Grenze International Journal of Engineering and Technology, January Issue

A Systematic Review on Weather Forecasting using Machine Learning Techniques

Katta Trinadha Ravi Kumar[1], Dr P Suresh Varma[2] and Dr M V Rama Sundari[3]
[1]Research Scholar,Dept of CSE, Adikavi Nannaya University,Rajamahendravaram,&Assoc Professor,Dept of Computer Science,SVKP&DrKSRaju Arts & Science College(A),Penugonda-534320
Email: trinadhaknu9@gmail.com
[2]Professor, Department of CSE, Adikavi Nannaya University, Rajamahendravaram, AP, India
Email: sureshvarmap@gmail.com
[3]Professor, Dept. of AIML, Gokaraju Ranga Raju Institute of Engineering and Technology, Hyderabad
Email: mvramasundari@gmail.com

***Abstract*—Numerical Weather Prediction (NWP) is a typical method for weather forecasting. Though sometimes it might result in unsatisfactory performance because to incorrect initial state setting. A data-driven strategy learns from historical data while integrating prior knowledge from NWP, augmented by an efficient information fusion mechanism. To predict something, a background study is required to understand the pattern. On Earth, every aspect of human life is impacted by nature. Because we cannot avoid the natural changes and conditions, we have chosen to reduce their effect on our lives. Hence, to achieve this, it is essential to have prior knowledge of the weather conditions to make things work in accordance with the changes in the environment. To perform prediction, an accurate classification of data is required. The development of a weather prediction model is the main goal of this paper. Agriculture is the field that is most influenced by the weather. Therefore, it can expand the scope to provide regional guidelines to farmers based on the conducted classification.**

***Index Terms*— Weather Forecasting, Cloud, Mobile, Collaborative Real-Time, Machine Learning.**

I. INTRODUCTION

In recent years, the world has seen quickly changing environmental conditions. All the primary sectors are influenced by weather conditions. Due to the sudden changes that might happen, environmental change is given a lot of consideration [11]. Therefore, weather forecasts are crucial. Weather forecasting involves the prediction of atmospheric conditions at a designated location and time in the future. Weather forecasting has a crucial role in various domains. Weather forecasting holds a significant role in the field of meteorology. Traditionally, weather predictions were performed using physical models of the atmosphere. The current atmospheric conditions are observed, and the future state is calculated by solving the fluid dynamics and thermodynamics equations numerically. In the presence of perturbations and uncertainties in the initial weather measurements, the system of ordinary differential equations governing this physical model becomes unstable [12]. However, it is possible to make it simpler and more reliable by using available technologies.
Machine learning can help in weather forecasting or prediction because it is more robust and does not necessarily require a clear understanding of the physical forecasting process, as stated in [15]. Hence, in this paper, two

Grenze ID: 01.GIJET.11.1.312

machine learning algorithms are implemented for weather prediction. In addition to prediction, weather classification based on the predicted values has also been incorporated. As discussed, predicting the weather plays a critical role in different fields. Agriculture is one of the primary fields where weather prediction can be highly beneficial. Therefore, the scope of the following paper is expanded for the application of this forecasting in agriculture. Based on the outcomes of the weather prediction model, essential guidelines will be generated for the farmers in that area. This will help them be aware of the forecast and take the necessary actions.

Meteorological elements, such as humidity, temperature, and wind, have a significant impact on various aspects of human livelihood. These elements provide analytical support for urban computing-related issues such as electric power generation planning, air quality analysis, and traffic flow prediction. In meteorology, the most general approach for predicting the weather is using physical models to simulate and forecast weather patterns. This is referred to as NWP, or numerical weather prediction. NWP offers the benefit of using numerical solutions for atmospheric hydro thermodynamic equations. The careful selection of the initial solution has improved the prediction accuracy. However, the instability of such differential equations may impact the reliability of NWP [18]. As meteorological big data becomes more accessible, researchers have come to understand that the introduction of data-driven methods into meteorology may lead to significant success.

Catastrophic events occur all around the world. Several unsafe events for human progress include waves, tornadoes, volcanic eruptions, seismic tremors, tropical storms, and extreme rainfalls. These are regarded as a few instances of destruction and demise, which have the potential to impact human progress. Exploring a method to visualize the impact of such natural phenomena is crucial importance for issuing warnings, mitigating disasters, and saving lives. In financial-related practices, investors also face similar challenges that restrict the trading of precipitation derivatives in financial markets. During the high temperature, high insolation, high stability (resulting in low mixing heights), low late morning relative humidity, and late spring, photochemical reactions occur that exhaust the cloud.

Several machine learning methods were utilized in weather forecasting [16]. Data-driven approaches have the benefit of rapidly modeling patterns through learning, thus removing the necessity for solving complex differential equations. However, the next difficulty arises from getting satisfactory performance by learning from previous observations, which requires large data and tedious feature engineering. To create a more effective and efficient solution, it combine the benefits of machine learning and NWP [17]. Simultaneously, single-value forecasting (also known as point estimate) lacks adaptability and reliability for a number of human decision. This paper's goal is to present a unified deep learning approach for addressing these problems through end-to-end learning. In particular, it will predict a range of meteorological variables at a range of weather stations for a wide range of future actions. The presented approach offers numerous benefits, including high accuracy, end-to-end learning, efficient data pre-processing, uncertainty quantification, and ease of deployment, each of which contribute to its considerable practical significance.

The goal of weather forecasting is to predict the atmospheric conditions at a specific location for a defined forecast period. Weather is defined as fluctuations of atmospheric factors like sunlight, temperature, humidity, wind direction, speed, precipitation and cloud cover. Due to the arrival of the Internet of Things (IoT), acquiring real-time weather forecasts has become possible. Weather data can be collected from various locations by deploying multiple weather sensors at a low cost. The gathered data can subsequently be processed using machine learning algorithms to achieve highly accurate predictions. Along with IoT, commercial cloud platforms' Infrastructure as a Service (IaaS) model is a suitable alternative for applications like weather forecasting with irregular computational resources. The potential for achieving cost-effectiveness can be realized by executing applications with bursty computational demands on rented high-end servers, as opposed to purchasing and maintaining specific hardware. because cloud customers only pay for the time that the resources were really used. A number of researchers have recently focused their efforts on the use of IoT, cloud computing, and machine learning for the creation of real-time weather forecasting systems.

The prediction of rainfall helps ensure proper aviation services and provides support in agricultural planning to facilitate the necessary arrangements for procuring, transporting, and distributing food grains during periods of below-normal rainfall. Machine learning offers a significant advantage in forecasting future weather conditions, thereby allowing for the implementation of preventive measures. While several techniques, such as satellite observation, ground observation, observations made from aircraft, radar, and ships are used by meteorological services to collect rainfall data [19]. The gathered datasets are organized by the meteorological department into graphs, charts, and sheets. However, the data was not properly processed and utilized by the Ethiopian National Meteorological Agency. The goal of this paper is to identify patterns within the dataset and apply them for predictive purposes by utilizing both conceptual machine learning and machine learning. Conversely, to create a seasonal prediction for Ethiopia, the Ethiopian Meteorological Agency consults a wide range of information sources abroad.

Predicting the air temperature accurately at any specific location and time is indeed a significant research challenge, with various types of uses from energy production to agriculture. According to climate experts, rising air temperatures might have a negative impact on the environment in the next decades. At present, accurately predicting air temperature is a challenging task when utilizing data-driven methods due to its inclusion in the chaotic and multidimensional weather system. In recent times, the difficulties in predicting the weather have increased due to the wide availability of weather data. To address this, a strong prediction model needs to be created that can find hidden patterns in the data. Additionally, it involves learning the connections between various weather observation variables. Collecting extensive data on various weather variables such as temperature, humidity, air pressure, etc., are essential for making accurate forecasts and gaining insights into the long-term evolution of the atmosphere. This is traditionally been achieved using computer models, specifically the Numerical Weather Prediction (NWP) models, based on physical equations. However, a large computational capacity and well-defined prior knowledge are required for these physics-based systems [9]. Deep learning techniques in particular preserve the benefit of identifying hidden patterns in the dataset without the need for prior knowledge. Therefore, in weather forecasting, such approaches may thus provide an appropriate alternative to the physical models that were previously utilized.

II. Literature Survey

Zheng Yan, Bin Wang, Tianrui Li, Huaishao Luo, Yu Zheng, Jie Lu, Guangquan Zhang, et al. [13] described that information fusion methods and motivation might be better understood through data exploration analysis. The modifications made to three specific climatic variables during the previous three years. Only temperature is seen to exhibit a significant seasonal change, whereas wind speed and relative humidity are exposed to significant noise. Therefore, considering the highly dynamic nature of weather changes, based on these observations, seasonal feature extraction from historical data won't produce the best outcomes. The frequent concept drift results in the loss of value for long-term historical meteorological data. It may be said that the conclusion is "For many time series tasks, only a few recent steps are necessary." In contrast, NWP is a relatively reliable forecasting technique, but wrong states might lead to undesired error bias. They suggest a balanced fusion mechanism to address.

• To begin with, modelling current meteorological dynamics should only use recent observations, i.e., Estimated Time En Route (ETE).

• Next, to make it simple to rectify bias in NWP, a smart NWP fusion technique should include NWP forecasting throughout the relevant forecasting time step.

However, an unwise fusion method that is not thoroughly considered may include NWP that is not useful in capturing important NWP signals.

Therefore, NWP forecasting is incorporated into Day-To-Dday (DTD), instead of being included in ETE or hidden coding. It aggregates historical statistics for 10 stations, from 3:00 afternoon to 15:00 (Universal Time Coordinated (UTC))), with the mean (solid line) and 90% confidence interval (shaded region). We observed the following: 1) The variance and mean statistics differ noticeably; for example, the mean value of station-ID 7 has a different trend than the other stations. 2) The variance and mean of the climatic characteristics are different for each hour at each site.

To train the prediction model, they developed a new loss function called Negative Log-likelihood Error (NLE). This approach has the capability to simultaneously infer prediction interval and sequential point estimation. An interesting experimental finding in this research is that training with the NLE loss significantly improves the generalization of point estimation. This might provide practitioners with new insights on the creation and application of learning algorithms for similar problems, including time series regression. The proposed technique obtained state-of-the-art performance on a real-world benchmark dataset for weather forecasting by using an effective deep ensemble strategy.

The complete approach was developed in Keras and has the adaptability needed for implementation in a real-world setting. The source codes and data will be made available, providing a benchmark for practitioners and researchers to explore weather forecasting.

Tulsi Pawan Fowdur, and Rosun Mohammad Nassir-Ud-Diin Ibn Nazir, et.al [3] presented the system model and methods. The whole system model for weather prediction comes first, which includes the web application, local server, cloud-hosted servlet, desktop and mobile edge devices, is described. The appropriate adjustments to traditional machine learning methods are then made in order to facilitate collaborative forecasting. The data flow

throughout the complete system is shown in the following diagram. By including a mobile client with an interactive interface created specifically for weather forecasting in Mauritius, the system builds on the one stated. Additionally, a local server capable of collaborative weather forecasting is substituted for the central local server. It accomplishes by forecasting the weather for a certain region using weather characteristics from several locations. Both the locally hosted servlet and the cloud servlet operate under the same general concept. The system consists of desktop and mobile edge devices that use the OpenWeather Application programming interface (API) to provide queries for real-time weather parameters for a specific region. Either the local server or the cloud receives the parameters directly. The central local server can send the values to a local My Structured Query Language (MySQL) database or an International Business Machines (IBM) Cloudant database. The servlet hosted in the cloud, the servlet hosted locally, or the local server can all be used for analytics. Users may view weather parameters and get weather predictions using a web-based application. In addition, a specialised Android client for Mauritius weather forecasting was created. This client displays a map of the country that is interactive and shows climate predictions that were collected from a servlet hosted in the cloud.
Local server centralised for cooperative weather forecasting. A Java programme was created to implement the centralised local server. It shows a screenshot of the centralised local server's design. The following functions are included in the desktop application:

- It keeps track of incoming connections from edge devices (desktop and Android), logs the data, and then stores the weather information in both a local Structured Query Language (SQL) database and a cloud-based NoSQL database.
- It displays a real-time graph that represents the alterations in several weather factors.
- It downloads information for a certain date from a local or cloud database.
- For a particular sample size, it downloads the most recent data from a local or cloud database.
- It uses five machine learning algorithms to provide predictions for a particular place.
- It carries out cross-validation of time series for different methods.
- For convenient comparison, it displays the cross-validation's Mean Aabsolute Percentage Error (MAPE) in a table and a graph.

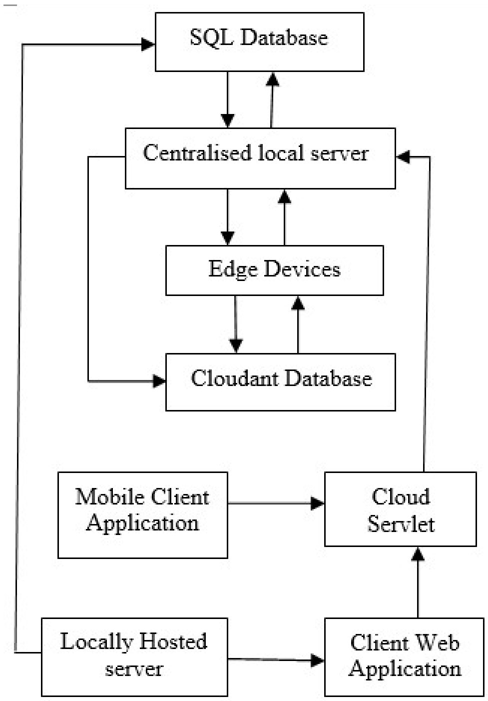

Fig.1: System Data Flow

The user clicks the "Predict" button to start predictions using one of the five potential machine learning algorithms for predicting the weather within a certain time range. The user chooses a date and a sample size before clicking the "Download" button to start the data download. The required information will be obtained from the appropriate source and stored to the local storage in Comma Separated Value (CSV) format. The desktop and mobile applications are represented by the edge devices for weather data collection. These applications are responsible for obtaining information from the OpenWeather API and sending it to the Cloudant database or the local server. The Cloudant database or the local server may also be contacted by the desktop and mobile applications to request data for download. The screen appears visible after starting the mobile client. The screen displays the most recent weather conditions that have been noted. There is also a graph that displays changes in a certain weather state in real time. The user only needs to click on the chosen variable in the second or third row to change the variable being shown. When people click on the servlet link, a webpage called the web application is shown to them. The website displays a graph depicting real-time changes in weather factors. Responding to requests from end users visiting the webpage is the main responsibility of the cloud-hosted servlet. The locally hosted servlet does the same tasks as the cloud-hosted servlet, except it runs on a personal Apache Tomcat server rather than the IBM Cloud infrastructure. To create a connection with the servlet, the user must supply the server's IP address. Data incoming from the desktop and mobile applications will be stored in a locally hosted MySQL database. The data from the local database is utilized for local predictions on the server, eliminating the necessity to interact with the cloud-hosted servlet. The Cloudant database, the system uses the IBM Cloud platform, which is used to store weather data on the cloud, providing global access to the information. To ensure that the information is kept in the Cloudant database, it is essential to incorporate a new Cloudant resource into the lists of IBM Cloud Resources. This establishes a Cloudant system that permits users to establish as many databases as required.

X. Chu, W. Bai, Y. Sun, W. Li, C. Liu, and H. Song, et al. [1], proposed a new technique for determining Earth environmental factors called the Global Navigation Satellite System (GNSS)/Low Earth Orbit Satellite (LEO) Radio Occultation Technology. The Global Navigation Satellite System of Russia, the Galileo system jointly run by China and the European Union (EU), China-based BeiDou Navigation Satellite System, and the Global Positioning System (GPS) are the four main global navigation satellite systems. As global navigation has improved, radio occultation technology has undergone further development.

The largest and most dynamic area of the Bohai Rim in Northeast China is located in the Beijing-Tianjin-Hebei. The Beijing, Tianjin, and Hebei area is essential for China's economic development. The Beijing-Tianjin-Hebei region's growth and people quality of life, have been impacted by the weather and pollution in the area. As a result, it is crucial to forecast the wind patterns in the Beijing-Tianjin-Hebei region since they have a big impact on the weather and air quality. Historical European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA)-Interim data were used in this study. Using a computer model, these data were used to estimate the wind patterns in the Beijing-Tianjin-Hebei area from June 2018 to July 2019. The European Centre for Medium-Range Weather Forecasts provided the first data. The dataset obtained from the numerical model was then used to build machine learning models. As inputs for the neural network, the programme obtains the temperature, pressure, and humidity data from cosmic occultation data. The wind field in the Beijing-Tianjin-Hebei region is then predicted using the machine learning model with the best performance among the several models.

After training and testing, the best machine learning model may be applied to provide useful predictions. When making actual predictions, the model uses data from a different source than during training. When the model is used for practical forecasting, GNSS-RO (Radio Occultation) data is utilised for prediction because it offers the benefits of high accuracy and all-weather in presenting worldwide uniform distribution, as well as high vertical resolution humidity, pressure profiles, and temperature. It is significant to highlight that the analysis of the temperature, pressure, and humidity data from GNSS-RO uses ERA-Interim data as a background field. Transfer learning is the process of building an effective model and using the rich data it generates as the initial training data for a machine learning model. The method is to build machine learning models utilizing a suitable quantity of data for some cases it may not have much data.

Implementing GNSS-RO data, a machine learning-based method for predicting regional wind fields. Observations made through remote sensing, there are instances as only thermodynamic parameters can be obtained because of the limitations of the instruments carried on satellites. A machine learning method was used to form a connection between kinetic attributes and thermodynamic attributes, enabling the prediction of kinetic attributes using only thermodynamic attributes, a capability not achievable solely through a numerical model. In terms of predicting the wind field, machine learning performs effectively. The difference between the numerical model and machine learning is quite small. In terms of neural networks, Convolutional Neural Network and Long Short-Term Memory (LSTM) show lower prediction errors than Deep Neural Networks (DNN). The wind speed and direction errors for LSTM and CNN's prediction of the wind field for the next three hours are 300 and 1.4 m/s, respectively. In comparison to DNN, CNN and LSTM exhibit a 57% reduction in wind speed error and a 50% reduction in wind direction error. As the prediction time gets longer, the prediction error becomes larger.

TABLE I: COMPARISON OF ERRORS BETWEEN PREDICTED AND MONITORED VALUES

Model	RMSE(m/s)		MAE(^{0})
	Zonal Wind Speed	Meridional Wind Speed	Wind Direction
WRF	2.3	2.1	26.2
CNN	2.6	2.5	30.5
LSTM	2.7	2.4	31.6

TABLE II: TRAINING TIME AND PREDICTION TIME OF DIFFERENT MODELS

	LSTM	**CNN**	**DNN**	**WRF**
Training Time	21min 3s	35min26s	14min 41s	-
Prediction Time	1.8s	2.2s	1.6s	15min32s
Relative Prediction Time	0.19%	0.24%	0.17%	100%

N. L. and M. H. S., et al. [10], described a technique for predicting weather at the neighborhood scale that involves the management of strategies based on scientific methodologies. Climate predictions were commonly established using practical techniques for constructing direct models. Evaluating variable estimation based on time arrangements is a fundamental responsibility in weather forecasting. However, accuracy and satisfaction in meteorological data were exhausted. Various forms of vulnerabilities and irritations can have an impact on data. More mind-boggling methods were suggested to address this challenge. In order to enhance the accuracy, wavelet change or Kalman channel is added to selected channels. Gathering estimation is another scientific philosophy applied in forecasting. Lot of the predictable capabilities are discovered through unusual situations in parallel rather than by a single predictable function.

However, models need to be parameterized in the context of conventional procedures, although this is not fundamentally necessary in reality. In the field of man-made mental ability, scholarly approaches aim to provide possible outcomes for the automatic modification of parameters. Artificial neural networks (ANN) are particularly suitable for analyzing the time course of actions in the context of forecasting, and they can be found in related research papers. Several proposals exist for enhancing climate prediction systems using different types of NNs, ranging from fundamental MultiLayer Perceptrons (MLP) to progressively complex iterative NNs. These proposals involve a mix of various techniques from the field of artificial intelligence, providing a foundational approach to enhance forecasting accuracy. This includes the remarkable Adaptive Network-Based Fuzzy Inference System (ANFIS) structure and the utilization of evolutionary computation to adjust NN parameters, such as NN with motion estimation or NN with Particle Swarm Optimization (PSO).

Its objective is to change the parameters of the prediction model as a Support Vector Machine (SVM) part of machine learning. Without regard to NN, SVM further needs data preparation for the backslide assessment conducted to choose SVM, the outputs of whose are the parameters previously mentioned. SVMs can be combined with different methods, such as PSO. Finally, determining the hybridization of conventional and subjective procedures for the culmination is also important. These methods can either assign each person for a particular task to light up, or one system can help the settings of another. The majority of the time, NNs can be used with wavelets, Markov chains, and Kalman channels.

The applications of weather prediction have a positive influence in various fields, including water systems, sun-based power determining, agriculture, and more, which were also analyzed. The examination is conducted concerning various execution measurements. In any case, improved hybrid approaches using NN can be carried out to achieve better results.

TABLE III: RMSE MEASURE FOR DIFFERENT METHODOLOGIES

Model	**RMSE**
RBF-NN	146.6148
RBF	92.0242
RBF	56.8353
Gaussian SVMs	57.044
Wavelet SVMs	39.071

R. S. Vejendla, T. Deepika, B. L. Pavuluri, P. Jithendra, and S. Bano, et al. [5], initiate this method by collecting weather data from a specific area. Subsequently, the collected data undergoes data pre-processing, which includes essential steps such as importing the necessary libraries once the weather dataset is imported. Following that, the process involves addressing missing data and noisy data using various techniques. Finally, the dataset is separated into training and testing data for scaling the result. To evaluate the method with the best weather forecast accuracy, run the three algorithms individually on the same weather dataset. When the decision tree algorithm is utilized on the data, the data is categorized into different classes based on specific criteria related to the given temperature data. It ultimately generates the weather prediction outcome as cold, hot, or rainy.
Additionally, the same weather dataset is subjected to the random forest algorithm, which effectively handles the large dataset due to its combination of many decision trees. When constructing individual trees within an uncorrelated forest, this algorithm employs both bagging and feature randomness. When compared to the predictions of individual trees, the resultant predictions are highly accurate. It determines the average forecast of a single tree to provide the weather prediction after taking into consideration all the bootstrap samples that were created.

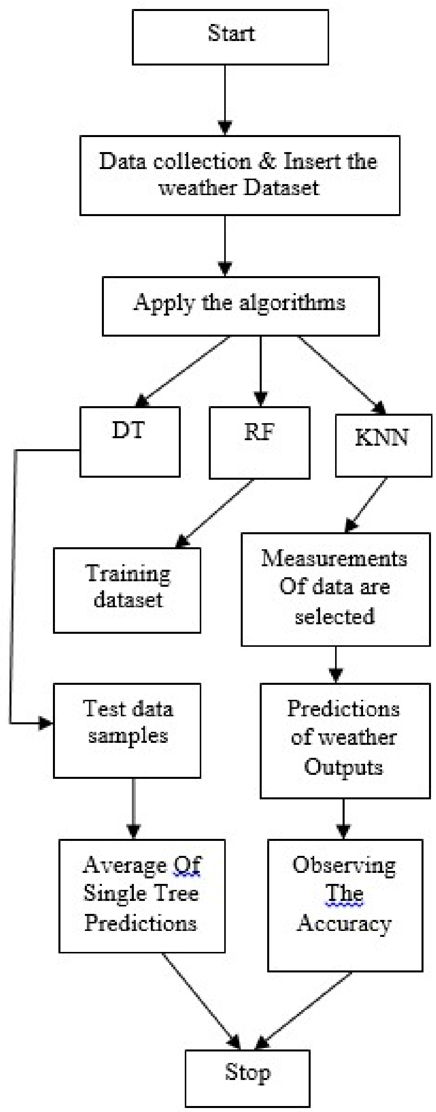

Fig 2: Machine Learning Flowchart

Furthermore, to evaluate the accuracy of alternative algorithms, neural networks were applied alongside k-nearest neighbor. This method involves calculating the distance between all data points and a set of initial k-points to identify any similarities among these points. After processing all data points through the input and hidden layers, the algorithm generates weather predictions as the output layer. Classifying a day as hot, wet, or cold involves the use of machine learning algorithms to predict the weather. Prediction was performed using a number of machine learning techniques, such as Decision Tree (DT), Random Forest (RF), and K-NN using Neural Networks. Using the same meteorological information that includes parameters like minimum temperature, maximum temperature, maximum cold, the forecast accuracy of each algorithm was assessed separately. The main goal of taking three algorithms is to evaluate their performance in terms of weather forecasting. It is clear from the information mentioned that the Random Forest algorithm has the best level of accuracy. Due to their categorization method, all of these algorithms fall under the category of supervised learning. They perform effectively with the training data but may not accurately predict other categories such as rainfall amount, temperature percentage, and outlook. Therefore, this approach cannot be recommended for future years for predicting weather changes in different areas. That is why there are plans to forecast different weather conditions by introducing additional attributes utilizing alternative algorithms such as SVM, ANN, and Naïve Bayes.

Debneil Saha Roy et al. [6] Consider the one-step approach, where models are trained to predict a single point in the future (one day ahead) using provided historical data. On the other hand, in the multi-step configuration, models are trained to predict values for a range of future values, which is 10 days into the future. In both cases, the models were fed weather information from the previous seven days. This data included various factors such as snowfall, precipitation, average wind speed, maximum temperature, average temperature, minimum temperature, and snow depth. By displaying the correlation matrix in Figure 3 as a heatmap, it can confirm the association between these characteristics. It is clear that the relationship between the average temperature. This correlation suggests that when building the models, these features should be taken into consideration. The models generate the average temperature for the next day and the 10 days in advance. All of the models were built with Keras, popular deep learning framework. The MLP has two thick layers with a total of 16 neurons each. One cell with 16 neurons makes up the LSTM, which is followed by a thick layer. One convolutional layer with 32 filters and a kernel size of five are included in the CNN+LSTM. This is then followed by a thick layer and an LSTM cell with 16 neurons. Rectified Linear unit (ReLU) is used as the activation function in each of these models. The Mean Absolute Error (MAE) loss function is selected, and Root Mean Squared Propagation (RMSprop) is used for optimisation with a suggested learning rate of 0.001. With a batch size of 32, all models were trained over a period of 10 epochs. Along with other meteorological variables including snowfall, precipitation, maximum temperature, minimum temperature, snow depth, average wind speed and the average air temperature was shown as a function of its historical values. The results suggest that forecasting accuracy improves as we use more complex deep neural networks.

S. Jain and D. Ramesh, et.al [7] described proposed methodology, which is consists of two phases: 1) The forecasting of seasonal weather, and 2) Selecting an appropriate crop. For seasonal weather forecasting, a Recurrent Neural Network (RNN) is utilized. Subsequently, the random forest classification technique is used to categorize appropriate crops. The

National Remote Sensing Agency (NRSA) Hyderabad station provides the data set used for predicting the weather. It includes information on meteorological characteristics throughout the duration of five years, such as humidity, temperature, wind speed, direction, and sun hours. The information was gathered at an hourly resolution between January 2014 and September 2019. Data preparation, including tasks such as filling in missing values, data normalization, and data transformation, constitutes the initial steps in training the prediction model.

The linear interpolation method is used to fill in the missing values of the data. Subsequently, the temperature data is used to calculate minimum and maximum temperatures, transforming the hourly data into daily average data. The dataset is then exposed to min-max normalisation to transform into a common scale. A Recurrent Neural Network (RNN) is used to predict the weather. This particular type of Neural Network (NN) is made with sequence dependence in consideration. RNN has the benefit of the output from the previous step to the current step over standard NN. As a result, the sequence may be remembered by the RNN's hidden state. All values in a conventional neural network are considered independent of one another.

A novel system for selecting crops based on predicted soil parameters and weather conditions to determine suitable crops for a given land. The categorization model is trained using the soil and weather data found in the Telangana state's agro-climatic zones. RNN is used in seasonal weather forecasting. The comparison of the prediction results for traditional ANN results demonstrates better prediction accuracy. An additional threshold

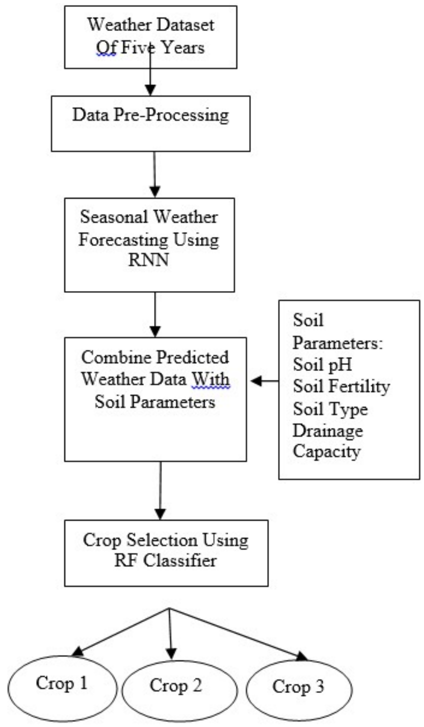

Fig.3: RNN Flowchart

parameter is incorporated into a random forest classifier to identify multiple suitable crop choices for a specific land tract. Furthermore, the proposed method additionally recommends the appropriate time for sowing each crop based on accordance with predicted weather parameters.

Yigezu Agonafir et al. [4] discussed about the application of a hybrid machine learning approach. This decision was made in light of the fact that CRISP only uses three feedback mechanisms, but the hybrid machine learning technique comprises five iterative phases. Additionally, Knowledge Discovery in Databases (KDD) and Sample, Explore, Modify, Model, and Assess(SEMMA) do not incorporate a feedback mechanism. Additionally, the processes are described in more generic and research-focused detail of hybrid machine learning technique. It replaces the 'deployment' in CRISP with the 'using discovered knowledge' step. A hybrid strategy was utilized to estimate rainfall in this study domain in order to speed up the knowledge discovery process.

Furthermore, SEMMA and Cross-Industry Process-Data Mining (CRISP-DM) are mostly company-oriented, with a specific emphasis on SEMMA, which is employed by Statistical Analysis System (SAS) Enterprise Miner and integrated into their software. On the other hand, However, CRISP-DM is more complete as compare to SEMMA. Additionally, all those process models guide will help individuals and experts in understanding how to utilize machine learning in practical situations. At the same time, CRISP-DM is expanded upon by the hybrid approach.

This paper, using the European Network on Monitoring Antisemitism (ENMA) database, aimed to explore the utilization of machine learning technology for rainfall predictive modeling. In this study, a hybrid and iterative methodology was applied, comprising six fundamental steps, includes issue domain understanding, data understanding, data preparation, machine learning, assessment, and use of the new information. In order to forecast rainfall, the efficacy as well as the effectiveness of several machine learning approaches, such as decision trees, the PART algorithm, IBK, and MLP, were studied in this work. The comparison of classification techniques using both a single algorithm and an ensemble method is the novelty of this study. In conclusion, this study found that month, year, humidity, maximum temperature, lowest temperature, sunlight, and wind speed are the key factors in forecasting rainfall. According to the results of experiments, the Ensemble technique in one-day-ahead prediction yields higher accuracy levels. On the Waikato Environment for Knowledge Analysis (Weka) experimenter, the Instance Based Learner (IBK) method generated prediction accuracies of 94.31% and 94.83%, respectively. When using all characteristics in 10 fold cross validation and scoring 75 percentage split, or 94.17% and 94.49%, respectively, on the WEKA experimenter, the improved J48 decision tree generated better results. Additionally, the J48 decision tree and the boosted PART method showed enhanced prediction accuracy with certain features, obtaining 95.46% and 95.44% prediction accuracy in the WEKA test, respectively. For a one-day advance forecast, the PART method has a prediction accuracy of 95.66% utilising a few key attributes from the WEKA explorer.

TABLE IV: EVALUATION ON SINGLE ALGORITHM & ENSEMBLE METHOD FOR ONE DAY AHEAD PREDICTION USING SELECTED ATTRIBUTES

Method	Learning Alogrithm	MAE	RMSE
Single Algorithm	PART	0.12	0.26
Boosting	PART	0.05	0.22

N. Anand, S. Srivastava, S. Sharma, L. K. Sinha, and S. Dhar et al. [8] performed a study with an emphasis on the creation of models for predicting rainfall. They used four various machine learning methods to anticipate the onset of landslides. Utilizing unbiased data, the models are calibrated and verified. To increase the models' ability to forecast outcomes accurately, normalization has been completed. Using the Gradient Descent optimization approach, the models are trained. Mean Square Error (MSE) is used as the loss function for training the linear regression, LSTM, and Back Propagation Neural Network (BPNN) models, whereas an insensitive loss function is used for training the Support Vector Regression (SVR) models.The models are assessed using the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) metrics. By computing the average distance between the actual curve and the fitted curve for each data point, RMSE reflects the difference between the expected output and the desired result. Furthermore, MAE shows how well the estimated values match the desired results. There is a high degree of confidence in the predicted values from the models when RMSE and MAE have low values. To get the best performance for each model, hyper-parameter optimisation is completed. Depending on the input set used, the models may forecast rainfall intensity for either the following month or a certain month in the following year. The created models work well at mapping low- to medium-intensity rainfall accurately, but they perform less well at forecasting low- to high-intensity rainfall accurately. By including other input variables like wind speed, temperature, humidity, etc., it is possible to improve the predicted accuracy of the models.

Q. -D. -E. -J. Ren, N. Li, and W. Zhang, et al. [2], explain the Inner Mongolian sand and dust storm statistics are collected over the last 54 years. The suggested model's validity is checked using this dataset. China's Surface Climate Data Daily Value Data Set China's Severe Dust Storm Sequence and Its Supporting Data Set are the two datasets used. The pre-processing of the data is the first stage. Both the label characteristics and the input attributes meet national requirements. The following categories are used in this article to classify the grades of sand and dust storms: super strong sand storm, strong sand storm, sand storm, floating dust, blowing sand, and no sand storm. As a result, this study uses five sand-dust storm grade labels in order of decreasing intensity: 0, 1, 2, 3, and 4. In this study, the input attribute consists of 15-day weather variables that have undergone preprocessing. This implies that the level of the sand and dust storm on the sixteenth day is predicted using weather information from the previous fifteen days. In this context, the model input sequence includes the key attributes.The beginning sample was the dataset of Inner Mongolian sand-dust storm data that has undergone preprocessing. A total of 100,000 samples were chosen based on certain time periods after down-sampling samples that did not contain sand-dust storms. A total of 110,853 sample data points were collected. Each level has a considerably different quantity of sand-dust storm samples, which might lead to the prediction model being overfit. The training data is balanced using the SMOTE (Synthetic Minority Over Sampling Technique) in order to address the incidence of these problems, while the test data is left unmodified. The training set's sand-dust storm samples were then analyzed, categorized as grades 0, 1, 2, and 3, are each increased to 100,000. The experimental results indicate that the LSTM-based sandstorm prediction model achieved an overall accuracy rate of approximately 0.8295. Additionally, the weighted average accuracy is 0.9118, the recall rate is 0.8296, and the F1-Score reached 0.8592, respectively.

TABLE V: COMPARISON OF FORECAST PERFORMANCE ON ORIGINAL TEST SET OF VARIOUS MODELS

Model	LSTM	CNN	Stacking FC	Stacking SVM
Overall accuracy	0.8296	0.833	0.8435	0.8405
Accuracy Of Sandstorm Occurrence	0.677	0.670	0.717	0.689
Accuracy Of No Sandstorm	0.85	0.85	0.86	0.86
Precision	0.91	0.91	0.93	0.93
Recall	0.83	0.83	0.84	0.84
F1-Score	0.86	0.87	0.88	0.87

Jakaria A H M, Md Mosharaf Hossain, and Mohammad Ashiqur Rahman, et.al [14] outlined methods for using machine learning to forecast the temperature in Nashville, Tennessee, at any specific hour on the next day. This forecast is based on weather information from Nashville's current day and many nearby cities. To create a single record, all of the cities under consideration's weather observations at a given timestamp are first combined. In this manner, every record in the data includes the atmospheric pressure, wind direction, humidity, condition, temperature, etc. from all the cities. The temperature at the same timestamp on the next day is chosen as the goal variable for each of these recordings. So, using the weather information from today, therefore we predict the temperature for the following day. The same methodology may also be used to forecast a wide range of other weather-related variables, such as rainfall, humidity, visibility, wind speed and direction, etc., for the following day and the following several days.A regression approach is chosen since the predicted results comprise continuous numerical variables, like the temperature in this case. Given that Random Forest Regression (RFR) uses an ensemble of many decision trees to make judgements, it gets to the conclusion that RFR is the better regressor. Additionally, the RFR method is contrasted with many other advanced ML approaches. The utilized regression techniques include SVR, Ridge Regression (Ridge), Extra-Tree Regression (ETR), and Multi-layer Perceptron (MLPR). It utilizes machine learning techniques to provide weather forecasts. Compared to traditional physical models, machine learning technology can offer intelligent models that are considerably simpler. These models are less demanding on resources and can be readily executed on nearly any computing device, such as mobile devices. The evaluation findings indicate that these machine learning models are capable of accurately forecasting weather attributes, which enables them to take on conventional models. It also uses previous information from the neighbourhood to predict weather for a certain region. It has been proven that this approach is more successful than focusing just on the region where weather forecasts are made

III. CONCLUSION

The results provided that well-selected and tuned machine learning algorithms can result in a decrease in prediction errors in various weather forecasting scenarios. The predictive data, including real-time satellite imagery or future weather forecasts, is an important component in predicting solar radiation over short- and medium-term prediction timeframes. The large-scale distribution and collection of this data are expected to improve prediction models that enhance resource allocation and operations. Simple time-of-day and day-of-year analysis uncovered highly statistically significant biases in the forecasts of eight weather variables (cloud cover, air temperature, humidity, visibility, wind direction, wind speed, air pressure, and dew point temperature) for five cities in Georgia. Hence, the review shows machine learning algorithms achieves better results.

REFERENCES

[1] X. Chu, W. Bai, Y. Sun, W. Li, C. Liu and H. Song, "A Machine Learning-Based Method for Wind Fields Forecasting Utilizing GNSS Radio Occultation Data," in *IEEE Access*, vol. 10, pp. 30258-30273, 2022, doi: 10.1109/ACCESS.2022.3159231.

[2] Q. -D. -E. -J. Ren, N. Li and W. Zhang, "Research on Sand-Dust Storm Forecasting Based on Deep Neural Network With Stacking Ensemble Learning," in *IEEE Access*, vol. 10, pp. 111855-111863, 2022, doi: 10.1109/ACCESS.2022.3216309.

[3] Tulsi Pawan Fowdur, Rosun Mohammad Nassir-Ud-Diin Ibn Nazir, A real-time collaborative machine learning based weather forecasting system with multiple predictor locations, Array, Volume 14, 2022, 100153, ISSN 2590-0056, https://doi.org/10.1016/j.array.2022.100153.

[4] Yigezu Agonafir, "Machine Learning Approach for Weather Forecasting", IEEE-SEM, Volume 9, Issue 9, September-2021 ISSN 2320-9151

[5] B. L. Pavuluri, R. S. Vejendla, P. Jithendra, T. Deepika and S. Bano, "Forecasting Meteorological Analysis using Machine Learning Algorithms," *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, 2020, pp. 456-461, doi: 10.1109/ICOSEC49089.2020.9215440.

[6] Debneil Saha Roy, Forecasting The Air Temperature at a Weather Station Using Deep Neural Networks, Procedia Computer Science, Volume 178, 2020, Pages 38-46, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.11.005.

[7] S. Jain and D. Ramesh, "Machine Learning convergence for weather based crop selection," *2020 IEEE International Students' Conference on Electrical,Electronics and Computer Science (SCEECS)*, Bhopal, India, 2020, pp. 1-6, doi: 10.1109/SCEECS48394.2020.75.

[8] S. Srivastava, N. Anand, S. Sharma, S. Dhar and L. K. Sinha, "Monthly Rainfall Prediction Using Various Machine Learning Algorithms for Early Warning of Landslide Occurrence," *2020 International Conference for Emerging Technology (INCET)*, Belgaum, India, 2020, pp. 1-7, doi: 10.1109/INCET49848.2020.9154184.

[9] Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., 2019. Deep learning and process understanding for data-driven earth system science. Nature 566, 195–204, DOI:10.1038/s41586-019-0912-1

[10] N. L. and M. H.S., "Atmospheric Weather Prediction Using various machine learning Techniques: A Survey," *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2019, pp. 422-428, doi: 10.1109/ICCMC.2019.8819643.
[11] S. Madan, P. Kumar, S. Rawat and T. Choudhury, "Analysis of Weather Prediction using Machine Learning & Big Data," *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, Paris, France, 2018, pp. 259-264, doi: 10.1109/ICACCE.2018.8441679..
[12] Biswas, Munmun & Dhoom, Tanni & Barua, Sayantanu. (2018). Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data. International Journal of Computer Applications. 182. 20-24. 10.5120/ijca2018918265.
[13] Bin Wang, Jie Lu, Zheng Yan, Huaishao Luo, Tianrui Li, Yu Zheng, Guangquan Zhang, “Deep Uncertainty Quantification: A Machine Learning Approach for Weather Forecasting”, arXiv:1812.09467, 2018, https://doi.org/10.48550/arXiv.1812.09467
[14] A H M Jakaria, Md Mosharaf Hossain, and Mohammad Ashiqur Rahman.2018. Smart Weather Forecasting Using Machine Learning: A Case Study in Tennessee. In Proceedings of ACMMid-Southeast conference (Mid-Southeast’18). ACM, NewYork,NY, USA, 4 pages, https://doi.org/10.48550/arXiv.2008.10789
[15] Mark Holmstrom, Dylan Liu, Christopher Vo, “Machine Learning Applied to Weather Forecasting” Stanford University (Dated: December 15, 2016)
[16] Hernandez, Emilcy & Sanchez-Anguix, Víctor & Julián, Vicente & Palanca, J. & Duque, Néstor. (2016). Rainfall Prediction: A Deep Learning Approach. 151-162. 10.1007/978-3-319-32034-2_13.
[17] Aditya Grover, Ashish Kapoor, and Eric Horvitz. 2015. A deep hybrid model for weather forecasting. In International Conference on Knowledge Discovery and Data Mining. ACM, 379–386, https://doi.org/10.1145/2783258.2783275
[18] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Neural Information Processing Systems. 802–810.
[19] O.F. Oyediran and A.B Adeyemo , “Performance Evaluation of Neural Network MLP and ANFIS models for Weather Forecasting Studies”, vol. 6, pp.147-164, March 2013, https://doi.org/10.1016/j.compeleceng.2019.07.024.