Indonesian Journal of Electrical Engineering and Computer Science
Vol. 43, No. 1, July 2026, pp. 345~354
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v43.i1.pp345-354

Hybrid machine learning framework for anomaly detection in industrial IoT environments

I Dewa Made Widia[1], Toni Anwar[2]
[1]Department of Digital and Creative Industries, Faculty of Vocational Studies, Universitas Brawijaya, Malang, Indonesia
[2]Department of Computer and Information Science, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia

Article Info

Article history:

Received Jan 23, 2026
Revised Jun 3, 2026
Accepted Jun 27, 2026

Keywords:

Anomaly detection
IDS
IIoT
Machine learning
Weighted fusion

ABSTRACT

The industrial internet of things (IIoT) has become a core component of Industry 4.0, enabling highly connected and data-driven industrial systems while simultaneously increasing exposure to cyber threats. Conventional intrusion detection systems (IDS), especially rule-based and signature-driven approaches, often struggle to cope with the dynamic, high-dimensional, and heterogeneous nature of IIoT traffic. This study proposes a hybrid anomaly detection framework that integrates autoencoder, isolation forest, and long short-term memory (LSTM) models using a weighted decision fusion strategy. Each component contributes complementary capabilities, including nonlinear feature learning, efficient outlier detection, and temporal pattern modeling. The framework is evaluated on the botnet of things (BoT-IoT) dataset and further validated using IoT-23. Experimental results show that the proposed hybrid approach achieves a precision of 0.999, recall of 0.970, and an F1-score of 0.985, while maintaining a false-negative rate below 0.001%. Although its area under the curve (AUC) is slightly lower than that of a standalone light gradient boosting machine (LightGBM) baseline, the hybrid framework consistently reduces missed detections, making it well suited for reliable real-time IIoT security monitoring.

Corresponding Author:

I Dewa Made Widia
Department of Digital and Creative Industries, Faculty of Vocational Studies, Universitas Brawijaya
12–14 Veteran Street, Malang, East Java 65145, Indonesia
Email: dewa_vokasi@ub.ac.id

1. INTRODUCTION

The industrial internet of things (IIoT) has become a key component of Industry 4.0, integrating sensing technologies, large-scale connectivity, and intelligent data processing to improve automation and operational efficiency across manufacturing, logistics, energy management, and industrial control systems [1]-[3]. While these advancements provide significant benefits, they also expand the cyber-attack surface of infrastructures, making cybersecurity a critical concern.

Unlike traditional enterprise networks, IIoT environments require high levels of reliability, real-time performance, and operational safety. Security breaches may not only compromise data but also disrupt physical processes and industrial operations. Moreover, heterogeneous communication protocols and resource-constrained devices limit the effectiveness of conventional security mechanisms, highlighting the need for adaptive intrusion detection solutions [2], [3]. Machine learning-based intrusion detection systems (IDS) have been widely explored to address these challenges. Classical approaches, including support vector machines (SVM), random forests (RF), and k-nearest neighbors (KNN), have shown promising performance in anomaly detection. However, they often struggle with high-dimensional data and temporal dependencies associated with sophisticated cyber-attacks [3]-[5].

To overcome these limitations, deep learning techniques such as autoencoders and long short-term memory (LSTM) networks have been increasingly adopted. Autoencoders detect anomalies through reconstruction errors, while LSTM models capture sequential attack patterns and temporal dependencies [6], [7]. Nevertheless, single-model approaches remain sensitive to parameter settings, computationally demanding, and less robust when applied to complex and highly imbalanced IIoT datasets [4], [5].

Recent studies have therefore explored hybrid intrusion detection frameworks that combine multiple learning paradigms. Autoencoders can capture nonlinear traffic characteristics, isolation forest identifies statistical outliers, and LSTM models learn temporal behaviors. Despite their advantages, many hybrid approaches rely on static fusion strategies and focus primarily on overall accuracy, while the reduction of false negatives remains relatively underexplored in highly imbalanced IIoT environments [8]-[10].

Recent advances in convolutional neural network (CNN)-LSTM architectures, transformer-based models, and ensemble learning techniques have further improved anomaly detection capability in complex IIoT scenarios [9]-[12]. In addition, benchmark datasets such as botnet of things (BoT-IoT) and IoT-23 have enabled more realistic evaluation of IDS. However, achieving reliable detection with minimal missed attacks remains a significant challenge [13], [14]. Furthermore, recent studies have also investigated optimized isolation forest models to improve intrusion detection performance in heterogeneous IIoT environments [15].

To address this issue, this study proposes a hybrid anomaly detection framework that integrates autoencoder, isolation forest, and LSTM models through a weighted decision fusion strategy [6]–[8]. Rather than focusing solely on maximizing metrics such as area under the curve (AUC), the proposed approach prioritizes detection reliability by reducing false negatives while maintaining balanced performance. The framework also incorporates feature normalization, preprocessing, and class imbalance handling to support stable model performance.

Experimental evaluation on the BoT-IoT dataset, with additional validation using IoT-23 [13], [14] demonstrates that the proposed framework achieves a favorable balance between detection accuracy and reliability. By combining complementary learning paradigms, the framework provides improved robustness for anomaly detection in challenging IIoT environments.

The main contributions of this study are as follows: i) a hybrid intrusion detection framework integrating autoencoder, isolation forest, and LSTM through a weighted fusion strategy; ii) a detection approach that explicitly prioritizes false-negative reduction in highly imbalanced IIoT environments; iii) a comprehensive evaluation using benchmark datasets, demonstrating improved detection reliability and robustness compared with standalone models. The remainder of this paper is organized as follows. Section 2 reviews related work, section 3 presents the proposed methodology, section 4 discusses the experimental results, and section 5 concludes the paper and outlines future research directions.

2. RELATED WORK

The increasing adoption of IIoT technologies has driven significant research on IDS for protecting industrial networks against cyber threats. Traditional signature-based approaches often struggle to cope with the dynamic, heterogeneous, and evolving nature of IIoT traffic. As a result, machine learning and deep learning techniques have become widely adopted for anomaly detection [2], [3].

Early studies primarily employed classical machine learning algorithms such as SVM, RF, and KNN [3], [5]. While these approaches are computationally efficient and relatively interpretable, their performance often depends on extensive feature engineering and may be limited when modeling complex nonlinear relationships and temporal dependencies in industrial traffic [4], [5]. To address these limitations, deep learning approaches have been increasingly explored [4]. Autoencoder-based methods learn compact representations of normal traffic and detect anomalies through reconstruction errors, whereas LSTM networks effectively capture temporal dependencies and sequential attack behaviors [6], [7]. Isolation forest has also gained attention due to its ability to detect anomalies without requiring labeled attack data [8]. However, these individual approaches often exhibit limitations related to computational complexity, parameter sensitivity, or restricted detection capability under highly imbalanced traffic conditions [15].

Table 1 summarizes representative intrusion detection approaches proposed for IIoT environments. The selected studies cover machine learning, deep learning, and hybrid frameworks, highlighting their strengths and limitations [9], [10]. This comparison provides a basis for identifying current research gaps and positioning the proposed framework within the existing literature [16], [17]. As shown in Table 1, recent studies increasingly employ hybrid intrusion detection frameworks to improve detection robustness by combining multiple learning paradigms. Although these approaches generally outperform single-model solutions, many still rely on static fusion mechanisms such as majority voting or simple averaging, which implicitly assume equal reliability among component models. This assumption may be unsuitable for highly imbalanced IIoT traffic, where different models contribute differently to anomaly detection performance.

Furthermore, most existing studies focus primarily on overall accuracy or AUC, while explicit false-negative reduction receives comparatively less attention. In industrial environments, missed attacks can result in operational disruption, safety risks, and financial losses, making detection reliability a critical design objective.

To address these limitations, this study proposes a hybrid intrusion detection framework that integrates autoencoder, isolation forest, and LSTM models through a weighted decision fusion mechanism [18]–[20]. Unlike conventional ensemble approaches, the proposed framework assigns different weights according to the anomaly detection capability of each model, thereby improving detection reliability under highly imbalanced IIoT conditions. Light gradient boosting machine (LightGBM) is employed as a baseline model for comparative evaluation. By emphasizing false-negative reduction and leveraging complementary detection paradigms, the proposed framework aims to bridge the gap between detection accuracy and operational reliability in real-world IIoT environments.

Table 1. Summary of related work on IIoT IDS

References	Method	Strength	Limitation
[6]	Autoencoder	Detects unseen anomalies	Weak temporal modeling
[7]	LSTM	Captures temporal patterns	High computational cost
[6], [8]	Autoencoder + isolation forest	Structural + outlier detection	Static fusion
[5], [15]	Ensemble machine learning	Improved robustness	Equal-weight assumption
[4], [5]	Hybrid machine learning IDS	Higher detection accuracy	Limited false-negative focus
[9]	CNN–LSTM	Spatial-temporal learning	Computationally intensive
[11]	Transformer IDS	Long-range dependency learning	Large training requirements
This work	Autoencoder + isolation forest + LSTM (weighted fusion)	False-negative reduction and detection reliability	Slightly lower AUC than LightGBM

3. METHOD

This section presents the architecture and operational workflow of the proposed hybrid IDS for IIoT environments. The framework integrates multiple learning paradigms to address the limitations of single-model intrusion detection approaches and improve detection reliability under highly imbalanced network traffic conditions. The overall detection process consists of five main stages: data acquisition; data preprocessing and feature transformation; model-specific anomaly detection; weighted decision fusion; performance evaluation [16], [21].

Figure 1 illustrates the overall architecture of the proposed framework. Network traffic data are first preprocessed through feature selection, categorical encoding, and feature normalization. The processed data are subsequently analyzed by multiple anomaly detection models operating in parallel. Finally, the outputs generated by the individual models are aggregated through a weighted decision fusion mechanism to produce the final classification result. By combining reconstruction-based learning, statistical anomaly detection, and temporal sequence modeling, the framework aims to improve robustness and detection reliability in heterogeneous IIoT environments [11], [15], [21].

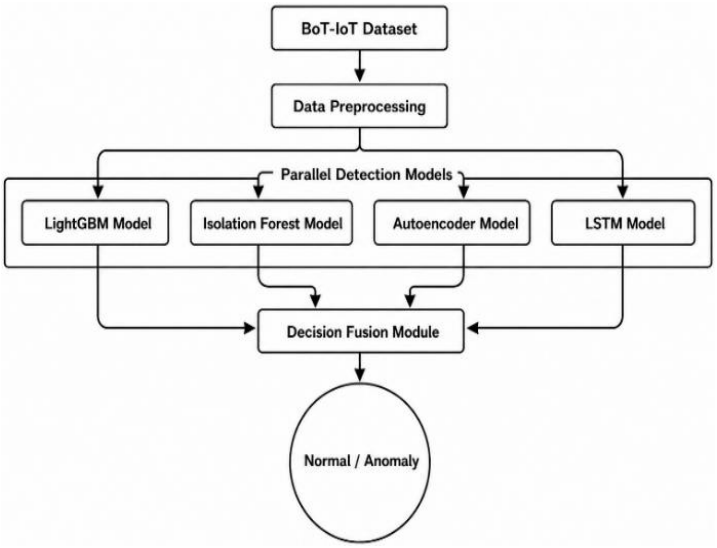

Figure 1. Hybrid IDS architecture

3.1. Dataset description

The proposed framework is evaluated using two publicly available IIoT datasets: BoT-IoT and IoT-23. BoT-IoT contains a wide range of attack scenarios, including denial-of-service, probing, and information theft activities, making it suitable for large-scale intrusion detection research [13], [14]. To evaluate the generalization capability of the proposed framework, additional validation is performed using the IoT-23 dataset, which includes real-world IoT malware traffic and benign network behavior. Due to the large size of the BoT-IoT dataset, a representative sample containing approximately 2.2 million records was generated using probabilistic sampling with a 3% inclusion rate. Both datasets exhibit significant class imbalance, reflecting realistic industrial network conditions.

Figure 2 presents the class distribution of the sampled BoT-IoT dataset. Distributed denial of service (DDoS) and denial of service (DoS) traffic dominate the dataset, whereas normal traffic and theft-related attacks represent only a small proportion of the total samples. A logarithmic scale is used to improve visualization of the large differences in class frequencies. Such imbalance poses a significant challenge for anomaly detection models and motivates the use of specialized preprocessing and learning strategies.

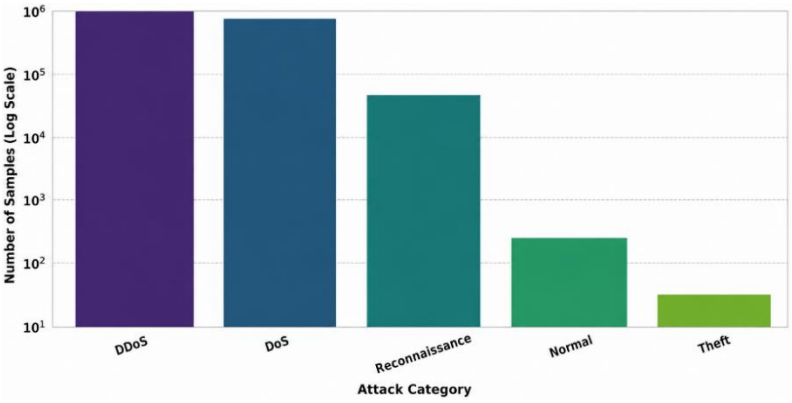

Figure 2. Class distribution of attack categories in the BoT-IoT dataset (log scale)

3.2. Dataset processing

To prepare the data for model training, a structured preprocessing pipeline is applied as illustrated in Figure 3. The preprocessing process includes data cleaning, feature selection, categorical encoding, and numerical scaling. High-cardinality identifiers such as IP addresses, MAC addresses, and port numbers are removed to improve model generalization and prevent memorization of specific traffic instances. Missing values are replaced with zero to ensure data consistency and compatibility across all models.

Categorical attributes, including protocol and connection-state information, are transformed using one-hot encoding. Numerical features are subsequently standardized using StandardScaler, which normalizes feature distributions by removing the mean and scaling to unit variance. This approach was selected because it is less sensitive to outliers and improves model stability when handling highly imbalanced network traffic. The resulting processed dataset serves as input for the anomaly detection models described in the following section.

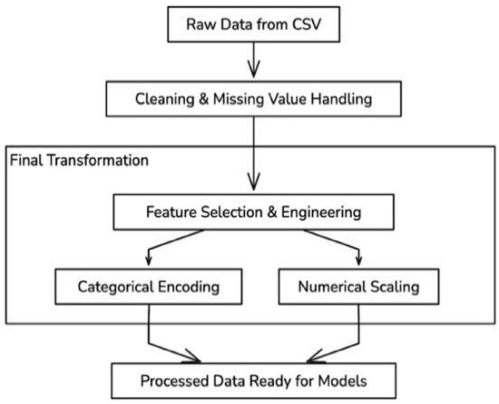

Figure 3. Data preprocessing pipeline for IIoT network traffic

3.3. Model architectures

The preprocessed dataset is divided into training and testing subsets using a 70:30 split, ensuring that the evaluation process reflects realistic detection scenarios. Four machine learning models are employed in this study, each representing a different detection paradigm.

- LightGBM (statistical baseline) is used as a strong baseline classifier due to its effectiveness in handling structured tabular data and its ability to capture nonlinear feature interactions. The model is trained on the full multi-class dataset to provide a comprehensive comparison with the proposed hybrid framework [22].
- Isolation forest (statistical anomaly detection) is an ensemble-based anomaly detection algorithm that isolates anomalies by recursively partitioning the feature space. Because anomalies are typically rare and different from normal observations, they can be isolated with fewer partitions, making isolation forest efficient for large-scale datasets [8].
- Autoencoder (reconstruction-based detection) focuses on learning representations of normal traffic behavior. The model consists of an encoder that compresses input features into a latent representation and a decoder that reconstructs the original input. Anomalies are detected when the reconstruction error exceeds a predefined threshold, indicating deviation from learned normal patterns [6].
- LSTM (temporal anomaly detection) networks are designed to capture long-term dependencies in sequential data and are therefore well-suited for analyzing network traffic flows [13]. By learning temporal patterns of normal traffic behavior, LSTM models can detect abnormal sequence patterns associated with cyber-attacks [7].

Figure 4 illustrates the autoencoder architecture used in the proposed framework. The encoder compresses high-dimensional traffic features into a latent representation, while the decoder reconstructs the original input. The model is trained exclusively on normal traffic samples, enabling it to learn the characteristics of legitimate network behavior.

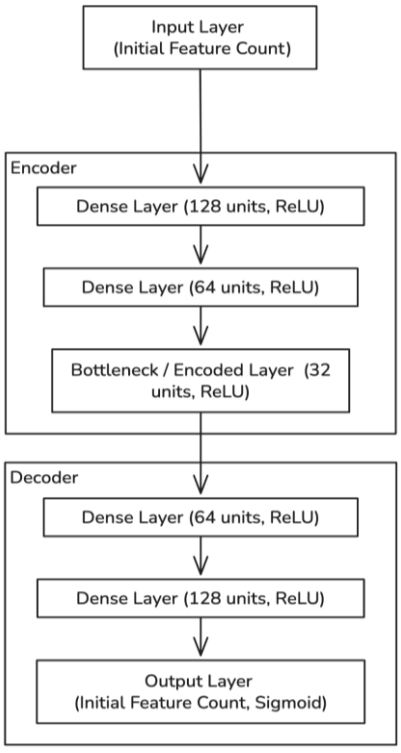

Figure 4. Autoencoder architecture for anomaly detection

The decoder then attempts to reconstruct the original input data from the latent representation. During the training phase, the autoencoder is trained using only normal traffic samples, allowing the model to learn the typical structure of legitimate network behavior. As a result, the model becomes highly effective in reconstructing normal traffic patterns while producing larger reconstruction errors when processing anomalous traffic. An anomaly score is calculated based on the reconstruction error between the original input vector x and the reconstruction output $\hat{x}$ which can be expressed [6]:

$$E = \|x - \hat{x}\|^2$$

where E represents the reconstruction error. If the reconstruction error exceeds a predefined threshold, the corresponding traffic instance is classified as anomalous. This mechanism allows the autoencoder to detect deviations from learned normal traffic patterns, making it particularly effective for identifying unknown or previously unseen attacks.

3.4. Decision fusion mechanism

The final stage of the proposed framework employs a decision fusion mechanism to combine the predictions of all base learners into a single outcome. Each individual model produces a binary decision, where 0 represents normal traffic and 1 indicates anomalous behavior**.** These outputs are aggregated through a weighted voting scheme, designed to leverage the complementary strengths of different models. Specifically, the LightGBM model is assigned a weight of 1.0**,** reflecting its balanced capability in handling structured data. The isolation forest is given a lower weight of 0.5, since it is effective for outlier detection but often generates higher false-positive rates in complex IIoT scenarios.

The autoencoder receives a weight of 1.2, as it captures nonlinear feature dependencies and demonstrates robustness under noise. Finally, the LSTM is assigned the highest weight of **1.5**, due to its ability to capture temporal dynamics across sequential traffic flows. A fusion threshold of 2.0 is applied: if the weighted sum of anomaly votes exceeds this value, the sample is classified as anomalous. This threshold is empirically chosen to prioritize recall**,** thereby reducing the probability of missed attacks, which is critical in IIoT security. At the same time, the weighted design helps maintain precision, providing a balanced and reliable detection strategy. The anomaly score is calculated using a weighted aggregation of model predictor [14]:

$$S = \sum_{i=1}^{n} w_i M_i$$

where M_i represent the binary output of the *i*-th model and w_i represent its assigned weight. In this study the weights are empirically determined based on validation performance.

$$w_{LightGBM} = 1.0,\ w_{IF} = 0.5,\ w_{AE} = 1.2,\ w_{LSTM} = 1.5$$

The final classification decision is defined as:

$$y = \begin{cases} 1, & S \geq T \\ 0, & S < T \end{cases}$$

where T represent the anomaly threshold. The threshold value is set to T = 2.0 to prioritize recall and reduce the likelihood of false negatives, which is critical in IIoT security environments.

3.5. Training and evaluation setup

The final component of the framework involves the training process and evaluation protocol used to validate system performance. Each model: autoencoder, isolation forest, and LSTM was trained and fine-tuned according to its specific architecture and learning characteristics. The autoencoder was optimized to minimize reconstruction error, the isolation forest was configured with appropriate tree depth and subsampling size, and the LSTM was trained on sequential traffic data to capture temporal dependencies. To ensure fairness and robustness, the dataset was partitioned into training, validation, and testing subsets using stratified sampling to preserve the imbalance ratio between normal and anomalous instances. Hyperparameter tuning was conducted through grid search and cross-validation to identify optimal configurations. Evaluation metrics included precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve AUC, with particular emphasis on minimizing false negatives due to their critical impact on intrusion detection [22]. The results of this setup provided a rigorous basis for assessing both the standalone models and the effectiveness of the proposed fusion strategy.

4. RESULTS AND ANALYSIS

This section presents the experimental results obtained from the proposed hybrid intrusion detection framework. The analysis focuses on evaluating the detection performance of individual models as well as the overall hybrid system. The experiments were conducted using the BoT-IoT dataset, which contains a large volume of network traffic generated from realistic IoT attack scenarios. The evaluation aims to assess how different detection paradigms perform under highly imbalanced IIoT traffic conditions.

4.1. Experimental environment

The experiments were conducted using Google Colab Pro+, a cloud-based computing platform that provides scalable computational resources for machine learning workloads. The platform was selected to efficiently process the large-scale BoT-IoT dataset, which contains millions of network traffic records representative of modern cyber threats [13].

Model training and evaluation were performed using Python 3.10. TensorFlow 2.x was employed for the autoencoder and LSTM models, while Scikit-learn was used for isolation forest, preprocessing, and evaluation metrics. LightGBM was utilized for the baseline classifier. All experiments were executed using a high-performance environment equipped with multi-core CPUs, substantial memory capacity, and GPU acceleration. This configuration ensured sufficient computational resources and reproducibility for evaluating the proposed framework under realistic operating conditions.

4.2. Performance evaluation

The performance of the proposed framework was evaluated by comparing four individual models—LightGBM, isolation forest, autoencoder, and LSTM—with the proposed hybrid intrusion detection framework. Approximately 70% of the dataset was used for training, while the remaining 30% was reserved for testing. All models were evaluated using a binary classification task that distinguishes normal traffic from anomalous traffic.

To provide a comprehensive assessment, precision, recall, F1-score, and area under the ROC curve were employed. These metrics evaluate not only overall classification performance but also the ability to minimize false positives and false negatives, both of which are critical in IIoT environments [23], [24]. Particular emphasis was placed on recall because missed attacks may lead to significant operational and security consequences. Table 2 summarizes the performance of all evaluated models.

Table 2. Performance comparison of individual models and the proposed hybrid detection framework

Model	Precision (anomaly)	Recall (anomaly)	F1-score (anomaly)	AUC
LightGBM	0.995	0.942	0.968	0.988
Isolation forest	0.879	0.500	0.640	0.129
Autoencoder	0.999	0.930	0.964	0.784
LSTM	0.950	0.920	0.935	0.500
Hybrid model (proposed)	0.999	0.970	0.985	0.787

The results show that LightGBM achieves the highest AUC (0.988), indicating strong ranking performance when distinguishing between normal and anomalous traffic. However, the proposed hybrid framework achieves the highest recall (0.970) and F1-score (0.985), demonstrating superior capability in identifying anomalous traffic while minimizing missed attacks.

Isolation forest exhibits the weakest overall performance, particularly in terms of recall and AUC, whereas autoencoder and LSTM provide complementary detection capabilities through reconstruction-based learning and temporal pattern analysis. By combining these heterogeneous detection paradigms through weighted fusion, the proposed framework achieves a more balanced performance than any individual model.

Although the hybrid model produces a lower AUC than LightGBM, this result reflects the design objective of the proposed framework. The weighted fusion strategy prioritizes recall and false-negative reduction rather than ranking performance alone. In highly imbalanced IIoT environments, reducing missed attacks is often more important than marginal improvements in AUC because undetected intrusions may cause severe operational and security impacts. Overall, the proposed framework achieves a favorable balance between detection sensitivity and classification reliability, making it suitable for practical anomaly detection in industrial IoT environments [25].

4.3. Hybrid model performance

The effectiveness of the proposed hybrid framework is further illustrated through the confusion matrix and ROC curve analysis shown in Figures 5 and 6. Figure 5 presents the confusion matrix of the proposed hybrid IDS. The model correctly identifies 659,753 anomalous instances while misclassifying only three anomalies as normal traffic. In addition, only 36 normal instances are incorrectly classified as anomalies. These results indicate a very low false-negative rate and demonstrate the framework's ability to detect malicious activities reliably while maintaining a low false-alarm rate.

Figure 6 compares the ROC curves of all evaluated models. LightGBM achieves the highest AUC (0.988), reflecting strong ranking capability. The proposed hybrid framework achieves an AUC of 0.787, comparable to the autoencoder (0.784), while providing improved recall and false-negative reduction. In contrast, isolation forest and LSTM exhibit substantially lower AUC values, indicating limited discriminative capability under the highly imbalanced conditions of the evaluated datasets.

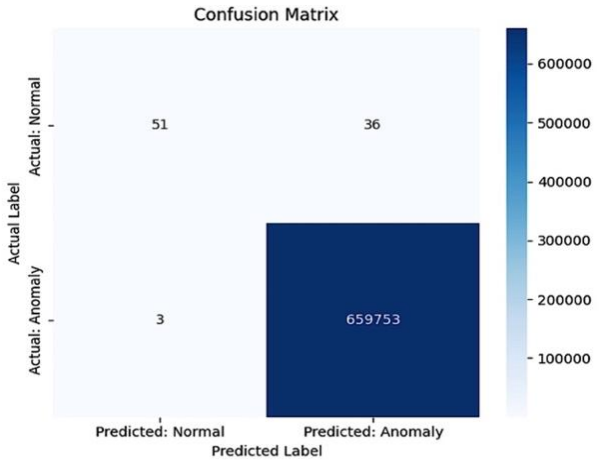

Figure 5. Confusion matrix of the proposed hybrid model

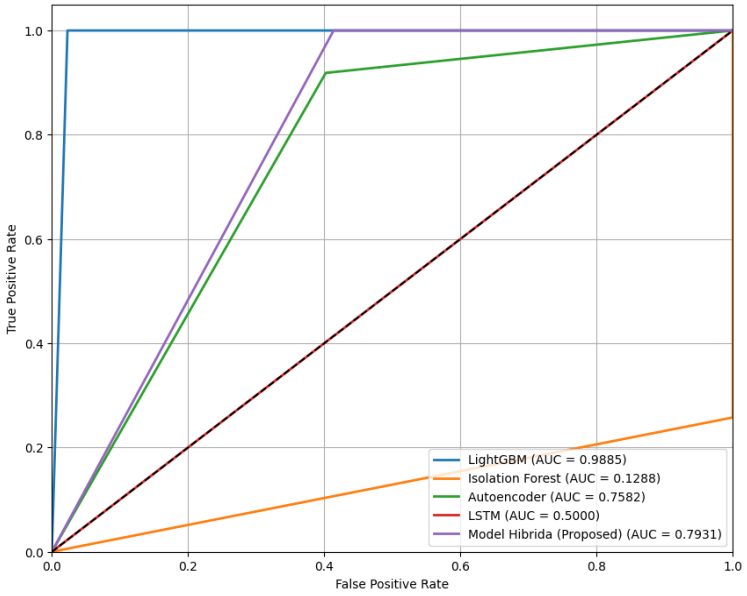

Figure 6. ROC curve comparison of all models

These findings suggest that the weighted fusion strategy successfully combines the complementary strengths of reconstruction-based, statistical, and temporal detection models. Although the hybrid framework does not achieve the highest AUC, it provides a more favorable balance between detection reliability and missed-attack reduction, which is particularly important in practical IIoT security environments.

5. CONCLUSION AND FUTURE WORK

This study proposed a hybrid anomaly detection framework for IIoT environments by integrating autoencoder, isolation forest, and LSTM models through a weighted decision fusion strategy. Experimental evaluation using the BoT-IoT dataset, with additional validation on IoT-23, demonstrates that the proposed framework achieves strong detection performance, particularly in terms of recall and false-negative reduction. Although LightGBM achieved a higher AUC, the hybrid framework provided a more balanced and reliable detection capability under highly imbalanced traffic conditions.

The results indicate that combining reconstruction-based, statistical, and temporal detection paradigms enables the framework to leverage complementary strengths while mitigating the limitations of individual models. This characteristic makes the proposed approach suitable for practical IIoT security monitoring, where detection reliability and missed-attack reduction are critical requirements.

Future work will focus on adaptive fusion mechanisms that dynamically adjust model weights according to changing traffic patterns. Additional evaluation using diverse real-world industrial datasets, multi-class attack scenarios, and explainable AI techniques will also be explored to further improve generalization, transparency, and operational usability.

ACKNOWLEDGMENTS

This research was conducted through collaborative efforts between Universitas Brawijaya and Universiti Teknologi PETRONAS. The authors would like to acknowledge the academic support and collaborative environment provided by both institutions, which significantly contributed to the development of this study.

FUNDING INFORMATION

This study was carried out without any external funding, and no financial support was received from governmental, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
I Dewa Made Widia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Toni Anwar	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no known financial or non-financial competing interests, including personal, professional, academic, ideological, or intellectual relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

REFERENCES

[1] L. D. Xu, E. L. Xu, and L. Li, "Industry 4.0: state of the art and future trends," *International Journal of Production Research*, vol. 56, no. 8, pp. 2941–2962, 2018, doi: 10.1080/00207543.2018.1444806.
[2] B. Alotaibi, "A survey on industrial internet of things security: requirements, attacks, AI-based solutions, and edge computing opportunities," *Sensors*, vol. 23, no. 17, Art. no. 7470, 2023, doi: 10.3390/s23177470.
[3] M. Nuaimi, L. C. Fourati, and B. B. Hamed, "Intelligent approaches toward intrusion detection systems for industrial internet of things: a systematic comprehensive review," *Journal of Network and Computer Applications*, vol. 215, Art. no. 103637, 2023, doi: 10.1016/j.jnca.2023.103637.
[4] A. Aldhaheri, F. Alwahedi, M. A. Ferrag, and A. Battah, "Deep learning for cyber threat detection in IoT networks: a review," *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 110–128, 2024, doi: 10.1016/j.iotcps.2023.09.003.
[5] T. Vaiyapuri, Z. Sbai, H. Alaskar, and N. Alaseem, "Deep Learning Approaches for Intrusion Detection in IIoT networks: a survey on methods, analysis, challenges and future directions," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, 2021, doi: 10.14569/IJACSA.2021.0120411.
[6] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proc. 2nd Workshop on Machine Learning for Sensory Data Analysis (MLSDA)*, 2014, pp. 4–11, doi: 10.1145/2689746.2689747.
[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[8] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in *Proc. 8th IEEE International Conference on Data Mining*, 2008, pp. 413–422, doi: 10.1109/ICDM.2008.17.

[9] H. C. Altunay and Z. Albayrak, "A hybrid CNN+LSTM-based intrusion detection system for industrial IoT security," *Engineering Science and Technology, an International Journal*, vol. 38, Art. no. 101322, 2023, doi: 10.1016/j.jestch.2022.101322.
[10] U. K. Lilhore *et al*., "HIDM: hybrid intrusion detection model for industry 4.0 networks using an optimized CNN-LSTM with transfer learning," *Sensors*, vol. 23, no. 18, Art. no. 7856, 2023, doi: 10.3390/s23187856.
[11] J. Casajús-Setién, C. Bielza, and P. Larrañaga, "Anomaly-Based Intrusion Detection in IIoT networks using transformer models," in *Proc. IEEE International Conference on Cyber Security and Resilience (CSR)*, 2023, pp. 72–77, doi: 10.1109/CSR57506.2023.10224965.
[12] Y. Wu, H.-N. Dai, and H. Tang, "Graph neural networks for anomaly detection in industrial internet of things," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9214–9231, 2022, doi: 10.1109/JIOT.2021.3094295.
[13] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: bot-IoT dataset," *Future Generation Computer Systems*, vol. 100, pp. 779–796, 2019, doi: 10.1016/j.future.2019.05.041.
[14] S. Garcia, A. Parmisano, and M. J. Erquiaga, "IoT-23: a labeled dataset with malicious and benign IoT network traffic," Zenodo, 2020, doi: 10.5281/zenodo.4743746.
[15] S. Elsaid and A. Binbusayyis, "An optimized isolation forest-based intrusion detection system for heterogeneous and streaming data in the industrial internet of things networks," *Discover Applied Sciences*, vol. 6, 2024, doi: 10.1007/s42452-024-06165-w.
[16] P. Ruzafa-Alcázar *et al.*, "Intrusion detection based on privacy-preserving federated learning for industrial IoT," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1145–1154, 2022, doi: 10.1109/TII.2021.3126728.
[17] Z. Wang, Z. Li, D. He, and S. Chan, "A lightweight approach for network intrusion detection in industrial cyber-physical systems based on knowledge distillation and deep metric learning," *Expert Systems with Applications*, vol. 206, Art. no. 117671, 2022, doi: 10.1016/j.eswa.2022.117671.
[18] 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. 36, pp. 295–300, 2022, doi: 10.1016/j.ifacol.2022.09.550.
[19] M. Bachar, A. Khiat, and K. El Guemmat, "Hybrid autoencoder and isolation forest for IoT anomaly detection with a novel model," *Engineering, Technology & Applied Science Research*, vol. 16, no. 1, pp. 31123–31129, 2026, doi: 10.48084/etasr.15288.
[20] C. Y. Priyanto, Hendry, and H. D. Purnomo, "Combination of isolation forest and LSTM autoencoder for anomaly detection," in *Proc. 2021 2nd International Conference on Innovative and Creative Information Technology (ICITech)*, 2021, doi: 10.1109/ICITech50181.2021.9590143.
[21] Y. K. Saheed, A. A. Usman, F. D. Sukat, and M. Abdulrahman, "A novel hybrid autoencoder and modified particle swarm optimization feature selection for intrusion detection in the internet of things network," *Frontiers in Computer Science*, vol. 5, 2023, doi: 10.3389/fcomp.2023.997159.
[22] G. Ke *et al.*, "LightGBM: a highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
[23] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.
[24] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006, doi: 10.1016/j.patrec.2005.10.010.
[25] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009, doi: 10.1109/TKDE.2008.239.

BIOGRAPHIES OF AUTHORS

I Dewa Made Widia received the B.Eng. degree in Electrical Engineering from Universitas Brawijaya, Malang, Indonesia, in 1991, and the M.Eng. degree in Electrical Engineering from Universitas Indonesia, Jakarta, Indonesia, in 1999. He obtained his Ph.D. in Environmental Science from Universitas Brawijaya, Malang, Indonesia, in 2025. He is currently a lecturer and researcher at the Faculty of Vocational Studies, Universitas Brawijaya. His research interests include machine learning, the IoT, network security, and the application of information technology for environmental sustainability. He can be contacted at email: dewa_vokasi@ub.ac.id.

Toni Anwar received the Ph.D. degree from RWTH Aachen University, Germany, in 1991. He is currently an associate professor with the Department of Computer and Information Science, Universiti Teknologi PETRONAS, Malaysia. His research interests include software engineering, IoT, artificial intelligence, data analytics, and smart applications in areas such as smart agriculture, smart cities, and connected healthcare. He can be contacted at email: toni.anwar@utp.edu.my.