Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00214-8

Original Article

Detecting and Preventing Cyberattacks in Internet of Things (IoT) Systems

Ruthwik Palem [1], Likhith Reddy Peketi [1], Vanathi M [1, ✉], Sandhiya B [1], Karthika J [2]

1. *Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, IND*
2. *Computer Science and Technology, Sathyabama Institute of Science and Technology, Chennai, IND*

Received: April 21, 2026 | Review began: May 17, 2026 | Review ended: July 21, 2026 | Published: August 10, 2026

Abstract

The rapid growth of Internet of Things (IoT) devices has significantly improved connectivity across smart environments. However, the resource-constrained nature of IoT devices and their limited built-in security mechanisms make them highly vulnerable to evolving cyberattacks. Traditional intrusion detection systems relying on signature-based or static rule sets are often ineffective against previously unseen attacks. This study proposes a hybrid machine learning-based intrusion detection and prevention framework for securing IoT networks. The framework integrates Isolation Forest, Autoencoder, Extreme Gradient Boosting, and Bidirectional Long Short-Term Memory models within a stacked ensemble architecture to improve attack detection while reducing false-positive predictions. The ensemble outputs are combined using a Logistic Regression meta-classifier to generate the final threat score. Experimental evaluation using the UNSW-NB15 dataset demonstrates that Extreme Gradient Boosting achieved the highest individual Receiver Operating Characteristic-Area Under the Curve score of 0.91, while the proposed framework further enhances detection robustness through ensemble learning and automated prevention. A Flask-based monitoring dashboard provides real-time visualization of detection results, blocked IP addresses, alerts, and system performance.

Categories: Network Security, IoT Security and Privacy, Machine Learning (ML)

Keywords: intrusion detection system, machine learning, cybersecurity, anomaly detection, ensemble learning, internet of things (iot)

Introduction

The Internet of Things (IoT) has emerged as one of the most significant technological advancements of the digital era, enabling seamless communication among billions of interconnected devices across various application domains, including healthcare, transportation, manufacturing, agriculture, industrial automation, and smart city infrastructures [1]. The rapid growth of IoT technologies has significantly improved operational efficiency, automation, and data-driven decision-making by facilitating continuous communication between sensors, embedded devices, cloud platforms, and intelligent applications. As the number of connected devices continues to increase, IoT ecosystems have become an integral component of modern digital infrastructure.

Despite these technological advancements, IoT networks remain highly vulnerable to cyberattacks due to their resource-constrained nature, heterogeneous communication protocols, and limited built-in security mechanisms [2,3]. Unlike conventional computing systems, many IoT devices possess limited processing power, memory capacity, and energy resources, making the implementation of complex security mechanisms particularly challenging. Consequently, IoT environments have become attractive targets for cyberthreats such as distributed denial-of-service attacks, botnet infections, ransomware, spoofing attacks, malware propagation, unauthorized access, and data manipulation [4].

Intrusion detection systems (IDSs) have traditionally served as one of the primary mechanisms for identifying malicious activities within computer networks. Signature-based IDSs compare observed network traffic against predefined attack signatures and rule sets to detect known threats. Although these approaches achieve high accuracy for previously identified attacks, they demonstrate poor performance when confronted with zero-day attacks, polymorphic malware, and continuously evolving cyberthreats. Similarly, anomaly-based detection methods improve the detection of previously unseen attacks but frequently suffer from higher false-positive rates due to the dynamic nature of legitimate network behaviour [5].

Recent developments in artificial intelligence (AI), machine learning, and deep learning have significantly improved the effectiveness of intrusion detection by enabling intelligent analysis of complex network traffic patterns [6]. Supervised learning algorithms such as Extreme Gradient Boosting (XGBoost) have demonstrated excellent performance in identifying known attack categories through labelled training data, whereas unsupervised techniques including Isolation Forest and Autoencoders effectively recognise anomalous network behaviour without requiring labelled attack samples [7,8]. Furthermore, recurrent neural network architectures such as Bidirectional Long Short-Term Memory (Bi-LSTM) networks are capable of learning temporal dependencies present within network traffic, enabling improved detection of sophisticated multi-stage attacks [9].

Although numerous machine learning-based intrusion detection frameworks have been proposed, many existing approaches rely on a single detection model, thereby limiting their capability to effectively detect both known and unknown attacks simultaneously. Individual learning algorithms often exhibit model-specific weaknesses, including reduced generalization capability, overfitting, higher false-positive rates, or limited adaptability to evolving cyberthreats. Moreover, several existing studies primarily focus on attack detection while providing limited support for automated prevention and response mechanisms capable of mitigating malicious activities immediately after detection [10-14].

Research gap

Current IoT intrusion detection research predominantly focuses on improving attack detection accuracy using individual machine learning or deep learning algorithms. However, relatively few studies integrate supervised learning, unsupervised anomaly detection, and deep learning within a unified ensemble framework that simultaneously supports automated prevention. Furthermore, many existing approaches provide limited implementation details regarding ensemble decision-making, reproducibility, and practical deployment considerations in real-world IoT environments. These limitations highlight the need for a comprehensive hybrid intrusion detection and prevention framework capable of accurately identifying diverse attack patterns while enabling intelligent automated response mechanisms.

Motivation

The increasing sophistication and frequency of cyberattacks targeting IoT infrastructures necessitate adaptive security mechanisms capable of analysing complex network behaviour in real time. Integrating multiple complementary detection models enables the strengths of supervised learning, anomaly detection, and sequential deep learning to be combined within a single decision-making framework. Such integration not only improves detection reliability but also reduces false-positive predictions and enhances resilience against previously unseen cyberthreats.

Major contributions

The primary contributions of this research are summarized as follows:

1. A hybrid intrusion detection and prevention framework integrating Isolation Forest, Autoencoder, XGBoost, and Bi-LSTM models within a stacked ensemble architecture.

2. A Logistic Regression-based ensemble learning strategy that combines the outputs of multiple detection models to improve overall intrusion detection performance and reduce false-positive rates.

3. An automated prevention mechanism capable of generating security alerts, dynamically blocking malicious IP addresses, and maintaining a real-time blocklist following intrusion detection.

4. A comprehensive experimental evaluation conducted using the UNSW-NB15 benchmark dataset with multiple performance metrics, including Precision, Recall, F1-score, Confusion Matrix, and Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) analysis.

5. A scalable cybersecurity framework that combines detection, prevention, and real-time monitoring through an integrated Flask-based dashboard suitable for intelligent IoT security applications.

Materials And Methods

Detection techniques

Rule-Based IDSs

Rule-based IDSs identify malicious activities by comparing network traffic against predefined attack signatures and manually created security rules. These systems are highly effective in detecting previously known attack patterns with minimal computational complexity and have been widely deployed in traditional network security environments [2]. Whenever incoming traffic matches a stored signature, the corresponding event is classified as malicious, enabling rapid identification of well-documented cyberthreats.

Despite their effectiveness against known attacks, rule-based IDSs exhibit several limitations in modern IoT environments. Since detection relies entirely on predefined signatures, these systems cannot effectively identify zero-day attacks, polymorphic malware, or evolving threat patterns. Furthermore, maintaining an up-to-date signature database requires continuous manual intervention, making rule-based approaches increasingly difficult to manage as the number and diversity of IoT devices continue to grow [3]. Consequently, rule-based IDSs are generally used as complementary security mechanisms rather than standalone solutions for securing large-scale IoT networks.

Statistical and Behaviour-Based Detection

Statistical and behaviour-based intrusion detection techniques identify cyberthreats by modelling the normal behavioural characteristics of network traffic rather than relying on predefined attack signatures. These approaches establish baseline profiles using statistical measures such as packet transmission rate, communication frequency, protocol usage, and connection duration. Significant deviations from the learned normal behaviour are considered potential indicators of malicious activity [4].

One of the primary advantages of behaviour-based detection is its capability to identify previously unseen attacks without requiring labelled attack signatures. This makes statistical detection particularly suitable for IoT environments where new attack variants frequently emerge. However, behaviour-based approaches may generate higher false-positive rates because legitimate changes in network behaviour can sometimes resemble anomalous activity. Consequently, statistical methods are often integrated with machine learning algorithms to improve detection reliability and reduce false alarms [5].

Machine Learning-Based Detection

Machine learning techniques have significantly improved intrusion detection by enabling intelligent analysis of complex network traffic patterns and automatic classification of malicious activities. Supervised learning algorithms are trained using labelled datasets to distinguish between normal and malicious network behaviour based on learned feature relationships [6].

Among various machine learning algorithms, XGBoost has demonstrated excellent performance in intrusion detection due to its ability to model nonlinear feature interactions while maintaining high computational efficiency and robustness against overfitting [12]. By constructing multiple decision trees through gradient boosting, XGBoost effectively learns complex attack patterns and provides accurate classification of known cyberthreats. Owing to these advantages, machine learning-based IDSs have become an essential component of modern cybersecurity solutions for IoT environments.

How to cite this article:
Palem R, Peketi L, M V, et al. (August 10, 2026) Detecting and Preventing Cyberattacks in Internet of Things (IoT) Systems.
Cureus J Comput Sci 3 : es44389-026-00214-8. DOI https://doi.org/10.7759/s44389-026-00214-8

Deep Learning for IoT Intrusion Detection

Deep learning techniques provide enhanced capability for detecting sophisticated cyberattacks by automatically learning hierarchical feature representations from network traffic data. Unlike conventional machine learning algorithms, deep neural networks can capture complex nonlinear relationships and temporal dependencies, enabling improved detection of advanced persistent threats and multi-stage attacks [8].

In the proposed framework, two complementary deep learning models are employed. The Autoencoder performs unsupervised anomaly detection by learning the normal characteristics of network traffic and identifying abnormal behaviour through reconstruction error analysis. Additionally, the Bi-LSTM network captures sequential dependencies within network communication by analysing traffic in both forward and backward directions. The combination of anomaly detection and temporal learning significantly improves the framework's ability to detect both known and previously unseen cyberattacks while reducing false-negative rates [13].

Hybrid and Ensemble Detection

Although individual detection techniques demonstrate strong performance under specific attack scenarios, no single algorithm consistently provides optimal results across all categories of cyberthreats. Supervised classifiers accurately recognise known attacks but may struggle with previously unseen threats, whereas anomaly detection techniques effectively identify novel attacks while often producing higher false-positive rates. Similarly, deep learning models successfully capture temporal attack behaviour but require greater computational resources [9-11].

To overcome these limitations, the proposed framework adopts a hybrid ensemble learning approach that integrates Isolation Forest, Autoencoder, XGBoost, and Bi-LSTM models. Each model independently analyses the incoming network traffic and generates an anomaly score or attack probability. These outputs are subsequently combined using a Logistic Regression-based stacked ensemble, enabling the framework to learn the optimal combination of predictions rather than relying on fixed voting strategies. This hybrid approach improves detection robustness, enhances generalisation capability, and reduces false-positive rates by leveraging the complementary strengths of supervised learning, unsupervised anomaly detection, and deep learning techniques. The final ensemble prediction is then utilised by the automated prevention module to initiate appropriate mitigation actions against malicious network traffic.

Proposed methodology

The proposed Intrusion Detection and Prevention System (IDPS) integrates supervised learning, unsupervised anomaly detection, deep learning, and ensemble learning within a unified cybersecurity framework to improve the detection and mitigation of cyberattacks in IoT environments. Unlike conventional IDSs that rely on a single detection technique, the proposed framework combines multiple complementary learning models to improve detection accuracy, reduce false positives, and enhance resilience against both known and previously unseen attacks.

The overall framework consists of six sequential stages:

(i) Network Traffic Acquisition
(ii) Data Preprocessing
(iii) Hybrid Intrusion Detection
(iv) Stacked Ensemble Learning
(v) Automated Prevention
(vi) Real-Time Monitoring

System Architecture

The proposed system follows a modular architecture in which network traffic is processed through multiple detection stages before the final intrusion decision is generated. Initially, network traffic records are collected from the UNSW-NB15 benchmark dataset and subjected to preprocessing operations, including duplicate removal, missing value handling, categorical feature encoding, feature normalization, and class balancing. These preprocessing steps improve data quality while ensuring compatibility across all machine learning and deep learning models.

The processed data are simultaneously analysed by four independent detection models, namely Isolation Forest, Autoencoder, XGBoost, and Bi-LSTM. Each model independently evaluates the incoming network traffic and generates an anomaly score or attack probability based on its learning mechanism.

The outputs generated by the four detection models are subsequently combined using a stacked ensemble learning strategy. Unlike conventional voting-based approaches, the proposed framework employs Logistic Regression as a meta-classifier to learn the optimal combination of individual model predictions. The ensemble model produces a final threat probability representing the likelihood that the analysed network traffic corresponds to malicious activity.

Based on the generated threat probability, the automated prevention module determines the appropriate mitigation strategy. Network traffic classified as malicious is blocked by adding the corresponding source IP address to the dynamic blocklist, while suspicious traffic generates security alerts for administrator review. Finally, all security events are visualised through a Flask-based monitoring dashboard, enabling continuous observation of system performance, blocked IP addresses, detection metrics, and feature importance analysis.

Figure *1* depicts the architecture of the proposed intrusion detection and prevention system.

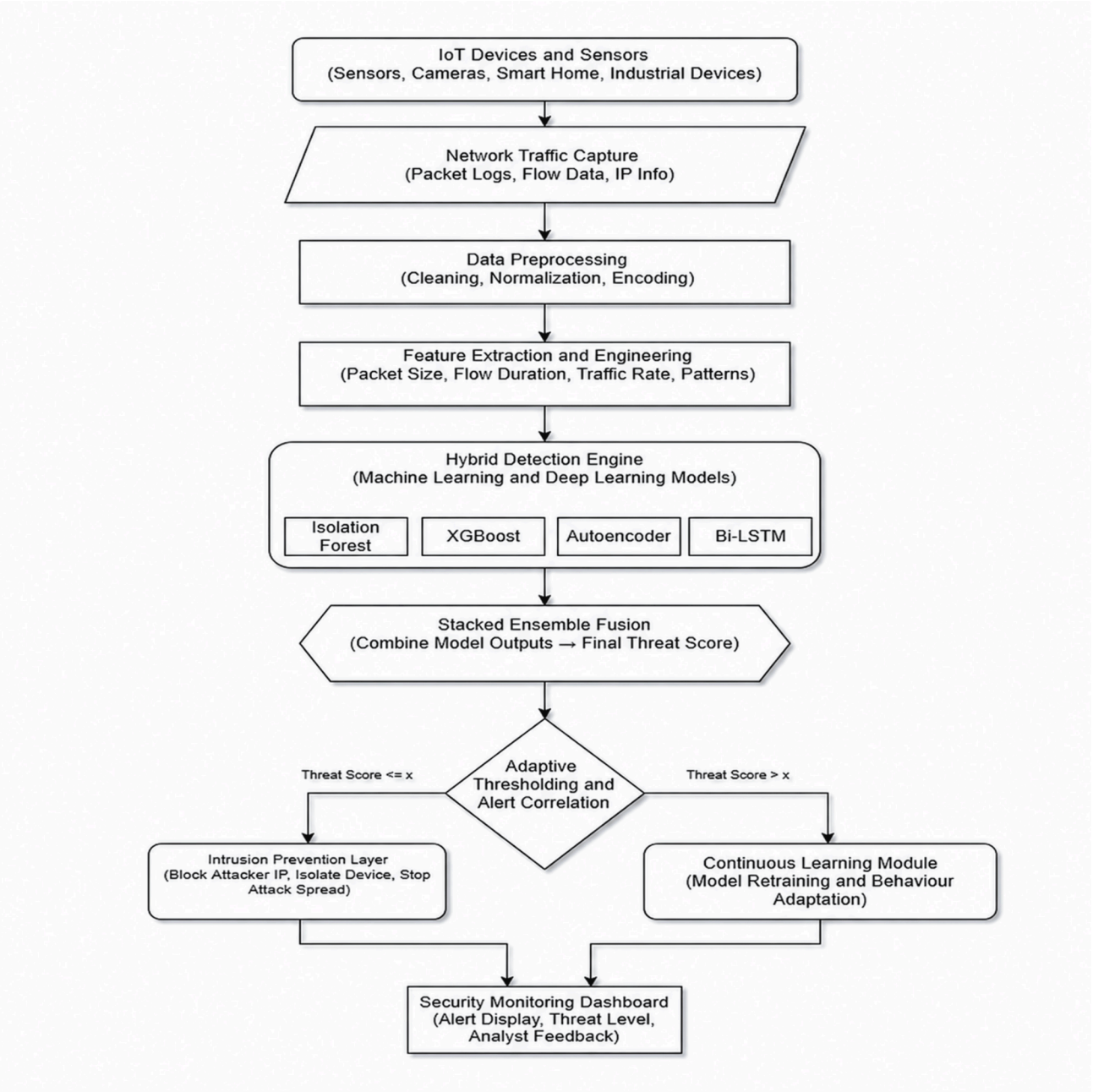

FIGURE 1: System Architecture of the Proposed IoT Intrusion Detection and Prevention Framework

Source: Author's own illustration
Bi-LSTM, Bidirectional Long Short-Term Memory; IoT, Internet of Things

Network Traffic Acquisition and Preprocessing

The proposed framework utilises the UNSW-NB15 benchmark dataset for model training and evaluation because it contains modern network traffic representing both legitimate communication and multiple categories of cyberattacks, including Exploits, Denial of Service, Generic attacks, Reconnaissance, Fuzzers, Analysis, Backdoors, Shellcode, and Worms [15].

Prior to model training, the collected data undergo several preprocessing operations to improve data quality and model performance. Duplicate records and missing values are removed to eliminate inconsistencies within the dataset. Categorical attributes, including protocol type, service type, and connection state, are transformed into numerical representations using Label Encoding, while numerical features are normalised using Min-Max Scaling to ensure that all features contribute equally during model training.

To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is applied only to the training dataset, thereby improving the classifier's ability to learn minority attack classes without affecting the evaluation dataset. Finally, the processed dataset is divided into 80% training data and 20% testing data to ensure consistent evaluation across all implemented detection models.

Hybrid Detection Model

To improve intrusion detection performance, the proposed framework combines four complementary learning models that analyse network traffic from different perspectives. Rather than relying on a single classifier, the hybrid approach integrates supervised learning, anomaly detection, and deep learning to improve robustness against diverse attack scenarios.

Isolation Forest is employed as the primary unsupervised anomaly detection model. It identifies abnormal network behaviour by recursively isolating observations within randomly generated decision trees. Since malicious traffic generally requires fewer partitions for isolation than legitimate traffic, higher anomaly scores indicate a greater likelihood of intrusion. The generated anomaly scores are forwarded to the ensemble model rather than being used for final classification independently.

The Autoencoder model provides an additional anomaly detection mechanism by learning the normal characteristics of network traffic through reconstruction. During inference, malicious samples typically produce larger reconstruction errors than legitimate communication because they differ significantly from the learned representation of normal behaviour. These reconstruction errors are subsequently supplied to the ensemble model as anomaly scores.

The XGBoost classifier serves as the primary supervised learning component of the framework. By employing gradient boosting over multiple decision trees, XGBoost effectively captures complex nonlinear relationships within network traffic and predicts the probability that an observed communication corresponds to malicious activity. Owing to its high classification accuracy, XGBoost contributes significantly to the final ensemble prediction.

To capture temporal dependencies present within network traffic, the framework further incorporates a Bi-LSTM network. Unlike conventional machine learning algorithms, the Bi-LSTM processes communication sequences in both forward and backward directions, enabling improved detection of sequential attack behaviour frequently observed in modern cyberthreats. The probability generated by the Bi-LSTM model is also utilised as an input to the ensemble learning stage.

Stacked Ensemble Learning

To improve intrusion detection performance, the proposed framework employs a stacked ensemble learning strategy that integrates the predictions generated by the four base detection models: Isolation Forest, Autoencoder, XGBoost, and Bi-LSTM. Rather than relying on a single classifier, the ensemble combines the strengths of supervised learning, anomaly detection, and deep learning to produce a more reliable final prediction.

Each base model independently analyses the same network traffic sample and generates either an anomaly score or an attack probability. These outputs are combined to form the ensemble feature vector:

$$E(x) = [S_{\mathrm{IF}}, S_{\mathrm{AE}}, P_{\mathrm{XGB}}, P_{\mathrm{BiLSTM}}] \quad (1)$$

where S_{IF} and S_{AE} represent the anomaly scores generated by the Isolation Forest and Autoencoder models, respectively, while P_{XGB} and P_{BiLSTM} denote the malicious traffic probabilities predicted by the XGBoost and Bi-LSTM models.

Before ensemble training, the prediction scores are standardised using StandardScaler to eliminate differences in numerical scale. The standardised outputs are then supplied to a Logistic Regression meta-classifier, which learns the optimal combination of model predictions from the training data. The resulting probability represents the final threat score used for intrusion detection.

1. Logistic Regression-Based Ensemble Prediction: The probability generated by the Logistic Regression meta-classifier is calculated using the logistic sigmoid function:

$$P(y=1|E)=\frac{1}{1+e^{-z}} \tag{2}$$

where

$$z=\beta_0+\beta_1 S_{\mathrm{IF}}+\beta_2 S_{\mathrm{AE}}+\beta_3 P_{\mathrm{XGB}}+\beta_4 P_{\mathrm{BiLSTM}} \tag{3}$$

where $P(y=1|E)P(y=1|E)P(y=1|E)$ represents the probability that the analysed network traffic is malicious, S_{IF} and S_{AE} denote the anomaly scores produced by the Isolation Forest and Autoencoder models, respectively, while P_{XGB} and P_{BiLSTM} represent the attack probabilities predicted by the XGBoost and Bi-LSTM models. The coefficients β_0, β_1, β_2, β_3, and β_4 are automatically learned during the training process, enabling the Logistic Regression meta-classifier to determine the optimal contribution of each base detection model when producing the final intrusion decision.

2. Threat Score Computation: The probability generated by the Logistic Regression meta-classifier is interpreted as the final threat score, which represents the likelihood that a network traffic instance is malicious. The threat score is defined as

$$T(x)=P(y=1|E)(3)T(x)=P(y=1|E)T(x)=P(y=1|E) \tag{4}$$

where $T(x)$ denotes the final threat score, E represents the ensemble feature vector, and $P(y=1|E)P(y=1|E)P(y=1|E)$ is the probability estimated by the Logistic Regression meta-classifier. The computed threat score is subsequently evaluated against predefined decision thresholds to determine whether the network traffic should be allowed, flagged as suspicious, or blocked by the automated prevention layer.

Compared with conventional voting-based ensemble methods, the proposed stacking approach improves generalisation capability by learning the relative contribution of each base learner automatically, thereby enhancing overall detection accuracy and reducing false-positive rates. This ensemble prediction is subsequently forwarded to the automated prevention layer for security policy evaluation.

Automated Prevention Layer

Following intrusion detection, the proposed framework automatically initiates appropriate mitigation actions based on the threat probability generated by the stacked ensemble model. This prevention layer enables the framework to function as an IDPS rather than a conventional IDS.

The proposed framework supports three security actions:

Allow: Traffic with a low threat probability is permitted to continue normal communication.

Alert: Suspicious traffic generates security notifications for administrator review.

Block: Traffic identified as malicious is automatically blocked by adding the corresponding source IP address to the dynamic blocklist.

The prevention module continuously records blocked IP addresses, generated alerts, and detection events for future analysis. In addition, an adaptive threshold mechanism is employed to monitor detection performance and adjust operational thresholds when necessary, thereby maintaining an effective balance between attack detection and false-

positive reduction. These automated mitigation strategies minimise administrator intervention while improving the overall security of IoT networks.

Real-Time Monitoring Dashboard

To facilitate continuous monitoring of network security, the proposed framework incorporates a Flask-based web dashboard that provides real-time visualisation of intrusion detection and prevention activities. The dashboard serves as an interface through which security administrators can observe system performance and analyse detected cyberthreats.

The dashboard displays key performance indicators, including overall detection accuracy, precision, recall, F1-score, confusion matrix, blocked IP addresses, generated alerts, and comparative performance of the individual detection models. In addition, feature importance visualisation is provided to improve the interpretability of the machine learning predictions and assist administrators in understanding the factors influencing intrusion detection decisions.

By integrating real-time monitoring with automated prevention, the dashboard improves situational awareness and enables efficient management of cybersecurity events within IoT environments. The graphical interface also supports rapid evaluation of model performance and facilitates timely response to emerging cyberthreats.

Results And Discussion

Experimental setup

The proposed hybrid IDPS was implemented using Python 3.9 on a workstation equipped with an Intel Core i7 processor, 16 GB RAM, and an NVIDIA GTX 1650 GPU running Ubuntu 20.04. The implementation utilised Scikit-learn, TensorFlow/Keras, and XGBoost libraries for machine learning, deep learning, and ensemble model development [11].

Performance evaluation was conducted using the UNSW-NB15 benchmark dataset, which contains both legitimate and malicious network traffic representing multiple contemporary cyberattack categories, including Exploits, DoS, Generic, Reconnaissance, Analysis, Backdoors, Shellcode, Fuzzers, and Worms [15]. Before training, duplicate records were removed, categorical attributes were encoded using Label Encoding, numerical features were normalised using Min-Max Scaling, and class imbalance was addressed using SMOTE. The dataset was partitioned into 80% training and 20% testing subsets to ensure consistent evaluation across all models.

Isolation Forest, Autoencoder, XGBoost, and Bi-LSTM models were independently trained using identical training and testing partitions. Their prediction scores were subsequently combined using a Logistic Regression-based stacked ensemble model, which generated the final threat probability used for intrusion detection and automated prevention. This experimental configuration ensured a fair comparison among the individual detection models while demonstrating the effectiveness of the proposed ensemble strategy.

Model performance evaluation

The effectiveness of the proposed framework was evaluated using standard classification metrics, including Precision, Recall, F1-score, and ROC-AUC. These metrics provide complementary information regarding the capability of each model to correctly distinguish malicious network traffic from legitimate communication while minimising false-positive predictions [9].

Table *1* shows the confusion matrix results obtained for each model during the evaluation process. The confusion matrix provides information about the number of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN), which are important indicators of detection performance in cybersecurity applications.

Models	True Positives	False Positives	True Negatives	False Negatives
Isolation Forest	260	95	85	510
Autoencoder	275	80	70	525
XGBoost	315	55	45	585
Bi-LSTM	300	65	50	570

TABLE 1: Confusion Matrix for IoT Cyberattack Detection Models

Bi-LSTM, Bidirectional Long Short-Term Memory; IoT, Internet of Things

The confusion matrix illustrates the number of TP, FP, TN, and FN generated by each detection model. TP represent correctly identified attacks, whereas FP correspond to legitimate traffic incorrectly classified as malicious. Similarly, FN indicate attacks that were not detected, while TN represent correctly classified legitimate traffic. Together, these values form the basis for calculating Precision, Recall, and F1-score, thereby providing a comprehensive assessment of model performance.

Among the evaluated models, XGBoost achieved the highest number of correctly detected attacks while simultaneously maintaining the lowest number of false-positive predictions. This behaviour demonstrates the model's ability to effectively distinguish between malicious and legitimate network traffic. The Bi-LSTM model also achieved competitive detection performance owing to its capability to capture temporal dependencies within sequential network traffic. In contrast, Isolation Forest and Autoencoder exhibited comparatively higher false-positive rates because anomaly detection models intentionally prioritise sensitivity to previously unseen attack behaviour.

ROC-AUC analysis

ROC analysis was performed to evaluate the discrimination capability of each detection model across varying classification thresholds. Figure *2* presents the ROC curve comparison of the evaluated models.

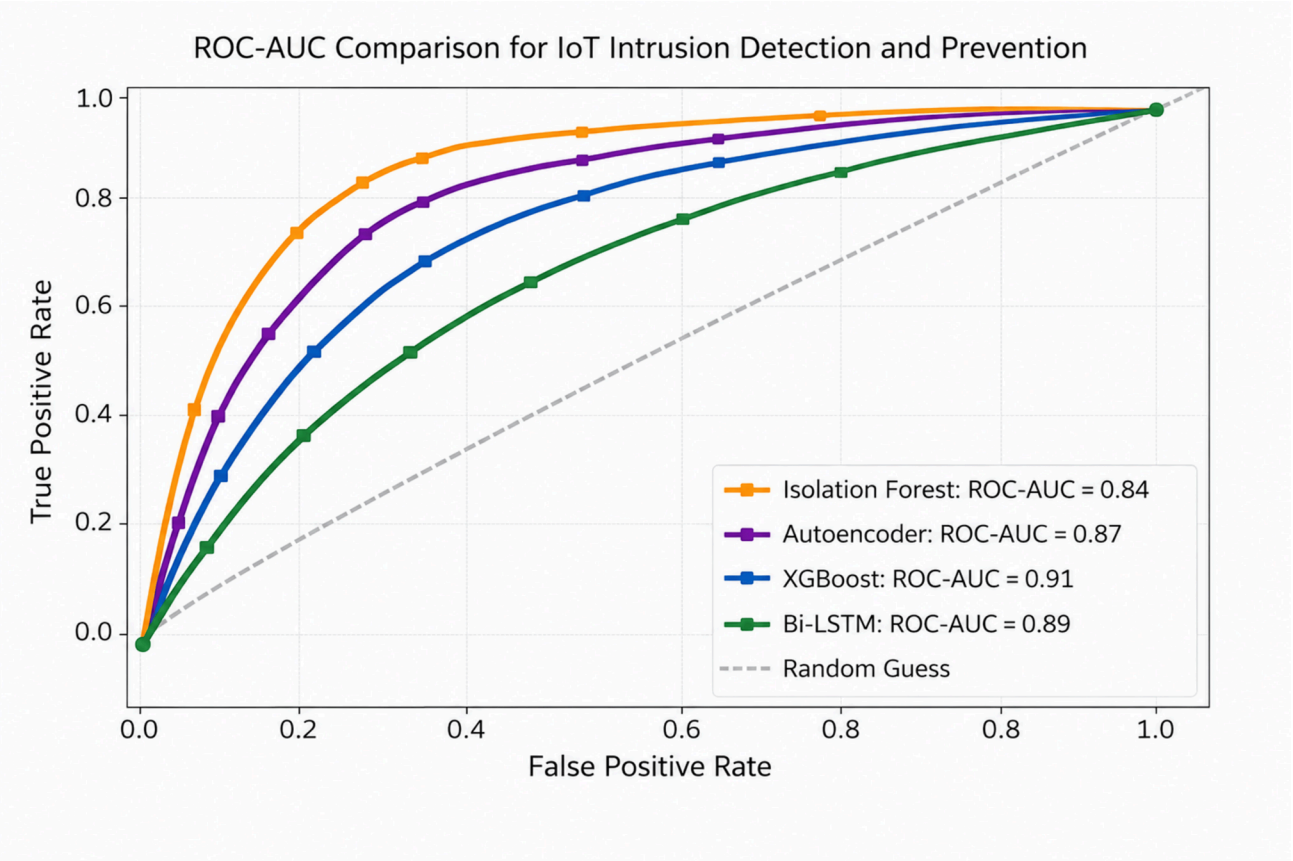

FIGURE 2: ROC Curve Comparison of the Evaluated Models

Bi-LSTM, Bidirectional Long Short-Term Memory; IoT, Internet of Things; ROC-AUC, Receiver Operating Characteristic - Area Under the Curve

The ROC-AUC values obtained during experimentation are summarised below:

Isolation Forest - 0.84

Autoencoder - 0.87

XGBoost - 0.91

Bi-LSTM - 0.89

The results demonstrate that XGBoost achieved the highest ROC-AUC score of 0.91, indicating superior classification capability compared with the remaining individual detection models. The Bi-LSTM model achieved an ROC-AUC value of 0.89, reflecting its effectiveness in identifying sequential attack patterns commonly observed within IoT communication. Autoencoder and Isolation Forest also demonstrated strong anomaly detection capabilities, particularly for identifying traffic that deviated significantly from normal communication patterns.

The ROC analysis confirms that supervised learning performs exceptionally well for known attacks, whereas anomaly detection models provide complementary capabilities for identifying previously unseen cyberthreats. This observation further justifies the adoption of the proposed hybrid detection framework.

Comparative analysis of individual detection models

Models	Precision	Recall	F1-Score	ROC-AUC
Isolation Forest	0.73	0.75	0.74	0.84
Autoencoder	0.78	0.80	0.79	0.87
XGBoost	0.88	0.90	0.89	0.91
Bi-LSTM	0.85	0.86	0.85	0.89

TABLE 2: Comparative Performance of Detection Models

Bi-LSTM, Bidirectional Long Short-Term Memory; ROC-AUC, Receiver Operating Characteristic - Area Under the Curve

Table *2* compares the Precision, Recall, F1-score, and ROC-AUC obtained by each individual detection model. Among all evaluated classifiers, XGBoost achieved the highest Precision (0.88), Recall (0.90), F1-score (0.89), and ROC-AUC (0.91), confirming its effectiveness for supervised intrusion detection. The Bi-LSTM network also demonstrated strong overall performance owing to its ability to capture temporal characteristics of network traffic. Isolation Forest and Autoencoder achieved comparatively lower Precision and F1-score values but contributed valuable anomaly detection capabilities by recognising deviations from legitimate communication patterns.

Although the individual models demonstrate satisfactory performance independently, each algorithm exhibits specific strengths and limitations. Consequently, integrating these heterogeneous prediction models through stacked ensemble learning provides a more robust intrusion detection framework capable of simultaneously identifying both known and previously unseen attacks.

Performance of the proposed stacked ensemble

Unlike conventional IDSs that rely on a single classifier, the proposed framework combines the outputs of Isolation Forest, Autoencoder, XGBoost, and Bi-LSTM using a Logistic Regression meta-classifier. The ensemble model learns the optimal contribution of each detector, thereby reducing model-specific prediction errors while improving overall detection reliability.

Although the individual detection models were evaluated separately to analyse their respective contributions, the final intrusion decision is generated by the stacked ensemble framework through the integration of their outputs using a Logistic Regression meta-classifier. Consequently, the overall performance of the proposed framework is determined by the collective behaviour of the integrated detection architecture rather than the performance of any individual model alone. By combining the complementary strengths of supervised learning, anomaly detection, and deep learning, the stacked ensemble improves the robustness and consistency of intrusion detection while reducing model-specific limitations. These findings demonstrate the effectiveness of the proposed hybrid framework for securing IoT environments against both known and previously unseen cyberthreats.

Discussion

The experimental results demonstrate that integrating multiple machine learning paradigms within a stacked ensemble architecture significantly improves intrusion detection capability for IoT environments. XGBoost achieved the highest individual classification performance, while the Autoencoder and Isolation Forest contributed valuable anomaly

detection capabilities for identifying previously unseen attack behaviour. The Bi-LSTM model further enhanced detection by learning temporal relationships present within sequential network traffic.

One of the primary strengths of the proposed framework lies in its ability to integrate supervised learning, anomaly detection, and deep learning into a unified decision-making architecture. Rather than depending exclusively on attack signatures or labelled datasets, the hybrid framework combines complementary prediction mechanisms to improve robustness against evolving cyberthreats.

The implementation further incorporates an automated prevention module capable of generating security alerts, dynamically blocking malicious IP addresses, maintaining a blocklist, and visualising security events through a real-time Flask dashboard. These capabilities demonstrate the practical applicability of the proposed framework beyond traditional intrusion detection by providing an integrated intrusion detection and prevention solution.

Despite these promising results, several limitations should be acknowledged. The framework was evaluated using the UNSW-NB15 benchmark dataset under controlled experimental conditions rather than in a live production IoT environment. Furthermore, although the proposed implementation demonstrates real-time monitoring and automated prevention, comprehensive evaluation of detection latency, mitigation delay, computational overhead, and attack-wise detection performance remains an important direction for future work. Lightweight deployment strategies based on edge computing and federated learning may further improve scalability for large-scale IoT ecosystems.

Overall, the experimental findings indicate that the proposed hybrid machine learning framework provides an effective, scalable, and adaptive cybersecurity solution capable of improving intrusion detection performance while supporting intelligent, automated response mechanisms for modern IoT environments.

Conclusions

This study presented a hybrid IDPS for securing IoT environments against increasingly sophisticated cyberthreats. The proposed framework integrates supervised learning, unsupervised anomaly detection, and deep learning techniques by combining Isolation Forest, Autoencoder, XGBoost, and Bi-LSTM models within a stacked ensemble learning architecture. Unlike conventional IDSs that rely on a single detection approach, the proposed framework exploits the complementary strengths of multiple learning paradigms to improve intrusion detection performance while simultaneously supporting automated prevention. Experimental evaluation was conducted using the UNSW-NB15 benchmark dataset following preprocessing through duplicate removal, label encoding, feature normalisation, and SMOTE. The individual detection models demonstrated promising classification performance, with XGBoost achieving the highest ROC-AUC score of 0.91, followed by Bi-LSTM (0.89), Autoencoder (0.87), and Isolation Forest (0.84). These findings indicate that supervised learning provides excellent detection capability for known attacks, whereas anomaly detection models contribute to identifying previously unseen malicious behaviour. The stacked ensemble framework combines these complementary outputs to improve overall detection robustness and reduce model-specific prediction errors.

Beyond intrusion detection, the proposed framework incorporates an automated prevention layer capable of generating security alerts, dynamically blocking malicious IP addresses, maintaining a blocklist, and supporting real-time monitoring through a Flask-based dashboard. The integration of detection, prevention, and visualisation within a unified framework enhances the practical applicability of the proposed system for intelligent IoT cybersecurity environments. Although the proposed framework demonstrates encouraging performance, several limitations should be acknowledged. The experimental evaluation was conducted using the UNSW-NB15 benchmark dataset under controlled laboratory conditions, and additional validation using real-world IoT deployments would further strengthen the practical applicability of the framework. Furthermore, comprehensive evaluation of attack-wise detection performance, statistical significance analysis, computational overhead, detection latency, and mitigation response time would provide additional evidence regarding large-scale deployment capability. Future research will focus on extending the proposed framework through lightweight edge-based intrusion detection, federated learning for distributed IoT environments, adaptive online learning for continuously evolving cyberthreats, and Explainable AI techniques to improve transparency and

interpretability of ensemble decision-making. Incorporating recent state-of-the-art deep learning architectures and evaluating the framework across multiple benchmark datasets will further enhance its scalability, robustness, and applicability to next-generation IoT cybersecurity systems.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Ruthwik Palem, Likhith Reddy Peketi, Karthika J, Vanathi M, Sandhiya B

Acquisition, analysis, or interpretation of data: Ruthwik Palem, Likhith Reddy Peketi, Karthika J, Vanathi M, Sandhiya B

Drafting of the manuscript: Ruthwik Palem, Likhith Reddy Peketi, Karthika J, Vanathi M, Sandhiya B

Critical review of the manuscript for important intellectual content: Ruthwik Palem, Likhith Reddy Peketi, Karthika J, Vanathi M, Sandhiya B

Supervision: Vanathi M

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request. The datasets supporting this study are publicly available in the UNSW-NB15 dataset at https://research.unsw.edu.au/projects/unsw-nb15-dataset.

Acknowledgements

The authors thank the Department of Computer Science and Engineering at Sathyabama Institute of Science and Technology for supporting this research.

References

1. Shone N, Ngoc TN, Phai VD, Shi Q: A deep learning approach to network intrusion detection. IEEE Transactions on Emerging Topics in Computational Intelligence. 2018, 2:41-50. 10.1109/tetci.2017.2772792
2. Bhuyan MH, Bhattacharyya DK, Kalita JK: Network anomaly detection: methods, systems and tools. IEEE Communications Surveys & Tutorials. 2014, 16:303-336. 10.1109/surv.2013.052213.00046
3. Pajouh HH, Javidan R, Khayami R, Dehghantanha A, Choo KKR: A two-layer dimension reduction and two-tier classification model for anomaly-based intrusion detection in IoT backbone networks. IEEE Transactions on Emerging Topics in Computing. 2019, 7:314-323. 10.1109/tetc.2016.2633228
4. Kotenko I, Saenko I, Branitskiy A: Framework for mobile Internet of Things security monitoring based on big data processing and machine learning. IEEE Access. 2018, 6:72714-72723. 10.1109/access.2018.2881998

5. Ferrag MA, Maglaras L, Ahmim A, Derdour M, Janicke H: RDTIDS: Rules and decision tree-based intrusion detection system for Internet-of-Things networks. Future Internet. 2020, 12:44. 10.3390/fi12030044

6. Buczak AL, Guven E: A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials. 2016, 18:1153-1176. 10.1109/comst.2015.2494502

7. Yin C, Zhu Y, Fei J, He X: A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access. 2017, 5:21954-21961. 10.1109/access.2017.2762418

8. Diro AA, Chilamkurti N: Distributed attack detection scheme using deep learning approach for Internet of Things. Future Generation Computer Systems. 2018, 82:761-768. 10.1016/j.future.2017.08.043

9. Tama BA, Nkenyereye L, Islam SMR, Kwak KS: An enhanced anomaly detection in web traffic using a stack of classifier ensemble. IEEE Access. 2020, 8:24120-24134. 10.1109/access.2020.2969428

10. Mirsky Y, Doitshman T, Elovici Y, Shabtai A: Kitsune: An ensemble of autoencoders for online network intrusion detection. Proceedings of the Network and Distributed System Security Symposium (NDSS). 2018, 1-15. 10.14722/ndss.2018.23204

11. Chen T, Guestrin C: XGBoost: A scalable tree boosting system. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016, 785-794. 10.1145/2939672.2939785

12. Karthika J, Loganathan S, Vanathi M: A hybrid machine learning based feature selection technique for attack detection in NIDS. Journal of Physics: Conference Series. 2022, 2335:012033. 10.1088/1742-6596/2335/1/012033

13. Meidan Y, Bohadana M, Mathov Y, Mirsky Y, Shabtai A, Breitenbacher D, Elovici Y: N-BaIoT—network-based detection of IoT botnet attacks using deep autoencoders. IEEE Pervasive Computing. 2018, 17:12-22. 10.1109/mprv.2018.03367731

14. Lundberg SM, Lee S-I: A unified approach to interpreting model predictions. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017, 4768-4777.

15. Moustafa N, Slay J: UNSW-NB15: A comprehensive data set for network intrusion detection systems. 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, ACT, Australia. 2015, 1-6. 10.1109/MilCIS.2015.7348942

How to cite this article:
Palem R, Peketi L, M V, et al. (August 10, 2026) Detecting and Preventing Cyberattacks in Internet of Things (IoT) Systems.
Cureus J Comput Sci 3 : es44389-026-00214-8. DOI https://doi.org/10.7759/s44389-026-00214-8