ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

COMPARATIVE ANALYSIS OF LONG SHORT-TERM MEMORY AND RANDOM FOREST FOR SOIL TEMPERATURE PREDICTION IN GREENHOUSE MICROCLIMATE

Muhammad Asyril Ar Rosyad, Marzuki Sinambela✉, Agustina Rachmawardani, Fadhli Luthfiansyah
Instrumentasi-MKG, Sekolah Tinggi Meteorologi Klimatologi dan Geofisika, Tangerang, Indonesia
Email: sinambela.m@gmail.com

DOI: https://doi.org/10.46880/jmika.Vol10No2.pp901-909

ABSTRACT

Soil temperature prediction is essential for greenhouse microclimate management because it influences plant growth and resource efficiency. This study proposes an LSTM-based prediction framework for one-hour-ahead soil temperature forecasting using greenhouse microclimate data collected from an Internet of Things (IoT)-based monitoring system. Random Forest (RF) was employed as a benchmark model to validate the proposed framework. The dataset, collected from March 23 to June 3, 2026, included air temperature, air humidity, soil temperature, soil moisture, and light intensity recorded at one-minute intervals. Data preprocessing consisted of cleaning, missing value interpolation, outlier detection, target construction, and feature engineering. Four experimental scenarios with and without the soil moisture feature were evaluated using MAE, RMSE, and R^2. The proposed LSTM model without the soil moisture feature achieved the best performance, with MAE = 0.3287, RMSE = 0.5881, and R^2 = 0.9662. Feature importance and correlation analyses further showed that the soil moisture feature contributed minimally to prediction accuracy. These results demonstrate that the proposed LSTM framework provides more accurate one-hour-ahead soil temperature prediction than the benchmark RF model.

Keyword: Soil Temperature Prediction, Greenhouse Microclimate, Random Forest, LSTM.

INTRODUCTION

One of the major challenges in modern greenhouse cultivation is maintaining stable soil temperature under rapidly changing microclimate conditions. Air temperature, relative humidity, soil moisture, and light intensity are some of the environmental elements that affect soil temperature. The instability of these conditions often causes disturbances in the plant root zone, which can lead to reduced crop productivity and lower nutrient absorption efficiency. (Karyati, Rani Octaviani Putri, 2018). Therefore, maintaining an optimal soil temperature is essential for ensuring healthy plant growth and improving greenhouse production efficiency.

In greenhouse cultivation, Internet of Things (IoT)-based monitoring systems have been widely adopted to continuously monitor environmental parameters in real time. However, conventional monitoring systems only provide information about current environmental conditions and are unable to predict future changes. Nevertheless, predictive capability is crucial to support automated control systems such as irrigation regulation, ventilation, and greenhouse heating (Mampitiya et al., 2024). Consequently, integrating prediction models with IoT-based monitoring systems can support more intelligent greenhouse management by enabling preventive actions before environmental conditions reach critical thresholds.

Greenhouse microclimate data are highly dynamic, nonlinear, and characterized by strong temporal dependencies, creating significant challenges for accurate soil temperature prediction (Haryoko et al., 2026). Conventional statistical approaches often produce suboptimal results because the relationships among variables such as air temperature, air humidity, soil moisture, and solar radiation are not always linear. Because of their capacity to model nonlinear relationships and enhance prediction accuracy, machine learning and deep learning techniques have garnered significant attention.

Among various prediction methods, Random Forest (RF) and Long Short-Term Memory (LSTM) are two of the most widely used approaches in environmental data modelling. Random Forest is a machine learning algorithm that provides robust performance in modelling nonlinear relationships and handling multivariate data through an ensemble of decision trees. However, the standard RF algorithm assumes that observations are independently sampled, which may limit its ability to directly model temporal

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

dependencies in sequential data without additional feature engineering or algorithmic extensions (Hu & Szymczak, 2023). In contrast, LSTM is specifically designed for time-series forecasting by learning temporal dependencies from historical observations through its memory mechanism. This capability enables LSTM to capture dynamic changes in sequential environmental data more effectively, making it a promising approach for soil temperature prediction (İnİk et al., 2022).

Although both Random Forest and LSTM have demonstrated promising performance in environmental prediction studies, studies that specifically compare their effectiveness for one-hour-ahead soil temperature prediction in greenhouse microclimate environments remain limited. Such a comparison is important because the two models employ different learning mechanisms, possess different computational characteristics, and may exhibit different predictive capabilities when modelling greenhouse microclimate data. To address this research gap, this study proposes an LSTM-based prediction framework for one-hour-ahead soil temperature prediction using greenhouse microclimate data collected from an Internet of Things (IoT)-based monitoring system. Random Forest (RF) is employed as a benchmark model to validate the effectiveness of the proposed LSTM framework. In addition, the contribution of the soil moisture feature is investigated through feature inclusion and exclusion scenarios to evaluate its influence on prediction performance. Model performance is assessed using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). The findings of this study are expected to provide practical guidance for selecting an appropriate prediction model and evaluating the contribution of soil moisture feature as an input variable for one-hour-ahead soil temperature forecasting in greenhouse environments.

RESEARCH METHOD

This study employed an experimental approach to develop and evaluate an LSTM-based model for one-hour-ahead soil temperature prediction in a greenhouse microclimate environment using environmental parameters collected from an Internet of Things (IoT)-based monitoring system. Random Forest (RF) was employed as a benchmark model to evaluate the performance of the proposed LSTM model. Furthermore, two feature scenarios, namely with and without soil moisture, were evaluated to investigate the contribution of soil moisture to prediction performance.

Data Collection

The research data were automatically collected using an IoT-based monitoring system installed inside the greenhouse. The monitoring system was deployed in the greenhouse of the UPT Produksi Benih Tanaman Pangan Hortikultura dan Perkebunan, Tangerang, Banten, Indonesia Data acquisition was conducted from March 23 to June 3, 2026, with a one-minute sampling interval operating continuously for 24 hours per day. A total of 57,828 observation records were obtained during the monitoring period.

The environmental parameters collected included air temperature (°C), air humidity (%), soil temperature (°C), soil moisture (%), and light intensity (lux). In this study, soil temperature was used as the main target variable, while the other parameters were used as predictor variables.

Data Cleaning

Before the modelling process, the dataset underwent a data cleaning stage to ensure the quality and continuity of the time-series data. The inspection was carried out based on the chronological order of observations to identify missing data caused by sensor communication failures or data transmission issues.

The inspection results showed that there were 835 minutes of missing data. To maintain the continuity of the time-series data, a reindexing process was performed using a one-minute interval, increasing the total number of records from 57,828 to 58,663. The missing values were then filled using the linear interpolation method to preserve the continuity of data patterns. The interpolation process was performed only after the chronological order of the observations had been preserved to avoid disrupting temporal continuity.

In addition to missing data, invalid data (outliers) were also detected in the air temperature and air humidity parameters using the Interquartile Range (IQR) method (Wayan, 2026), based on the following equations:

$$\text{IQR} = \text{Q_3} - \text{Q_1} \quad (1)$$

$$\text{LowerBound} = \text{Q}_1 - 1.5 \times \text{IQR} \quad (2)$$

$$\text{UpperBound} = \text{Q}_3 + 1.5 \times \text{IQR} \quad (3)$$

Data outside this range were categorized as outliers. The identification results showed 339 outliers in air temperature and 108 outliers in air humidity, resulting in a total of 447 outliers. All detected outliers were converted into missing values (NaN) and then corrected using linear interpolation. After this process

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

was completed, the total number of records remained 58,663

Target Variable Construction

For prediction purposes, the target variable was constructed using the time shifting method. In this study, soil temperature was predicted 60 minutes ahead, so the target data were generated by shifting the soil temperature values by 60 time steps.

Mathematically, the target variable is expressed as:

$$\text{Target}_{(\text{SoilTemp})} = \text{TempSoil}_{(t+60)} \quad (4)$$

With a one-minute recording interval, the model is expected to provide soil temperature predictions one hour before the actual conditions occur

Correlation Analysis

Before the feature construction process was carried out, correlation analysis was performed to identify the relationships between environmental parameters and soil temperature. This analysis aimed to determine which features had the most significant contribution to the target variable.

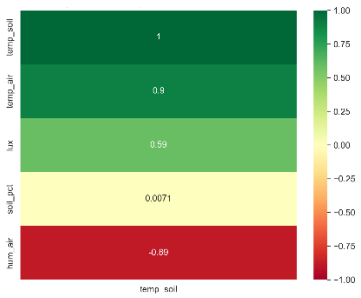


Figure 1. Correlation Soil Temperature with Environmental Parameters

The analysis results showed that air temperature had a very strong positive correlation with soil temperature, with a correlation value of 0.90, while air humidity showed a strong negative correlation of −0.89. The light intensity parameter had a moderate correlation of 0.59, whereas soil moisture (soil_pct) only showed a very weak correlation of 0.0071.

Although soil moisture (soil_pct) showed a relatively weak linear correlation with soil temperature, it was retained in the experimental scenarios because correlation analysis only reflects linear relationships between variables. Complex nonlinear relationships can be learned by machine learning and deep learning models like Random Forest and LSTM; therefore, soil moisture may still contribute to improving prediction performance.

Feature Engineering

The feature engineering stage was performed to construct informative input features that enable the prediction models to better learn the characteristics of greenhouse microclimate data. Based on the correlation analysis, two feature scenarios were evaluated: (1) features including soil moisture and (2) features excluding soil moisture. These scenarios were designed to investigate the contribution of soil moisture to one-hour-ahead soil temperature prediction.

For the Random Forest model, the predictor variables consisted of air temperature, relative humidity, soil temperature, light intensity, and soil moisture (for the corresponding feature scenario). To incorporate temporal information, lag features were generated from previous observations at 1, 5, 10, 15, 30, and 60-minute intervals. These lag features provide historical information because Random Forest does not explicitly model temporal dependencies, allowing the model to utilize previous observations for one-hour-ahead soil temperature prediction.

For the LSTM model, the same environmental variables were used together with two cyclical time features, namely hour_sin and hour_cos, to represent daily periodic patterns. These transformed features help the model recognize cyclic variations in greenhouse microclimate conditions. Furthermore, the data were arranged into sequential windows of 30 consecutive time steps (lookback = 30), enabling the LSTM model to automatically learn temporal dependencies from historical observations before predicting soil temperature 60 minutes ahead.

After feature engineering, the total number of records was reduced to 58,543 because observations at the beginning and end of the time series could not be used during the construction of lag features, sequence windows, and one-hour-ahead prediction targets.

Data Normalization

The normalization process was applied only to the LSTM model using the Min-Max Scaling method. The purpose of normalization was to transform the feature values into a range between 0 and 1, allowing the neural network training process to be more stable and helping the model converge faster (Kusnaidi et al., 2022).

The normalization process was performed using the following equation:

$$X_{(\text{norm})} = \left(X - X_{(\text{min})}\right) / \left(X_{(\text{max})} - X_{(\text{min})}\right) \quad (5)$$

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

Meanwhile, the Random Forest model did not require normalization because tree-based algorithms are generally not sensitive to differences in data scale.

Dataset Splitting

After all pre-processing stages were completed, A chronological time-series split was utilised to separate the dataset into training and testing sets, using 80% of the data for training and 20% for testing. This resulted in 46,834 training records and 11,709 testing records.

Experimental Environment

All experiments were conducted on a personal computer equipped with an Intel Core i7 processor, 16 GB RAM, and the Windows operating system. The prediction models were implemented using Python 3.11.7, with Scikit-learn 1.2.2 for the Random Forest model and TensorFlow 2.18.0 (Keras) for the LSTM model. Data processing and analysis were performed using Pandas 2.1.4 and NumPy 1.26.4.

Modelling

This study employed two prediction models for one-hour-ahead soil temperature forecasting using greenhouse microclimate data. LSTM was developed as the primary prediction model, while Random Forest (RF) was employed as a benchmark model for performance comparison. Both models were trained using the same dataset and evaluated under identical experimental conditions.

Random Forest (RF)

Random Forest (RF) was employed as the benchmark machine learning model because of its robustness in modelling nonlinear relationships among environmental variables. To produce a reliable prediction, RF builds several decision trees utilising bootstrap samples of the training data and aggregates their results via ensemble averaging. (Sládek, 2023). This approach enables the model to capture non-linear relationships among environmental parameters while reducing the risk of overfitting.

The final prediction of Random Forest is obtained by averaging the outputs of all decision trees, which can be expressed as:

$$f_{RF}(x) = \frac{1}{K}\sum_{k=1}^{K} T_k(x) \quad (6)$$

where K represents the number of regression trees and $T_k(x)$ represents the prediction generated by the k-th decision tree. The Random Forest model was trained using the engineered environmental and lag features to predict soil temperature 60 minutes ahead. The hyperparameter configuration is presented in Table 1.

Table 1. Hyperparameter Configuration of the Random Forest Model

Parameter	Value
n_estimators	150
max_depth	5
min_samples_split	30
min_samples_leaf	10
max_features	sqrt
bootstrap	True
random_state	42

Long Short-Term Memory (LSTM)

LSTM was employed as the primary deep learning model because of its ability to learn temporal dependencies from sequential observations (Sitompul et al., 2024). The input data were arranged into sequential windows consisting of 30 consecutive observations (lookback = 30), allowing the model to learn historical greenhouse microclimate patterns before predicting soil temperature 60 minutes ahead.

The main mechanism of LSTM consists of the forget gate, input gate, and output gate. The memory update process can be expressed as follows:

Forget gate:

$$f_t = \sigma\left(W_f\left[h_{(t-1)}, x_t\right] + b_f\right) \quad (7)$$

Input gate:

$$i_t = \sigma\left(W_i\left[h_{(t-1)}, x_t\right] + b_i\right) \quad (8)$$

Cell state:

$$C_t = f_t * C_{(t-1)} + i_t * \widetilde{C}_t \quad (9)$$

Output:

$$h_t = o_t * \tanh(C_t) \quad (10)$$

The LSTM model learned temporal relationships directly from the sequential input data to generate one-hour-ahead soil temperature predictions. The hyperparameter configuration of the LSTM model used in this study is summarized in Table 2.

Table 2. Hyperparameter Configuration of the LSTM Model

Hyperparameter	Value
Lookback window	30
LSTM layers	128 and 64 units
Dropout rate	0.2
Epochs	100

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

Batch size	64
Optimizer	Adam
Initial learning rate	0.001
Loss function	Mean Squared Error (MSE)
EarlyStopping	patience = 20
ReduceLROnPlateau	factor = 0.5, patience = 5

Experimental Scenarios

To investigate the contribution of soil moisture to prediction performance, both prediction models were evaluated under two feature scenarios, resulting in four experimental models, as presented in Table 3.

Table 3. Experimental Scenarios

Scenario	Model	Soil Moisture
S1	Random Forest	Included
S2	Random Forest	Excluded
S3	LSTM	Included
S4	LSTM	Excluded

Model Evaluation

The performance of the four models was evaluated using three main metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2) (Ihzaniah et al., 2023).

MAE was used to measure the average absolute prediction error:

$$MAE = (1/n)\sum(i = 1)^n|y_i - \hat{y_i}| \quad (11)$$

RMSE was used to measure the model's sensitivity to larger prediction errors:

$$RMSE = \sqrt{\left((1/n)\sum(i = 1)^n(y_i - \hat{y_i})^2\right)} \quad (12)$$

Meanwhile, R^2 was used to measure the model's ability to explain the variability of the actual data:

$$R^2 = 1 - \left(\sum(i = 1)^n(y_i - \hat{y_i})^2\right) / \left(\sum(i = 1)^n(y_i - \bar{y})^2\right) \quad (13)$$

The evaluation results of the four experimental models were compared to identify the best-performing model for one-hour-ahead soil temperature prediction and to analyse the contribution of soil moisture as an input variable in greenhouse microclimate forecasting.

RESULTS AND DISCUSSION

Model Prediction Results

After the training process was completed, the prediction results from each model were compared with the actual data to evaluate the ability of the models to follow changes in soil temperature patterns in the validation dataset. The visualization of the comparison between actual and predicted data was used to observe the level of pattern similarity and the capability of the models to capture the dynamic behaviour of soil temperature in the greenhouse environment.

Random Forest + Soil Moisture

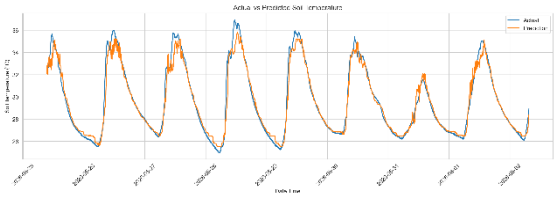


Figure 2. Rf Soil Temperature Prediction with Soil Moisture Feature

Figure 2 illustrates the prediction results of the Random Forest (RF) model using soil moisture as one of the input features. Overall, the predicted soil temperature closely followed the general trend of the observed data throughout the validation period. The model was able to capture both increasing and decreasing temperature patterns, indicating that Random Forest successfully learned the nonlinear relationships among the environmental variables. However, slight deviations were observed at several temperature peaks, where the predicted values tended to underestimate the actual temperature. These deviations indicate that although Random Forest produced stable predictions, its ability to respond to rapid temporal changes remained limited.

Random Forest - Soil Moisture

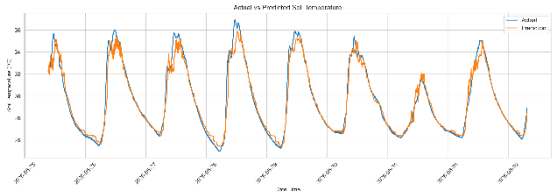


Figure 3. Rf Soil Temperature Prediction without Soil Moisture Feature

Figure 3 presents the prediction results of the Random Forest model without the soil moisture feature. The predicted curve remained highly consistent with

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

the observed soil temperature pattern and showed performance comparable to the previous scenario. Only minor differences were observed between the two feature scenarios, suggesting that the exclusion of soil moisture had little influence on the predictive capability of the Random Forest model. This observation indicates that the model mainly relied on other environmental variables, particularly air temperature and historical soil temperature, to estimate future soil temperature.

LSTM + Soil Moisture

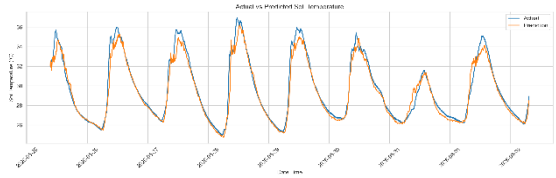


Figure 4 LSTM Soil Temperature Prediction with Soil Moisture feature

Figure 4 shows the prediction results of the Long Short-Term Memory (LSTM) model using soil moisture as an input feature. The predicted curve generally followed the observed soil temperature pattern and exhibited smoother transitions than the Random Forest model. Since LSTM is specifically designed to learn temporal dependencies from sequential observations, it was able to capture continuous changes in greenhouse microclimate conditions. Nevertheless, slight prediction delays were still observed during periods of rapid temperature increase, indicating that the model required additional time to adapt to abrupt environmental changes.

LSTM - Soil Moisture

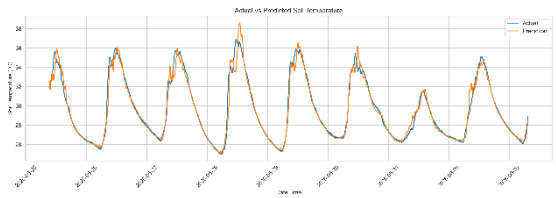


Figure 5. LSTM Soil Temperature Prediction without Soil Moisture Feature

Model Evaluation Results

To quantitatively measure model performance, three evaluation metrics were used: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2). The evaluation results on the validation dataset are presented in Table 4.

Table 4. Evaluation Results of the Prediction Models

Models	Soil Moisture Feature	MAE	RMSE	R^2
RF	Yes	0.3765	0.6186	0.9626
LSTM	Yes	0.4141	0.6578	0.9577
RF	No	0.3733	0.6209	0.9623
LSTM	No	0.3287	0.5881	0.9662

Based on the evaluation results presented in Table 4, all prediction models achieved satisfactory performance, with coefficient of determination (R^2) values exceeding 0.95. These results indicate that the developed models were able to explain more than 95% of the variability in soil temperature, demonstrating their effectiveness for one-hour-ahead prediction under greenhouse microclimate conditions.

Among the Random Forest models, the inclusion of soil moisture resulted in a slightly lower RMSE (0.6186) and a marginally higher R^2 (0.9626) than the model without soil moisture. However, the differences between the two scenarios were minimal, indicating that soil moisture contributed only marginally to the predictive performance of the Random Forest model.

In contrast, the LSTM model responded differently to the inclusion of the soil moisture feature. The model trained without soil moisture achieved the best overall performance, with an MAE of 0.3287, RMSE of 0.5881, and R^2 of 0.9662. Compared with the model that included soil moisture, the exclusion of this feature improved all evaluation metrics. This finding suggests that soil moisture provided limited additional information for one-hour-ahead soil temperature prediction in the greenhouse environment.

Overall, the LSTM model without soil moisture outperformed the other three experimental models. This result indicates that learning temporal dependencies from historical greenhouse microclimate data was more beneficial for soil temperature prediction than incorporating the soil moisture variable, which exhibited only a weak relationship with the target variable.

Computational Performance

Table 5. Computational Performance of the Prediction Models

Model	Training Time (s)	Inference Time (s)
Random Forest	2.16	0.21
LSTM	2249.81	9.95

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

In addition to prediction accuracy, the computational performance of both models was evaluated. As shown in Table 5, Random Forest required 2.16 s for training and 0.21 s for inference, whereas the proposed LSTM model required 2249.81 s for training and 9.95 s for inference. The higher computational cost of LSTM is expected because it learns temporal dependencies through a deep sequential architecture, while Random Forest constructs independent decision trees with lower computational complexity. Despite the higher computational cost, the proposed LSTM model achieved the highest prediction accuracy among all experimental scenarios, demonstrating that the additional computational effort was compensated by improved predictive performance.

Analysis Feature Importance

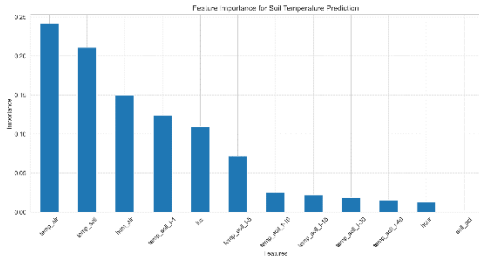


Figure 6. Feature Importance with Soil Moisture (Soil_pct)

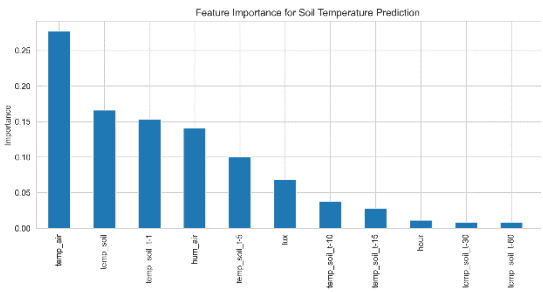


Figure 7. Feature importance without Soil Moisture (Soil_pct)

To further investigate the contribution of each predictor variable, feature importance analysis was performed using the Random Forest model. Figures 6 and 7 present the feature importance values obtained under the two experimental scenarios.

The results showed that air temperature (temp_air) was the most influential predictor in both scenarios, followed by soil temperature (temp_soil) and air humidity (hum_air). These variables consistently exhibited the highest feature importance values, indicating that they contributed more substantially to the prediction of soil temperature than the other input variables.

In contrast, soil moisture (soil_pct) exhibited the lowest feature importance among all predictor variables. This finding is consistent with the correlation analysis presented in Section 3, where soil moisture showed a Pearson correlation coefficient of only 0.0071 with soil temperature. The consistency between the correlation analysis and the feature importance results indicates that soil moisture contributed only limited additional information for one-hour-ahead soil temperature prediction in the greenhouse microclimate environment.

Model Comparison Analysis

Based on the evaluation results and prediction visualizations, the Random Forest model showed relatively stable performance under both feature scenarios, indicating that the inclusion of soil moisture had only a marginal effect on its predictive performance. In contrast, the LSTM model exhibited a more noticeable difference between the two scenarios. The model trained without soil moisture achieved better performance than the model including soil moisture, suggesting that soil moisture provided limited additional predictive information for one-hour-ahead soil temperature prediction.

Overall, the LSTM model without soil moisture achieved the best performance among all experimental models, with the lowest MAE and RMSE values and the highest R^2 value. The superior performance of LSTM can be attributed to its ability to learn temporal dependencies directly from sequential greenhouse microclimate observations through its memory mechanism. This capability enables LSTM to capture dynamic soil temperature patterns more effectively than the Random Forest model.

These findings are consistent with both the correlation analysis and the feature importance analysis, which showed that soil moisture exhibited only a weak relationship with soil temperature and the lowest contribution among the predictor variables. Therefore, the inclusion of additional features does not necessarily improve prediction performance when those features contribute limited predictive information.

The findings of this study are consistent with previous studies reporting that LSTM generally outperforms conventional machine learning algorithms for environmental time-series prediction because of its ability to learn temporal dependencies from sequential observations. Similar conclusions were reported by İnİk et al. (2022) and Sitompul et al. (2024), who demonstrated that LSTM achieved higher prediction accuracy than conventional machine learning approaches for environmental forecasting tasks. Meanwhile, the stable performance of Random Forest

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

observed in this study agrees with Sládek (2023), who reported that ensemble tree models effectively capture nonlinear relationships among environmental variables but rely heavily on feature engineering to represent temporal information. These results further confirm that deep learning approaches provide greater advantages when modelling highly dynamic greenhouse microclimate data.

From a practical perspective, the results indicate that excluding the soil moisture parameter does not reduce prediction performance for one-hour-ahead soil temperature forecasting. This finding suggests that greenhouse monitoring systems dedicated to soil temperature prediction may reduce sensor deployment, maintenance, and calibration requirements by omitting soil moisture sensors without sacrificing prediction accuracy. However, soil moisture measurements remain important for irrigation management and other crop management applications.

CONCLUSION

The study's findings and analysis allow for the following deductions to be made:

1) Both the Random Forest (RF) and Long Short-Term Memory (LSTM) models achieved satisfactory performance for one-hour-ahead soil temperature prediction in the greenhouse microclimate environment, as indicated by coefficient of determination (R^2) values exceeding 0.95 across all experimental scenarios.
2) The inclusion of the soil moisture parameter had only a limited effect on the performance of the Random Forest model, as evidenced by the minimal differences in MAE, RMSE, and R^2 between the scenarios with and without soil moisture.
3) For the LSTM model, excluding the soil moisture parameter resulted in the best prediction performance. This finding indicates that soil moisture provided limited additional predictive information for one-hour-ahead soil temperature prediction under the greenhouse conditions observed in this study.
4) The feature importance analysis identified air temperature (temp_air), soil temperature (including its lag features), and air humidity (hum_air) as the most influential predictor variables, whereas soil moisture (soil_pct) exhibited the lowest contribution. This result is consistent with the correlation analysis, which showed a very weak relationship between soil moisture and soil temperature.
5) Overall, the LSTM model without soil moisture was identified as the best-performing model among the four experimental scenarios. These findings demonstrate that learning temporal dependencies from historical greenhouse microclimate data can provide more accurate one-hour-ahead soil temperature predictions than the benchmark Random Forest model for the dataset used in this study.
6) Although the proposed LSTM model required longer training and inference times than Random Forest, it achieved the highest prediction accuracy. These findings demonstrate the trade-off between computational efficiency and predictive performance when selecting prediction models for real-time greenhouse IoT applications.

REFERENCES

Haryoko, H., Farida, L. D., Aminuddin, A., & Ghozali, B. (2026). Analisis Komparatif Algoritma Random Forest dan Multiple Linear Regression untuk Prediksi Suhu Udara dengan Rekayasa Fitur Siklik. *The Indonesian Journal of Computer Science Research, 5*(1), 99-107. https://doi.org/10.59095/ijcsr.v5i1.255

Hu, J., & Szymczak, S. (2023). A review on longitudinal data analysis with random forest. *Briefings in bioinformatics, 24*(2). https://doi.org/10.1093/bib/bbad002

Ihzaniah, L. S., Setiawan, A., & Wijaya, R. W. N. (2023). Perbandingan kinerja metode regresi k-nearest neighbor dan metode regresi linear berganda pada data Boston housing. *Jambura Journal of Probability and Statistics, 4*(1), 17-29. https://doi.org/10.34312/jjps.v4i1.18948

İnik, O., İnik, Ö., Öztaş, T., & Yuksel, A. (2022). Soil temperature prediction with long short-term memory (LSTM). *Turkish Journal of Agricultural and Natural Sciences, 9*(3), 779-785. https://doi.org/10.30910/turkjans.1101753

Karyati, K., Putri, R. O., & Syafrudin, M. (2018). Suhu dan kelembaban tanah pada lahan revegetasi pasca tambang di PT Adimitra Baratama Nusantara, Provinsi Kalimantan Timur. *AGRIFOR: Jurnal Ilmu Pertanian dan Kehutanan, 17*(1), 103-114. https://doi.org/10.31293/af.v17i1.3280

Kusnaidi, M. R., Gulo, T., & Aripin, S. (2022). Penerapan normalisasi data dalam mengelompokkan data mahasiswa dengan menggunakan metode K-means untuk menentukan prioritas bantuan uang kuliah tunggal. *Journal of Computer System and Informatics (JoSYC), 3*(4), 330-338. https://doi.org/10.47065/josyc.v3i4.2112

Mampitiya, L., Rozumbetov, K., Rathnayake, N., Erkudov, V., Esimbetov, A., Arachchi, S., Kantamaneni, K., Hoshino, Y., & Rathnayake, U. (2024). Artificial intelligence to predict soil temperatures by development of novel model.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

Scientific Reports, 1–16. https://doi.org/10.1038/s41598-024-60549-x

Wayan, J. (2026). Perbandingan Metode IQR vs Isolation Forest dalam Deteksi Outlier pada Data Klaim Asuransi di PT Askrindo Cabang Semarang. *Jurnal Transformatika, 23*(2), 268-277.

Sitompul, P. A. U., Sujaini, H., Yacoub, R. R., & Ananda, N. (2024). Penerapan Model Long Short-Term Memory (LSTM) Untuk Prediksi Kualitas Udara dan Analisis Korelasi Terhadap Unsur Cuaca di Jongkat Kabupaten Mempawah. *Jurnal Sains & Teknologi Modifikasi Cuaca, 25*(1), 33-42. https://doi.org/10.55981/jstmc.2024.4990

Sládek, D. (2023). Application of the Random Forest method on the observation dataset for visibility nowcasting. *G eofizika, 40*(1), 75-101. https://doi.org/10.15233/gfz.2023.40.5

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.