Journal of Discrete Mathematical Sciences & Cryptography
ISSN 0972-0529 (Print), ISSN 2169-0065 (Online)
Vol. 28 (2025), No. 7, pp. 2889–2898
DOI : 10.47974/JDMSC-2559

CyberShield-H : A hybrid sequential deep learning model for robust DDoS mitigation in critical healthcare IoT systems

P. R. Anisha *
Kaniti Anjali †
Department of Computer Science and Engineering
Stanley College of Engineering and Technology for Women
Hyderabad 500001
Telangana
India

Abstract

The growing reliance on Healthcare IoT (H-IoT) systems creates the potential for Distributed Denial of Service (DDoS) attacks on critical services. This research introduces a hybrid deep learning model with a combination of GRU and LSTM networks for real-time, multi-class DDoS detection from flow-aware temporal traffic consecutive data. To enhance robustness, adversarial training with the Fast Gradient Sign Method (FGSM) was used with PCA for efficient operational efficiency with less dimensionality while retaining discriminative features. The adversarially trained GRU-LSTM model achieved a clean data accuracy level of 94.99% and an adversarial accuracy level of 91.17% which is a good measure of robustness. Additionally, interpretability was articulated and accomplished using LIME to show the contributions of the features used in the model. Overall, this work introduces an interpretable and stable framework to provide an insight into the security implications and considerations for H-IoT facing demonstrated DDoS attacks.

Subject Classification: *Primary 68T07, Secondary 68M25.*

Keywords: *Adversarial training, Deep learning, DDoS detection, Healthcare IoT, LIME interpretability, Temporal sequence modeling.*

1. Introduction

The digitization of healthcare systems, which ranges from electronic health records (EHR) to artificial intelligence (AI)-based platforms, has increased the likelihood of Distributed Denial-of-Service (DDoS) attacks.

* *E-mail:* **pranisha@stanley.edu.in (Corresponding Author)**
† *E-mail:* **anjalikaniti02@gmail.com**

Nearly 90% of hospitals report being at risk, with costs exceeding $125 billion by 2025 [1]. Also, there are large-scale attacks; for instance, the attack on Boston Children's Hospital [2] highlighted the need for a more robust defense. Although there are techniques based on deep learning that support DDoS detection, there are still various issues, including false alarms, lack of generalizability, and adjustment to real-time attacks [3]. Recently developed machine learning approaches and hybrid optimization frameworks have aimed to improve the accurate and adaptive detection of DDoS attacks in dynamic network settings [9-10].

The key contributions of this paper include:

1. A hybrid GRU-LSTM architecture that captures both short-term and long-term traffic dependencies to enhance effective DDoS detection for H-IoT networks.
2. The use of principal component analysis (PCA) for efficient preprocessing, and adversarial training to allow resilience against dynamically planned and manipulated amplification of attack behaviors.
3. The use of LIME-based XAI to enhance human-understandability and engender trust in a critical health care context.

2. Materials and Methods

2.1 *Description of Dataset*

In order to create a strong and interpretable deep learning framework for real-time Detection of DDoS in Healthcare IoT, the CIC-DDoS2019 dataset, developed by the Canadian Institute for Cybersecurity [4] was used as the benchmark. A detailed overview is presented in Table 1, which summarizes its key attributes.

Table 1
An overview of CIC-DDoS2019

Attributes	Count	Description
Total Samples	528,970	Number of flow-based instances in the dataset
Total Features	87	Original number of features
Flow-based Features	4	Flow-level statistics (Flow Duration, Flow Bytes/sec etc.)

Contd...

Packet-based Features	30	Packet-related statistics (Packet Size, Fwd Length etc.)
Flag-based Features	12	Flag information (TCP flags, ACK, SYN, PSH etc.)
Traffic-based Features	8	Number of active connections, window size
IAT-based Features	14	Inter-arrival time features (mean, std, min/max) for forward and backward packets
Target Labels	7	BENIGN, Syn, Portmap, LDAP, UDP, MSSQL, NetBIOS

2.2 *Data preprocessing*

Prior to modeling on the CIC-DDoS2019 dataset, several preprocessing steps were performed. Fixed-length temporal sequences were created through normalization to manage class imbalance. Redundant and highly correlated were removed to decrease noise; categorical features (i.e. labels, IPs) were label-encoded. The features underwent standardizing with Z-score normalization:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

Where, x represents the original value, μ is the mean of the feature, σ is the standard deviation of the feature and z is the standardized value.

To substantiate feasibility, PCA was applied to reduce dimensionality while retaining over 95% of the variance, thereby improving efficiency and generalization. The effect of PCA in compressing feature space while

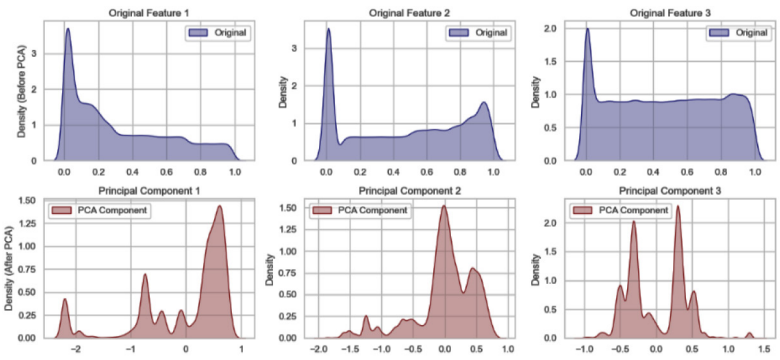

Figure 1

Comparison of Selected Features Before and After PCA

preserving key patterns is illustrated in Figure 1. These were then were then structured into temporal sequences and passed to the proposed model.

2.3 *Sequence Modeling Preparation*

To facilitate temporal learning, the PCA-reduced features were segmented into overlapping fixed-width windows (10 timesteps × 9 features) for the GRU-LSTM model with adversarial training. Each fixed-width window included continuous network activity where the class label was based on the last timestep to reflect the real-world practice of making a decision (Figure 2). This format preserved temporal dependencies that are critical to the model for classification.

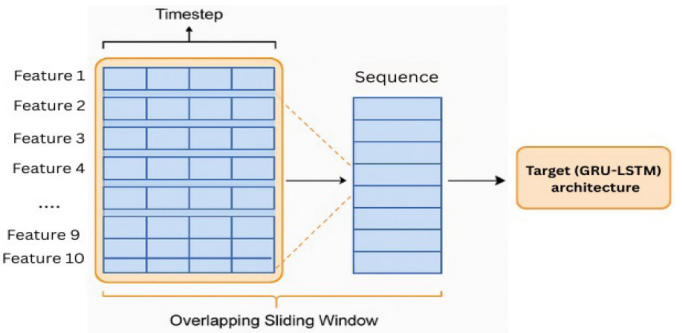

Figure 2

Sequence Construction from Temporally Balanced Network Traffic

2.4 *Model Construction for Adversarial Training*

This study presents a sequential deep learning model for multi-class DDoS detection using GRU and LSTM layers on PCA-transformed traffic. The model represents input as sequences of 10 timesteps × 9 features to incorporate short-term and long-term dependencies before feeding to dense layers for classification. The model is implemented in Keras using GRU and LSTM layers with selected units, followed by a Dense layer with sigmoid activation to classify subjects into two classes. A Dropout layer was added to reduce overfitted outcomes, and the final structure has an Adam optimizer, and Binary Crossentropy as a loss function to fulfill the process as effectively as possible. The GRU-LSTM model (Figure 3) combines each component into a solely sequential flow:

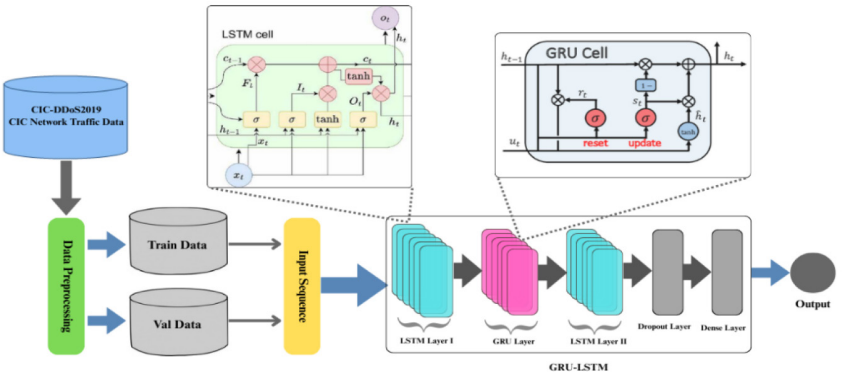

Figure 3

Proposed Model Algorithm

1. The GRU layers consider short-term dependencies relevant to PCA-reduced sequences of traffic, with minimal parameter overhead.
2. The LSTM layers are designed to consider the long-term temporal relationships that are important for delayed or burst-based attacks.
3. The combined representation from the GRU and LSTM layers is output to an intermediate dense layer for regularization and a dense softmax layer for a multi-class probability distribution.

Model robustness was improved using FGSM-based adversarial training, where small ε-scaled perturbations were applied to inputs. The GRU–LSTM was trained on both clean and adversarial sequences, vulnerability to adversarial attacks (Figure 4).

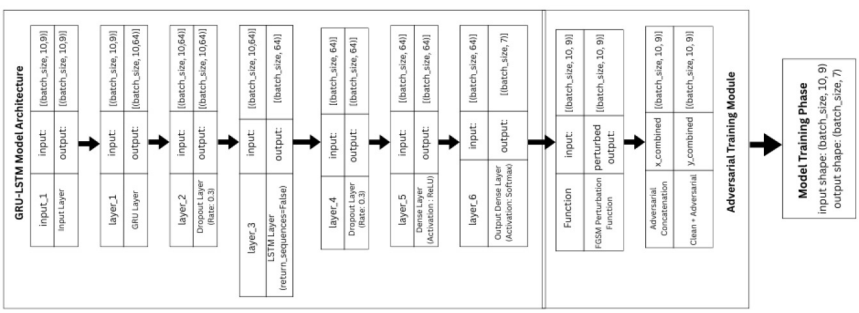

Figure 4

Proposed GRU-LSTM Model: Architecture and Pipeline Overview

3. Results and analysis

The adversarially trained GRU-LSTM model was evaluated on stratified CIC-DDoS2019 traffic, allowing the evaluation to be time-aware and no data leakage to occur.

3.1 *Evaluation of Model Training Process*

The training of adversarial GRU-LSTM model (Figure 5) displays a successful learning, no overfitting and a good generalization. The GRU-LSTM is able to capture temporal dependencies in network traffic; showing a stable convergence needed for DDoS detection.

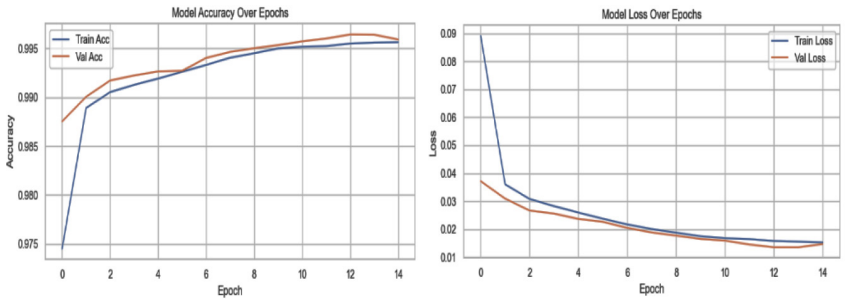

Figure 5

Evaluation of Model Accuracy and Loss

3.2 *Evaluation of Model Performance*

The proposed GRU-LSTM model accurately distinguishes benign and multiple classes of DDoS attacks, effectively captures both major and

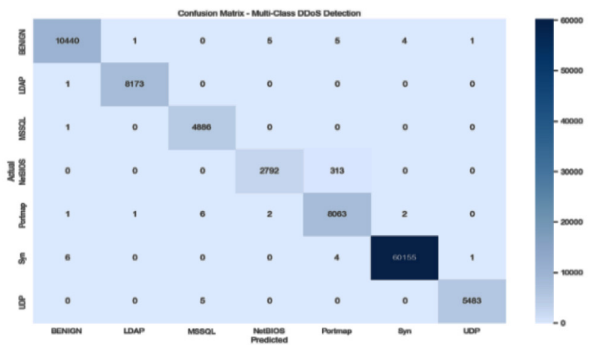

Figure 6

Confusion Matrix for Adversarial GRU-LSTM model

minority classes, with misclassifications tending to occur among temporally similar traffic (Figure 6).

The GRU-LSTM model competently distinguished benign from multiple DDoS attack classes achieving an accuracy of 94.99% when classifying inputs from clean data and an accuracy of 91.17% when classifying adversarial FGSM inputs. This demonstrates robust and stable classification of benign and attack classes with small input perturbations.

3.3 *Post-hoc Interpretability of Adversarially Trained GRU-LSTM Model*

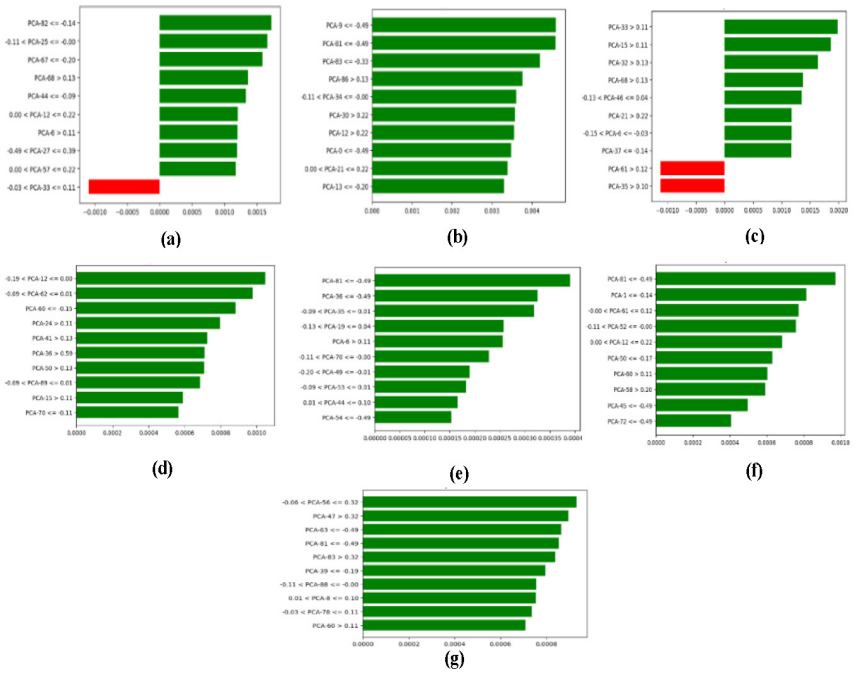

Figure 7

LIME-based Interpretability for Individual Classes in the proposed model

(6a) BENIGN: Strong positive contributions for normal traffic.

(6b) SYN DDoS: Key features detect aggressive packet bursts.

(6c) Portmap: Features reflect consistent port scanning.

(6d) MSSQL: Mixed contributions with overlapping traffic patterns.

(6e) LDAP: High-impact features for repeated directory queries.

(6f) NetBIOS: Sparse but decisive features guide prediction.

(6g) UDP: Compact signal activations shape confidence.

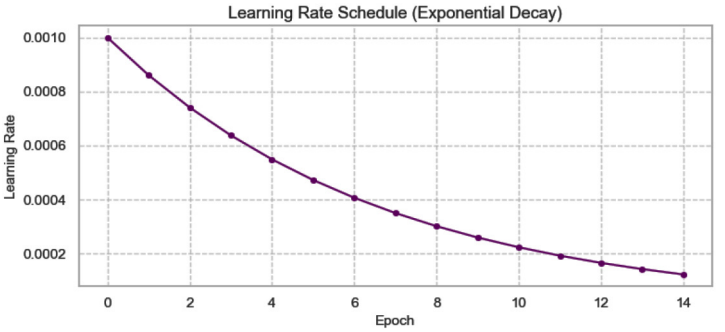

Figure 8

Learning Rate Schedule over the epochs

3.4 *Learning Rate Schedule Analysis*

An Exponential Decay Learning Rate schedule was implemented by starting at 0.001 and reducing the rate through epochs to achieve faster convergence in a more stable manner as displayed in Figure 8.

3.5 *Comparative Evaluation with Baseline Models*

In order to evaluate the performance of the adversarial GRU - LSTM model, evaluation with the baseline of stand-alone LSTM and GRU models were performed under the same training settings. As it can be seen in Table 2, the hybrid model has consistently performed better than the baselines despite different evaluation measures, confirming its better performance in learning time dependencies for DDoS detection.

Table 2

Baseline Model Comparison

Model	Accuracy (%)	Macro-F1	Weighted-F1
Standard LSTM	85.15	0.48	0.80
Standard GRU	90.72	0.74	0.89
Proposed GRU-LSTM	94.99	0.80	0.94

3.6 *Comparative Analysis of various DDoS detection methodologies*

To contextualize performance, we compare the proposed GRU-LSTM model to existing DDoS detection methods from the recent literature, which employ deep learning, machine learning, and hybrid approaches, using standard benchmark datasets. In Table 3, we summarize the

architecture of each method, its associated dataset, strengths/weaknesses, and distinguishing features.

Table 3

Comparative Analysis with other DDoS detection methodologies

S. No.	Methodology	Accuracy	Overall Effectiveness and Concerns
1	Mousa et al. [5]	93.76%	High accuracy; complex for real-time use.
2	Martinez et al. [6]	94.85%	Strong performance on few-shot samples; scalability to large-scale traffic patterns untested.
3	Elsayed et al. [7]	92.54%	Efficient deep architecture; tested only on CIC-IDS2017 dataset. Adversarial noise not considered.
4	Bolodurina et al. [8]	93.51%	Focused on imbalance mitigation; Emphasized relevance of balanced data sampling.
5	Proposed (GRU-LSTM + Adversarial Training) Model	94.99%	Multi-class robust architecture; preserves flow-sequence semantics.

4. Discussion

The results suggest that the approach is viable for healthcare systems, offering a dependable service even amidst noise or adversarial attacks across facilities, telemedicine, and patient monitoring. LIME-based interpretability assists providers in integrity checks of alerts, conservatively biasing against threats while filtering out false alarms, and offers transparent event management, all in ways that fortify trust and build safety in practice.

5. Conclusion

This study presents a GRU-LSTM model that is adversarially trained for DDoS detection in healthcare systems that is both high-performance and interpretable. The model was able to achieve 94.99% accuracy in detecting the more complex attack patterns, and maintained over 91% accuracy in the adversarial cases, thereby ensuring reliability in a noisy or uncertain environment. Using LIME allows for the process of transparent

decision-making, which is an important aspect when it comes to sensitive healthcare applications. The adversarial training, temporal modeling, and explainability provide a secure and interpretable class of model appropriate for real-world healthcare and other critical infrastructure settings.

References

[1] R. Udayakumar, A4. Joshi, S. and R. Sugumar, Deep Fraud Net: A Deep Learning Approach for Cyber Security and Financial Fraud Detection and Classification, *Journal of Information Systems and Information Security*, vol. 13, no. 4, pp. 138–157, (2023).

[2] DDoS Attacks in Healthcare, Center for Internet Security, (2025).

[3] M. Mittal and S. Behal, Deep learning approaches for detecting DDoS attacks - a systematic review, *Soft Computing*, vol. 27, no. 1, pp. 1–37, (2021).

[4] CIC- DDoS2019 dataset, Kaggle, (2025).

[5] A. K. Mousa and M. N. Abdullah, An Improved Deep Learning Model for DDoS Detection Based on Hybrid Stacked Autoencoder and Checkpoint Network, *Future Internet,* vol. 15, no. 8, pp. 278, (2023).

[6] F. Martinez, M. Mapkar, and A. Alfatemi, Redefining DDoS Attack Detection Using a Dual-Space Prototypical Network-Based Approach, *Computer Communications and Networks*, pp. 279–287, (2024).

[7] M. S. Elsayed, N.-A. Le-Khac, S. Dev, and A. D. Jurcut, DDoSNet: A Deep-Learning Model for Detecting Network Attacks, *Computers & Security,* vol. 89, pp. 101-129, (2020).

[8] I. Bolodurina, A. Shukhman, and D. Parfenov, Classifying Unbalanced Datasets in Identifying DDoS Attacks, *Journal of Physics: Conference Series,* vol. 1679, no. 4, pp. 042020, (2020).

[9] J. Mante and K. Kolhe, A novel DDoS detection framework leveraging hybrid binary grey wolf optimization and explainable AI, *Journal of Information & Optimization Sciences*, vol. 46, no. 2, pp. 563–578, (2025).

[10] P. K. Swain, A. Patnaik, and S. Satpathy, Hybrid machine learning model for network anamoly detection using time-series forecasting, *Journal of Information & Optimization Sciences,* vol. 46, no. 6, pp. 2311-2322, (2025).

Received July, 2025