INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN SOCIAL SCIENCE (IJRISS)
ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

Learning Approaches for Weather Forecasting: A Comparative Study

Turgut Pura[1,*] , Duygu Cebeci[2], Elif Sıla Demireli[3] and Emine Baki[4]

[1]Department of Software Engineering, İstanbul Gedik Üniversitesi

[2,3,4]Department of Computer Engineering, İstanbul Gedik Üniversitesi

[*]Corresponding Author

DOI: https://doi.org/10.47772/IJRISS.2026.100600303

Received: 01 June 2026; Accepted: 06 June 2026; Published: 23 June 2026

ABSTRACT

This study investigates daily average air temperature forecasting using meteorological data collected in Istanbul, Türkiye, between January 2015 and December 2024. The dataset consists of 3,653 daily observations including temperature, precipitation, and wind-related variables. To improve forecasting performance, several preprocessing and feature engineering techniques were applied, including seasonal encoding, lagged temperature features, moving averages, and Min–Max normalization. The forecasting task was formulated as a multivariate time-series regression problem, and the temporal structure of the data was preserved through chronological train–test splitting.

Six prediction models representing statistical, machine learning, and deep learning approaches were evaluated: Linear Regression, Random Forest, Support Vector Regression (SVR), XGBoost, LightGBM, and Long Short-Term Memory (LSTM). Model performance was assessed using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). Experimental results showed that Linear Regression achieved the best overall performance with an RMSE of 0.0442 and an R^2 value of 0.9543. Random Forest produced comparable results with an RMSE of 0.0452 and an R^2 value of 0.9523. In contrast, SVR, XGBoost, LightGBM, and LSTM exhibited substantially lower predictive performance and negative R^2 values on the test dataset.

The findings indicate that daily temperature data in Istanbul exhibit strong temporal continuity and seasonality. Models capable of effectively utilizing lagged temperature information and seasonal patterns achieved the highest forecasting accuracy. The results also demonstrate that increasing model complexity does not necessarily improve predictive performance when working with relatively limited local meteorological datasets. Overall, the study highlights the effectiveness of simple and interpretable models for short-term temperature forecasting and provides a comparative evaluation of different machine learning approaches under real-world forecasting conditions.

Keywords: forecasting, machine learning, time series analysis, deep learning, regression

INTRODUCTİON

Weather forecasting involves analyzing and predicting atmospheric conditions for a specific location and time period using both empirical observations and analytical methods [1]. Variables such as temperature, precipitation, wind, and humidity directly influence many aspects of daily life, including agriculture, transportation, and disaster management. Accurate forecasts play a key role in protecting human life, supporting economic activities, and improving operational planning. In contrast, inaccurate predictions may lead to serious consequences, such as economic losses, ineffective disaster response, and poor decision-making in critical sectors. In recent years, the growing impact of climate change has increased the variability and uncertainty of weather conditions, making reliable short-term forecasting more important than ever.

Modern weather forecasting systems rely on large-scale computational models that combine data from ground stations, satellites, and radar observations [2]. Among these, Numerical Weather Prediction (NWP) models have been widely used to estimate future atmospheric states by modeling the physical relationships between current conditions and their evolution over time [1]. However, despite their effectiveness, these models have several limitations. Previous studies have highlighted issues such as high computational requirements, increasing data complexity, limited resolution at local scales, and sensitivity to initial conditions [4]. As a result, even small measurement errors can propagate over time and reduce forecasting accuracy.

Due to these challenges, there has been a growing interest in the use of machine learning and deep learning techniques for weather prediction. Unlike traditional physics-based models, these approaches can learn patterns directly from historical data and capture complex, nonlinear relationships among variables. In addition, the high dimensionality and temporal nature of meteorological data make machine learning methods a suitable alternative. Models such as Random Forest, Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Long Short-Term Memory (LSTM) networks have been widely applied in recent studies and have shown promising results in handling time-dependent weather data.

In this study, we develop a machine learning-based forecasting framework using daily meteorological data collected in Istanbul between 2015 and 2024. The main objective is to compare the performance of different models in terms of their predictive accuracy and error characteristics. To achieve this, the time-series nature of the data is explicitly considered, and all stages of the workflow, including preprocessing, feature engineering, and model evaluation, are handled in an integrated manner. The proposed system is also designed with a simple interface to provide practical daily temperature forecasts.

Overall, this study aims not only to report model performance but also to identify which approaches are more suitable for local-scale forecasting problems. Unlike many existing studies that rely on datasets from different regions, this work focuses on a single geographical area and examines how regional characteristics affect model behavior. By combining data-driven analysis with an application-oriented design, the study provides a practical contribution to the field of weather forecasting.

LİTERATURE REVİEW

In recent years, the use of machine learning techniques in weather forecasting has attracted increasing attention, resulting in a growing number of studies in this field. Existing research differs in terms of datasets, modeling approaches, and problem definitions.

Earlier work mainly focused on classical machine learning methods such as Linear Regression, Random Forest, and Support Vector Machines. These approaches generally performed well when applied to relatively simple datasets with a limited number of variables. However, weather data often exhibit strong temporal dependencies, seasonal patterns, and nonlinear relationships, which make long-term and multivariate forecasting more challenging for these traditional models.

To overcome these limitations, more recent studies have turned to deep learning-based approaches. Deep learning models are capable of learning complex patterns directly from data and are particularly suitable for time-series problems. Among these, Long Short-Term Memory (LSTM) networks have become one of the most widely used techniques, mainly because they can capture long-term dependencies and retain information from previous time steps. Compared to classical models, these architectures can better represent both static and temporal relationships within meteorological data.

Chen et al. [5] evaluated several advanced machine learning and deep learning models, including PanGu, GraphCast, FourCastNet, MetNet, DeepESD, and NNCAM, for weather and climate prediction. Their results showed that while these models perform well for short-term forecasting, their performance decreases in medium-

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

and long-term scenarios due to the complexity of climate systems. This finding highlights the importance of models that can effectively handle long-term dependencies.

Gupta et al. [6] compared SVM, Artificial Neural Networks, Random Forest, and Gradient Boosting methods for temperature and precipitation prediction. According to their findings, Gradient Boosting achieved the highest accuracy, mainly because of its ability to iteratively reduce errors and capture nonlinear relationships. In addition, its robustness against overfitting makes it suitable for complex datasets.

Zhang et al. [7] explored a wide range of machine learning and deep learning approaches, including CNN, RNN, Transformers, and graph-based neural networks. Their results indicated that CNN and Transformer-based models are effective for high-dimensional data, while Graph Neural Networks perform better in capturing spatial relationships. However, challenges such as computational cost and the prediction of rare weather events still remain.

Gangula et al. [8] focused on classification-based weather prediction and emphasized that the choice of algorithm plays a crucial role in achieving accurate results. Their study suggests that model selection should depend on both the characteristics of the dataset and the specific problem rather than relying on a single method.

Sevinç and Kaya [9] compared LSTM and ARIMA models using meteorological data from Diyarbakır. Their results showed that LSTM slightly outperformed ARIMA, especially in capturing temporal patterns. This finding supports the idea that deep learning models are more suitable for time-series forecasting tasks.

Nigam et al. [10] developed a weather classification system based on sensor data and compared multiple machine learning models. Their study highlights that not only model selection but also data collection strategies play a significant role in prediction performance.

Rahman et al. [11] evaluated several machine learning models for weather prediction in data-scarce regions. Interestingly, their results showed that a relatively simple model, Gaussian Naive Bayes, achieved the highest accuracy. This suggests that simpler models can still be effective when properly matched with the data.

Maheswari and Gomathi [12] applied Decision Tree and SVM models for long-term forecasting and found that machine learning methods generally outperform traditional numerical approaches. However, they also noted that long-term prediction remains a challenging task.

Das et al. [13] analyzed multiple regression-based and machine learning models using long-term meteorological data and reported that Random Forest achieved the best performance for short-term predictions. Their findings underline the importance of having sufficiently large datasets.

Finally, Hennayake et al. [14] applied an LSTM-based approach for short-term forecasting using multivariate data. The model achieved low error rates, demonstrating the effectiveness of deep learning in capturing complex meteorological patterns, particularly in regions with dynamic climate conditions.

DATASET AND DATA PREPROCESSİNG

Data source and dataset description

This study utilizes meteorological observations collected for Istanbul over a ten-year period from January 1, 2015, to December 31, 2024. The dataset consists of a total of 3,653 daily records. Each observation includes multiple variables such as temperature measurements (2m_mean, 2m_max, 2m_min), precipitation components (precipitation_sum, rain_sum, snowfall_sum), and wind dynamics (windspeed, windgusts). The continuity of the dataset provides a solid foundation for enabling the models to learn seasonal cycles and recurring weather patterns effectively.

Data cleaning and structuring

This study utilizes meteorological observations collected for Istanbul over a ten-year period from January 1, 2015, to December 31, 2024. The dataset consists of a total of 3,653 daily records. Each observation includes multiple variables such as temperature measurements (2m_mean, 2m_max, 2m_min), precipitation components (precipitation_sum, rain_sum, snowfall_sum), and wind dynamics (windspeed, windgusts). The continuity of the dataset provides a solid foundation for enabling the models to learn seasonal cycles and recurring weather patterns effectively.

Data transformation

To prepare the data for the modeling process, data type consistency was first verified. Time-related information (timestamps) was converted into datetime objects, allowing appropriate indexing for time-series analysis. Statistical inspections confirmed the absence of missing values (NaN) within the dataset; therefore, no imputation techniques were required, and the original structure of the raw data was preserved.

Dimensional transformation

To enhance the predictive capability of the models, additional features were derived from the raw dataset. The feature engineering process was conducted with a focus on three main components.

- First point Seasonal Decomposition: A seasonal variable was extracted from the date information and transformed into numerical vectors using the one-hot encoding method. This approach enabled the models to mathematically capture seasonal transitions.
- Lagged Features: To capture temporal correlations within the time series, lagged values of the target variable from the previous day (T−1) as well as 3-day and 7-day moving averages were introduced as additional input features.
- Normalization: Since the variables possess different measurement scales, all features were normalized to the range [0, 1] using the Min–Max scaling technique. This step prevented scale-induced bias during the model training process and ensured stable convergence

METHODOLOGY AND TECHNİCAL PREPARATİON

Data source and data analysis

The study is formulated as a multivariate time series regression problem aimed at predicting daily average air temperature based on historical meteorological data. To preserve the temporal order of the dataset, the data were split without random shuffling. Accordingly, the first 70% of the observations (2,552 samples) were allocated for model training, while the remaining 30% (1,094 samples) were reserved for evaluating model performance.

Exploratory Data Analysis (EDA) and visualization

Prior to model development, the distributions of variables and their interrelationships were examined using the Seaborn and Matplotlib libraries. Seasonal fluctuations in temperature data were visualized, and correlation coefficients between independent variables (such as precipitation and wind-related features) and the target variable (average temperature) were computed. Figure 1 presents the correlation matrix heatmap illustrating these relationships. This exploratory analysis provided valuable insights into the underlying data structure and guided subsequent feature selection and model configuration.

Algorithmic configuration and training

In this study, six different models based on statistical, machine learning, and deep learning approaches were selected to comprehensively analyze the weather data of Istanbul.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

Linear regression

Linear Regression was employed as a baseline model in weather forecasting to establish a fundamental reference. It models the relationship between independent variables and the target variable using a linear equation. This model plays a critical role in identifying general trends within the data and in benchmarking the performance improvements achieved by more complex models.

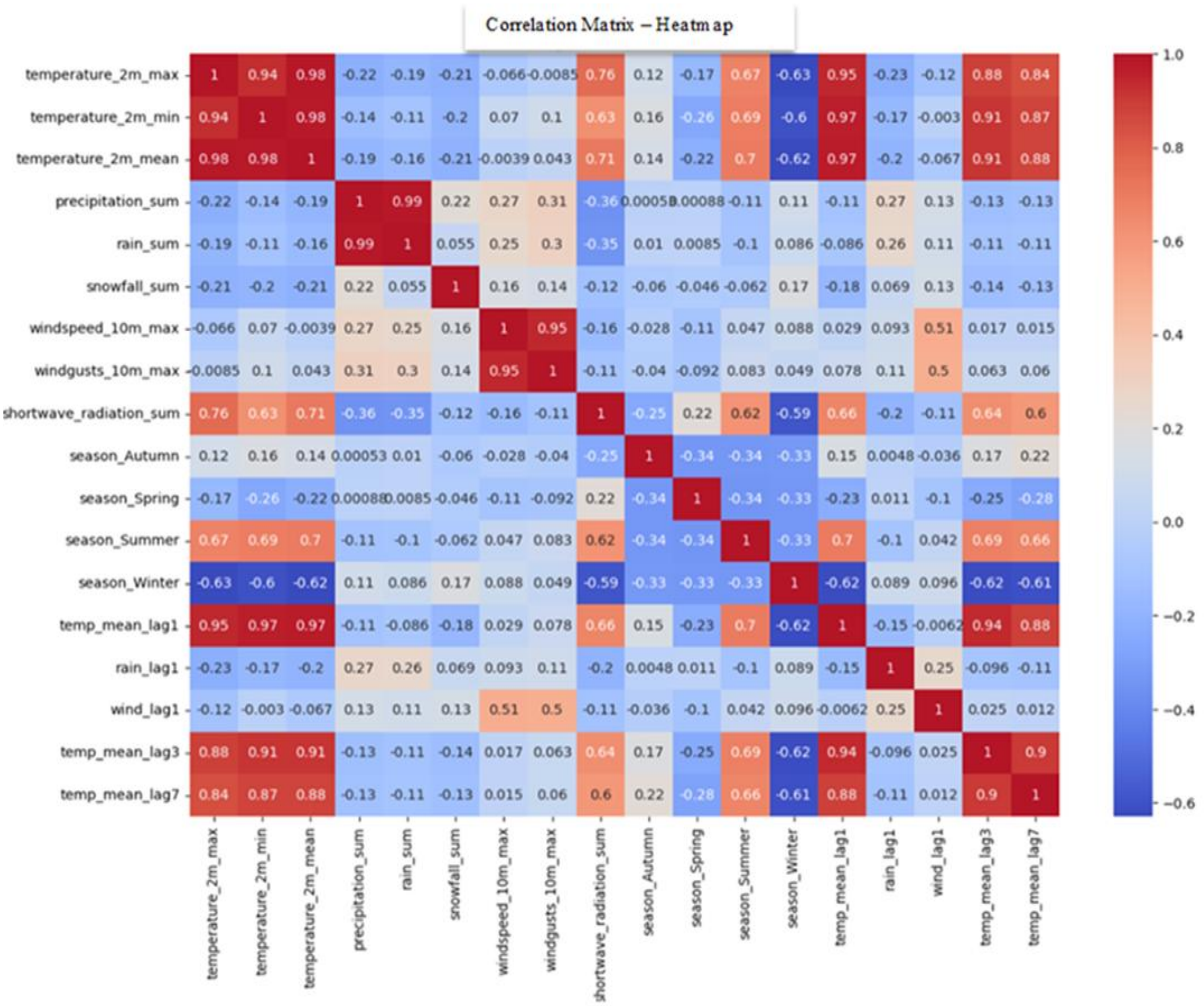

Figure 1. Correlation matrix heatmap of meteorological variables

Support Vector Regression (SVR)

Support Vector Regression (SVR), the regression adaptation of Support Vector Machines, maps the data into a high-dimensional feature space to determine the optimal decision boundary. Owing to its kernel functions, SVR is effective in capturing nonlinear relationships within atmospheric data and in minimizing the influence of outliers.

Random Forest (RF)

Random Forest is an ensemble learning method composed of multiple decision trees. It generates predictions by training numerous decision trees on different data subsets and aggregating their outputs through averaging. This structure reduces the overfitting problem commonly observed in individual decision trees while demonstrating strong robustness in modeling complex interactions among meteorological parameters.

XGBoost (Extreme Gradient Boosting)

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

XGBoost is a high-performance implementation of the gradient boosting algorithm that iteratively optimizes prediction errors. The regularization techniques integrated into the algorithm enhance its generalization capability. Due to its ability to handle missing values efficiently and its high computational speed, XGBoost is well suited for weather forecasting tasks.

LightGBM (LGBM)

Developed by Microsoft, LightGBM adopts a leaf-wise tree growth strategy rather than the traditional level-wise approach. This enables the construction of deeper trees and often results in lower error rates. LightGBM is particularly efficient for large-scale meteorological datasets due to its low memory consumption and high processing speed.

Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are an advanced type of recurrent neural network within deep learning architectures. Unlike classical algorithms, LSTM models can retain long-term temporal dependencies and seasonal patterns through their internal gate mechanisms. This capability allows LSTM to offer enhanced predictive performance for time-series problems, such as weather forecasting, where past observations directly influence future conditions.

All models were optimized using the TensorFlow and Keras libraries in a Google Colab environment with T4 GPU support.

Evaluation metrics

To measure model performance and provide a comparative analysis, evaluation metrics commonly used in the literature were selected.

MAE (Mean Absolute Error): MAE represents the arithmetic mean of the absolute differences between predicted values and actual observations. It measures the average magnitude of prediction errors regardless of their direction. Since MAE assigns equal weight to all errors, it is more robust to outliers compared to MSE. The Mean Absolute Error (MAE) is calculated as given in Equation (1).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|yi - \hat{y}i|$$

MSE (Mean Squared Error): MSE computes the average of the squared differences between predicted values and actual observations. By squaring the errors, larger deviations are penalized more severely than smaller ones, making MSE a useful measure for assessing a model's sensitivity to large prediction errors. This property allows MSE to serve as an indicator of model stability against substantial deviations. However, the resulting value is expressed in squared units of the target variable, which may limit direct interpretability. The Mean Squared Error (MSE) is computed as defined in Equation (2).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2$$

RMSE (Root Mean Squared Error): RMSE is calculated by taking the square root of the Mean Squared Error. While preserving MSE's advantage of penalizing larger errors more heavily, RMSE converts the error magnitude back to the original unit of the target variable. This property facilitates a more intuitive and physically meaningful interpretation of the model's prediction error. RMSE is calculated as defined in Equation (3), providing an error measure expressed in the same unit as the target variable.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2}$$

R^2 (Coefficient of Determination): The R^2 score is a statistical measure that indicates the proportion of variance in the dependent variable (temperature) that is explained by the independent variables in the model. It takes values between 0 and 1, where values closer to 1 indicate that the model effectively captures the underlying patterns in the data, while values near 0 suggest that the model performs no better than a simple mean-based prediction. The coefficient of determination (R^2) is calculated as given in Equation (4).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - \hat{y}i)^2}{\sum_{i=1}^{n}(yi - \hat{y})^2}$$

Experimental results

In this study, the problem of predicting daily average temperature was addressed using meteorological observations collected for Istanbul over the period from 2015 to 2024. Due to the inherent temporal dependency of the dataset, the time series nature of the data was explicitly considered during the modeling process. Consequently, instead of adopting a conventional random train–test split, a chronological partitioning strategy was employed. This approach enables the models to learn exclusively from historical observations and generate forecasts for future periods, thereby providing a more realistic and reliable evaluation of predictive performance under real-world conditions.

Data source and raw data analysis

During the modeling stage, daily average temperature values were selected as the target variable. The input features were constructed exclusively from information that would be available at the time of prediction, thereby preventing the model from indirectly accessing future information. Within this framework, the following features were included in the model:

- Lagged values of average temperature at 1, 3, and 7 days
- Precipitation amount and wind speed from the previous day
- Seasonal variables encoded using the one-hot encoding method to represent seasonal effects

This feature set effectively eliminated the risk of data leakage by ensuring that the models relied solely on historical information. All numerical variables were normalized using the Min–Max scaling technique, thereby minimizing the adverse effects of differing feature scales on model training.

The dataset was divided into training and test sets with an approximate ratio of 70%–30%. The test set was selected from the final portion of the dataset and intentionally included more volatile and challenging periods to forecast. This strategy enabled a more robust assessment of the models' generalization performance under demanding conditions.

Applied models

Both classical machine learning algorithms and deep learning-based approaches were jointly evaluated in the experimental study. The models employed are listed below:

- Linear Regression
- Random Forest
- Support Vector Regression (SVR)
- XGBoost

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

- LightGBM
- Long Short-Term Memory (LSTM)

For each model, hyperparameter optimization was conducted, and the best-performing configurations were used to obtain test results. This procedure ensured that each model's predictive potential was fairly and effectively represented.

Performance metrics

To comprehensively evaluate model performance in regression-based forecasting tasks, the following performance metrics were employed:

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

These metrics enabled a joint analysis of both the magnitude of prediction errors and the extent to which the models explained the variance in the target variable.

Model performance comparison

The results obtained on the test dataset are presented in Table 1.

Table 1. Test dataset performance of all models.

Model	RMSE	MSE	MAE	R^2
Linear Regression	0.0442	0.00196	0.0323	0.9543
Random Forest	0.0452	0.00204	0.0331	0.9523
SVR	0.2840	0.08068	0.2305	-0.8857
XGBoost	0.2807	0.07879	0.2277	-0.8415
LightGBM	0.2801	0.07846	0.2272	-0.8337
LSTM	0.2636	0.06947	0.2144	-0.6236

As shown in Table 1, the Linear Regression model achieved the lowest values across all error metrics while simultaneously producing the highest R^2 score. This finding indicates that past temperature values and seasonal components exhibit a strong and predominantly linear structure in daily average temperature forecasting.

Although the Random Forest model demonstrated performance comparable to that of Linear Regression, it did not provide a significant improvement in terms of error metrics. Despite the inherent capability of tree-based models to capture nonlinear relationships, the largely linear nature of the relationships within this dataset limited the potential advantages of Random Forest.

The negative R^2 values observed for the SVR, XGBoost, and LightGBM models represent a statistically critical outcome. A negative R^2 score indicates that the predictions generated by these models result in a higher squared error than a simple mean-based baseline. This behavior suggests that these complex models tended to learn noise present in the training data, leading to overfitting and a loss of generalization capability when exposed to even minor distributional shifts in the test dataset.

Although SVR, XGBoost, and LightGBM produced low error values during the training phase, they experienced a pronounced performance degradation during testing. This decline indicates that the complex structures of these models were insufficiently adaptive to the volatile temperature fluctuations observed in the test period.

Despite being trained with optimized hyperparameters, the LSTM model exhibited weak performance on the test dataset. The negative R^2 value implies that the model performed worse than a simple average-based prediction during the testing phase. The primary reason for this outcome is the limited size of the dataset, which was insufficient for effectively learning long-term sequential dependencies required by deep learning architectures such as LSTM.

RESULTS AND DİSCUSSİON

In The experimental results demonstrate that, for the problem addressed in this study, increasing model complexity does not necessarily lead to improved performance; on the contrary, performance deteriorates markedly in certain cases. In particular, the superior performance of the Linear Regression model clearly indicates that the temperature forecasting problem exhibits strong temporal continuity and pronounced seasonality.

The primary reason for the discrepancy between validation scores obtained during training and the final comparative results on the test set lies in the use of a chronologically separated evaluation strategy, namely out-of-time validation. While complex models achieved high performance under randomly split validation schemes, they struggled to adapt to seasonal shifts and abrupt temperature fluctuations encountered during the test period. This difficulty resulted in increased error metrics. These findings highlight the critical importance of learning not only random patterns present in the data but also the structural changes introduced by temporal dynamics.

This study provides a concrete illustration of the well-known bias–variance tradeoff frequently emphasized in the literature. High-capacity models, such as LSTM and gradient boosting-based approaches, faced challenges in effectively optimizing their parameters given the relatively limited dataset size of approximately 3,650 observations. In the absence of the large-scale data typically required by deep learning models, simpler and more rigid models proved to be less sensitive to noise and consequently produced more stable and reliable predictions.

The performance of the models was further evaluated by comparing the actual and predicted values on the test dataset. As shown in Figure 2, the relationship between actual values and model predictions varies significantly across the six models. In contrast to the dispersed patterns and lower predictive performance observed for the LightGBM, LSTM, SVR, and XGBoost models, the Linear Regression (R^2 = 0.954) and Random Forest (R^2 = 0.952) models demonstrate clearly superior performance. For these two models, the data points are tightly clustered around the ideal 45-degree reference line (y = x), visually confirming their high explanatory power.

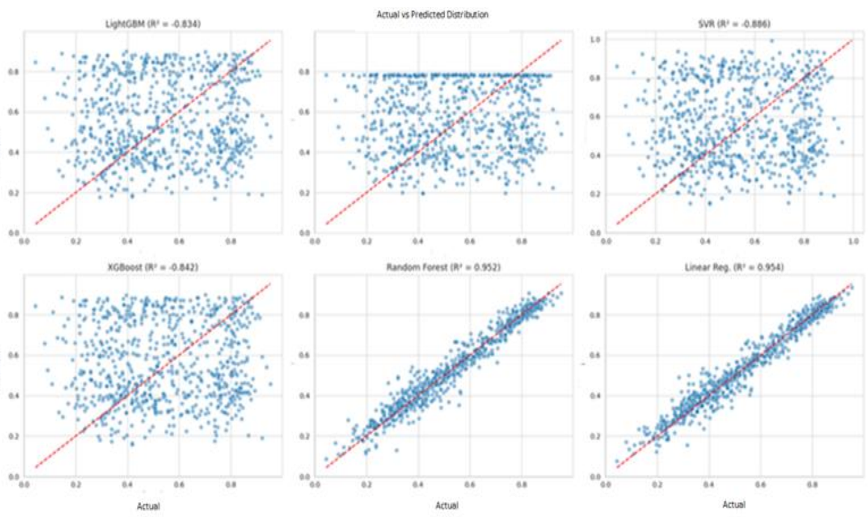

Figure 2. Scatter plot of actual and predicted values

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

Furthermore, this distribution indicates that these models maintain consistent predictive performance across both low and high value ranges, while exhibiting relatively stable variance. The narrow dispersion around the reference line suggests low heteroscedasticity, reinforcing the robustness and reliability of the Linear Regression and Random Forest models.

In the Figure 3, the actual values are represented by a solid black line. The predictions generated by the models are illustrated as follows: LightGBM with a blue dashed line, LSTM with a green dotted line, SVR with an orange dash–dot line, XGBoost with a red dashed line, Random Forest with a solid purple line, and Linear Regression with a brown dotted line.

As shown in Figure 3, the actual and normalized predicted values corresponding to the last 100 days of the test dataset were analyzed along the time axis to evaluate the temporal alignment between observations and model outputs. The results reveal that the best-performing models not only capture the overall seasonal trend but also accurately follow abrupt increases (peaks) and sharp decreases in temperature values. This behavior indicates a high level of sensitivity to short-term dynamics, which is critical for reliable daily weather forecasting.

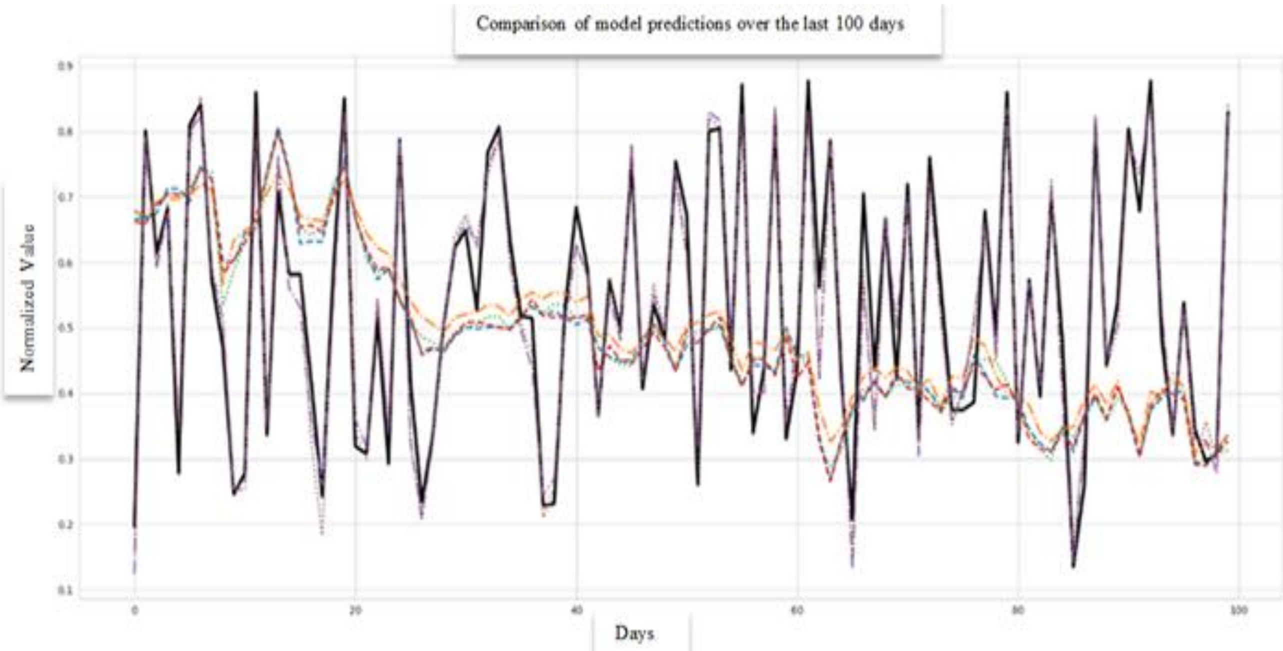

Figure 3. Time series forecast plot for the last 100 days

In particular, the Linear Regression and Random Forest models demonstrate a close synchronization with the actual values, successfully tracking rapid temperature fluctuations without introducing significant temporal delays. For these models, the phase shift phenomenon—commonly observed in time series forecasting, where predictions lag behind real observations—is largely mitigated. The strong alignment between predicted and observed series confirms that these models effectively learn temporal dependencies while maintaining responsiveness to sudden changes.

In contrast, the remaining models exhibit smoother prediction trajectories that tend to dampen sharp variations, resulting in noticeable deviations during periods of high volatility. Although these models approximate the general trend, their limited ability to respond to abrupt changes leads to reduced accuracy over short horizons. Overall, the time series analysis presented in Figure 3 reinforces the quantitative findings reported in Table 1 and highlights that models with simpler structures and strong temporal continuity assumptions provide more robust and reliable performance in short-term temperature forecasting tasks.

As shown in Figure 4, the distribution of prediction errors across models was examined using boxplot analysis. The error distributions of the LightGBM, LSTM, SVR, and XGBoost models exhibit substantially higher variance, with residuals spread over a wide range, approximately from −0.8 to +0.8. This wide dispersion indicates unstable predictive behavior and a higher sensitivity to fluctuations in the data, particularly during periods of abrupt temperature changes.

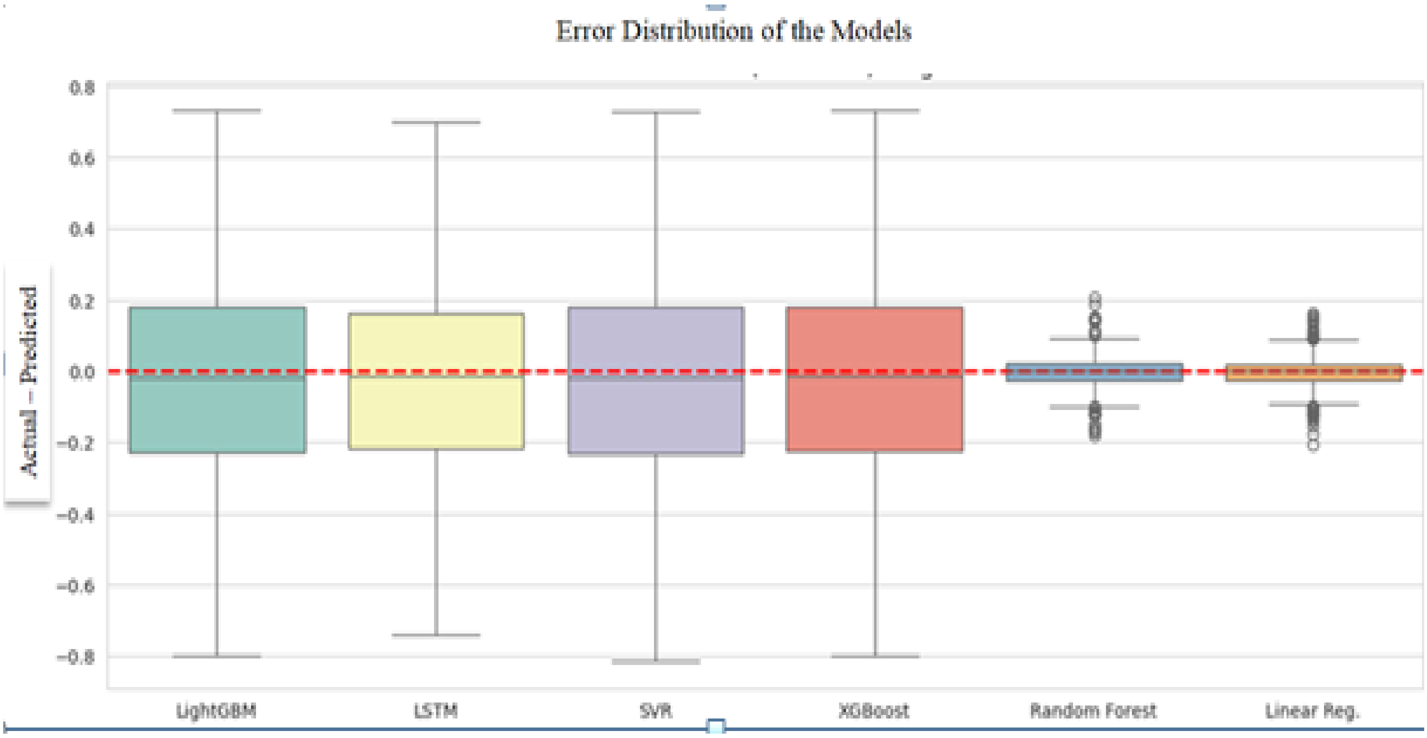

Figure 4. Residual error distribution plot

In contrast, the Random Forest and Linear Regression models display error distributions that are tightly concentrated around the zero line. For these models, the median residual values are effectively centered at zero, and the interquartile ranges (IQRs) are notably narrow. This pattern provides strong evidence of low bias and well-controlled variance, indicating that these models generate consistent and reliable predictions across the test dataset.

Overall, the residual error analysis confirms that Random Forest and Linear Regression achieve superior stability compared to more complex models. The minimized spread of residuals and the absence of systematic bias further support the robustness of these models, reinforcing the findings obtained from the performance metrics and time series evaluations presented earlier.

All tables should be numbered with Arabic numerals. Every table should have a caption. Headings should be placed above tables, left justified. Only horizontal lines should be used within a table, to distinguish the column headings from the body of the table, and immediately above and below the table. Tables must be embedded into the text and not supplied separately. Below is an example which the authors may find useful.

As illustrated in Figure 5, the predictive performance of the six trained machine learning models was evaluated using the Root Mean Squared Error (RMSE) metric. Since lower RMSE values indicate higher forecasting accuracy, this comparison provides a clear assessment of model effectiveness. The Linear Regression model achieved the best performance with an RMSE value of 0.0442, indicating highly accurate predictions with minimal error. Closely following this result, the Random Forest model yielded an RMSE of 0.0452, demonstrating comparable predictive capability.

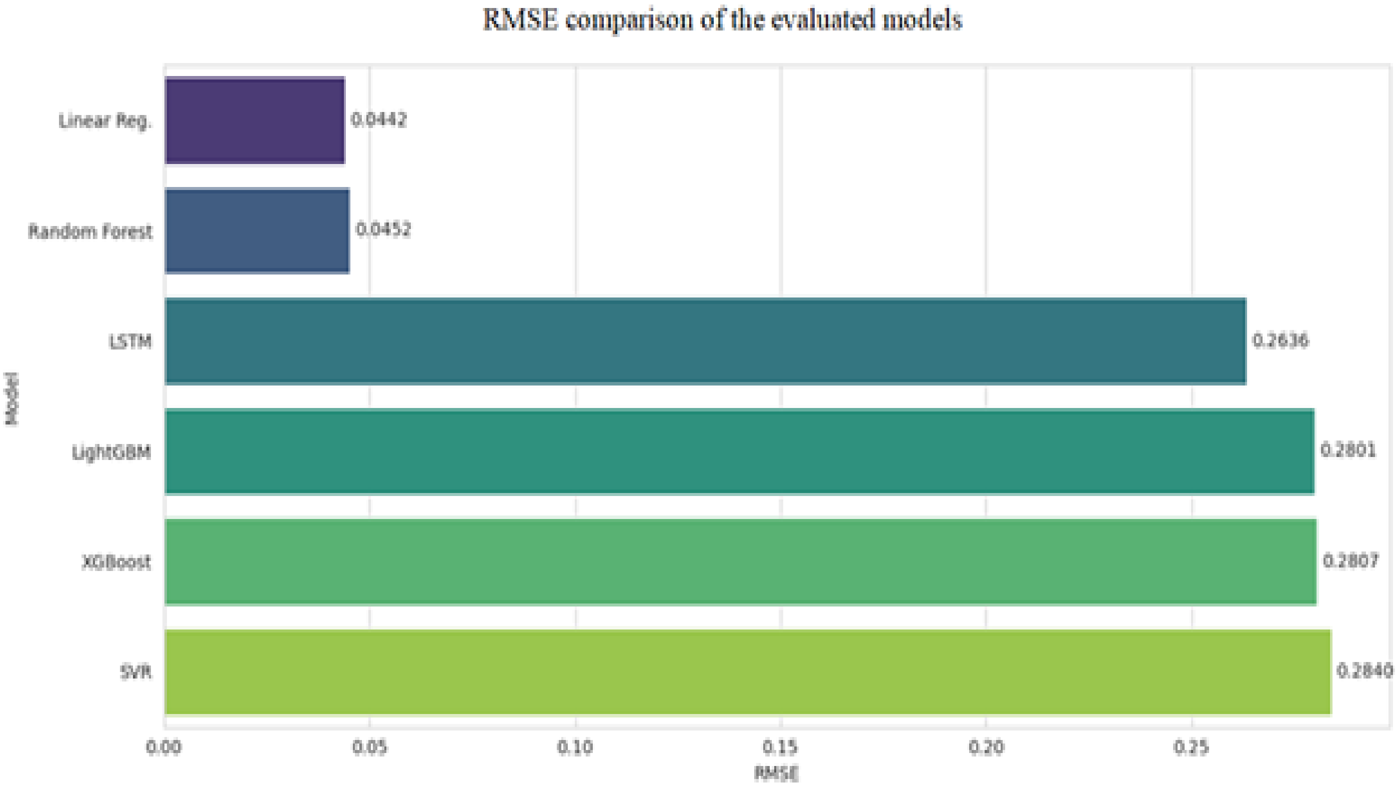

Figure 5. Comparison of models based on their RMSE values

In contrast, the remaining models—SVR, XGBoost, LightGBM, and LSTM—exhibited substantially higher RMSE values, ranging approximately between 0.26 and 0.28. These elevated error levels suggest that more complex model architectures were less effective in capturing short-term temperature variations under the given data conditions. The pronounced performance gap observed in Figure 5 reinforces the conclusion that, for this dataset, simpler models with strong temporal continuity assumptions outperform more sophisticated learning algorithms in daily average temperature forecasting.

Detailed analysis of the best-performing model

The superior performance of the Linear Regression model can be attributed to three primary factors. First, the lagged temperature variables exhibit strong explanatory power, effectively capturing the high temporal dependency present in daily temperature data. Second, the seasonal variables contribute directly and significantly to the model by representing recurring seasonal patterns in a structured manner. Third, the relatively simple structure of the model results in a low risk of overfitting, thereby enabling stable and reliable generalization to unseen data.

Coefficient analysis indicates that the highest weight is assigned to the temperature value of the previous day (lag-1). This finding is consistent with the Persistence Method widely used in meteorology, which is based on the assumption that "tomorrow's weather is likely to be similar to today's." For the examined region and time period, this assumption serves as a strong predictive indicator. By directly and transparently reflecting the strong autocorrelation inherent in temperature time series, the Linear Regression model achieves the lowest RMSE value among all evaluated models.

Comparison with the Literature

Previous studies in the literature have shown that linear and tree-based methods are frequently preferred for short-term temperature forecasting tasks. The approximate R^2 value of 0.95 obtained in this study is consistent with reported findings in related research and supports the effectiveness of the adopted feature engineering strategy. These results further suggest that when temporal dependencies and seasonal structures are appropriately modeled, relatively simple algorithms can outperform more complex approaches in local-scale weather forecasting applications.

CONCLUSİONS

In this study, the problem of daily average temperature forecasting was examined in depth using historical meteorological data, and a comprehensive comparison of multiple machine learning approaches was conducted. The experimental findings demonstrate that the Linear Regression and Random Forest models provide the most balanced and reliable performance for this task. These results indicate that temperature time series exhibit strong temporal continuity and seasonality, and that models capable of directly leveraging these characteristics can achieve high predictive accuracy.

The superior performance of Linear Regression and Random Forest further suggests that increased model complexity does not necessarily lead to improved forecasting outcomes, particularly when the available data volume is limited. While advanced machine learning and deep learning models possess strong representational capacity, their effectiveness may be constrained by data sparsity and the risk of overfitting. In contrast, simpler models with lower variance appear to generalize more robustly under real-world forecasting conditions.

Despite the promising results, this study has several limitations. First, the dataset is restricted to a single geographical region, which may limit the generalizability of the findings to other climatic zones. Second, the forecasting framework is based on single-step predictions, focusing solely on short-term temperature estimation. Additionally, the relatively limited data volume poses challenges for training deep learning architectures effectively.

Future research directions include extending the analysis to datasets covering longer time horizons and multiple geographical regions to improve model generalization. Exploring multi-step forecasting strategies would enable the assessment of longer-term predictive performance. Incorporating additional meteorological variables, such as humidity and atmospheric pressure, may further enhance model accuracy by capturing more complex atmospheric interactions. Moreover, adopting dynamic training strategies based on rolling window approaches and periodic retraining mechanisms would allow the models to adapt to concept drift over time. Such strategies are expected to reduce prediction errors, particularly during seasonal transitions, and to improve the long-term robustness and adaptability of weather forecasting systems.

Data Availability Statement

The data used in this study are publicly available and can be obtained from the mgm database.

ACKNOWLEDGEMENTS

The authors declare that there are no acknowledgements for this study.

Ethics Approval

This study does not involve human participants, animals, or any personal data. The research was conducted using publicly available meteorological data. Therefore, ethics committee approval was not required.

REFERENCES

1. Meteoroloji Genel Müdürlüğü, “Hava tahmini hakkında sıkça sorulan sorular,” Ankara, Türkiye, 2025. [Online]. Available: https://www.mgm.gov.tr (accessed: Dec. 2025).
2. Climavision, “The ultimate guide to weather forecast models,” 2025. [Online]. Available: https://www.climavision.com (accessed: Dec. 2025).

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

3. Meteoroloji Genel Müdürlüğü, "Meteoroloji nedir?," Ankara, Türkiye, 2025. [Online]. Available: https://www.mgm.gov.tr (accessed: Dec. 2025).
4. J. A. Brotzge et al., "The evolution of numerical weather prediction and machine learning," Bull. Amer. Meteorol. Soc., pp. E698–E705, 2023.
5. L. Chen et al., "Machine learning methods in weather and climate applications," Int. J. Appl. Sci., 2023.
6. R. Gupta et al., "Comparative analysis of machine learning algorithms for temperature and rainfall prediction," Environ. Model. Softw., vol. xx, pp. 105123, 2023.
7. H. Zhang et al., "Machine learning methods for weather forecasting," Atmosphere, vol. 82, 2025.
8. P. R. Gangula et al., "A comparative study of machine learning techniques for nuanced weather prediction," in Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI), 2023.
9. A. Sevinç and B. Kaya, "Derin öğrenme yöntemleri ile sıcaklık tahmini: Diyarbakır ili örneği," J. Comput. Sci., pp. 217–225, 2021.
10. S. Nigam et al., "Comparative study to determine accuracy for weather prediction using machine learning," in Proc. Int. Conf. Comput. Commun. Inform. (ICCCI), 2023.
11. M. S. Rahman et al., "Comparative evaluation of weather forecasting using machine learning models," in Proc. Int. Conf. Comput. Inf. Technol. (ICCIT), 2023.
12. K. B. Maheswari and S. Gomathi, "A comprehensive analysis of weather prediction using machine learning," in Proc. Int. Conf. Sci. Technol. Eng. Math. (ICONSTEM), 2024.
13. P. Das et al., "An intelligent regression approach for weather forecasting system using machine learning," in Proc. Int. Conf. Cogn. Green Ubiquitous Comput. (IC-CGU), 2024.
14. K. M. S. A. Hennayake et al., "Machine learning based weather prediction model for short term weather prediction in Sri Lanka," in Proc. Int. Conf. Inf. Autom. Sustain. (ICIAfS), 2021.