A Comparative Study of Deep Learning and Modern Machine Learning Methods for Predicting Australia's Precipitation

Hira Farman[1], Qurat-ul-ain Mastoi[2], Qaiser Abbas[3],
Saad Ahmad[4], Abdulaziz Alshahrani[5], Salman Jan[6], Toqeer Ali Syed[7]
Department of Computer Science, Iqra University, Pakistan[1]
Department of Computer Science, Iqra University,
Karachi Institute of Economics and Technology, Karachi, 74600, Pakistan[1]
School of Computing and Creative Technologies University of the West of England Bristol, United Kingdom[2]
Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah, 42351, Saudi Arabia[3]
Faculty of Computer Science, IQRA University Karachi, Pakistan[4]
Islamic University of Madinah, CO 42351 Madinah, Saudi Arabia[5, 7]
Faculty of Computer Studies, Arab Open University, 512, A'ali Kingdom of Bahrain[6]

***Abstract*—Floods are chaotic weather patterns that cause irreversible and devastating harm to people's lives, crops, and the socioeconomic system. It causes extensive property damage, animal mortality, and even human fatalities. To mitigate the risk of flooding, it is imperative to create an early warning system that can accurately forecast the amount of rain that will fall tomorrow. Rainfall forecasting is essential to the lives of people and is absolutely important everywhere in the world. The rainfall prediction model reduces risk and helps to prevent further human deaths. Statistics cannot reliably forecast rainfall since the atmosphere is dynamic. Due to the preceding factors, this study uses machine learning and deep learning techniques to estimate precipitation. The purpose of this study is to develop and evaluate a prediction model for forecasting rainfall of 5 cities of Australia (Darwin, Sydney, Perth airport, Melbourne, Brisbane). The Dataset was gathered from the national meteorological organization of Australia is the Australian Government Bureau of Meteorology, also known as the BOM. To monitor and forecast meteorological conditions, climatic trends, and natural calamities like cyclones, storms, floods, the Bureau of Meteorology is essential. The dataset includes 14, 5460 size, 23 features detailed city-specific monthly averages for Australia from 2008 to 2017(10 years). An effective rainfall forecasting was produced by integration of a number of Machine Learning and Deep Learning techniques, including Random Forest model (RF), Decision Tree (DT) and Gradient Boosting classifier (GBC), Artificial Neural Network (ANN), and Recurrent Neural Network (RNN). The models were trained to forecast rainfall, reducing the potential impact of floods. Results indicate that combining neural networks and Random Forests provides the most accurate predictions.**

Keywords*—Machine learning; rainfall prediction; neural network; Random Forest; deep learning

I. Introduction

Natural disasters including floods, hurricanes, and earthquakes are becoming more frequent and intense due to environmental changes such as deforestation, urbanization, and climate change. Floods pose serious threats to lives, agriculture, and economies, usually it results from heavy rainfall and poor drainage system in the some of the regions. Given the growing impact of extreme weather, this study compares three LSTM-based neural network architectures to identify the most suitable model for forecasting hourly rainfall volumes. The leading purpose of such study is to use deep learning algorithms in developing flood prediction models based on meteorological data [1]–[3]. Publics and governments may be enabled to both prevent the flood occurrence with immediate and long term actions and prepare for evacuation and rescue operations with the help of the early warning in advance.The studies were conducted using data from crowdsourcing, geospatial, hydrological, and meteorological sources. In this study [4]–[6], successfully machine learning is used to build a rain forecast model. This work [7] focuses on the machine learning (ML) methods of prediction for six selected stations per semi-annual cycle in Bangladesh in order to create a new achieving format of the monthly dry days (MDD). Through the use of machine learning approaches, the study analyzes the unanticipated effects of flood protection in Bangladesh [8].The main contribution of this research [9] is to determine the state-of-the-art machine learning methods for flood prediction together with the significant parameters that were fed into the model. This will make it possible for flood management and/or researchers to compare the prediction findings to one another when assessing machine learning techniques for early flood forecasting. The study investigated the possibility of developing a probabilistic forecasting model through the employment of various machine learning techniques, including the k nearest neighbors (also called KNN)method, the fuzzy inferential model (FIM), and the support vector regression methodology (SVR). Modeling of flood conditions is a difficult task that requires an in-depth analysis of the situations that affect flooding. This study proposes an Internet of Things based flood state prediction (IOT-FSP) model to assist river flooding conditions forecast [10]–[18].

This study explores the use of machine learning (ML) and deep learning (DL) techniques to improve rainfall forecast accuracy. Grey Relation Analysis (GRA) is used to identify influential factors, while Support Vector Machines (SVM) and

Artificial Neural Networks (ANN) are applied for prediction. Specifically, a model combining Feed Forward Neural Networks, the Levenberg-Marquardt algorithm, and backpropagation is used to forecast monthly and bi-monthly rainfall in Northern India. Performance is evaluated using MSE, MRE, and regression analysis. The research not only aims to enhance prediction but also supports disaster preparedness and response planning.

A. *Purpose of the Study / Research Questions*

The following are the research primary goals:

- Feature cleaning and visual representation for datasets features
- Univariate and multivariate analysis of datasets with regard to several features.
- Several models of machine learning have been experimented and have shown their accuracy in forecasting Rainfall.
- Validating deep learning's efficacy in precipitation prediction.
- Comparative analysis for different algorithms for accuracy and error for Rainfall prediction.
- In order to classify the objects with extreme speed and accuracy, the architecture of the LSTM-Deep learning network was developed.
- An extended experiment was conducted to provide a detailed examination of the proposed model.

Rainfall prediction in Australia is challenging due to regional differences, monsoon dynamics, and climate change. This study investigates five ML and DL models to improve forecasting accuracy by identifying key atmospheric variables and evaluating seasonal rainfall severity over the past decade. The study aims to (1) support early warning systems with relevant statistics, (2) provide metrics for policymaking related to rainfall management, soil degradation, and drought, and (3) enhance climate models by deepening the understanding of factors influencing weather patterns.

B. *Justification for Model Selection*

In this study, we selected advanced machine learning (ML) and deep learning (DL) models such as LSTM, CNN, and SVR due to their proven ability to capture nonlinear patterns in time-series data. LSTM networks, in particular, are well-suited for sequential data such as rainfall records because they can retain long-term dependencies, which are essential for understanding delayed rainfall effects due to climate shifts. CNNs, though originally developed for spatial features, have recently shown promise in extracting local patterns in time-series signals, making them a good fit for rainfall fluctuations. Meanwhile, SVR provides a robust baseline with good generalization ability, especially in high-dimensional meteorological datasets.

Traditional statistical models, such as ARIMA or linear regression, often assume linearity and stationarity, which are major limitations when applied to rainfall data that exhibit strong temporal variability and noise. These models also struggle with multi-feature dependencies and are less adaptive to changing climate patterns. Furthermore, simpler ML models like decision trees may lack the depth required to model complex seasonal transitions or capture hidden trends across time steps, leading to reduced prediction accuracy.

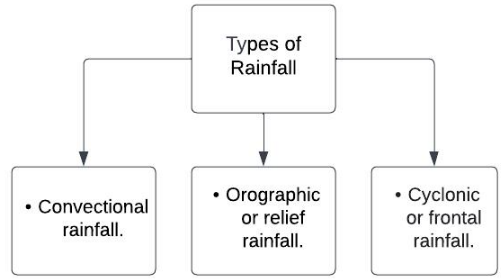

Fig. 1. Types of rainfall.

The other part of the research is organized as follows: On the other side, the theoretical context, comparative analysis, the data set and Australia Location utilized in study, and experimental methods are all presented. Following the Machine learning and Deep learning algorithms designing, which are the main part of the early rainfall prediction of the dataset based on accuracy and other measures like F1 score, accuracy, and recall, is the Results and discussion section of work, the Last section Summary and future research is a conclusion part.

II. Related Work

A. *Types of Rainfall*

Rainfall can be classified into various types based on different criteria such as duration, intensity, spatial distribution, and the mechanisms responsible for its occurrence. Here are some common types of rainfall in Fig. 1.

1) Convectional rainfall: As the Earth's surface warms, air rises, resulting in convectional rainfall. As it increases, the air cools and releases moisture as rain. In regions with high temperatures and high humidity, convectional rainfall is frequent, frequently occurring in tropical climates in the late afternoon or evening [19], [20].

2) Orographic rainfall: Orographic rainfall is the result of moist air being lifted as it is pushed over high ground, such as mountains or hills. On the mountain's windward side, moisture and precipitation result from the rising, cooling air. Due to descending dry air, the leeward side gets a rain shadow effect, which reduces rainfall [21].

3) Cyclonic rainfall: Rainfall associated with cyclones or low-pressure systems is referred to as cyclonic rainfall. It is characterized by persistent, widespread rain that is frequently accompanied by high winds. Forecasting weather, preparing for disasters, and managing water resources all depend on an understanding of cyclonic rainfall patterns [20].

B. *Effects of Rainfall*

The understanding of rainfall effects is of great importance for several businesses, p.e. agriculture, water resources management, disaster preparedness, and climate research. For sustainable growth and risk reduction related to extreme weather events, monitoring rainfall patterns and managing its effects are essential. Here are some notable effects of rainfall, Water Supply [22], Agriculture, Moisture in the Soil and The process of erosion, Flood [4], [9], [10], [23]–[26], Hydroelectric Power Generation, Weather and Climate Patterns [2], [27].

C. Motivation and Impact of Flood Events

Existing flood prediction models rely on complex statistical computations that are either very costly financially and computationally or aren't applicable to applications in the future. To circumvent these issues, approaches that mix time-series data with machine learning algorithms are being researched. For various downstream applications, it is necessary to compare the efficiency of deep learning architectures and conventional machine learning methods.

Here the rainfall is estimated by using a simplified model [28]. It is Southern India, Kerala that faced the flood of the century. Often, it also means a huge loss of peoples' lives and properties. This consequently prompted us to investigate the variations in the rainfall pattern of Kerala. In Bangladesh, floods kill people, destroy household properties, crops and means of livelihood of the masses constantly. Flooding is caused by lakes, rivers and other water bodies discharging water and incorporating adjacent land into the flooded region. Every year, flooding becomes the worst enemy for more than 4. In India have the 3. 84 crores people. 84 million in the case of Bangladesh and the rest in 3. 29 million in China has realized that [1].The recent event which took place in Kerala in August 2018 is one of the most remarkable incidents followed by massive floods in India , which is currently considered as one of the most affected nations with respect to catastrophic floods across the globe.

D. Machine Learning Applications in Flood Forecasting

Much has been said in past times on Flood probability, which can be done through IoT and ML based on rainfall, humidity, water temperature, water velocity, and other variables. There are some limitations in the research since we have not tried to establish the probability range for flood based on the levels of temperature and rain intensity. Scientific communities all over the globe are interested in developing flood forecasting methods. Thus, there is a demand for flood forecasting that must be high-quality and reliable so that those living near the flooded areas can have the warning signs that they need to evacuate. Thus, a 5-hours flood simulation of rainfall area for Kuala Lumpur city in this study was provided using neural network's Autoregressive Structure with Extended Input (NNARX) and its Enhanced Modeling. This work [29] combines ML technology with established ones to generate models that have more prediction yield than other existing models. The main value of this study is in the fact that it highlights the current trend of using ML models for flood prediction as well as giving suggestions of the best kinds of models to use. This analysis is mainly focused on the empirical research of ML approaches in the area where the models have been evaluated for robustness, accuracy, utility, and speed.

E. Model Evaluation, Datasets, and Techniques

To make the research more detailed and accurate, the area of interest is divided into grids delimited by latitude and longitude, with precipitation and flow readings merged into vector organizations based on coordinates. The input property consists of a two-dimensional timeline containing spatial information instead of a one-dimensional timeline. The first step which is the focus to extracting spatial and temporal components of hydrological information is comprised of.In particular the author introduces the convolutional LSTM (Con LSTM), which combines the Convolutional Neural Network (CNN) and Long Short Term Memory Network (LSMN) to boost performance. The aim of this work is to design a flood prediction system which can be used as an advantageous instrument for urban administration and resilience management through an integration of machine learning algorithms and GIS techniques [30]. This method will lead to development of long-term strategic policies for the growth of smart cities that will be workable and suitable flood indicators at the level of the municipality. At the Random Forest algorithm with the accuracy 0% . 0.96, and a Pearson's (or linear) coefficient of 0. 77 was an excellent choice as the input layer or the hidden layer of machine learning as it can identify errors. This requires summarizing different past research studies that examine the ML techniques for flood forecasting and the characteristics which are used for the forecast. Goal of the research is to find flood forecasting approaches which mainly involve ML techniques and to determine flood prediction parameters that have been used as input parameters to the models for forecasted flooding in order to achieve that. The most valuable aspect of this work is the list of critical variables and recent ML methods employed for flood prediction. They may be able to then both carry out short- and long-term preventive measures, be proactive and rescue people and provide relief for flood victims, with the help of an early warning of a flood disaster [31]. For example, one of the main factors in most flood management is the geographic location of the affected areas and their respective severity.

F. Recent Developments and Future Directions

Floods cannot yet be reliably predicted in advance with the help of any approach. Data that was prepared and manually entered was typically used by earlier technologies. It was not possible to make early and real-time estimates due to the lengthy processes. In order to forecast flood events in particular regions and time periods, this research proposes a novel approach to flood forecasting that integrates meteorological, hydrological, spatial data, and big data from crowdsourcing sources. System that takes into account both historical data and climatic conditions produced the most precise estimation by the setup using an MLP ANN for the Correct proportions, Kappa, MAE, and RMSE 97%. 0. 89, 0. 93, 0., and 0. 10, respectively. In this [32] research cutting-edge operating methods have been investigated. The current move towards data-driven strategies for flood prediction is something that the writers see and discuss. Forecasting tasks are becoming more and more relevant for machine learning-based models that were trained using historical data for climatic parameters.

The main objective of this work is to demonstrate recent advances in machine learning-based flood forecasting. To develop their conclusions, the authors looked at various widely used flood prediction techniques that different specialists might use. In study [23] create a probabilistic forecasting model, this study used a variety of machine learning techniques, such as k nearest neighbors (KNN), fuzzy inference models (FIM), and support vector regression (SVR).In order to lessen the harm caused by flooding, an examination of the utilization of information gleaned from urban rivers to forecast floods is done in this work [24]. The Artificial Neural Networks were examined to determine their level of accuracy in the forecasting models after the immersion theorem had proven the interdependence of the data. Whenever there have been significant flooding-related difficulties, WSNs have been installed. The study's methodology [33] may help improve the early warning systems that are in place now and generate risk-based development plans. Uncertainties in machine learning-based geospatial algorithms for flood prediction are resolved using a unique method. This paper proposes a method for decreasing regional disparity along with four distinct and hybridized ML based flood susceptibility models (the FSMs). In order to forecast and identify the flooding sites or flood sensitive zones in the Teesta River a basin, this study [4] applied cutting-edge revolutionary ensemble machine learning algorithms. The purpose of the work being done is to construct a rain prediction model using the successful machine learning Random Forest [5]. The goal of this study [6] is to reduce the significant hazards associated with this natural disaster while also making recommendations for policy. To generate an accurate forecast, this study will make use of a Decision Tree classification technique, K Nearest neighbor(KNN), a Support Vector Classifier(SVC), and Binary Logistic Regression(BLR).The findings will be compared to identify the model with the highest level of accuracy. In this study [34], the author used the k-nearest neighbor's algorithm to predict a flood using various correlation coefficients for feature selection. It is well-known that estimating flood risk and making informed decisions [35] depend greatly on quantifying and reducing the uncertainty related to the hydrologic forecast. This article provides a thorough analysis of Bayesian forecasting techniques used in flood forecasting. In this research, 180 independent models based on five diverse machine learning algorithms that include the exponential back propagation neural network('EBPNN'), multilayer perceptron(MLP), support vector regression(SVR), Decision Tree regression(DT Regression) and extreme gradient boosting(XG-Boost), were developed. The "Someshwari Kangsa" sub-watershed of the Bangladeshi arterial north-central hydrological zone containing an area of 5772 square kilometers was used by models. Indeed, it is a difficult task to make a forecast on the upcoming rain by means of classic machine learning algorithms. Besides that, various attempts have been also made by employing different computer techniques to forecast rainfall. To build the long term memory rainfall module of Bangladesh, this article applies the technique of the feedforward architecture driven long short term memory (LSTM) networks that gets rid of the chaos related problems of the different methods. In Bangladesh, this work proposes a novel approach to forecasting monthly dry days (MDD) at six specified recourse stations and then examining the outcomes of the models. The MDD and MWD datasets in terms of monthly dry and wet days, respectively, were done using different rainfall thresholds. This research, although suggested by simply posing the question of 'what will be the consequences of flood mitigation methods in Bangladesh in the long-run' [8] is highly recommended. Data from the historical events (represented by the emigrations and mortality rates) and economic surveys accumulated from 1983 to 2014. The primary objective in this study [36] is the significance of being responsible for and taking care of disasters which humans bring on themselves. Technology that is now in use, software Artificial (AI) can be used for these tasks. Review work and comparison of different approaches and algorithms used by researchers to estimate rainfall are presented in tabular form [37]. Making methods and procedures used in rainfall forecasting understandable to non-experts is the aim of this endeavor. One notable illustration of how India is currently among the nations in the world that have experienced the most severe flooding is the most recent disaster that occurred in Kerala in August 2018. Much work has been done in the past to use Internet of Things, or IoT, and ML (machine learning) approaches to assess the likelihood of flooding based on rainfall, humidity, water temperature, water velocity, and other characteristics. The prime output of this study is to review the ML models used for flood forecasting as current models and give advice on how to choose the best models. India [27] is in more danger for floods that cause grave losses now; last August extreme floods happened in Kerala and it illustrates the disasters from this catastrophe. The problem is that the flood frequency model is based solely on historical data from the past – number of rainfall events and their water temperature. The neural network has been employed that is Deep Learning to calculate probabilities of flooding considering temperature and rainfall intensity. Consequently, with the flooding having taken over as one of the most well-known subjects of research in hydrology, flood prediction has become one of the principal areas of focus of hydrologists. The issue has been addressed by a lot of researchers with different methods, starting from image processing to physical models, still, are not at the level that can accommodate all applications due to imprecision and insufficient time steps. It studies deep learning methods in gauge height prediction, and also the error in gauge height prediction is assessed. For constructing and verifying the model, the measured height data from Valley Park, Missouri's Meramec River was implemented. According to an analysis based on previous research articles, this study looks at whether the present machine learning (ML) algorithms for flood forecasting are effective given the variables that are used to predict floods. The manifold model is represented here [38] as a machine learning substitute to hydraulic modeling of flood waters. All model achieves performance standards that are tuned to operational use, provided historical data has been used as a benchmark. The article proposes a system that predicts possible floods in a river basin using machine learning and the Internet of Things (IoT) for the research [25].The model connects the Wireless Sensor Network, or WSN, to a personalized mesh network via a ZigBee connection, and it then uses a GPRS module to transmit data over the internet.For predicting the occurrence floods occurrences in the Pattani River, this work [26] examines applying possible machine learning algorithms using open data. This study [23] employed a combination of machine learning tools, such as support vector regression (SVR), fuzzy inference model (FIM), and the k-nearest neighbors (k-NN) method in order to develop a probabilistic forecasting model. This study takes a probabilistic forecasting model with the help of a number of machine learning approaches, including SVR model (a support vector regression model), fuzzy inferential models (FIM), and the KNN technique (the K nearest neighbors technique). In total, three multi criteria decision making evaluation approaches, namely VIKOR, SAW, and TOPSIS, along with two machine learning methods, NB and NBT, were utilized to assess their ability to simulate flood vulnerability in the Ningdu Catchment[Complex Flow of Words] which is one of China's most flood-prone regions [39]. Two techniques, Extensive gradient boosting This experiment is based on the Deep Belief Network (DBN) for forecasting the Daya and Bhargavi river banks, which flow towards the Indian state of Odisha. A comparison study that is based on other machine learning methods helps to demonstrate the beneficial impacts of dams in detail. This study [40] focuses on the application of ant colony optimization (ACO), Genetic algorithms (GA), artificial neural networks (ANN), and Particle swarm optimization (PSO) approaches to flood hydrograph prediction. In the current study [41], the relative accuracy of the RB FNN, SVM, and Firefly Algorithm (FA) models compared to the regular ANN, RB FNN, and SVM algorithms for river flood discharge forecasting in the Barak River was examined. Urban flooding is becoming increasingly common [42], which is detrimental to both the economy and quality of life for people. However, the existing flood prediction algorithms have grown too primitive or insufficient to accurately capture the details of flood evolution.This study uses deep neural networks to quicken the computation of a physics-based 2D urban flood forecasting approach that uses the Shallow Water Equation (SWE). Using data modeled by the use of a partial differential equation (PDE) solver, convolution neural networks (CNN) and generative adversarial networks with conditions (CGANs) are utilized to identify flood dynamics. The four ML-based FSMs Fandom Forest (RF), K nearest neighbor (KNN), multilayer perceptron (MLP), and hybridized genetic algorithm–Gaussian radial basis function–support vector regression (GA;RBF;SVR) shown in this article [43]—present a framework for reducing spatial disagreement. The outcomes of those four models were combined to generate an enhanced model as well. The approach presented in this study may be useful in developing risk based development plans and enhancing current early warning systems. This paper [44] utilized a deep learning-based model to predict the water level flood phenomenon of a river in Taiwan. The experimental results showed that the Conv GRU neural network model performed better than other current methods. The trial's outcomes showed that the suggested method could correctly identify the wrong water levels. The purpose of this study project [45] is to use artificially intelligent neural network (ANN) modeling techniques and Lavenberg Marquardt (MLR) multiple linear regression to determine long-term seasonal rainfall patterns in Western Australia.This study [46] demonstrates how RBFs, which are both linear and nonlinear kernel functions, can produce superior results in the same catchment under various conditions. Lighter rainfalls would provide quite different responses from bigger ones, which is a highly helpful technique to disclose the behavior of an SVM model. The study also demonstrates an unexpected result in the SVM reaction to various rainstorm inputs. The process of predicting flood status is difficult [10] and necessitates thorough investigation of the causes of flooding. In order to make it easier to foresee the situation with rivers flooding, this research suggests an Internet of Things-based flood status prediction (IOT-FSP) model. The IoT-FSP model utilizes the Internet of Things architecture to facilitate the collection of flood data as well as algorithms for machine learning (ML) for flood prediction: Decision Tree (DT),Random Forest (RF).This research predicted the flood prone areas in Nigeria using historical flood records [11] from 1985 to 2020 and a number of variables that were confounding. Both logistic regression (LR) and ANN (artificial neural network) algorithms were trained and evaluated to determine the relationship between flood occurrence and the fifteen (15) explanatory factors, which include topographic, meteorological land utilization, and proximity information. This resulted in the creation of a flood susceptibility map. This study [12] explores how several machine learning algorithms can be used to create the most accurate flood determining model. This work proposed three novel machine learning models: the multivariate adaptable regressed splines (MARS), the boosted regression model (the BTR), and the generalized additive model (the GAM) [13]. The province of Ardabil, one of the lands near the Caspian Sea coast, which is regularly affected by flooding, was chosen for applying the methodology that the study referred to. The objective of this study is to figure out how the rainfall Figure time-series data from eight stations along the Kelantan River and the corresponding discharge values influence water level accuracy at Kuala Krai downstream [14]. Another approach of pre-processing involves using of Data Unpredictability and Mutual Information (MI) to recover necessary information to be used as attributes for the forecast model. In this study, the author developed an early flood warning model by using the input and the output layers of multilayer perceptron with stream specification to forecast incoming water level. This research aims to discuss the incorporation of machine learning to trace rain patterns [16]. To do this, a collection of data representing the estimates of rainfall seen in Australia's major cities over the past ten years was subjected to the four main machine learning techniques: K-nearest neighbors (KNN), Decision Tree (DT), Random Forest (RF), and Neural Networks (NN). This study [17] represents the accuracy of rainfall forecasting models engaged by modern machine learning algorithms in forecasting rainfall volume of hours using weather series time information from UK cities. The results clearly indicate that neural nets perform best. This work [18] examines the flood hazard analysis in the Turkish province of Bitlis using the analytical hierarchy approach, a multi-parameter modeling tool.

The main goal of this [47] is to estimate rainfall by utilizing machine learning and deep learning techniques to identify trends in historical meteorological data. The results of this study showed that long short-term memory (LSTM), polynomial regression, and Random Forest regression performed at the greatest levels. The R2 values for polynomial regression and Random Forest are 0.76 and 0.09, respectively, whereas LSTM has a loss value of 0.09. The three different algorithms seem data mining approaches that are often employed in weather prediction; they are successful and have a solid theoretical basis in the computing model for hourly forecasting of rainfall [48]. The author of this study [49] employed three algorithms to forecast rain, using ROC curves, Brier scores, and confusion matrices as validation parameters. The input data is a ten-year panoramic set comprising 3528 datasets and 8 features from the Kemayoran Meteorological Station in Jakarta (96745).The Indian Meteorological Department in Pune contributed data from many meteorological stations in

North India, which were used in this study [50]to analyzes rainfall records spanning 141 years. Using monthly rainfall data, the Artificial Neural Network (ANN) method has been used to create forecasting models for rainfall prediction one to two months in advance. These models make use of the Levenberg-Marquardt training function and the Feed Forward Neural Network (FFNN) with Back Propagation approach. Regression analysis, Mean Square Error (MSE), and Magnitude of Relative Error (MRE) have all been used to evaluate the performance of both models. .The study [51] probably explores the methods used in the rainfall forecasting model, outlining the fundamental ideas and workings of SVR, regression, and the hybrid SVR-PSO methodology. The SVR model's incorporation of Particle Swarm Optimization offers an optimization method to enhance the prediction power. For this particular meteorological application, the comparison might shed light on the possible benefits of adopting SVR and the SVR-PSO hybrid over conventional regression models. According to studies, the ANOVA RBF Kernel offers the best forecasting accuracy with the minimum RMSE value, making it an excellent kernel to employ with the SVR-PSO approach for rainfall forecasting. This study [52] uses ensemble models, optimized artificial neural network models, and large climate indicators to predict rainfall. Accordingly, the new MLP and RBF NN models, as well as the novel hybrid GT and ensemble, were the primary innovations of this study. Not only can the ensemble models of the current work be utilized to predict rainfall, but they can also be employed to predict other meteorological data. Also examined was the uncertainty of the input data and model parameters. A hybrid gamma test was used to pick the inputs, which is a novel approach to input selection (GT).To develop a new test for selecting the optimal input situation, the GT was combined with the NMR method. The hybrid approaches [53] utilizing ACO and three different neural network architectures are presented in this study. ACO+ Feed-Forward back propagation, ACO+cascade-Forward back propagation, and ACO+ Pattern Recognition NN Classifier were the hybrid methods that were put out. The ACO Method and Neural Network are combined to create the techniques. Results of a comparison of the performance of the suggested and current models were given. It has been discovered that the suggested techniques outperform the current Feed-Forward, cascade-Forward, and Pattern Recognition NN Classifiers in terms of performance. This study [54] proposes an improved technique for creating daily short-term and monthly long-term ensemble weather forecasting models for rainfall predictions. This is achieved by combining five rainfall prediction models (Naïve Bayes, C4.5, neural network, support vector machine, and Random Forest) using three linear algebraic combinations: maximum probability, average probability, and majority vote. Using the Malaysian state of Selangor, daily weather data over a six-month period (2010–2015) yielded 1581 occurrences, which were categorized into two groups. There are two classes of rainfall: "active rainfall," which has 428 instances, and "no rainfall," which has the remaining instances. This initiative [55] aims to use feature selection and machine learning approaches to create the most accurate rainfall forecast model possible. Prior to and following feature selection, the Artificial Neural Network (ANN) attains a maximum accuracy of 90% and 91%, respectively.This research in [19] primary contribution is to identify the most recent machine learning techniques for flood prediction as well as the noteworthy parameters that were used as model input. This will allow scientists and/or flood managers to use the prediction results as a reference when evaluating ML methods for early flood prediction.

G. Comparative Analysis

This research is intended to provide the basic framework for using machine learning and deep learning algorithms for rainstorm forecasting. To illustrate an instance, a dataset for rainfall indicators, weather information and related variables from capital cities of Australia in the last ten years is given. Table I represents the sum up of the benchmarking machine learning algorithms, strategy and input parameters that have been used in different rainfalls and floods predicting events.

H. Discussion on Past Studies

Over time, the scientific community has paid close attention to rainfall prediction due to its complexity. Previous research on rainfall prediction has used a variety of approaches, from complex machine learning algorithms to statistical models. Here's a summary of some important findings from earlier research on rainfall prediction. In the past, researchers have used a variety of data sources, such as satellite imaging, climate models, and meteorological measurements, to forecast rainfall. Features including wind speed, humidity, temperature, the land, atmospheric pressure, and oceanic conditions are often extracted from these data sources. Rainfall data's temporal and spatial characteristics must be considered in order to fully represent its complex patterns.. Because machine learning approaches can capture nonlinear correlations and manage enormous datasets, they have become popular for rainfall prediction. Popular Deep learning model Neural networks, and supervised machine learning algorithm Decision Tree (DT), support vector machines(SVM), and Random Forests(RF) are some of the methods that easily capture non linearity from data and are utilized for rainfall prediction. Rainfall prediction is inherently uncertain due to the chaotic nature of atmospheric processes and the influence of various factors such as climate change, El Niño-Southern Oscillation (ENSO), and local topography Deep learning models LSTM and ANN are presently the main techniques for rainfall forecasting, with a focus on machine learning. Using meteorological radar data, this study [57] used LSTM networks to predict short-term rainfall in mountainous areas. The results showed promise in terms of lead time and prediction accuracy. In contrast to conventional methods, this study [58] showed how feed forward neural networks (FNNs) can be used to capture complicated rainfall patterns and improve forecast accuracy when used for rainfall prediction in arid environments. Regardless many difficulties, supervised machine learning has several potential applications in rainfall prediction hourly, seasonally, daily, monthly. To fully achieve the potential of machine learning, ongoing research, data collection, stakeholder involvement, and the integration of ML with conventional modeling techniques will be required. This research applies a combination of pre-trained convolutional neural networks and long short-term memory networks to predict rainfall. Along with the achievements and discoveries described above, there are also a lot of other innovations in terms of implementation of deep learning or DL and machine learning or ML towards Rainfall forecasting. However, the prior research had a number of limitations and

TABLE I. A COMPARISON OF WORKS OF LITERATURE THAT MAKE USE OF RAINFALL OR FORECASTS OF THE WEATHER

Ref	Country	Region	Dataset Description	Algorithm	Accuracy	Best Accuracy	Prediction Type
[28]	UK	Bath, Bristol, Cardiff, Newport, Swindon	Past data from the UK cities Open Weather 5 dataset from Jan 2000 to Apr 2020	Xg Boost Auto ML, LSTM-Network	Loss: 0.0014–0.0001 RMSE: 0.037 MAE: 0.009 RMSLE: 0.0072–0.0015	Stacked-LSTM RMSE: 0.037–0.0084 MAE: 0.0071–0.001 RMSLE: 0.015–0.0037	Rainfall Prediction
[2]	Bangladesh	Dhaka	Yearly flood data near 34 stations (1980-2020)	Logistic Regression (LR), SVC, KNN, DT	LR: 0.8676, SVC: 0.8088, KNN: 0.8235, DT: 0.8088	Logistic Regression	Flood
[3]	Kuala Lumpur	Kelang River at Petaling Bridge	Real-time Rainfall data (19/11/2010 - 21/11/2010)	NNARX, Gradient Descent Back Propagation	Best Fit: 89.822%, Prediction Error: 0.0041 m	Loss Function RMSE: 0.0634 m, (V) 0.0040 m	Water Level Check
[56]	Urban	Lisbon	Data from Jan 2013 to Dec 2018 (52584 observations)	Random Forest (RF), GIS	RF: Accuracy 0.96, MCC: 0.77	Combined Hot Spot with RF Model	Flood Prediction (Hourly Data)
[31]	Thailand	Surat Thani and Nakhon Si Thammarat	Excessive 5-year and 100-year return period	DT, RF, Naïve Bayes, MLP, RBF, SVM, Fuzzy Logic	MLP ANN: 97.83%, SVM: 96.67%, RF: 96.67%	MLP ANN, SVM, RF	Flood Forecasting
[23]	Taiwan	Yilan River basin (Liwu station)	Hourly Rainfall Data (2012-2018, 6 gauges)	KNN, Support Vector Regression (SVR), Fuzzy Inference Model	RMSE: 0.07, CE: 0.99 (1-hour)	SVR, Fuzzy Inference Model	Hourly Forecasting
[24]	Brazil	São Carlos, São Paulo	Rainfall in April 2014 (days:hours:min format)	Chaos Theory, MLP, E-RNN	MLP: R^2=0.994	MLP (Multi-layer Perceptron)	Flood
[33]	Bangladesh	Southwestern Coastal Region	Yearly Flood Data (BARC)	MLP, KNN, RF, Genetic Algorithm (GA;RBF;SVR)	MLP: 0.967, KNN: 0.956, RF: 0.984	Optimized Model: 0.987	Flood
[4]	Bangladesh	Teesta River basin	206 Non-Flood Locations selected randomly	Bagging Classifier (RF, RT, M5P, REPtree)	M5P: AUC=0.945	Bagging Models (RF, REPtree, RT)	Flood Spot Detection
[5]	Bangladesh	Multiple Cities	Dataset from 2016-2019 (2391 records)	DT, KNN, LR, NB, RF	RF: 87.68%	Random Forest (RF)	Rainfall Prediction
[6]	Bangladesh	Gazipur, Rangpur, Barisal Districts	Rainfall Data (2011–2020)	DT, RF, SVM, NN	BLR: 0.8676	Binary Logistic Regression (BLR)	Rainfall
[34]	Bangladesh	32 Districts	65 Years of Meteorological Data	KNN	K=2: 92.8%, K=3: 93.4%, K=4: 93.7%, K=5: 94.2%, K=6: 94.5%, K=9: 94.7%	Best Accuracy: 94.91%	Flood Prediction
[23]	Taiwan	Yilan River Basin (Liwu station)	Hourly River Data (2012-2018, 15 floods)	SVR, Fuzzy Inference Model, KNN	90% CI: Acceptable Results	Probabilistic Forecasting	Real-Time Probabilistic Flood Forecasting
[45]	Western Australia	Marradong, Quanbun Downs, Sturt Creek	Rainfall Data (1957-2013)	MLR, ANN	Coefficients: 0.35 to 0.93 (MLR)	Superiority of Non-Linear Modeling	Seasonal Precipitation Forecast
[15]	France	Gardon_d'Anduze River	Hydrometric Data (2002-2018)	MLR, ANN	Nash Criterion: 0.9381	Satisfying Outcomes	Flood Prediction
[55]	Australia	Nationwide	10 Years Data (145460 rows, 23 attributes)	NB, DT, SVM, RF, Logistic Regression, ANN, PCA	ANN: Accuracy 91%	High Accuracy ANN	Rainfall Classification

flaws. Limited availability of high-quality and spatially dense rainfall data poses challenges for model training and validation. Overfitting, especially in complex machine learning models, can lead to poor generalization performance, particularly when dealing with short and noisy time series data. Many models perform poorly in novel contexts because they over fit to datasets

ML models have the ability to accurately predict rainfall; nevertheless, there can be a lack of interaction between them and decision support systems (DSS) to facilitate real-time decision-making. -User-friendly interfaces, interoperability, and seamless integration are essential to guaranteeing the practical applicability of machine learning-based forecasting systems. Interpretability and explain ability are often lacking in machine learning models, particularly deep learning models, which makes it difficult for users to comprehend how predictions are made. This restriction impedes decision-making, adoption, and trust in operational forecasting applications.

III. METHODOLOGY

A. Dataset Description

A dataset gathered from the kaggle platform constituted a basis for the work discussed in the article. As indicated in Table II, the data collection covers a sample of 145,460 entries with information on 23 research variables. The values provide meteorological information that was compiled over a ten-year period from 49 distinct Australian cities. The target parameter for the machine learning and deep learning algorithms' prediction task is a Boolean variable named "Rain Tomorrow" indicating yes or no as to whether it will rain tomorrow. Similarly Table II displays the dataset's dimensionality as compared to benchmark studies. Comparison Benchmark Dataset (Rainfall Prediction Attribute) is presented in Table II .

TABLE II. COMPARISON OF INPUT PARAMETERS/ATTRIBUTES USED IN PREVIOUS PAPERS

Ref No	Input parameters/Attribute Used in Previous Paper	Dimensionality of the datasets was downsized	Our work features
[28]	Pressure, moisture, wind speed, wind level, visibility of clouds, temperature, and time zone, Snow, rain, and snow all in three hours.	11 features	• ‘Date’ • ‘Min Temp’ • ’Max Temp’ • ‘Rainfall’ • ’Evaporation’ • ‘Sunshine’ • ‘WindGustSpeed’ • ‘WindSpeed9am’ • ‘WindSpeed3pm’ • ‘Humidity 9am’ • ’Humidity 3pm’ • ‘Pressure 9am’ • ‘Cloud9am’ • ‘Cloud3pm’ • ’Temp 9am’ • ’Temp 3pm’ • ‘Location’ • ‘WindGustDir’ • ’WindDir9am’ • ’WindDir3pm’ • ’RainToday’ • ’RainTomorrow’ • ‘Pressure 3pm’
[2]	State, district, year, month, rainfall, max temp, min temp, flood occurrence	8 features	
[56]	Measurements of humidity, temperature, exposure to the sun, rainfall, velocity of the wind, and speed of wind are included in the dataset	6 features	
[33]	Aspect, Elevation, Slope, Curvature, Land Subsidence, Precipitation, Flow accumulation, SPI, TWI, Land cover, Soil texture, Soil permeability, Distance to drainage channels, Distance to rivers	14 features	
[4]	Topographic factors such as elevation, slope, curvature, aspect, STI (Topographic Wetness Index), SPI (Standardized Precipitation Index), and TWI (Topographic Wetness Index), LULC (Land Use and Land Cover), rainfall, distance to the river, and soil type are the twelve factors which can be selected.	12 parameters	
[5]	MaxTemp, MinTemp, Actual Evaporation, Relativehumidity9am, Relative humidity 2 pm, Sunshine, Cloudy, Solar Radiation, Rainfall	9 features	
[6]	On an annual rate, from January to December: Flood, Station; April; May; June; July; August; September; October; November; and December.	16 features	
[34]	Precipitation, the amount of cloud cover, the humidity level, the lowest temperature, the speed of the wind, etc.	5 features	
[27]	Temperature and Rainfall	2 features	
[45]	Seasonal rainfall and climate	2 features	
[15]	Flooding incidents are used to train and test models: (i) 09-10 November 2018, (ii) 09-10 September 2002	25 events	
[59]	Water cut, saturation perforation, supplied petroleum radius, density of perforations, controlling area, controlled reserves, thickness of reservoirs, degree of drilling process, hole radius, flow bottom, and hole pressure, Permeability	13 features	

B. Numerical and Categorical Weather Feature Used as a Predictor

Fig. 2, displays the missing values in each feature of the dataset in a more comprehensive manner. The yellow bars indicate missing values, while the purple bars represent available data. The columns correspond to the features in the dataset, and the figure illustrates the distribution of missing values across all variables.

Fig. 3 shows the behavior of all features, and Table III provide descriptions of the numerical characteristics together with details on their kind, availability of data, and units. Table IV define the numerical feature description with type and unit.

In this context, the eight compass points - North (N), Northeast (NE), East (E), Southeast (SE), South (S), Southwest (SW), West (W), Northwest (NW), - as well as the points in between are the wind features. Table IV has one more feature that includes two ways to display data i.e. unit, kind, category, and having data or not.

C. Correlated Features in the Dataset

Breaking the date feature in month day year and the Fig. 5 showing the strong and weak correlation with the features. Tuples of Highly positively & strongly Correlated Features in overall Continent are provided in Table V.

D. Locations of Study Area

In this research the effect of the data’s location was looked at. The weather may vary too much in places that

TABLE III. METEOROLOGICAL NUMERICAL FEATURES DESCRIPTION

No	Name	Features Brief Description with Units	Type	Unit	Missing Value	Available Data
1	'Date'	The complete day date and of rainfall occurrence	string	(No unit)	0	145460
2	'Min Temp'	The lowest temperature that a certain day might experience.	decimal	Celsius (°C)	1485	143975
3	'Max Temp'	The maximum temperature recorded on a particular day.	decimal	Celsius (°C)	1261	144199
4	'Rainfall'	Rainfall on a specific day.	decimal	millimeters (mm)	3261	142199
5	'Evaporation'	Drying on a specific day.	decimal	millimeters (mm)	62790	82670
6	'Sunshine'	On a certain day there was bright sunshine.	decimal	hours	69835	75625
7	'Wind Gust Speed'	Strongest wind gust's speed on a given day.	decimal	kilometers per sec	10263	135197
8	'WindSpeed9am'	Wind speed for 10 minutes before 9 am.	decimal	kilometers per sec	1767	143693
9	'WindSpeed3pm'	Wind speed for ten minutes before three o'clock.	decimal	kilometers per hour (km/h)	30622	142398
10	'Humidity 9am'	The percentage of the wind's humidity at 9:00 am.	decimal	percentage (%)	2654	142806
11	'Humidity 3pm'	The percentage of the wind's humidity at 3 PM.	decimal	percentage (%)	4507	140953
12	'Pressure 9am'	Atmospheric pressure at the time 9am it was observed	decimal	hectopascals (hPa)	15065	130395
13	'Pressure 3pm'	Atmospheric pressure at 3 PM the observed time	decimal	hectopascals (hPa)	15028	130432
14	'Cloud9am'	Areas of the sky that are clouded in at 9:00 am.	decimal	(No unit)	55888	89572
15	'Cloud3pm'	Areas of the sky that are clouded in at 3 PM.	decimal	(No unit)	59358	86102
16	'Temp 9am'	Temperature of rainfall at 9 am	decimal	Celsius (°C)	1767	143693
17	'Temp3pm'	Temperature of rainfall at 3 PM	decimal	Celsius (°C)	3609	141851

TABLE IV. METEOROLOGICAL WIND FEATURES DESCRIPTION

No	Name	Description	Categories in Specific Feature	Percentage	Type	Unit	Missing Value	Available Data
1	Wind Gust Dir	The wind's direction over the 24 hours leading up to midnight (sixteen compass points)	ENE, ESE, N, NE, NNE, NNW, NW, S, SE, SSE, SSW, SW, W, WNW, WSW, NaN	NA 7%, W 7%, Other (125219) 86%	string	(No unit)	10326	135134
2	Location	Specific name of the Australian city where rainfall was recorded	Newcastle, Albury, Badgerys Creek, Cobar, Coffs Harbour, Moree, Wagga, Williamtown, Wollongong, Canberra, Tuggeranong, Mount Ginini, Ballarat, Bendigo, Sale, Melbourne Airport, Melbourne, Woomera, Albany, Witchcliffe, Pearce RAAF, Perth Airport, Perth, Salmon Gums, Walpole, Hobart, Launceston, Alice Springs, Darwin, Katherine, Uluru	Canberra 2%, Sydney 2%, Other (40676) 95%	string	(No unit)	0	145460
3	Wind Dir 9am	The direction of the wind in the first ten minutes before 9 am	WNW, ENE, NE, SSW, ESE, NW, S, W, SW, NNE, NNW, N, SE, E, SSE, WSW	SE & WW 0.07%, Others 0.06%	string	(No unit)	10566	134894
4	Wind Dir 3pm	10 minutes before 3 o'clock the wind's direction	SSE, NNW, ENE, NNE, WENE, WSW, SSE, SW, NW, N, ESE, ENE, SSW, others	N 0.087%, Other less than 1%	string	(No unit)	3062	142398
5	Rain Today	'Yes' if it rains today. 'No' if not raining today.	Yes, No	Total 3261	string	(No unit)	3261	141851
6	Rain Tomorrow	If it rains tomorrow, then 1 (Yes). If it doesn't rain tomorrow, then 0 (No).	Yes, No	Total 3267	string	(No unit)	3267	142193

TABLE V. TUPLES OF HIGHLY POSITIVELY AND STRONGLY CORRELATED FEATURES IN OVERALL CONTINENT

No	Tuple	Correlation Coefficient
1	The attributes Max and Min temperatures have a significant positive correlation.	0.74
2	There is a significant positive link between the minimum temperature and the temp3pm.	0.71
3	The attribute 9am temperature and Min Temp are strongly positively correlated.	0.90
4	Max Temperature and Temp 9am exhibit a strong positive association.	0.89
5	Maximum temperature and temperature at 3 p.m. are both fairly high.	0.98
6	Wind Gust Speed and WindSpeed3pm variables are highly positively correlated.	0.69
7	Pressure 9am and Pressure 3pm variables are strongly positively correlative.	0.96
8	The variables Temp 9am and Temp3pm have a high positive correlation.	0.86

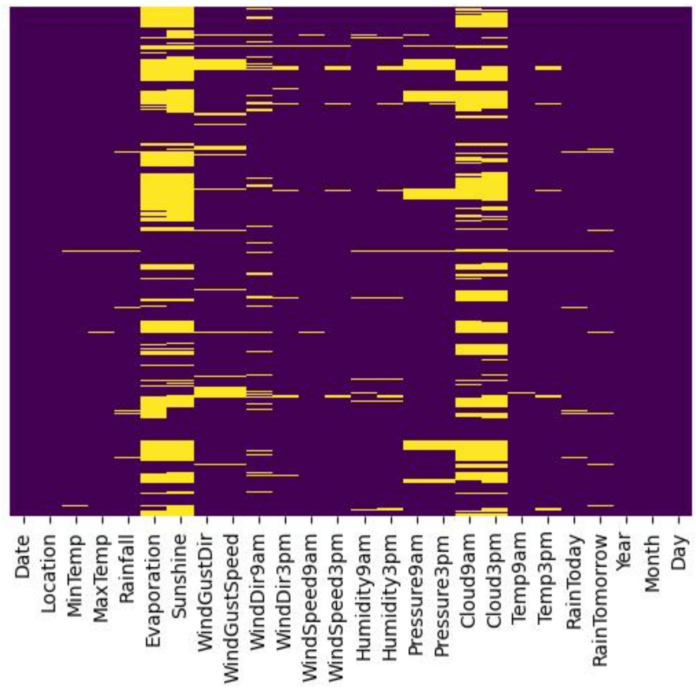

Fig. 2. Visual representation of missing values for each feature in the dataset.

are predominantly in different parts of Australia. The earlier study made use of a dataset that had a large number of cities. There are data gaps in several cities. For this reason, consider the following five Australian research regions((Darwin, Perth Airport, Sydney, Brisbane, and Melbourne). Given the aforementioned and the fact that local weather patterns and microclimates can frequently differ greatly, especially when taking into consideration weather forecasts for the entire continent, it should be more sensible to construct unique models for various locations. As a result, this study produces weather predictive models for various areas. To comprehend the numerous factors that affect whether it rains the following day, this study consider each city's specific model separately. In this study, it was also investigated whether applying machine learning and deep methods may help with rainfall predictions in this particular Australian location. Fig. 5 shows the locations of five Australian cities (Drawn, Perth Airport, Sydney, Brisbane, and Melbourne).

Fig. 6 displays the city-specific rain forecast for tomorrow. Fig. 7 shows that the month has less of an impact on the distribution of rainy days in Sydney and Melbourne. However, count plots for Brisbane, Perth, and Darwin indicate that these locations experience both wet and dry months (particularly Darwin).Darwin had more wet days than Perth did over the course of the research period, even though the count plot indicates that there are much more days with rain than days without

In particular, this study will be used to find out whether the application of deep learning and machine learning algorithms can lead to a higher precision grade and a drop in errors.. Everyone who is now alive in the country will gain something from this endeavor.Four machine learning methods are used in the suggested method to forecast rainfall Random Forest(RF), Gradient boosting (GB) etc. The framework incorporates a number of crucial processes, pre-processing, data normalization and feature engineering including feature selection, feature encoding, model training, and prediction evaluation. The proposed work involves the optimization of a classification model for the rain prediction with the help of supervised machine learning (ML) and deep learning(DL) methods. The primary objective of the research is to achieve maximum accuracy in rainfall prediction. This objective is pursued through the application of supervised learning and deep learning methodologies. The methodology revolves around supervised learning, where the model learns patterns from labeled training data. Specifically, deep learning is highlighted, resulting in the application of multiple layer neural networks that exhibit an ability to capture complicated patterns in the data. Fig. 8 indicates the overall framework of the proposed scheme for forecasting precipitation. It may illustrate the different components of the model, such as data preprocessing, feature extraction, model training, and evaluation. The architecture of the proposed research work is presented in Fig. 8.

1) Data cleaning and pre processing missing data: Another significant part of data preprocessing in machine learning generally is management of missing data in a dataset. Fig. 2 demonstrates the number of blank samples available for two variables. Through a total of 145,460 samples, for 23 variables.Therefore, nearly 45% of the samples would need to be deleted if the samples with no data for any of their variables were also eliminated. In order to avoid throwing away a lot of data(there are a total of 49 cities dataset contain) e. g ['albury' ;'Badgerys Creek' ,'cobar' ;'coffs harbor' 'moree' 'Newcastle'; 'Norah Head'; 'Norfolk Island'; 'penrith' 'Richmond' ,'Sydney' ,'Sydney Airport' 'waggaWagga' ,'Williamstown' ,'Wollongong'; 'Canberra' 'Tuggeranong'; 'mount_Ginini' ,'ballarat' ,'bendigo', 'Sale' 'Melbourne AirPort' 'Melbourne'; 'Mildura'; 'nhil' ,'Portland';'Watsonia'; 'Dartmoor' ,'Brisbane' 'cairns', 'gold Coast' ;'Townsville' ,'Adelaide' ,'mount Gambier' ,'nuriootpa' ,'woomera', 'Albany',;'Witchcliffe' ,'Pearce').

In total, seven (7) different stations in the dataset. In particular, less than 10 percent of string text data is missing or absent. The variables that lacked data were examined and categorized according to the cities. The examination of factors for which there are data produced the following results. In some of the cities, certain parameters do not have any data at all. There are samples for which certain variables have no information. It is assumed that this is due to a lack of a corresponding sensor at the city's weather site or data not recorded. A malfunction in the sensors' communication with each other could be the reason. As with the previous case, two different scenarios were shown to exist: data loss for one day and data loss for several days in a row [16]. It was decided to remove all null values for all features in this case. Using only 41% of observations is possible. Detecting and removing outliers from the dataset finally, in the case of objective variables, In this work balance the imbalance data. Fig. 9 and Fig. 10 depict the balance and imbalance target variable.

2) Data normalization: Scaling variable values to give them the same quantitative weight and place them on the same interval or scale is known as data normalization. Rescaling the

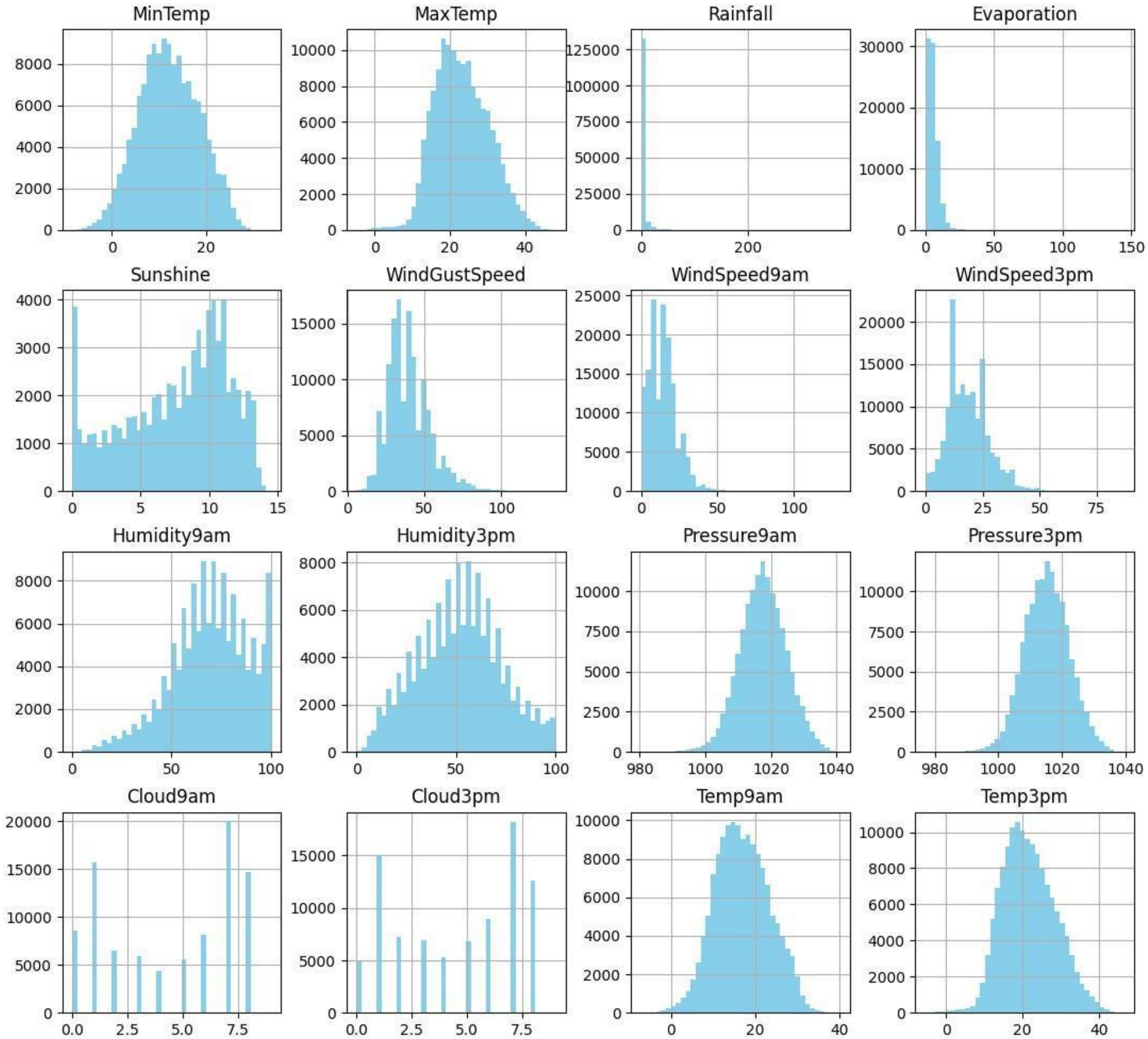

Fig. 3. Histogram showing the dataset's attribute statistics.

complete dataset to a standard distribution or range. Eq. (1) provides a method for normalizing data using the min-max scaling strategy.

$$\mathrm{x}_{(normalized)} = \frac{x - min(x)}{max(x) - min(x)} \tag{1}$$

Where: x stands for the dataset's original value, min(x) for its lowest value, max(x) for its highest value, and x_normalizedfor its normalized value falls between 0 and 1. The values of the complete dataset are scaled by Eq. (1) to the range [0, 1], where the smallest value is made to equal 0 and the largest value is produced to equal 1.

3) Feature engineering: First, the map reveals that Watsonia, Perth, and Melbourne airports are all close to one another. Based on this, assume that it makes sense to select Melbourne and Perth airports for rain prediction because they have fewer null variables than Watsonia and Perth. Although the date field itself does not contain any meteorological information, it is possible to get the month and utilize it to study weather patterns. In this experiment, the month feature that had been removed is replaced by (January, February, March, April, May, June, July, August, September, October, November and December). Since the data in the region as a whole had to be investigated, the data pertaining to each place has not been divided to create distinct subsets. For instance, there are some cities with heat and humidity circumstances that more or less favor rain, based on their location. Similarly, on the day the data is collected, several weather events may take place that affect the rain.

4) Feature encoding: In our dataset some features are categorical and some numerical. The category variables were then converted into numerical values. Two distinct sets of data variables had to be used in this process On the other hand, because the parameters WindGustDir, WindDir9am, and WindDir3pm show the direction of the wind, the data transform

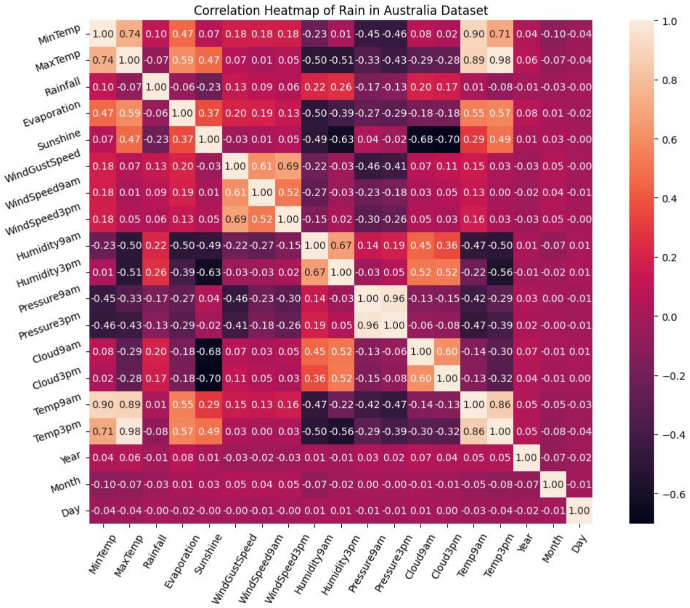

Fig. 4. Correlation with variables.

into numeric form before using them because they contain a categorical data string. Our study's goal variable is called Rain Tomorrow, and its values of "1, 0" are applied to Boolean representations of type "string" (which only take YES/NO responses). Fig. 4 shows correlation with variables.

5) Feature scaling: The process of scaling individual features within a dataset to have comparable magnitudes is known as feature scaling. The Standard Scalar was used to scale the dataset's features in order to guarantee that it was equitable and appropriate for the models that were used. The data are scaled by placing a unit standard deviation around the mean. The feature scaling using the z-score scaling or standardization technique is represented in Eq. (2). The feature scaling takes place element wise and for every feature.

$$\mathrm{x}_{(Scalled)} = \frac{x - \mu}{\theta} \tag{2}$$

Where: x indicates the feature's original value, is the dataset's mean (average), is the feature's standard deviation, and ,The scaled value of the feature, x_scaled, has a standard deviation of 1 and a mean of 0. Eq. (2) is applied to alter the attribute values to have a mean of 0 and a standard deviation of 1.

6) After preprocessing result: A total of 145460 samples were collected over a ten-year period in 49 Australian cities for the initial set of variables, which included 23 ('Date', 'Min Temp', 'Max Temp', 'Rainfall', 'Evaporation', 'Sunshine', 'Wind Gust Speed', 'Wind Speed 9am', 'Wind Speed 3pm', 'Humidity 9am', 'Humidity 3pm', 'Pressure 9am', 'Cloud 9am', 'Cloud 3pm', 'Temp 9am', 'Temp 3pm', 'Location', 'Wind Gust Dir', 'Wind Dir. 9am', 'Rain Today', 'Rain Tomorrow', 'Pressure 3pm',. Unlike Salvia Main, the site's dataset is made of 79 columns, resulting in a total of 14 727 samples. In addition to that, approximately 80% of the data gathered out of these 14727 samples is used for training the models and remaining 20% of the data is used for checking the functioning of the newly developed models.

In the present Australia rainfall prediction, the integration of some basic normalization methods such as min-max scaling and z-score standardization enhance the model prediction. Most of the rainfall prediction models involve various properties such as temperature, humidity and wind speed, all of which have different units and scales. Scaling makes sure that all different features are equally important, thus their values are transformed into the same range. This restricts the longer feature space ranges from becoming dominant over other small feature space ranges and aids in learning algorithms like neural networks and Decision Tree learn faster and better during the training process. Normalization also enhances numerical stability and prevents the model from having a bias towards features with large variances, thus enhancing the accuracy of the rainfall estimation.

IV. Results

Using the location segment, for analyzing and dividing the dataset into various regions so that it could create a variety of unique models, Fig. 7 represents the different locations of Australia. To see the differences between the causes triggering rain on subsequent days, analyze the models independently. Machine learning (ML) techniques have considerably improved prediction systems over the past two decades by offering more efficient and approachable means of replicating the intricate mathematical representations of the physical processes causing floods. Analyzing the possibilities that machine learning and deep learning algorithms offer to conventional forecasting methodologies for the prediction of rain is the aim of this study. This was accomplished using the techniques of neural networks (NN), Decision Tree (DT), Random Forests (RF), and Gradient Boosting Classifiers (GBC). Examine the benefits of rainfall probability in Australia's five biggest cities: Darwin, Sydney, Brisbane, Perth, and Melbourne, using the methodology given.A more detailed description of the algorithms may be found below. The values from the training dataset have been predicted by using a specific location.

1) Decision tree: Using a Decision Tree machine learning approach [60]–[62], problems with regression and classification are addressed. Its organization is comparable to a flowchart, where a decision rule is represented by each branch, an attribute or characteristic by each internal node, and a result or class label by each leaf node.

2) Random forest: During training, a massive number of decision trees are constructed by an ensemble learning system known as Random Forest [62], which then outputs the average forecast for regression tasks or the majority vote for classification tasks. Random Forest is resistant to over fitting and performs well on a variety of datasets. It can handle missing values and remain accurate even with a large feature set. The Random Forest approach is depicted in Fig. 11

3) Gradient boosting classifier: A powerful ensemble learning method for classification and regression applications is gradient boosting. Gradient Boosting constructs Decision Tree sequentially, with each tree learning from the mistakes of its predecessors, in contrast to typical decision tree algorithms like Random Forest, which generate many trees individually

Fig. 5. Geographical location of five specific regions of Australia.

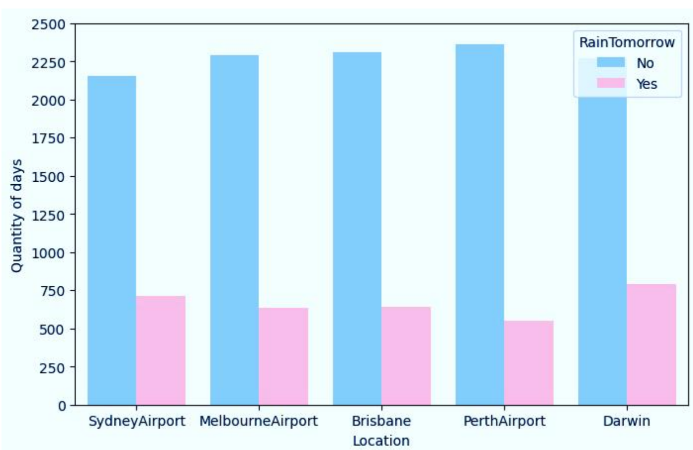

Fig. 6. Tomorrow's forecast for a specific city.

[63] as shown in Fig. 12. Initialize the model with a simple model, such as a single leaf (constant) value for regression or a constant probability for classification. Then it calculates the residuals or pseudo-residuals for each data point, which represent the errors made by the initial model.

$F_m((x)$ as the current ensemble model (sum of first m weak learners) $h_m((x)$ as the m-th weak learner (e.g., Decision Tree), Φ as the learning rate, L as the loss function. At each iteration, update the model as follows:

At each iteration, update the model through Eq. (3) as follows:

$$F_m(x) = F(-1)(x) + \rho \cdot h_m(x) \tag{3}$$

Then, the residuals (or pseudo-residuals) are updated the Eq.4:

$$r_{im} = \frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F_{m-1}(x_i)} \tag{4}$$

Finally, the prediction at each iteration is given by Eq. 5:

$$y_{\text{pred}}(x) = F_M(x) = \sum_{m=1}^{M} \rho \cdot h_m(x) \tag{5}$$

4) Recurrent Neural Network: Recurrent Neural Networks (RNNs), on the other hand, serve as the foundation for sequential data processing within neural networks. These architectures operate iteratively through sequences, updating hidden states at each step to encapsulate contextual information. However, RNNs often encounter challenges when attempting to retain information over prolonged sequences, commonly referred to as the vanishing gradient problem. Despite their limitations, RNNs remain widely used for various sequential data tasks, such as language modeling, sentiment analysis, and machine translation. While they may struggle with long-term dependencies, RNNs offer simplicity and computational efficiency, making them suitable for applications where shorter-term relationships are predominant.

The RNN's computations are governed by the following formulas. Eq. (6) is utilized in the computation of hidden states:

The function $h(t)$ is calculated as follows:

$$h(t) = f\left(W_{xh} \cdot x(t) + W_{hh} \cdot h(t-1) + b_h\right) \tag{6}$$

Eq. (6) is used for output calculation:

$$y(t) = f\left(W_{hy} \cdot h(t) + b_y\right) \tag{7}$$

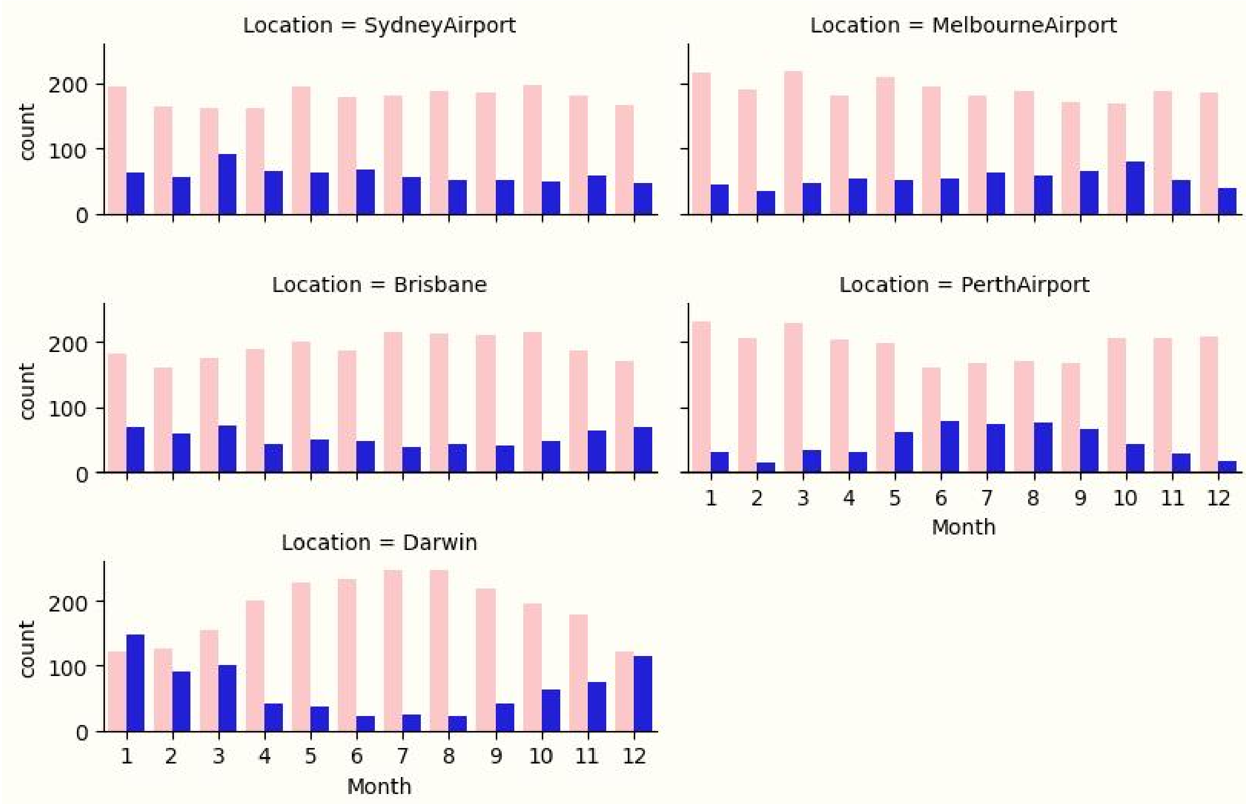

Fig. 7. Tomorrow's forecast for a specific city.

Where t is the time step, $x(t)$ is the input at time step t, and $h(t)$ is the hidden state at time step t. The weight matrices W_{xh} and W_{hh} control the flow of information, and b_h and b_y are bias vectors.

The following equations are used for the input gate:

$$i(t) = \text{sigmoid}\left(W_i \cdot [h(t-1)\, x(t)] + b_i\right) \tag{8}$$

The candidate cell state $\hat{C}(t)$ is computed as:

$$\hat{C}(t) = \tanh\left(W_c \cdot [h(t-1)\, x(t)] + b_c\right) \tag{9}$$

The cell state updating function is given by:

$$c(t) = f(t) \cdot c(t-1) + i(t) \cdot \hat{C}(t) \tag{10}$$

The output gate computations are as follows:

$$o(t) = \text{sigmoid}\left(W_o \cdot [h(t-1)\, x(t)] + b_o\right) \tag{11}$$

Finally, the hidden state is calculated as:

$$h(t) = o(t) \cdot \tanh\left(c(t)\right) \tag{12}$$

A. *Criteria for Evaluating Models*

The metrics (or key indicators of performance (Key Performance Factors)) that will be used to evaluate the algorithms' output are described in this section [64].

1) Accuracy: Number reflecting how well the predicted model performed. The formula shown in Eq. (13)

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \tag{13}$$

Where, TP designates it as "true positives." "Result where the model outputs a positive class and correctly classifies it. FP is called a False Positive. Lead to a case in which the positive class is erroneously designated by the model as a negative class. TN, or the true negative, connotes the outcome where the model predicted the negative class to be. False negative, i.e. FN is a concept in detection that is associated with negative. is a situation where the model predicts the other class to be wrong.

2) Precision: The percentage of instances that are correctly identified as positive is known as precision. Which is, whether a model forecasts positive numbers. The formula shown in Eq. (14).

$$\text{Precision} = \frac{TP}{TP + FN} \tag{14}$$

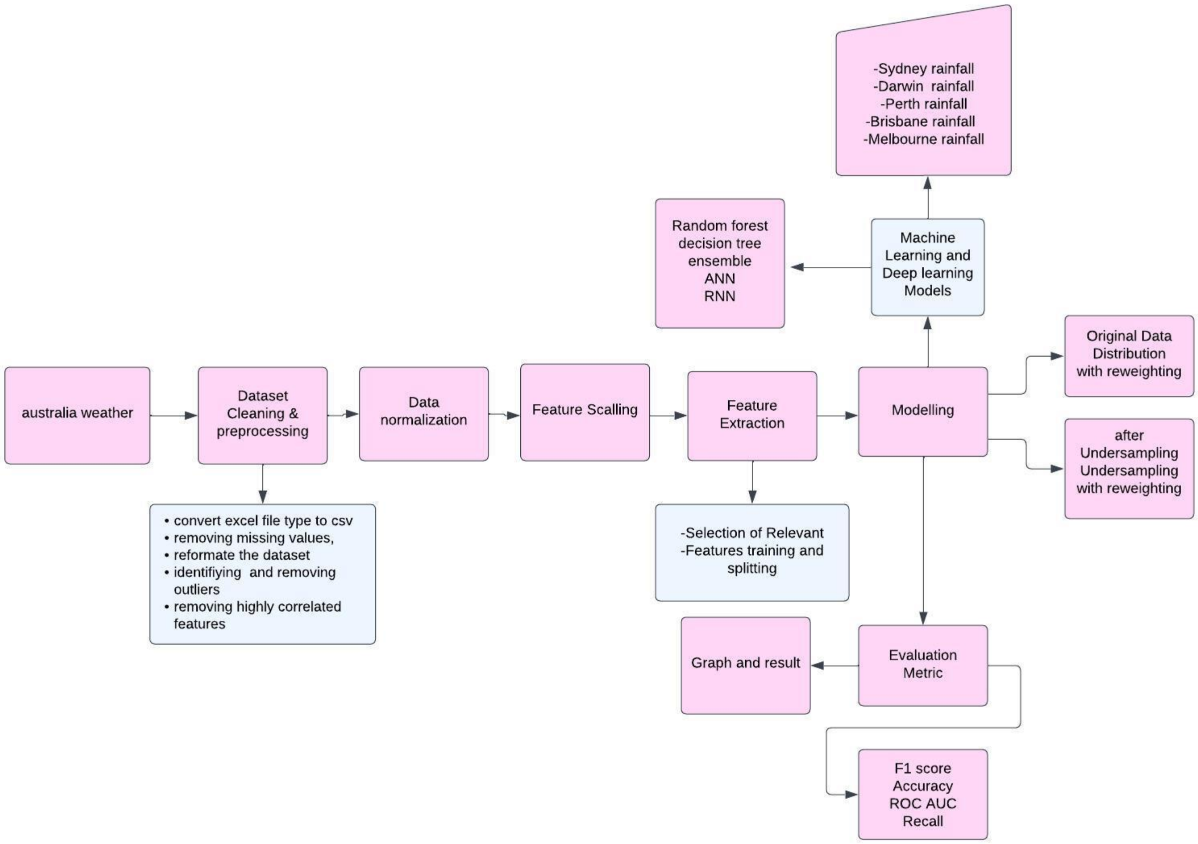

Fig. 8. Architecture of the proposed work.

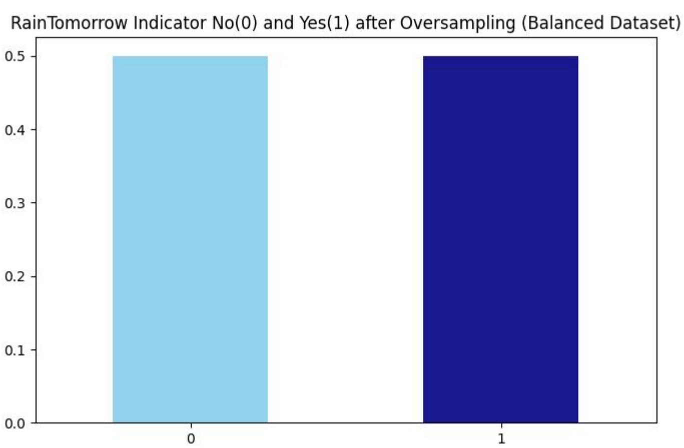

Fig. 9. Rain tomorrow indicator no(0) and yes(1) after oversampling (balanaced dataset).

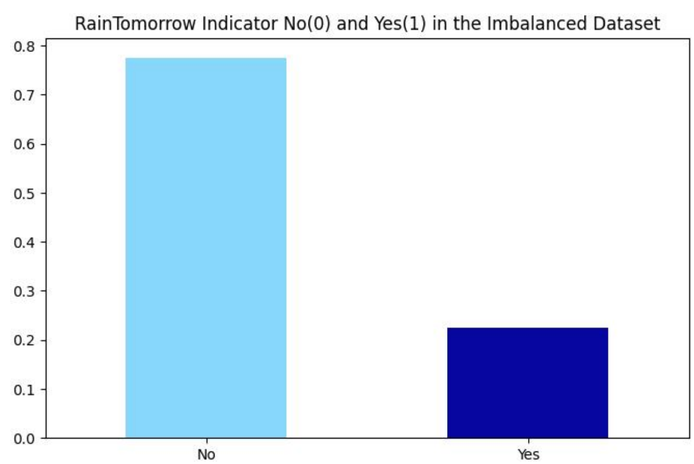

Fig. 10. Rain tomorrow indicator no(0) and yes(1) in the imbalanaced dataset.

3) Recall: The percentage of correctly detected positives to all positives is known as recall. The sensitivity formula and this formula are identical as shown by Eq. (15).

$$\text{Recall} = \frac{TP}{TP + FP} \tag{15}$$

4) F1 score: When precision and recall are insufficient for evaluating performance, for as when one mining method has better accuracy but worse recall than another, the question of which algorithm is superior may come up. The F-measure, which gives the mean of recall and precision, can be used to address this problem. An industry standard for evaluating

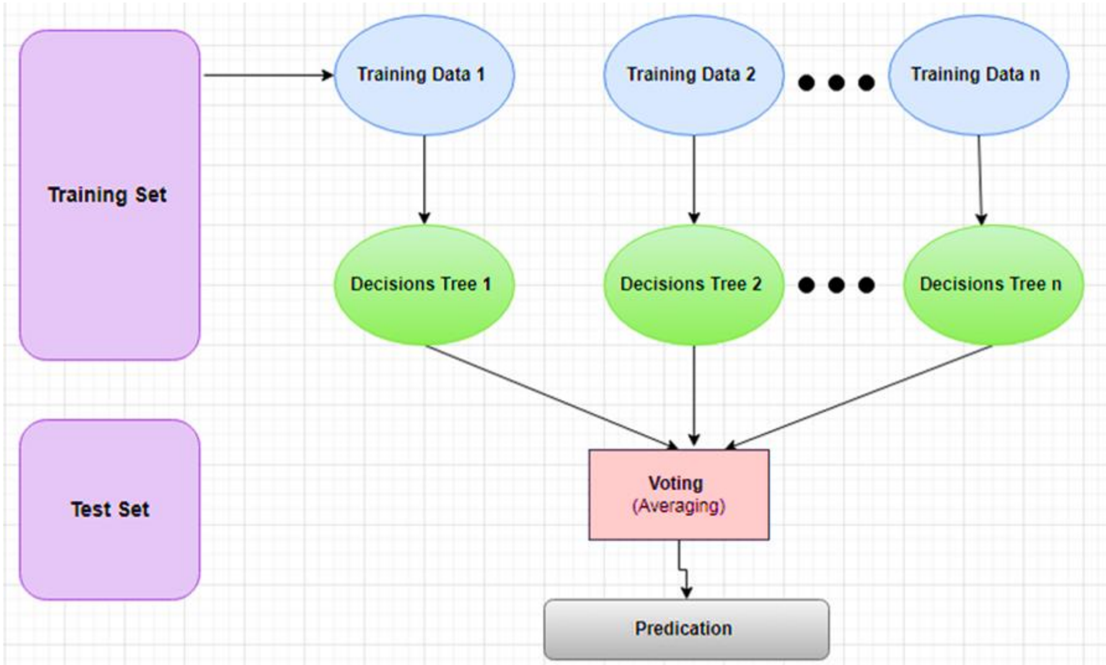

Fig. 11. Representation of random forest algorithm.

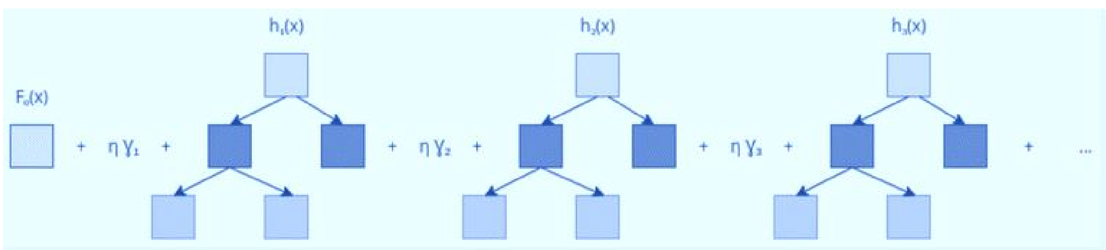

Fig. 12. Representation of gradient boosted trees.

the performance of a classification model is the F1 score. Eq. (16) shows how the computation appears. It provides an equitable evaluation of a model's accuracy by merging recall and precision into a single metric.

$$\text{F1 score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{16}$$

5) ROC AUC: AUC ROC stands for areas under the curve of the "Receiver Operating Characteristics" curve. The performance of an ML model is commonly assessed using the AUC ROC curve. The ROC curve's AUC definition measures how well a binary classifier can differentiate between different categories.

B. Experimental Results

Due to the fact that factors affect rainfall in different regions differently, it is impossible to produce a model that would be able to predict rainfall across the whole Australia. Models can be constructed only for restricted areas to know what roles with probability of rain are the variables being assigned. The data set contains daily weather observations from several locations in Australia for about 10 years. The impact of the data's location was examined in this experiment. The data in this set includes around ten years' worth of daily weather observations made in various parts of Australia. However, This study concentrated on five specific cities (Perth Airport, Malbulane, Brisbane, Sydney, and Darwin) for the objectives of our study. In order to predict whether it would rain on a certain day, machine learning methods were tried on the Rain Tomorrow feature.

1) Sydney decision tree results: Table VI presents Sydney result using Decision Tree.

Table VI represents the model's result with the initial data and the undersampling. It can be said that the results of the Random Forest and Gradient Boosting classifiers are very near to each other with the Decision Tree and Ensemble classifiers as shown in Table VII. In this study Random Forest and Gradient boosting models are built using undersampling and assume that the RF classifier, given its somewhat superior performance, is the best model for the dataset.

Table VI presents Sydney Comparison of all results By using Random forest and Gradient Boosting.

TABLE VI. SYDNEY RESULT USING DECISION TREE

Metric	Original Data Distribution	Original Data Distribution with Reweighting	Under Sampling	Under Sampling with Reweighting
Accuracy	0.83	0.78	0.77	0.70
F1 score	0.60	0.60	0.61	0.58
ROC_AUC	0.72	0.75	0.75	0.74
Recall class 1	0.50	0.69	0.71	0.84

TABLE VII. SYDNEY COMPARISON OF ALL RESULTS USING RANDOM FOREST AND GRADIENT BOOSTING

Metric	Original Data Distribution	Original Data Distribution with Reweighting	After Under Sampling	Under Sampling with Reweighting	Random Forest	Gradient Boosting
Accuracy	0.83	0.78	0.77	0.70	0.79	0.78
F1 score	0.60	0.60	0.61	0.58	0.64	0.64
ROC_AUC	0.72	0.75	0.75	0.74	0.78	0.77
Recall class 1	0.50	0.69	0.71	0.84	0.76	0.76

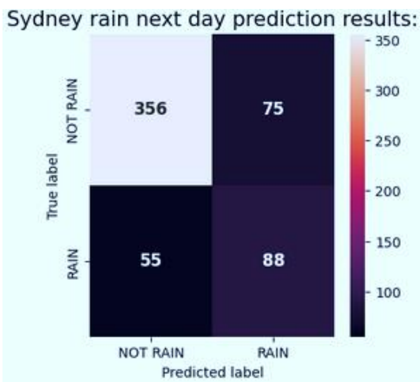

Fig. 13. Sydney next day prediction result.

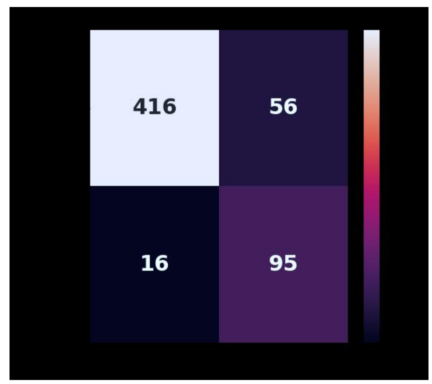

Fig. 14. Perth airport next day prediction result.

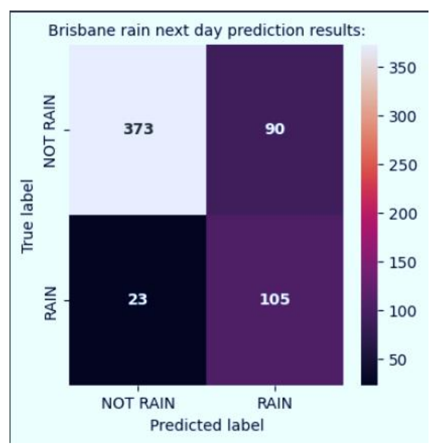

Fig. 15. Brisbane next day prediction result.

C. Perth Airport, Brisbane, Melbourne, Darwin Comparison Accuracy using ML Algorithm

Building a model that can forecast rain for the entire Australian continent is unachievable because different factors have varying effects depending on where they are. Only for specific areas models build and look at how the variables affect the probability of rain. In this study created models for specific locations and examined the variables' effects on the likelihood of rain. Without comparing it to a single tree, Random Forest model with under sampled training data is the best choice for Perth Airport, Brisbane, Melbourne, Darwin. Table VIII represents the prediction result for all 5 locations of Australia and also shows us that it's possible to predict rain for next day with different accuracy depending on the location (using one model type - RF classifier).

The confusion matrix of five specific regions of Australia for rainfall prediction are presented in Fig. 13, 14, and 15.

D. Estimation of Feature Importance

The estimation of feature importance for different cities is presented in this subsection as follows:

1) Estimation of feature importance for Sydney: Here only a couple of weather factors affect Sydney weather. The main factors are sunshine, wind speed 3 p.m., humidity 3 p.m., maximum temperature, air pressure 9 a.m. and air pressure 3 p.m. TableIX provides Feature Importance for Sydney. All other features have importance less than 1 percent.

2) Estimation of feature importance for Perth airport: The most essential factors causing rain next day in Perth (Airport) are Pressure 3pm, Pressure 9am, Wind Gust Speed, Evaporation, Wind Speed 3pm, Sunshine, Rainfall. Table IX describe some factors with F1 score. Table X provides Feature Importance for Perth Airport city. All other features have importance less than 1 percent.

3) Estimation of Feature Importance for Brisbane city: The key elements causing rain the following day in Brisbane are the following: humidity, 3 p.m. sunshine, humidity, 9 p.m.,

TABLE VIII. Prediction Results for All 5 Locations of Australia Using Random Forest

Metric	(i) Sydney	(ii) Darwin	(iii) Perth	(iv) Brisbane	(v) Melbourne
Accuracy	0.79	0.86	0.88	0.82	0.76
F1 score	0.64	0.76	0.73	0.66	0.58
ROC AUC	0.78	0.87	0.87	0.82	0.76
Recall	0.76	0.89	0.86	0.81	0.77

TABLE IX. Feature Importance for Sydney

Feature	Sunshine	Wind Gust Speed	Humidity 3pm	Wind Speed 3pm	Max Temp	Pressure 9am	Pressure 3pm
F1 Score	0.04	0.02	0.018	0.017	0.014	0.014	0.013

TABLE X. Feature Importance for Perth Airport City

Feature	Pressure 3pm	Pressure 9am	Wind Gust Speed	Evaporation	Wind Speed 3pm	Sunshine	Rainfall
F1 Score	0.06	0.04	0.03	0.021	0.014	0.013	0.01

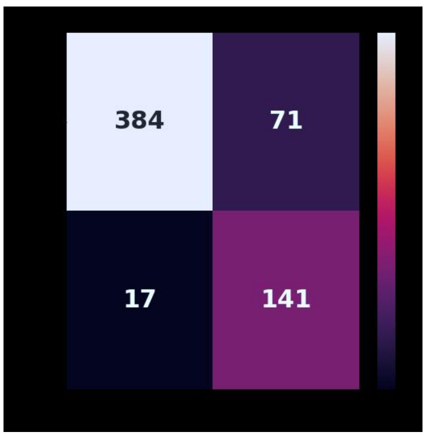

Fig. 16. Darwin next day prediction result.

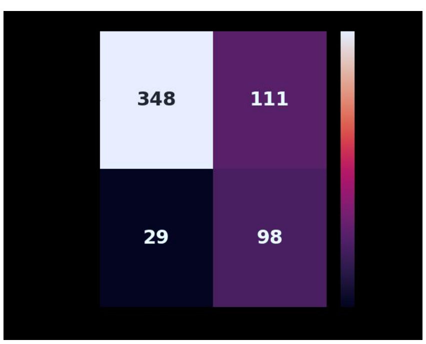

Fig. 17. Melbourne next day prediction result.

clouds, maximum temperature, 9 a.m., minimum temperature, and pressure at 3 p.m. Table IX outlines the F1 score-related elements. Table XI provides Feature Importance for Brisbane. All other features have importance less than 1 percent.

4) Estimation of feature importance for Darwin city: The most essential factors causing rain next day in Darwin are Humidity 3pm, Wind Gust Speed, sunshine,Temp 3pm. Table 9d describe some important attribute with F1 score. Table XII provides feature importance for Darwin. All other features have importance less than 1 percent (Pressure 3pm, Min Temp, Rainfall, Wind Speed 9 am, Wind speed) (see Fig. 16).

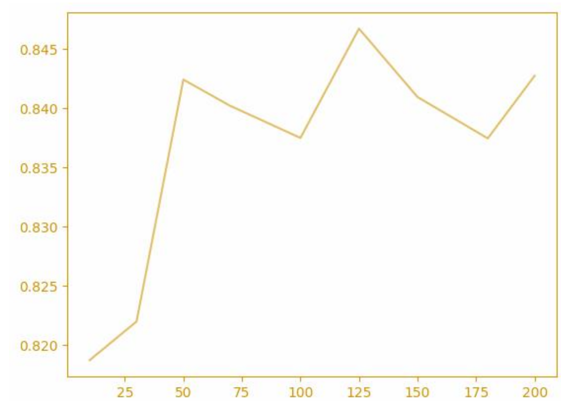

Fig. 18. Darwin ROC curve representation.

5) Estimation of feature importance for Melbourne airport city: Through extensive research, three notable factors in determining the rain tomorrows forecast for Melbourne (Airport) – Humidity 3pm, Sunshine, Pressure 3pm – were discovered. Table IX describe the f1 score of these important features. Table XIII provides Feature Importance for Melbourne Airport. All other features have importance less than 1 percent (see Fig. 17).

E. Ranking of the Most Significant Elements for Various Locations

After modeling, feature importance was analyzed, and the feature importance boxplot shows that certain models with poor outcomes might have been over fitted. Table XIV provides ranking of the most important factors for different regions. The table further shows that the most important variables at various locations are air pressure, air humidity, wind gust speed, and wind speed.

It seems like a good idea to gather more information on these qualities for each area and to keep better track of variables like clouds, sunshine, temperature, and so forth depending on the location. Fig. 18 to 22 represent the curve result of the cities.

TABLE XI. Feature Importance for Brisbane

Feature	Humidity 3pm	Sunshine	Humidity 9pm	Cloud 3pm	Max Temp	Temp 9am	Min Temp	Pressure 3pm
F1 Score	0.07	0.023	0.016	0.016	0.013	0.012	0.012	0.011

TABLE XII. Feature Importance for Darwin

Feature	Humidity 3pm	Wind Gust Speed	Sunshine	Temp 3pm
F1 Score	0.037	0.03	0.0185	0.01

TABLE XIII. Feature Importance for Melbourne Airport

Feature	Humidity 3pm	Sunshine	Pressure 3pm	Cloud 9am	Pressure 9am	Wind Gust Speed	Wind Speed 3pm	Temp 3pm	Max Temp
F1 Score	0.036	0.03	0.025	0.019	0.017	0.016	0.014	0.01	0.01

TABLE XIV. Ranking of the Most Important Factors for Different Regions

Rank	Sydney	Darwin	Perth Airport	Brisbane	Melbourne Airport
1	Sunshine	Humidity 3pm	Pressure 3pm	Humidity 3pm	Humidity 3pm
2	Wind Gust Speed	Wind Gust Speed	Pressure 9am	Sunshine	Sunshine
3	Humidity 3pm	Sunshine	Wind Gust Speed	Humidity 9am	Pressure 3pm
4	Wind Speed 3pm	Temp 3pm	Evaporation	Cloud 3pm	Cloud 9am
5	Humidity 9am	Pressure 3pm	Wind Speed 3pm	Max Temp	Pressure 9am

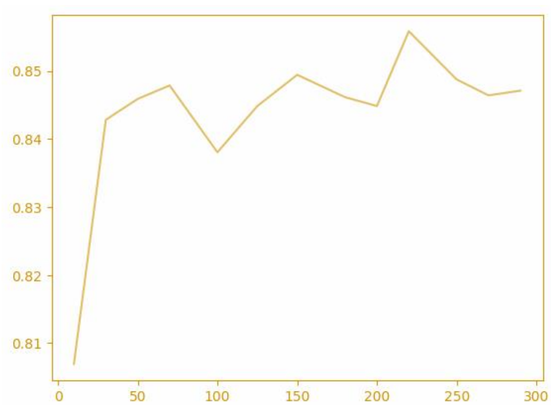

Fig. 19. Perth Airport ROC curve representation.

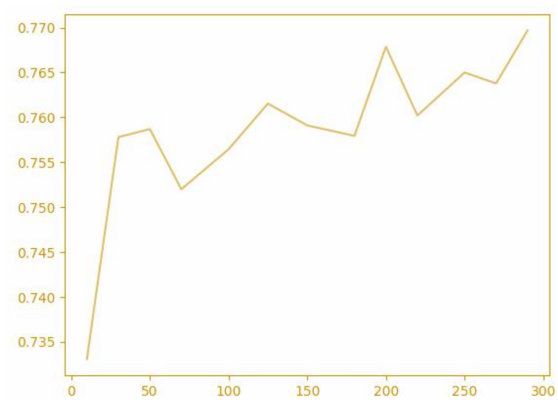

Fig. 21. Melbourne airport ROC curve representation.

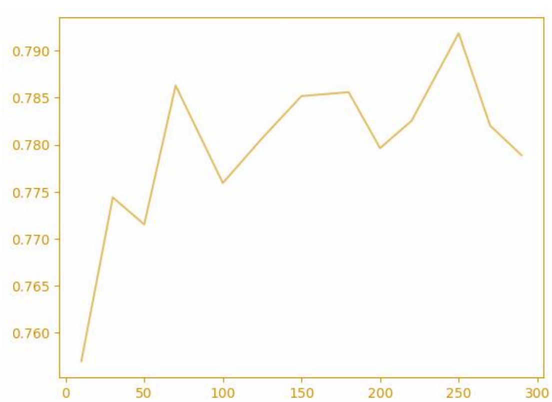

Fig. 20. Brisbane ROC curve representation.

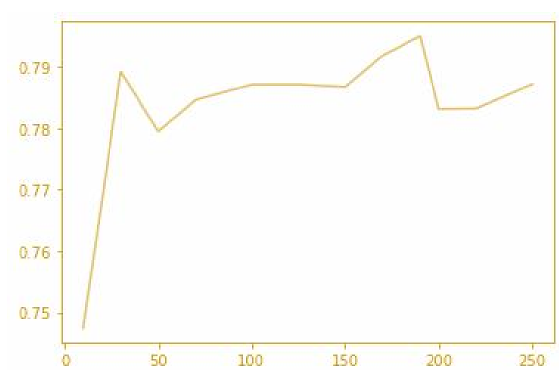

Fig. 22. Sydney ROC curve representation.

```
Model: "sequential_17"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 simple_rnn_14 (SimpleRNN)   (None, 100)               10200

 dropout_14 (Dropout)        (None, 100)               0

 dense_14 (Dense)            (None, 1)                 101

=================================================================
Total params: 10,301
Trainable params: 10,301
Non-trainable params: 0
_________________________________________________________________
```

Fig. 23. RNN Model configuration.

F. Perth Airport, Brisbane, Melbourne, Darwin Accuracy using Recurrent Neural Network

Fig. 23 presents the default architecture of the RNN network that was applied. The table shows the RNN model efficacy in Table XII below. The parameters are displayed in Table XV for the Creation of DL model, as shown below. This table provides parameter value used for training.

TABLE XV. PARAMETER VALUES USED FOR TRAINING

Parameter	Values
Epochs	150
Batch Size	256
Learning Rate	0.0003
Optimizer	Adam
Loss	Binary Cross Entropy
Metrics	Accuracy

This study makes use of a unique, three-layer ANN model. Assess and assemble the model using 150 epochs. Adam Optimization used in this experiment with a batch size of 256. Table XVI provides five cities Result by using RNN and compares Darwin's validation loss to others and shows that it is better. And the graph shows that loss in training and testing both rapidly lowers as the number of epochs rises. Fig. 24 to Fig. 28 shows the loss over iterations for RNN model of five city of Australia (Sydney, Darwin, Perth airport, Brisbane, Melbourne airport.

V. COMPARISON OF RESULTS WITH EXISTING FLOOD AND RAINFALL PREDICTION METHODS

Our applied model outperformed several well-known algorithms across various locations, showing notable accuracy gains. For example, the Random Forest model used in our study achieved an accuracy of 0.83 for Sydney, which is higher than the 0.78 accuracy reported in other studies using comparable datasets. Additionally, we obtained a validation loss of 0.5523 for Sydney with our Recurrent Neural Network (RNN) methodology, which is significantly lower than the losses reported in previous approaches, ranging from 0.63 to 0.71. Furthermore, our proposed machine learning approach not only matched but also exceeded the results of prior studies on flood prediction in Bangladesh, where logistic regression had an accuracy of 0.8676. Previous studies had primarily employed advanced techniques like reweighting and undersampling to enhance performance on imbalanced datasets. Overall, the results of our study demonstrate a significant improvement in the ability to predict floods and rainfall, proving the efficacy of our approach compared to earlier research.

VI. NOVELTY OF THE PROPOSED METHODOLOGY

The proposed method is novel as it combines the structural assignment method with the use of a spin-glass model. This work introduces a Rainfall Forecasting Model that leverages state-of-the-art artificial intelligence and machine learning techniques, specifically employing Recurrent Neural Networks alongside advanced machine learning algorithms. The novelty of our approach lies in the following aspects:

A. Integration of RNNs

Unlike most other studies that primarily focus on classical machine learning approaches or simpler RNN structures, this study leverages RNNs. This choice enables the model to capture temporal dependencies present in rainfall data. RNNs are particularly useful for processing sequential data, making them well-suited for forecasting models where past observations strongly influence future values.

B. Comparative Analysis with Diverse Algorithms

In addition to comparing it with other deep learning models, we also evaluate RNN against traditional machine learning methods, including Random Forest (RF), Decision Tree (DT), and Gradient Boosting Classifier (GBC). This comprehensive cross-comparison reveals the relative advantages and disadvantages of each method in forecasting rainfall, providing valuable insights into their performance.

C. Feature Importance Ranking

This study not only performs algorithmic comparisons but also provides a combined analysis and ranking of key meteorological factors across various regions. This analysis helps explain why, despite similar overall causes, different variables influence the amount of rainfall in various cities across Australia.

D. Undersampling with Reweighting

We adopt undersampling with reweighting procedures to address the class imbalance problem in our dataset, which is common in precipitation forecasting. This method improves the accuracy of our predictions by ensuring that minority classes, such as instances of high rainfall occurrence, are given adequate weight and not overlooked due to their rarity.

E. Improved Accuracy and Reliability

In this study, after addressing the class imbalance, this approach improves prediction accuracy by ensuring that minority classes, such as cases of high rainfall, receive appropriate weight during model training. As a result, rare but significant occurrences are not overlooked, leading to more reliable and accurate rainfall predictions.

TABLE XVI. FIVE CITIES RESULT USING RNN

Metric	(a) Sydney	(b) Darwin	(c) Brisbane	(d) Melbourne	(e) Perth Airport
Loss	0.8106	0.6433	0.7182	1.2848	0.8355
Accuracy	0.6740	0.6553	0.5827	0.5157	0.6165
Val_Loss	0.5523	0.5156	0.6429	0.6482	0.6851
Val_Accuracy	0.7296	0.7740	0.6322	0.6217	0.5712

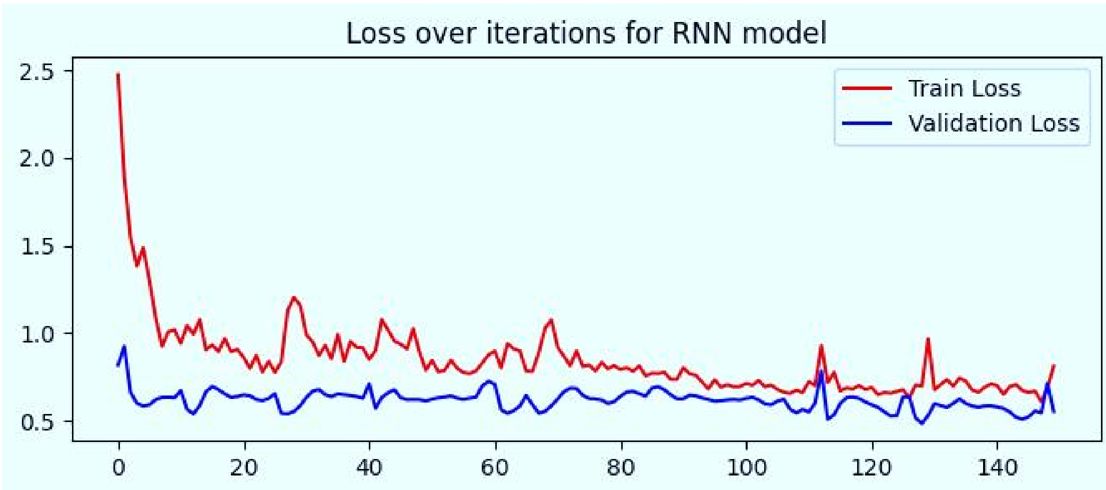

Fig. 24. Loss over iterations for Sydney city.

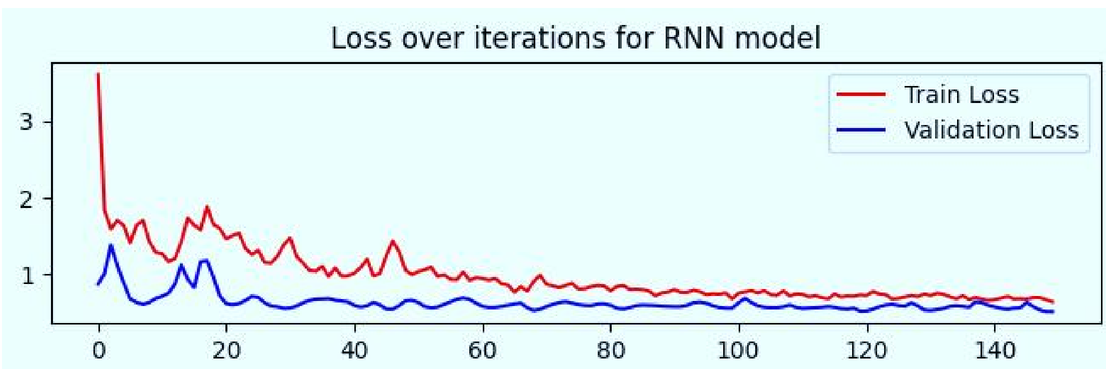

Fig. 25. Loss over iterations for Darwin city.

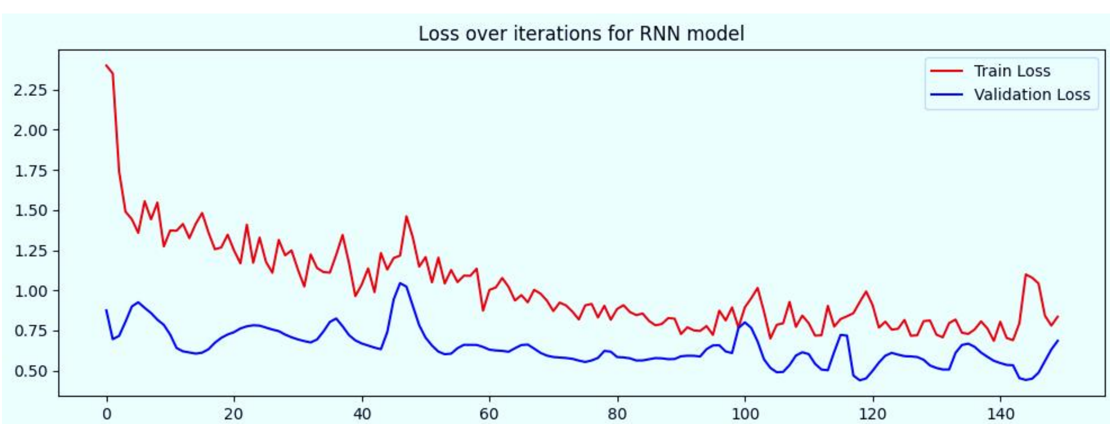

Fig. 26. Loss over iterations for Perth Airport city.

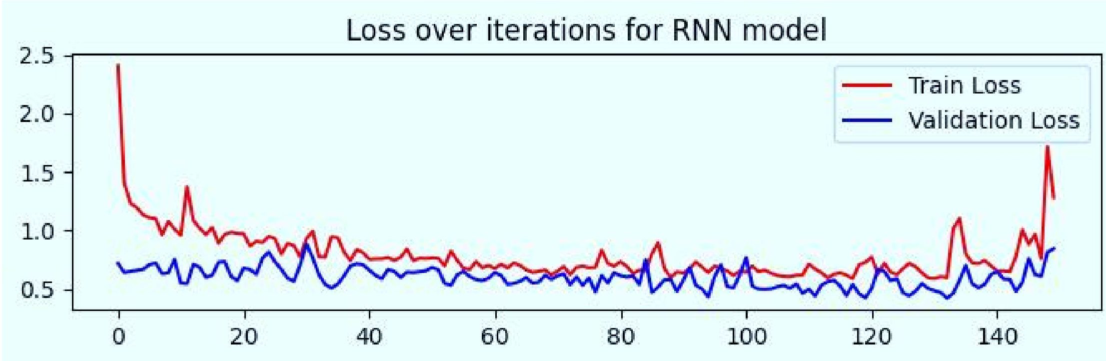

Fig. 27. loss over iterations for Brisbane city

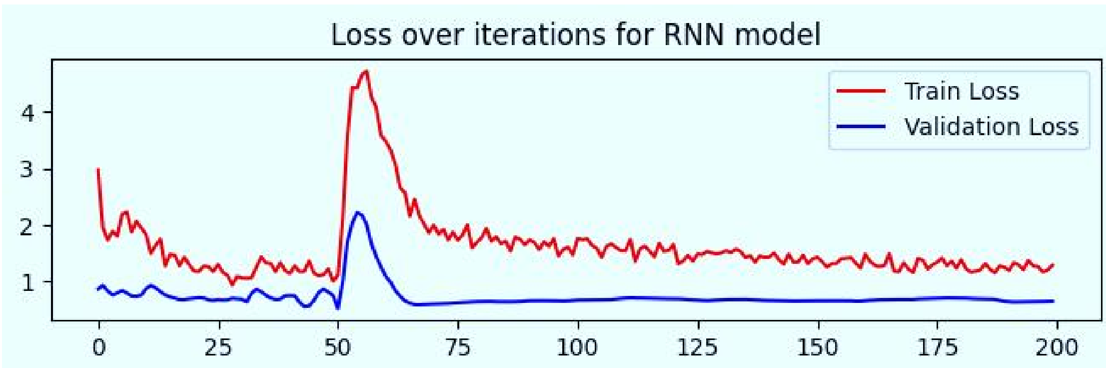

Fig. 28. Loss over iterations for Melbourne airport city.

VII. Conclusion

Estimating rainfall is important for managing water supplies, preserving human life, and protecting the environment. Because geographic and regional factors and changes have an impact on rainfall estimation, problems with inaccurate or insufficient estimation may arise. In this research work, data analytics are used in the field of weather prediction. The research will analyze the effectiveness of machine learning and deep learning methods in addressing the issue of precipitation forecasting, which is confined to Australia, The study's predicted variable is "rain tomorrow. " Several machine learning-powered models for forecasting, e.g. Random Forest, Neural Networks, were employed to predict rainfall after the datasets were obtained. Moreover, the paper vividly explains how machine learning algorithms, unlike neural networks, can accurately imitate the nonlinear nature of natural processes. Finally, the algorithms work and are more successful when the data is broken down by city, which makes it possible to understand how the phenomenon is localized. There are numerous ways to continue the task. Therefore, it would be interesting to examine the outcomes of the study of data from various nations as well as the weather observations from 2019 to the current. It would be a good idea to gather more information on these qualities for each region and to keep better track of variables like clouds, sunshine, temperature, and so forth depending on the location. In this study, that many AI models developed, more specifically deep learning convolutional neural networks, have performed better than traditional machine learning models due to their high level of prediction accuracy and robustness. This is due to the models' ability to recognize complicated patterns and dependencies in the input data used for rainfall prediction models. This work also revealed the potential of feature engineering and data preprocessing strategies in improving rainfall prediction model performance. By retaining relevant input characteristics to carefully handle missing values, outliers and temporal dependencies in the data, we could improve the data predictive power of these models.To conclude, the discovered results are an extension to the prevailing knowledge on weather forecasting and provide an insight into the relevance of machine learning and deep learning techniques in environmental solutions. This research achieves improvements in the fields of climate modeling, disaster planning, agriculture, and water resource management, which all demand high precision in rain forecasts, with the purpose of risk assessment and decision-making.

VIII. Future Work

The measures that can be taken in the future include hyperparameter adjustment to increase model accuracy, live dataset prediction, and forecasting rainfall several days in advance. Fine tuning hyperparameters including the learning rate, number of layers, and activation functions would increase the precision and robustness of employed models. Incorporating live or real-time datasets into prediction systems will enable models to adapt dynamically to changing weather conditions, thereby improving responsiveness and reliability. Extending the forecast window to predict rainfall several days in advance could be valuable for disaster preparedness, agricultural planning, and water resource management. Together, these improvements can contribute to building a more adaptive and proactive rainfall prediction framework.

ACKNOWLEDGMENT

This work is supported by the Deanship of Research, Islamic University Madinah.

IX. CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] T. Luo, A. Maddocks, C. Iceland, P. Ward, and H. Winsemius, "World's 15 countries with the most people exposed to river floods," *World Resources Institute*, vol. 5, 2015.

[2] S. Sankaranarayanan, M. Prabhakar, S. Satish, P. Jain, A. Ramprasad, and A. Krishnan, "Flood prediction based on weather parameters using deep learning," *Journal of Water and Climate Change*, vol. 11, no. 4, pp. 1766–1783, 2019.

[3] F. A. Ruslan, A. M. Samad, Z. M. Zain, and R. Adnan, "5 hours flood prediction modeling using nnarx structure: Case study kuala lumpur," in *2014 IEEE 4th International Conference on System Engineering and Technology (ICSET)*, 2014, pp. 1–5.

[4] S. Talukdar, B. Ghose, Shahfahad, R. Salam, S. Mahato, Q. B. Pham, N. T. T. Linh, R. Costache, and M. Avand, "Flood susceptibility modeling in teesta river basin, bangladesh using novel ensembles of bagging algorithms," *Stochastic Environmental Research and Risk Assessment*, vol. 34, pp. 2277–2300, 2020.

[5] N. J. Ria, J. F. Ani, M. Islam, and A. K. M. Masum, "Standardization of rainfall prediction in bangladesh using machine learning approach," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2021, pp. 1–6.

[6] M. M. A. Syeed, M. Farzana, I. Namir, I. Ishrar, M. H. Nushra, and T. Rahman, "Flood prediction using machine learning models," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2022, pp. 1–5.

[7] S. A. Osmani, J.-S. Kim, C. Jun, M. W. Sumon, J. Baik, and J. Lee, "Prediction of monthly dry days with machine learning algorithms: a case study in northern bangladesh," *Unpublished*, 2022.

[8] A. Manandhar, A. Fischer, D. J. Bradley, M. Salehin, M. S. Islam, R. Hope, and D. A. Clifton, "Machine learning to evaluate impacts of flood protection in bangladesh," *Water*, vol. 12, p. 483, 2020.

[9] M. Goto, F. Cheros, N. A. Haron, and M. N. M. Nawi, "Evaluation of machine learning approach in flood prediction scenarios and its input parameters: A systematic review," in *IOP Conference Series: Earth and Environmental Science*, vol. 498, 2020, p. 012001.

[10] F. M. Aswad, A. N. Kareem, A. M. Khudhur, B. A. Khalaf, and S. A. Mostafa, "Tree-based machine learning algorithms in the internet of things environment for multivariate flood status prediction," *Journal of Intelligent Systems*, vol. 30, pp. 1–16, 2021.

[11] E. H. Ighile, H. Shirakawa, and H. Tanikawa, "Application of gis and machine learning to predict flood areas in nigeria," *Sustainability*, vol. 14, pp. 1023–1042, 2022.

[12] K. Kunverji, K. Shah, and N. Shah, "A flood prediction system developed using various machine learning algorithms," in *Proceedings of the 4th International Conference on Advances in Science and Technology (ICAST 2021)*, 2021, pp. 145–152.

[13] E. Dodangeh, B. Choubin, A. N. Eigdir, N. Nabipour, M. Panahi, S. Shamshirband, and A. Mosavi, "Integrated machine learning methods with resampling algorithms for flood susceptibility prediction," *Science of The Total Environment*, vol. 744, pp. 140–150, 2020.

[14] N. M. Khairudin, N. Mustapha, T. N. M. Aris, and M. Zolkepli, "A study to investigate the effect of different time-series scales towards flood forecasting using machine learning," *Journal of Theoretical and Applied Information Technology*, vol. 99, pp. 2582–2592, 2021.

[15] F. Y. Dtissibe, A. A. A. Ari, C. Titouna, O. Thiare, and A. M. Gueroui, "Flood forecasting based on an artificial neural network scheme," *Natural Hazards*, vol. 102, pp. 857–876, 2020.

[16] A. Sarasa-Cabezuelo, "Prediction of rainfall in australia using machine learning," *Information*, vol. 13, pp. 98–111, 2022.

[17] C. M. Liyew and H. A. Melese, "Machine learning techniques to predict daily rainfall amount," *Journal of Big Data*, vol. 8, p. 27, 2021.

[18] A. not provided in your list, "Flood risk analysis using gis-based analytical hierarchy process: a case study of bitlis province," *Journal of Environmental Management*, vol. 305, p. 114379, 2022.

[19] N. A. Maspo, A. N. B. Harun, M. Goto, F. Cheros, N. A. Haron, and M. N. M. Nawi, "Evaluation of machine learning approach in flood prediction scenarios and its input parameters: A systematic review," in *IOP Conference Series: Earth and Environmental Science*, vol. 479, no. 1. IOP Publishing, 2020, p. 012038.

[20] L. Chen, Y. Li, and Z. Cheng, "An overview of research and forecasting on rainfall associated with landfalling tropical cyclones," *Advances in Atmospheric Sciences*, vol. 27, pp. 967–976, 2010.

[21] Y. L. Lin, S. Chiao, T. A. Wang, M. L. Kaplan, and R. P. Weglarz, "Some common ingredients for heavy orographic rainfall," *Weather and Forecasting*, vol. 16, no. 6, pp. 633–660, 2001.

[22] V. Gude, S. Corns, and S. Long, "Flood prediction and uncertainty estimation using deep learning," *Water*, 2020.

[23] D. T. Nguyen and S.-T. Chen, "Real-time probabilistic flood forecasting using multiple machine learning methods," *Water*, vol. 12, no. 3, p. 787, 2020.

[24] G. Furquim, G. Pessin, B. S. Faiçal, E. M. Mendiondo, and J. Ueyama, "Improving the accuracy of a flood forecasting model by means of machine learning and chaos theory," in *EANN 2015*, 2015.

[25] P. Mitra, R. Ray, R. Chatterjee, R. Basu, P. Saha, S. Raha, R. Barman, and S. Raha, "Flood forecasting using internet of things and artificial neural networks," in *2016 IEEE Annual India Conference (INDICON)*, 2016, pp. 1–6.

[26] J. Noymanee, N. O. Nikitin, and A. V. Kalyuzhnaya, "Urban pluvial flood forecasting using open data with machine learning techniques in pattani basin," *Procedia Computer Science*, vol. 119, pp. 66–74, 2017.

[27] S. Sankaranarayanan, M. Prabhakar, S. Satish, P. Jain, A. Ramprasad, and A. Krishnan, "Flood prediction based on weather parameters using deep learning," *Journal of Water and Climate Change*, 2020.

[28] A. Y. Barrera-Animas, L. O. Oyedele, M. Bilal, T. D. Akinosho, J. M. D. Delgado, and L. A. Akanbi, "Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting," *Machine Learning with Applications*, vol. 7, p. 100204, 2022.

[29] A. Mosavi, P. Ozturk, and K.-w. Chau, "Flood prediction using machine learning models: Literature review," *Water*, vol. 10, no. 11, p. 1536, 2018.

[30] N.-A. Maspo, A. N. B. Harun, M. Goto, F. Cheros, N. A. Haron, and M. N. M. Nawi, "Evaluation of machine learning approach in flood prediction scenarios and its input parameters: A systematic review," in *IOP Conference Series: Earth and Environmental Science*, vol. 479, 2019, p. 012038.

[31] S. Puttinaovarat and P. Horkaew, "Flood forecasting system based on integrated big and crowdsource data by using machine learning techniques," *IEEE Access*, vol. 8, pp. 130 327–130 344, 2020.

[32] P. Ghorpade, A. Gadge, A. Lende, H. Chordiya, G. Gosavi, A. Mishra, B. Hooli, Y. S. Ingle, and N. Shaikh, "Flood forecasting using machine learning: A review," in *2021 8th International Conference on Smart Computing and Communications (ICSCC)*, 2021, pp. 1–6.

[33] M. S. Gani, A. Zakaria, S. Siam, I. Kabir, Z. Kabir, M. R. Ahmed, Q. K. Hassan, R. M. Rahman, and A. Dewan, "A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction," *Journal of Environmental Management*, vol. 326, Part B, p. 116813, 2023.

[34] N. Gauhar, S. Das, and K. S. Moury, "Prediction of flood in bangladesh using k-nearest neighbors algorithm," in *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, 2021, pp. 1–5.

[35] S. H. P. Coulibaly, "Bayesian flood forecasting methods: A review," *Journal of Hydrology*, vol. 551, pp. 340–351, 2017.

[36] F. Faisal, "Artificial intelligence for flood prediction and management: Lessons for pakistan," *ISSI*, 2022.

[37] A. Parmar, K. Mistree, and M. Sompura, "Machine learning techniques for rainfall prediction: A review," in *2017 International Conference on Innovations in Information Embedded and Communication Systems (ICIIECS)*, 2017, pp. 1–5.

[38] E. Morin, "Flood forecasting with machine learning models in an operational framework," *Hydrology and Earth System Sciences*, vol. 25, pp. 3671–3685, 2021.

[39] K. Khosravi, H. Shahabi, B. T. Pham, J. Adamowski, A. Shirzadi, B. Pradhan, J. Dou, H.-B. Ly, G. Gróf, H. L. Ho, H. Hong, K. Chapi, and I. Prakash, "A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods," *Science of The Total Environment*, vol. 658, pp. 573–590, 2019.

[40] G. Tayfur, V. P. Singh, T. Moramarco, and S. Barbetta, "Flood hydrograph prediction using machine learning methods," *Water*, vol. 10, no. 10, p. 1436, 2018.

[41] A. Sahoo, S. Samantaray, and D. K. Ghose, "Prediction of flood in barak river using hybrid machine learning approaches: A case study," *Journal of the Geological Society of India*, vol. 97, pp. 1077–1085, 2021.

[42] K. Qian, A. Mohamed, and C. Claudel, "Physics informed data driven model for flood prediction: Application of deep learning in prediction of urban flood development," *Arxiv*, vol. 1907.11818, pp. 1–8, 2019.

[43] M. S. G. Adnan, Z. S. Siam, I. Kabir, Z. Kabir, M. R. Ahmed, Q. K. Hassan, R. M. Rahman, and A. Dewan, "A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction," *Journal of Environmental Management*, vol. 326, p. 116813, 2023.

[44] S. Miau and W.-H. Hung, "River flooding forecasting and anomaly detection based on deep learning," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, pp. 105–111, 2020.

[45] I. Hossain, H. Rasel, M. A. Imteaz, and F. Mekanik, "Long term seasonal rainfall forecasting using linear and non linear modeling approaches; a case study for western australia," *Meteorology and Atmospheric Physics*, vol. 131, pp. 1221–1231, 2019.

[46] D. Han, L. Chan, and N. Zhu, "Flood forecasting using support vector machines," *Journal of Hydroinformatics*, vol. 9, pp. 267–276, 2007.

[47] A. Rajab, H. Farman, N. Islam, D. Syed, M. A. Elmagzoub, A. Shaikh, and M. Alrizq, "Flood forecasting by using machine learning: A study leveraging historic climatic records of bangladesh," *Water*, vol. 15, no. 22, p. 3970, 2023.

[48] B. N. Lakshmi, "A comparative study of classification algorithms for risk prediction in pregnancy," in *Proceedings of the International Conference on Advances in Computing and Communication Engineering (ICACCE)*, 2015, pp. 0–5.

[49] R. Prasetya and A. Ridwan, "Data mining application on weather prediction using classification tree, naive bayes and k-nearest neighbor algorithm with model testing of supervised learning probabilistic brier score, confusion matrix and roc," *Journal of Applied Communication and Information Technology*, vol. 4, no. 2, pp. 25–33, 2019.

[50] N. Mishra, H. K. Soni, S. Sharma, and A. K. Upadhyay, "Development and analysis of artificial neural network models for rainfall prediction by using time-series data," *International Journal of Intelligent Systems and Applications*, vol. 10, no. 1, pp. 16–23, 2018.

[51] F. Yulianto, W. F. Mahmudy, and A. A. Soebroto, "Comparison of regression, support vector regression (svr), and svr-particle swarm optimization (pso) for rainfall forecasting," *Journal of Information Technology and Computer Science*, vol. 5, no. 3, pp. 235–247, 2020.

[52] S. Mohamadi, Z. Sheikh Khozani, M. Ehteram, A. N. Ahmed, and A. El-Shafie, "Rainfall prediction using multiple inclusive models and large climate indices," *Environmental Science and Pollution Research*, vol. 29, no. 56, pp. 85 312–85 349, 2022.

[53] K. V. Rajkumar and K. Subrahmanyam, "A novel method for rainfall prediction and classification using neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, pp. 250–259, 2021.

[54] N. S. Sani, A. H. Abd Rahman, A. Adam, I. Shlash, and M. Aliff, "Ensemble learning for rainfall prediction," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 11, pp. 450–460, 2020.

[55] M. M. Hassan, M. A. T. Rony, M. A. R. Khan, M. M. Hassan, F. Yasmin, A. Nag, and W. El-Shafai, "Machine learning-based rainfall prediction: Unveiling insights and forecasting for improved preparedness," *IEEE Access*, vol. 11, pp. 132 196–132 222, 2023.

[56] M. Motta and M. N. P. S. de Castro, "A mixed approach for urban flood prediction using machine learning and gis," *International Journal of Disaster Risk Reduction*, vol. 56, p. 102154, 2021.

[57] P. A. Ascierto, R. Accorona, G. Botti, D. Farina, P. Fossati, G. Gatta, and V. Vanella, "Mucosal melanoma of the head and neck," *Critical Reviews in Oncology/Hematology*, vol. 112, pp. 136–152, 2017.

[58] Z. R. Chong, S. H. B. Yang, P. Babu, P. Linga, and X. S. Li, "Review of natural gas hydrates as an energy resource: Prospects and challenges," *Applied Energy*, vol. 162, pp. 1633–1652, 2016.

[59] J. Fan and X. Li, "Prediction of the productivity of steam flooding production wells using gray relation analysis and support vector machine," *Journal of Computational Methods in Sciences and Engineering*, vol. 15, no. 3, pp. 499–506, 2015.

[60] M. Mohammed, R. Kolapalli, N. Golla, and S. S. Maturi, "Prediction of rainfall using machine learning techniques," *International Journal of Scientific and Technology Research*, vol. 9, no. 01, pp. 3236–3240, 2020.

[61] M. M. Sanadi, "Rainfall prediction using logistic regression and support vector regression algorithms," in *Advances in Computing and Data Sciences: 5th International Conference, ICACDS 2021, Nashik, India, April 23–24, 2021, Revised Selected Papers, Part I.* Springer International Publishing, 2021, pp. 617–626.

[62] S. Biruntha, B. S. Sowmiya, R. Subashri, and M. Vasanth, "Rainfall prediction using knn and decision tree," in *2022 International Conference on Electronics and Renewable Systems (ICEARS)*. IEEE, March 2022, pp. 1757–1763.

[63] V. S. Monego, J. A. Anochi, and H. F. de Campos Velho, "South america seasonal precipitation prediction by gradient-boosting machine-learning approach," *Atmosphere*, vol. 13, no. 2, p. 243, 2022.

[64] M. Badawy, A. A. Abd El-Aziz, A. M. Idress, H. Hefny, and S. Hossam, "A survey on exploring key performance indicators," *Future Computing and Informatics Journal*, vol. 1, pp. 47–52, 2016.