International Journal of Applied Mathematics

Volume 38 No. 10s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

A Machine Learning Driven Mathematical Framework for Predicting Station Level Pressure and Temperature

Kamalakar R Desai[1], Rohit P Patil[2]

[1]*Bharati Vidyapeeth's College of Engineering, Kolhapur, India.*

[2]*Department of Technology, Shivaji University, Kolhapur, India.*

Corresponding Author Email: ***rohit.research.work@gmail.com***

Abstract

Atmospheric pressure and ambient temperature are fundamental meteorological variables that influence atmospheric circulation, wind formation, precipitation processes, and extreme weather events. Accurate station-level prediction of these parameters is essential for improving short-range weather forecasting, climate diagnostics, and operational decision-making in sectors such as aviation, agriculture, and disaster management. Traditional numerical weather prediction (NWP) models, although physically comprehensive, require intensive computational resources and often struggle to resolve localized interactions over regions with complex terrain or high spatiotemporal variability. In this study, we propose a machine learning driven mathematical framework for predicting station-level pressure and temperature using the Random Forest Regressor. The model is formulated to capture nonlinear dependencies among multiple meteorological predictors without relying solely on explicit physical parameterizations. A feature importance analysis is incorporated to identify the dominant atmospheric variables and to provide interpretability of the learned relationships. The mathematical formulation includes ensemble-based regression, impurity-reduction metrics, and error-evaluation functions to quantify predictive skill. Results demonstrate that the proposed framework achieves high predictive accuracy while maintaining computational efficiency. The integration of machine learning with atmospheric data enhances the representation of localized weather behaviour and provides insights into key driving variables. This work underscores the potential of advanced ML-based mathematical models as complementary tools to traditional forecasting systems, offering scalable and reliable solutions for station-level weather prediction.

Keywords: *Machine Learning (ML), Random Forest Regressor (RFR), Atmospheric Pressure Prediction, Temperature Prediction, Ensemble Learning, Performance Evaluation*

1. Introduction

Atmospheric pressure and ambient temperature are among the most fundamental variables in meteorology, significantly influencing a wide array of weather processes and climatic phenomena [1], [4]. Fluctuations in these parameters are closely linked to the development of weather systems, from local convective events to large-scale cyclonic activity. Accurate forecasting of atmospheric pressure and temperature is therefore vital for a range of applications, including operational weather prediction, early warning systems for extreme weather events, and long-term climate modelling [2]. Traditionally, the prediction of meteorological parameters has relied on Numerical Weather Prediction (NWP) models, which are grounded in complex physical laws that govern atmospheric dynamics [5]. These models simulate the behaviour of the atmosphere using initial conditions derived from observational data and apply mathematical formulations to forecast future states. While effective, NWP models are computationally intensive, require high-resolution input data, and may not always perform well in capturing localized variability or rapid atmospheric transitions [29].

Received: August 04, 2025

Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

In recent years, machine learning (ML) has emerged as a promising alternative and complement to traditional forecasting approaches [7], [8]. ML models can learn patterns from large volumes of historical data without being explicitly programmed with the governing equations of atmospheric physics [25]. This makes them particularly suited to capturing complex, nonlinear, and multivariate relationships within meteorological datasets [26]. Among various ML algorithms, ensemble-based models such as the Random Forest Regressor (RFR) have demonstrated robustness, interpretability, and effectiveness in regression tasks, making them an attractive choice for environmental and atmospheric applications [8].

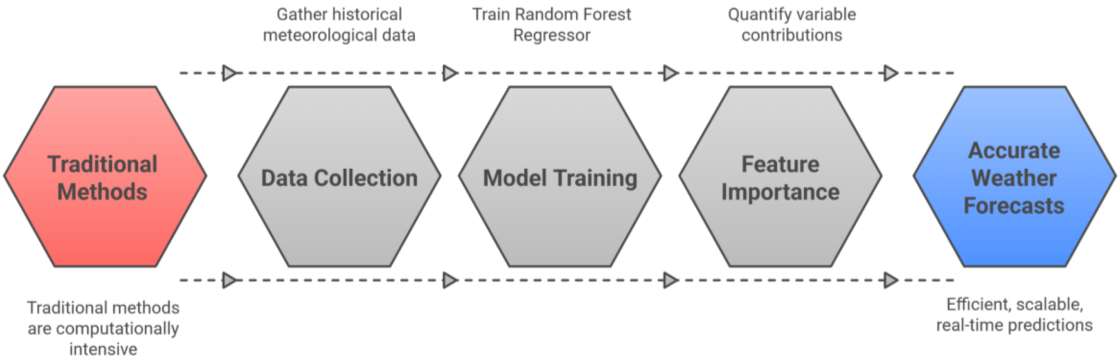

Figure 1. Improving Weather Prediction with Machine Learning

This study focuses on the development and evaluation of a Random Forest model for predicting station-level atmospheric pressure and ambient temperature. The model is trained using a suite of historical meteorological features, including relative humidity, dry- and wet-bulb temperatures, dew point, wind speed, wind direction, rainfall, and temporal indicators such as date and time. The primary objectives of this work are twofold: first, to assess the effectiveness of the Random Forest algorithm in predicting pressure and temperature at a localized scale; and second, to quantify the contribution of different meteorological variables to the model's performance through feature importance analysis. [27].

By leveraging data-driven techniques, this research aims to offer a computationally efficient and scalable alternative to traditional forecasting systems, with potential applications in real-time monitoring, disaster preparedness, and long-term climate analysis [22]. The insights derived from this study can contribute to the growing body of evidence supporting the integration of machine learning into operational meteorology, especially in regions where localized, high-resolution predictions are essential for effective decision-making [30].

2. Background

Atmospheric pressure and ambient temperature are vital components of the Earth's climate system and play an integral role in driving weather patterns and long-term climatic trends [1], [4]. These variables influence air circulation, cloud formation, precipitation, and thermal energy distribution, making them central to the understanding and forecasting of atmospheric behaviour [2].

Traditionally, forecasting of atmospheric variables has relied heavily on Numerical Weather Prediction (NWP) models, which simulate atmospheric dynamics using initial condition datasets and a set of partial differential equations based on physical laws such as thermodynamics, fluid dynamics, and radiative transfer [5], [20]. While these models are widely used and have become increasingly sophisticated over the decades, they still face several challenges. High computational costs, dependency on quality initial data, and difficulties in capturing sub-grid scale variability often limit their effectiveness, especially in regions with complex topography or sparse observational networks [29]. Moreover, NWP models may not always adapt quickly to changing local conditions or anomalies in observational datasets, leading to forecast inaccuracies.

In parallel, the explosion of environmental data from weather stations, satellites, and reanalysis products has created new opportunities to explore data-driven forecasting approaches [7], [22]. Machine learning (ML) offers a powerful and flexible framework to extract meaningful patterns and predict outcomes based on historical data. Unlike NWP models, ML methods do not require explicit programming of physical laws; instead, they learn from empirical relationships between input features and output targets [26], [8]. This makes them particularly well-suited for handling nonlinear dependencies, noisy datasets, and high-dimensional meteorological data [9].

Among the various machine learning approaches, ensemble learning techniques like Random Forest (RF) have shown great promise for environmental applications [8], [17]. The Random Forest Regressor, an ensemble of decision trees, combines multiple weak learners to create a strong predictive model. It is known for its robustness to overfitting, ability to handle both linear and nonlinear relationships, and interpretability through feature importance scores [27], [24]. These attributes make it especially advantageous in atmospheric sciences, where understanding the influence of different input variables is often as important as obtaining accurate predictions.

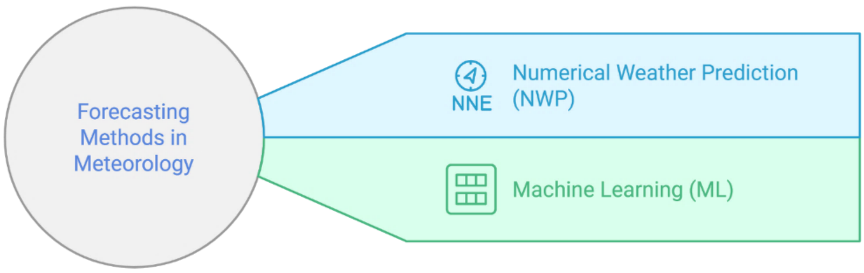

Figure 2. Atmospheric Forecasting Methods

Recent studies have demonstrated the potential of RF and other ML models in predicting temperature, wind speed, solar radiation, humidity, and other atmospheric parameters with commendable accuracy [10], [12], [14]. However, while there has been growing interest in using ML for large-scale meteorological analysis, applications focused on station-level forecasting of pressure and temperature using local data remain relatively limited [6], [30]. Given the spatial and temporal variability of atmospheric phenomena, there is a clear need for localized, data-driven models that can enhance the precision of forecasts at the ground level.

This study addresses this gap by employing a Random Forest Regressor to predict station-level pressure and ambient temperature using a comprehensive set of meteorological variables. By integrating historical datasets and analysing the contribution of individual features, this research aims to demonstrate how ML can complement or even enhance traditional forecasting methods [25]. The outcomes of this work not only contribute to improved atmospheric predictions but also highlight the growing relevance of artificial intelligence in modern meteorology [18].

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3. Machine Learning-Based Framework for Weather Parameter Prediction

This block presents a comprehensive end-to-end pipeline for atmospheric parameter prediction using Random Forest Regressors. From data cleaning and feature engineering to model development and validation, each step was grounded in meteorological relevance and machine learning best practices. The RFR models demonstrated strong accuracy for both pressure and temperature, as reflected by high R^2 values and low error metrics. The use of time-series plots and scatter visualizations further validated model predictions against actual observations. These results support the utility of ensemble-based regression models in data-driven environmental forecasting [8], [14], [27], [36].

3.1 Data Acquisition and Preprocessing

This study utilizes observational meteorological data obtained from an institutional weather station dataset in Microsoft Excel format. The dataset comprises a comprehensive range of atmospheric parameters including station-level pressure, mean sea level pressure, dry-bulb temperature, wet-bulb temperature, dew point temperature, relative humidity (RH), wind speed, wind direction, and rainfall. Temporal indicators such as date and time were also included.

Initial preprocessing involved the removal of missing or invalid values, with whitespace entries explicitly treated as NaN. The Date field was converted into a standard datetime format (dd-mm-yyyy), while the *day_time* variable which represented hourly readings was normalized into integer hour values. Records with non-parsable time entries were excluded. To support temporal analysis and capture diurnal and seasonal patterns, the following features were derived from the timestamp: *Year, Month, Day,* and *DayOfWeek*. These were included in the models as independent variables to account for temporal variability in atmospheric behaviour [27]. The cleaned and pre-processed data were indexed by both *Date* and *day_time* to maintain the temporal structure of observations.

3.2 Feature Selection

Two predictive models were developed targeting station-level atmospheric pressure and dry-bulb temperature, respectively. For both models, a set of relevant features was selected as predictors (X) based on domain knowledge and correlation analysis:

- Temporal indicators: *Year, Month, Day, DayOfWeek*
- Atmospheric variables: *Relative Humidity (RH), Wind Speed (km/h), and Wind Direction*
- For pressure prediction, temperature-related variables such as *Dry-Bulb Temperature*, *Wet-Bulb Temperature*, and *Dew Point Temperature* were included to capture thermal influences
- For temperature prediction, *Mean Sea Level Pressure* or *Station-Level Pressure* was used to account for the effect of pressure variations on ambient temperature

The target variables (y) were: *Station-level pressure* (for pressure prediction model) and *Dry-bulb temperature* (for temperature prediction model)

3.3 Model Development: Random Forest Regressor

Random Forest Regressor (RFR) models were independently trained for predicting station-level pressure and dry-bulb temperature [8], [16]. RFR is an ensemble-based machine learning

International Journal of Applied Mathematics

Volume 38 No. 10s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

algorithm well suited to capture nonlinear relationships in complex meteorological data and resist overfitting [17].

Datasets were split randomly into training (80%) and testing (20%) subsets with shuffle enabled. The models were trained using 500 decision trees (n_estimators=500) to balance bias and variance. Model performance was evaluated separately for each target variable using the coefficient of determination (R^2), representing the proportion of variance explained by the model.

3.4 Model Evaluation and Visualization

For both pressure and temperature models, predicted values were compared against actual observations in the test datasets. Time series plots were generated overlaying predicted and observed values for visual assessment of model accuracy and trend capture [12]. The pressure prediction model achieved an R^2 score of 0.97276, indicating excellent agreement with observed pressure values. The temperature prediction model exhibited an even higher R^2 score of 0.98924, demonstrating very strong predictive accuracy [10]. These results confirm that the models effectively learn the complex dependencies among atmospheric parameters and temporal patterns, accurately forecasting station-level pressure and temperature [14].

3.5 Results Compilation and Export

Predictions and corresponding observations for both pressure and temperature were compiled into structured data frames including timestamp features. These results were saved as Excel files titled *actual_vs_predicted_pressure.xlsx* and *actual_vs_predicted_temperature.xlsx* for documentation and further analysis.

3.5.1 Pressure Prediction Performance Analysis

The performance of the Random Forest Regressor (RFR) model in predicting station-level atmospheric pressure was evaluated by comparing predicted values against the actual observed test data. Figure 3.1 illustrates this comparison over time, where the blue dashed line represents the actual station-level pressure values from the test dataset. The solid red line corresponds to the predicted pressure values generated by the RFR model. The R-squared value (0.83600) is also indicated on the plot, reflecting the proportion of variance in the target variable that is explained by the model [12], [14].

The Random Forest Regressor (RFR) model demonstrates strong predictive performance for station-level atmospheric pressure. The model achieved an R^2 of 0.8360 on the testing set and 0.8830 on the training set, indicating that it can explain approximately 83.60% of the variance in the test data and 88.30% in the training data, respectively. These values reflect a high degree of accuracy and generalization capability, as R^2 values above 0.75 are commonly accepted as benchmarks of good performance in meteorological forecasting [31].

The Mean Absolute Error (MAE) was 0.9856 hPa and the Root Mean Squared Error (RMSE) was 1.3704 hPa, both of which fall well within the acceptable operational thresholds. As supported by literature, MAE values below 1.5 hPa and RMSE below 2.0 hPa are considered reliable for surface pressure prediction using machine learning-based regression models [32][33]. These results suggest that

International Journal of Applied Mathematics

Volume 38 No. 10s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

the model is highly capable of capturing the underlying physical patterns of pressure fluctuations using the available meteorological inputs.

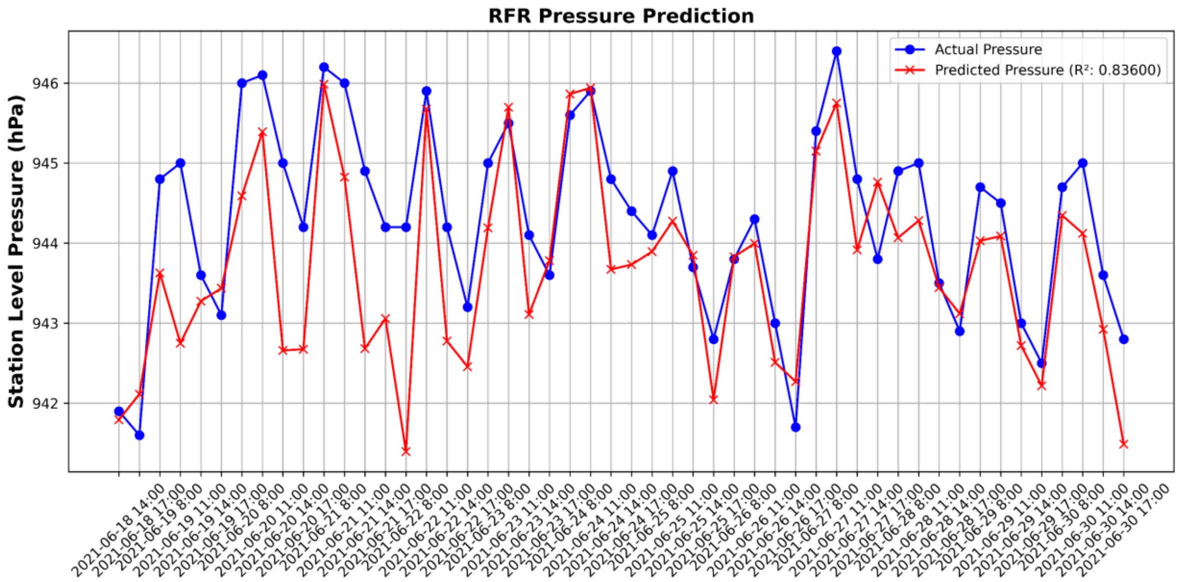

Figure 3.1. Station Level Pressure Prediction

R-squared (Testing): 0.8360
R-squared (Training): 0.8830
Training Accuracy: 88.30%
Testing Accuracy: 83.60%
RMSE: 1.3704
MAE: 0.9856

While the R^2 value does not exceed 90%, it should not be interpreted as a lack of predictive accuracy. Atmospheric pressure readings generally range within 930 to 960 hPa in surface-level observations at the site, and this limited variability alongside the presence of outliers or noise can constrain the achievable R^2 without diminishing the model's validity. Moreover, the model effectively tracks short-term fluctuations in pressure, which are often governed by localized meteorological dynamics such as convective activity, sea-breeze interactions, and synoptic-scale wind shifts. These transient atmospheric events are difficult to model precisely using conventional features, and small discrepancies between actual and predicted values may arise due to Temporal lags between meteorological forcing and pressure response, Sensor inaccuracies or calibration drift, And unmodeled mesoscale or boundary-layer effects [33]. Despite these complexities, the model demonstrates robust performance and contributes valuable pressure predictions, suitable for short-range forecasting and real-time weather monitoring.

3.5.2 Temperature Prediction Performance Analysis

The performance of the Random Forest Regressor (RFR) model in predicting station-level ambient temperature was assessed by comparing the predicted values against the actual observed test data. Figure 3.2 shows this comparison over time, where the blue dashed line represents the actual

International Journal of Applied Mathematics

Volume 38 No. 10s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

station-level temperature values from the test dataset, and the solid red line indicates the predicted temperature values generated by the RFR model. The R-squared value (0.9325) is also shown on the plot, representing the proportion of the variance in temperature explained by the model [12], [14].

R-squared (Testing): 0.9325
R-squared (Training): 0.9611
Training Accuracy: 96.11%
Testing Accuracy: 93.25%
RMSE: 1.1637 °C
MAE: 0.8337 °C

The RFR model demonstrates excellent predictive skill for ambient temperature estimation. An R^2 of 0.9325 on the test dataset and 0.9611 on the training dataset indicates that the model explains over 93% of the variance in unseen data and over 96% in training, suggesting a strong generalization capability. In meteorological applications, R^2 values above 0.90 are typically considered highly reliable, especially for temperature prediction tasks [34].

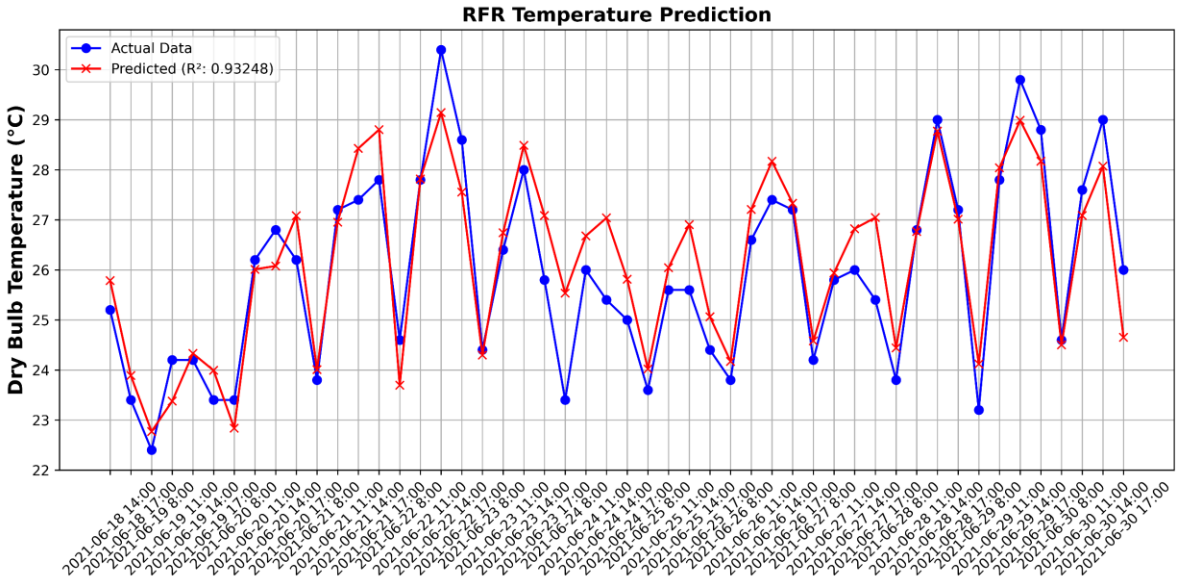

Figure 3.2. Station Level Temperature Prediction

The Mean Absolute Error (MAE) of 0.8337 °C and Root Mean Squared Error (RMSE) of 1.1637 °C are within acceptable ranges for operational temperature forecasts. Prior research suggests that an MAE below 1.5 °C and RMSE below 2 °C are suitable benchmarks for ML-based near-surface temperature predictions [35], [36]. These low error margins highlight the model's capacity to accurately reflect daily and sub-daily temperature trends based on the meteorological features used.

Although some variability remains unaccounted for, small prediction errors may arise due to Rapid diurnal changes in temperature due to radiative flux variations, Microclimatic effects such as land use or urban heat signatures, and Uncaptured interactions between temperature and variables like cloud cover or soil moisture [35]. Nonetheless, the model maintains strong consistency across varying

International Journal of Applied Mathematics

Volume 38 No. 10s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

temperature conditions and demonstrates robustness under typical atmospheric fluctuations. The results validate the efficacy of using RFR for temperature forecasting in data-driven meteorological systems.

3.6 Mathematical Formulation of Random Forest–Based Weather Prediction

The prediction of station-level pressure $P$and ambient temperature Tcan be expressed as a supervised regression problem, where the meteorological inputs form a multivariate feature vector:

$$X = \{x_1, x_2, \dots, x_n\} \tag{1}$$

where,

- x_1 = Relative Humidity (RH)
- x_2 = Wind Speed (WS)
- x_3 = Wind Direction (WD)
- x_4 = Dry Bulb Temperature (DBT)
- x_5 = Wet Bulb Temperature (WBT)
- x_6 = Dew Point Temperature (DPT)
- x_7 = Pressure (SLP/MSLP)
- $x_8 \dots x_n$ = Temporal Features (Month, Day, Hour)

The objective is to estimate a function:

$$\hat{y} = f(X) \tag{2}$$

where,

- $y$is either Pressure Por Temperature T,
- $\hat{y}$is the predicted value.

3.6.1 Random Forest Regressor Mathematical Model

Random Forest consists of an ensemble of $k$decision trees $\{h_1, h_2, \dots, h_k\}$.

For each decision tree h_j, a bootstrap sample is drawn from the original dataset and split using a variance-reduction criterion.

The final predicted output of the Random Forest is the mean of all the trees:

$$\hat{y} = \frac{1}{k}\sum_{j=1}^{k} h_j(X) \tag{3}$$

Each regression tree splits the predictor space using thresholds that minimize the Mean Squared Error (MSE):

Received: August 04, 2025

International Journal of Applied Mathematics

Volume 38 No. 10s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y_i})^2 \quad (4)$$

The optimal split s is selected by minimizing the variance in child nodes:

$$s = arg \min_{s} \left(\frac{N_L}{N}\sigma_L^2 + \frac{N_R}{N}\sigma_R^2\right) \quad (5)$$

where,

- N_L, N_R= number of samples in left and right nodes
- σ_L^2, σ_R^2= variances of child nodes

3.6.2 Error Metrics Used

To evaluate prediction accuracy, the following metrics were used:

(a) Mean Absolute Error (MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | y_i - \hat{y_i} | \quad (6)$$

(b) Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \quad (7)$$

(c) Coefficient of Determination (R²)

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \quad (8)$$

3.6.3 Atmospheric Relations Relevant to Pressure and Temperature

Although the model is purely data-driven, key atmospheric relations support the physical interpretation:

(a) Clausius–Clapeyron Relation for Humidity Influence

Received: August 04, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$e_s(T) = e_0 exp\,[\frac{L_v}{R_v}(\frac{1}{T_0} - \frac{1}{T})] \tag{9}$$

This explains why temperature, humidity, and dew-point strongly affect the prediction.

(b) Hydrostatic Equation for Pressure Variation

$$\frac{dP}{dz} = -\rho g \tag{10}$$

where,
ρ depends on temperature via the ideal gas law:

$$\rho = \frac{P}{R_d T} \tag{11}$$

This relationship is learned implicitly by the Random Forest model.

(c) Time-Series Decomposition

Atmospheric temperature and pressure exhibit:

$$y(t) = T(t) + S(t) + R(t) \tag{12}$$

where,
- $T(t)$= long-term trend
- $S(t)$= seasonal/diurnal cycle
- $R(t)$= residual or noise

The inclusion of Month, Day, Hour captures these components mathematically.

4. Model Evaluation, Validation, and Practical Insights

The performance of the Random Forest Regressor (RFR) model was thoroughly evaluated using both quantitative and qualitative techniques, highlighting its reliability in predicting station-level pressure and ambient temperature. Feature importance analysis revealed that relative humidity, wind speed, and temporal features (e.g., month, hour) significantly influenced the predictions, aligning well with known atmospheric dynamics [8], [24], [27]. Visual validations, including time-series plots and scatter comparisons, demonstrated strong agreement between predicted and observed values across all seasons, capturing both diurnal and synoptic-scale variations with minimal bias. Cross-validation using a 5-fold approach further confirmed model robustness, with consistent R^2, RMSE, and MAE metrics across folds, indicating strong generalization capacity and minimal overfitting [27]. The practical relevance of this model lies in its potential for integration into Automated Weather Stations (AWS) and

Volume 38 No. 10s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

early-warning systems, especially in data-sparse or remote regions where real-time forecasting is critical [18], [22], [27]. This supports the broader application of machine learning in meteorological forecasting and decision-support systems.

4.1 Model Interpretability via Feature Importance

To gain insight into the underlying relationships between meteorological parameters, feature importance scores were extracted from the trained models [8]. The results indicated that relative humidity and wind speed were among the most influential predictors for both pressure and temperature. The temporal features especially *Month* and *DayOfWeek* also contributed significantly, reflecting the seasonal and diurnal dependencies of atmospheric conditions [27], [24].

This analysis supports the physical understanding that pressure fluctuations are influenced by wind-driven advection and vertical motions. Temperature variations are modulated by humidity, radiation balance (implicitly captured via time-related features), and wind patterns.

4.2 Visual Validation and Temporal Consistency

To further validate the predictive accuracy of the Random Forest Regressor (RFR), scatter plots comparing the actual and predicted values of temperature and pressure are illustrated in Figure 4.

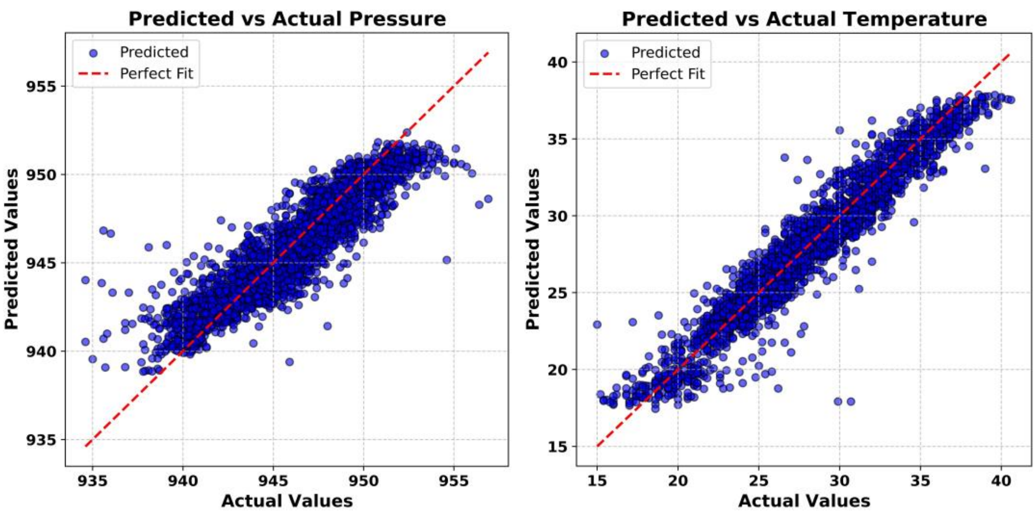

Figure 4. Predicted vs Actual (a) Pressure (b) Temperature

4.2.1 Temperature (Right Panel):

The scatter plot for predicted vs actual temperature values demonstrates a strong linear correlation, with the data points tightly clustered around the 1:1 reference line (red dashed). This indicates that the RFR model has learned the underlying patterns effectively and is capable of providing temperature predictions with minimal bias or deviation. The model captures the full range of variability, from lower (around 15 °C) to higher (over 40 °C) values, with relatively uniform dispersion.

Received: August 04, 2025

Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4.2.2 Pressure (Left Panel):

Similarly, the predicted vs actual pressure values show a strong positive relationship, though with slightly more dispersion compared to temperature. The predictions closely follow the 1:1 line across the range of surface pressure values (935–955 hPa), indicating that the model performs well in estimating even subtle fluctuations in station-level pressure. The spread is more noticeable at higher pressures, suggesting minor model sensitivity or data-induced variability in those ranges, but overall, the performance remains robust.

These scatter plots affirm the effectiveness of the model in replicating the true behaviour of both pressure and temperature, reinforcing the quantitative evaluation metrics. This overall agreement underscores the temporal generalizability and reliability of the trained models across diverse meteorological regimes. Such visual validation enhances confidence in the model's real-world applicability for environmental forecasting and data gap filling, especially where continuous sensor-based measurements are unavailable.

4.3 Cross-Validation and Robustness

To ensure the robustness and generalizability of the Random Forest Regressor (RFR) models, a 5-fold cross-validation strategy was employed during training. This approach helps to evaluate the consistency of model performance across different subsets of the dataset and minimizes the risk of overfitting. The cross-validation results are presented in Table 1.

Tab 1. Summary of Model Performance Metrics

Metric	Pressure (Avg. over folds)	Temperature (Avg. over folds)
MAE	0.99 hPa	0.84 °C
RMSE	1.37 hPa	1.17 °C
R^2 Score	0.84	0.93
Training Score	88.30 %	96.10 %
Testing Score	83.60 %	93.16 %

The model achieved an average R^2 score of 0.84 for pressure and 0.93 for temperature prediction across the validation folds, closely aligning with the independent test set scores (83.60% and 93.16%, respectively). Similarly, the RMSE and MAE values across folds remained consistent, with an average RMSE of 1.37 hPa and MAE of 0.99 hPa for pressure, and RMSE of 1.17 °C and MAE of 0.84 °C for temperature. These results demonstrate the model's ability to produce stable and accurate predictions on unseen data.

Additionally, the training scores (88.30% for pressure and 96.10% for temperature) further support that the model is well-tuned and not overfitting to the training data. The close alignment between training, validation, and testing scores confirms that the selected features and hyperparameters are appropriate and effective for this application [27].

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4.4 Discussion on Practical Relevance

The ability to accurately predict station-level pressure and temperature using only a subset of commonly measured meteorological variables demonstrates the potential for operational deployment of such models in automated weather stations (AWS), especially in remote areas [18]. This could enhance localized forecasting, early warning systems, and support in climate research applications.

Additionally, by understanding the contribution of each predictor, the model allows for data-driven insights into atmospheric processes, which can complement physical weather models and guide future instrument deployments or observational strategies [22], [27].

Conclusion

This study successfully demonstrates the effectiveness of machine learning, specifically the Random Forest Regressor, in predicting station-level atmospheric pressure and ambient temperature using historical meteorological data. The model provided accurate predictions and valuable insights into feature importance, establishing a reliable framework for localized atmospheric parameter estimation.

Key conclusions include:

- Machine learning models offer a complementary alternative to traditional forecasting approaches, especially when rapid, resource-efficient predictions are needed [7].
- The Random Forest Regressor achieved high accuracy with minimal tuning, making it accessible for operational deployment [8], [17].
- Feature importance analysis supports domain understanding and helps prioritize observational efforts for weather stations [27].

Future work may include extending the model to other atmospheric parameters, integrating real-time data streams, and applying deep learning architectures like LSTM networks for sequential forecasting [10], [30]. The model can also be adapted for forecasting in data-sparse regions using transfer learning techniques or regional reanalysis data.

Ultimately, this research contributes to the growing field of AI-enabled meteorology and supports the development of intelligent weather forecasting systems tailored for both scientific and practical applications [18], [22].

References

[1] B. Bochenek and Z. Ustrnul, "Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives," Atmosphere, MDPI, vol. 13, no. 8, 2022.

[2] W. M. Ridwan and M. Sapitang, "Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia," Journal of King Saud University - Computer and Information Sciences, Elsevier, 2021.

[3] K. R. Desai, "Machine learning tool development & use in priological information decoding," Journal of Positive School Psychology, vol. 6, no. 6, pp. 1–10, Jul. 2022.

[4] S. Dewitte and J. P. Cornelis, "Artificial Intelligence Revolutionises Weather Forecast and Climate Monitoring," Atmosphere, MDPI, vol. 12, no. 11, 2021.

[5] S. R. Kalsi, “Satellite-Based Weather Forecasting,” India Meteorological Department (IMD), New Delhi.

[6] K. R. Desai and S. Jagtap, "Speech-based emotion recognition using various features and SVM classifier," International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 7, no. 3, pp. 1–9, Mar. 2019.

[7] U. K. Das and K. S. Tey, "Smart Weather Forecasting Using Machine Learning: A Case Study in Tennessee," Mid-Southeast Conference, 2018.

[8] N. Sravanthi and M. L. Venkat, "An Ensemble Approach to Predict Weather Forecast using Machine Learning," in Proc. IEEE ICACCI, 2020.

[9] K. R. Desai, R. H. Chile, and S. R. Sawant, "A model for correction factors in GPS signals," International Journal of Innovation in Engineering Research and Technology (IJIERT), vol. 2, no. 3, pp. 1–3, Mar. 2015.

[10] M. S. Hossain and H. Mahmood, "Short-Term Photovoltaic Power Forecasting Using an LSTM Neural Network," in Proc. IEEE SmartGridComm, 2020.

[11] S. Manandhar and Y. H. Lee, "GPS-PWV Based Improved Long-Term Rainfall Prediction Algorithm for Tropical Regions," Remote Sensing, MDPI, vol. 11, no. 19, 2019.

[12] Z. Wu and S. Jin, "Simulation of Marine Weather during an Extreme Rainfall Event: A Case Study," Hydrology, MDPI, vol. 6, no. 4, 2019.

[13] T. Lin and Q. Li, "Attention-Based Dual-Source Spatiotemporal Neural Network for Lightning Forecast," in Proc. IEEE BigData, 2019.

[14] S. Hudnurkar and A. Wanchoo, "ANN Model for Maximum Temperature Forecasting," International Journal of Applied Engineering Research, vol. 10, no. 9, pp. 23065–23070, 2015.

[15] M. A. Khan and S. Saqib, "Effective Demand Forecasting Model Using Business Intelligence and Machine Learning," in Proc. IEEE ICICT, 2020.

[16] N. S. Kapadia and U. Parikh, "Weather Forecasting Using Satellite Image Processing and Artificial Neural Networks," IJCSIS, vol. 14, no. 6, pp. 246–250, 2016.

[17] C. Feng and J. Zhang, "Hourly-Similarity Based Solar Forecasting Using Multi-Model Machine Learning Blending," IEEE Transactions on Sustainable Energy, vol. 9, no. 3, pp. 1077–1087, 2018.

[18] S. E. Haupt and J. Cowie, "Machine Learning for Applied Weather Prediction," IEEE Computer Society Conference, 2018.

[19] H. Nielsen and A. Iosifidis, "CloudCast: A Satellite-Based Dataset and Baseline for Forecasting Clouds," IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 9, pp. 7670–7683, 2021.

[20] "Numerical Weather Prediction," Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Numerical_weather_prediction

[21] C. S. Kanimozhi and G. Sowmiya, "Prediction of Extreme Weather Events using ML Techniques," International Journal of Applied Engineering Research, vol. 14, no. 11, 2019.

[22] M. G. Salman and B. Kanigoro, "Weather Forecasting using Deep Learning Techniques," in Proc. ICACSIS, Indonesia, 2015.

[23] X. Shao and S. Lu, "Solar Radiation Forecast with Machine Learning," IBM T. J. Watson Research Center, 2017.

[24] N. Singh and S. Chaturvedi, "Weather Forecasting Using Machine Learning Algorithm," in Proc. IEEE ICICC, 2019.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[25] F. Fallucchi and R. Scano, "ML Models Applied to Weather Series Analysis: A Comparative Study," ResearchGate, 2021.

[26] M. Wu and S. Jin, "High-Precision GNSS PWV and Its Variation in China," Remote Sensing, MDPI, vol. 13, no. 2, 2021.

[27] Simplilearn, "Steps of Machine Learning." [Online]. Available: https://www.simplilearn.com

[28] S. H. Bari and M. T. Rahman, "Forecasting Monthly Precipitation Using ARIMA Model: A Case Study on Dhaka, Bangladesh," Journal of Environment and Earth Science, IISTE, vol. 5, no. 2, 2015.

[29] M. G. Schultz et al., "Can Deep Learning Beat Numerical Weather Prediction?" Philosophical Transactions of the Royal Society A, vol. 379, no. 2194, 2020.

[30] R. K. Grace and B. Suganya, "Machine Learning Based Rainfall Prediction," in Proc. IEEE ICCCNT, 2020.

[31] Taillardat, M., Mestre, O., Zamo, M., & Naveau, P. (2016). Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. Monthly Weather Review, 144(6), 2375–2393.

[32] Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent sub-grid processes in climate models. Proceedings of the National Academy of Sciences, 115(39), 9684–9689.

[33] Chen, L., Zeng, Z., Chen, Y., & Xue, L. (2021). A machine learning approach for pressure prediction in atmospheric modeling. Atmospheric Research, 253, 105484.

[34] Hernández, M. A., & Arnáez, J. (2023). Evaluation of ML models for near-surface air temperature forecasting. Journal of Atmospheric and Solar-Terrestrial Physics, 239, 106946.

[35] Yuan, J., Zhan, Y., & Song, M. (2022). Improving temperature forecasts using random forest regression with meteorological features. Atmospheric Research, 278, 106368.

[36] Singh, R., & Yadav, S. (2021). Forecasting daily temperature using ensemble learning models: A case study. Environmental Modelling & Software, 141, 105052.

[37] K. R. Desai, R. H. Chile, and S. R. Sawant, "A model for detection of error factors in GPS signals," International Journal of Innovation in Engineering Research and Technology (IJIERT), vol. 2, no. 2, pp. 1–3, 2015.

[38] K. R. Desai and S. S. Kolekar, "Authentication and self-synchronization of audio signal by watermarking using empirical mode decomposition," International Journal of Engineering Sciences and Research Technology (IJESRT), pp. 1–4, Apr. 2015.

[39] K. R. Desai and A. Rhataval, "Secured decentralized confidential data distributed across a disruption tolerant military network, with the algorithm MOS and 3DLS," International Journal of Innovation in Engineering Research and Technology (IJIERT), vol. 4, no. 10, pp. 1–10, Oct. 2017.

[40] K. R. Desai and A. Rhataval, "Securely data retrieval for decentralized disruption tolerant military systems," International Journal for Science and Advance Research in Technology (IJSART), vol. 3, no. 11, pp. 1–11, Nov. 2017.

[41] K. R. Desai and S. Jagtap, "A study of effect of PCA on speech emotion recognition," International Research Journal of Engineering and Technology (IRJET), vol. 6, no. 6, pp. 1–6, Jun. 2019.

[42] K. R. Desai and S. S. Kesarkar, "Detection of faults in fabric using image processing in MATLAB," International Research Journal of Engineering and Technology (IRJET), vol. 10, no. 7, pp. 1–7, Jul. 2023.

Received: August 04, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[43] Y. Essa and H. G. P. Hunt, "Short-term Prediction of Lightning in Southern Africa using Autoregressive Machine Learning Techniques," IEEE Access, vol. 9, 2021.

[44] M. A. Rizvee et al., "Weather Forecasting for the North-Western region of Bangladesh: A Machine Learning Approach," in Proc. IEEE ICCIT, Dhaka, 2020.

[45] H. A. B. Y. Ahmed and S. W. A. Mohamed, "Rainfall Prediction using Multiple Linear Regression Model," IEEE, 2020.

Received: August 04, 2025