Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 5s (2025)

Advanced Weather Forecasting with Machine Learning: Leveraging Meteorological Data for Improved Predictions

Sunil Khatri[1,2]*, Rajani P.K.[3]

[1]Research Scholar, Dept. of EXTC, Pimpri Chinchwad College of Engineering, Maharashtra, India

[2]Assistant Professor, Dept. of IoT, Thakur College of Engineering and Technology, Maharashtra, India

[3]Associate Professor, Dept. of EXTC, Pimpri Chinchwad College of Engineering, Maharashtra, India

*Corresponding email: skhatri1909@gmail.com

Article History:

Received: *01-10-2024*

Revised: *22-11-2024*

Accepted: *02-12-2024*

Abstract:

Objectives: This study proposes a machine learning approach for improving weather forecasting accuracy with reduced resource requirements, focusing on rainfall and flooding predictions in urban regions.

Methods: The study used historical meteorological data (2009–2023) from Indian regions, applying supervised learning models, including regression and ensemble methods. Data preprocessing steps included handling missing values, outlier detection, and normalization. The models were evaluated using Accuracy, MAE and MSE value to determine prediction accuracy and reliability.

Findings: The results indicate that AI models, especially LSTM, provide significant accuracy improvements with 80.11% accuracy. These models performed well in predicting short-term weather phenomena such as rainfall in urban flood-prone areas like Mumbai. The method demonstrated the potential to produce reliable forecasts with limited computational resources. The findings complement existing research, adding value by showcasing the adaptability and scalability of resource-efficient ML models for local meteorological applications. This work highlights the practical implications for urban planning and flood preparedness.

Novelty: A cost-effective machine learning framework for accurate local weather forecasting, addressing scalability and computational constraints.

Keywords: Weather Forecasting, Machine Learning, Meteorological Data, Predictive Modelling, Flood Prediction

1. Introduction

Accurate weather forecasting plays a critical role in various sectors such as agriculture, transportation, urban planning, and disaster management. As climate variability and extreme weather events continue to rise, the need for precise and reliable forecasting systems has become more important than ever. These systems enable individuals and organizations to make informed decisions, mitigate risks, and plan effectively for future events.

Traditional weather prediction models rely on complex physical simulations that require significant computational resources and extensive data inputs. Despite their sophistication, these models often struggle with declining accuracy due to the increasing complexity of atmospheric conditions and inherent limitations in their algorithms. Additionally, the high costs associated with running these

systems on advanced infrastructure can make them inaccessible, particularly in resource-constrained settings.

In recent years, machine learning has emerged as a promising alternative for weather prediction. By leveraging historical meteorological data, machine learning models can identify patterns and trends that traditional methods may overlook. These models not only provide faster predictions but also reduce the computational burden, making them suitable for deployment on low-cost systems. Such advancements are particularly relevant for regions like India, where diverse climatic conditions and frequent extreme weather events necessitate localized and efficient forecasting solutions.

This study explores the application of machine learning techniques to predict weather conditions, including rainfall and temperature, with a focus on urban regions prone to flooding. By utilizing historical data from multiple Indian locations, the proposed system demonstrates its potential to provide accurate, resource-efficient forecasts. The findings of this research highlight the transformative potential of machine learning in modern weather forecasting, offering a practical and scalable solution for addressing challenges posed by traditional methods.

The key contributions of this work include:

• Implementation of machine-learning algorithms for efficient and accurate weather prediction.

• Development of a resource-optimized forecasting framework that can operate on low-cost computational systems.

• Comparative evaluation of different machine-learning models to determine their suitability for predicting specific meteorological conditions.

1.1 Machine Learning Architecture

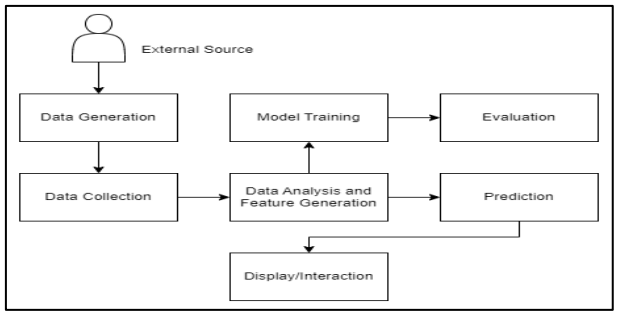


Fig 1. Architecture of Machine Learning

The architecture of the machine learning model is illustrated in Fig. 1. The process consists of several key stages to ensure accurate and efficient weather predictions:

1. *Data Collection*: Relevant meteorological data is gathered from external sources, focusing on historical weather records. This stage involves accumulating large datasets that will be used for subsequent analysis.

2. *Data Preprocessing*: In this stage, the collected data undergoes thorough cleaning and conditioning. Measures are taken to eliminate issues such as redundancy, missing values, and inconsistencies to ensure the quality and reliability of the data.

3. *Data Analysis*: Once the data is prepared, an analysis is conducted to identify the key features or independent variables that contribute to weather predictions. These include factors such as cloud cover, wind chill, atmospheric pressure, and sunlight hours, all of which are used to predict the target variable, such as temperature.

4. *Model Training*: During the model training stage, machine learning algorithms are employed to fit both independent and dependent features to the model. The collected data is used to teach the model the underlying patterns, enabling it to make accurate predictions.

5. *Model Evaluation*: This stage involves assessing the performance of the trained model using various evaluation metrics. By comparing different model configurations, the most effective approach is selected based on accuracy and performance.

6. *Prediction*: In the prediction stage, the model uses current meteorological data as input to generate forecasts. The predicted weather conditions, such as temperature and rainfall, are derived based on the accuracy achieved during the training phase.

7. *Display and Output*: Finally, the forecasted values are presented to end-users in an accessible format. The predicted weather conditions are displayed, allowing users to make informed decisions based on the forecasts.

1.2 Machine Learning Algorithms:

Machine learning algorithms can be grouped into three main types: supervised learning, unsupervised learning, and reinforcement learning. Each type serves specific purposes depending on the data characteristics and the application. This work utilizes supervised learning to construct models for regression and classification tasks.

1.2.1 Supervised Learning

Table 1: Examples of Algorithms in Supervised Learning: Random Forest, Decision Trees, Linear Regression, Logistic Regression, Boosting Algorithms

Type	Practical Example
Neural Network	Financial outcome prediction, fraud detection
Classification & Regression	Spam filtering, fraud detection
Decision Tree	Risk management, threat detection

Supervised learning relies on labelled datasets, where both input features and corresponding outputs are available. Models are trained to map inputs to outputs by minimizing the error between predicted and actual values. Tasks such as classification and regression are commonly addressed using this method.

1.2.2 Unsupervised learning

Table 2: Examples of Algorithms in Unsupervised Learning: K-Means Clustering, Apriori Algorithm, Adaptive Resonance Theory, Self-Organizing Maps (SOM).

Type	Practical Example
Cluster Analysis	Detecting fraudulent transactions, spam filtering
Pattern Recognition	Object detection, people detection
Association rule learning	Bioinformatics, manufacturing, and assembly

Unsupervised learning involves analysing unlabelled data to identify hidden patterns or structures. This type of learning is effective for tasks such as clustering, dimensionality reduction, and anomaly detection.

1.2.3 Reinforcement Learning

Reinforcement learning focuses on training models through interaction with an environment. Algorithms learn by receiving rewards for favourable actions and penalties for unfavourable ones, gradually improving decision-making strategies.

Example: Markov Decision Process (MDP): Utilized for modelling tasks where outcomes are influenced by both random factors and decision-making processes.

Reinforcement learning is particularly suitable for dynamic and uncertain environments, requiring iterative updates to improve strategies.

1.2.4 Challenges and Considerations

Applying machine learning methods effectively requires addressing several key challenges:

• Understanding the problem domain, including constraints and objectives.

• Ensuring data quality by addressing issues such as missing values, redundancy, and inconsistent variables.

• Selecting appropriate algorithms and tuning their parameters to optimize performance for the specific task.

2. Literature Review

Recent advancements in machine learning have significantly enhanced the accuracy and efficiency of weather prediction systems. Several studies have demonstrated the potential of these techniques to address limitations in traditional forecasting methods.

Research comparing deep learning and numerical weather prediction models underscored the strengths and weaknesses of both approaches. Deep learning demonstrated superior adaptability to complex, non-linear patterns in atmospheric data, whereas traditional numerical models provided greater reliability in structured forecasting scenarios [1].

Another investigation evaluated machine learning models for predicting meteorological variables such as temperature, humidity, and precipitation in Indian regions. This study emphasized the importance

of feature selection and data preprocessing in enhancing model performance, offering valuable insights for localized forecasting [2].

A comparative study on decision trees, random forests, and artificial neural networks demonstrated their effectiveness in weather prediction. Random forests and neural networks outperformed decision trees in terms of accuracy and computational efficiency, making them suitable for high-dimensional meteorological datasets [3].

Time-series analysis techniques, including ARIMA and exponential smoothing, were compared to modern deep learning methods. The findings highlighted the superiority of deep learning in handling long-term dependencies and irregularities in weather patterns, particularly when applied to large datasets [4].

The use of clustering algorithms combined with regression models was investigated for weather prediction. This hybrid approach improved forecasting accuracy by identifying distinct weather patterns within datasets before applying regression analysis [5].

Research on ensemble methods, such as integrating support vector regression and linear regression, revealed significant improvements in prediction accuracy. Ensemble models effectively reduced errors by leveraging the strengths of individual algorithms, demonstrating their potential for meteorological applications [6].

The implementation of back-propagation algorithms was explored for weather classification and forecasting. This method achieved moderate accuracy levels but highlighted the need for further refinement to handle complex meteorological data [7].

A comparison between artificial neural networks and linear regression models revealed that neural networks excel in modelling non-linear weather patterns, whereas regression models remained more interpretable for simpler scenarios [8].

An ensemble approach combining predictions from multiple regression models, including support vector regression, was shown to enhance forecasting precision. This methodology effectively minimized individual model errors, making it a robust choice for weather prediction [9].

Research on support vector regression (SVR) models demonstrated their ability to handle high-dimensional meteorological datasets. The study highlighted the advantages of SVR in achieving higher accuracy compared to traditional regression models [10].

The integration of clustering techniques with regression models was proposed to improve weather prediction accuracy. This approach identified distinct weather patterns within datasets and applied targeted regression techniques, yielding promising results [11].

A study on flood prediction in Mumbai utilized various machine learning techniques, including Logistic Regression, K-Nearest Neighbours, Random Forest, and Gradient Boosting. The findings revealed that Random Forest and Gradient Boosting outperformed other methods, accurately identifying flood-prone areas and contributing to disaster preparedness efforts [12].

Multiple ML algorithms were examined (Elkhrachy et al., 2022) for their potential to improve the accuracy of water depth estimates during a flash flood incident in New Cairo City, Egypt. Features

were extracted from the SAR data's backscattering spectral band and fed into ML algorithms separately. Integrating several training datasets and machine learning algorithms was suggested as a means of improving the accuracy of water depth forecasts using satellite data in order to construct an emergency plan in the event of floods [13].

As reported by Abdullah et al. (2021), the author discovered a rising pattern of use for Multiple-criteria decision analysis (MCDA) methods in the management of flood and drought events in the twenty-first century. This work surveyed the literature on MCDA methods for flood and drought management from 2000 to 2020, drawing from 149 articles in journals, conference proceedings, and other scholarly publications. Decision-makers in the Data management platform (DMP) have shifted their attention to flood occurrences because of their significant consequences, which has prompted a greater number of research projects on the topic [14].

The influences of both the natural and social settings on flood risk have been researched (Chen et al., 2021) and found to make the problem of flood risk complicated and difficult to conceptualize. Targeted flood control measures are required due to the unique features of each risk category. Future efforts to stop flooding should put more attention on the Digital Elevation Model, M1DP, and RD, which are the three most important driving factors [15].

Kalantar et al. (2021) employed ANNs, DLNNs, and DLNNs that had been improved with particle swarm optimization (PSO) to predict and estimate flood-prone areas. Methods such as sensitivity, specificity, area under the curve (AUC), and the true skill statistic (TSS) were employed to evaluate the performance of the models. In order to study and analyse complex occurrences like floods, large datasets from disciplines like remote sensing and earth observation provide an invaluable starting point. It appears that optimization and ML algorithms can be used in contexts such as crisis management and urban planning, which call for the rapid analysis, visualization, and information extraction from enormous data sets [16].

Mane et al., (2020) tried out a variety of ML algorithms on the available rainfall data, including SVM, KNN, Logistic Regression, Naive Bayes, and others. The author used these ML models to create a comprehensive flood prediction and warning system, including a website and an Android application, to notify worried citizens and government officials. Moreover, the system encourages real-time monitoring via the built-in website as a simple medium for distributing data, especially in outlying regions [17].

The significance of training data was recognised in a research study (Katiyar et al., 2020). Due to the limitations of SAR photos, such as distorted geometry, shadow regions, and uncertainty in the representation of urban flood areas, the author suggests including additional data in addition to SAR images in future research. By combining DEM data, geometry and shadow areas can be taken care of. Separating the flooded paddy fields from other flooded regions may also benefit from the use of multi-temporal SAR images and various polarizations. By comparing the image taken before the flood with the one taken during the flood, a change-detecting mechanism can be used, greatly improving the accuracy of the procedure [18].

The studies reviewed provide a comprehensive foundation for applying machine learning to weather prediction. However, challenges such as data quality, feature selection, and high accuracy remain areas

for ongoing research. The integration of robust algorithms with scalable systems offers promising directions for future advancements in the field.

3. Methodology

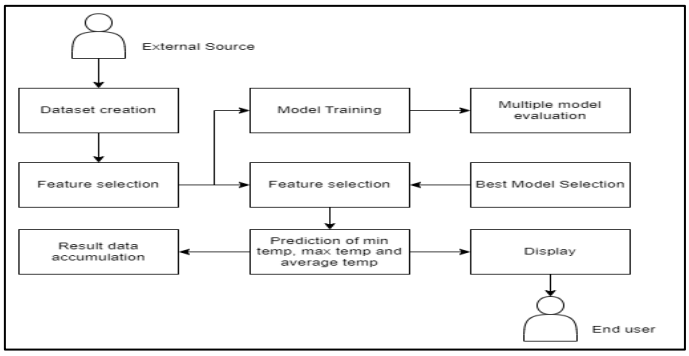


Fig 2: Stages for weather prediction.

The system operates through a comprehensive process involving data collection, analysis, and evaluation to identify the most precise prediction model as denoted in Fig 2. The initial phase focuses on gathering historical weather data spanning from 2009 to 2023. During this phase, issues such as redundancies and missing values in the collected data are systematically addressed to ensure data integrity and reliability.

Subsequently, the data is analysed to identify patterns and trends that inform the prediction model's characteristics. Key features in the dataset include variables such as sunlight hours, date-time, atmospheric pressure, precipitation levels, and other meteorological parameters. These features serve as the foundation for building and refining the model.

The training phase utilizes this curated dataset to develop machine learning models capable of making accurate predictions. These models are assessed through rigorous evaluation processes to determine the most optimal configuration. The evaluation involves comparing various models based on performance metrics to ensure the highest level of prediction accuracy.

Once finalized, the model predicts weather parameters such as average, minimum, and maximum temperatures using current independent input data. These predictions are then transmitted to the end-user via a user-friendly interface, ensuring practical applicability for real-world use cases.

3.1 Flowchart

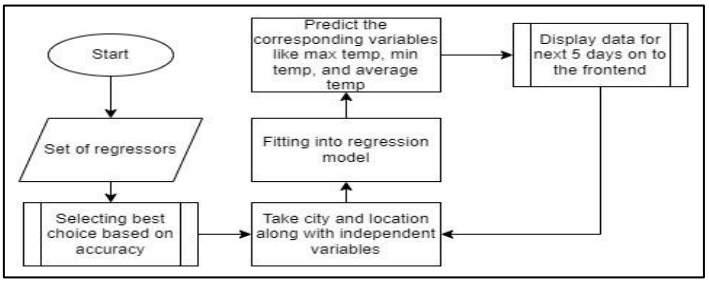


Fig 3: Flowchart for weather prediction.

The flow of the system, as illustrated in Fig. 3, begins with the selection of regressors and identifying the most suitable model from the available set. The selected regressors are applied to the collected data for model training and accuracy testing. Model accuracy serves as a critical parameter for evaluating performance, guiding the selection of the optimal regressor that offers the best combination of accuracy and computational efficiency.

Once the most suitable model is identified, it is utilized to predict the average, minimum, and maximum temperatures for the next five days based on the provided independent input values. The process incorporates location-specific data and relevant meteorological variables, which are fitted into the model to generate accurate predictions.

The predicted values are subsequently displayed through a user-friendly interface, enabling end-users to access the forecasted weather conditions with ease. This streamlined approach ensures that predictions are both accurate and accessible, enhancing practical applicability.

3.2 Data Analysis for Weather Prediction

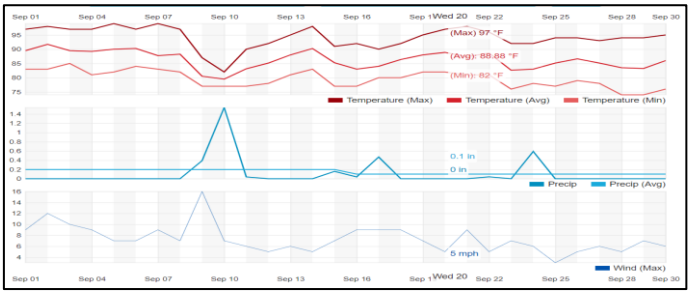


Fig 4: Data for the month of September 2023 for Temperature, Precipitation, Wind for Mumbai

Throughout September, temperature variations were observed, with significant highs and lows. Maximum temperatures ranged from 82°F to 99°F, with the peak of 99°F recorded on the 5th and 7th. The average temperature for the month was 88.6°F, reflecting relatively warm conditions. Minimum temperatures fluctuated between 74°F and 85°F, with the lowest temperature of 74°F recorded on multiple occasions.

The month commenced with temperatures ranging from 83°F to 89°F, gradually rising to a mid-month peak of 97°F to 99°F over several days. As September progressed, temperatures began to decline, with the lowest values of 74°F to 78°F occurring towards the month's end. These trends reflect the transition from late summer to early autumn, characterized by residual summer heat giving way to cooler fall conditions.

Precipitation data for September indicates varied rainfall levels. The initial eight days were marked by dry conditions, with no measurable rainfall. On the 9th, precipitation increased to 0.39 inches, followed by a significant spike of 1.54 inches on the 10th. The remainder of the month featured minor rainfall events, including 0.04 inches on the 11th, 0.16 inches on the 15th, and another 0.04 inches on the 16th.

Later in the month, precipitation increased notably to 0.59 inches on the 23rd. However, the majority of days after this remained dry, resulting in a total monthly precipitation of 3.27 inches. The mid-month rainfall significantly influenced the overall weather and water availability during September.

Wind speed data for the month highlights fluctuations in intensity. Maximum wind speeds ranged from 3 mph to 16 mph, with the highest speed of 16 mph recorded on the 8th. The average wind speed for September was 4.1 mph, indicating generally moderate conditions. Some days recorded minimal wind activity, including calm conditions (0 mph) at the start and end of the month.

Wind intensity increased gradually during the month, with sporadic peaks, such as the 16 mph recorded on the 8th. Towards the latter part of September, wind speeds diminished, with several days experiencing little to no wind, reflecting calmer conditions.

These wind speed variations indicate shifts in atmospheric dynamics typical of the late summer to early autumn transition, driven by changes in air pressure and prevailing weather systems.

3.3 Working Methodology

The system can be divided into three parts as given below:

3.3.1 The Prediction Model

The prediction model forecasts the temperature for the next five days. Similar to other machine learning models, the process begins with dataset selection. The datasets, sourced from Kaggle, cover various locations such as Jaipur, Mumbai, Pune, Bengaluru, Hyderabad, and others. Python libraries such as Matplotlib, Pandas, and NumPy are utilized for data manipulation and visualization.

A pre-processed dataset is employed in conjunction with the scikit-learn library to develop the prediction model. The model is designed to handle independent values and predict the target variable, temperature. Linear regression is used to fit the independent variables, which include factors such as date/time, maximum temperature, minimum temperature, solar hours, UV index, sunrise and sunset times, pressure, humidity, heat index, and wind-chill.

The dataset typically consists of 25 columns and approximately 96,000 entries, spanning hourly temperature records from 2009 to 2023. A data visualization tool is used to show that measurements are taken across a range of latitude and longitude values for each date.

The model's performance is evaluated using metrics such as mean absolute error (MAE) and mean squared error (MSE) from the scikit-learn library.

The following methodology outlines the steps for weather prediction using machine learning models:

1. *Data Collection*: The required dataset is gathered, containing parameters such as minimum and maximum temperature, sun hours, actual temperature, cloud cover, wind speed, heat index, humidity, precipitation, wind direction, sunrise, sunset, and UV index. Data is collected from January 2009 to December 2023 on an hourly basis.

2. *Data Preprocessing*: The dataset undergoes preprocessing to ensure it is clean and ready for analysis. This step involves checking for missing values, removing outliers, and normalizing the data as needed.

3. *Feature Selection*: Relevant features are selected from the pre-processed dataset. Feature selection is performed using techniques such as correlation analysis, mutual information, and recursive feature elimination to improve model accuracy.

4. *Model Training*: Various machine learning models, including decision trees, random forests, linear regression, and artificial neural networks, are trained on the pre-processed dataset. Hyperparameter tuning is performed to identify the optimal model for weather forecasting.

5. *Model Evaluation*: The trained models are evaluated using performance metrics such as MAE, MSE, and the coefficient of determination (R^2). The models are compared to determine the best model for accurate weather prediction.

6. *Model Deployment*: Once the best model is identified, it is deployed for making predictions on future weather conditions. The model is integrated with a user-friendly interface for easy access to predictions.

7. *Results Analysis*: The deployed model's performance is analysed and compared with other existing models. The evaluation is based on accuracy, reliability, and usefulness in predicting weather conditions.

This structured methodology ensures the development of a robust and accurate machine learning model for weather prediction.

3.3.2 The Backend

The backend of the system is implemented using the Python Flask module to create a web framework for weather prediction. A local host application is developed to take user input, including the city for which weather predictions are required. The corresponding latitude and longitude values for the selected city are retrieved.

Once the form is submitted, the model receives the city-specific data and predicts the temperature for the next five days. The predicted maximum temperature, minimum temperature, and rainfall quantity are then passed to the frontend for display. The data is rendered on the webpage using Jinja templating.

3.3.3 The Frontend

The frontend is designed using HTML, CSS, and Jinja templating. The interface consists of a form with a dropdown menu for selecting cities. After the user selects a city, the ‘predict’ button triggers a POST request to activate the backend, which performs the necessary functions to predict the temperature and rainfall for the next five days.

The predictions, including the maximum and minimum temperatures, are passed from the backend to the frontend and displayed using HTML cards. These cards show the weather forecast for each day. The average temperature is calculated from the maximum and minimum temperatures, while the rainfall prediction determines whether rain is expected. If rain is forecasted, the model further classifies it as a drizzle or a thunderstorm.

Icons representing four weather possibilities—sunny, cloudy, drizzle, and thunderstorm—are displayed for each prediction, providing a visual representation of the forecast. Users can then select any other city to generate further predictions.

3.4 Mathematical Modelling

A dataset containing historical weather data, including parameters such as temperature, precipitation, wind speed, and other relevant meteorological variables, is used for model development. The model equation is expressed as follows:

$$\text{Temperature}=\beta 0+\beta 1\times\text{Humidity}+\beta 2\times\text{Wind Speed}+\beta 3\times\text{Cloud Cover}+\varepsilon \quad (1)$$

Where:

Temperature is the predicted temperature.

β0 is the intercept of the regression model.

β1, β2, and β3 are coefficients for the predictor variables (Humidity, Wind Speed, Cloud Cover).

Humidity is the humidity level.

Wind Speed is the wind speed.

Cloud Cover is the extent of cloud cover.

ε is the error term, accounting for unexplained variability.

To develop this model, a dataset containing historical values of temperature, humidity, wind speed, and cloud cover is utilized. Multiple linear regression analysis is performed using statistical software or programming libraries to estimate the coefficients (β0, β1, β2, and β3) that best fit the data.

Let y represent the dependent variable (the target to be predicted), and x represent the independent variable (the predictor). The model assumes a linear relationship between y and x, which is represented as:

$$y=\beta 0+\beta 1x+\varepsilon \quad (2)$$

Where:

y is the dependent variable (the value you want to predict).

x is the independent variable (the input used for prediction).

β0 is the intercept of the regression line, representing the value of y when x is 0.

β1 is the slope of the regression line, representing the change in y for a unit change in x.

ε is the error term, accounting for the variability that is not explained by the linear relationship.

In the case of multiple linear regression, where there are multiple independent variables (x1, x2, etc.), the model equation is extended as follows:

$$y=\beta 0+\beta 1x1+\beta 2x2+\ldots+\beta pxp+\varepsilon \quad (3)$$

where,

p is the number of independent variables.

Root Mean Squared Error (RMSE) is used as the performance metric to evaluate the machine learning model. The formula for calculating RMSE is as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(\tilde{y}_t - y_t)^2}{n}} \quad (4)$$

Where n is the test $\tilde{y}_t$ is the predicted parameter & y_t is the actual parameter, respectively.

It is important to recognize that while this model captures the relationships between temperature and the three predictor variables, real-world weather prediction models are considerably more complex. These models typically incorporate a larger set of predictor variables, account for intricate interdependencies and interactions among variables, and employ advanced techniques such as feature engineering and regularization to mitigate overfitting.

Additionally, weather prediction models are often updated and refined based on new data and improved understanding of atmospheric phenomena. Operational weather prediction requires the use of sophisticated mathematical models that include advanced numerical simulations and data assimilation methods, extending beyond the capabilities of basic regression models.

4. Results & Discussion

The objective of developing a prediction model and integrating it into a website with an intuitive user interface was successfully achieved. The model forecasts the maximum and minimum temperatures, as well as the probability of precipitation, for a specified number of days based on current data. The model's effectiveness was evaluated over a five-day period, with a deviation of approximately 0.5, demonstrating a high level of precision.

To achieve optimal accuracy, linear regression was chosen, utilizing conventional coding practices and a limited dataset. While other models, such as random forest and decision tree regression, are capable of making predictions, a thorough evaluation led to the selection of linear regression.

The user-friendly interface was developed using Python, with HTML, CSS, and Jinja2 employed to create an aesthetically pleasing design as denoted in Fig 5, Fig 6 and Fig 7.

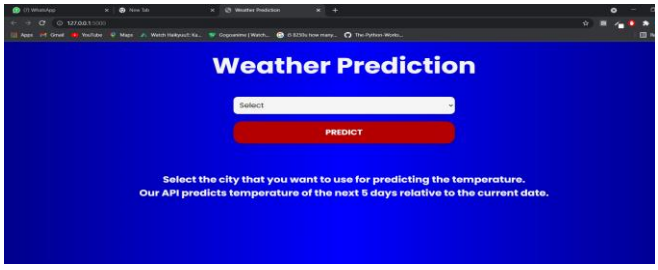


Fig 5: Home Page

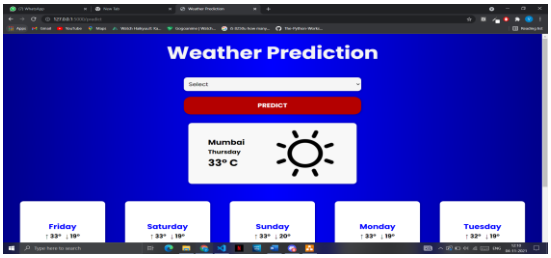


Fig 6: Prediction page 1

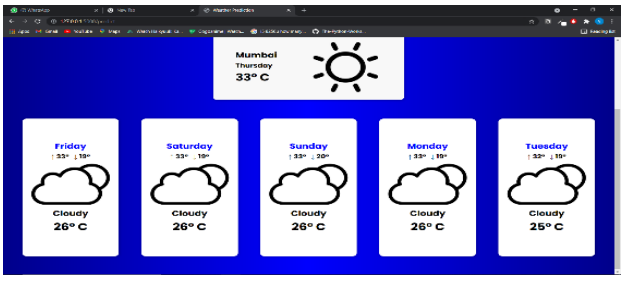


Fig 7: Prediction page 2

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 5s (2025)

4.1 PERFORMANCE COMPARISON OF MODELS

The performance of the models was compared by observing the trend in the Root Mean Squared Error (RMSE) on the test data. Initially, the RMSE was high, when only data from Northern India was used. However, accuracy improved as data from multiple years were incorporated. The RMSE continued to decrease as additional latitudes were included. The lowest RMSE, when only Northern India data was used, was achieved when full geographic coverage was attained.

The RMSE was calculated on test data with progressively increasing amounts of training data from across the country. When only one year of data was used, the RMSE was high, but it decreased as more years were added with eight years of data. However, the RMSE increased again when more years were added, due to abrupt weather changes in certain years that impacted the model's training. A comparison of different models was conducted.

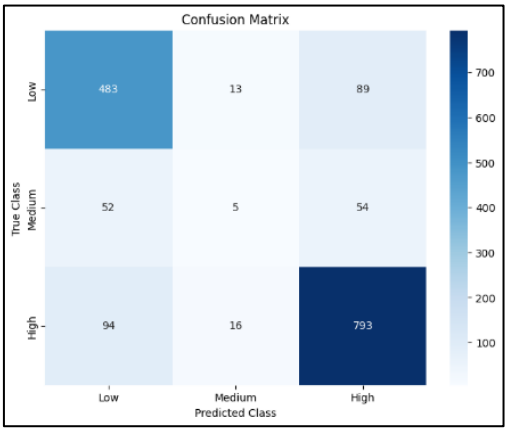


Fig 8: Confusion Matrix using LSTM

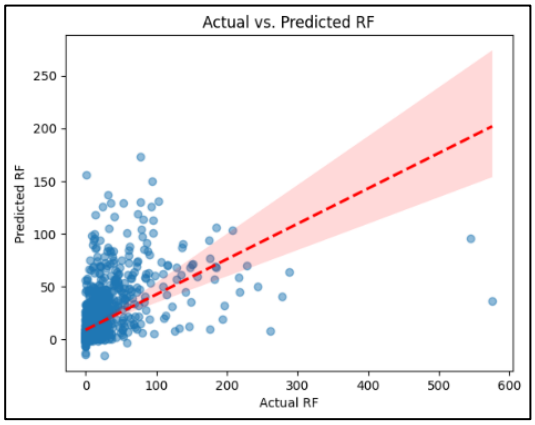


Fig 9: Prediction Plot Image using XGboost

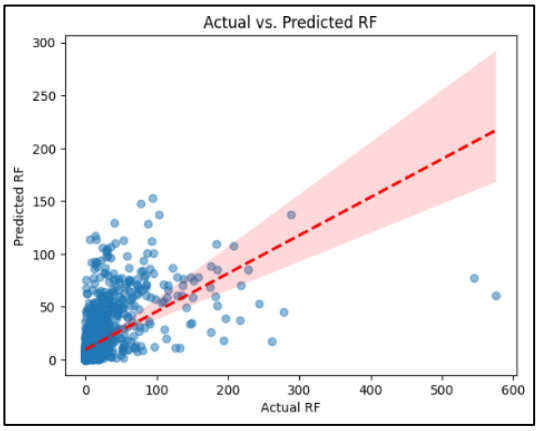


Fig 10: Prediction Plot Image using Random Forest

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 5s (2025)

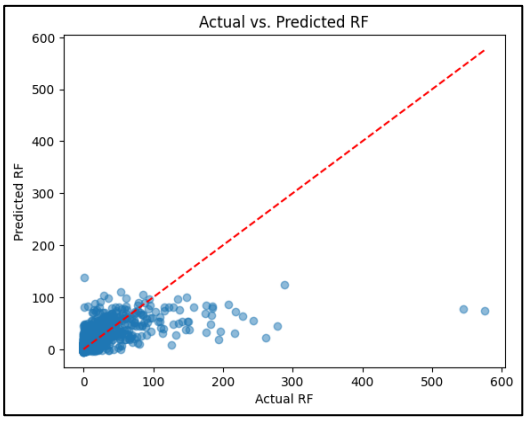


Fig 11: Prediction Plot Image using Polynomial Regression

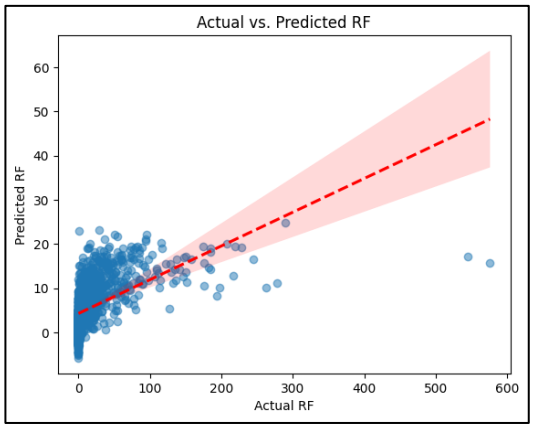


Fig 12: Prediction Plot Image using SVM

Table 3: Evaluation Rubrics of various ML algorithms for rainfall prediction

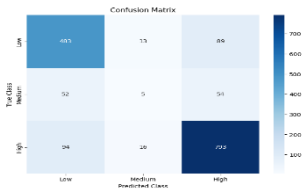
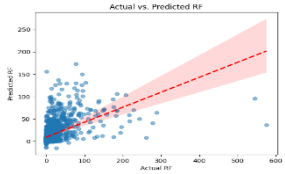
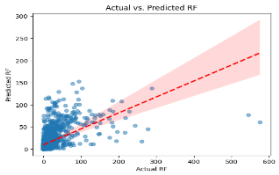
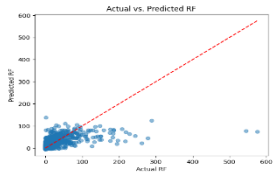
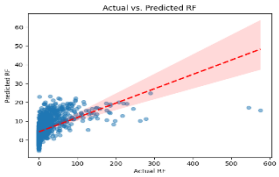
Algorithm	**Prediction Plot Image**	**Accuracy value**
LSTM		**80.11 %**
XGboost		65.65 %
Random forest		68.14 %
Polynomial regression		69.89 %
SVM		70.96 %

Fig. 8 to Fig. 12 illustrate evaluation plots of several models, and the findings for each model are observed accordingly as mentioned in Table 3.

The best results are achieved with LSTM offering the most precise and reliable weather forecasts, balancing model flexibility with resilience.

5. Conclusion

In conclusion, the use of machine learning algorithms for weather prediction represents an emerging field with the potential to significantly transform how weather patterns and conditions are forecasted. Various prediction Models show promise in this complex domain. Nonlinear ensemble models demonstrate exceptional ability to capture the complexities of weather patterns from historical data.

However, challenges remain, particularly in acquiring sufficient high-quality training data and adapting to abrupt changes in weather systems. While linear regression models lack the flexibility and accuracy needed for complex predictions, advanced machine learning algorithms have the potential to revolutionize weather forecasting. This is possible through meticulous data preprocessing and regularization, which help mitigate the risk of overfitting. These algorithms can learn from vast amounts of past data, enabling them to identify patterns and generate precise forecasts.

These capabilities offer reliable predictions that can guide decision-making for individuals and businesses across various sectors. The adoption of machine learning in weather forecasting could significantly impact industries such as transportation, energy production, agriculture, and disaster management. Furthermore, these algorithms can be continuously updated with new meteorological data in real-time, enhancing their usefulness in providing up-to-date forecasts.

Despite the promising benefits, several obstacles need to be addressed. The volume of data required for algorithm training remains a significant challenge. Ensuring that the algorithms are trained on accurate, reliable data requires extensive preprocessing due to the fragmented and inconsistent nature of weather data. Additionally, the inherent complexity of weather systems, which can change unexpectedly, poses a challenge to the model's adaptability.

Nonetheless, the application of machine learning in meteorological forecasting holds the potential to substantially increase prediction accuracy. With continued innovation and investment, machine learning algorithms could fundamentally change how weather is predicted, providing valuable insights for businesses, governments, and individuals. Addressing the challenges of this field will require a multidisciplinary approach that integrates expertise in machine learning, meteorology, and data science.

References

[1] Schultz M. G., Betancourt C., Gong B., Kleinert F., Langguth M., Leufen L. H., Mozaffari A. and Stadtler S. 2021, "Can deep learning beat numerical weather prediction?", PHILOSOPHICAL TRANSACTIONS A, ROYAL SOCIETY PUBLISHING, A.37920200097, http://doi.org/10.1098/rsta.2020.0097

[2] Jitcha Shivang, S Sridhar, "Weather Prediction for Indian Location using Machine Learning", International Journal of Pure and Applied Mathematics, Volume 118 No. 22 2018, 1945-1949.

[3] L. Zhang and J. Xia, "Flood detection using multiple Chinese satellite datasets during 2020 China summer floods," Remote Sensing, vol. 14, no. 1, Jan. 2022, doi: 10.3390/rs14010051.

[4] S. Pathak, M. Liu, D. Jato-Espino, and C. Zevenbergen, “Social, economic and environmental assessment of urban sub-catchment flood risks using a multi-criteria approach: A case study in Mumbai City, India,” Journal of Hydrology, vol. 591, Dec. 2020, doi: 10.1016/j.jhydrol.2020.125216.

[5] D. L. Chang, S. H. Yang, S. L. Hsieh, H. J. Wang, and K. C. Yeh, “Artificial intelligence methodologies applied to prompt pluvial flood estimation and prediction,” Water (Switzerland), vol. 12, no. 12, Dec. 2020, doi: 10.3390/w12123552.

[6] R. Bentivoglio, E. Isufi, S. Nicolaas Jonkman, and R. Taormina, “Deep Learning Methods for Flood Mapping: A Review of Existing Applications and Future Research Directions,” 2021, doi: 10.5194/hess-2021-614.

[7] D Sanjay, Sawaitful, KP Prof. Wagh, Dr. Chatur, “Classification and Prediction of Future Weather by using Back-Propagation Algorithm – An Approach”.

[8] Soheila Dehghani, Ali Jafari, and Hamidreza Zareipour, “A Comparison of Linear Regression and Artificial Neural Network Models for Weather Prediction”

[9] Wei Ma, Bingliang Liu, and Yiran Zhao Wei Ma, Bingliang Liu, and Yiran Zhao, "An Ensemble Method for Accurate Weather Prediction using Regression Models."

[10] Hongmei Zhang, Rongli Wang, and Tao Liu, "Weather Forecasting using Support Vector Regression Models".

[11] Shanshan Liu, Hao Liu, and Zongmin Li, “Integrating Regression Models with Clustering Techniques for Improved Weather Prediction”.

[12] Khatri, S., Kokane, P., Kumar, V. et al. Prediction of waterlogged zones under heavy rainfall conditions using machine learning and GIS tools: a case study of Mumbai. GeoJournal (2022). https://doi.org/10.1007/s10708-022-10731-3.

[13] I. Elkhrachy, “Flash Flood Water Depth Estimation Using SAR Images, Digital Elevation Models, and Machine Learning Algorithms,” Remote Sensing, vol. 14, no. 3, Feb. 2022, doi: 10.3390/rs14030440.

[14] M. F. Abdullah, S. Siraj, and R. E. Hodgett, “An overview of multi-criteria decision analysis (Mcda) application in managing water-related disaster events: Analyzing 20 years of literature for flood and drought events,” Water (Switzerland), vol. 13, no. 10. MDPI AG, May 02, 2021. doi: 10.3390/w13101358.

[15] J. Chen, G. Huang, and W. Chen, “Towards better flood risk management: Assessing flood risk and investigating the potential mechanism based on machine learning models,” Journal of Environmental Management, vol. 293, Sep. 2021, doi: 10.1016/j.jenvman.2021.112810.

[16] B. Kalantar et al., “Deep neural network utilizing remote sensing datasets for flood hazard susceptibility mapping in Brisbane, Australia,” Remote Sensing, vol. 13, no. 13, Jul. 2021, doi: 10.3390/rs13132638.

[17] “Early Flood Detection and Alarming System Using Machine Learning Techniques,” 2020.

[18] V. Katiyar, N. Tamkuan, and M. Nagai, “FLOOD AREA DETECTION USING SAR IMAGES WITH DEEP NEURAL NETWORK DURING, 2020 KYUSHU FLOOD JAPAN,” 2020.