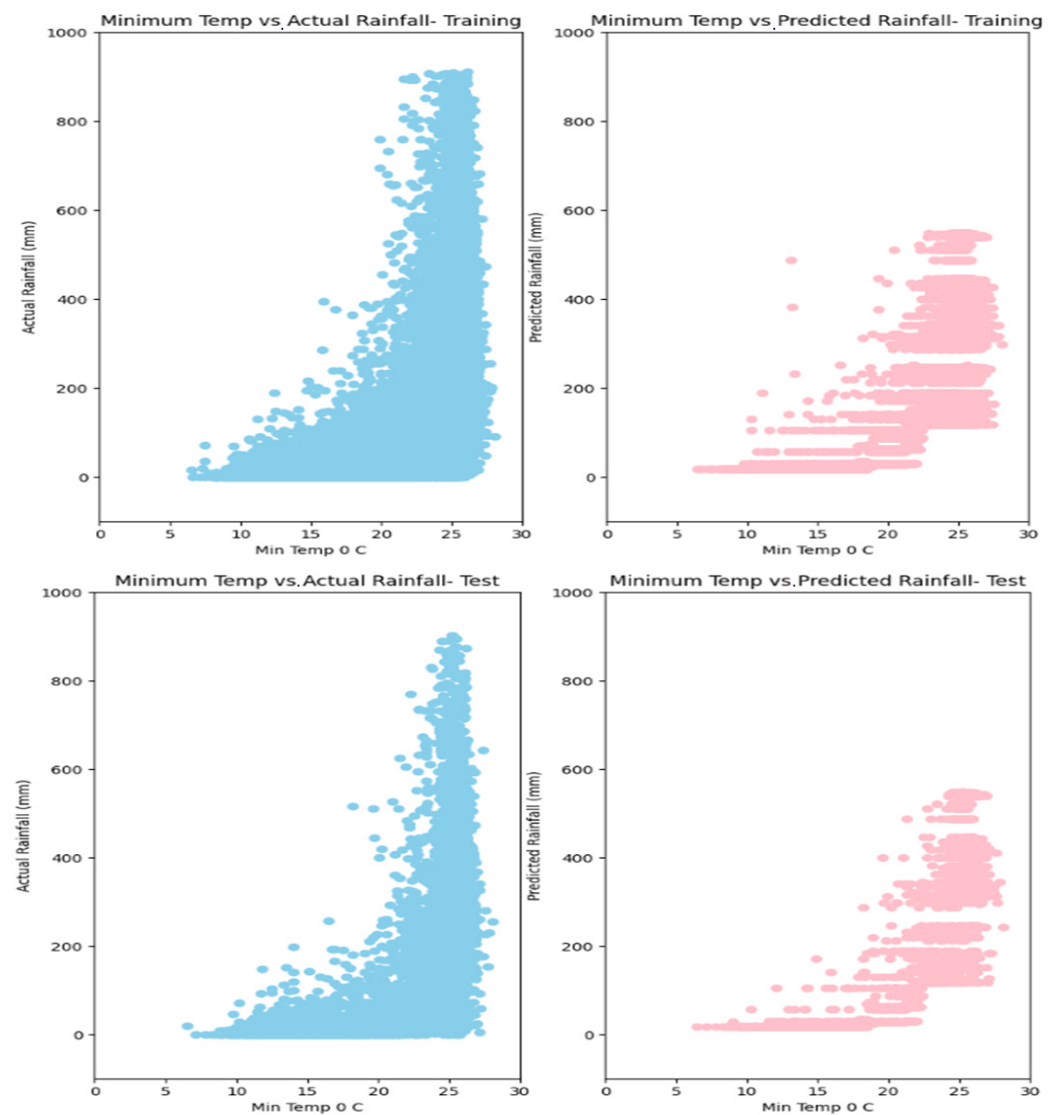


Figure 20. Illustration of actual and predicted rainfall using AdaBoost.

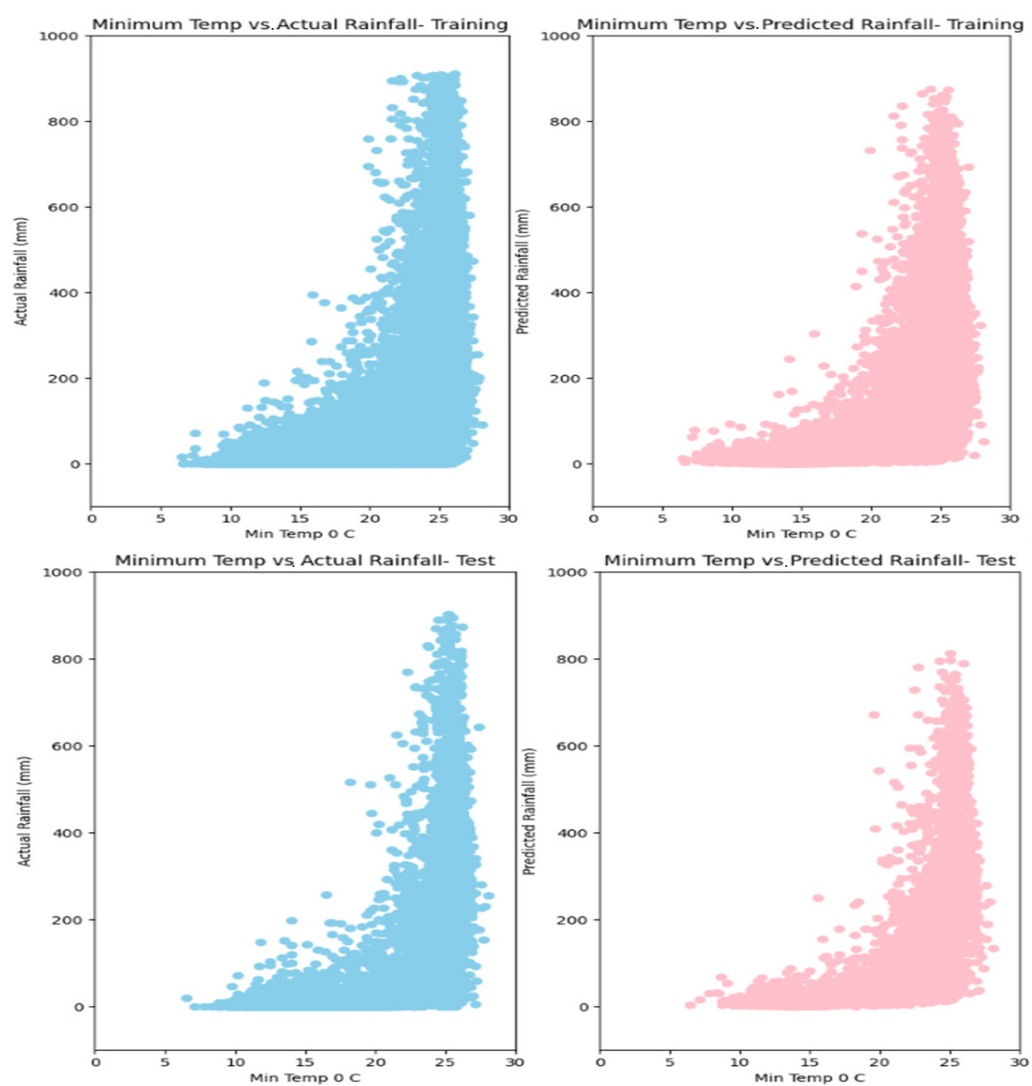


Figure 21. Illustration of actual and predicted rainfall using the Stacking Regressor model.

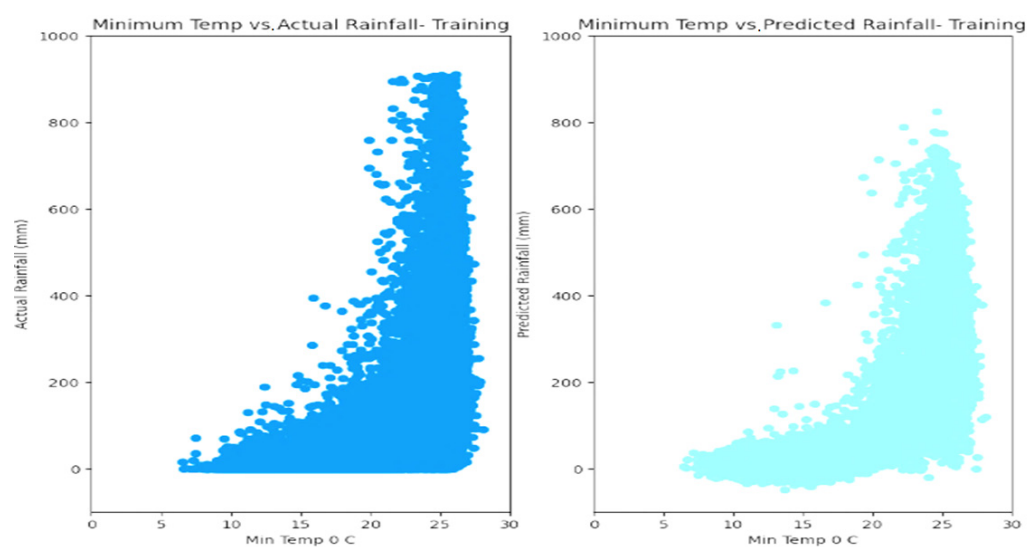


Figure 22. Cont.

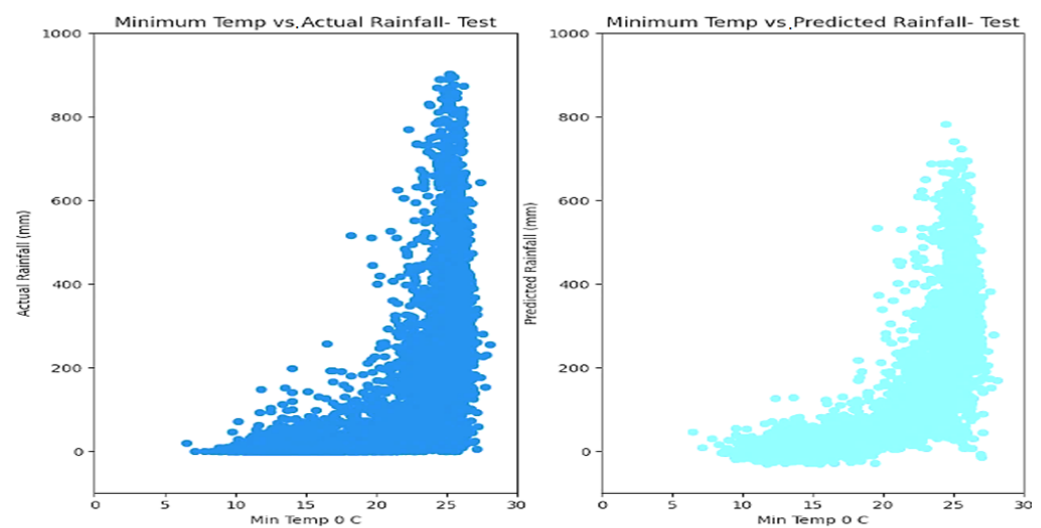


Figure 22. Illustration of actual and predicted rainfall using ANN.

Table 5. Results of implemented machine learning models.

Machine and Deep Learning Model		Evaluation Metrics R^2 and RMSE			
S.No.	ML Model	R^2 Score Training	R^2 Score Testing	RMSE Score Training	RMSE Testing
1.	Multiple Linear regression	0.6643	0.6687	118.3231	118.279217
2.	Polynomial regression	0.773177	0.7642164	99.12397	99.844
3.	Decision Tree mode	0.75	0.72	101.195	123.27715
4.	k-nearest neighbours	0.9992	0.74723	5.5840	103.31968
5.	Support vector machine	0.654139	0.6583	120.108182	120.12110
6.	Random Forest	0.96417	0.768234	38.656	99.5790
7.	AdaBoostRegressor	0.7047	0.710915	110.9689	110.49437
8.	Stacking Regressor	0.74631	0.738501	102.88535	106.1608
9.	Artificial Neural Network	0.763247	0.75847	100.911	100.77041

Table 6 shows the results obtained via LSTM and RNN. Various statistics, such as loss, validation loss, RMSE and testing, are also shown. It can be seen that the loss values for LSTM are significantly better than RNN.

Table 6. Results obtained using various deep learning models (LSTM and RNN).

S. No	Model	Architecture	Parameters	Value
1.	LSTM In order to address and overcome the shortcomings of conventional RNNs, the LSTM approach was specifically developed for learning long-term dependencies.	Refer to Figure 23	Loss	0.0904
			RMSE	0.3007
			Val_loss	0.0906
			Testing set loss	93,260.7188
2.	RNN The main feature of an RNN is its ability to maintain a hidden state or memory, which is revised at each time step and passed as input to the next, allowing the network to consider previous information while processing the current input.	Refer to Figure 24	Loss	126.5478
			mean_absolute_error:	126.5478
			Val_loss	124.1010
			Val_mean_absolute_error	124.1010

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 64)	20224
dense (Dense)	(None, 14)	910

```

Total params: 21,134
Trainable params: 21,134
Non-trainable params: 0

```

Figure 23. Architecture of LSTM model used for training.

```
Model: "sequential_3"
```

Layer (type)	Output Shape	Param #
simple_rnn_3 (SimpleRNN)	(None, 100)	10200
dropout_5 (Dropout)	(None, 100)	0
dropout_6 (Dropout)	(None, 100)	0
dense_3 (Dense)	(None, 1)	101

```

Total params: 10,301
Trainable params: 10,301
Non-trainable params: 0

```

Figure 24. Architecture of RNN model used for training.

6. Discussion

This paper provided a comparison of various machine learning and deep learning models. It can be seen that polynomial regression, random forest and LSTM provide the best results. We have primarily used sk-learn and TensorFlow along with k-fold cross-validation. The other machine learning models employed in previous studies gave poor results primarily because they failed to capture the complex trends available in the data. These models (linear regression, decision trees and support vector machines) are primarily linear and cannot capture the non-linear trends in the data. Such models have shallow capacity. In contrast, polynomial regression employs a higher capacity model to capture the variance in data. The hypothesis space of polynomial regression is much richer and it can capture non-linear patterns in the data. The same goes for the random forest that uses trees in parallel to model the complex data patterns. Finally, LSTM is a deep neural network that has several layers along with the activation functions to model trends in the data. In addition, the LSTM model can capture the sequential patterns in the data. Therefore, the results obtained from these models are suitable. Finally, several similar studies have been reported in the literature. Similar results have been reported in [71], where R^2 values were obtained to be close to 0.87 and 0.92 [72]. The authors of [72] also reported similar results.

Before proceeding towards the end of the discussion, let us explore the results in depth. Table 5 shows the results of various implemented machine learning models. Their R^2 and RMSE values are shown for both training and testing. The primary concern is the values of these parameters during testing. A higher R^2 value shows the supremacy of the model. We can see that polynomial regression provides the highest R^2 value of 0.76. Similarly, a lower error values (RMSE) also shows the model is performing well. Again, the RMSE value of 99.844 is the lowest for polynomial regression. There are other models as well, such as multiple linear regression, decision tree, k-nearest neighbour and support vector machines. However, these models have R^2 values of less than 0.76. Figures 14–22 show the actual and predicted rain fall with respect to temperature during both training and testing. It can be seen that the trend/plot for polynomial regression is similar/closed for both training and testing, whereas for other models, the plot differs significantly for training and testing. Hence, we can say that polynomial regression provides a better modelling of rainfall. Table 6 shows the results obtained via deep learning models. The loss and RMSE values for LSTM are significantly better than those obtained for RNN because LSTM captures long-term dependencies and RNN suffers from vanishing gradient and exploding gradient problems. Therefore, LSTM has performed with better results.

There are certain limitations in this work. The dataset was limited, and we have only tested a few deep learning models. More work can be carried out to employ pre-trained models and transfer learning to improve performance. We have only evaluated this model for specific parameters. Other parameters can also be considered for extensive evaluations.

7. Conclusions and Future Work

The development of a machine-learning-based rain prediction model is the aim of this study. The model used is based on a dataset of 2391 observations that the Bangladesh Meteorological Department (BMD) in Dhaka has gathered. We combined nine machine learning models and two deep learning algorithms for this research. Before the rainfall forecasts were verified, every model was trained using 16 input characteristics. The simulation's performance outcomes were all satisfactory. Future versions of this model will have a system for early warning and include additional parameters like humidity, wind gusts, pressure in the air, atmospheric pressure, sun exposure to radiation, etc. This will enable the early detection of natural disasters and enhance real-time global warming predictions. The difficulty in guessing which models are most appealing or how many days in advance are ideal for making a prediction would be another highly intriguing future study related to the one just mentioned. It is also planned to test the approach for other countries along with state-of-the-art machine learning models. One can also develop a disaster response system once a disaster has happened.

Finally, data from several meteorological departments in various nations or areas may be included in future research. These would offer a broader range of information, capturing various climatic trends and improving the model's suitability. Hydrological model integration, in addition to mathematical flood channel modelling, may provide a better understanding of land–water interactions and increase the accuracy of flood forecasts. In the near future, researchers can look at the prospect of adding additional characteristics that can have a significant impact on flooding, such as soil moisture, changes in land use, or urbanisation measures. Furthermore, temporal patterns in greater detail while taking seasonality and long-term climate change into account could be examined. This could lead to better performance from the LSTM model or indicate the need for further temporal-based neural network topologies. Future researchers should examine how man-made constructions like levees, reservoirs, and dams affect flood forecasting. Predictions in areas with substantial human involvement may become more accurate due to understanding these interventions.

Moreover, historical accounts and first-hand experience from flood-prone communities might offer insightful information. Subsequent studies can use input from the community or personal experiences to improve further prediction algorithms. Further research could

concentrate on creating a complete end-to-end early warning system that combines prediction with communication channels to notify locals in flood-prone areas, building on the predictive skills.

Author Contributions: Conceptualisation, A.R., H.F., N.I. and D.S.; methodology, M.A.E., A.S., M.A. (Muhammad Akram) and M.A. (Mesfer Alrizq); software, A.R., H.F., N.I. and D.S.; validation, M.A.E., A.S., M.A. (Muhammad Akram). and M.A. (Mesfer Alrizq); formal analysis, A.R., H.F., N.I. and D.S.; investigation, M.A.E., A.S., M.A. (Muhammad Akram) and M.A. (Mesfer Alrizq); resources, H.F. and N.I.; data curation, M.A.E. and M.A. (Mesfer Alrizq); writing—original draft preparation, A.R., H.F., N.I. and D.S.; writing—review and editing, M.A.E., A.S., M.A. (Muhammad Akram). and M.A. (Mesfer Alrizq); visualisation, M.A. (Muhammad Akram).; supervision, A.S.; project administration, A.R. and N.I.; funding acquisition, M.A.E. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to acknowledge the support of the Deputy for Research and Innovation, Ministry of Education, Kingdom of Saudi Arabia, for funding this research through a grant (NU/IFC/2/SERC/-/48) under the Institutional Funding Committee at Najran University, Kingdom of Saudi Arabia.

Data Availability Statement: Data available in a publicly accessible repository on Kaggle and can be found at the following link: <https://www.kaggle.com/datasets/emonreza/65-years-of-weather-data-bangladesh-preprocessed>, accessed on 20 October 2023.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Syeed, M.M.A.; Farzana, M.; Namir, I.; Ishrar, I.; Nushra, M.H.; Rahman, T. Flood prediction using machine learning models. In Proceedings of the 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 9–11 June 2022; IEEE: New York, NY, USA, 2022.
2. Kumar, V.; Azamathulla, H.M.; Sharma, K.V.; Mehta, D.J.; Maharaj, K.T. The state of the art in deep learning applications, challenges, and future prospects: A comprehensive review of flood forecasting and management. *Sustainability* **2023**, *15*, 10543. [CrossRef]
3. Gude, V.; Corns, S.; Long, S. Flood prediction and uncertainty estimation using deep learning. *Water* **2020**, *12*, 884. [CrossRef]
4. Nguyen, D.T.; Chen, S.-T. Real-time probabilistic flood forecasting using multiple machine learning methods. *Water* **2020**, *12*, 787. [CrossRef]
5. Furquim, G.; Pessin, G.; Faical, B.S.; Mendiondo, E.M.; Ueyama, J. Improving the accuracy of a flood forecasting model by means of machine learning and chaos theory: A case study involving a real wireless sensor network deployment in brazil. *Neural Comput. Appl.* **2016**, *27*, 1129–1141. [CrossRef]
6. Talukdar, S.; Ghose, B.; Shahfahad; Salam, R.; Mahato, S.; Pham, Q.B.; Linh, N.T.T.; Costache, R.; Avand, M. Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 2277–2300. [CrossRef]
7. Maspo, N.-A.; Bin Harun, A.N.; Goto, M.; Cheros, F.; Haron, N.A.; Nawawi, M.N.M. Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2020.
8. Mitra, P.; Ray, R.; Chatterjee, R.; Basu, R.; Saha, P.; Raha, S.; Barman, R.; Patra, S.; Biswas, S.S.; Saha, S. Flood forecasting using Internet of things and artificial neural networks. In Proceedings of the 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 13–15 October 2016; IEEE: New York, NY, USA, 2016.
9. Noymanee, J.; Nikitin, N.O.; Kalyuzhnaya, A.V. Urban pluvial flood forecasting using open data with machine learning techniques in pattani basin. *Procedia Comput. Sci.* **2017**, *119*, 288–297. [CrossRef]
10. Aswad, F.M.; Kareem, A.N.; Khudhur, A.M.; Khalaf, B.A.; Mostafa, S.A. Tree-based machine learning algorithms in the Internet of Things environment for multivariate flood status prediction. *J. Intell. Syst.* **2021**, *31*, 1–14. [CrossRef]
11. Sankaranarayanan, S.; Prabhakar, M.; Satish, S.; Jain, P.; Ramprasad, A.; Krishnan, A. Flood prediction based on weather parameters using deep learning. *J. Water Clim. Change* **2020**, *11*, 1766–1783. [CrossRef]
12. Wang, G.; Yang, J.; Hu, Y.; Li, J.; Yin, Z. Application of a novel artificial neural network model in flood forecasting. *Environ. Monit. Assess.* **2022**, *194*, 125. [CrossRef]
13. Puttinaovarat, S.; Horkaew, P. Flood forecasting system based on integrated big and crowdsource data by using machine learning techniques. *IEEE Access* **2020**, *8*, 5885–5905. [CrossRef]

14. Ria, N.J.; Ani, J.F.; Islam, M.; Masum, A.K.M. Standardization Of Rainfall Prediction In Bangladesh Using Machine Learning Approach. In Proceedings of the 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 6–8 July 2021; IEEE: New York, NY, USA, 2021.
15. Osmani, S.A.; Kim, J.-S.; Jun, C.; Sumon, W.; Baik, J.; Lee, J. Prediction of monthly dry days with machine learning algorithms: A case study in Northern Bangladesh. *Sci. Rep.* **2022**, *12*, 19717. [CrossRef] [PubMed]
16. Manandhar, A.; Fischer, A.; Bradley, D.J.; Salehin, M.; Islam, M.S.; Hope, R.; Clifton, D.A. Machine learning to evaluate impacts of flood protection in Bangladesh, 1983–2014. *Water* **2020**, *12*, 483. [CrossRef]
17. Aydin, M.C.; Sevgi Birincioğlu, E. Flood risk analysis using gis-based analytical hierarchy process: A case study of Bitlis Province. *Appl. Water Sci.* **2022**, *12*, 122. [CrossRef]
18. Msabi, M.M.; Makonyo, M. Flood susceptibility mapping using GIS and multi-criteria decision analysis: A case of Dodoma region, central Tanzania. *Remote Sens. Appl. Soc. Environ.* **2021**, *21*, 100445. [CrossRef]
19. Shafizadeh-Moghadam, H.; Valavi, R.; Shahabi, H.; Chapi, K.; Shirzadi, A. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. *J. Environ. Manag.* **2018**, *217*, 1–11. [CrossRef]
20. Elmagzoub, M.; Syed, D.; Shaikh, A.; Islam, N.; Alghamdi, A.; Rizwan, S. A survey of swarm intelligence based load balancing techniques in cloud computing environment. *Electronics* **2021**, *10*, 2718. [CrossRef]
21. Al Reshan, M.S.; Syed, D.; Islam, N.; Shaikh, A.; Hamdi, M.; Elmagzoub, M.A.; Muhammad, G.; Talpur, K.H. A Fast Converging and Globally Optimized Approach for Load Balancing in Cloud Computing. *IEEE Access* **2023**, *11*, 11390–11404. [CrossRef]
22. Islam, N.; Raza, E.; Mohsin, S.; Ansari, A.; Shuja, R.; Syed, D. Forecasting on COVID-19 Data Using ARIMAX Model. In *Data Science with Semantic Technologies*; CRC Press: Boca Raton, FL, USA, 2023; pp. 95–113.
23. Islam, N.; Khan, S.K.; Rehman, A.; Aftab, U.; Syed, D. Stock Prediction for ARGAM Companies Dataset. *KIET J. Comput. Inf. Sci.* **2023**, *6*, 1–13. [CrossRef]
24. Bui, D.T.; Pradhan, B.; Nampak, H.; Bui, Q.-T.; Tran, Q.-A.; Nguyen, Q.-P. Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modeling in a high-frequency tropical cyclone area using GIS. *J. Hydrol.* **2016**, *540*, 317–330.
25. Chatterjee, S.; Datta, B.; Sen, S.; Dey, N.; Debnath, N.C. Rainfall prediction using hybrid neural network approach. In Proceedings of the 2018 2nd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom), Ho Chi Minh, Vietnam, 29–31 January 2018; IEEE: New York, NY, USA, 2018.
26. Islam, M.N.; van Amstel, A.; Ghosh, B.K.; Sarker, K.R. Climate Change and Living with Floods: An Empirical Case from the Saghata Union of Gaibandha District, Bangladesh. In *Bangladesh II: Climate Change Impacts, Mitigation and Adaptation in Developing Countries*; Springer: Cham, Switzerland, 2021; pp. 459–478.
27. Luo, T.; Maddocks, A.; Iceland, C.; Ward, P.; Winsemius, H. *World's 15 Countries with the Most People Exposed to River Floods*; World Resources Institute: Washington, DC, USA, 2015.
28. Kumari, S.; Tripathy, K.K.; Kumbhar, V. *Data Science and Analytics*; Emerald Publishing Limited: Bingley, UK, 2020.
29. Thirumalai, C.; Harsha, K.S.; Deepak, M.L.; Krishna, K.C. Heuristic prediction of rainfall using machine learning techniques. In Proceedings of the 2017 International Conference on Trends in Electronics and Informatics (ICEI), Tirunelveli, India, 11–12 May 2017; IEEE: New York, NY, USA, 2017.
30. Adnan, R.; Zain, Z.M.; Ruslan, F.A. 5 hours flood prediction modeling using improved NNARX structure: Case study Kuala Lumpur. In Proceedings of the 2014 IEEE 4th International Conference on System Engineering and Technology (ICSET), Bandung, Indonesia, 24–25 November 2014; IEEE: New York, NY, USA, 2014.
31. Mosavi, A.; Ozturk, P.; Chau, K.-W. Flood prediction using machine learning models: Literature review. *Water* **2018**, *10*, 1536. [CrossRef]
32. Chen, C.; Jiang, J.; Liao, Z.; Zhou, Y.; Wang, H.; Pei, Q. A short-term flood prediction based on spatial deep learning network: A case study for Xi County, China. *J. Hydrol.* **2022**, *607*, 127535. [CrossRef]
33. Motta, M.; de Castro Neto, M.; Sarmiento, P. A mixed approach for urban flood prediction using Machine Learning and GIS. *Int. J. Disaster Risk Reduct.* **2021**, *56*, 102154. [CrossRef]
34. Ghorpade, P.; Gadge, A.; Lende, A.; Chordiya, H.; Gosavi, G.; Mishra, A.; Hooli, B.; Ingle, Y.S.; Shaikh, N. Flood forecasting using machine learning: A review. In Proceedings of the 2021 8th International Conference on Smart Computing and Communications (ICSCC), Kerala, India, 1–3 July 2021; IEEE: New York, NY, USA, 2021.
35. Adnan, M.S.G.; Siam, Z.S.; Kabir, I.; Kabir, Z.; Ahmed, M.R.; Hassan, Q.K.; Rahman, R.M.; Dewan, A. A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction. *J. Environ. Manag.* **2023**, *326*, 116813. [CrossRef] [PubMed]
36. Gauhar, N.; Das, S.; Moury, K.S. Prediction of flood in Bangladesh using K-nearest neighbors algorithm. In Proceedings of the 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Dhaka, Bangladesh, 5–7 January 2021; IEEE: New York, NY, USA, 2021.
37. Han, S.; Coulibaly, P. Bayesian flood forecasting methods: A review. *J. Hydrol.* **2017**, *551*, 340–351. [CrossRef]
38. Hamidul Haque, M.; Sadia, M.; Mustaq, M. Development of Flood Forecasting System for Someshwari-Kangsa Sub-watershed of Bangladesh-India Using Different Machine Learning Techniques. *EGU General Assembly Conference Abstracts*; EGU: Virtual, 2021. Available online: <https://ui.adsabs.harvard.edu/abs/2021EGUGA..2315294H/abstract> (accessed on 20 October 2023).

39. Billah, M.; Adnan, N.; Akhond, M.R.; Ema, R.R.; Hossain, A.; Galib, S.M. Rainfall prediction system for Bangladesh using long short-term memory. *Open Comput. Sci.* **2022**, *12*, 323–331. [\[CrossRef\]](#)
40. Yaseen, M.W.; Awais, M.; Riaz, K.; Rasheed, M.B.; Waqar, M.; Rasheed, S. Artificial Intelligence Based Flood Forecasting for River Hunza at Danyor Station in Pakistan. *Arch. Hydro-Eng. Environ. Mech.* **2022**, *69*, 59–77. [\[CrossRef\]](#)
41. Parmar, A.; Mistree, K.; Sompura, M. Machine learning techniques for rainfall prediction: A review. In Proceedings of the International Conference on Innovations in Information Embedded and Communication Systems, Coimbatore, India, 17–18 March 2017.
42. Khosravi, K.; Panahi, M.; Golkarian, A.; Keesstra, S.D.; Saco, P.M.; Bui, D.T.; Lee, S. Convolutional neural network approach for spatial prediction of flood hazard at national scale of Iran. *J. Hydrol.* **2020**, *591*, 125552. [\[CrossRef\]](#)
43. Kovalchuk, S.V.; Krikunov, A.V.; Knyazkov, K.V.; Boukhanovsky, A.V. Classification issues within ensemble-based simulation: Application to surge floods forecasting. *Stoch. Environ. Res. Risk Assess.* **2017**, *31*, 1183–1197. [\[CrossRef\]](#)
44. Nevo, S.; Morin, E.; Rosenthal, A.G.; Metzger, A.; Barshai, C.; Weitzner, D.; Voloshin, D.; Kratzert, F.; Elidan, G.; Dror, G.; et al. Flood forecasting with machine learning models in an operational framework. *arXiv* **2021**, arXiv:2111.02780. [\[CrossRef\]](#)
45. Khosravi, K.; Shahabi, H.; Pham, B.T.; Adamowski, J.; Shirzadi, A.; Pradhan, B.; Dou, J.; Ly, H.-B.; Gróf, G.; Ho, H.L.; et al. A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *J. Hydrol.* **2019**, *573*, 311–323. [\[CrossRef\]](#)
46. El-Magd, S.A.A.; Pradhan, B.; Alamri, A. Machine learning algorithm for flash flood prediction mapping in Wadi El-Laqeita and surroundings, Central Eastern Desert, Egypt. *Arab. J. Geosci.* **2021**, *14*, 323. [\[CrossRef\]](#)
47. Nayak, M.; Das, S.; Senapati, M.R. Improving Flood Prediction with Deep Learning Methods. *J. Inst. Eng. Ser. B* **2022**, *103*, 1189–1205. [\[CrossRef\]](#)
48. Tayfur, G.; Singh, V.P.; Moramarco, T.; Barbetta, S. Flood hydrograph prediction using machine learning methods. *Water* **2018**, *10*, 968. [\[CrossRef\]](#)
49. Sahoo, A.; Samantaray, S.; Ghose, D.K. Prediction of flood in Barak River using hybrid machine learning approaches: A case study. *J. Geol. Soc. India* **2021**, *97*, 186–198. [\[CrossRef\]](#)
50. Qian, K.; Mohamed, A.; Claudel, C. Physics informed data driven model for flood prediction: Application of deep learning in prediction of urban flood development. *arXiv* **2019**, arXiv:1908.10312.
51. Miaou, S.; Hung, W.-H. River flooding forecasting and anomaly detection based on deep learning. *IEEE Access* **2020**, *8*, 198384–198402. [\[CrossRef\]](#)
52. Hossain, I.; Rasel, H.M.; Alam Imteaz, M.; Mekanik, F. Long-term seasonal rainfall forecasting using linear and non-linear modelling approaches: A case study for Western Australia. *Meteorol. Atmos. Phys.* **2020**, *132*, 131–141. [\[CrossRef\]](#)
53. Ighile, E.H.; Shirakawa, H.; Tanikawa, H. Application of GIS and machine learning to predict flood areas in Nigeria. *Sustainability* **2022**, *14*, 5039. [\[CrossRef\]](#)
54. Kunverji, K.; Shah, K.; Shah, N. A flood prediction system developed using various machine learning algorithms. In Proceedings of the 4th International Conference on Advances in Science & Technology (ICAST2021), Mumbai, India, 7 May 2021.
55. Dodangeh, E.; Choubin, B.; Eigdir, A.N.; Nabipour, N.; Panahi, M.; Shamshirband, S.; Mosavi, A. Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. *Sci. Total Environ.* **2020**, *705*, 135983. [\[CrossRef\]](#)
56. Khairudin, N.M.; Mustapha, N.O.; Aris, T.N.; Zolkepli, M.A. A study to investigate the effect of different time-series scales towards flood forecasting using machine learning. *J. Theor. Appl. Inform. Technol.* **2021**, *99*, 5687–5699.
57. Dtissibe, F.Y.; Ari, A.A.A.; Titouna, C.; Thiare, O.; Gueroui, A.M. Flood forecasting based on an artificial neural network scheme. *Nat. Hazards* **2020**, *104*, 1211–1237. [\[CrossRef\]](#)
58. Sarasa-Cabezuelo, A. Prediction of rainfall in Australia using machine learning. *Information* **2022**, *13*, 163. [\[CrossRef\]](#)
59. Liyew, C.M.; Melese, H.A. Machine learning techniques to predict daily rainfall amount. *J. Big Data* **2021**, *8*, 153. [\[CrossRef\]](#)
60. Singh, P. Indian summer monsoon rainfall (ISMR) forecasting using time series data: A fuzzy-entropy-neuro based expert system. *Geosci. Front.* **2018**, *9*, 1243–1257. [\[CrossRef\]](#)
61. Mishra, N.; Soni, H.K.; Sharma, S.; Upadhyay, A.K. Development and analysis of artificial neural network models for rainfall prediction by using time-series data. *Int. J. Intell. Syst. Appl.* **2018**, *12*, 16. [\[CrossRef\]](#)
62. Chitwatkulsiri, D.; Miyamoto, H. Real-Time Urban Flood Forecasting Systems for Southeast Asia—A Review of Present Modelling and Its Future Prospects. *Water* **2023**, *15*, 178. [\[CrossRef\]](#)
63. Kumar, V.; Sharma, K.V.; Caloiero, T.; Mehta, D.J.; Singh, K. Comprehensive overview of flood modeling approaches: A review of recent advances. *Hydrology* **2023**, *10*, 141. [\[CrossRef\]](#)
64. Mosaffa, H.; Sadeghi, M.; Mallakpour, I.; Jahromi, M.N.; Pourghasemi, H.R. Application of Machine Learning Algorithms in Hydrology. In *Computers in Earth and Environmental Sciences*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 585–591.
65. Jehanzaib, M.; Ajmal, M.; Achite, M.; Kim, T.-W. Comprehensive review: Advancements in rainfall-runoff modelling for flood mitigation. *Climate* **2022**, *10*, 147. [\[CrossRef\]](#)
66. Mistry, S.; Parekh, F. Flood Forecasting Using Artificial Neural Network. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2022.
67. Xu, Y.; Hu, C.; Wu, Q.; Jian, S.; Li, Z.; Chen, Y.; Zhang, G.; Zhang, Z.; Wang, S. Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation. *J. Hydrol.* **2022**, *608*, 127553. [\[CrossRef\]](#)

68. Cho, M.; Kim, C.; Jung, K.; Jung, H. Water level prediction model applying a long short-term memory (lstm)-gated recurrent unit (gru) method for flood prediction. *Water* **2022**, *14*, 2221. [[CrossRef](#)]
69. Qadeer, K.; Rehman, W.U.; Sheri, A.M.; Park, I.; Kim, H.K.; Jeon, M. A long short-term memory (LSTM) network for hourly estimation of PM_{2.5} concentration in two cities of South Korea. *Appl. Sci.* **2020**, *10*, 3984. [[CrossRef](#)]
70. Available online: <https://www.kaggle.com/datasets/emonreza/65-years-of-weather-data-bangladesh-preprocessed> (accessed on 20 October 2023).
71. Wong, T.-T.; Yeh, P.-Y. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans. Knowl. Data Eng.* **2019**, *32*, 1586–1594. [[CrossRef](#)]
72. Rahman, M.; Chen, N.; Elbeltagi, A.; Islam, M.M.; Alam, M.; Pourghasemi, H.R.; Tao, W.; Zhang, J.; Shufeng, T.; Faiz, H.; et al. Application of stacking hybrid machine learning algorithms in delineating multi-type flooding in Bangladesh. *J. Environ. Manag.* **2021**, *295*, 113086. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.