International Journal of Recent Research and Review, Special Issue 2, 2026

ISSN 2277 – 8322

An Intelligent Machine Learning Framework for Accurate Weather Forecasting Using Historical Meteorological Data

Neelam Soni, Dr. Pradeep Jha, Pankaj Jain

Department of Computer Science and Engineering, Global Institute of Technology, Jaipur

neelamggc16@gmail.com, pradeep.jha@gitjaipur.com, pankaj.jain@gitjaipur.com

Abstract: Weather forecasting is a critical task in environmental science with significant applications in agriculture, transportation, urban planning, disaster management, and energy systems. This study presents a machine learning-based framework for accurate maximum temperature prediction using historical weather data collected from Bengaluru, India. The dataset consists of 96,432 hourly observations with 25 meteorological attributes. A comprehensive preprocessing pipeline, including data cleaning, categorical encoding, feature scaling, and outlier removal, produced a refined dataset of 63,703 records. Feature selection techniques, namely Sequential Feature Selection (SFS) and Recursive Feature Elimination (RFE), identified the most influential predictors, including minimum temperature, solar hours, UV index, wind chill, wind gust speed, and ambient temperature. Multiple regression models, including Linear Regression, K-Nearest Neighbors (KNN), Decision Tree, Multilayer Perceptron (MLP), and Random Forest, were trained and evaluated using MAE, MSE, RMSE, and the coefficient of determination (R^2). Experimental results demonstrate that ensemble and nonlinear machine learning models outperform conventional linear approaches, with the Random Forest Regressor achieving the highest prediction accuracy ($R^2 > 0.93$), followed closely by KNN. The findings indicate that effective feature engineering and data preprocessing significantly enhance forecasting performance, making machine learning a reliable and accurate solution for short-term weather prediction.

Keywords: Weather Forecasting, RFE, KNN, Machine Learning, Accuracy.

1. INTRODUCTION

Weather forecasting is the scientific process of predicting atmospheric conditions at a specific location and time based on the analysis of meteorological observations [1]. Weather data are collected from multiple sources, including

ISSN 2277 – 8322

satellites, radar systems, ground weather stations, weather balloons, and ocean buoys, which continuously monitor parameters such as temperature, atmospheric pressure, humidity, wind speed, precipitation, and solar radiation [2]. Accurate weather forecasting plays a vital role in modern society by supporting decision-making across diverse sectors, including agriculture, transportation, energy management, disaster preparedness, and urban planning [3]. In agriculture, reliable weather predictions assist farmers in planning irrigation, crop cultivation, fertilizer application, and harvesting, thereby improving productivity and reducing losses caused by adverse weather conditions [4], [5]. Similarly, the aviation, maritime, and road transportation sectors rely heavily on weather forecasts to enhance operational efficiency and ensure public safety.

Weather forecasting is also essential for disaster management, providing early warnings for extreme weather events such as cyclones, hurricanes, floods, thunderstorms, and heatwaves. Timely forecasts enable government agencies and emergency responders to implement evacuation plans, allocate resources effectively, and minimize both human casualties and economic losses. Recent advances in remote sensing, satellite technology, high-performance computing, and numerical weather prediction have significantly improved forecasting capabilities. Furthermore, the rapid development of artificial intelligence (AI) and machine learning (ML) has transformed modern weather prediction by enabling the analysis of large-scale historical meteorological datasets and the discovery of complex nonlinear relationships among weather variables. Compared with conventional statistical approaches, machine learning models offer improved predictive performance, adaptability, and scalability for short-term weather forecasting.

Motivated by these advancements, this study develops a machine learning-based framework for predicting maximum temperature using historical weather data collected from Bengaluru, India. A comprehensive data preprocessing pipeline, feature selection techniques, and multiple regression models, including Linear Regression, K-Nearest Neighbors (KNN), Decision Tree, Multilayer Perceptron (MLP), and Random Forest, are employed and comparatively evaluated. The proposed framework aims to improve forecasting accuracy while demonstrating the effectiveness of machine learning techniques for reliable short-term weather prediction.

ISSN 2277 – 8322

2. PROPOSED METHODOLOGY

A. Data Collection and Preprocessing

Historical weather data consisting of 96,432 hourly observations with 25 meteorological attributes was used for this study. The dataset includes temperature, humidity, atmospheric pressure, wind characteristics, precipitation, solar radiation, cloud cover, and astronomical variables. The target variable selected for prediction is maximum temperature (maxtempC) due to its importance in weather forecasting, agriculture, and energy management.

The preprocessing stage involved removing duplicate records, eliminating irrelevant features (such as constant-value attributes), handling categorical variables through Label Encoding, and standardizing numerical features using the StandardScaler to ensure uniform feature scaling. Outliers were detected and removed using the Interquartile Range (IQR) method, resulting in a refined dataset containing 63,703 high-quality observations.

B. Feature Analysis and Selection

A Pearson correlation analysis was performed to examine relationships among meteorological variables and identify significant predictors. The analysis revealed strong correlations between maximum temperature and variables such as minimum temperature, solar hours, UV index, ambient temperature, wind chill, and wind gust speed.

To select the most informative features, two feature selection techniques were employed:

- Sequential Feature Selection (SFS)
- Recursive Feature Elimination (RFE)

Both methods consistently identified the most influential predictors, reducing redundancy while improving model performance.

C. Model Development

The processed dataset was divided into 75% training and 25% testing subsets to ensure unbiased evaluation. Five regression-based machine learning algorithms were implemented and compared:

- Linear Regression (LR)
- K-Nearest Neighbors (KNN) Regressor
- Multilayer Perceptron (MLP) Regressor
- Decision Tree (DT) Regressor
- Random Forest (RF) Regressor

These models were trained using the selected meteorological features to predict maximum temperature.

ISSN 2277 – 8322

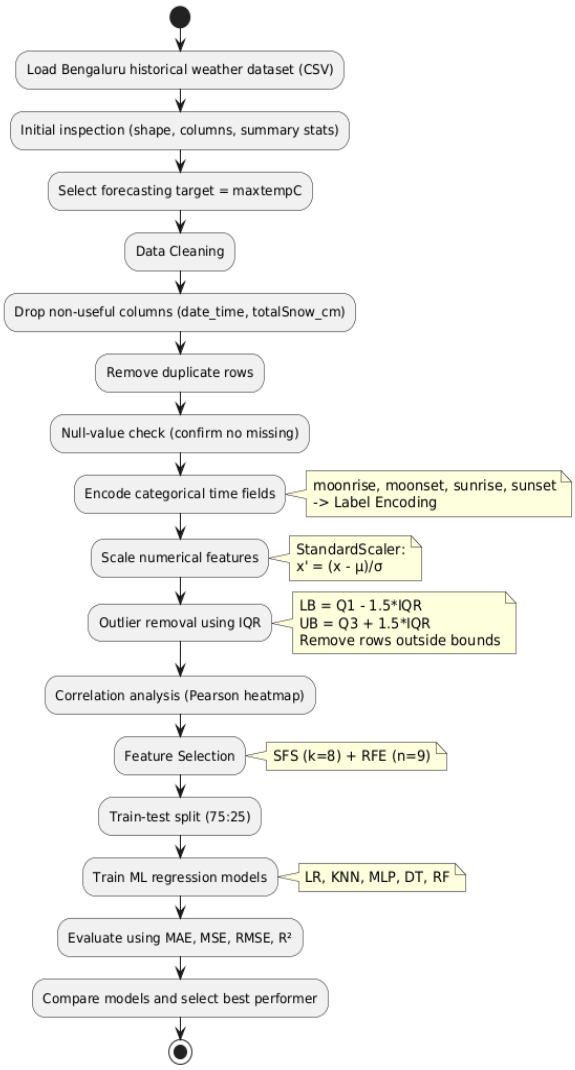

Figure 1: Overall Proposed Methodology

D. Performance Evaluation

The predictive performance of each model was evaluated using standard regression metrics, including:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Coefficient of Determination (R^2)

A comparative analysis was conducted to determine the most effective forecasting model. The methodology combines comprehensive data preprocessing, feature engineering, correlation analysis, and machine learning-based regression to develop an accurate and reliable weather prediction system suitable for real-world forecasting applications.

3. RESULTS AND DISCUSSION

Error-based metrics further quantify prediction quality.

Table 41: Error Metrics (Linear Regression Baseline)

Metric	Value
MAE	0.346
MSE	0.149
RMSE	0.447
R^2 (%)	82.66

Lower RMSE indicates good predictive stability for baseline models. A comparative ranking is performed based on accuracy, generalization, and robustness.

Table 2: Comparative Assessment and Ranking

Rank	Model	Strength
1	Random Forest	Highest accuracy, strong generalization
2	KNN Regressor	Excellent local prediction

ISSN 2277 – 8322

3	MLP Regressor	Stable nonlinear learning
4	Decision Tree	High accuracy but overfitting
5	Linear Regression	Interpretable but limited

The experimental results indicate that nonlinear and ensemble machine learning models outperform conventional linear models in forecasting maximum temperature. Among all the evaluated models, the Random Forest Regressor achieved the highest prediction accuracy by effectively capturing complex meteorological relationships through an ensemble of decision trees. The K-Nearest Neighbors (KNN) model also demonstrated strong performance by exploiting local similarities in weather patterns, while the Multilayer Perceptron (MLP) effectively modeled nonlinear relationships within the data. These findings highlight the superiority of advanced machine learning techniques for accurate and reliable weather forecasting.

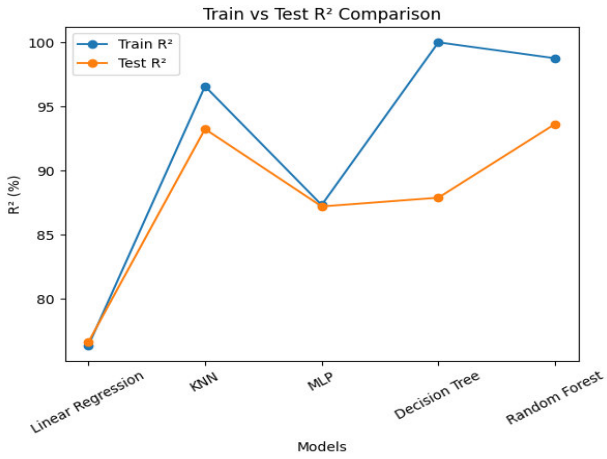

Figure 2: Comparative Analysis of R^2 Scores (Training and Testing)

Correlation analysis and feature selection confirmed that temperature-related variables, solar radiation, and wind characteristics are the most influential predictors of maximum temperature. Strong positive correlations among variables such as tempC, FeelsLikeC, HeatIndexC, and WindChillC demonstrate their physical consistency while highlighting the need for feature selection to reduce multicollinearity. Maximum temperature also showed significant positive correlations with UV index and sun hours, reflecting the impact of solar radiation, whereas its negative correlation with humidity aligns with the cooling effects of moisture and cloud formation. Wind-related variables exhibited moderate correlations, indicating their contribution to temperature variation, particularly in nonlinear models. Overall, the preprocessing pipeline, correlation

ISSN 2277 – 8322

analysis, and feature selection improved model stability, robustness, and forecasting accuracy, making the proposed framework suitable for real-world weather prediction applications.

Effect of the Feature Selection on Model Stability

Sequential Feature Selection (SFS) and Recursive Feature Elimination (RFE) were employed to identify the most significant predictors for maximum temperature forecasting. Both techniques consistently selected key features, including minimum temperature (mintempC), solar hours (sunHour), UV index (uvIndex), ambient temperature (tempC), and wind-related variables, demonstrating the robustness and reliability of the feature selection process.

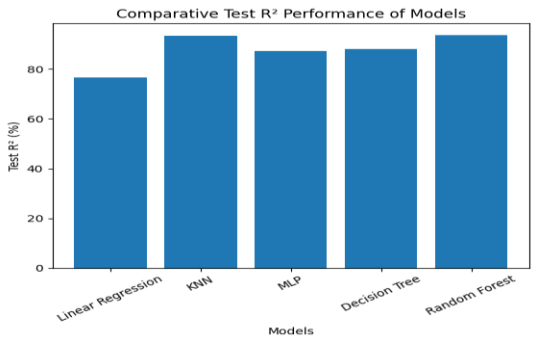

Figure 3: Comparative Bar Chart for Model Performance

The reduced feature subset minimized dimensionality and overfitting while improving computational efficiency and model interpretability. This enhancement was particularly evident in the Random Forest and K-Nearest Neighbors (KNN) models, which achieved strong generalization performance without a significant gap between training and testing accuracy. Consequently, the proposed feature selection framework provides an efficient and reliable foundation for real-time weather forecasting applications.

4. CONCLUSION

Weather forecasting is an essential application that supports decision-making in agriculture, transportation, disaster management, energy planning, and public safety. This study presented a machine learning-based framework for predicting maximum temperature using a large historical weather dataset from Bengaluru, India. The dataset was preprocessed through data cleaning, feature scaling, outlier removal, and feature selection to improve model performance and reliability.

Five regression models Linear Regression, K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Decision Tree, and Random Forest—were developed and evaluated using MAE, MSE, RMSE, and R^2 metrics. Experimental results demonstrated that nonlinear and ensemble learning approaches significantly outperformed conventional linear models. Among all models, the Random Forest Regressor achieved the highest prediction

accuracy and generalization capability, making it the most effective model for short-term weather forecasting.

The findings confirm that machine learning techniques, when combined with effective preprocessing and feature selection, provide accurate, reliable, and scalable weather prediction solutions. Such models can support real-world applications in climate-sensitive sectors, contributing to improved planning, resource management, and sustainable decision-making.

REFERENCES

[1] M. Biswas, T. Dhoom, and S. Barua, "Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data," International Journal of Computer Applications, vol. 182, no. 34, pp. 20–24, 2018.

[2] H. B. Bluestein, F. H. Carr, and S. J. Goodman, "Atmospheric Observations of Weather and Climate," Atmosphere-Ocean, vol. 60, 2022.

[3] H. Zhang, Y. Liu, C. Zhang, and N. Li, "Machine Learning Methods for Weather Forecasting: A Survey," Atmosphere, vol. 16, no. 1, Art. no. 82, 2025.

[4] B. Collins, Y. Lai, U. Grewer, S. Attard, J. Sexton, and K. G. Pembleton, "Evaluating the Impact of Weather Forecasts on Productivity and Environmental Footprint of Irrigated Maize Production Systems," Science of The Total Environment, vol. 954, 2024.

[5] P. Jha, G. K. Soni, H. Dogra, D. Goswami, K. Choudhary, and H. Vaishnav, "Plant Disease Detection and Classification Using Convolutional Neural Network," in Proc. 4th Int. Conf. on Automation, Computing and Renewable Systems (ICACRS), 2025, pp. 1442–1446.

[6] Aayushi, L. Yadav, P. Paliwal, A. Johari, R. Ajmera, and G. K. Soni, "A Simulation-Based Evaluation of Machine Learning Models for Algorithmic Trading in Equity Markets," in Proc. 6th Int. Conf. on Expert Clouds and Applications (ICOECA), 2026, pp. 1049–1054.

[7] M. K. Jha, G. K. Soni, G. Jain, S. Tiwari, K. Gupta, and B. Singhal, "Comparative Analysis of Classical Machine Learning Models for Twitter Sentiment Classification," in Proc. 5th Int. Conf. on Communication, Computing and

ISSN 2277 – 8322

Electronics Systems (ICCCES), 2026, pp. 1949–1954.

[8] S. A. Saiyed, N. Sharma, H. Kaushik, P. Jain, G. K. Soni, and R. Joshi, "Transforming Portfolio Management with AI and ML: Shaping Investor Perceptions and the Future of the Indian Investment Sector," in Proc. Parul University Int. Conf. on Engineering and Technology (PiCET), 2025, pp. 1108–1114.

[9] P. Jha, D. Dembla, and W. Dubey, "Deep Learning Models for Enhancing Potato Leaf Disease Prediction: Implementation of Transfer Learning Based Stacking Ensemble Model," Multimedia Tools and Applications, vol. 83, pp. 37839–37858, 2024.

[10] P. Jha, D. Dembla, and W. Dubey, "Comparative Analysis of Crop Diseases Detection Using Machine Learning Algorithm," in Proc. 3rd Int. Conf. on Artificial Intelligence and Smart Energy (ICAIS), 2023, pp. 569–574.

[11] P. Jha, D. Dembla, and W. Dubey, "Crop Disease Detection and Classification Using Deep Learning-Based Classifier Algorithm," in Emerging Trends in Expert Applications and Security, Lecture Notes in Networks and Systems, vol. 682, 2023.

[12] P. Jha, D. Dembla, and W. Dubey, "Implementation of Machine Learning Classification Algorithm Based on Ensemble Learning for Detection of Vegetable Crops Disease," International Journal of Advanced Computer Science and Applications, vol. 15, no. 1, 2024.

[13] P. Jha, D. Dembla, and W. Dubey, "Implementation of Transfer Learning Based Ensemble Model Using Image Processing for Detection of Potato and Bell Pepper Leaf Diseases," International Journal of Intelligent Systems and Applications in Engineering, vol. 12, pp. 69–80, 2024.

[14] N. Soni and N. Nigam, "Recent Advances in Artificial Intelligence and Machine Learning: Trends, Challenges, and Future Directions," International Journal of Engineering Trends and Applications (IJETA), vol. 12, no. 1, pp. 9–12, 2025.