*Journal of Intelligent System and Telecommunications 2(*1) (2025) 54-63

Journal of Intelligent System and
Telecommunications

Journal homepage: https://journal.unesa.ac.id/index.php/jistel/index

Advanced Machine Learning–Based Temperature Forecasting for the Jombang Region Using ERA5 Reanalysis Data (2020–2025)

Zainal Arifin[1*]*,Chamdan Mashuri*[2]*, Ginanjar Setyo Permadi*[3]

[1]Jurusan Teknik Elektro, Fakultas Teknik,Universitas Negeri Surabaya, Indoneisa
[2]Jurusan Manajemen Informatika, Fakultas Teknologi Informasi, Universitas Hasyim Asy'ari Tebuireng Jombang, Indonesia
[2]Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Universitas Hasyim Asy'ari Tebuireng Jombang, Indonesia
*Correspondence: E-mail: 25051505014@mhs.unesa.ac.id

ARTICLE INFO

Article History:
Submitted/Received 22 December 2025
First Revised 22 April 2026
Accepted 28 April 2026
First Available Online 25 May 2026
Publication Date 25 May 2026

Keyword:
ERA5 Reanalysis, Temperatur Forecasting, Machine Learning Model, LSTM Neural Network, XGBoost Regression, Random Forest

ABSTRACT

Accurate temperature forecasting plays a crucial role in supporting climate-sensitive sectors such as agriculture, environmental management, and public health, particularly in tropical regions with complex atmospheric dynamics. This study presents a machine learning-based framework for short-term air temperature forecasting in the Jombang region, Indonesia, utilizing ERA5 reanalysis data from 2020 to 2025. The dataset was preprocessed through temporal alignment, anomaly handling, and lag-based feature engineering to effectively capture diurnal temperature variations.

Three machine learning models—Long Short-Term Memory (LSTM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost)—were developed and evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The experimental results demonstrate that XGBoost achieved the best performance, with an MAE of 0.5982 °C, RMSE of 0.8279 °C, and MAPE of 2.2320%, outperforming both LSTM and Random Forest.

These findings suggest that boosting-based ensemble learning is more effective in modeling nonlinear temperature patterns compared to recurrent and bagging-based approaches. This study contributes to the advancement of localized temperature forecasting models in tropical regions and provides a practical reference for selecting suitable machine learning methods for regional climate applications.

1. INTRODUCTION

Air temperature is one of the most essential meteorological variables, as it directly affects agriculture, water resources, energy demand, ecosystem stability, and human thermal comfort. Accurate temperature forecasting is therefore crucial for planning climate-sensitive activities and minimizing the impacts of weather variability. Price et al. [1] demonstrated that machine learning-based forecasting systems are increasingly capable of delivering accurate probabilistic weather predictions, highlighting the expanding role of data-driven methods in operational meteorology.

Traditional forecasting approaches, including statistical regression techniques and physics-based numerical weather prediction models, have long been employed to estimate atmospheric variables. Nevertheless, these methods often face challenges in representing nonlinear interactions and rapidly changing local weather conditions. Yang et al. [2] explained that modern machine learning techniques are able to learn complex relationships among meteorological predictors and can improve prediction performance when conventional models are constrained by uncertainty or simplified assumptions.

Recent developments in artificial intelligence have introduced a variety of algorithms for environmental time-series forecasting. Among these, Long Short-Term Memory (LSTM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) have attracted considerable attention. Hochreiter and Schmidhuber [16] introduced LSTM as a recurrent neural network architecture specifically designed to retain long-term temporal dependencies, making it highly suitable for sequential forecasting tasks. Breiman [17] proposed Random Forest as an ensemble of decision trees that enhances predictive stability through bootstrap aggregation and randomized feature selection. Chen and Guestrin [18] developed XGBoost as an efficient gradient boosting framework with regularization capabilities, enabling high predictive accuracy and strong computational efficiency.

Several recent studies have confirmed the effectiveness of these models for atmospheric applications. Fister et al. [9] reported that machine learning models can achieve accurate long-term air temperature forecasting when supported by appropriate feature reduction strategies. Li et al. [12] further showed that deep learning approaches are effective for short-term outdoor air temperature prediction, particularly when temporal dependencies are dominant. These findings suggest that both deep learning and ensemble learning methods are strong candidates for temperature forecasting tasks.

Despite these global advancements, localized forecasting studies in tropical regions remain relatively limited. Tropical climates are characterized by intense solar radiation, high humidity, strong convective activity, and pronounced diurnal cycles, all of which produce highly dynamic temperature behavior. Haji-Aghajany et al. [5] emphasized that weather forecasting in such environments remains challenging because atmospheric processes can change rapidly over short temporal and spatial scales. Consequently, models developed in temperate regions may not always generalize effectively to tropical microclimates.

Jombang Regency, located in East Java, Indonesia, represents an important case study because its economy depends heavily on agriculture while also experiencing urban development and seasonal climate variability. Accurate local temperature forecasting can support planting schedules, irrigation management, public health preparedness, and regional planning. However, no previous study has comprehensively compared multiple machine learning algorithms for short-term temperature forecasting in Jombang using long-term reanalysis data.

The availability of ERA5 reanalysis data offers an opportunity to address this research gap. Hersbach et al. [21] described ERA5 as a high-quality global atmospheric reanalysis product with consistent temporal coverage and reliable meteorological variables, making it highly suitable for machine learning-based forecasting studies. By integrating ERA5 data with advanced predictive algorithms, localized temperature models can be developed even in regions with limited ground observation networks.

Therefore, this study aims to develop and compare LSTM, Random Forest, and XGBoost models for short-term air temperature forecasting in the Jombang region using ERA5 reanalysis data from 2020 to 2025. Model performance is evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The main contribution of this research is the identification of the most effective machine learning approach for localized tropical temperature forecasting, which may support future climate services and environmental decision-making in regional areas.

2. METHODS

This study employed a structured experimental framework consisting of data acquisition, preprocessing, feature engineering, model development, and performance evaluation. The proposed methodology was designed to ensure reproducibility and enable a fair comparison among the selected machine learning models.

2.1. Study Area and Data Source

The study focused on Jombang Regency, East Java, Indonesia, which is characterized by a tropical climate with distinct diurnal temperature variations and seasonal rainfall patterns. Air temperature data were obtained from the ERA5 reanalysis dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) through the Copernicus Climate Data Store. Hersbach et al. [21] described ERA5 as a fifth-generation global atmospheric reanalysis that combines numerical weather models with observational data, thereby providing temporally consistent and high-resolution meteorological information.

The primary variable used in this study was 2-meter air temperature, which represents near-surface atmospheric temperature. Data were collected from January 2020 to March 2025 at a two-hour temporal resolution. Grid cells corresponding to the geographical boundaries of Jombang Regency were extracted and aggregated to represent regional temperature conditions.

2.2. Data Preprocessing

Raw ERA5 data were processed through several stages prior to model training. First, all temperature values originally expressed in Kelvin were converted to degrees Celsius. Missing records and abnormal values were then examined, and linear interpolation was applied where necessary to maintain temporal continuity. Outliers resulting from data irregularities were filtered using statistical thresholding based on the interquartile range.

After the cleaning stage, the time series data were chronologically ordered and resampled into a uniform two-hour interval. Because machine learning models are sensitive to feature scale, numerical predictors were normalized using Min-Max scaling to a range between 0 and 1. This normalization process improves convergence stability, particularly for neural network models such as LSTM [19].

2.3. Feature Engineering

To capture temporal dependencies and periodic patterns, lag-based input features were generated from historical temperature observations. Previous temperature values at lag-1, lag-3, lag-6, and lag-12 time steps were used as predictors. These lag variables represent short-term memory effects as well as daily cyclical behavior.

Additional calendar-based predictors were also incorporated, including hour of day, day of month, month, and day of year. These variables enable the models to learn diurnal and seasonal temperature patterns. Yang et al. [2] noted that engineered temporal features can substantially improve machine learning performance in meteorological forecasting tasks.

The target variable for prediction was the temperature value at the next time step, corresponding to a two-hour-ahead forecasting horizon.

2.4. Machine Learning Models

Three machine learning models were developed and evaluated in this study: Long Short-Term Memory (LSTM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost).

2.4.1. Long Short-Term Memory (LSTM)

LSTM is a recurrent neural network architecture specifically designed to model sequential data while preserving long-term temporal information. Hochreiter and Schmidhuber [16] introduced memory cells and gating mechanisms that enable relevant information to be retained or forgotten during the training process. This capability makes LSTM well-suited for temperature forecasting, where current values are strongly influenced by previous observations.

The LSTM model implemented in this study consisted of two hidden layers with 64 units each, tanh activation functions, the Adam optimizer, a batch size of 32, and 50 training epochs.

2.4.2. Random Forest (RF)

Random Forest is an ensemble learning algorithm that combines multiple decision trees through bootstrap aggregation. Breiman [17] explained that averaging a large number of trees reduces variance and improves predictive robustness. Random Forest is widely used because it can effectively handle nonlinear relationships, noisy data, and complex feature interactions.

The Random Forest model implemented in this study used 200 trees, a maximum tree depth of 10, and a minimum samples split of 2.

2.4.3. Extreme Gradient Boosting (XGBoost)

XGBoost is an optimized gradient boosting framework that constructs decision trees sequentially while minimizing prediction error through gradient-based optimization. Chen and Guestrin [18] stated that XGBoost incorporates regularization mechanisms that enhance generalization performance and reduce overfitting.

The XGBoost model implemented in this study used 300 estimators, a maximum depth of 6, a learning rate of 0.05, and squared error as the objective function.

2.5. Experimental Setup

The dataset was chronologically divided into training and testing subsets in order to preserve temporal order. The first 80% of the timeline was used for model training, while the remaining 20% was reserved for testing. Random shuffling was not applied, as time-series forecasting requires sequential consistency.

All models were implemented using Python-based machine learning libraries. Hyperparameters were determined through preliminary tuning based on validation performance and computational efficiency.

2.6. Performance Evaluation

Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), which are widely used in regression forecasting studies. Makridakis et al. [20] highlighted that combining multiple error metrics provides a more reliable assessment of predictive quality.

$$MAE = \frac{1}{\eta} \sum_{i=1}^{\eta} |y_i - \hat{y}_i| \tag{1}$$

where y_i is the observed temperature, $\hat{y}_i$ is the predicted temperature, and n is the total number of observations.

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

$$MAPE = \frac{100}{\eta} \sum_{i=1}^{\eta} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{3}$$

Lower values of MAE, RMSE, and MAPE indicate better forecasting performance. These evaluation metrics were used to identify the most reliable model for short-term temperature prediction in the Jombang region.

2.7. Model Training Procedure

All predictive models were trained using the same chronological training dataset to ensure a fair comparative evaluation. The training process was conducted after data normalization and feature construction had been completed. For the LSTM model, training was performed iteratively using backpropagation through time with the Adam optimizer to minimize prediction loss. To enhance generalization capability and reduce overfitting, an early stopping mechanism was implemented by monitoring validation loss during training. The training process was automatically terminated when no improvement was observed for ten consecutive epochs, and the model parameters corresponding to the best validation performance were retained.

For the Random Forest model, training involved constructing multiple decision trees using bootstrap samples randomly drawn from the training dataset. Final predictions were obtained by averaging the outputs of all trees, which helps reduce variance and improve predictive robustness. In the XGBoost model, trees were built sequentially, with each new tree optimized to correct residual errors produced by the previous iteration. Regularization control and learning rate adjustment were applied to prevent excessive model complexity and enhance generalization performance.

To maintain consistency across experiments, all models were trained using identical training and testing partitions. Hyperparameter values were determined through preliminary experiments based on forecasting accuracy and computational efficiency.

2.8. Research Workflow

The complete research workflow, including data acquisition, preprocessing, feature engineering, model training, and performance evaluation, is presented in Figure 2.

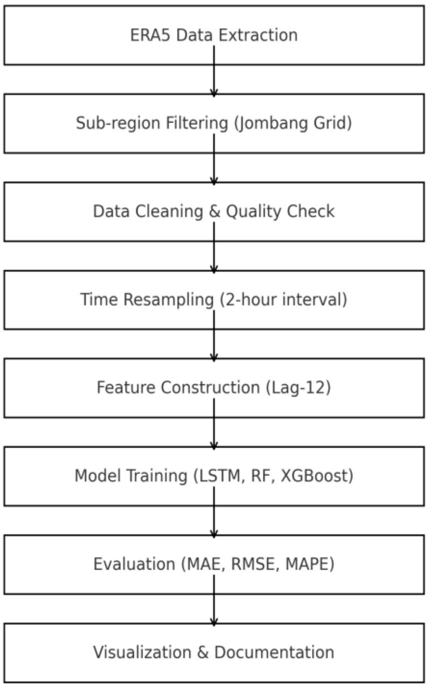

Figure 1. Illustrates the overall workflow in this research.

(1)

3. RESULTS AND DISCUSSION

3.1. Comparative Model Performance

To quantitatively assess the forecasting performance of the developed models, three widely used error metrics—Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE)—were computed using unseen testing data. These metrics offer complementary perspectives on model accuracy by capturing absolute deviations, sensitivity to squared errors, and relative prediction errors, respectively. The comparative performance of the LSTM, Random Forest, and XGBoost models is summarized in Table 1.

Table 1. the forecasting performance of the evaluated models based on MAE, RMSE, and MAPE metrics using unseen testing data.

Model	MAE (°C)	RMSE (°C)	MAPE (%)
LSTM	0.6705	0.8949	2.5274
Random Forest	0.6803	0.9227	2.5513
XGBoost	0.5982	0.8279	2.2320

As shown in Table 1, XGBoost consistently produced the lowest error values across all evaluation metrics, highlighting its strong capability in modeling nonlinear temperature dynamics. In particular, XGBoost outperformed both LSTM and Random Forest in terms of MAE, RMSE, and MAPE, demonstrating superior predictive accuracy and stability. LSTM exhibited competitive performance due to its strength in capturing temporal dependencies, while Random Forest resulted in slightly higher errors, likely reflecting its limitation in modeling sequential relationships.

The consistent outperformance of XGBoost across all metrics underscores the effectiveness of boosting-based ensemble learning for structured time-series data. Figure 4 further illustrates the comparative performance of the evaluated models in terms of MAE and RMSE, offering a clearer visual representation of their relative accuracy.

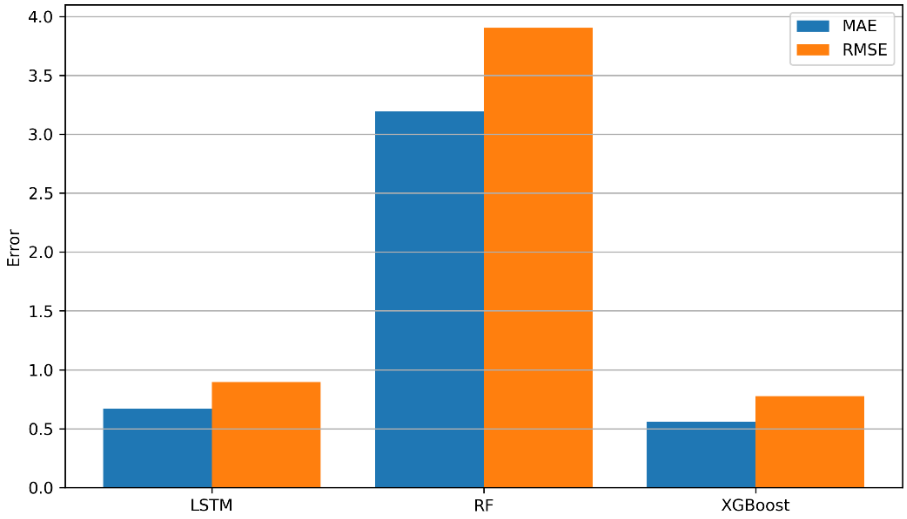

Figure 2. erformance comparison of machine learning models based on MAE and RMSE metrics.

The results indicate that XGBoost achieved the best overall performance among the evaluated models, yielding the lowest error values across all metrics. This suggests that XGBoost was more effective in capturing nonlinear relationships within the temperature time-series data.

LSTM showed competitive performance, particularly in modeling temporal dependencies, but produced slightly higher error values compared to XGBoost. In contrast, Random Forest recorded the highest error among the models, indicating that bagging-based ensemble learning was less effective than boosting methods for this dataset.

3.2. Prediction Behavior Analysis

Figure 3 presents a comparison between the actual temperature observations and the predicted values generated by the LSTM, Random Forest, and XGBoost models over a representative testing period.

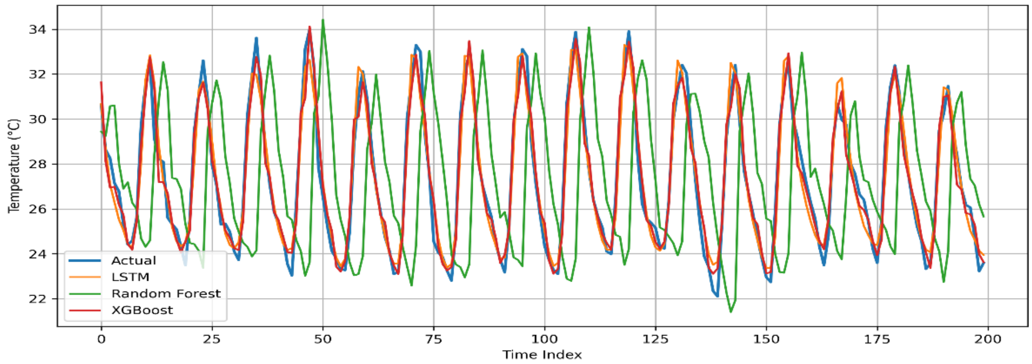

Figure 3. Comparison of actual and predicted temperature values using LSTM, Random Forest, and XGBoost models.

All models were able to capture the general diurnal pattern of temperature variation, characterized by rising temperatures during the daytime and decreasing temperatures at night. However, XGBoost showed the closest agreement with the observed temperature curve, particularly in accurately capturing peak and trough values.

LSTM produced relatively smoother predictions, which at times led to underestimation of rapid temperature changes. This behavior is typical of sequence-based neural networks, where sudden fluctuations may be moderated due to temporal smoothing. Meanwhile, Random Forest generated stable predictions but exhibited a slight lag in responding to abrupt transitions, likely due to the averaging effect across multiple decision trees.

3.3. Residual Error Analysis

Residual analysis shows that prediction errors for all models were generally centered around zero, indicating the absence of significant systematic bias. However, larger residuals were observed during periods of rapid temperature variation, such as transitions between daytime heating and nighttime cooling. Figure 4 presents the residual error distribution of the XGBoost model.

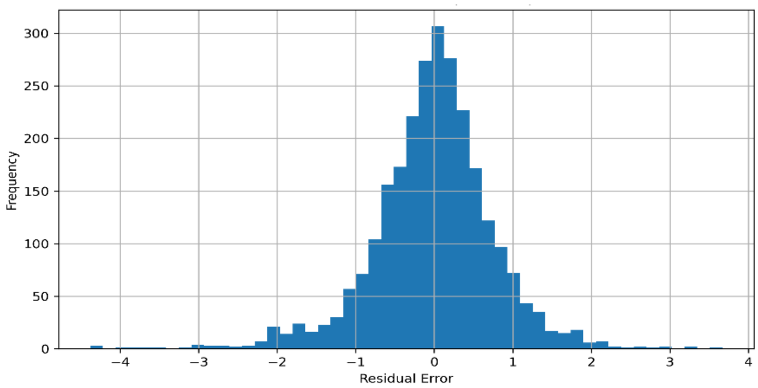

Figure 4. esidual error distribution of the XGBoost model.

Among the evaluated models, XGBoost produced the most concentrated (narrowest) residual distribution, reflecting more consistent prediction accuracy. This observation is consistent with the findings of Chen and Guestrin [18], who reported that gradient boosting methods effectively minimize residual errors through iterative optimization.

3.4. Comparison with Previous Studies

The superior performance of XGBoost observed in this study is in line with previous research on machine learning-based temperature forecasting. Fister et al. [9] demonstrated that ensemble-based models can outperform both traditional and standalone machine learning approaches in air temperature prediction. Likewise, Zhao et al. [15] reported that boosting algorithms achieve strong predictive performance due to their ability to capture complex nonlinear relationships.

The comparatively lower performance of LSTM suggests that deep learning models may require additional meteorological inputs, such as humidity, wind speed, or solar radiation, to fully leverage their temporal modeling capabilities. When only temperature data are available, tree-based ensemble methods may provide a more efficient and accurate alternative.

3.5. Practical Implications

The findings of this study offer important practical implications for regional climate analysis and decision-making. Accurate short-term temperature forecasting can support agricultural planning, optimize irrigation scheduling, and enhance preparedness for extreme temperature events.

Given its strong performance and computational efficiency, XGBoost can be considered a reliable approach for operational temperature forecasting in tropical regions such as Jombang.

3.6. Study Limitations

Despite the encouraging results, several limitations should be acknowledged. First, the study relies on ERA5 reanalysis data, which may not fully capture highly localized temperature variations. Second, the input variables were limited to historical temperature and temporal features. Incorporating additional meteorological parameters could potentially improve prediction accuracy.

Future work may focus on exploring hybrid modeling approaches, multi-step forecasting strategies, and the integration of real-time observational data to further enhance predictive performance.

4. CONCLUSION

This study developed and evaluated a machine learning-based framework for short-term air temperature forecasting in the Jombang region using ERA5 reanalysis data. Three models—Long Short-Term Memory (LSTM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost)—were implemented and assessed using standard evaluation metrics, namely MAE, RMSE, and MAPE.

The experimental results indicate that XGBoost consistently outperformed the other models, achieving the lowest error values across all evaluation metrics. This outcome suggests that boosting-based ensemble learning is highly effective in capturing nonlinear temperature dynamics in tropical environments. LSTM also demonstrated competitive performance due to its strength in modeling temporal dependencies, although its predictions tended to be smoother and less responsive to rapid fluctuations. In contrast, Random Forest produced relatively higher error values, indicating that bagging-based methods may be less suitable for sequential temperature forecasting compared to boosting approaches.

The primary contribution of this study lies in the development of a localized temperature forecasting framework based on ERA5 reanalysis data, along with a comprehensive comparison of multiple machine learning algorithms under tropical climate conditions. The findings offer practical insights for selecting appropriate predictive models in regions characterized by limited observational data and complex atmospheric dynamics.

Despite these promising results, several limitations should be acknowledged. The study relies solely on historical temperature data and temporal features, without incorporating additional meteorological variables such as humidity, wind speed, or solar radiation. Moreover, the forecasting task is limited to short-term prediction, and the use of reanalysis data may not fully capture highly localized microclimate variations.

Future research may focus on integrating multi-variable atmospheric datasets, exploring hybrid deep learning architectures, and extending the framework to multi-step forecasting scenarios. Additionally, incorporating real-time observational data could further improve prediction accuracy and support the development of operational climate forecasting systems for regional applications..

5. ACKNOWLEDGMENT

The author would like to express sincere gratitude to all individuals and institutions who contributed to the completion of this research. Appreciation is extended to the Copernicus Climate Data Store for providing access to ERA5 datasets, as well as colleagues and academic mentors who offered invaluable insights throughout the development of this work. The author also acknowledges the technical assistance received in data processing, model evaluation, and manuscript preparation.

6. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. The authors confirmed that the paper was free of plagiarism.

7. AUTHORS' CONTRIBUTION/ROLE

The first author was responsible for the overall research design, including dataset acquisition, data preprocessing, model development, and experimental analysis. The second author contributed as a machine learning expert, providing guidance on model selection, optimization strategies, and data analysis. The third author was responsible for manuscript preparation, including writing, structuring, and refining the scientific presentation of the study.

8. AI USE AND DECLARATION OF GENERATIVE AI USE

During the preparation of this work, the authors used Grammarly in order to improve the readability and language of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

8. REFERENCES

[1] [1] I. Price et al., "Probabilistic weather forecasting with machine learning," Nature, vol. 637, pp. 84–90, 2025, doi: 10.1038/s41586-024-08252-9.

[2] R. Yang et al., "Interpretable machine learning for weather and climate prediction: A review," Atmospheric Environment, vol. 338, p. 120797, 2024, doi: 10.1016/j.atmosenv.2024.120797.

[3] C. O. de Burgh-Day and T. Leeuwenburg, "Machine learning for numerical weather and climate modelling," Geoscientific Model Development, vol. 16, pp. 6433–6463, 2023, doi: 10.5194/gmd-16-6433-2023.

[4] H. Zhang et al., "Machine learning methods for weather forecasting: A review," Atmosphere, vol. 16, no. 1, p. 82, 2025, doi: 10.3390/atmos16010082.

[5] S. Haji-Aghajany et al., "Beyond the horizon: A comprehensive analysis of AI-based weather forecasting models," Engineering Applications of Artificial Intelligence, vol.

138, p. 112335, 2025, doi: 10.1016/j.engappai.2025.112335.

[6] C. Kent et al., “Skilful global seasonal predictions from a machine learning weather model trained on reanalysis data,” npj Climate and Atmospheric Science, vol. 8, p. 52, 2025, doi: 10.1038/s41612-025-01198-3.

[7] M. G. Schultz et al., “Can deep learning beat numerical weather prediction?,” Philosophical Transactions of the Royal Society A, vol. 379, 2021, doi: 10.1098/rsta.2020.0097.

[8] S. Ravuri et al., “Skillful precipitation nowcasting using deep generative models of radar,” Nature, vol. 597, pp. 672–677, 2021, doi: 10.1038/s41586-021-03854-z.

[9] D. Fister et al., “Accurate long-term air temperature prediction with machine learning models and data reduction techniques,” Applied Soft Computing, vol. 141, p. 110118, 2023, doi: 10.1016/j.asoc.2023.110118.

[10] F. Şevgin, “Machine learning-based temperature forecasting for sustainable climate change adaptation and mitigation,” Sustainability, vol. 17, no. 5, p. 1812, 2025, doi: 10.3390/su17051812.

[11] C. Voyant et al., “Machine learning methods for solar radiation forecasting: A review,” Renewable Energy, vol. 105, pp. 569–582, 2017, doi: 10.1016/j.renene.2016.12.095.

[12] Y. Li et al., “Deep learning-based short-term outdoor air temperature forecasting,” Energy and Buildings, vol. 250, p. 111281, 2022, doi: 10.1016/j.enbuild.2021.111281.

[13] X. Huang et al., “Urban air temperature prediction using machine learning approaches,” Atmospheric Research, vol. 256, p. 105576, 2021, doi: 10.1016/j.atmosres.2021.105576.

[14] J. Chen et al., “Hybrid machine learning models for hourly temperature forecasting,” Applied Energy, vol. 264, p. 114710, 2020, doi: 10.1016/j.apenergy.2020.114710.

[15] Z. Zhao et al., “Short-term thermal environment forecasting using ensemble learning,” Sustainable Cities and Society, vol. 89, p. 104327, 2023, doi: 10.1016/j.scs.2022.104327.

[16] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[17] L. Breiman, “Random forests,” Machine Learning, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[18] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[19] B. Lim and S. Zohren, “Time-series forecasting with deep learning: A survey,” Philosophical Transactions A, vol. 379, 2021, doi: 10.1098/rsta.2020.0209.

[20] S. Makridakis et al., “The M4 competition: Results, findings, conclusion and way forward,” International Journal of Forecasting, vol. 36, no. 1, pp. 54–74, 2020, doi: 10.1016/j.ijforecast.2019.04.014.

[21] H. Hersbach et al., “The ERA5 global reanalysis,” Quarterly Journal of the Royal Meteorological Society, vol. 146, pp. 1999–2049, 2020, doi: 10.1002/qj.3803.

[22] J. Muñoz-Sabater et al., “ERA5-Land: A state-of-the-art global reanalysis dataset for land applications,” Earth System Science Data, vol. 13, pp. 4349–4383, 2021, doi: 10.5194/essd-13-4349-2021.

[23] Copernicus Climate Data Store, “ERA5 hourly data on single levels from 1940 to present,” ECMWF, 2025. [Online]. Available: https://cds.climate.copernicus.eu/