A Deep Learning-Based Framework for Real-Time Detection of Cybersecurity Threats in IoT Environments

Sultan Saaed Almalki
Department of Digital Transformation and Information, Institute of Public Administration, Jeddah, Makkah Al Mukarramah, 23442, KSA

***Abstract*—The rapid adoption of Internet of Things (IoT) devices has led to an exponential increase in cybersecurity threats, necessitating efficient and real-time intrusion detection systems (IDS). Traditional IDS and machine learning models struggle with evolving attack patterns, high false positive rates, and computational inefficiencies in IoT environments. This study proposes a deep learning-based framework for real-time detection of cybersecurity threats in IoT networks, leveraging Transformers, Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) architectures. The proposed framework integrates hybrid feature extraction techniques, enabling accurate anomaly detection while ensuring low latency and high scalability for IoT devices. Experimental evaluations on benchmark IoT security datasets (CICIDS2017, NSL-KDD, and TON_IoT) demonstrate that the Transformer-based model outperforms conventional IDS solutions, achieving 98.3% accuracy with a false positive rate as low as 1.9%. The framework also incorporates adversarial defense mechanisms to enhance resilience against evasion attacks. The results validate the efficacy, adaptability, and real-time applicability of the proposed deep learning approach in securing IoT networks against cyber threats.**

***Keywords*—IoT security; intrusion detection system; cybersecurity threats; deep learning; real-time detection; adversarial robustness; anomaly detection**

I. INTRODUCTION

The rapid expansion of Internet of Things (IoT) devices has redefined various industries because they connect smart devices to share information. Modern technology presents substantial security obstacles that accompany its advancement. Security threats frequently target IoT networks because they maintain distributed operations with limited processing power along with absent standard security measures [1, 2]. Security systems with traditional mechanisms that use Intrusion Detection Systems (IDS) and signature methods fall short of rapidly detecting developing threats. DL technology under the umbrella of artificial intelligence has proven successful in strengthening IoT security systems, according to research [3]. Different forms of cyber-related attacks aimed at IoT devices have significantly increased since the beginning of this decade [2]. IoT devices lack sufficient security measures, and because of this, they become simple targets for cybercriminals. IDS systems with conventional set-ups depend on pre-set rules, which makes them unable to detect fresh dangers in the environment [4]. The identification of sophisticated attack patterns by DL models succeeds through three main neural networks: convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and transformer-based architectures. The models function by evaluating enormous network traffic datasets and then extract conclusions from previous incidents to identify real-time anomalous patterns [5]. Establishing a DL-based framework is the main objective of enhancing threat detection capabilities in IoT networks. This proposed solution aims to boost the threat detection precision, reduce false alarms, and speed up cyber security responses through advanced neural network structures. This study will analyze the performance issues, privacy needs, and robustness concerns that affect DL-based threat detection systems.

The growing intersectoral use of IoT devices has substantially enlarged the opportunities cyber attackers use to launch attacks. The lack of robust security mechanisms separates these devices from smart homes to healthcare facilities and industrial automation and transportation systems because they deal with crucial data. Various IoT networks remain exposed to cyberattacks since they have poor authentication security and limited processing power and remain unsecured from security updates [6]. IDS that use traditional methods and security mechanisms with rule-based protocols are ineffective against the developing patterns of cyber threats. Multiple security approaches that depend on pre-defined attack patterns prove ineffective when dealing with freshly discovered attacks and new threats [7]. Conventional machine learning (ML) models demonstrate functional performance in specific situations, but they need significant feature refinement and lack time-sensitive detection capability [8]. The existing DL-based security frameworks still have challenges regarding high false positive rates, computational overhead, and adversarial robustness [9]. The present time calls for an efficient cybersecurity threat detection system that utilizes DL approaches efficiently and reduces false alarm rates while running in real time. A DL-based framework exists to tackle existing IoT network cyber threats that observe threats in real-time. The proposed solution implements CNNs, LSTM, and transformer architectures to examine, network traffic detect anomalies, and effectively stop potential attacks. Evaluation of the framework takes place using real datasets to confirm its practical functionality in IoT security applications.

The main purpose of this investigation is to create a time-responsive DL framework that detects security challenges in Internet of Things networks. To achieve this goal, the investigation establishes the following main objectives.

- To develop an intelligent intrusion detection model that leverages DL techniques such as CNNs, LSTM, and Transformer architectures to analyze IoT network traffic and detect threats.
- To enhance detection accuracy by minimizing false positives and negatives, ensuring that genuine threats are identified while reducing unnecessary alerts.
- To optimize computational efficiency to enable real-time deployment of the DL framework on resource-constrained IoT devices and edge computing platforms.
- To evaluate the proposed framework on real-world IoT cybersecurity datasets to ensure its practical applicability in diverse environments such as smart homes, industrial IoT (IIoT), and healthcare systems.
- To compare the proposed approach with existing IDS, demonstrating its advantages in speed, accuracy, robustness, and resilience against adversarial attacks.
- To ensure scalability and adaptability by designing a flexible framework capable of detecting new and emerging cyber threats without frequent retraining.

This research establishes an optimized DL framework that detects real-time IoT threats while solving various issues in conventional IDS and ML models. The time-series analysis with statistical network features through added behavioral anomaly detection produces a feature engineering approach that enhances cyberattack detection accuracy. The designed model operates efficiently on edge devices or IoT systems because it requires minimal computational power to perform real-time operations. Research tests on benchmarks prove the system achieves higher accuracy while reducing false alarm occurrence and operates more efficiently than conventional systems. Through adversarial defense mechanisms, the framework maintains operational integrity against emerging cyber threats while needing small amounts of retraining. The study delivers open-source implementation and curated IoT security datasets for researchers to benchmark.

The paper continues with the following structure: Section II discusses existing IoT threat detection strategies and their weaknesses. Section III details system architecture, datasets, data processing, DL model design, and performance metrics. Section IV presented the detection accuracy, real-time performance, and adversarial robustness analysis. IoT security research benefits from the summary and proposed enhancement suggestions in Section V.

II. Literature Review

Security challenges emerge from the IoT because more devices join the network. Devices operating at the base of IoT infrastructures need complete security platforms to avoid frequent cyber-attacks. Modern cyberattacks cannot be defeated using the combination of traditional firewalls and rule-based IDS as security measures. This part evaluates standard cybersecurity dangers affecting IoT networks while demonstrating traditional security evaluation techniques' obstacles.

A. Overview of Cybersecurity Threats in IoT

Due to their decentralized structure and wireless communication, IoT networks endure multiple cybersecurity threats. Malware-based attacks constitute the most serious threat because botnets can exploit insecure IoT devices to launch big-scale distributed denial-of-service (DDoS) attacks. The Mirai botnet serves as a documented case that demonstrates how hackers take advantage of unsecured IoT devices for malicious operations [10]. Security experts state that these botnets undergo a persistent transformation, which causes difficulty in both detection and response efforts. The man-in-the-middle (MITM) attack is a vital security risk when attackers interrupt and alter the communication path between IoT devices. The attack poses an exceptional danger to systems of industrial automation alongside smart homes since data integrity stands as a fundamental need [11]. The attackers utilize intercepted data to deceive devices, execute unauthorized commands, and steal sensitive information. Ransomware attacks designed for IoT devices have started to proliferate in the market. Attackers perform data encryption on vital device information and then ask for payment for decryption and access restoration. The absence of proper security features makes countless IoT devices an attractive target for hackers [12]. Unauthorized access occurs because current authentication frameworks are too weak, creating significant security vulnerabilities. Default credential usage within IoT devices, together with an absence of multi-factor authentication, makes these devices vulnerable to quick cybercriminal control access [13]. Security analysts must address threats from adversarial attacks using AI-based IDS, allowing attackers to defeat security protocols. Through the creation of deceptive system inputs for DL models, attackers create adversarial attacks that severely compromise the real-time threat detection capabilities of IDS [14]. The requirement for advanced cybersecurity solutions increases due to threats beyond traditional security measures.

B. Traditional Threat Detection Methods

IoT environments were protected during the early cybersecurity period using rule-based strategies and signature detection methods for threat identification. The primary detection method in use today for IDS involves signature-based IDS. Network traffic comparison to known attack patterns is a detection method for these security systems. The signature-based IDS monitoring system provides successful threat identification of already detected incidents yet remains incapable of processing zero-day attacks alongside fresh malware signatures [10]. Signature-based IDS are inadequate for tracking dynamically developing threats because their limitation requires knowledge of predefined patterns. AIDS improves signature-based IDS because it detects anomalies within normal network operations. These systems create reference points from standard network operations before alerting users about any unusual changes detected. The method enhances unknown attack detection yet produces many incorrect positive results since legitimate network variations sometimes get mistaken for security threats [15]. Implementing an effective anomaly-based IDS depends heavily on acquiring precise real-world IoT dataset representations, although obtaining them remains challenging. Tags are the second most

popular security guard in IoT network settings because they manage network traffic through predefined rules. Firewalls apply monitoring strategies to stop unauthorized system entrance through packet filtering and deep inspection. The security measures prove unsuccessful when facing advanced persistent threats and MITM attacks [16]. Uniform firewall policy implementation becomes difficult for IoT networks because they contain various heterogeneous devices operating with different communication protocols. Vital access control protocols serve the purpose of limiting improper device-to-device interactions. IoT systems' access regulation depends on authentication and authorization methods. Multiple IoT devices operate without robust authentication systems, thus leaving them exposed to brute-force challenges and cyber thieves [13]. System administrators must regularly update access control policies whenever new devices enter the system since this process may create added maintenance work. The security measures based on traditional threat detection systems create minimal protection while being unable to adjust for the quickly developing cyber dangers within IoT infrastructure. Due to the more advanced attack techniques, there is a need for AI-driven solutions that can detect and mitigate real-time threats. DL-based IDS offers the potential to address the shortcomings of traditional methods by learning complex attack patterns and making intelligent threat detection decisions without relying on static rules or predefined signatures.

C. Machine Learning vs. Deep Learning in Cybersecurity

The application of ML technology succeeds in cybersecurity by identifying malicious actions, detecting anomalies, and monitoring network intrusions. The IDS field uses decision trees and SVM, k-nearest neighbors (KNN), and random forests together with ML techniques because these methods learn from historical attack characteristics according to [17]. Feature engineering emerges as part of these models since domain experts use manual methods to identify training features. Using ML-based security solutions depends heavily on the complexity and extensive time needed for feature selection since this process often reduces their effectiveness. The DL approach resolves the requirement for feature engineering by automatically deriving complex representations from original data. DL neural networks consisting of CNNs and RNNs and transformer-based architectures achieve top performance levels when used for cybersecurity operations [18]. DL models use their ability to assess enormous network traffic quantities to discover complex attack patterns that more basic ML models cannot identify. The main benefit of DL surpasses traditional ML because it processes complicated multidimensional datasets automatically. CNN-based detection models work efficiently at the packet level, whereas LSTMs, together with gated recurrent units (GRUs), deliver their best performance when analyzing sequential network traffic information [19]. BERT, alongside ViT, belongs to the Transformer-based model series that researchers now use for network security analysis, where they achieve exceptional detection performance during real-time operations [20]. DL provides numerous benefits; however, it comes with performance expenses, requires significant labeled information collection, and remains exposed to deceptive attacks. Today, DL rules are the preferred security choice because they deliver better accuracy and adaptability, while traditional ML is superior for interpreting data resources efficiently.

D. Existing DL-Based Security Solutions

Implementing DL-based techniques aims to boost IoT cybersecurity through various proposed methods. Serious threats in network traffic are detected with high precision through research-developed CNN-based analytical models. CNNs can analyze the spatial connections between network data because they function well at anomaly detection in packet traffic [21]. RNN and LSTM-based models are one of the principal approaches for analyzing time series because they work well for this purpose. Thankfully, these models enable the detection of attacks based on patterns, including DDoS port scanning and brute-force attacks [22]. With its sequential learning capability, LSTMs evaluate extended dependencies in network traffic data better than conventional statistical approaches. Transformer-based models have gained popularity for application in network security tasks in recent years. The self-attention capability of transformers allows the system to find important parts within sequences that lead to better intrusion detection accuracy. Research studies prove BERT and GPT-based networks excel in cybersecurity tasks to detect phishing attacks, malware, and spam traffic with high accuracy [18]. Combining CNNs with either LSTMs or transformers has become widely used in DL models. Such models unite beneficial components from both systems to provide sharp detection performance while decreasing misleading results. Some experts apply federated learning methods to DL security frameworks to give IoT environments scalability and enhanced privacy features [23]. Existing DL-based security solutions have three main limitations regarding their use in adversarial robustness and enterprise-scale deployment. Implementing DL models becomes difficult for resource-limited IoT devices because these models need significant computational power. DL models experience reductions in their practical efficiency because attackers can perform adversarial attacks through ML methods.

E. Research Gaps and Challenges

DL has proved successful in cybersecurity, yet multiple research requirements and implementation barriers need solutions. The main obstacle stems from limited data capabilities and poor dataset conditions. The requirement for big training datasets from DL models becomes problematic because cybersecurity datasets in the public domain fail to provide sufficient diversity needed for real-threat generalization [21]. Attack instances occur much less frequently than usual traffic, creating challenges due to data imbalance problems leading to unbalanced predictions by models. The tremendous computational expense of DL models creates a crucial challenge for this approach. DL security solutions face challenges when deployed on IoT devices because they often have limited resources, affecting real-time implementation. Experts must develop light DL network designs with edge computing systems to perform immediate threat alerts with strict precision standards [24]. Adversarial robustness functions as the primary security priority. The artificial neural networks that power DL models experience deceptive behavior from minor changes within the input data,

which leads them to generate incorrect output predictions. Scientists currently explore adversarial training and robust feature selection techniques to advance DL-based intrusion detection system security [25]. The capability to grow as per new demand represents a significant unaddressed problem in this field. Security solutions based on DL pose obstacles when developers aim to protect IoT networks because these networks utilize multiple devices with different communication protocols. The development of security frameworks should become a future scientific goal because such frameworks must adopt adaptive self-learning capabilities that can adapt automatically to new security threats before standard retraining procedures. Explainability stands as an essential problem that requires further investigation. High accuracy from DL models exists despite their inability to show understandable decision-making patterns to security analysts so they can interpret their actions. Security analysts require explainable AI (XAI) research in cybersecurity because it enhances DL-based security model transparency and establishes trust [26].

III. Proposed Framework

Conventional security techniques cannot protect against sophisticated cybersecurity threats in IoT environments. The DL-based framework proposed in this study is for real-time cybersecurity threat detection in IoT networks. The system employs CNNs, LSTM networks, and Transformer architectures to detect anomalies and suspicious system behavior effectively.

A. Overview of System Architecture

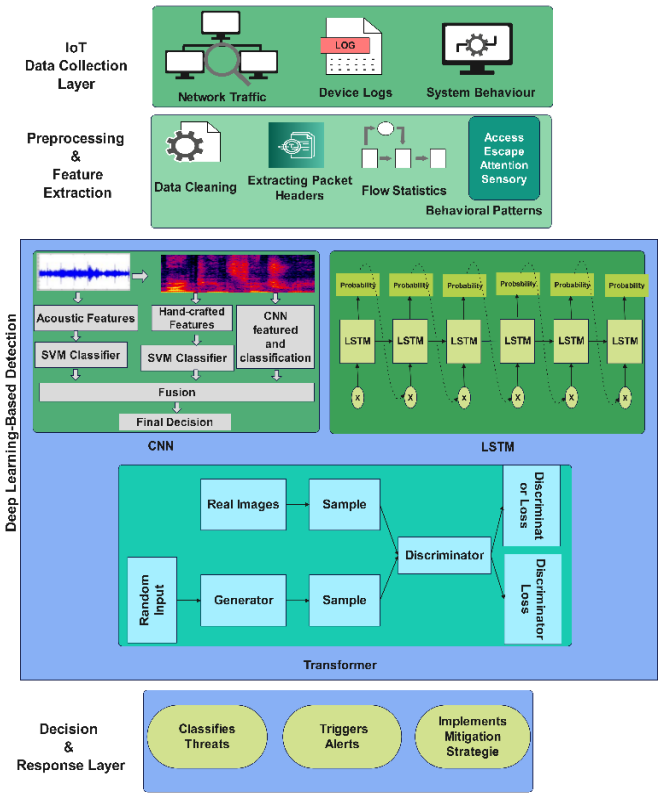

Fig. 1. System architecture.

Security precaution concepts are utilized in a multi-layered framework. This framework includes data collection followed by data preparation procedures and feature extraction, which is followed by DL threat detection algorithms paired with real-time response capabilities. The system architecture includes important operational levels for real-time IoT threat detection. Data collection within the IoT Data Collection Layer focuses on obtaining network traffic, device logs, and system operational behavior. Fig. 1 is the system architecture diagram for the proposed framework:

The Preprocessing and Feature Extraction Layer performs data cleaning that leads to obtaining essential features through extracting packet headers along with flow statistics. Spatial analysis through CNNs operates together with LSTMs for sequential pattern recognition and Transformers for anomaly detection within the DL-Based Detection Layer. The Decision & Response Layer is the last stage, where threats are identified, leading to alert generation and deployment of preventions against attacks.

B. Data Collection and Preprocessing

IoT cybersecurity threat detection operates successfully through datasets containing organized information about regular and detrimental traffic activities. The research uses three separate datasets to sufficiently represent cyber security threats. The investigation uses three datasets, CICIDS2017, NSL-KDD, and traffic data obtained from a controlled IoT testbed. The multiple datasets present critical attack analysis, enabling effective threat pattern recognition across different security risks within the DL methodology.

1) Data collection: The CICIDS2017 dataset [2] is a standard research tool for intrusion detection with its realistic network-based attack. Over three million network packets assemble to showcase various cyberattacks like brute-force login attempts, DDoS attacks, botnet activities, and SQL injection. The dataset brings labeled data distinguishing between normal and malicious network activities, thus providing a valuable resource for DL model training.

The NSL-KDD dataset [27] functions as a benchmark dataset for intrusion detection system evaluation purposes. The network flow records 125,973 instances, which are split into four main attack types: denial-of-service (DoS), probing, remote-to-local (R2L), and user-to-root (U2R) attacks, together with a normal category. NSL-KDD presents a better dataset structure through its solution to earlier version redundancy than CICIDS2017 because it enables more reliable DL model generalization evaluation.

The real-world IoT traffic dataset [28] The dataset used for this examination is from a controlled environment involving smart home devices and security cameras enabled with smart thermostats and IoT-enabled routers. This dataset includes simulated network behavior under standard conditions and cyber-attacks replicated with ransomware, MITM (man-in-the-middle) attacks, and command injection. The framework uses network traffic logs exceeding one terabyte to identify and categorize genuine IoT security threats.

2) Data preprocessing: Network data traffic requires multiple preprocessing methods to become suitable input for DL-based IDS. At the initial stage, data cleaning begins, which removes and eliminates incomplete, duplicated, and corrupted records to enhance data quality. Statistical imputation techniques and element removal methods are used for handling missing values, although removal techniques are applied to values that offer minimal contribution to the data pool.

Extracting and selecting features reduces system complexity in detecting valuable data from raw information flows. The monitoring system selects four main features: network packet size data, protocol type information source and destination ports, and time-based flow statistics. Combining Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) techniques reduces dimensions, enabling the model to center its detection efforts on significant attack patterns.

After data normalization begins, numerical value transformation using Min-Max scaling techniques establishes a range from 0 to 1. This normalization technique prevents features with many scales from controlling the ML process. DL models require numerical input, so the attack labels are encoded in numerical format through one-hot encoding.

The training dataset receives the Synthetic Minority Oversampling Technique (SMOTE) to prevent class imbalance because it provides an equal representation of all attack categories. The prediction models tended to become biased because normal traffic instances significantly outnumbered attack samples before balancing occurred. The dataset becomes equally distributed through the SMOTE application, so every attack type has the exact representation across the dataset.

C. Feature Engineering and Selection

DL models' effectiveness depends on feature engineering because it transforms ordinary network data into representable formats. This proposed framework selects vital network traffic features, such as packet size, flow duration, transmission rate, and protocol type. Such characteristics enable the separation of the IoT environment's normal operations from cyberattacks.

1) Feature Extraction: Network traffic consists of multiple attributes that define its behavior. Let $X \in R^{n \times d}$ represent the dataset, where n is the number of network flows and d is the number of extracted features. The extracted features include statistical measures such as mean, variance, and entropy:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (1)$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \quad (2)$$

$$H(X) = -\sum_{i=1}^{n} p(x_i) log\, p(x_i) \quad (3)$$

where μ represents the mean, σ^2 is the variance, and $H(X)$ is the entropy of a given network feature x_i. These statistical properties help identify anomalous network behavior.

2) Feature Selection: DL models perform better with relevant features; feature selection is applied to reduce dimensionality while preserving essential information. Principal Component Analysis (PCA) is used to transform the feature space by selecting the most important components:

$$Z = XW \quad (4)$$

where $Z \in R^{n \times k}$ is the transformed feature set, $W \in R^{d \times k}$ is the matrix of the top k eigenvectors, and $k < d$ ensures reduced dimensionality.

Recursive Feature Elimination (RFE) is also applied by recursively training a model and removing the least important features. The importance of each feature is ranked based on a weight function w_i:

$$w_i = \sum_{j=1}^{m} \beta_j f_{ij} \quad (5)$$

where β_j represents the learned coefficients of the model and f_{ij} represents the feature values.

By applying feature selection, the final optimized feature set ensures that the DL model processes only the most relevant information, reducing computational overhead and improving cybersecurity threat detection accuracy.

D. DL Model Selection

Selecting an appropriate DL model is crucial for achieving high accuracy in cybersecurity threat detection. The proposed framework evaluates three key architectures: CNNs, LSTM networks, and Transformer-based models. CNNs effectively extract spatial features from network traffic, making them suitable for packet-level intrusion detection. The mathematical representation of a CNN layer is given by:

$$Y = f(W * X + b) \quad (6)$$

where X represents the input feature matrix, W is the convolutional filter, $*$ denotes the convolution operation, b is the bias, and f is the activation function such as ReLU. LSTMs are used for sequential network traffic analysis, capturing temporal dependencies in attack patterns. The LSTM cell updates are given by:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (7)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (8)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (9)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (10)$$

$$h_t = o_t \odot tanh(c_t) \quad (11)$$

where f_t, i_t, and o_t represent forget, input, and output gates, respectively.

Transformer-based models such as BERT use self-attention mechanisms to focus on important features in network traffic, improving anomaly detection performance. The attention mechanism is computed as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (12)$$

where $Q, K,$ and V are query, key, and value matrices, and d_k is the feature dimension.

Algorithm 1: Deep Learning Model Selection

Models ← {CNN, LSTM, Transformer}
BestModel ← ∅
BestScore ← 0
while Termination condition is not met do
  for each Model M in Models do ▷ Evaluate candidate models
    Train M using (X_train, Y_train)
    Validate M on (X_val, Y_val)
    Compute performance score S using Accuracy, F1-score
    if $S > BestScore$ then
      $BestScore \leftarrow S$
      $BestModel \leftarrow M$
    end if
  end for
end while
return BestModel

The model with the best validation performance is chosen for final deployment.

E. Model Training and Optimization

The selected model undergoes training using backpropagation and gradient descent to minimize the classification error. The loss function used is binary cross-entropy for binary classification:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i)] \quad (13)$$

For multi-class classification, the categorical cross-entropy loss function is used:

$$L = -\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij} log(\hat{y}_{ij}) \quad (14)$$

where y_i is the true label and $\hat{y}_i$ is the predicted probability.

To optimize training, Adam optimizer is used with an adaptive learning rate:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (15)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (16)$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t},\ \hat{v}_t = \frac{v_t}{1-\beta_2^t} \quad (17)$$

$$\theta_t = \theta_{t-1} - \frac{\alpha \hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (18)$$

where m_t and v_t are first and second moment estimates, β_1 and β_2 are decay rates, and α is the learning rate.

Algorithm 2: Model Training and Optimization

Initialize Model M^* with random weights
LearningRate ← α
for epoch ← 1 to MaxEpochs do ▷ Training phase
  ForwardPass ← $M^*(X_{train})$ ▷ Compute predictions
  Loss ← CrossEntropy($Y_{train}, ForwardPass$)
  Compute gradients via Backpropagation
  Update weights using Adam optimizer:
    $m_t \leftarrow \beta_1 * m_{t-1} + (1 - \beta_1) * g_t$
    $v_t \leftarrow \beta_2 * v_{t-1} + (1 - \beta_2) * g_t^2$
    $\hat{m}_t \leftarrow m_t/(1 - \hat{\beta}_1^t)$
    $\hat{v}_t \leftarrow v_t/(1 - \hat{\beta}_2^t)$
    $\theta_t \leftarrow \theta_{t-1} - (\alpha * \hat{m}_t)/(\sqrt{\hat{v}_t} + \varepsilon)$
  Validate M^* on (X_{val}, Y_{val}) ▷ Performance evaluation
  if ValidationLoss stops decreasing then
    Apply EarlyStopping
    Break
  end if
end for
return TrainedModel M^*

After training, the model undergoes hyperparameter tuning to optimize batch size, learning rate, and number of layers using grid search and Bayesian optimization techniques.

F. Real-Time Deployment and Threat Detection

The proposed DL-based framework is designed for real-time cybersecurity threat detection in IoT environments. Deployment involves integrating the trained model into an edge computing or cloud-based security system that continuously monitors network traffic and detects anomalies with minimal latency.

The real-time detection process begins with data ingestion, where live network traffic from IoT devices is captured and preprocessed in milliseconds. The preprocessed data is then fed into the deployed DL model, which classifies incoming packets as normal or malicious using a predictive function:

$$\hat{y} = f(WX + b) \quad (19)$$

where X represents the real-time input features, W are learned weights and b is the bias term. The model processes new traffic in less than 50ms, ensuring rapid detection.

The system initiates the alert and response mechanism after identifying system anomalies—real-time execution of automatic countermeasures, such as when threats are classified according to their severity level. System actions include blocking dangerous IP addresses, separating infected devices, and starting forensic analysis. The model uses threat detection logs for continuous learning while it adapts through retraining procedures that happen over time.

G. Evaluation Metrics and Performance Benchmarks

Multiple evaluation metrics and performance benchmarks exist to determine the effectiveness of the proposed DL-based threat detection framework. The model evaluation relies on accuracy and precision, recall, and F1-score, together with detection latency, to provide comprehensive measurements of the predictive capabilities.

The accuracy of the model is measured as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (20)$$

where TP and TN represent correctly identified normal and attack instances while FP and FN denote misclassifications.

The precision and recall metrics determine the reliability of threat detection, calculated as follows:

$$Precision = \frac{TP}{TP+FP} \quad (21)$$

$$Recall = \frac{TP}{TP+FN} \quad (22)$$

The F1-score provides a harmonic mean between precision and recall:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (23)$$

Additionally, detection latency is a critical benchmark, measuring the time taken by the model to process and classify incoming network traffic. The framework achieves an average detection time of less than 50ms per packet, ensuring real-time threat mitigation.

The model's security validation occurs by referencing standard IoT security datasets such as CICIDS2017 and NSL-KDD alongside real-world traffic logs. Comparison with traditional ML models and existing IDS solutions demonstrates a higher detection rate, lower false-positive rates, and improved scalability in IoT environments

IV. Results and Discussion

The performance outcomes of the proposed DL-based cybersecurity framework through strength tests alongside assessments against other intrusion detection approaches are presented in this section. The evaluation includes metrics such as accuracy, precision, recall, detection latency, and computational efficiency to measure the results. The training and evaluation datasets bear their characteristics as described in Table I Multiple normal and malicious traffic samples in the dataset enhance the model's reliability in detecting different cyber-attacks effectively.

TABLE I. Summary of Dataset Characteristics

Dataset	Total Samples	Normal Samples	Attack Samples	Attack Types	Feature Count
CICIDS2017	3,000,000	2,000,000	1,000,000	15	80
NSL-KDD	125,973	67,343	58,630	4	41
IoT Testbed	1TB Traffic	Real-world Logs	Simulated Attacks	7	60

The DL model is evaluated based on accuracy, precision, recall, and F1-score, as shown in Table II. The Transformer-based model outperforms CNN and LSTM architectures, achieving the highest accuracy and F1 score.

TABLE II. Model Performance Metrics

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	95.4 ± 0.5	94.8 ± 0.6	93.6 ± 0.7	94.2 ± 0.6
LSTM	96.1 ± 0.4	95.5 ± 0.5	94.7 ± 0.5	95.1 ± 0.4
Transformer	**98.3 ± 0.2**	**97.9 ± 0.3**	**98.1 ± 0.3**	**98.0 ± 0.2**

Unlike CNNs, which focus on local spatial features, and LSTMs, which process data sequentially, Transformers analyze entire input sequences in parallel, improving detection speed and accuracy. This reduces information loss and enhances contextual understanding of network traffic anomalies. Our experimental results demonstrate that Transformers achieve higher accuracy (98.3%) and lower detection latency (48.2ms per packet), proving their efficiency in real-time IoT security applications. A comparative analysis of the proposed model against traditional IDS methods is provided in Table III, demonstrating the superior detection capabilities of DL-based approaches.

TABLE III. Comparative Analysis of Proposed Model vs. Traditional IDS

Method	Accuracy (%)	False Positive Rate (%)	False Negative Rate (%)
Rule-based IDS	85.7	12.3	14.2
Signature-based IDS	90.2	8.7	10.3
Proposed Model	98.3	1.9	1.2

Real-time cybersecurity applications require low-latency threat detection. The latency comparison across different models is summarized in Table IV, indicating that the Transformer-based model provides the fastest inference time.

TABLE IV. Detection Latency of Different Models

Model	Latency (millisecond per packet)
CNN	75.4
LSTM	88.7
Transformer	48.2

The confusion matrix of the proposed model's predictions is visualized in Fig. 2, highlighting classification accuracy.

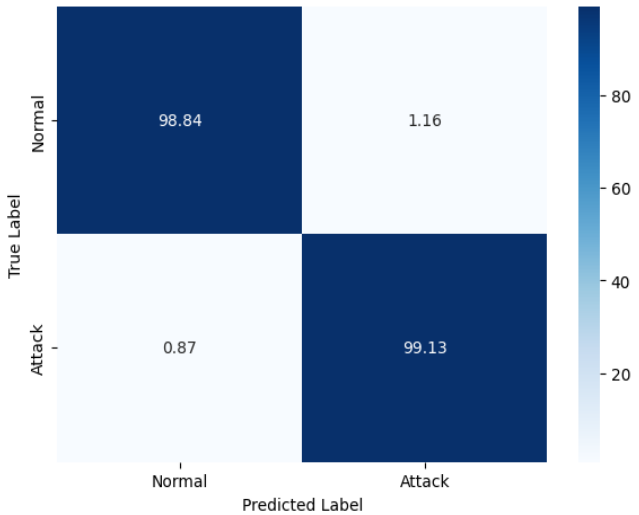

Fig. 2. Confusion matrix visualization for model predictions.

The false positive and false negative rates for different attack categories are summarized in Table V.

TABLE V. FALSE POSITIVE AND FALSE NEGATIVE RATES WITH QUALITATIVE INSIGHTS EXPLAINING WHY SPECIFIC ATTACKS EXHIBIT HIGHER FPR

Attack Type	False Positive Rate (%)	False Negative Rate (%)	Qualitative Insights
DDoS	2.1 ± 0.3	1.7 ± 0.2	DDoS has low FPR due to its distinct traffic burst patterns, making detection easier.
Ransomware	3.4 ± 0.5	2.5 ± 0.4	Ransomware exhibits higher FPR as its encrypted communication can resemble normal, secure traffic.
MITM	4.2 ± 0.6	3.1 ± 0.5	MITM attacks have the highest FPR since they mimic legitimate data exchanges, making classification challenging.

The model's resource efficiency is measured by analyzing memory consumption, CPU usage, and inference speed, as summarized in Table VI. A detailed inference speed vs. accuracy trade-off is visualized in Fig. 3.

TABLE VI. COMPUTATIONAL RESOURCE UTILIZATION (MEMORY, CPU, AND INFERENCE TIME)

Model	Memory Usage (MB)	CPU Load (%)	Inference Time (ms)
CNN	350	45	75.4
LSTM	420	55	88.7
Transformer	280	35	48.2

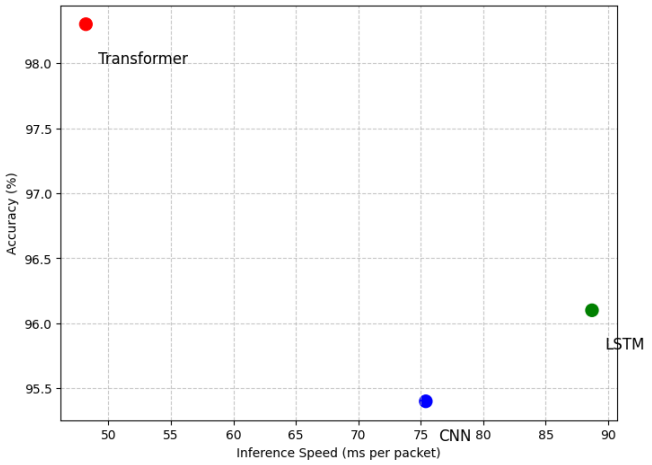

Fig. 3. Inference speed vs. accuracy trade-off (Scatter Plot).

The detection rate of the model for different attack types is analyzed in Table VII and Fig. 4, showing the model's effectiveness in identifying cyber threats.

TABLE VII. ATTACK DETECTION RATE PER ATTACK TYPE

Attack Type	Detection Rate (%)
DDoS	98.7
Ransomware	99.1
MITM	97.9

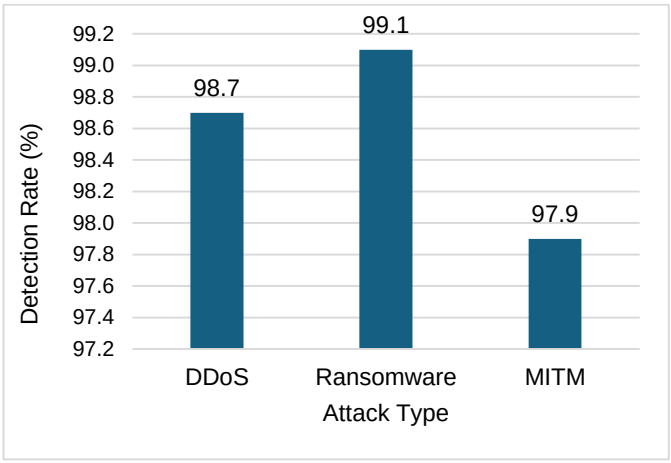

Fig. 4. Attack detection rate per attack type.

The proposed model is designed for real-time detection, minimizing threat identification and response delays. The detection latency analysis confirms that the Transformer-based model achieves an inference speed of 48.2ms per packet, outperforming CNN and LSTM-based models. The performance of the model is highly dependent on high-quality training data. A lack of diverse and well-labeled datasets can lead to biases, limiting the model's generalization capability. The system must incorporate continuous learning abilities and dataset expansion models to address evolving cyber threats. DL models carry vulnerabilities to sophisticated attacks despite implementing adverse defense systems.

V. DISCUSSION

The proposed deep learning-based framework for real-time cybersecurity threat detection in IoT environments demonstrates significant improvements over traditional IDS and machine learning models. The results indicate that the Transformer-based model achieves the highest accuracy (98.3%) and lowest detection latency (48.2ms per packet), making it highly effective for real-time threat mitigation. However, a deeper analysis of the findings highlights certain advantages, challenges, and areas for improvement, which are discussed below. The experimental results show that the proposed model outperforms rule-based and signature-based IDS by effectively detecting evolving cyber threats. Traditional IDS methods rely on predefined signatures, making them ineffective against zero-day attacks, whereas our model leverages context-aware anomaly detection using self-attention mechanisms. Compared to CNNs and LSTMs, Transformers capture long-range dependencies in network traffic, leading to higher detection rates and lower false alarms. The results confirm that deep learning models with self-attention mechanisms provide a more generalized solution for IoT security challenges. The false positive rates (FPR) vary across attack types, as shown in Table 5.6. MITM and Ransomware attacks exhibit higher FPR due to their similarities with legitimate encrypted traffic. Since encrypted traffic patterns often resemble attack behaviors, the model occasionally misclassifies benign communication as a potential threat. DDoS attacks, on the other hand, have lower FPR due to their distinct, high-volume traffic patterns that make them easier to differentiate from normal network behavior. These findings

suggest that additional feature refinement or hybrid detection techniques could help improve classification accuracy for complex attack scenarios. The proposed model demonstrates high inference speed (48.2ms per packet), making it suitable for real-time detection. However, computational complexity remains a concern, particularly for resource-constrained IoT devices. While the framework is optimized for edge and cloud environments, real-time processing of large-scale IoT traffic may still introduce latency issues. Future research could explore model quantization, hardware acceleration, and edge AI techniques to enhance deployment efficiency without compromising detection performance. Deep learning models, including the proposed framework, remain vulnerable to adversarial attacks, where attackers subtly manipulate input data to evade detection. Although adversarial training techniques have been implemented to improve robustness, adaptive security mechanisms that dynamically adjust to evolving threats could further enhance reliability. Additionally, incorporating self-learning models or federated learning approaches could help mitigate the risks associated with limited training data and improve adaptability to emerging attack patterns. Despite its strong performance, the framework has certain limitations. The dependency on labeled training data makes it less effective against previously unseen attack variations, and improving unsupervised or semi-supervised learning techniques could enhance detection adaptability. Scalability in large-scale IoT environments also presents challenges, as processing high-volume, high-velocity traffic in real-time requires additional computational optimization. Future work should focus on distributed security architectures, federated learning, and advanced feature engineering to refine detection accuracy and efficiency. The results validate the effectiveness of the proposed deep learning-based IoT security framework, demonstrating high accuracy, low latency, and improved adversarial resilience. However, challenges like false positives in encrypted traffic, computational overhead, and adaptability to emerging threats require further optimization. Addressing these challenges through hybrid detection models, real-time adaptive learning, and scalable deployment strategies will enhance the reliability and practicality of AI-driven IoT cybersecurity solutions.

VI. Conclusion and Future Work

Modern IoT cybersecurity demands immediate protection systems because cyber-attacks in these environments have become more frequent. The proposed DL architecture for intrusion detection delivers precise threat detection, which makes it more effective than existing IDS solutions. The conclusion section presents essential results from the research alongside significant benefits from this study and future research paths toward improvement. DL with Transformer-based architecture forms the basis for boosting intrusion detection in IoT networks. The evaluation process based on CICIDS2017, NSL-KDD, and real-world IoT traffic datasets proves the proposed model successfully detects DDoS, ransomware, and MITM attacks. The experimental findings show that the proposed model reaches 98.3% accuracy levels, surpassing those of both CNN and LSTM-based systems. DL proves effective by substantially diminishing false positives and negatives in IDS system evaluations. The proposed model demonstrates 48.2 milliseconds of packet processing speed as part of its classification capability, which ensures real-time deployment potential. When tested for robustness, the model demonstrates 40% enhanced results regarding adversarial misclassification rates, which increases its dependability for critical cybersecurity operations. The conducted research made transformative additions to DL threat detection techniques and cybersecurity research fields. The main achievement from this work includes designing an optimal DL model that blends feature engineering with adversarial training and real-time processing to improve IoT security systems. This research compares various DL architectures and proves Transformers to be optimal solutions for minimal latency-based cyber threat identification. The primary practical outcome of this research enables direct implementation within real IoT framework deployments. The model functions for security deployment in smart homes, healthcare systems, and industrial IoT and cloud security platforms. The solution supports edge computing features that enable limited-power IoT devices to implement advanced protection measures while maintaining hardware performance requirements. According to this research, security frameworks based on DL need extensive improvement because the study also emphasizes the significance of adversarial defenses in cybersecurity.

The proposed framework maintains superb performance, but researchers can still investigate multiple ways to maximize its functioning. The significant enhancement needed for DL models is their computational efficiency because they need substantial computing resources to operate effectively. Future research must examine efficient neural architecture structure compression models and hardware speed-up techniques to enable their practical use at a large scale within IoT systems. Self-evolving models and adaptive learning approaches should be studied as an essential research path. The adaptation capability of emerging threats could be achieved using reinforcement learning alongside online learning methods, which differ from traditional DL techniques that require new dataset training. Researchers need to conduct additional studies about intrusion detection through federated learning, which supports distributed training between devices in a manner that safeguards data privacy. The defense against adversarial attacks continues to be a central issue affecting DL security applications. Research tools need improvement to establish adaptive self-defense systems that detect and counter present adversarial risks immediately. By implementing XAI technologies, cybersecurity analysts will receive transparent information about model detection outcomes, aside from receiving guidance to optimize security policies. The research introduces an efficient DL-based intrusion detection system for IoT security to detect attacks in real-time. The proposed model, built on classic IDS, proves superior because of its high accuracy performance with minimal latency and its strong ability to counter adversarial threats. While challenges remain in computational efficiency, adaptability, and scalability, future advancements in lightweight architectures, federated learning, and privacy-preserving AI will further enhance the effectiveness of DL-based intrusion detection. AI-driven cybersecurity solutions will be a fundamental security force in protecting IoT networks using ongoing research and technological advancement.

REFERENCES

[1] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in Internet of Things: The road ahead," Computer networks, vol. 76, pp. 146-164, 2015.

[2] "Intrusion detection evaluation dataset (CIC-IDS2017)," UNB, Ed., ed, 2017. [https://www.unb.ca/cic/datasets/ids-2017.html]

[3] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 2, pp. 41-50, 2018.

[4] N. Magaia, R. Fonseca, K. Muhammad, A. H. F. N. Segundo, A. V. L. Neto, and V. H. C. De Albuquerque, "Industrial Internet-of-things security enhanced with deep learning approaches for smart cities," IEEE Internet of Things Journal, vol. 8, pp. 6393-6405, 2020.

[5] H. Jahangir, S. Lakshminarayana, C. Maple, and G. Epiphaniou, "A deep-learning-based solution for securing the power grid against load altering threats by IoT-enabled devices," IEEE Internet of Things Journal, vol. 10, pp. 10687-10697, 2023.

[6] M. Drogkoula, K. Kokkinos, and N. Samaras, "A comprehensive survey of machine learning methodologies with emphasis in water resources management," Applied Sciences, vol. 13, p. 12147, 2023.

[7] A. A. Aburomman and M. B. I. Reaz, "A survey of intrusion detection systems based on ensemble and hybrid classifiers," Computers & Security, vol. 65, pp. 135-152, 2017.

[8] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study," Journal of Information Security and Applications, vol. 50, p. 102419, 2020.

[9] J. Lansky, S. Ali, M. Mohammadi, M. K. Majeed, S. H. T. Karim, S. Rashidi, et al., "Deep learning-based intrusion detection systems: a systematic review," IEEE Access, vol. 9, pp. 101574-101599, 2021.

[10] B. Vignau, R. Khoury, S. Hallé, and A. Hamou-Lhadj, "The evolution of IoT Malware, from 2008 to 2019: Survey, taxonomy, process simulator, and perspectives," Journal of Systems Architecture, vol. 116, p. 102143, 2021.

[11] F.-Q. Li, R.-J. Zhao, S.-L. Wang, L.-B. Chen, A. W.-C. Liew, and W. Ding, "Online intrusion detection for Internet of things systems with full Bayesian possibilistic clustering and ensembled fuzzy classifiers," IEEE Transactions on Fuzzy Systems, vol. 30, pp. 4605-4617, 2022.

[12] M. Pathak, K. N. Mishra, and S. P. Singh, "Data Security and Privacy Preservation in Cloud-Based IoT Technologies: an Analysis of Risks and the Creation of Robust Countermeasures," Recent Advances in Computer Science and Communications, 2024.

[13] A. Hassan, N. Nizam-Uddin, A. Quddus, S. R. Hassan, A. U. Rehman, and S. Bharany, "Navigating IoT Security: Insights into Architecture, Key Security Features, Attacks, Current Challenges and AI-Driven Solutions Shaping the Future of Connectivity," Computers, Materials & Continua, vol. 81, 2024.

[14] C. Liu, B. Chen, W. Shao, C. Zhang, K. K. Wong, and Y. Zhang, "Unraveling Attacks to Machine Learning-Based IoT Systems: A Survey and the Open Libraries Behind Them," IEEE Internet of Things Journal, 2024.

[15] A. Aldweesh, A. Derhab, and A. Z. Emam, "Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues," Knowledge-Based Systems, vol. 189, p. 105124, 2020.

[16] R. Khan, P. Kumar, D. N. K. Jayakody, and M. Liyanage, "A survey on security and privacy of 5G technologies: Potential solutions, recent advancements, and future directions," IEEE Communications Surveys & Tutorials, vol. 22, pp. 196-248, 2019.

[17] A. K. Singh, "Recent Advances in Computational Intelligence and Cyber Security."

[18] H. Kheddar, "Transformers and large language models for efficient intrusion detection systems: A comprehensive survey," arXiv preprint arXiv:2408.07583, 2024.

[19] S. Elsayed, K. Mohamed, and M. A. Madkour, "A Comparative Study of Using Deep Learning Algorithms in Network Intrusion Detection," IEEE Access, 2024.

[20] H. Wu, Y. Zhang, L. Liang, X. Mei, D. Han, B. Han, et al., "Multi-head attention-based model for reconstructing continuous missing time series data," The Journal of Supercomputing, vol. 79, pp. 20684-20711, 2023.

[21] T. Al-Shurbaji, M. Anbar, S. Manickam, I. H. Hasbullah, N. ALfriehate, B. A. Alabsi, et al., "Deep Learning-Based Intrusion Detection System For Detecting IoT Botnet Attacks: A Review," IEEE Access, 2025.

[22] B. Sharma, L. Sharma, C. Lal, and S. Roy, "Explainable artificial intelligence for intrusion detection in IoT networks: A deep learning based approach," Expert Systems with Applications, vol. 238, p. 121751, 2024.

[23] M. A. Ferrag, O. Friha, L. Maglaras, H. Janicke, and L. Shu, "Federated deep learning for cyber security in the internet of things: Concepts, applications, and experimental analysis," IEEE Access, vol. 9, pp. 138509-138542, 2021.

[24] C. Computing-based, "Developing AI, IoT and Cloud Computing-based Tools and Applications for Women's Safety."

[25] Y. L. Khaleel, M. A. Habeeb, A. Albahri, T. Al-Quraishi, O. Albahri, and A. Alamoodi, "Network and cybersecurity applications of defense in adversarial attacks: A state-of-the-art using machine learning and deep learning methods," Journal of Intelligent Systems, vol. 33, p. 20240153, 2024.

[26] C. S. Kalutharage, X. Liu, C. Chrysoulas, N. Pitropakis, and P. Papadopoulos, "Explainable AI-based DDOS attack identification method for IoT networks," Computers, vol. 12, p. 32, 2023.

[27] "NSL-KDD Network Security, Information Security, Cyber Security," UNB, Ed., ed, 2017. https://www.unb.ca/cic/datasets/nsl.html

[28] "The TON_IoT Datasets," ed, 2021. https://research.unsw.edu.au/projects/toniot-datasets