JOIN (Jurnal Online Informatika)
p-ISSN: 2528-1682, e-ISSN: 2527-9165
Volume 10 Number 1 | June 2025: 227-238
DOI: 10.15575/join.v10i1.1571

Enhancing Weather Monitoring for Agriculture with Deep Learning: Anomaly Detection in East Java Using LSTM Autoencoder and OCSVM

Maulana Ahsan Fadillah[1], Yenni Angraini[2], Rahma Anisa[3]
[1,2,3]Study Program of Statistics and Data Science, IPB University, Indonesia

Article Info

Article history:

Received February 13, 2025
Revised April 20, 2025
Accepted April 21, 2025
Published June 05, 2025

Keywords:

Anomaly Detection
LSTM autoencoder
LSTM
OCSVM
Weather

ABSTRACT

Agricultural productivity in East Java is under threat from unpredictable and harsh weather patterns, particularly rapid variations in sunlight length and rainfall intensity. These abnormalities can interrupt agricultural cycles, lower yields, and make farming communities more vulnerable to climatic calamities. However, current weather monitoring systems frequently fall short of detecting small anomalies in time series weather data that could serve as early warning signs of such disasters. This study seeks to close this gap by creating a robust anomaly detection methodology adapted to time-dependent weather variables important to agriculture. In this study, a hybrid model combining Long Short-Term Memory (LSTM) autoencoder and One-Class Support Vector Machine (OCSVM) is proposed. The LSTM autoencoder's structure reconstructs time series data and signifies anomalies through reconstruction errors (MSE), while OCSVM validates these anomalies to reduce false positives. The model was applied to daily weather data from East Java spanning 2015–2024. The results showed that the model effectively detected 11 anomalies in sunlight duration and 7 in rainfall, with F1-scores of 0.71 and 0.82, respectively. Several of these anomalies corresponded to actual disaster events such as floods, landslides, and droughts. This research contributed to the field by demonstrating the effectiveness of combining deep learning and machine learning for weather anomaly detection. The proposed framework offers valuable insights for early warning systems and can support local governments and farmers in improving disaster preparedness and enhancing agricultural resilience in East Java.

Corresponding Author:

Yenni Angraini,
Study Program of Statistics and Data Science, School of Data Science, Mathematics and Informatics, IPB University
Jl. Raya Darmaga Kampus IPB, Babakan, Kec. Dramaga, Kabupaten Bogor, Jawa Barat 16680
Email: y_angraini@apps.ipb.ac.id

1. INTRODUCTION

Especially in climate-sensitive areas like East Java, Indonesia, unpredictable weather patterns, especially with regard to rainfall and sunlight length, have become a significant obstacle to agricultural productivity. These weather irregularities have the potential to alter agricultural cycles, lower yields, and make people more susceptible to natural calamities like landslides, floods, and droughts. Therefore, it is essential to identify these abnormalities early on in order to guarantee agricultural sustainability and enhance readiness for disasters.

p-ISSN: 2528-1682
e-ISSN: 2527-9165

Anomaly detection, defined as the identification of rare or unexpected patterns within data[1], has shown promising applications in climate monitoring. However, a lot of conventional methods have trouble efficiently extracting temporal features, particularly when working with multivariate time-series data that captures the intricate and ever-changing character of weather systems. Present studies emphasize the value of machine learning techniques such as One-Class Support Vector Machine (OCSVM) for anomaly detection and deep learning models, especially Long Short-Term Memory (LSTM) networks, for capturing temporal relationships. A review of existing methods highlights their strengths and limitations. For instance, [2] effectively used OCSVM to detect groundwater anomalies but faced challenges with data dimensionality and temporal modeling. Deep learning methods have shown significant potential in time series anomaly detection. However, as noted in [3], these models still face limitations when dealing with high-dimensional multivariate datasets with complex temporal dynamics. To address this, [4] suggested integrating LSTM networks with anomaly detection for better temporal feature extraction.

Although each model has its advantages, OCSVM approaches may perform poorly with sequential data, and LSTM-based models sometimes lack strong decision bounds for outlier identification[5]. Recent studies have suggested combining OCSVM, which more successfully validates these anomalies, with LSTM autoencoders, which compress and reconstruct sequential data to highlight anomalies depending on reconstruction error, in order to overcome these constraints. By utilizing both robust classification and temporal feature extraction, this hybrid technique allows for more accurate anomaly detection in agriculture[4].

To identify anomalies in multivariate meteorological time-series data in East Java, this study suggests a hybrid system that combines an LSTM autoencoder with an OCSVM. It focuses on rainfall and sunlight length, two important factors that have a big impact on agriculture. The fact that this study uses actual, unlabeled agricultural weather data rather than labeled or single-variable datasets makes it more useful for real-world early warning systems than many other studies. In addition to identifying anomalies when they arise, the model is made to examine trends over time in a variety of variables.

This study illustrates how cutting-edge AI techniques can be used to address actual environmental and agricultural issues from the standpoint of information systems. The suggested method provides improved accuracy in identifying anomalies inside intricate, high-dimensional datasets by fusing deep learning and machine learning approaches. This aligns with the current technology revolution, where intelligent systems play a growing role in decision-making. The novelty of this study lies in its ability to support early disaster risk mitigation and promote agricultural resilience, especially in regions like East Java that are vulnerable to extreme weather.

2. METHOD

2.1. Data Source

The data used in this study is secondary data on sunlight duration (in hours) and rainfall (in millimeters) in East Java, sourced from the Meteorology, Climatology, and Geophysics Agency (BMKG) at the East Java Climatology Station. The data set consists of daily records of sunlight duration and rainfall, with a total of 3,288 entries covering the period from May 1, 2015, to April 30, 2024.

2.1.1 Data Exploration

To learn more about the features of the dataset, data exploration was proceeded. This procedure involved creating descriptive statistics to describe important characteristics such as the mean, standard deviation, minimum, and maximum values of each meteorological variable. Time series analysis was also performed to investigate any trends, seasonal patterns, and possible anomalies throughout time. The purpose of this exploratory stage was to determine the data's natural behavior, identify any missing or excessive values, and ensure the data was suitable for further anomaly detection modeling.

2.1.2 Data Preprocessing

Perform data preprocessing by imputing missing values using data from the Pasuruan and Malang Geophysics Stations. Linear interpolation (equation 1) [6] was applied for imputing single missing value, while the LSTM algorithm was for consecutive missing data, as outlined in Table 1 and based on studies [7] and [5].

$$f(x) = f(x_o) + \frac{f(x_1)-f(x_0)}{x_1-x_0}(x - x_0) \quad (1)$$

Table 1. Initialization of LSTM Model Hyperparameter for Imputation

Component	Parameter	Value
General	Optimizer	Adam
	Learning rate	0.0001
	Batch size	32
	Number of epochs	50
	Time steps	180
LSTM	LSTM units	50
	Activation function	Tanh

If there exists any indication of overfitting, dropout or L2 regularization will be applied. Dropout (0.1) will randomly deactivate 10% of the units during training to reduce dependence on specific connections and improve generalization [8]. If dropout alone is insufficient, L2 regularization (0.1) will be used to penalize large weights in the loss function, keeping them small to limit model complexity. This prevents the model from overfitting to minor details in the training data. L2 regularization is defined in Equation (2).

$$L_2 = \Sigma_t w_t^2 \quad (2)$$

Equation (2) calculates the L2 regularization penalty by summing the squares of all weights w_t in the model. The total loss function, which combines the original loss function (MSE) with the L2 regularization penalty, is expressed in Equation (3).

$$L_{total} = MSE + \lambda L_2 \quad (3)$$

L_{total} represents the total loss function, MSE is the original loss function that measures the model's prediction error, λ is the hyperparameter that controls the magnitude of the penalty for large weights, and $L2$ is the L2 regularization penalty [9].

Early stopping is also applied to halt training when the validation loss shows no improvement over several epochs, preventing overtraining of the model and potential overfitting [10]. Model evaluation is performed by ensuring there are no signs of overfitting, with close attention to a minimal difference between training loss and validation loss. Additionally, the lowest RMSE is the main consideration, indicating that the model performs well in imputing data, as shown in Equation (4) [11].

$$MSE = \sqrt{\frac{\sum_{t=1}^{n}(\hat{r}_t - r_t)^2}{n}} \quad (4)$$

$\hat{r}_t$ represents the predicted value, r_t is the observed value, and n is the number of data points.

The data is split into the following parts, May 1, 2015 – April 30, 2022, as the training set and May 1, 2022 – April 30, 2024, as the test set. Data standardization is also performed using a robust scaler to ensure all variables have a consistent scale. [12].

2.2. The LSTM autoencoder model

Initializing LSTM autoencoder hyperparameters, such as the number of layers, neurons, activation functions, and window size, is required to ensure the model's capability to accurately extract and reconstruct data. Dropout or L2 regularization will be applied if there are indications of overfitting. Referring to studies by [13], [5], and [4], the LSTM autoencoder hyperparameters are shown in Table 2.

Table 2. Initialization of Hyperparameter for the LSTM Autoencoder Model

Component	Parameter	Hyperparameter
General	Optimizer	Adam
	Learning rate	0.0001
	Batch size	32
	Number of epochs	50
	Time steps	180
Encoder	LSTM units	50
	Activation function	Tanh
Decoder	LSTM units	50
	Activation function	Tanh

If overfitting occurs, dropout values between 0.1 and 0.2 will be applied to improve generalization. If overfitting persists, L2 regularization (0.1–0.2) will penalize large weights, reducing model complexity. Early stopping will also be used to halt training when validation loss stops improving, preventing further overfitting.

The process is then continued by training the LSTM autoencoder model on each window using the training data within the window as input, with a predetermined window length. The model reconstructs the data following the window, using the best-trained hyperparameter combination on the test data. The reconstruction error, calculated as |reconstructed data – original data| [19], is then used as input for anomaly detection with the OCSVM method.

Finally, the model will be evaluated using loss plots and RMSE. Evaluation is conducted by ensuring a minimal difference between training loss and validation loss to avoid overfitting, while also considering the smallest RMSE to ensure accurate predictions. A low RMSE value further validates the model's strong capability to reconstruct data and detect anomalies effectively [20].

2.3. The OCSVM model

Building the OCSVM model involves determining the threshold for early anomaly detection using the LSTM autoencoder through hyperparameter tuning. Reconstruction errors guide the selection of Nu (ν) and Gamma (γ) via grid search, with the kernel function influencing data separation. Gamma controls the radius of influence, while Nu defines the proportion of anomalies. Percentiles are used to set thresholds for suspected anomalies in the LSTM autoencoder. Hyperparameter selection, based on studies [14] and [13], ensures optimal model performance, as outlined in Table 3.

Table 3. Initialization of Hyperparameter for the OCSVM Model

Component	Parameter
Optimizer	RBF
Learning rate	0.01, 0.03, 0.05, 0.10, 0.50
Batch size	0.01, 0.001
Number of epochs	98, 99

The hyperparameter tuning process optimizes the values of nu, gamma, and the threshold to achieve the best results, taking into account precision, recall, and F1-score, as demonstrated in Equations (5), (6), and (7). A high F1-score is prioritized as the main evaluation metric, reflecting the balance between precision and recall for optimal anomaly detection [13].

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FP}} \quad (5)$$

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FN}} \quad (6)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

TP (True Positive) is the number of correctly identified anomalies, TN (True Negative) is the number of correctly identified normal, FP (False Positive) is the number of incorrectly identified anomalies, and FN (False Negative) is the number of incorrectly identified normal.

2.4. Detecting Anomalies

The LSTM autoencoder, which finds anomalies based on reconstruction error, was used to detect anomalies early. Any notable differences between the original and rebuilt data (shown as a high reconstruction error) are identified as possible anomalies in this approach, which uses a model to try to recreate the input time-series data. Following a hyperparameter tuning procedure, detection thresholds were established using percentile values from the reconstruction error distribution. These first irregularities functioned as suspect labels for additional examination. Time-series graphics that highlight the points identified as aberrant were developed to help comprehend the temporal distribution of these anomalies [13].

By using One-Class Support Vector Machine (OCSVM) on the reconstruction error output from the LSTM autoencoder, anomaly detection was further refined. This method employs OCSVM to enhance the reliability point classification in high-dimensional space, however it continues to interpret high reconstruction error values as possible anomalies, consistent with the original method. To confirm whether the abnormalities were consistent with actual events, the detected anomalies were then compared to historical records of disaster events, including droughts and floods. This step offered a contextual assessment of the model's efficacy. For a better comparison, time-series representations were recreated using the OCSVM-based anomalous points in addition to the disaster events [14].

The performance of the combined LSTM autoencoder–OCSVM model was assessed quantitatively as the last stage. Precision, recall, and F1-score—standard classification metrics—were computed using the early suspected anomaly labels. The ROC curve and the Area Under the Curve (AUC) were also used to evaluate the model's overall performance. The capacity of the model to differentiate between normal and anomalous data is demonstrated by a high AUC value, which is close to 1.00 [15]. Plotting the anomalous points from the early LSTM detection and the improved OCSVM method alongside real-world disaster events allowed for visual comparisons and further confirmed the model's applicability in environmental and agricultural monitoring.

3. RESULT AND DISCUSSION

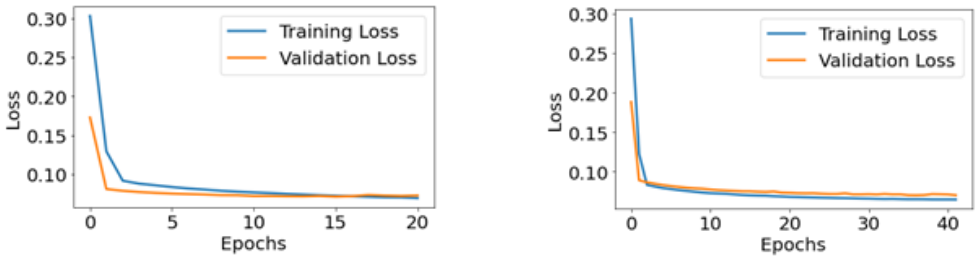
3.1. Data preprocessing and exploration

The initial data preprocessing involved imputing missing values for sunlight duration and rainfall in East Java from May 1, 2015, to April 30, 2024. The original dataset consists of 3,228 records for sunlight duration and 2,952 for rainfall, which increased to 3,288 after imputation. The process began with data from two nearby units, imputing 37 missing sunlight duration values and 230 for rainfall. Linear interpolation then imputed 93 missing rainfall points, while no interpolation was needed for sunlight duration. Finally, the LSTM algorithm estimated 13 consecutive missing points for both variables, successfully completing the imputation.

Table 4. Optimal Imputation Model Hyperparameters for LSTM of Sunlight Duration and Rainfall

Component	Parameter	Hyperparameter	
		Sunlight Duration	Rainfall
General	Optimizer	Adam	Adam
	Learning rate	0.0001	0.0001
	Batch size	32	32
	Number of epochs	50	50
	Time steps	180	180
LSTM	LSTM units	50	50
	Activation function	Tanh	Tanh
	Regularization	-	L2 (0.1)

p-ISSN: 2528-1682
e-ISSN: 2527-9165



(a) (b)

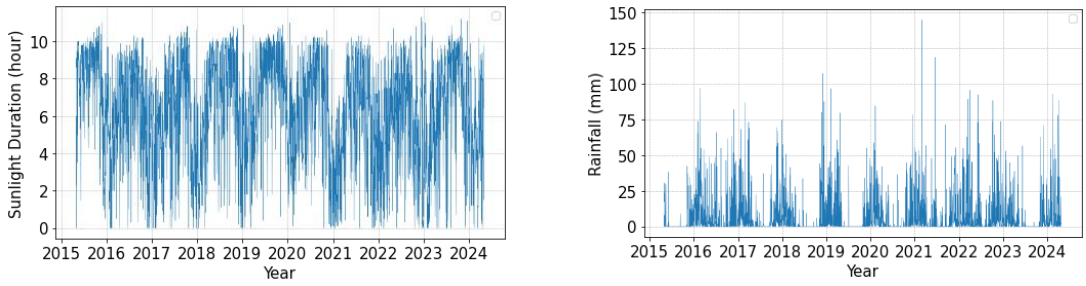
Figure 1. Training and validation of the best LSTM imputation model for (a) sunlight duration and (b) rainfall

The LSTM autoencoder model used in this study, as shown in Table 4, was selected from 3 combinations of hyperparameters for sunlight duration and 9 combinations for rainfall, based on the lowest RMSE and stable loss plots with no signs of overfitting. In Figures 1(a) and 1(b), both the training loss and validation loss consistently decrease with only small differences, indicating that the selected LSTM model effectively learns the patterns of sunlight duration and rainfall data, minimizing imputation errors. The evaluation results of the best model in Table 4 show an RMSE of 0.2590 for sunlight duration and 0.1324 for rainfall, indicating that the model can reconstruct the data with relatively small errors.

Table 5. Descriptive Statistics of Data Before and After Imputation

Component	Parameter	Hyperparameter	
		Sunlight Duration	Rainfall
General	Optimizer	Adam	Adam
	Learning rate	0.0001	0.0001
	Batch size	32	32
	Number of epochs	50	50
	Time steps	180	180
LSTM	LSTM units	50	50
	Activation function	Tanh	Tanh
	Regularization	-	L2 (0.1)

Table 5 shows descriptive statistics before and after imputation. The average daily sunshine duration slightly changed from 6.2185 hours (SD: 2.8246) to 6.2166 hours (SD: 2.8173), while average rainfall decreased from 6.6178 mm (SD: 14.0414) to 6.4101 mm (SD: 13.7273). Minimum values remained at 0.0000, indicating days without sunshine or rainfall. The median sunshine duration stayed at 6.6000 hours, and median rainfall at 0.0000 mm. Maximum values were 11.3000 hours for sunshine and 145.0000 mm for rainfall, indicating extreme weather. After imputation, the dataset remained consistent with 3,288 data points for both variables.



(a) (b)

Figure 2. Time series plot (a) sunlight duration and (b) rainfall

Data exploration in Figure 2 shows that sunshine duration and rainfall in East Java display clear seasonal patterns and significant daily variations. Sunshine duration varies from extremely low to extremely high, while rainfall fluctuates from light to heavy, both reflect seasonal trends. Recognizing

these patterns is a critical step for detecting weather anomalies affecting East Java's agriculture. After imputation and exploration, the data was split into training and testing sets and standardized using a robust scaler to ensure consistent scale and minimize outlier influence.

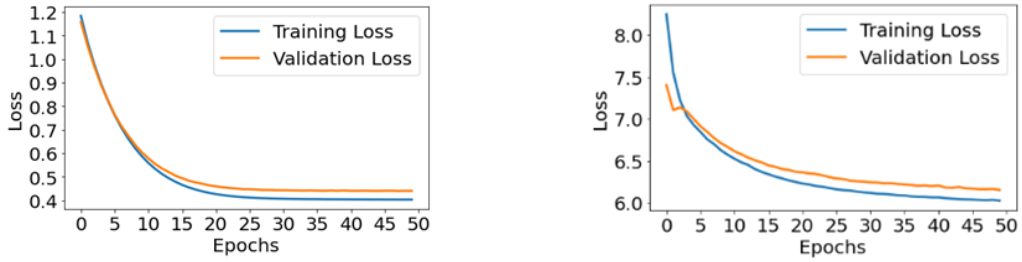
3.2. LSTM Autoencoder Modeling

The duration of sunlight and rainfall are used as inputs in the process of modeling an LSTM autoencoder, where the training data is utilized to train the model. At this stage, time steps are an important hyperparameter in determining the amount of data used as input for prediction (reconstruction). Time steps enable the network to leverage information from previous days to create more accurate predictions [16].

Table 6. Architecture and Best Hyperparameters of the LSTM Autoencoder Model

Component	Parameter	Hyperparameter	
		Sunlight Duration	Rainfall
General	Optimizer	Adam	Adam
	Learning rate	0.0001	0.0001
	Batch size	32	32
	Number of epochs	50	50
	Time steps	30	180
Encoder	LSTM units	50	50
	Activation function	Tanh	Tanh
	Regularization	L2 (0.01)	L2 (0.1)
Decoder	LSTM units	50	50
	Activation function	Tanh	Tanh
	Regularization	L2 (0.01)	L2 (0.2)

The LSTM autoencoder model, selected from 15 hyperparameter combinations based on the lowest RMSE and stable loss (Table 6), has two LSTM layers in both the encoder and decoder, extended by one additional layer. It uses the Adam optimizer with a 0.0001 learning rate, a batch size of 32, and 50 epochs. Time steps are 30 for sunlight duration and 90 for rainfall, with 50 units and Tanh activation in each LSTM layer. L2 regularization is 0.01 for sunlight duration and 0.02 for rainfall.



(a) (b)

Figure 3. Plot of training and validation loss (a) sunlight duration and (b) rainfall

Figure 3 illustrates that both the training loss and the validation loss consistently decrease for sunlight duration and rainfall, indicating that the LSTM autoencoder model effectively learns patterns and minimizes reconstruction errors. The validation loss being slightly higher than the training loss, along with the small difference between the two, suggests that the model is not overfitting and is capable of generalizing well. The model's RMSE is 0.6625 for sunlight duration and 2.4559 for rainfall, indicating relatively small reconstruction errors and good anomaly detection performance.

3.3. Hyperparameter Tuning for OCSVM Model and Threshold for Early Anomaly Detection Using LSTM Autoencoder

The OCSVM (One-Class Support Vector Machine) modeling was conducted using the reconstruction error data, in parallel with the hyperparameter tuning via grid search to select the best gamma and Nu values. The kernel function employed is the RBF (Radial Basis Function). The Nu value controls the number of data points considered anomalous, while Gamma influences the RBF kernel function. Grid search is utilized to explore various combinations of Nu and Gamma to identify the optimal configuration that maximizes anomaly detection.

Table 7. Best Hyperparameter for the OCSVM Model and LSTM Autoencoder Threshold

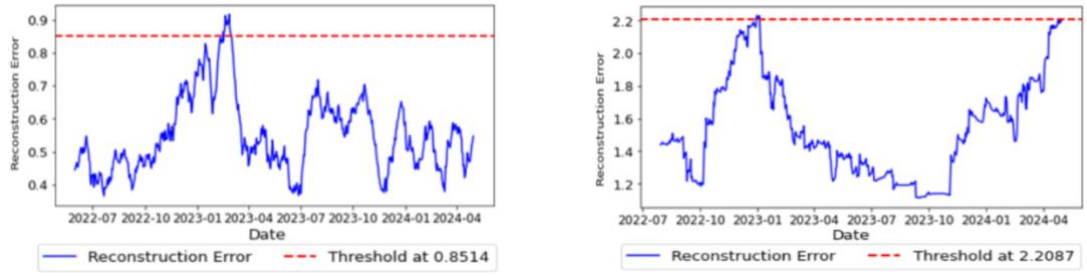
p-ISSN: 2528-1682
e-ISSN: 2527-9165

Data	Parameter			Evaluation Metrics		
	Percentile (%)	Gamma	Nu	F1-score	Recall	Precision
Sunlight Duration	98	0.01	0.03	0.7096	0.7857	0.6470
Rainfall	99	0.001	0.03	0.8235	1.0000	0.7000

Hyperparameter tuning via grid search for the OCSVM model (Table 7) identified the best parameters from 20 combinations. For sunlight duration, the optimal parameters were Gamma 0.01 and Nu 0.03, while for rainfall, Gamma 0.001 and Nu 0.03 were best. A smaller gamma value implies a broader influence of the training examples, and a Nu value of 0.03 indicates that 3% of the data are classified as anomalous. The optimal percentiles were 98% for sunlight duration and 99% for rainfall, capturing 98% and 99% of the data as normal, respectively.

3.4. Initial Anomaly Detection Using LSTM Autoencoder

The determination of the reconstruction error threshold for the LSTM autoencoder is carried out by considering the final evaluation metrics using OCSVM to ensure accurate anomaly detection. The threshold is set using the percentile from the distribution of reconstruction errors, based on the results of hyperparameter tuning.

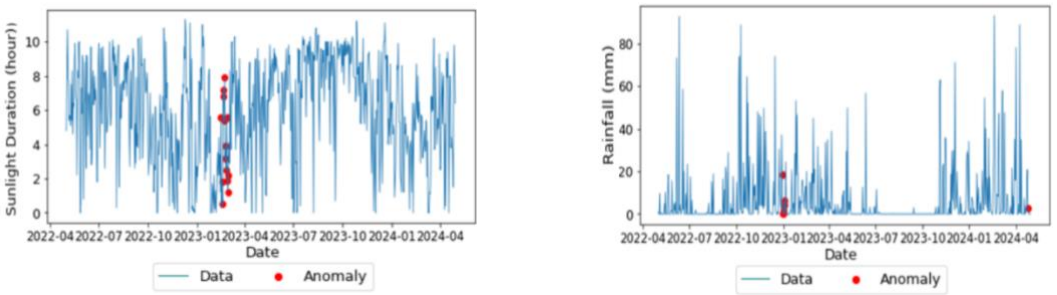


(a) (b)

Figure 4. Reconstruction error produced by the LSTM autoencoder model on data: (a) sunlight duration and (b) rainfall

Figure 4 illustrates the reconstruction error from sunlight duration and rainfall data. The hyperparameter tuning yields a threshold of 0.8831 (98th percentile) for sunlight duration, identifying the top 2% of errors as anomalies, while the threshold for rainfall is 0.2087 (99th percentile), identifying the top 1% of errors. Figure 5 illustrates that the LSTM autoencoder detected 14 anomalies in sunlight duration and 7 in rainfall. These anomalies represent data points that deviate from the overall pattern, rather than merely reflecting extreme values.

As shown in Figure 5, the time series plot, which includes the anomaly points detected by the LSTM autoencoder for both datasets, demonstrates that the model is able to identify points that significantly deviate from the normal pattern. This is indicated by high reconstruction errors, suggesting that the model struggled to reconstruct the data.

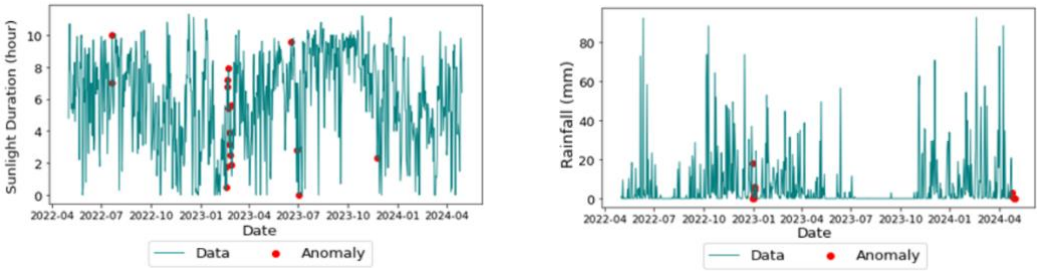


(a) (b)

Figure 5. Time series plot showing LSTM autoencoder anomaly points on data: (a) sunlight duration and (b) rainfall

3.5. Anomaly Detection Using LSTM Autoencoder Reconstruction Error in OCSVM

The hyperparameters used in anomaly detection with LSTM autoencoder reconstruction error in OCSVM, such as nu and gamma, were determined through a hyperparameter tuning process to ensure optimal model performance in detecting anomalies. Evaluation with OCSVM shows advanced anomaly detection results. As seen in Figure 6, the OCSVM model detected a total of 17 anomalies for the sunlight duration data, while 10 anomalies were detected for the rainfall data.



(a) (b)

Figure 6. Time series plot including OCSVM anomaly points on data (a) sunlight duration and (b) rainfall

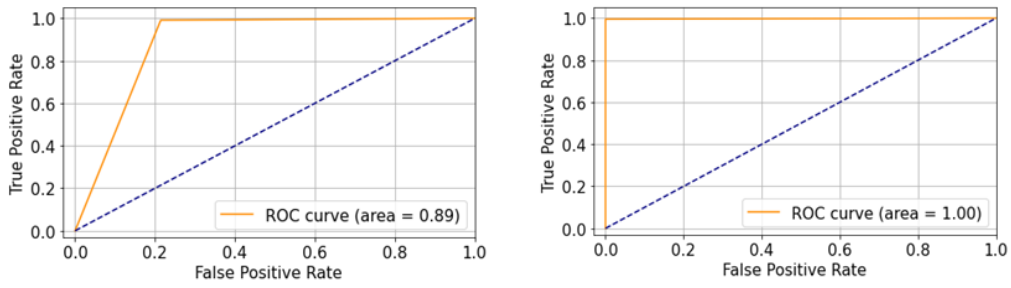
In Figure 6, a time series plot is displayed, which includes the anomaly points detected by OCSVM for both datasets. This plot shows that OCSVM successfully identified more anomaly points compared to the initial detection using the LSTM autoencoder. This indicates the model's ability to detect anomalies that may have gone undetected during the initial anomaly detection stage.

3.6. Evaluating of Anomaly Detection Results

Table 8 presents the evaluation metrics for initial anomaly detection using the LSTM autoencoder and OCSVM on sunlight duration and rainfall data. The model achieved a precision of 0.65, recall of 0.79, and an F1-score of 0.71 for sunlight duration anomalies at the 98th percentile. For rainfall anomalies at the 99th percentile, it achieved a precision of 0.70, perfect recall of 1.00, and an F1-score of 0.82.

Table 8. Best Hyperparameter for the OCSVM Model and LSTM Autoencoder Threshold for Rainfall

Data	Percentile (%)	Precision	Recall	F1-score
Sunlight Duration	98	0.65	0.79	0.71
Rainfall	99	0.70	1.00	0.82



(a) (b)

Figure 7. ROC curves and AUC scores for OCSVM anomaly detection on data (a) sunlight duration and (b) rainfall

The ROC curve is key in evaluating anomaly detection performance. In Figure 7, the AUC for sunlight duration is 0.89, showing strong model performance, while the AUC for rainfall is 1, indicating perfect accuracy in detecting anomalies. This highlights the model's excellent results for both rainfall and sunlight duration.

An AUC close to 1 shows the model's strong ability to detect anomalies with few false positives. The ROC curve reflects the balance between true and false positive rates at different thresholds. In agriculture, accurate anomaly detection aids decision-making and risk management. These results

p-ISSN: 2528-1682
e-ISSN: 2527-9165

confirm that the LSTM autoencoder and OCSVM combination is effective for detecting weather anomalies, supporting adaptive monitoring and planning in East Java.

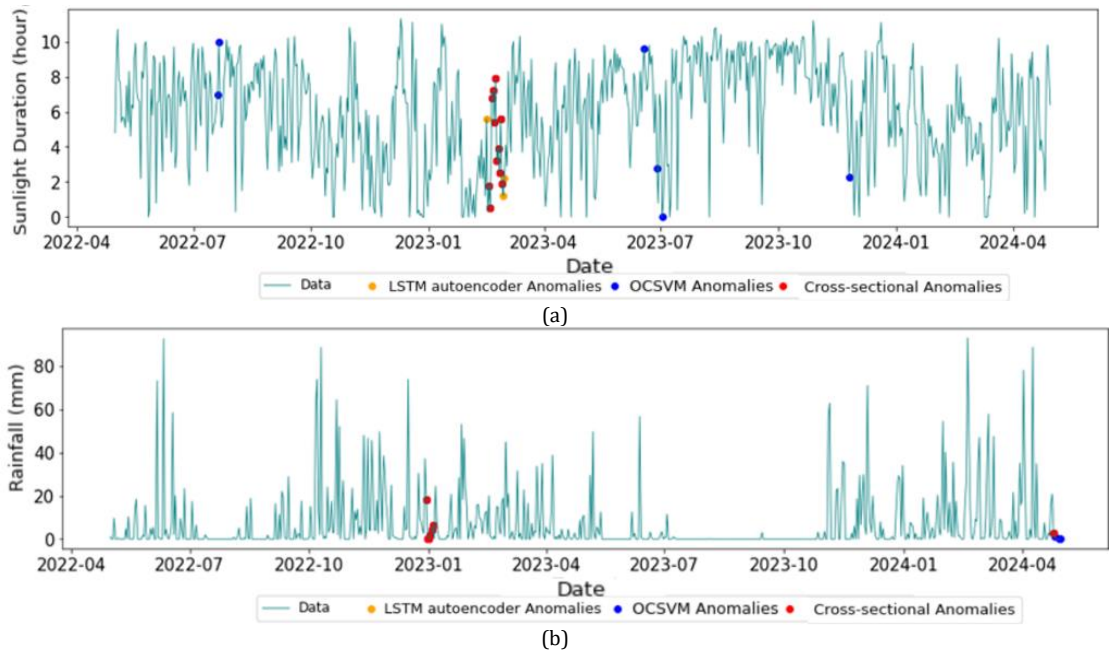


Figure 8. Time series plot including anomaly points from LSTM autoencoder, OCSVM, and their intersection on data: (a) sunlight duration and (b) rainfall

The final evaluation of anomaly detection using OCSVM (Figure 8) identified 17 anomalies in sunlight duration and 10 in rainfall. Within this subset, 11 sunlight anomalies and 7 rainfall anomalies were also detected by the LSTM autoencoder, demonstrating strong concordance between the methods. However, OCSVM detected 6 additional sunlight anomalies and 3 additional rainfall anomalies that the LSTM autoencoder missed. This highlights OCSVM's ability to identify subtle or isolated anomalies that may not generate large reconstruction errors in the LSTM autoencoder.

Table 9. Disasters in East Java Associated with Peak Monthly Anomalies in Rainfall and Sunlight Duration

Month	Disaster			
	Flood	Landslide	Extreme Weather	Total
January 2023	2	2	9	13
February 2023	9	14	11	34

The results show a clear temporal alignment between weather anomalies and disaster occurrences in East Java (Table 9). The detected anomalies are contextual, coinciding with increased disaster frequency. In January 2023, 4 rainfall anomalies were detected, coinciding with 13 reported disaster events, consisting of 2 floods, 2 landslides, and 9 extreme weather incidents. In February 2023, 11 sunlight duration anomalies were identified, during which 34 disaster events occurred, including 9 floods, 14 landslides, and 11 extreme weather events. These findings show the need for continuous weather monitoring to anticipate future risks.

Weather-related disasters significantly impact agriculture, particularly in regions highly dependent on climatic conditions. Floods and landslides damage farmland, causing crop failures and reducing soil fertility [17]. Extreme weather events, including heavy rainfall, strong winds, and heatwaves, disrupt plant growth and lower agricultural productivity[18]. These findings emphasize the importance of ongoing weather analysis to mitigate disaster impacts, ensuring food security and supporting farmers in East Java.

3.7. Discussion

This study was designed to address the challenge of identifying weather anomalies that can disrupt agricultural activities in East Java. Existing monitoring systems are often limited in anticipating irregular weather conditions early enough to support effective response. The proposed framework combines LSTM autoencoder and OCSVM to detect anomalies in two key variables—sunlight duration and rainfall—each tested independently.

The main contribution of this research lies in its two-stage anomaly detection mechanism. By using LSTM autoencoder to capture temporal deviations and refining the output with OCSVM, the model reduces the risk of missing critical anomalies, which is crucial in an agricultural context where delayed responses can cause major losses. Unlike fixed-threshold models, this approach adapts to underlying data patterns and is more resilient to noise.

The results showed that the majority of detected anomalies occurred in months where disaster reports increased significantly, indicating that this method is capable of highlighting irregular weather conditions that align with potential agricultural disruptions. Although not statistically tested for correlation, this temporal match strengthens the practical value of the findings.

Limitations of this study include the absence of spatially distributed weather data and the unavailability of labeled anomaly datasets for more robust validation. Future research should explore applying the model to additional climate variables such as temperature or humidity, deploying it across different geographic areas, and integrating it into real-time monitoring platforms for early warning systems.

4. CONCLUSION

This study addresses the growing need for accurate weather anomaly detection to support agricultural decision-making and reduce the risk of losses due to extreme weather in East Java. By employing a two-stage framework—LSTM autoencoder for reconstruction-based anomaly scoring and OCSVM for refined classification—this research offers a novel approach that enhances anomaly detection accuracy for univariate weather variables. The model achieved strong performance, with an F1-score of 0.71 for sunlight duration and 0.82 for rainfall. Notably, 11 sunlight anomalies were detected in February 2023 and 4 rainfall anomalies in January 2023, periods that also recorded high disaster frequencies. These findings highlight the model's capability to capture weather irregularities that are relevant for early warning systems and agricultural planning. The novelty of this research lies in its integration of deep learning and anomaly detection techniques to create a more adaptive and threshold-independent system. This approach is particularly useful for regions with limited labeled data and high seasonal variability. Future work should explore additional climate variables such as temperature or humidity, test the model's generalizability in other regions, and implement real-time deployment for operational use in agriculture and disaster risk management.

REFERENCES

[1] M. R. Kamal and M. A. Setiawan, "Deteksi Anomali dengan Security Information and Event Management (SIEM) Splunk pada Jaringan UII," *J. Ajang Unjuk Tugas Akhir oleh Mhs Inform.*, vol. 4, pp. 1–6, 2021.

[2] H. Shi, J. Guo, Y. Deng, and Z. Qin, "Machine learning-based anomaly detection of groundwater microdynamics: case study of Chengdu, China," *Sci. Rep.*, vol. 13, no. 1, pp. 1–19, 2023. doi: 10.1038/s41598-023-38447-5.

[3] Z. Z. Darban, G. I. Webb, S. Pan, C. Aggarwal, and M. Salehi, "Deep Learning for Time Series Anomaly Detection: A Survey," *ACM Comput. Surv.,* vol. 57, no. 1, 2024.

[4] D. T. Ha, N. X. Hoang, N. V. Hoang, N. H. Du, T. T. Huong, and K. P. Tran, "Explainable Anomaly Detection for Industrial Control System Cybersecurity," *IFAC-PapersOnLine.*, vol. 55, no.10, pp. 1183-1188, 2022.

[5] P. Mobtahej, X. Zhang, M. Hamidi, and J. Zhang, "An LSTM-Autoencoder Architecture for Anomaly Detection Applied on Compressors Audio Data," *Comput. Math. Methods.*, vol. 32, pp. 1–22, 2022. doi: 10.1155/2022/3622426.

[6] P. Saeipourdizaj, P. Sarbakhsh, and A. Gholampour," Application of Imputation Methods for Missing Values of PM10 and O3 data: Interpolation, Moving Average, and K-Nearest Neighbor Methods," *Environmental Health Engineering and Management Journal.*, vol. 8, no.3, pp. 215-226, 2021.

[7] Y. Wei, J. Jang-Jaccard, W. Xu, F. Sabrina, S. Camtepe, and M. Boulic, "LSTM-Autoencoder-Based Anomaly Detection for Indoor Air Quality Time-Series Data," *IEEE Sens. J.*, vol. 23, no. 4, pp. 3787–3800, 2023. doi: 10.1109/JSEN.2022.3230361.

[8] Z. Liu, Z. Xu, J. Jin, Z. Shen, and T. Darrell, "Dropout Reduces Underfitting," *Proceedings of Machine Learning Research,* Vol.202, pp. 22233-22248, 2023.

[9] Z. Farhadi, H. Bevrani, and M. R. Feizi-Derakhshi, "Combining Regularization and Dropout Techniques for Deep Convolutional Neural Network," in *Proc. IEEE Glob. Energy Conf. (GEC)*, pp. 335–339, 2022. doi: 10.1109/GEC55014.2022.9986657.

[10] M. K. Anam, S. Defit, Haviluddin, L. Efrizoni, M. B. Firdaus, "Early Stopping on CNN-LSTM Development to Improve Classification Performance," *Journal of Applied Data Sciences.*, vol. 5, no. 3, pp. 1175-1188, 2024.

[11] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021.

[12] L. B. V. Amorim, G. D. C. Cavalcanti, and R. M. O. Cruz, "The choice of scaling technique matters for classification performa nce," *Appl. Soft Comput.*, vol. 133, pp. 1–37, 2023. doi: 10.1016/j.asoc.2022.109924.

[13] S. Maleki, S. Maleki, N. R. Jennings, "Unsupervised Anomaly Detection with LSTM Autoencoders using Statistical Data-Filtering," *Applied Soft Computing.,* vol. 108, 2021.

[14] Y. Qiao, K. Wu, and P. Jin, "Efficient Anomaly Detection for High-Dimensional Sensing Data with One-Class Support Vector Machine," *IEEE Transactions on Knowledge and Data Engineering.,* vol. 35, no. 1, pp. 404-417, 2021.

[15] Ş. K. Çorbacıoğlu, and G. Aksel, "Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value," *Turkish journal of emergency medicine.*, vol. 23, no. 4, pp. 195-198, 2023.

[16] Z. Xu, "Factors Influencing the Predictive Performance of the LSTM Model on Stock Prices and Its Application in Forecasting US Technology Sector Stock Prices—Sample Time Span, Time Window Length, Feature Selection, and Prediction Days.," *In Proceedings of the International Conference on Image Processing, Machine Learning and Pattern Recognition.,* pp. 370-375, 2024.

[17] L. R. E. Malau, K. R. Rambe, N. A. Ulya, and A. G. Purba, "Dampak Perubahan Iklim terhadap Produksi Tanaman Pangan di Indonesia," *Jurnal Penelitian Pertanian Terapan.,* vol. 23, no. 1, pp. 34-36, 2024.

[18] D. Auliya, A. H. Rosandi, and W. T. Subroto, "Analisis Perubahan Iklim terhadap Produktivitas Padi di Jawa Timur," *Diponegoro Journal of Economics.*, vol. 13, no. 3, pp. 55-65, 2024. https://doi.org/10.14710/djoe.47595.

[19] F. Angiulli, F. Fassetti, and L. Ferragina, "Reconstruction Error-based Anomaly Detection with Few Outlying Examples," *arXiv preprint arXiv*:2305.10464, 2023. doi: 10.48550/arXiv.2305.10464.

[20] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geosci. Model Dev.*, vol. 15, pp. 5481–5487, 2022. Doi: https://doi.org/10.5194/gmd-15-5481-2022.