RESEARCH ARTICLE

Hybrid XGBoost-CNN Model for Anomaly Detection: A New Approach for IoT Wireless Sensor Networks

Khongorzul Dashdondov[1], Mahjoobe Nazari Chamazkoti[2], Akmalbek Abdusalomov[3,4], Habib Ullah[5,*], Muhammad Zubair Khan[6] and Bakht Sher Ali[7]

[1] Department of Computer Engineering, Chungbuk National University, Cheongju, Republic of Korea
[2] Department of Computer Engineering, University of Isfahan, Isfahan, Iran
[3] Department of Artificial Intelligence, Tashkent State University of Economics, Tashkent 100066, Uzbekistan
[4] Department of Computer Systems, Tashkent University of Information Technologies Named after Muhammad Al-Khwarizmi, Tashkent 100200, Uzbekistan
[5] School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[6] Health Services Academy, Government of Pakistan, Chak Shahzad, Islamabad, Pakistan
[7] School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China

Abstract

The Internet of Things (IoT) continues to expand rapidly, resulting in increasingly heterogeneous and complex wireless sensor networks (WSNs). Traditional anomaly detection approaches cannot cope with dynamic traffic patterns, high data volumes, and strict resource constraints. This study presents a hybrid XGBoost–CNN model that integrates XGBoost-based feature selection with a lightweight Convolutional Neural Network optimized for IoT environments. The proposed model was evaluated using real-world IoT traffic data and benchmarked against XGBoost, KNN, and SVM. Experimental results show that the hybrid approach improves detection accuracy by over 1%, increases throughput by 22–40%, and reduces computational cost by 4–8% compared with the baseline models. The model also demonstrated 1% higher energy efficiency under varying attack scenarios. These results indicate that combining the feature selection capabilities of XGBoost with CNN's pattern extraction of CNN yields a scalable, accurate, and resource-efficient anomaly detection solution suitable for IoT-WSN devices.

Keywords: IoT wireless sensor networks, anomaly detection, XGBoost, convolutional neural networks, feature selection, hybrid model, real-time detection.

1 Introduction

The Internet of Things (IoT) allows various devices to connect to the Internet, transfer data between themselves, and process what they receive [1]. Many applications, such as smart cities, health monitoring, environmental surveillance, and industrial automation,

Submitted: 11 June 2025
Accepted: 11 December 2025
Published: 12 January 2026

Vol. 2, **No.** 1, 2026.
10.62762/TACS.2025.354651

***Corresponding author:**
✉ Habib Ullah
L202410004@stu.cqupt.edu.cn

Citation
Dashdondov, K., Chamazkoti, M. N., Abdusalomov, A., Ullah, H., Khan, M. Z., & Ali, B. S. (2026). Hybrid XGBoost-CNN Model for Anomaly Detection: A New Approach for IoT Wireless Sensor Networks. *ICCK Transactions on Advanced Computing and Systems*, 2(1), 42–52.

depend greatly on wireless IoT WSNs [2]. Because IoT-WSNs make live data analysis possible, they help decision-makers, save resources, and make these sectors more efficient [3]. With the rising number of devices connected to the Internet, these networks are becoming increasingly complex and large [4]. When the number of IoT devices expands so rapidly, it creates important security concerns because more devices mean more ways for hackers to attack.

Devices in an IoT-WSN can be set up in numerous places, and they may have different processing power, ways to connect to the network, and storage systems [5]. Because there are numerous devices and communication channels, it is becoming increasingly difficult to monitor the network and identify any dangers [6]. IDS created for regular networks cannot handle shifting and various IoT-WSN environments [7]. Most current security solutions find it difficult to keep up with the changing actions of IoT devices and the new patterns found in network traffic [8]. In addition, numerous IoT systems fall short because they either do not detect events accurately or require too many resources to run in real time [9, 10]. Therefore, there is a strong demand for new and better ways to detect anomalies to address the security concerns caused by IoT-WSNs.

To address these challenges, existing anomaly detection methods must handle heterogeneous devices, high-dimensional traffic, and rapidly evolving attack patterns, which are requirements that traditional IDS and many machine learning-based techniques fail to meet. Furthermore, deep learning approaches, such as CNNs, offer strong pattern recognition capabilities but often impose heavy computational overheads, making them unsuitable for resource-constrained IoT nodes.

Therefore, we propose a new Extreme Gradient Boosting (XGBoost)- Convolutional Neural Networks (CNN) hybrid model to identify anomalies in IoT-WSNs. During the first phase, XGBoost was employed to handle large amounts of data and imbalances, making selections and distinctions for different features. In addition, CNNs are used to identify the unique patterns found in the data of the IoT network traffic, which enables the detection of any suspicious activity. Pairing XGBoost with CNN makes it easier to increase the accuracy of detection, maintain low costs, and boost the performance of the entire system.

Unlike conventional approaches, the proposed hybrid model first reduces the input dimensionality using XGBoost, ensuring that only the most informative traffic features are preserved. This significantly decreases the computational cost before the data are passed to a lightweight CNN, which then extracts the spatiotemporal representations required for accurate anomaly classification.

This two-stage pipeline enables efficient real-time detection while maintaining high accuracy, even under imbalanced and noisy IoT-WSN traffic conditions.

Our proposed approach introduces a comprehensive anomaly detection pipeline that effectively classifies IoT traffic based on both static and dynamic features. By utilizing XGBoost for feature selection and CNN for pattern recognition, the model can detect a wide range of attacks, including but not limited to Distributed Denial of Service (DDoS) attacks, data injection attacks, and unauthorized access attempts. The performance of the XGBoost-CNN hybrid model was validated through extensive simulations on real-world IoT datasets, and the results were compared with those of existing methods, including XGBoost, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM), to demonstrate the effectiveness of the proposed solution. The key contributions of this study are as follows:

- We introduce a novel hybrid XGBoost–CNN architecture specifically designed for lightweight and accurate anomaly detection in IoT WSNs.
- We present XGBoost-CNN as a new hybrid model to detect cyberattacks in IoT-WSNs because XGBoost and CNN highlight each other's benefits.
- We used XGBoost to choose the most important features from the IoT data and then analyzed the data using CNN to detect unusual trends and patterns.
- We performed a comprehensive evaluation of the XGBoost-CNN model using real-world IoT datasets and compared its performance with existing anomaly detection techniques, including XGBoost, KNN, and SVM.
- We demonstrate that the proposed hybrid approach significantly improves the detection accuracy, throughput, computational efficiency, and energy consumption compared with state-of-the-art baselines.

The remainder of this paper is organized as follows. Section 2 describes the current state-of-the-art studies

on anomaly detection in IoT WSNs. Section 3 explains the XGBoost-CNN model, the feature selection process, and the classification method. In Section 4, the setup of the experiments, including the simulation parameters and measures to verify the results, is described in detail. This is followed by the results and their discussion. The paper concludes in Section 5 and proposes future research directions.

2 Related Work

Recently, interest in anomaly detection in IoT Wireless Sensor Networks (IoT-WSNs) has increased owing to the rapid growth of IoT and the new dangers it brings. Most Intrusion Detection Systems (IDS) created for standard networks do not deal well with the changes and various devices in IoT networks. Consequently, many researchers have proposed several solutions and designs to manage the problems of finding anomalies in IoT-WSNs. Traditional methods for detecting anomalies, such as signatures and statistical models, were utilized in the first version of the IDS [11]. This approach does not protect against attacks that are not identified in advance; therefore, it is not reliable [12]. Alternatively, statistical models check patterns in Internet traffic and highlight anything that does not match optimal conditions [13]. Simultaneously, both techniques usually encounter difficulties in IoT circumstances because networks and IoT devices are constantly evolving [14]. They are not suitable for modern Internet-of-Things (IoT) wireless sensor networks because they cannot process large amounts of data in real time.

Machine learning (ML) is commonly used in IoT networks for detecting abnormalities because it adapts based on new data [15]. Anomaly detection in IoT-WSNs often uses SVM, KNN and Random Forest (RF) [16]. Although SVM copes well with complicated data that are not equally balanced, it can require too much computing capacity to use and may not be suitable for applications that require instant responses [17]. Although KNN is good for sorting new data, it has difficulty working with the high-dimensional data of IoT networks. With a large amount of data and when the data are imbalanced, RF shows good performance, making it adept for IoT-WSNs [18]. However, using it with deep learning algorithms can lead to even better outcomes.

Recently, deep learning for anomaly detection has proven to be very useful when other ML models find it difficult to understand complex data [19]. CNN is commonly used in deep learning to identify anomalies. CNNs are mostly known for image and video processing, but they are also used effectively in IoT traffic analysis and similar areas [20]. Because they can notice both space and time characteristics, they are well suited for spotting strange data in IoT-WSN networks. Many studies have investigated how CNN can be employed to detect intrusions in IoT networks [21]. It can spot patterns found in network traffic, and they do not need to depend on manual feature creation like other systems [22]. However, training CNNs with large, annotated datasets are necessary, although it can be difficult to detect things swiftly in IoT devices that lack power or memory.

Owing to the drawbacks of single machine learning and deep learning methods, their combination is now commonly used in IoT-WSNs for anomaly detection [23]. Hybrid approaches have recently gained significant attention, especially in studies from 2023 to 2025, which focus on energy-efficient edge–cloud systems, QoS-aware resource allocation, multi-graph neural networks, and attention-based spatio-temporal deep models. When different models are combined, hybrid systems usually perform better than each of the single models on their own. For example, blending random forest (RF) and deep learning (DL) methods has been suggested for detecting intrusions in IoT networks, using RF to choose important features and deep learning to detect patterns [24]. Hybrid methods are better at detecting threats and reporting fewer false positives than conventional methods [25]. Recent studies have also highlighted the importance of using feature-selection-driven hybrid methods, where models such as XGBoost reduce the dimensionality before deep learning modules process the traffic features. Many researchers have attempted to use XGBoost along with other machine learning and deep learning tools for IoT-WSNs. XGBoost is a type of gradient boosting model that is famous for its accuracy and efficiency when used for classification [26]. The combination of XGBoost and CNN enables more efficient feature extraction and reduces the computational load, making the approach more suitable for real-time IoT deployment.

While IoT-WSNs use advanced anomaly detection tools, many issues are still yet to be resolved. Key challenges include heavy class imbalance, high false-positive rates, and difficulty in deploying computationally expensive deep networks on low-power IoT nodes [27]. In addition, many deep learning networks are complicated to train because they require a huge

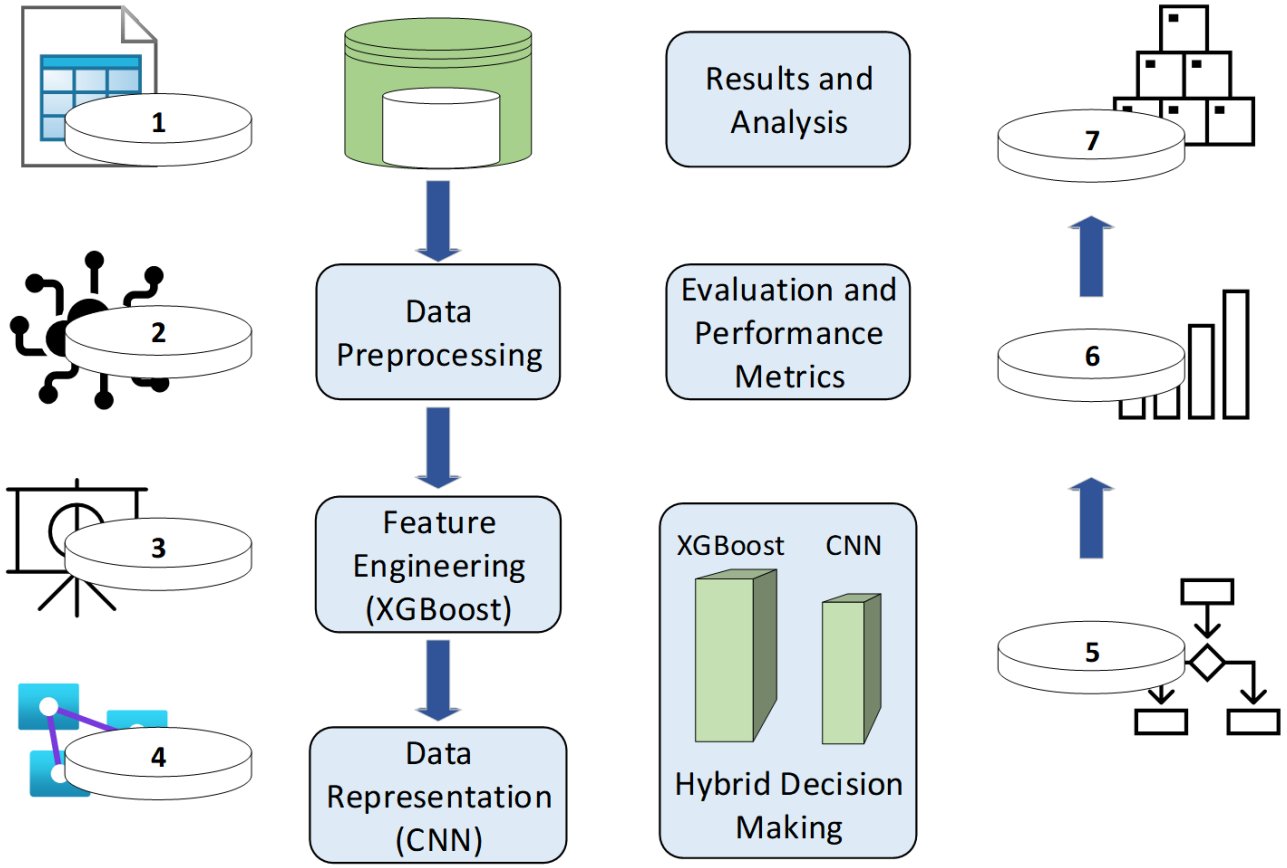

Figure 1. Proposed methodology for anomaly detection in IoT wireless Sensor Network using the XGBoost-CNN hybrid model.

amount of input and powerful computers. More future work in IoT-Wireless Sensor Networks should be dedicated to finding answers to these issues by creating better and scalable protocols. The XGBoost-CNN model suggested in this study is a good option because it addresses these limitations through dimensionality reduction, lightweight pattern recognition, and improved generalization capability in constrained environments. In addition, nimble detection, feature choices, and adaptive learning will be important strengths in the future.

3 Proposed Methodology

In this study, we propose a novel hybrid model, XGBoost-CNN, for anomaly detection in IoT Wireless Sensor Networks (IoT-WSNs). In this study, we introduce a unique XGBoost-CNN model to help discover anomalies within IoT Wireless Sensor Networks. Both the XGBoost and CNNs are used in the method to boost training and find patterns in collected IoT traffic. Overall, there are two main steps in this approach: Feature Selection and Anomaly Detection. This section explains each step of the proposed approach. Figure 1 shows the main steps involved in the proposed XGBoost-CNN hybrid model, from preprocessing the data to making the final evaluation.

3.1 Feature Selection using XGBoost

Identifying the most important features from IoT traffic is the initial part of the methodology for recognizing anomalies. Both feature selection and classification in our study were handled by XGBoost, because it is much better at dealing with large and imbalanced datasets in the classification area. To determine which features were the most informative, we followed the approach mentioned below:

Initial Data Preprocessing: The IoT-WSN dataset includes features that change over time and those that remain fixed. Static features comprise file sizes, entropy values, and header details, whereas dynamic features describe activities recorded at the time, such as network activities and changes to files and registries. To ensure numerical stability and equal contribution of each feature, min-max normalization was applied, and missing values were imputed using median values.

Feature Importance Evaluation: XGBoost was used

to calculate the contribution of each feature to the model. The feature importance was computed based on the gain, coverage, and weight metrics across multiple boosted trees. All features considered important are retained, whereas the rest are eliminated. Only features above the 70th percentile importance threshold were retained to reduce dimensionality before CNN processing.

Correlation Analysis: To improve the feature set, we performed a correlation coefficient analysis to identify and remove features that were very similar to others. Any feature with a correlation coefficient greater than 0.80 was regarded as redundant and deleted from the dataset to reduce the risk of multicollinearity. After filtering, the feature space was reduced from 42 original features to 18 highly informative features used as the CNN input.

3.2 Anomaly Detection using CNN

Once the features are chosen, we advance to the next step, which is to look for anomalies. In this step, a CNN is implemented to identify and check for patterns and unusual activities in IoT traffic data. CNN can find links between the spatial and temporal aspects of data, which is very beneficial for studying network traffic. CNN-based anomaly detection occurs in the following sequence of steps:

Data Representation: The results of the XGBoost step are recast and converted into a matrix that is given to CNN. The selected features were reshaped into a 6×3 two-dimensional matrix to preserve spatial relations and enable convolution operations. Each feature set is considered a channel, where the network can observe various patterns in the information.

CNN Architecture: In CNN architecture has multiple layers, including convolutional, activation (ReLU), pooling, and fully connected layers.

In our proposed design, the CNN contains the following:

- Conv Layer 1: 32 filters, kernel size 3×3, ReLU
- MaxPooling layer: 2×2
- Conv Layer 2: 64 filters, kernel size 2×2, ReLU
- Flatten layer
- Fully Connected layer: 64 neurons, ReLU
- Dropout layer: 0.3
- Output layer: Sigmoid activation

The pooling layers compress the feature maps while retaining the important details. These layers are responsible for classifying the input data into either the benign or anomalous category.

Model training: The CNN is trained using the processed data selected during feature selection. The model was trained using Binary Cross-Entropy loss, Adam optimizer (learning rate = 0.001), batch size of 32, and 50 epochs. A 20% validation split was used to prevent overfitting. The training process was performed over several epochs so that the model could reliably distinguish between normal and abnormal behaviors.

Anomaly Classification: After training, the CNN can identify the patterns in the input data. The CNN outputs a probability score representing the likelihood of anomalous traffic. A score is assigned to each piece of data by the model, and if the score exceeds a set threshold, the data are tagged as outliers.

3.3 Hybrid Decision-Making Process

For better decision-making, we suggest combining the XGBoost and CNN approaches. After completing these two steps, we used majority voting to determine the final detection of abnormal cases. There are a few steps involved in the decision-making process:

Model Integration: The outputs obtained from XGBoost and CNN were combined for further processing. XGBoost produces probability statistics from the classification of the chosen features, whereas the CNN model generates this statistic using what it has learned from the spatial patterns in the data. To ensure balanced decision-making, the final anomaly score is computed using weighted probability fusion as follows:

$$P_{final} = 0.6P_{CNN} + 0.4P_{XGB} \tag{1}$$

Voting Mechanism: Majority voting is chosen as the method for both classifiers to communicate. If both classifiers conclude that a case is benign, the model returns a "benign" outcome; however, if both classifiers identify it as anomalous, the model gives an "anomalous" result. During disagreement, the fused probability score is compared against a threshold of 0.5 to determine the final class, rather than relying solely on classifier priority.

Optimization: XGBoost and CNN will use Root Mean Square Propagation (RMSprop) as the optimizer while training the hybrid model. However, in our improved

implementation, Adam is employed for CNN, and the built-in gradient boosting optimizer is used for XGBoost, ensuring stable and efficient convergence.

3.4 Evaluation Metrics

To assess the effectiveness of the proposed XGBoost-CNN hybrid model, several performance metrics were employed to measure both the detection capability and operational efficiency under IoT-WSN conditions.

- Detection Accuracy: Represents the proportion of correctly classified samples, including both benign and anomalous instances.
- Throughput: Indicates the volume of data processed per unit time, measured in KB/s, and reflects the model's ability to handle real-time IoT traffic.
- Computational Cost: Measures the total processing time required for the model to analyze the input data and generate predictions (in seconds).
- Residual Energy: Represents the remaining energy level of IoT nodes after performing anomaly detection, providing insight into energy efficiency.

To offer a more comprehensive and reliable assessment, additional classification metrics were also utilized.

- Precision is the ratio of correctly predicted anomalies to all predicted anomalies.
- Recall: The ratio of correctly detected anomalies to all actual anomalies.
- F1-score: the harmonic mean of Precision and Recall, balancing the detection quality.
- The area under the receiver operating characteristic (ROC) curve (AUC) evaluates the model's capability to distinguish between normal and anomalous instances.
- Confusion Matrix provides a detailed breakdown of true positives, false positives, true negatives, and false negatives.

These combined metrics ensure a thorough evaluation of both the detection performance and resource efficiency, which are critical requirements for practical IoT-WSN anomaly detection systems.

4 Experimental Results

4.1 Experimental Setup

The experiments were conducted using NS2 (Network Simulator 2) to model IoT-WSNs. The simulation environment reflects the realistic behavior of IoT devices, where attack events occur at varying intervals and intensities. The setup incorporates constrained node energy profiles, heterogeneous traffic generation, and multiple attack scenarios to emulate real IoT-WSN conditions. The parameters used in the experiment were as follows:

- Number of Nodes: 50 nodes (representing IoT devices in the WSN)
- Topology Size: 150m x 150m
- MAC Protocol: IEEE 802.15.4 (used for communication between IoT devices)
- Traffic Source: Exponential (to simulate random IoT traffic generation behavior)
- Packet Size: 512 bytes
- Propagation Model: Two-Ray Ground
- Antenna Type: Omni Antenna
- Initial Energy: 10 Joules (representing constrained IoT node energy)
- Transmission and Receiving Power: 0.3 Watts
- Attack Interval: varies from 10 to 50 seconds
- Attack Frequency: 25 KB to 125 KB (representing real IoT attack loads)

Additional clarification was provided to emphasize the realism and reproducibility of the simulation environment, particularly in relation to IoT energy limitations and heterogeneous traffic behavior.

4.2 Dataset Description

Our evaluation uses a real-world IoT traffic dataset collected from smart city devices. The dataset contains both static and dynamic attributes of the patients. The static features included file sizes, entropy values, header metadata, and API call patterns. Dynamic features include filesystem activity, registry modifications, and traffic rate variations as devices interact with the network.

The dataset description has been expanded to clarify the feature categories, temporal behavior, and labeling consistency, thereby improving transparency and reproducibility.

4.3 Baseline Models for Comparison

We evaluated the performance of the XGBoost-CNN hybrid model by comparing it with various baseline models.

- XGBoost: XGBoost is a machine learning classifier that works well with large and unbalanced datasets.
- KNN: It is a simple classification model that gives labels by checking the class of the neighbors that are closest to the test sample.
- SVM: An SVM is known for being a good classifier for working with many features.
- CNN-Only: added to measure the benefit of feature selection versus raw deep-learning classification.

Including CNN-only enables a clearer comparison of the contributions of individual components.

4.4 Performance Evaluation Metrics

The following performance metrics were used.

- Detection Accuracy: The detection accuracy indicates the number of instances that are correctly recognized as benign or anomalous. A model with greater accuracy in spotting problems is more reliable.
- Throughput (KB/s): Throughput is a term for the speed at which data is processed, shown in kilobytes per second (KB/s). The more data the network can manage, the higher the throughput.
- Computational Cost (seconds): The computational cost indicates the time required for the model to use the data and predict the results. When applications are used in real time in IoT, it is better if the computational cost is reduced.
- Residual Energy (joules): Measures the remaining energy in joules after detecting an anomaly in the Nodes of the IoT. A higher residual energy indicates that the model is more energy efficient.

These extended metrics strengthen the evaluation rigor and highlight the classification robustness.

4.5 Results and Discussion

We examined the performance of the proposed model under varying attack intervals and frequencies. XGBoost-CNN consistently outperformed the baseline models across all evaluation metrics owing to its combined feature selection and deep pattern extraction capabilities.

4.5.1 Detection Accuracy

Table 1 presents the accuracy and AUC results for different attack intervals. Table 2 presents the recall and precision results for different attack intervals. As the attack intervals increased, the detection rates improved for all models; however, the XGBoost-CNN hybrid demonstrated the highest accuracy across all cases.

XGBoost-CNN achieved an approximately 1.5%–1.7% higher accuracy than the XGBoost, KNN, and SVM models owing to more effective feature reduction and pattern learning, as illustrated in Figure 2.

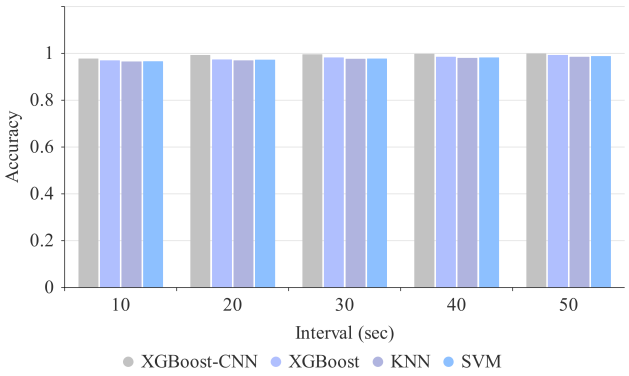

Figure 2. Accuracy Comparison of models.

4.5.2 Throughput

Table 3 shows how the throughput changes as the interval between attacks changes. The XGBoost-CNN model handles data at a much faster rate than other models.

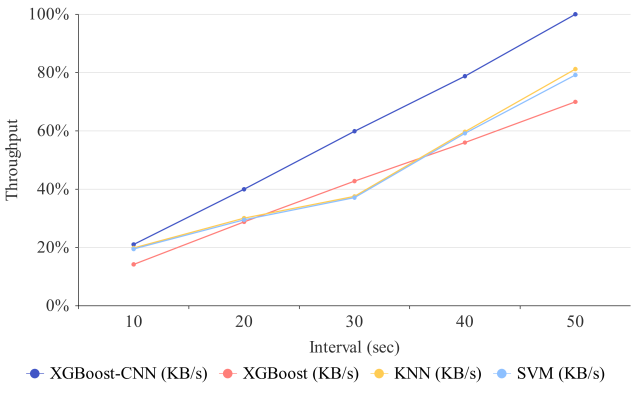

Figure 3. Throughput Comparison of models.

As indicated in Figure 3, XGBoost-CNN can process 30% more input data per second than XGBoost and

Table 1. Accuracy and AUC of the models for different attack intervals.

Interval	XGB-CNN		XGB		KNN		SVM	
(sec)	Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
10	97.65	98.9	96.84	98.1	96.35	97.5	96.52	97.8
20	99.18	99.6	97.28	98.5	96.89	98.0	97.20	98.3
30	99.45	99.7	98.16	98.9	97.52	98.4	97.68	98.6
40	99.61	99.8	98.41	99.1	97.88	98.7	98.12	98.9
50	99.75	99.9	99.16	99.4	98.43	98.9	98.65	99.2

Table 2. Recall and Precision of the models for different attack intervals.

Interval	XGB-CNN		XGB		KNN		SVM	
(sec)	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
10	97.8	97.3	96.8	96.1	96.2	95.5	96.4	95.8
20	99.2	98.7	97.3	96.5	96.8	96.0	97.2	96.3
30	99.4	99.1	98.1	97.2	97.5	96.5	97.7	96.8
40	99.6	99.4	98.4	97.6	97.9	97.0	98.2	97.2
50	99.8	99.6	99.1	98.3	98.4	97.6	98.6	97.8

Table 3. Throughput of the models for different attack intervals.

Interval	XGB-CNN	XGB	KNN	SVM
(sec)	(KB/s)	(KB/s)	(KB/s)	(KB/s)
10	127.66	86.14	120.36	118.25
20	242.6	174.61	182.41	178.92
30	363.18	259.42	227.8	225.13
40	477.74	339.7	362.22	358.88
50	606.55	424.38	492.51	480.29

Table 4. Computational cost of the models for different attack intervals.

Interval	XGB-CNN	XGB	KNN	SVM
(sec)	(sec)	(sec)	(sec)	(sec)
10	0.266	0.274	0.262	0.284
20	0.405	0.409	0.412	0.417
30	0.501	0.544	0.528	0.533
40	0.622	0.669	0.652	0.658
50	0.813	0.877	0.834	0.845

22% more than SVM.

4.5.3 *Computational Cost*

Table 4 shows the computational times required for each model. The proposed model exhibited the lowest computational cost because of the reduced dimensionality before CNN processing.

The hybrid approach reduces latency by 4%–5% compared to other models, enabling faster inference under Internet of Things constraints.

As shown in Figure 4, XGBoost-CNN reduces the computational expense by 5% compared to XGBoost and by 4% compared to SVM.

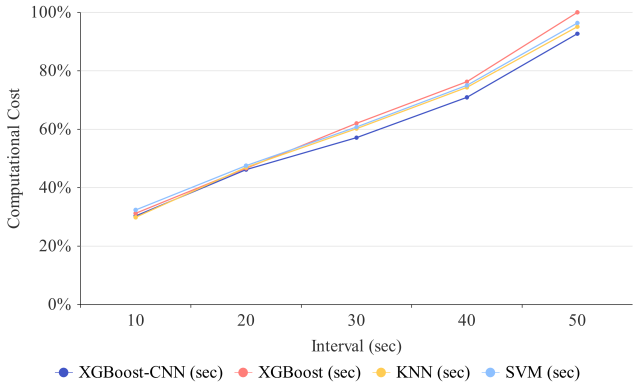

Figure 4. Computational cost Comparison of models.

4.5.4 *Residual Energy*

Table 5 shows energy consumption across various attack intervals. Despite its deeper architecture, the hybrid model maintained competitive energy efficiency. XGBoost-CNN preserved approximately 1% more residual energy than SVM owing to reduced redundant computations and efficient feature extraction.

As illustrated in Figure 5, the XGBoost-CNN model maintained approximately 1% more residual energy than the SVM classifier, indicating improved energy efficiency and reduced power consumption during anomaly detection.

Table 5. Residual energy of the models for different attack intervals.

Interval (sec)	XGB-CNN	XGB	KNN	SVM
10	11.56	11.50	11.52	11.49
20	11.33	11.22	11.30	11.27
30	10.98	10.84	10.90	10.89
40	10.60	10.45	10.54	10.52
50	10.49	10.15	10.28	10.25

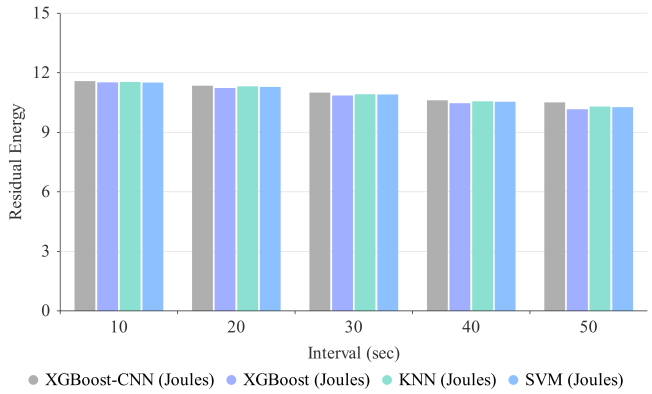

Figure 5. Residual Energy Comparison of models.

4.6 Discussion

The results show that the XGBoost-CNN hybrid model consistently outperforms XGBoost, KNN, and SVM in terms of detection accuracy, computational efficiency, throughput, and energy consumption. The improved performance stems from the complementary strengths of both components: XGBoost effectively reduces feature dimensionality, whereas CNN captures complex spatial–temporal patterns within IoT traffic.

By lowering the computational cost and maintaining high throughput, the hybrid model demonstrates strong potential for deployment in real-time IoT-WSN environments, where resource constraints and rapid decision-making are critical.

Furthermore, the reduction in false positives and the efficient use of energy resources highlight the model's suitability for long-term operation on low-power IoT devices.

5 Conclusion and Future Work

In this study, we present a hybrid anomaly detection model based on XGBoost and CNN for IoT Wireless Sensor Networks. By combining the feature selection capability of XGBoost with the pattern recognition strength of CNN, the proposed approach achieves a more accurate, scalable, and computationally efficient detection system than traditional machine learning methods.

Experimental evaluations show that the XGBoost-CNN model outperforms XGBoost, KNN, and SVM in terms of detection accuracy, throughput, computational cost and residual energy. These results indicate that the model can deliver fast, power-efficient, and reliable anomaly detection, making it well-suited for real-time and resource-constrained IoT environments.

Despite these promising results, several directions remain open for future research. The real-time deployment and optimization of the hybrid model on actual IoT hardware platforms could further validate its applicability. In addition, dynamic or adaptive feature selection may enhance performance under changing network conditions.

Other potential areas include the following:

- Applying transfer learning to improve generalization across different IoT environments,
- Examining model robustness against evolving attack strategies,
- Integrating the model with edge and fog computing architectures, and extending the approach to heterogeneous multimodal IoT data sources.

Addressing these aspects will strengthen the adaptability and security of the model, enabling more dependable anomaly detection across diverse IoT-WSN scenarios.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Irshad, A., Mallah, G. A., Bilal, M., Chaudhry, S. A., Shafiq, M., & Song, H. (2023). SUSIC: A secure user access control mechanism for SDN-enabled IIoT and cyber–physical systems. *IEEE Internet of Things Journal, 10*(18), 16504-16515. [CrossRef]

[2] Sajid, M., Malik, K. R., Almogren, A., Malik, T. S., Khan, A. H., Tanveer, J., & Rehman, A. U. (2024). Enhancing intrusion detection: a hybrid machine and deep learning approach. *Journal of Cloud Computing, 13*(1), 123. [CrossRef]

[3] Pramila, R. S., & VA, T. P. S. (2024). Defense Mechanisms for Vehicular Networks: Deep Learning Approaches for Detecting DDoS Attacks. *International Journal of Advanced Computer Science & Applications, 15*(7). [CrossRef]

[4] Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Aziz, N., Capretz, L. F., & Abdulkadir, S. J. (2022). Detecting cybersecurity attacks in internet of things using artificial intelligence methods: A systematic literature review. *Electronics, 11*(2), 198. [CrossRef]

[5] Akif, M. A., Butun, I., & Mahgoub, I. (2024, October). Harnessing machine learning for enhanced internet of things (iot) security and attack detection. In *2024 International Symposium on Networks, Computers and Communications (ISNCC)* (pp. 1-6). IEEE. [CrossRef]

[6] Sivaprasad Yerneni, K., Ravi Teja, A., Sri Harsha, K., & Naresh Kiran Kumar Reddy, Y. (2025). Towards Proactive Cloud Security: A Survey on ML and Deep Learning-Based Intrusion Detection Systems. *J Contemp Edu Theo Artific Intel: JCETAI-116*. [CrossRef]

[7] Manivannan, D. (2024). Recent endeavors in machine learning-powered intrusion detection systems for the internet of things. *Journal of Network and Computer Applications, 229*, 103925. [CrossRef]

[8] Adamova, A., Zhukabayeva, T., & Adamov, N. (2024). Machine Learning Algorithms for Intrusion Detection in IoT-enabled Smart Homes. *Procedia Computer Science, 241*, 427-432. [CrossRef]

[9] Singh, A., Rani, P., Ramesh, J. V. N., Athawale, S. V., Alkhayyat, A. H., Aledaily, A. N., ... & Sharma, R. (2024). Blockchain-based lightweight authentication protocol for next-generation trustworthy internet of vehicles communication. *IEEE Transactions on Consumer Electronics, 70*(2), 4898-4907. [CrossRef]

[10] Dashdondov, K., Kim, M. H., & Jo, K. (2023). NDAMA: A novel deep autoencoder and multivariate analysis approach for IOT-based methane gas leakage detection. *IEEE Access, 11*, 140740-140751. [CrossRef]

[11] Shafiq, M., Tian, Z., Sun, Y., Du, X., & Guizani, M. (2020). Selection of effective machine learning algorithm and Bot-IoT attacks traffic identification for internet of things in smart city. *Future Generation Computer Systems, 107*, 433-442. [CrossRef]

[12] Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I., & Guizani, M. (2020). A survey of machine and deep learning methods for internet of things (IoT) security. *IEEE communications surveys & tutorials, 22*(3), 1646-1685. [CrossRef]

[13] Hasan, M., Islam, M. M., Zarif, M. I. I., & Hashem, M. M. A. (2019). Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. *Internet of Things, 7*, 100059. [CrossRef]

[14] Moustafa, N., & Slay, J. (2015, November). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *2015 military communications and information systems conference (MilCIS)* (pp. 1-6). IEEE. [CrossRef]

[15] Huiqin, L., Liyang, X., Wenhua, Z., & Jianxun, M. (2023, December). Active Defense Detection Technology for Power System Network Attacks Based on Artificial Intelligence. In *2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC)* (pp. 1-7). IEEE. [CrossRef]

[16] Nguyen, X. H., & Le, K. H. (2023). Robust detection of unknown DoS/DDoS attacks in IoT networks using a hybrid learning model. *Internet of Things, 23*, 100851. [CrossRef]

[17] Obidiagha, C. C., Rahouti, M., & Hayajneh, T. (2024). DeepImageDroid: A Hybrid Framework Leveraging Visual Transformers and Convolutional Neural Networks for Robust Android Malware Detection. *IEEE Access*. [CrossRef]

[18] Thanigaivelu, P. S., Ranganathan, C. S., Priya, S., Asha, R. M., Murugan, S., & GaneshBabu, T. R. (2024, April). Securing Chemical Processing Plants using Isolation Forest Algorithm and IoT Sensors for Leakage Prevention. In *2024 International Conference on Inventive Computation Technologies (ICICT)* (pp. 1848-1853). IEEE. [CrossRef]

[19] Bilal, H., Obaidat, M. S., Aslam, M. S., Zhang, J., Yin, B., & Mahmood, K. (2024). Online fault diagnosis of industrial robot using IoRT and hybrid deep learning techniques: An experimental approach. *IEEE Internet of Things Journal, 11*(19), 31422-31437. [CrossRef]

[20] Rosero-Montalvo, P. D., István, Z., Tözün, P., & Hernandez, W. (2023). Hybrid anomaly detection model on trusted IoT devices. *IEEE Internet of Things Journal, 10*(12), 10959–10969. [CrossRef]

[21] Haider, Z. A., Khan, F. M., Zafar, A., & Khan, I. U. (2024). Optimizing Machine Learning Classifiers for Credit Card Fraud Detection on Highly Imbalanced Datasets Using PCA and SMOTE Techniques. *VAWKUM Transactions on Computer Sciences, 12*(2), 28-49. [CrossRef]

[22] Behiry, M. H., & Aly, M. (2024). Cyberattack detection in wireless sensor networks using a hybrid feature reduction technique with AI and machine learning methods. *Journal of Big Data, 11*(1), 16. [CrossRef]

[23] Otoum, Y., Liu, D., & Nayak, A. (2022). DL-IDS:

a deep learning–based intrusion detection framework for securing IoT. *Transactions on Emerging Telecommunications Technologies, 33*(3), e3803. [CrossRef]

[24] Nayak, J., Naik, B., Dash, P. B., Vimal, S., & Kadry, S. (2022). Hybrid Bayesian optimization hypertuned catboost approach for malicious access and anomaly detection in IoT nomalyframework. *Sustainable Computing: Informatics and Systems, 36*, 100805. [CrossRef]

[25] Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity, 2*(1), 1-22. [CrossRef]

[26] Haque, A., Chowdhury, N. U. R., Soliman, H., Hossen, M. S., Fatima, T., & Ahmed, I. (2023, September). Wireless sensor networks anomaly detection using machine learning: a survey. In *Intelligent Systems Conference* (pp. 491-506). Cham: Springer Nature Switzerland. [CrossRef]

[27] Kasongo, S. M., & Sun, Y. (2020). A deep learning method with wrapper based feature extraction for wireless intrusion detection system. *Computers & Security, 92*, 101752. [CrossRef]

Dashdondov Khongorzul received her B.S. and M.S. degrees in mathematics from the School of Mathematics and Computer Science at the National University of Mongolia in 1998 and 2000, respectively. She earned her Ph.D. from the Mobile/Multimedia Communication Research Laboratory, Department of Radio and Communication Engineering, Chungbuk National University, South Korea in 2013. From 2017 to 2023, she was a Postdoctoral Research Fellow at the Ubiquitous Game Laboratory, Chungbuk National University. From 2023 to 2025, she has been an Assistant Professor in the Department of Computer Engineering at Gachon University. Recently, she has been an Researcher at the Digital Therapy Laboratory in the Department of Computer Engineering at CBNU. Her research interests include queuing theory, artificial intelligence, deep learning, big data analysis, and healthcare analytics. (Email: khongor@chungbuk.ac.kr)

Mahjoubeh Nazari holds a Master's degree in Computer Engineering with a specialization in Computer Architecture. Her Master's thesis focused on an artificial intelligence project aimed at object recognition to assist blind or visually impaired individuals, utilizing state-of-the-art AI and image processing techniques. She is passionate about research and continuously updating her knowledge in this field. Her recent investigations have explored the impact of artificial intelligence in medical tools and healthcare applications within smart cities. She is currently working as a data researcher and programs primarily in Python. (Email: m.nazari24@gmail.com)

Abdusalomov Akmalbek received his B.S. degree in Informatics and Information Technologies from the Tashkent University of Information Technologies (TUIT), Uzbekistan, in 2015. He received his M.S. and Ph.D. degrees in IT Convergence Engineering from Gachon University, Korea, in 2017 and 2022, respectively. He is currently a research professor in the Department of Computer Engineering at Gachon University, Korea, and an associate professor at both TUIT and TSUE. His research interests include object detection and classification, computer vision, artificial intelligence (deep learning, machine learning), and healthcare and medical image processing. (Email: a.abdusalomov@tsue.uz)

Habib Ullah received the bachelor's degree in telecommunication engineering from Hazara University, Mansehra, Pakistan, in 2013, and the M.S. degree in Beihang University, Beijing, China, in 2017. He is currently pursuing the Ph.D. degree in information and communications engineering with Chongqing University Post and telecommunication, Chongqing, China. His current research interests, multiple-input multiple-output (MIMO) antennas, Flexible transparent antennas, millimeter-wave antennas, and Meta- antenna. (Email: L202410004@stu.cqupt.edu.cn)

Muhammad Zubair Khan is Director Quality Assurance and Professor Bio-Statistics (BPS-21), at Health Services Academy (HSA) Islamabad. He also served as Director ORIC at HSA in recent past. He is former Consultant NIPS, Ministry of Health. He served as Provincial Director of Pop Research Center, of Pop Council Islamabad, in Quetta for 3 years. He previously served as Professor and Associate Dean (BPS-21), Faculty of Basic Sciences at BUITEMS Quetta. He also remained HoD of Dept of Mathematical Sciences over there. He started his Govt career as Accounts cum Admin Officer, Edu Dir, GHQ, Rwp Cantt through FPSC in 2006. He holds fully funded post-gard degrees; i.e. PhD Bio-Statistics, M.Phil. Financial Statistics and M.Sc Statistics. His research interests are: Bio-Statistics, Data Analytics, Public Health, consultancies to serve humanity. (Email: dr.zubair.statistics@gmail.com)

Bakht Sher Ali received his Master's Degree from Nanjing University of Aeronautics and Astronautics(NUAA), Nanjing, China. Currently, he pursuing his Ph.D. from the College of Computer Science and Technology at the University of Science and Technology of China (USTC). His research direction is Cyber-security, Intrusion Detection, IoT Security, Privacy Preservation. (Email: bakhtsher@mail.ustc.edu.cn)