JUITA: Jurnal Informatika e-ISSN: 2579-8901; Vol. 13, Issue 2, July 2025

Simulation Study to Identify Factors Affecting the Performance of LSTM and XGBoost for Anomaly Detection on Labeled Time Series Data

Muhammad Rizky Nurhambali[1*], Yenni Angraini[2], Anwar Fitrianto[3]
[1,2,3] *Study Program of Statistics and Data Science, School of Data Science, Mathematics and Informatics, IPB University, Bogor, Indonesia*
*corr-author: rizkynurhambali@apps.ipb.ac.id

Abstract - Time series analysis has evolved to include forecasting and anomaly detection, which can be applied in various fields. Machine learning methods, such as long short-term memory (LSTM) and extreme gradient boosting (XGBoost), are widely developed because they are considered superior to conventional methods. Both use a forecasting approach for anomaly detection. However, the limitations of both methods on anomalies, such as data length, labeling method, and number of anomalies have not been explored. Therefore, this study aims to identify factors that affect the performance of LSTM and XGBoost in forecasting and anomaly detection through various scenarios and compare their metrics evaluation. The study utilizes Jakarta's air quality index data for 2018–2023, which was preprocessed and augmented for simulation purposes. The study shows that the LSTM method is superior to XGBoost, as shown by the lower MAPE (14.7024%), lower RMSE (13.9909), and higher balanced accuracy (0.9935). These results are reinforced by the significant Mann-Whitney test between the two methods, indicating a difference in the method's accuracy. In addition, the Kruskal-Wallis test for each combination of method and treatment showed significant results. These results indicate that data length, labeling method, and number of anomalies affect the method's accuracy.

Keywords: anomaly detection, forecasting, LSTM, time series, XGBoost

I. INTRODUCTION

Time series analysis is a subset of regression that considers the temporal aspects of the data. It is often associated with forecasting, and many methods have been developed, especially machine learning methods. Machine learning methods are widely developed because it is considered capable of processing large and complex data and increasing the model's accuracy compared to conventional methods. They are broadly used in time series forecasting include long short-term memory (LSTM) and extreme gradient boosting (XGBoost). Several studies have shown the superiority of both methods over conventional methods (ARIMA), such as the use of LSTM in the study of [1] on various financial data and [2] on sales data, as well as the use of XGBoost in [3] on human brucellosis in mainland China and [4] in the case of annual rice production in Bangladesh.

The use of time series analysis can be extended to detect anomalies in data. It is gaining attention due to its increasing application in various fields, such as environmental and urban management, medical risk, and natural disasters. For example, anomaly detection has been developed for heart impulse analysis based on electrocardiogram (ECG) to detect heart abnormalities [5]. In both ECG cases and cases in other fields, time series anomaly detection is also widely developed using machine learning and deep learning methods because they can learn expressive representations of complex time series [6]. These methods detect various anomalies, such as point, contextual, collective, and others.

One approach to time series anomaly detection is the use of forecasting methods. Forecasting methods are generally trained in semi-supervised, i.e., training data without anomalies is used for the learning process. Afterwards, the difference between the predicted and actual values is observed to determine the presence of an anomaly. The anomalies are values with significant differences. In this approach, LSTM and XGBoost methods are included. Research using LSTM and XGBoost for anomaly detection in time series has been widely used, such as the research of [7] with LSTM on Chinese earthquake precursor data and [8] with XGBoost on data for steam turbine health prognostics. Both methods demonstrated the ability to detect anomalies well.

Based on previous studies by [1-4,7-8], LSTM and XGBoost perform well in forecasting and anomaly detection. However, these studies are still limited in their application to empirical data and have not examined the characteristics of the method or the anomalies detected, even though LSTM and XGBoost are known to be robust to noise. Recent research, such as [9], is also limited to combining XGBoost and LSTM to improve the method's performance on Industrial Internet of Things (IIoT) imbalanced data with variable selection by XGBoost. Therefore, this study aims to examine and compare the robustness of LSTM and XGBoost under various conditions of labeled time series data. The evaluation is conducted systematically by considering robustness of both models unique combinations of data period length, labeling, and number of anomalies. This research provides a comparative analysis of the two models across various scenarios, addressing a significant gap in existing research and guiding model selection for practical applications.

II. METHOD

A. Data

The study utilized Jakarta's air quality index (AQI) data as the response variable obtained from AirNow (www.airnow.gov). Meanwhile, meteorological data in the form of humidity, wind, and temperature were obtained from NASA Power (https://power.larc.nasa.gov/data-access-viewer/) as explanatory variables. These meteorological variables were chosen because they are the main factors affecting the dispersion process, removal mechanism, and formation of atmospheric particles [10,11]. Both are hourly data from 2018-2023 and were utilized for simulation study. However, the simulation only focused on the AQI data by preprocessing and augmenting it with treatments. The detailed process is described in the following steps.

1) Data imputation: The AQI data has 2,281 missing values with varying lengths of missing data. Therefore, various imputation methods were used, including linear interpolation, seasonal decomposition, seasonal splitting, disaggregation, and forecasting with LSTM.

2) Anomaly labeling: Empirical data anomalies are labeled with a moving range (MR) using range magnitudes of 2 (MR (2)) and 3 (MR (3)). However, a modification is needed when dealing with the negative values of the changes in the AQI. This technique is similar to differencing in time series data. In addition, the 4-sigma rule is used as a threshold. Furthermore, the observed value labeled as an anomaly is replaced with the average observed value in that hour so that it is not an anomaly.

3) Randomized selection of observations and treatment augmentation: A total of 0.25% (132 observations), 0.50% (263 observations), 0.75% (395 observations), and 1.00% (526 observations) of the overall data were randomly selected. These are the values around the number of anomalies in the labeling process (0.50% or 264 observations of the overall data). Next, a 5-sigma distance is assigned to the selected observations, which becomes the lower anomaly if the observed difference is negative and the upper anomaly if the observed difference is positive. The observation is then labeled as an anomaly. Despite the addition of anomalies, data without anomalies (0% anomaly) was also used as a control treatment.

4) Repetition: Repeating steps (1)-(3) ten times to obtain ten different data sets.

B. Simulation Study Procedure

The simulation study was conducted following the procedure in Fig. 1 using R and Python software with the following detailed steps.

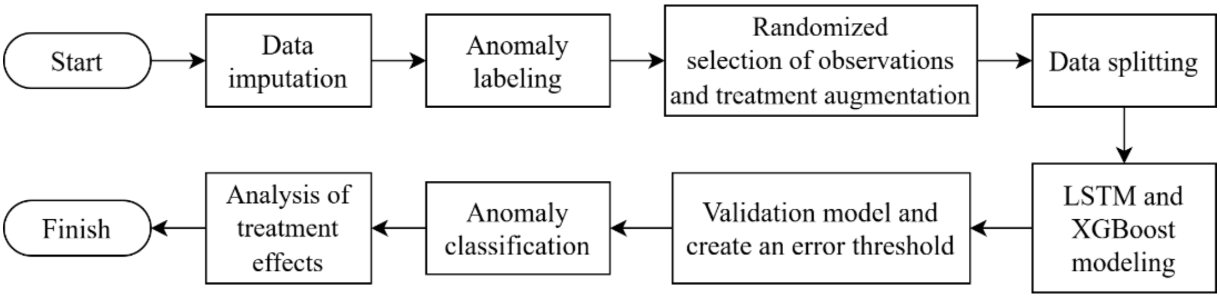

Fig. 1 Flowchart of the analysis procedure from data generation in one repetition

TABLE I
HYPERPARAMETER VALUES USED IN EACH METHOD

Method	Hyperparameter	Tested Value	Optimized Value
LSTM	Batch size	72	72
	Epoch	50	50
	Optimizer	Adam, rmsprop	Adam
	Learning rate	0.01, 0.001, 0.0001	0.001
XGBoost	Learning rate	-	0.2
	max_depth	-	4
	min_child_weight	-	7
	gamma	-	0
	colsample_bytree	-	0.5
	n_estimators	-	100

1) Data splitting: The data is split into yearly and half-yearly (semesterly), instead of the whole data. All parts of the data from each scenario will be used as training and test data. This aims to see the limitations of each method in forecasting. This categorization results in 60 combinations between the length of the data period (all data, annual, and semester), moving range (MR (2) and MR (3)), method (LSTM and XGBoost), and the number of anomalies in the data (0.00%, 0.25%, 0.50%, 0.75%, and 1.00%).

2) LSTM and XGBoost modeling: The modeling stage is carried out by setting the hyperparameters. The best hyperparameter is determined by using sliding window cross-validation and expanding window cross-validation as done by [12]. LSTM uses an architecture consisting of two main layers and a dense layer. The best LSTM hyperparameters used are listed in Table I. However, XGBoost did not find a better hyperparameter combination than the research [13] that modeled air quality with various machine learning approaches, namely with the hyperparameters in Table I.

3) Validating and calculation of the residuals: The residuals are used to calculate the method's accuracy with mean absolute percentage error (MAPE) and root mean square error (RMSE) in (1) and (2), where y_t is t-time value, $\hat{y}_t$ is t-time forecast value, and n is number of forecasting periods. Then, create an error threshold (T) based on the sigma rule using (3), where $\bar{e}$ is mean errors and σ is the standard deviation of errors.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100\% \tag{1}$$

$$RMSE = \sum_{t=1}^{n}\left(\frac{(y_t - \hat{y}_t)^2}{n}\right)^{\frac{1}{2}} \tag{2}$$

$$T = \bar{e} \pm 4\sigma \tag{3}$$

4) Anomaly classification: The data will be classified as an anomaly if it exceeds the threshold. In addition, the classification accuracy is calculated with balanced accuracy (BACC) following (4) where true positive (TP) is a positive class that is correctly classified as a positive class, true negative (TN) is a negative class that is correctly classified as a negative class, false positive (FP) is a negative class that is classified as a positive class, and false negative (FN) is a positive class that is classified as a negative class. BACC is used because of the data imbalance between anomaly and non-anomaly classes. According to [14], using the BACC metric will be more meaningful than the accuracy in imbalanced classes.

$$BACC = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \tag{4}$$

5) Analysis of treatment effects: Non-parametric tests with Kruskal-Wallis and Mann-Whitney were used to examine the effect between treatment groups. The Kruskal-Wallis test looks at differences for three or more groups, while the Mann-Whitney test looks at differences between two groups. The null hypothesis is that the groups have the same median

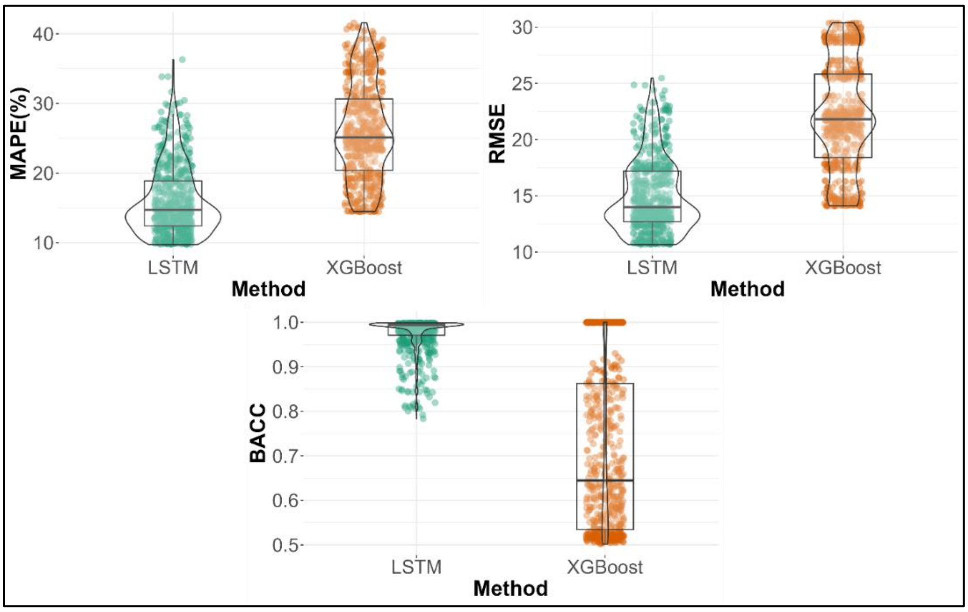

Fig. 2 Comparison of metric evaluation values between methods with boxplot and violin plot (lower MAPE, lower RMSE, and higher BACC indicate better performance)

III. RESULT AND DISCUSSION

A. *Performance Evaluation Method*

LSTM and XGBoost are anomaly detection methods with a forecasting approach. Forecasting is a validation process to find the difference between the predicted data and the actual data used in anomaly detection. Therefore, before looking at the anomaly detection performance, both models' forecasting performance is also part of the study. The performance of LSTM and XGBoost for anomaly detection, especially on AQI data, is studied using various scenarios.

Fig. 2 shows that LSTM generally performs better than XGBoost in validation and anomaly detection. This is indicated by the lower position of the LSTM evaluation metric plot for validation metrics and higher for classification metrics. Many of the forecasting metric values provided by XGBoost are greater than the maximum metric values provided by LSTM. However, the classification metric value is the opposite; the classification metric value of LSTM is never lower than XGBoost. This shows that the accuracy of LSTM is better than that of XGBoost. In addition, the violin diagram in the boxplot shows that the RMSE metric value of XGBoost can be more easily clustered than LSTM. This outcome may be influenced by unconsidered factors.

Table II further reinforces the superiority of LSTM, where the MAPE validation metric value of LSTM can be classified as good forecasting, while XGBoost is classified as fair forecasting. This illustrates the advantages of the LSTM method over XGBoost in capturing patterns in the data. In addition, with the Mann-Whitney formal test, a p-value of 0 is obtained, so there is a significant difference in the performance of the two methods. The sizable difference in metric values indicates that LSTM is superior to XGBoost. All these results are in line with the research of [15] and [16] which showed the superiority of LSTM over XGBoost in terms of forecasting with Chinese stock market data and Covid-19 transmission, as well as [17] research which showed the superiority of LSTM over XGBoost in terms of anomaly detection in internet of things (IoT) devices. However, the comparison of these methods ignores the factors of data length, moving range, and the number of anomalies in the data, so that not much can be explained, even though based on Table III, it appears that there is an influence of other factors. Therefore, the evaluation of the influence of each factor is discussed further in the other subsections.

TABLE II
COMPARISON OF MEDIAN METRIC VALUES BETWEEN METHODS

Methods	**Validation Metrics**		**Classification Metrics**
	MAPE (%)	**RMSE**	**BACC**
LSTM	14,7024	13,9909	0,9935
XGBoost	25,1152	21,8074	0,6439

TABLE III
NON-PARAMETRIC TEST RESULTS OF METRIC EVALUATION VALUES AGAINST TREATMENT FACTORS

Validation Metrics						
Factor	**Test**	**Degree of Freedom**	**MAPE**		**RMSE**	
			P-value	**Conclusion**	**P-value**	**Conclusion**
Methods	Mann-Whitney	-	0,0000	Significant	0,0000	Significant
Data length	Kruskal-Wallis	2	0,0000	Significant	0,0000	Significant
Number of anomalies	Kruskal-Wallis	4	0,0000	Significant	0,0000	Significant
Combination of methods & data length	Kruskal-Wallis	5	0,0000	Significant	0,0000	Significant
Combination of methods & number of anomalies	Kruskal-Wallis	9	0,0000	Significant	0,0000	Significant
Combination of methods, data length, number of anomalies	Kruskal-Wallis	29	0,0000	Significant	0,0000	Significant

Classification Metric (BACC)				
Factor	**Test**	**Degree of Freedom**	**P-value**	**Conclusion**
Methods	Mann-Whitney	-	0,0000	Significant
Data length	Kruskal-Wallis	2	0,6131	Not significant
Moving range	Mann-Whitney	-	0,0000	Significant
Number of anomalies	Kruskal-Wallis	4	0,0000	Significant
Combination of methods & data length	Kruskal-Wallis	5	0,0000	Significant
Combination of methods & moving range	Kruskal-Wallis	3	0,0000	Significant
Combination of methods & number of anomalies	Kruskal-Wallis	9	0,0000	Significant
Combination of methods, data length, moving range, number of anomalies	Kruskal-Wallis	59	0,0000	Significant

Note: α = 0.05 is used as the significance level

B. Effect of Data Length on Method Performance

Fig. 3 shows the effect of data length on the methods, especially for XGBoost. XGBoost has better accuracy when used in shorter data, while LSTM has better accuracy when used in longer data. However, the accuracy of LSTM remains stable with shorter data and is never worse than that of XGBoost. These contradictory results indicate an effect between the method and data length as shown in Table III with a significant Kruskal-Wallis test.

In forecasting, LSTM works better on long-period data because they have memory cells and gate mechanisms to process information. The architecture allows LSTM to filter out relevant information and delete those that are not. However, in XGBoost, this architecture does not exist, so the ability to store long-term information is also different. According to [18], the advantage of LSTM comes from its ability to capture sequential information, while the advantage of XGBoost comes from ensemble learning. The ensemble learning in XGBoost utilizes time-related features to replace sequential information. The lack of time-related features may be why XGBoost does not perform better than LSTM.

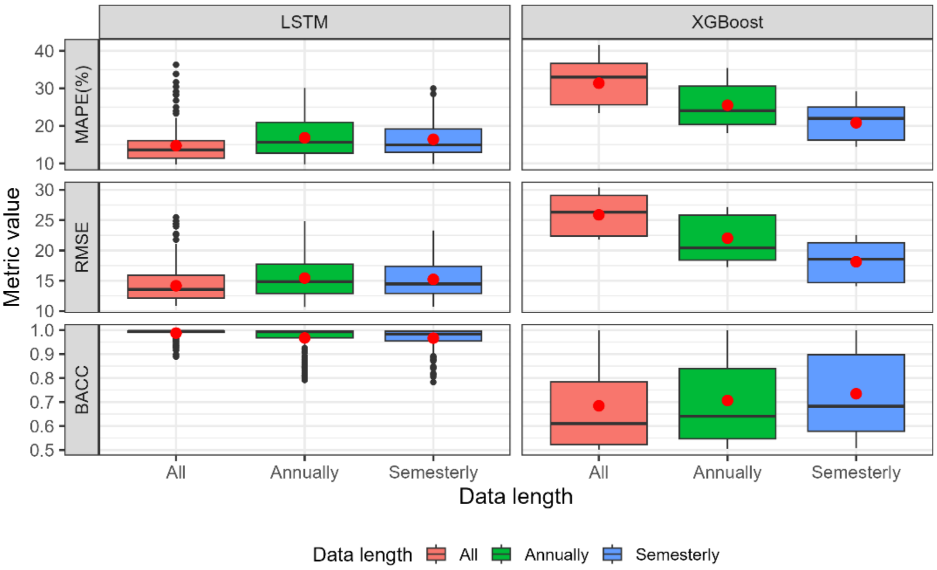

Fig. 3 Comparison of metric evaluation values between combinations of methods across varying data lengths (lower MAPE, lower RMSE, and higher BACC indicate better performance)

C. Effect of Moving Range on Method Performance

Moving range is used in the labeling process, so it will only affect the classification process of anomaly detection results. The difference in moving range affects the performance of both methods, as shown in Fig. 4. This result is reinforced by Table III, which shows a significant effect between the method and MR. LSTM is a more suitable method with MR labeling than XGBoost. Both methods perform better with MR (3). However, the difference in accuracy between the two methods remains high.

The use of MR (3) in the simulation can indicate anomalous observations that are more appropriate to be replaced due to changes in mean or variance, as in [19]. In addition, according to [20], if the data has a seasonal (cyclic) pattern, the use of MR is not recommended because it causes observations outside the boundaries difficult to distinguish during peak times in the cycle. Therefore, using MR in actual anomaly labeling for anomaly detection still needs further study.

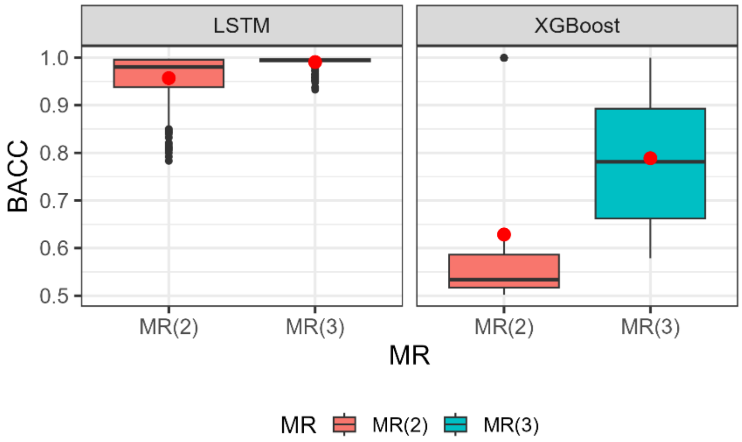

Fig. 4 Comparison of classification metric values between combinations of methods across MR(2) and MR(3) (higher BACC indicate better performance)

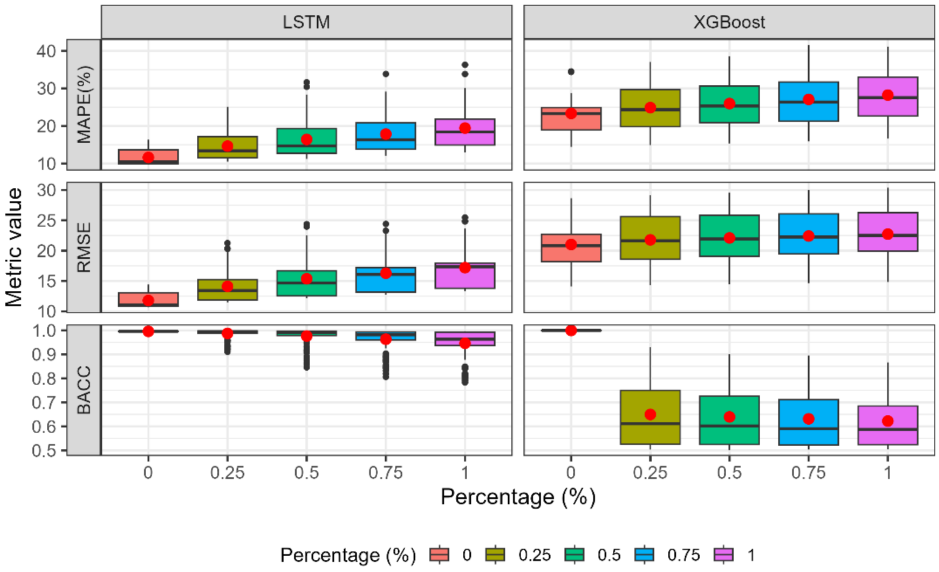

Fig. 5 Comparison of metric evaluation values between combinations of methods across varying number of anomalies (lower MAPE, lower RMSE, and higher BACC indicate better performance)

D. Effect of Number of Anomalies on Method Performance

LSTM and XGBoost are two methods that have the advantage of being robust to noise [21,22]. However, both methods will certainly have limitations in accepting such noise. Fig. 5 shows the effect of the number of anomalies on both methods, where the method's accuracy worsens as the number of anomalies in the data increases (0% to 1%). Nevertheless, the accuracy of LSTM is still better than that of XGBoost. The accuracy of LSTM does not immediately decrease drastically with the increase in the number of anomalies, unlike XGBoost, which decreases drastically, especially during the anomaly detection process. The larger anomaly value in the data may cause the model to capture the condition as a normal condition, so the model works better on clean data from anomalies. In addition, the Kruskal-Wallis test results in Table III reinforce that there is a significant influence between the number of anomalies and the method's accuracy, both validation and anomaly detection. This result is in line with the research of [23], with LSTM on network data showing the influence of anomalies resulting in larger forecasting errors between predicted and actual values. Although using different evaluation metrics, the classification metric evaluation results also align with the research of [24], who used hierarchical clustering and LSTM on California vehicle traffic and air pollution data.

E. Comparison of Anomaly Detection Results

LSTM and XGBoost anomaly detection are forecasting approaches, so anomaly detection depends on the value of the residual. The residual value is a change in the extreme value labeled as an anomaly with a point anomaly type if it crosses the threshold, as shown in Fig. 6c and Fig. 6d. The anomaly label results are then matched with the actual values, so that in Fig. 6a and Fig. 6b, there are anomalies in the middle of the data value range. The difference in the definition and approach of anomalies used makes these results different from the research of [25]. Ref [25] mentioned that methods that are robust to noise, such as LSTM, will make point anomaly detection difficult.

Furthermore, the number of anomalies detected as shown in Fig. 6c and Fig. 6d indicate the difference in sensitivity of the methods. This difference is not only due to the method's ability to forecast, but is also influenced by the threshold of the residuals generated by the two methods during validation. LSTM produces a lower validation error because it can capture the data pattern better than XGBoost. The low error narrows the interval and makes the method more sensitive so that the method will detect more anomalies. This also applies vice versa, as shown in the XGBoost results.

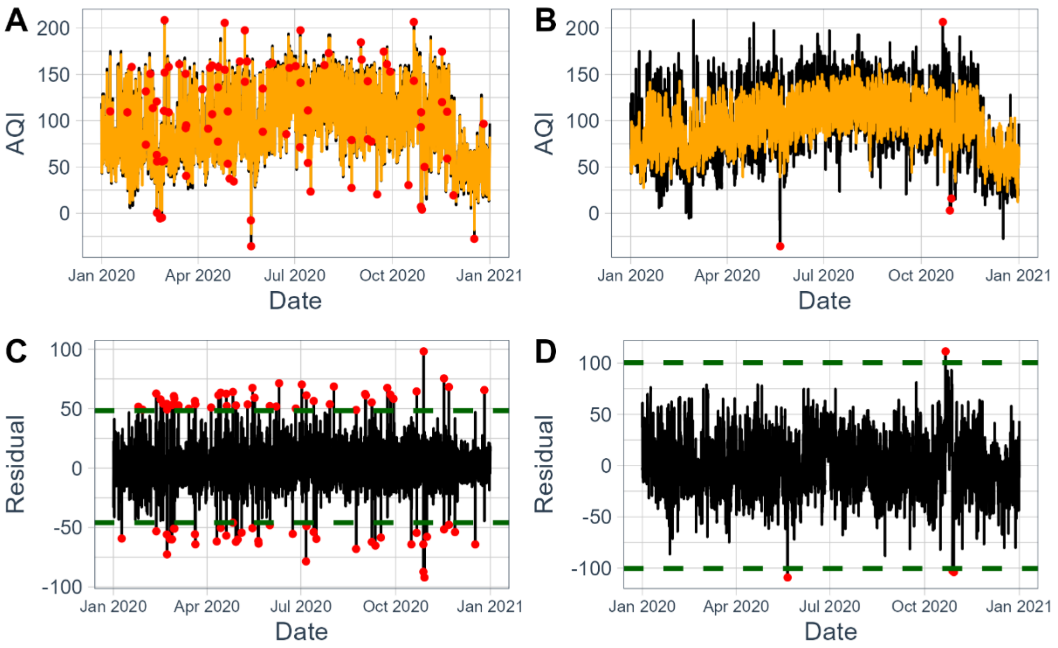

Fig. 6 Examples of anomaly detection results (red dots) and residuals of LSTM (A, C) and XGBoost (B, D) for 2020 data with 0.5% anomaly treatment

IV. CONCLUSION

Simulation studies show that LSTM performs better in validation and anomaly detection than XGBoost for each treatment condition. In addition, other factors such as data length, number of anomalies, and combinations between factors also affect the method's accuracy. The limitation of this study is limited to generalizability because it only uses one kind of dataset, only tries one kind of anomaly type, and uses a fixed hyperparameter. According to [25], the application to various data sets and the determination of hyperparameters can deepen the characteristics of the method in anomaly detection and make future research more comprehensive. In addition, future research can consider other labeling approaches, such as the use of different range values in MR (>3) because it is intuitively better at detecting diversity changes [19], labeling by rank [26], distance measure [27,28], and by probability [29,30].

ACKNOWLEDGEMENT

The authors gratefully acknowledge the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia, which provided financial support for this research with contract number 027/E5/PG.02.00.PL/2024.

REFERENCES

[1] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, “A Comparison of ARIMA and LSTM in Forecasting Time Series,” in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL: IEEE, Dec. 2018, pp. 1394–1401. doi: 10.1109/ICMLA.2018.00227.

[2] U. M. Sirisha, M. C. Belavagi, and G. Attigeri, “Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison,” *IEEE Access*, vol. 10, pp. 124715–124727, 2022, doi: 10.1109/ACCESS.2022.3224938.

[3] M. Alim, G. H. Ye, P. Guan, D. S. Huang, B. Sen Zhou, and W. Wu, “Comparison of ARIMA model and XGBoost model for prediction of human brucellosis in mainland China: a time-series study,” *BMJ Open*, vol. 10, no. 12, p. e039676, Dec. 2020, doi: 10.1136/BMJOPEN-2020-039676.

[4] M. Noorunnahar, A. H. Chowdhury, and F. A. Mila, “A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh,” *PLOS ONE*, vol. 18, no. 3, p. e0283452, Mar. 2023, doi: 10.1371/JOURNAL.PONE.0283452.

[5] E. J. da S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, “ECG-based heartbeat classification for arrhythmia detection: A survey,” *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 144–164, Apr. 2016, doi: 10.1016/J.CMPB.2015.12.008.

[6] Z. Z. Darban, G. I. Webb, S. Pan, C. Aggarwal, and M. Salehi, “Deep Learning for Time Series Anomaly Detection: A Survey,” *ACM Computing Surveys*, vol. 57, p. 42, Jan. 2024, doi: 10.1145/3691338.

[7] Y. Cai, M. L. Shyu, Y. X. Tu, Y. T. Teng, and X. X. Hu, “Anomaly detection of earthquake precursor data using long short-term memory networks,” *Applied Geophysics*, vol. 16, no. 3, pp. 257–266, Sep. 2019, doi: 10.1007/S11770-019-0774-1/METRICS.

[8] Z. Que and Z. Xu, “A Data-Driven Health Prognostics Approach for Steam Turbines Based on Xgboost and DTW,” *IEEE Access*, vol. 7, pp. 93131–93138, 2019, doi: 10.1109/ACCESS.2019.2927488.

[9] Z. Chen, Z. W. Li, J. Huang, S. Z. Liu, and H. X. Long, “An effective method for anomaly detection in industrial Internet of Things using XGBoost and LSTM,” *Scientific reports*, vol. 14, no. 1, p. 23969, Dec. 2024, doi: 10.1038/s41598-024-74822-6.

[10] S. K. Goyal and C. V. C. Rao, “Assessment of atmospheric assimilation potential for industrial development in an urban environment: Kochi (India),” *Science of The Total Environment*, vol. 376, no. 1–3, pp. 27–39, Apr. 2007, doi: 10.1016/J.SCITOTENV.2007.01.067.

[11] H. Zhang, Y. Wang, J. Hu, Q. Ying, and X. M. Hu, “Relationships between meteorological parameters and criteria air pollutants in three megacities in China,” *Environmental Research*, vol. 140, pp. 242–254, Jul. 2015, doi: 10.1016/J.ENVRES.2015.04.004.

[12] M. R. Nurhambali, Y. Angraini, and A. Fitrianto, “Implementation of Long Short-Term Memory for Gold Prices Forecasting,” *Malaysian Journal of Mathematical Sciences*, vol. 18, no. 2, pp. 399–422, 2024, doi: 10.47836/mjms.18.2.11.

[13] M. A. Haq, “SMOTEDNN: A Novel Model for Air Pollution Forecasting and AQI Classification,” *Computers, Materials & Continua*, vol. 71, no. 1, pp. 1403–1425, Nov. 2021, doi: 10.32604/CMC.2022.021968.

[14] S. Bej, N. Davtyan, M. Wolfien, M. Nassar, and O. Wolkenhauer, “LoRAS: an oversampling approach for imbalanced datasets,” *Machine Learning*, vol. 110, no. 2, pp. 279–301, Feb. 2021, doi: 10.1007/S10994-020-05913-4/FIGURES/2.

[15] Y. Zhu, “Stock Price Prediction based on LSTM and XGBoost Combination Model,” *Transactions on Computer Science and Intelligent Systems Research*, vol. 1, pp. 94–109, Oct. 2023, doi: 10.62051/Z6DERE47.

[16] J. Luo, Z. Zhang, Y. Fu, and F. Rao, “Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms,” *Results in Physics*, vol. 27, p. 104462, Aug. 2021, doi: 10.1016/J.RINP.2021.104462.

[17] X. Wang and X. Lu, “A host-based anomaly detection framework using XGBoost and LSTM for IoT devices,” *Wireless Communications and Mobile Computing*, vol. 2020, 2020, doi: 10.1155/2020/8838571.

[18] Z. Wang, T. Hong, and M. A. Piette, “Building thermal load prediction through shallow machine learning and deep learning,” *Applied Energy*, vol. 263, p. 114683, Apr. 2020, doi: 10.1016/J.APENERGY.2020.114683.

[19] S. E. Rigdon, E. N. Cruthis, and C. W. Champ, “Design Strategies for Individuals and Moving Range Control Charts,” *Journal of Quality Technology*, vol. 26, no. 4, pp. 274–287, 1994, doi: 10.1080/00224065.1994.11979539.

[20] P. M. Berthouex, “Constructing Control Charts for Wastewater Treatment Plant Operation,” *Research Journal of the Water Pollution Control Federation*, vol. 61, no. 9, pp. 1534–1551, 1989.

[21] O. M. Osama, K. Alakkari, M. Abotaleb, and E. S. M. El-Kenawy, “Forecasting Global Monkeypox Infections Using LSTM: A Non-Stationary Time Series Analysis,” in *ICEEM 2023 - 3rd IEEE International Conference on Electronic Engineering*, Menouf: Institute of Electrical and Electronics Engineers Inc., Oct. 2023. doi: 10.1109/ICEEM58740.2023.10319532.

[22] P. H. Vuong, T. T. Dat, T. K. Mai, P. H. Uyen, and P. T. Bao, “Stock-Price Forecasting Based on XGBoost and LSTM,” *Computer Systems Science and Engineering*, vol. 40, no. 1, pp. 237–246, Aug. 2021, doi: 10.32604/CSSE.2022.017685.

[23] X. Yidan, H. Shaolin, and Y. Guotao, “Analysis and Improvement Approach of the Impact of Data Disturbance on LSTM Prediction Algorithm,” *Transactions on Engineering and Computing Sciences*, vol. 11, no. 5, pp. 1–15, Sep. 2023, doi: 10.14738/TECS.115.15411.

[24] R. M. Shukla and S. Sengupta, “Scalable and Robust Outlier Detector using Hierarchical Clustering and Long Short-Term Memory (LSTM) Neural Network for the Internet of Things,” *Internet of Things*, vol. 9, p. 100167, Mar. 2020, doi: 10.1016/J.IOT.2020.100167.

[25] S. Schmidl, P. Wenig, and T. Papenbrock, “Anomaly detection in time series,” *Proceedings of the VLDB Endowment*, vol. 15, no. 9, pp. 1779–1797, 2022, doi: 10.14778/3538598.3538602.

[26]H. Huang, “Rank Based Anomaly Detection Algorithms,” Syracuse University, 2013.

[27] K. G. Mehrotra, C. K. Mohan, and H. Huang, *Anomaly Detection Principles and Algorithms*. in Terrorism, Security, and Computation. Cham: Springer International Publishing, 2017. doi: 10.1007/978-3-319-67526-8.

[28] W. Wu, “Developing an Unsupervised Real-time Anomaly Detection Scheme for Time Series with Multi-seasonality,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 9, pp. 4147–4160, Aug. 2019, doi: 10.1109/TKDE.2020.3035685.

[29] A. M. Committee, “Robust statistics–how not to reject

outliers. Part 1. Basic concepts," *Analyst*, vol. 114, no. 12, pp. 1693–1697, Jan. 1989, doi: 10.1039/AN9891401693.

[30] T. Iwata, M. Toyoda, S. Tora, and N. Ueda, "Anomaly detection with inexact labels," *Machine Learning*, vol. 109, no. 8, pp. 1617–1633, Aug. 2020, doi: 10.1007/S10994-020-05880-W/TABLES/4.