Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FEDERATED INTRUSION DETECTION IN IOT ENVIRONMENTS USING HGS-CS-BASED FEATURE SELECTION AND A GATED TRANSFORMER-BILSTM ARCHITECTURE WITH SWARM AGGREGATION

P. UMA DEVI [1,2], GURPREET SINGH CHHABRA [1*]

[1] Department of Computer Science and Engineering, GITAM Deemed to be University, Vishakhapatnam, Andhra Pradesh, India
[2] Department of Game Design Technologies, Dr.YSR Architecture and Fine Arts University, Kadapa, Andhra Pradesh, India
[1*]Corresponding Author Email id: gchhabra@gitam.edu

ABSTRACT

Introduction The increasing complexity of cyber threats in Internet of Things (IoT) environments requires the development of efficient and privacy preserving Intrusion Detection Systems (IDS). IoT networks generate high-dimensional and heterogeneous traffic data, making effective feature selection and temporal attack detection challenging, particularly in distributed and resource-constrained environments.

Objectives: This research has the following objectives: (1) to develop a hybrid feature selection method using Hybrid Grey Wolf Optimization integrated with Cuckoo Search and chaotic Lévy flight (HGS-CS), (2) to reduce the dimensionality of IoT network traffic while retaining highly discriminative features, (3) to develop a TSTformer-LSTM model combining a Transformer Encoder, Bidirectional LSTM, and Gated Fusion mechanism for capturing global contextual and temporal dependencies, and (4) to develop a privacy-preserving federated IDS using SwarmFed with attention-based aggregation, Stable Focal Loss, and centralized fine-tuning.

Methods: This research proposed a novel framework of hybrid feature selection using Hybrid Grey Wolf Optimization integrated with Cuckoo Search enhanced with chaotic Levy flight (HGS-CS) and a TSTformer-LSTM model integrated with a Transformer Encoder with a Bidirectional LSTM and a Gated Fusion mechanism that captures both the global dependencies in context and also the temporal dependencies which validated on the datasets UNSW-NB15 and BoTNeTIoT-L01. The HGS-CS performs the multi-objective feature selection which results in a dimensionality reduction of 86.96% by selecting only 3 highly discriminative features in BoTNeTIoT-L01 dataset and 8 features in UNSWNB-15 dataset. The model is trained based on a SwarmFed federated learning framework with attention-based aggregation, Stable Focal Loss and centralized fine-tuning.

Results: Experimental results show performance which achieves an accuracy of 96.9% on BoTNeTIoT-L01 dataset and 89% on UNSWNB-15 dataset. The proposed framework provides a robust, efficient, and privacy-aware IDS solution suitable for distributed and resource-constrained IoT environments

Keywords: *Transformer Encoder, Bidirectional LSTM, UNSW-NB15, BoTNeTIoT-L01, Federated learning.*

1. INTRODUCTION

Intrusion Detection Systems (IDS) have become an essential part of today's cybersecurity and they are developed to monitor, analyse and respond to malicious activity throughout networks and host environments. With the increase in digital communication and the sophistication of cyber-attacks, traditional rule-based and signature-driven IDS approaches have become increasingly ineffective in addressing evolving threats. As a recent development in the area of IDS performance improvements, the application of deep learning (DL) methods has been studied as an effective technique that offers flexibility, scalability, and superior performance in terms of detection accuracy [1]. Author provide a study of the application of machine learning techniques for IDS in a systematic manner, discussing their capabilities, as well as limitations of using practical in real world environments limitations of their application to a real world scenario. The survey indicates that the ML-based IDS can result in a considerable improvement of the detection rates and reduction of the false positives compared to the traditional

algorithms, the use of the algorithm, the quality of the data set and its features selection matter in the performance of the IDS, the research indicates that the balance between the detection rates and the computational efficiency should be maintained, particularly in real time settings where the IDS should have the ability to handle massive amounts of traffic without introducing a reduction in latency. This should analysis provides a thorough explanation of why hybrid and ensemble approaches are necessary that incorporate more than one learning paradigm to gain resilience to various and changing attacks [2].

Research Questions

(a) RQ1: How does the HGS-CS effectively reduce the dimensionality of IoT network traffic while retaining discriminative features for intrusion detection?

(b) RQ2: How does the proposed Gated TSTformer-LSTM architecture effectively capture both global contextual dependencies and temporal dependencies in IoT network traffic?

(c) RQ3: How does the SwarmFed framework with attention-based aggregation and Stable Focal Loss provide effective privacy-preserving distributed intrusion detection?

(d) RQ4: How does the proposed HGS-CS–TSTformer-LSTM–SwarmFed framework perform compared with conventional baseline models in terms of detection performance and feature reduction?

The important contribution of this research is as follows:

1. This research paper proposed an Advanced Meta Heuristic Feature Engineering method as Hybrid Grey Wolf Optimization with Cuckoo Search (HGS-CS). The combination of the local exploitation capabilities of GWO and global exploration of Cuckoo Search can reach better multi-objective feature selection. This leads to extreme dimension reduction while maintaining the most discriminative attributes needed for attack identification.
2. Novel TSTformer-LSTM Architecture: This research proposes a parallel neural architecture that utilizes the Gated Fusion mechanism. Transformer Encoders to capture global contextual dependencies, with Bidirectional Long Short-Term Memory (Bi-LSTM) networks to capture fine-grained temporal sequences, the model is able to detect complicated, multi-stage attack patterns that are not identified by traditional models.
3. Robust SwarmFed Learning approach to make data privacy not a trade-off with the model utility, Attention-based Aggregation, Stable Focal Loss and Centralized Fine-tuning. The proposed framework was validated on the UNSW-NB15 and BoTNeTIoT-L01 datasets. Experimental results indicate that BoTNeTIoT-L01 and UNSW-NB15 96.9% and 89% respectively indicate that the model has high resource levels of resistance and performance in the constrained setting of the distributed environment.

The paper is organized as follows: Section 2 reviews related work. Section 3 describes datasets description. Section 4 details research methodology Section 5 presents the proposed model Architecture TSTformer-LSTM and Swarm Fed framework. Section 6 reports results. Section 7 concludes with future directions

2. RELATED WORKS

Intrusion Detection Systems (IDS) have become important in the field of cybersecurity, within Internet of Things (IoT) environments, where limited computation and privacy preserving make detection challenges. Previous studies have shown in the field of machine learning provided the theoretical background for adaptive IDS that can differentiate between normal and malicious network activities [3].The Previous studies have shown the limitations of the traditional rule-based and signature-driven methods shows the necessary importance to combine machine learning (ML) and deep learning (DL) methods to obtain a better detection accuracy, scalability, and flexibility in dynamic network conditions [4]. In recent years Several solutions have been proposed to enhance the performance of IDS. The feature selection methods which use optimization as, grey wolf optimization (GWO), particle swarm optimization (PSO) have proved to be effective in reducing the dimensionality of features and the computational cost without compromising upon the performance of the classification [5]. All of these models are often focus on detection accuracy but the other important factors such as model interpretability and computational efficiency not considering. Concurrently, deep learning architectures like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid ensemble models have shown better performance in capturing complex patterns within network traffic. Despite of all these advancements, current models often have limitations to address long-range contextual

dependencies and temporal dynamics in IoT traffic data Transformer-based architectures have recently been shown attention in their ability to model global dependencies, their combination with temporal sequence learners and efficient fusion mechanisms is not fully investigated [7].

The increasing importance of data privacy has led to a further to adopt the federated learning (FL) frameworks in IDS design. Federated methods enable

model training in a secure way without access to all data, which eliminates privacy concerns. Nevertheless, the majority of the existing FL-based IDS rely on the traditional system of aggregation, including Federated Averaging (FedAvg), which is susceptible to the poisoning attack and cannot be adaptable to heterogeneous IoT settings [8]. Advanced models like attention-based aggregation or adaptive loss optimization are not typically integrated, and, consequently, they lack to provide robustness and equity among distributed nodes. In addition, the evaluation methodologies applied in IDS research are still dependent on benchmark datasets such as UNSW-NB15 and BoTNeTIoT-L01 with minimal focus on cross-dataset generalization and deployment. This insufficiency of comprehensive assessment restricts the applicability and scalability of suggested models into practical use in real-world IoT ecosystems [9]. The recent research ML and DL methods, requires the ability to effectively secure context-sensitive and IoT-specific IDS models. The new literature considers the implementation of lightweight models and optimization algorithms that should be deployed with resource-constrained IoT systems [10].

satellite and IoT communication networks have been suggested as one of the primary areas where IDS should be capable of balancing between detection accuracy and low latency and energy efficiency [11]. sensor-based IDS frameworks have been proposed to capitalize on the real-time data streams, which demonstrate greater adaptability in dynamic Recent developments in the field of deep learning architectures have also proposed new hybrid designs where convolutional, recurrent, and transformer building blocks are integrated to learn spatial and temporal network traffic dependencies [13]. These methods demonstrate potential to improve the detection ability, but there are still issues related to scalability and interpretability. Further, it has been observed that federated learning is made more complex to offer privacy-conserving IDS solutions, and distributed training heterogeneous IoT devices without centralizing sensitive data [14].

Most federated IDS implementations make use of simple aggregation schemes that limit resistance to adversarial attacks and poisoning attacks. Recent surveys indicate that cross-domain validation and benchmarking is also a major concern to be taken into consideration as models of the IDS are often highly accurate in a single-dataset scenario but not in a broad spectrum of traffic conditions [15]. This raises the question of their capability to be used in real world implementation of IoT where traffic patterns are highly heterogeneous and dynamic. Moreover, feature selection method has been studied widely, and it has been demonstrated that it can be effectively used in dimensionality reduction which can also preserve detection accuracy and can reduce computational cost [16].

To address the limitations, present in the above existing research we have proposed novel framework in this study combines Hybrid Grey Wolf Optimization with Cuckoo Search (HGS-CS) which is a multi-objective feature selection method and found (86.9%) percent dimensionality reduction as well as preserving discriminatory features. A new TSTformer-LSTM model has been introduced that uses Transformer encoders with BiLSTM and gated fusion to learn both global and temporal dependencies, the incorporation of the model in a SwarmFed federated learning system with attention-based aggregation and Stable Focal Loss guarantees privacy, scalability, and resilience.

The experimental validation on both UNSW-NB15 and BoTNeTIoT-L01 datasets shows better performance, making it a significant solution of an efficient, robust, and privacy-conscious IDS to a distributed IoT environment. This study proposed the following contribution as designing the privacy-preserving and efficient intrusion detection system (IDS) for the modern IoT environments. Although existing studies have demonstrated the effectiveness of federated learning, optimization-based feature selection, and deep learning for IoT intrusion detection, most studies investigate these components separately or use relatively large feature sets. The comparative studies also report different experimental settings, preprocessing procedures, feature subsets, and training paradigms, making direct performance comparison difficult. The main gap addressed in this study is the

integration of compact feature selection, global and temporal feature learning, and attention-based federated aggregation within a single framework. The proposed approach therefore focuses not only on detection accuracy but also on reducing feature dimensionality and supporting distributed IoT intrusion detection.

2.1 Problem Statement

The rapid growth of the Internet of Things (IoT) has brought significant advancements across various domains. Traditional Intrusion Detection Systems (IDS) struggle to handle high-dimensional and noisy IoT data effectively. Federated Learning (FL) is a decentralized approach to training models using local data, The Existing IoT IDS systems have challenges in high-dimensional noisy data, lack of feature selection methods, and poor temporal-contextual dependency models. Despite their potential, Federated Learning methods often struggle with limitations such as model leakage, inconsistent architectures, and resource weaknesses in the resource-constrained IOT environment. To address these gaps, this paper suggests a hybrid HGS-CS feature selection and Gated TSTformer-LSTM model with in a Swarm Fed framework, which ensures dimensionality reduction, privacy protection, and high detection accuracy on benchmark UNSW-NB15, and BoTNeTIoT-L01 datasets.

3. DATASET DESCRIPTION

This study evaluates the efficiency of the proposed framework as HGS-CS-TSTformer-Swarm Fed architecture (on benchmark datasets as UNSW-NB15 and BoTNeTIoT-L01) by providing a detailed network traffic logs (with labeled attacks) hence facilitating efficient training and evaluation of the deep learning

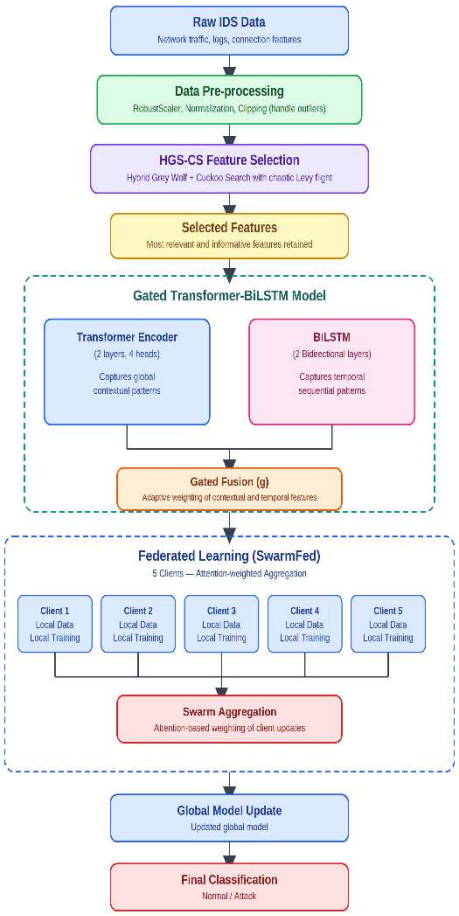

Figure 1: Proposed Methodology

Table 1 Illustrates The Aim And Limitations Of The Previous Research

Approach	Reference	Description	Advantages	Limitations
Centralized ML/DL	[1],[2]	Data aggregated at central server (SVM, RF, CNN)	High accuracy, easy optimization	Privacy risk, high bandwidth, single point of failure
Standard FL	[3],[5],[6],[15]	Local training + share weights to server	Privacy-preserving, low data transfer	Poisoning risk, assumes IID data
Weighted/Hybrid FL	[8],[10]	Weighted updates based on data quality/node reliability	Handles non-IID, improved performance	Extra computation for weighting
RF-FedAvg	[4]	RF + FedAvg for WSNs	Low effic	
Federated Transfer Learning	[9]	Uses pre-trained models across domains	Fast for s	
Domain-Adversarial	[12]	Aligns different protocol distributions (IP vs BLE)	Robu perfo	
Privacy-Enhanced FL	[16]	+ DP/encryption	Stron proto	
(FedPyIDS)				

3.1 UNSWNB-15 Dataset

The UNSW-NB15 dataset is a collection of intrusion detection system evaluations produced by the Cyber Range Lab at UNSW Canberra to provide a benchmark on intrusion detection systems. It was created with the IXIA Perfect Storm tool which created an approximation of a realistic network traffic, with both regular network activity and contemporary cyberattacks. The dataset has

approximately 2.5 million labeled records, each having 49 features that consist of basic attributes, content attributes, time-based attributes and statistical attributes. It includes 9 types of attacks (DoS, Exploits, Fuzzers, and Reconnaissance) as well as normal traffic, which makes it appropriate to multiclass classification and anomaly detection missions. The data is given in PCAP and CSV formats and separated into a training set of 175,341 records and a testing set of 82,332 records. UNSW-NB15 dataset provides a range of features and can be used in both traditional and deep learning methods, it is also challenging imbalance of classes and limited generalization [17].

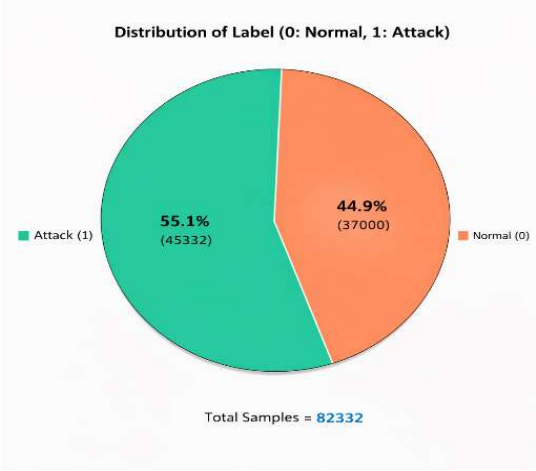

Figure 2: Distribution of attacks

3.2 BoTNeTIoT-L01 Dataset

The BoTNeTIoT-L01 dataset represents a recent benchmark designed for IoT-based intrusion detection research. It consists of network traffic gathered from nine IoT devices within a local network environment where packet capture was implemented using Wireshark through the centralized switching infrastructure. The dataset incorporates realistic botnet attack situations, specifically the Mirai and Gafgyt variants, along with the normal traffic.A total of twenty-three statistical features are computed on the raw packet capture (pcap) files so as to characterize the network behavior. These features are generated through statistical analysis for a fixed window of time (10 seconds). Important statistical metrics including mean, variance, count, magnitude, radius, covariance and correlation coefficient are computed that captures temporal and relational properties of the traffic. A decay factor of 0.1 is used to give weight to recent observations in the time window, and this setting is called the L0.1 setting. In addition to statistical measures, basic traffic attributes such as the packet count, jitter, and packet size (for outbound traffic as well as combined inbound and outbound flows) are extracted. Multiple statistical descriptors are computed for each of these attributes, hence a comprehensive set of twenty-three descriptive features results. This organized model can be analyzed and modeled effectively to understand IoT network traffic to detect intrusion [18].

Table 2: UNSW-NB15 Dataset: Key Feature Categories

Feature Category	Description
Flow Identifier	A unique tag assigned to each network flow.
Source/Destination IP & Ports	Specifies the IP addresses and port numbers for both sender and receiver.
Protocol	Specifies the communication protocol in use, such as TCP, UDP, or ICMP.
Service Type	Identifies the application-level service involved, for example, HTTP, FTP, SSH, or SMTP.
Connection State	Indicates the current status of the network session, such as Established, Closed, or Reset.
Intrusion Type	Classifies the detected attack, including categories like Fuzzers, DoS, Exploits, or Backdoors
PacketSize Metrics	Statistical measures of packet sizes including minimum, maximum, mean, and standard deviation.
Connection Duration	Total time span of the network session.
Byte Transfer Stats	Amount of data sent and received by source and destination.

4.RESEARCH METHODOLOGY

4.1 Data Preprocessing

The BoTNeTIoT-L01 dataset preprocessed to ensure data quality and consistency. An irrelevant index-related attribute (*Unnamed: 0*) was removed as it did not contribute to the learning process. The dataset was examined for data integrity, confirming the absence of duplicate records and missing values. Furthermore, all input features were found to be in numerical format (*float64*), while the target variable was encoded as an integer (*int64*), eliminating the need for categorical encoding or transformation. This ensured that the dataset was directly suitable

for building models without additional preprocessing overhead.

UNSWNB-15 dataset is used to ensure that the dataset could be used for further Data pre-processing, a structured pipeline was followed. First, categorical attributes (proto, service, and state) were transformed using one hot encoding and a total of 196 features were obtained. Non informative identifiers and labels (id, attack_cat and label) were removed to avoid information leakage and redundancy. then Variance Threshold filter with a threshold value of 0.01 was used in order to remove constant and low variance features so that the dimensionality was reduced (51 retained attributes). This step ensured that features with adequate variability were only retained to train the model. To normalize, the standard Scaler method was used on the filtered feature set. To normalize the feature space and ensure uniform contribution of variables, the **Standard Scaler** was applied. This technique rescales features by removing the mean and scaling to unit variance, defined as:

$$x' = \frac{x-\mu}{\sigma} \quad (1)$$

Where

- x = original feature value, μ = mean of the feature
- σ = standard deviation of the feature, x′ = standardized feature value

This transformation ensures that each feature has a mean of zero and a standard deviation of one, thereby improving the stability and convergence of deep learning models. For baseline models (Logistic Regression, Random Forest, and DNN), a subset of 23 features was selected from the reduced feature set to ensure computational efficiency and comparison, for the proposed model employs an adaptive feature selection mechanism (HGS-CS), selecting a smaller optimal subset (8 features), thereby improving efficiency while maintaining high detection performance.

4.1.1 Exploratory Data Analysis and Feature Characterization

An Exploratory data analysis (EDA) was conducted to attain a more comprehensive understanding of the BoTNeTIoT-L01 dataset. Essential statistical metrics, such as mean, standard deviation, minimum, maximum, and quartile values, were calculated to examine feature distributions. The features were classified into continuous and discrete kinds, comprising 22 continuous variables and one discrete variable denoting the class label.

4.1.2 Handling Class Imbalance

UNSW-NB15 and BoTNeTIoT-L01 datasets both exhibited class imbalance, potentially negatively impacting model performance by biasing prediction towards the majority class. A random undersampling method was used on the majority class (normal traffic) to achieve a balanced distribution of classes. The process resulted in a sample of 1,026,994 samples, with both classes being equally represented (50:50 ratio). The dataset was randomly shuffled, and the benchmark index was reset to ensure an unbiased training and enhance the generalization of the proposed model

2.HGS-CS Feature Selection ALGORITHM

Algorithm 1: HGS-CS Feature Selection (Hybrid Grey-Wolf + Chaotic Cuckoo Search)

Input: $X_{\text{train}} \in \mathbb{R}^{m\times n}$, y_{train}, $n_{\text{features}} = n$, pop_size = 10, max_iter = 20, $\lambda = (0.4,0.2,0.2,0.2)$, pa = 0.25, $\alpha = 0.01$, $\beta = 1.5$, stability_thresh = 0.5

Output: Selected feature indices $S \subset \{0,\dots,n-1\}$

1.Chaotic Initialization (10 iterations of logistic map):

$$P^{(0)} = \text{rand}(pop\_size, n) \Rightarrow P = \begin{cases} \mu P & \text{if } P < 0.5 \\ \mu(1-P) & \text{else} \end{cases}, \mu = 2.0$$

$$P \leftarrow (P > 0.5)$$

2. For each fold in 3-fold StratifiedKFold:

Compute fitness for every individual p_i:

$$\text{Fitness}(p) = \lambda_1(1-F1) + \lambda_2 R + \lambda_3 \frac{k}{n} + \lambda_4 \min(T,1)$$

where

$R = \frac{1}{k(k-1)} \sum_{i\neq j} |\text{corr}(X[:,i], X[:,j])|$(average absolute correlation)

$$k = |\{j : p_j > 0.5\}|$$

T: training time (s)

3.Levy Flight Step (for exploration):

$$\sigma_u = \left(\frac{\Gamma(1+\beta)\sin(\pi\beta/2)}{\Gamma((1+\beta)/2)\beta 2^{(\beta-1)/2}}\right)^{1/\beta}, \text{Levy} = \alpha \frac{u}{|v|^{1/\beta}}, u \sim \mathcal{N}(0,\sigma_u), v \sim \mathcal{N}(0,1)$$

4.HGS Update Rule (Harris Hawks inspired):

$$P_i^{\text{new}} = P_i + \alpha_{\text{rand}} \cdot \nabla + \beta_{\text{rand}} \cdot (\text{best} - P_i) + \gamma_{\text{rand}} \cdot (P_j - P_k)$$

Clip: $P_i^{\text{new}} \leftarrow \text{clip}(P_i^{\text{new}}, 0, 1)$.

5. Replacement & Abandonment:
If stagnant: $P_i \leftarrow P_i + \text{Levy}$
Abandon worst pa% individuals → random binary reset
Update best position if fitness improvesStability Weighting (across folds):

$$\text{stab}_j = \frac{1}{F}\sum_{f=1}^{F} \mathbb{I}(P_f^j > 0.5), S = \{j: \text{stab}_j \geq 0.5\}$$

4.4 Proposed Advanced Framework

4.5 HGS-CS Based Feature Selection

To enhance the performance of the model and reduce the computational complexity. The hybrid method of feature selection based on Hybrid Grey Wolf Optimization and Cuckoo Search (HGS-CS) was introduced This algorithm is a combination of the analytical and exploitation capabilities of the Grey Wolf Optimization (GWO) and Cuckoo Search (CS) methods with the addition of a chaotic Lévy Flight mechanism to enhance convergence and to prevent local optima. In the given framework, the population size of 10 candidate solutions was launched and the optimization was performed in 20 iterations. Every candidate solution corresponds to a subset of features, and the fitness was determined according to the performance in classification. In order to have robustness and generalization, 3-fold stratified cross-validation was employed in the evaluation phase. To minimize the computational load, but at the expense of representativeness, the selection of the features was conducted on a stratified subsample of 20,000 instances. The proposed HGS-CS methodology was evaluated using two benchmark datasets, the BoTNeTIoT-L01 and UNSW-NB15, to prove its efficacy in both IoT and traditional network environments. The proposed method effectively identified a highly compact subset of discriminative features within the BoTNeTIoT-L01 dataset. The proposed model effectively identified high-dimensional feature space, with targeted attributes, namely, H L 0.1 weight, HH L 0.1 mean and HH L 0.1 jit mean were chosen, providing a dimensionality reduction of about 86.96%.

These features capture both statistical behavior and temporal dynamics of IoT network traffic, which are critical for distinguishing between normal and malicious activities. on the UNSW-NB15 benchmark dataset, which has diverse network traffic parameters, both numerical and categorical, the HGS-CS methodology revealed its effectiveness in identifying a highly selective set of features. From the reduced feature space, eight features, selected as dttl, djit, dtcpb, ct_dst_src_ltm, proto_arp, proto_tcp, proto_unas, and service_http, were selected, and they achieved a high level of dimensionality reduction without losing any useful information.

These features capture key protocol behaviors and traffic statistics, which are very predictive of network intrusions. Data preprocessing, including categorical encoding, variance thresholding, and feature scaling, provided data stability and improved feature quality optimization before Training the model. The findings of the two datasets show that HGS-CS strategy is an effective method to reduce dimensionality space and minimize computational complexity, enhance model interpretability. The method increases training efficiency, rapid inferencing, and achieves better generalization performance by emphasizing a minimum important number of features. The technique improves training efficiency, reduces inference time, and achieves better generalization performance by selecting relevant set of features. This demonstrates the strength and adaptability of the suggested feature selection method that it is highly suitable to be applied in real-world intrusion detection frameworks performing in heterogeneous and resource-restricted environments.

4.5 MODEL ARCHITECTURE

4.5.1 Transformer Architecture

The Transformer architecture uses self-attention to store relationships between terms, allowing each token to dynamically attend to other tokens in the sequence. A positional encoding layer is used to retain positional relationships within feature sequences. This layer generates position vectors, which are then included into word embeddings, using the model to reflect relative distances and structural relationships more precisely.
Feature Transformation Equations

Multi-Head Attention Layer

$$X_1 = \text{MHA}(QX, KX, VX) = \text{Concat}(\text{head}_1, \text{head}_n)W_0 \quad (1)$$

Feature Fusion:

$$X_2 = \text{Concat}(X_2, X_1) \quad (2)$$

Gated Feature Selection:

$$X_L = X_t \cdot X_s + (1 - X_s) \cdot X_2 \quad (3)$$

Here, $W_1, W_2$are learnable weight matrices, and b_1, b_2represent bias terms. The gate vector X_s dynamically regulates the passage of feature information, enabling the model to preserve

important semantic and contextual signals while filtering out less significant data[19].

4.5.2 Bi-LSTM Model

A BiLSTM network calculates the hidden state sequences for both layers that operate in both the backward and forward directions. The forward hidden state sequence is computed by iterating from time step t=1 to T, whereas the backward hidden state sequence is computed from t=T to 1. The final output is then updated using these sequences. The outputs from both the backward and forward layers are measured using the standard Long Short-Term Memory equations, as described in Equations 1 and 2. The bidirectional long short-term memory layer produces an output vector, YT, which is computed using a combination function, σ, that merges both sequences.

This σ function can operate in various ways, such as concatenation, summation, averaging, or multiplication. By integrating Bidirectional with LSTM units, that can effectively capture long-term dependencies in both directions. Merging the backward and forward Long Short-Term Memory layers results in a single layer of BiLSTM, which effectively captures information about both past and future contexts in the sequence of multiple domains, such as speech recognition [20] . Within the federated learning framework, the BiLSTM layer enhances the robustness of the intrusion detection system by enabling each client to learn temporal representations locally while preserving data privacy. When integrated with the SwarmFed aggregation framework, these locally learned BiLSTM features contribute to improved global model performance without requiring raw data sharing. Consequently, BiLSTM serves as an effective component for modeling sequential traffic patterns in resource-constrained and privacy-sensitive IoT environments.

4.6 TSTFORMER-LSTM HYBRID MODEL

Algorithm 2: TSTformer–LSTM Forward Pass

Input: $X \in \mathbb{R}^{b\times n_{\text{feat}}}$(batch) **Output**: Logits $\in \mathbb{R}^{b\times 2}$

$$
\begin{aligned}
H_{\text{in}} &= \text{BatchNorm}(\text{Linear}(X)) \in \mathbb{R}^{b\times d_{\text{model}}} \\
H_{\text{in}} &= H_{\text{in}}[:,\text{None},:] \text{(add sequence dim)} \\
H_{\text{trans}} &= \text{TransformerEncoder}(H_{\text{in}})[:,0,:] \in \mathbb{R}^{b\times d_{\text{model}}} \\
H_{\text{lstm},-} &= \text{BiLSTM}(H_{\text{in}})[:,-1,:] \in \mathbb{R}^{b\times 2h_{\text{lstm}}} \\
H_{\text{fused}} &= \text{GatedFusion}(H_{\text{trans}}, H_{\text{lstm}}) \\
&= g \odot \text{LN}(\text{proj}_t(H_{\text{trans}})) + (1-g) \odot \text{LN}(\text{proj}_l(H_{\text{lstm}})) \\
g &= \sigma(W_1 H_t + W_2 H_l + b) \\
\text{logits} &= \text{Classifier}(H_{\text{fused}}) = \text{Linear}(\text{ReLU}(\text{Linear}(H_{\text{fused}})))
\end{aligned}
$$

The suggested Hybrid Transformer–BiLSTM with Gated Fusion model initiates by transferring the decreased feature set into a 64-dimensional space using a linear input layer, succeeded by batch normalization to provide stable scaling and enhanced convergence. The altered features are subsequently processed concurrently by two branches: a Transformer encoder of two layers and four attention heads that captures global contextual dependencies, and a stacked bidirectional LSTM with two layers that learns sequential and temporal patterns in network traffic. A gated fusion technique is used to integrate these complementary representations, where each branch output is projected, normalized, and adaptively weighted through a sigmoid-based gating function. This enables the model to dynamically prioritize both contextual or temporal data based upon the traffic environment. The integrated representation is then transmitted through fully linked layers using nonlinear activation functions, resulting in a classifier that generates predictions for normal and attack categories. This architecture efficiently combines global attention with sequential modeling while maintaining a lightweight design, rendering it appropriate for deployment resource-constrained environments such as IoT networks. The integration of feature selection with the hybrid deep learning model significantly reduces input dimensionality, thereby lowering computational overhead and improving training efficiency without compromising detection accuracy.

4.7 Federated Learning with Swarm-Based Aggregation

The proposed system integrates a federated learning technique as SwarmFed to mitigate privacy concerns and facilitate distributed learning across many data sources. This method involves distributing the training data among several client nodes, each of which separately trains a local version of the TSTformer-LSTM model. This decentralized model provides that sensitive network data remains localized, effectively protecting data privacy. In each communication round, the global

model parameters are passed on to all participating clients. Each client does local training with its designated data subset and modifies the model parameters. The suggested methodology utilizes an attention-based swarm aggregation mechanism rather than traditional aggregation approaches like Federated Averaging. This method assigns weights to each client's contribution according to its validation performance, enabling more reliable models to use greater significance on the global update. The aggregation method calculates adaptive weights by a softmax-based attention mechanism, guaranteeing stable and efficient convergence of the global model. Furthermore, several communication cycles are conducted to progressively enhance the global model. Subsequent to federated training, a centralized fine-tuning phase is executed to augment model performance. This federated learning system enhances scalability and tackles issues associated with non-independent and identically distributed (non-IID) data among clients. To further strengthen model performance under class imbalance conditions, a Stable Focal Loss function is utilized during training. This loss function emphasizes hard-to-classify samples while preventing instability commonly associated with standard focal loss, thereby improving detection sensitivity for minority attack classes. The federated training process is conducted over 10 communication rounds, where local models are iteratively updated and aggregated. Following this, a **c**entralized fine-tuning phase of 20 epochs is performed to refine the global model and enhance convergence. The suggested approach enhances generality and robustness in intrusion detection tasks by integrating swarm intelligence principles with federated optimization.

$$\mathcal{L}_{\text{focal}} = \frac{1}{N}\sum_{i=1}^{N}(1-p_t)^{\gamma}\cdot \text{CE}(z_i, y_i) \quad (2)$$

where $p_t = \exp(-\text{CE}), \gamma = 2.0$, CE includes label smoothing (0.05). NaN-safe fallback to plain CE.
where $p_t = \exp(-\text{CE}),(z_i, y_i))$ represents the predicted probability for the true class, γ is the focusing parameter that controls the strength of down-weighting easy samples,
$\text{CE}(z_i, y_i)$ denotes the cross-entropy loss between predicted logits z_i and true labelsy_i
N is the total number of samples.
The focusing parameter (γ=2.0\gamma = 2.0γ=2.0) balances hard and easy samples, while label smoothing (0.05) reduces overconfidence and improves calibration.

A fallback to standard cross-entropy ensures numerical stability during training. This Stable Focal Loss enhances detection of minority attack classes while maintaining overall performance, making it suitable for robust intrusion detection systems. The proposed SwarmFed framework enhances privacy preservation, robustness, and generalization, making it well-suited for real-world IoT intrusion detection scenarios.

Algorithm 3: SwarmFed Training (Attention-based Federated Learning)

Input: $X_{\text{train}}, y_{\text{train}}$(sharded into $C = 5$clients), rounds = 10, local_epochs = 3 **Output**: Global model

For round $r = 1$**to 10**:

1.Broadcast global parameters $\theta^{(r-1)}$to all clients.

2.Local Training (each client c):

$$\theta_c^{(r)} = \arg\min_{\theta} \mathcal{L}_{\text{StableFocal}}(\text{TSTformer-LSTM}(X_c), y_c)$$

$$\mathcal{L}_{\text{Focal}} = \frac{1}{N}\sum(1-p_t)^{\gamma}\cdot \text{CE}(z,y), \gamma = 2.0$$

2.Compute deltas and client scores:

$\Delta_c = \theta_c^{(r)} - \theta^{(r-1)}, s_c = -\text{val\_loss}_c$

3.Attention-weighted Aggregation:

$$\alpha = \text{softmax}(s), \theta^{(r)} = \theta^{(r-1)} + \sum_{c=1}^{C}\alpha_c \cdot \Delta_c$$

4.Centralized Fine-tuning (20 epochs on full selected data):$\theta \leftarrow \text{Adam}(\theta, \mathcal{L}_{\text{Focal}}, \text{lr} = 10^{-3}$,scheduler = StepLR($\gamma = 0.5$))

5. RESULTS AND ANALYSIS

5.1 Experimental Setup and Evaluation

The proposed framework was run on PyTorch on a workstation with an NVIDIA GPU and an Intel Core i7 processor, the test was done with the BoTNeTIoT-L01 and unswnb-15 dataset that was balanced through under sampling to reduce class bias. The final classification phase was done on a standard 80:20 train-test split and a 3-fold cross-validation was incorporated in the feature selection process to ensure the strength of the subsets selected. To measure performance, five major metrics were used, including Accuracy, Precision, Recall, F1-score and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC).

5.2 Performance of HGS-CS Feature Selection

The proposed HGS-CS framework was shown to be very efficient in dimensionality reduction. The process of optimization was characterized by a high rate of convergence of the fitness function that became stable between the 10th and 15th iterations in all folds. As shown in Table 2, the algorithm was able to continuously decrease the feature space. The

23 statistical features of the BoTNeTIoT-L01 dataset were narrowed down to 3 important features using a stability voting mechanism, whereas the feature space of the UNSW-NB15 dataset contained initially 51 (following preprocessing) features, which were narrowed down to 8 features. This amounts to dimensionality reduction of about 86.96 percent in BoTNeTIoT-L01 and 84.31 percent in UNSW-NB15, which greatly improves the computational efficiency without majorly reducing the discriminative performance.

Table 3: List of Selected Features using HGS-CS Method

Dataset	Original Features	Selected Features (Method)	No. of Selected Features	Reduction (%)	Selected Feature Names
BoTNeTIoT-L01	23	HGS-CS	3	86.96%	H_L0.1_weight, HH_L0.1_mean, HH_jit_L0.1_mean
UNSW-NB15	51	HGS-CS	8	84.31%	dttl, djit, dtcpb, ct_dst_src_ltm, proto_arp, proto_tcp, proto_unas, service_http

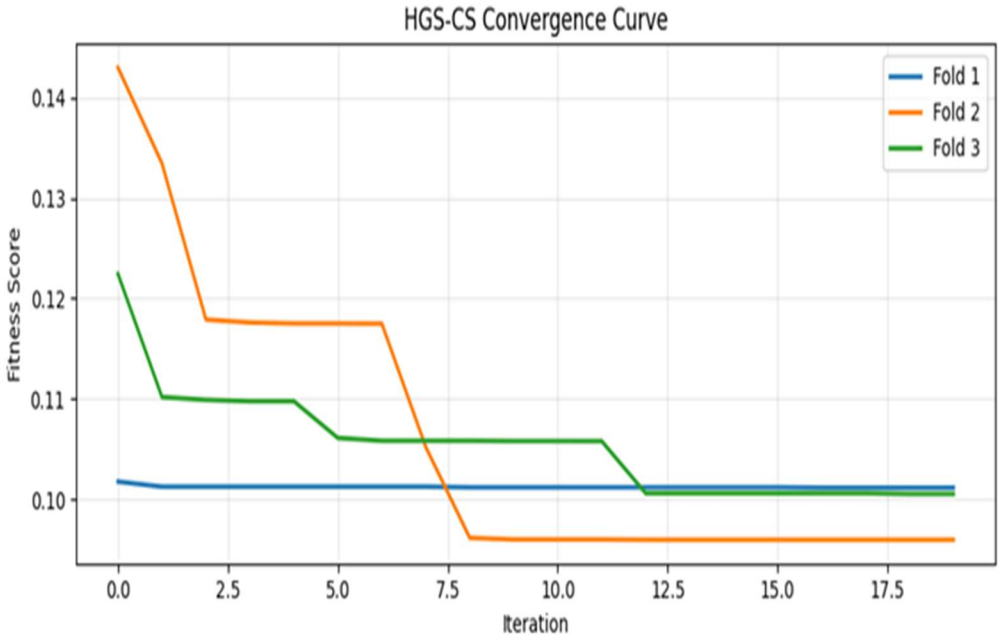

Figure 3: HGS-CS Convergence Curve on BoTNeTIoT-L01

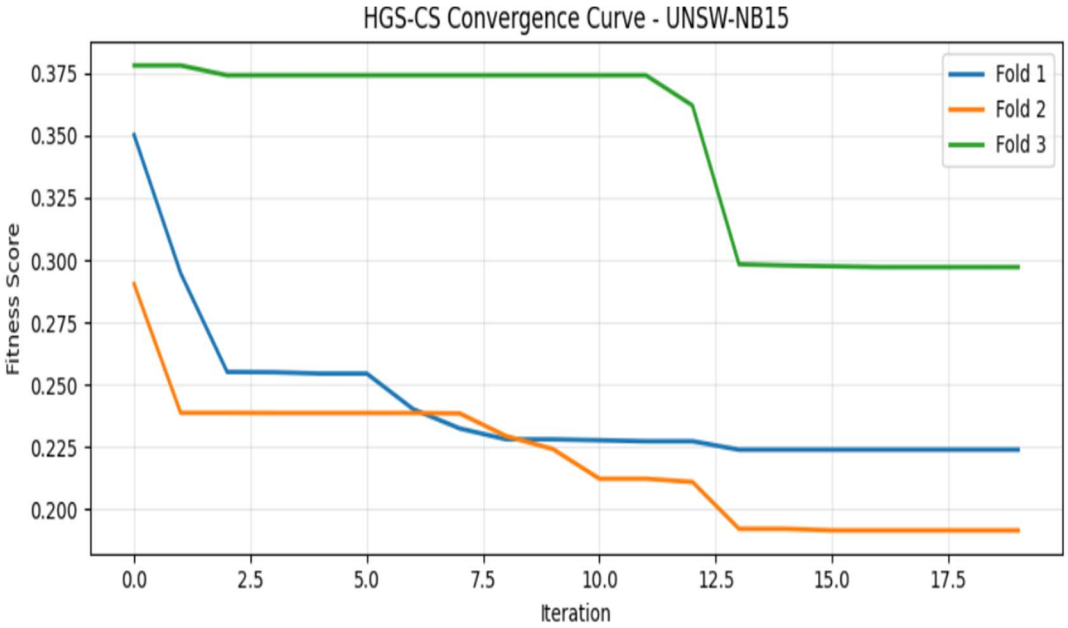

Figure 4: HGS-CS Convergence Curve on UNSWNB-15 Dataset

Table 4: HGS-CS Feature Selection Results Across Folds

Dataset	Fold	Iter 1 Fitness	Iter 6 Fitness	Iter 11 Fitness	Iter 16 Fitness	Best Fitness	Features Selected
BoTNeTIoT-L01	Fold 1	0.1017	0.1012	0.1012	0.1012	0.1011	6
	Fold 2	0.1430	0.1175	0.0960	0.0959	0.0959	5
	Fold 3	0.1224	0.1061	0.1058	0.1006	0.1005	6
UNSW-NB15	Fold 1	0.3501	0.2545	0.2277	0.2240	0.2240	13
	Fold 2	0.2904	0.2387	0.2123	0.1916	0.1916	14
	Fold 3	0.3781	0.3741	0.3741	0.2975	0.2972	17

Table 4 presents the feature subsets selected independently across different cross-validation folds. Due to variations in training data distribution, the number of selected features varies across folds (5 to 6 features for BoTNeTIoT-L01 and 13 to 17 features for UNSW-NB15). To obtain a robust and stable feature subset, a stability-based aggregation strategy was employed. Specifically, the frequency of selection of each feature across folds was computed, and features selected in at least 50% of the folds were retained in the final subset. This process reduces variability in feature selection and ensures that only consistently important features are chosen. As a result, the final feature sets were reduced to 3 features for BoTNeTIoT-L01 and 8 features for UNSW-NB15, as reported in Tables 3 and 5.

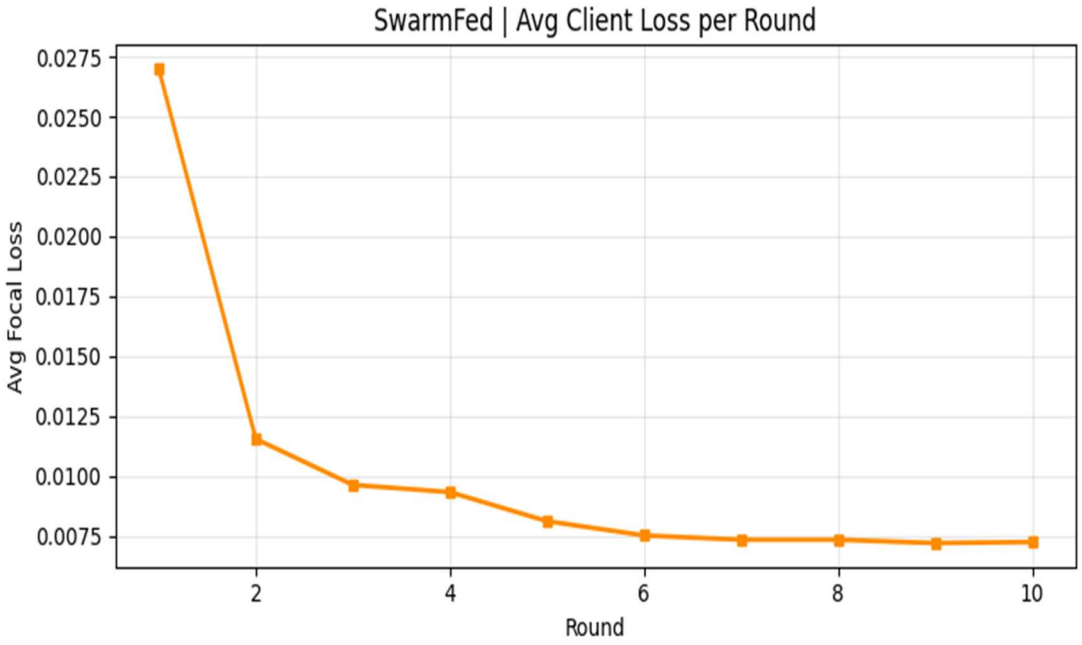

Figure 5: SwarmFed Avg Client Loss per round on BoTNeTIoT-L01 Dataset

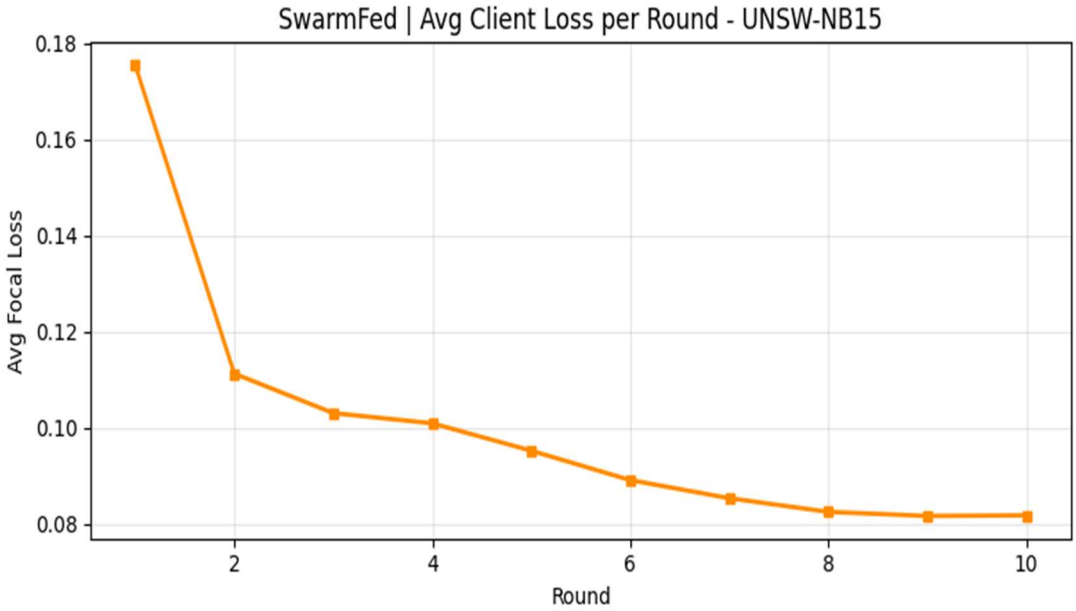

Figure 6: SwarmFed Avg Client Loss per round on UNSWNB-15 DATASET

Table 6: Comparative Performance Analysis on Base Line Models against Proposed Model On BoTNeTIoT-L01 and UNSWNB-15Dataset

Reference	Dataset	Features used	Accuracy	Precision (Normal/Attack)	Recall (Normal/Attack)	F1-Score (weighted)	ROC-AUC	Training Paradigm
[7]	UNSW-NB15	30	0.96	0.95 / 0.96	0.95 / 0.96	0.95	0.97	Federated
[9]	UNSW-NB15	30	0.96	0.95 / 0.95	0.95 / 0.95	0.95	0.96	Fed. Transfer Learning
[10]	UNSW-NB15	49	0.93	0.92 / 0.92	0.91/ 0.91	0.92	0.96	Federated
[6]	BoTNeTIoT-L01	10	0.95	0.94 / 0.97	0.95 /0. 98	0.97	0.98	Federated
[13]	BoTNeTIoT-L01	30	0.94	0.94 / 0.95	0.93 / 0.95	0.94	0.95	Domain-Adversarial
[16]	BoT-IOT	30	0.96	0.95 /0. 96	0.95 / 0.96	0.95	0.96	Federated
Proposed	**UNSW-NB15**	**8**	**0.89**	**0.92 / 0.88**	**0.87 / 0.92**	**0.89**	**0.95**	**Fed. + Centralised (GPU)**
Proposed	**BoTNeTIoT-L01**	**3**	**0.96**	**0.95 / 0.96**	**0.95 / 0.97**	**0.96**	**0.97**	**Fed. + Centralised (GPU)**

Table 7: Component-wise Ablation Study Based on F1-score

Model Configuration	HGS-CS	Transformer	BiLSTM	Fusion	SwarmFed	Focal Loss	Fine-Tuning	F1-score
Without Feature Selection	✗	✓	✓	✓	✓	✓	✓	0.9510
Without Transformer	✓	✗	✓	✓	✓	✓	✓	0.9535
Without BiLSTM	✓	✓	✗	✓	✓	✓	✓	0.9548
Without Fusion	✓	✓	✓	✗	✓	✓	✓	0.9560
Without SwarmFed	✓	✓	✓	✓	✗	✓	✓	0.9572
Without Focal Loss	✓	✓	✓	✓	✓	✗	✓	0.9585
Without Fine-Tuning	✓	✓	✓	✓	✓	✓	✗	0.9590
Full Proposed Model	✓	✓	✓	✓	✓	✓	✓	**0.9600**

Table 7 shows an ablation study was carried out to assess the component-wise was conducted to evaluate the contribution of each module in the proposed architecture using F1-score as the primary evaluation metric. The results show that removing any individual component leads to a decrease in performance. In particular, the absence of HGS-CS feature selection results in the largest performance drop (F1-score = 0.9600) to 0.9510, highlighting its importance in selecting the most informative features. Similarly, the Transformer and BiLSTM modules contribute to effective temporal feature extraction, while the fusion mechanism enhances feature integration. SwarmFed improves distributed learning performance, and focal loss helps address class imbalance. Centralized fine-tuning further refines the model. The complete model achieves the highest F1-score of 0.9600, confirming the effectiveness of the integrated architecture.

5.3 SwarmFed: Federated Training Convergence

The proposed SwarmFed framework was used to perform the federated learning stage to allow privacy-preserving model training across decentralized clients. The training across clients shows a stable and consistent convergence, as the global loss goes down to 0.0270 in the first communication round to 0.0072 in the 10th communication round. It is important to note that convergence curve has a smooth decreasing trend with no sharp oscillations, which suggests consistent aggregation of client updates through the attention-based weighting mechanism used in SwarmFed. This may be especially relevant in practical IoT settings, where data distributions tend to be non-independent and non-identically distributed (non-IID).

The above findings prove that this federated learning approach is efficient and robust, fits in the IOT and Network intrusion detection systems that are decentralized and privacy-sensitive.

5.4 Centralized Fine-Tuning and Classification Performance

In this Centralized Fine-Tuning After the federated training process, the proposed model was further refinement in centralized fine-tuning to achieve the predictive performance. The TSTformer-LSTM model was able to converge fastly with a final training loss of 0.0045 in 20 epochs, which implies effective learning and reliable optimization.

5.5 Discussion

The results demonstrate that the proposed framework provides competitive intrusion detection performance compared with existing approaches. On the UNSW-NB15 dataset, the proposed framework achieved 89% accuracy using only eight selected features, while the compared methods reported 93–96% accuracy using 30–49 features. On the BoTNeTIoT-L01 dataset, the proposed framework achieved 96.9% accuracy using only three selected features, compared with 94–96% accuracy reported by the compared approaches using 10–30 features. The major advantage of the proposed framework is therefore not accuracy alone, but the combination of competitive detection performance and substantial feature reduction. HGS-CS reduced the BoTNeTIoT-L01 feature space from 23 to 3 features, corresponding to an 86.96% reduction, and reduced the UNSW-NB15 feature space from 51 to 8 features, corresponding to an 84.31% reduction. This reduction can potentially decrease the computational and input-processing requirements of the intrusion detection system. The proposed framework also integrates feature selection, Transformer-based contextual learning, BiLSTM-based temporal learning, and federated learning within a unified architecture. The ablation study supports the contribution of these components, as removing individual components resulted in lower F1-scores than the complete model, which achieved an F1-score of 0.96 However, the comparison with previous studies should be interpreted cautiously. The compared studies may use different preprocessing methods, feature subsets, class-balancing techniques, train-test splitting procedures, model architectures, and evaluation protocols. Therefore, the reported results demonstrate competitive performance but should not be interpreted as absolute superiority under identical experimental conditions.

The present study also has limitations. The experiments were conducted using balanced binary classification and benchmark datasets under controlled conditions. Real-world IoT environments may involve multiclass attacks, naturally imbalanced traffic, heterogeneous non-IID client data, evolving traffic patterns, and resource constraints. Communication overhead, inference latency, energy consumption, and formal privacy guarantees were also not comprehensively evaluated. These limitations provide important directions for future research.

6.CONCLUSION AND FUTURE SCOPE

This study proposed a hybrid intrusion detection framework integrating HGS-CS feature selection, a TSTformer-LSTM deep learning model, and SwarmFed federated learning to address challenges of high dimensionality, computational complexity, and data privacy in IoT environments. The proposed framework achieved 96.9% accuracy on the BoTNeTIoT-L01 dataset and 89% accuracy on the UNSW-NB15 dataset while reducing the feature space from 23 to 3 features (86.96%) for BoTNeTIoT-L01 and from 51 to 8 features (84.31%) for UNSW-NB15. Dataset respectively demonstrating the effectiveness of the HGS-CS algorithm in selecting highly discriminative features., further demonstrating the applicability of the proposed framework across different network intrusion datasets while maintaining a compact feature representation. The hybrid Transformer BiLSTM architecture effectively captured both global and temporal patterns, while SwarmFed enabled privacy-preserving distributed training with stable convergence. The framework significantly reduces feature dimensionality and computational complexity, making it a promising candidate for deployment in resource-constrained IoT environments.The ablation study further confirmed that the proposed feature selection method maintains high performance with minimal feature usage. However, the present study is limited to balanced binary classification **a**nd evaluates the proposed framework using benchmark datasets under controlled experimental conditions. Real-world IoT environments are often characterized by multiclass attacks, naturally imbalanced data distributions, non-IID client data, and temporal evolution of network traffic**,** which were beyond the scope of the current work. Future work will focus on extending the proposed framework to multiclass intrusion detection**,** evaluating it on naturally imbalanced IoT datasets, investigating non-IID federated client settings**,** as model size, inference latency, communication overhead, and energy consumption Evaluating it on additional IoT datasets such as TON_IoT, and deploying it on edge devices. Furthermore, incorporating differential privacy and online learning mechanisms can enhance adaptability and robustness in dynamic environments

Data availability

https://research.unsw.edu.au/projects/unswnb15-dataset

https://www.kaggle.com/datasets/azalhowaide/iot-dataset-for-intrusion-detection-systems-ids

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper's conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing original draft preparation, writing review and editing, visualization have been done by 1st. The supervision and project administration have been done by the 2nd author

REFERENCES

[1] H. Liu and B. Lang, ‘‘Machine learning and deep learning methods for intrusion detection systems: A survey,’’ Appl. Sci., vol. 9, no. 20, p. 4396, Oct. 2019.

[2] UmaDevi, P., & Chhabra, G. S. (2026). Enhancing Intrusion Detection Using Semi-supervised Learning and Knowledge Distillation Method. *International Journal of Intelligent Engineering & Systems*, *19*(5), 1090. Doi:1090-1108. 10.22266/ijies2026.0531.61.

[3] Tanyıldız, H., Batur Şahin, C., & Batur Dinler, Ö. (2024). Federated Learning for Attack Prediction on UNSW-NB15 Training Data. NATURENGS, 5(2), 1-7. https://doi.org/10.46572/naturengs.1524394

[4] Khraisat, A., Talukder, M.A., Uddin, M.A. *et al.* RF-FedAvg: Federated learning-based random forest model for intrusion detection in wireless sensor networks. *Cluster Comput* 28, 873 (2025). https://doi.org/10.1007/s10586-025-05591-8

[5] C. S. Abhijit, R. Himmatramka, A. S. R. Bishoyi, L. Prasath and T. Subbulakshmi, "Privacy-Preserving Network Intrusion Detection using Federated Learning," *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India, 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10307648

[6] Devi, P. Uma, and Gurpreet Singh Chhabra. "A Comprehensive Survey of Deep Learning Methods in Network Intruder Detection System." *AI-Driven Competitive Intelligence and Next-Generation Security* (2025): 89-105.

[7] Bai, Ye, Jiang, Weiwei, Mu, Jianbin, Liu, Shang, Gu, Weixi, Wang, Shuke, Enhancing IoT Security via Federated Learning: A Comprehensive Approach to Intrusion Detection, IET Information Security, 2025, 8432654,16pages,2025. https://doi.org/10.1049/ise2/8432654

[8] M. A. Azad, A. K. Bashir, R. M. A. Azad and S. A. Shah, "WFL: Edge-Enabled Weighted Federated Learning for Securing Heterogeneous Networks," in *IEEE Communications Standards Magazine*, vol. 10, no. 1, pp. 10-17,March2026,doi: 10.1109/MCOMSTD.2025.3584703

[9] Andrea Bellmunt, Beatriz Otero, Eva Rodríguez, Xavier Masip-Bruin,Federated transfer learning-based intrusion detection system in 5G networks,Expert Systems with Applications,Volume 305,2026,130868,ISSN 0957-4174,https://doi.org/10.1016/j.eswa.2025.130868.

[10] Alamro S (2026) AnomLocal: A hybrid local-global anomaly detection model for network security using federated learning. PLoS One 21(2): e0339981. https://doi.org/10.1371/journal.pone.0339981

[11] R. Kumar, J. Dutta, N. Vamsi, U. Sankararao Varri and D. Puthal, "Next-Generation Security in the 6G Era: The Role of AI in Safeguarding Future Networks," in *IEEE Access*, vol. 14, pp. 17347-17380, 2026, doi: 10.1109/ACCESS.2025.3650208.

[12]R.Parada, V.Monzon Baeza, C.H.F.deGamboa , R.Serrano Camacho,and C.Monzo, "Integrated LEO Constellation and In-Cabin Distribution System for Continuous Aircraft 5G Connectivity," International Journal of Satellite Communications and Networking (2026): 1–13, https://doi.org/10.1002/sat.70046.

[13] Jin, H. Cross-Protocol Domain Gap in Internet of Things Intrusion and Anomaly Detection: An Empirical Internet Protocol-to-Bluetooth Low Energy Study of Domain-Adversarial Training. *Sensors* 2026, *26*, 1184. https://doi.org/10.3390/s26041184

[14] Ilias, L.; Doukas, G.; Lamprou, V.; Mouzakitis, S.; Ntanos, C.; Askounis, D. A Federated Learning-Based Network Intrusion Detection System for 5G and IoT Using Mixture of Experts. *Electronics* 2026, *15*, 1057. https://doi.org/10.3390/electronics15051057

[15] N. F. Hasan and A. Saad, "Federated Learning-Based Intrusion Detection System Using Convolutional Neural Networks for IoT Networks," *2026 5th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*, Rajshahi, Bangladesh, 2026, pp. 1-6, doi: 10.1109/ICECTE69292.2026.11429438.

[16] Mankotia, S.; Conte de Leon, D.; Rimal, B.P. FedPrIDS: Privacy-Preserving Federated Learning for Collaborative Network Intrusion Detection in IoT. *J.*

[17] Moustafa, Nour, and Jill Slay. *"UNSW-NB15: A comprehensive dataset for network intrusion detection systems (UNSW-NB15 network dataset)."* In **2015 Military Communications and Information Systems Conference (MilCIS)**, pp. 1–6. IEEE, 2015. doi: 10.1109/MilCIS.2015.7348942

[18] A. Alhowaide, I. Alsmadi, J. Tang. "Towards the design of real-time autonomous IoT NIDS", Cluster Computing (2021), pages 1-14, Jan 2021.

[19] U. C. Akuthota and L. Bhargava, "Transformer-Based Intrusion Detection for IoT Networks," in *IEEE Internet of Things Journal*, vol. 12, no. 5, pp. 6062-6067, 1 March1, 2025, doi: 10.1109/JIOT.2025.3525494

[20] G. P.V., N. S. Sethumadhavan and L. Vishnu Namboothiri, "Exploring a Hybrid BiLSTM-Transformer Model for Detecting Intrusions in Federated Learning Settings," *2025 5th International Conference on Pervasive Computing and Social Networking (ICPCSN)*, Salem, India, 2025,pp.367-374,doi: 10.1109/ICPCSN65854.2025.11034920.