International Journal of Computer Networks and Applications (IJCNA)
DOI: 10.22247/ijcna/2025/20 Volume 12, Issue 3, May – June (2025)

RESEARCH ARTICLE

Deep Learning Approach for Protecting IoT Smart Home Devices Against Multiclass Botnet Attacks

Haifa Ali Saeed Ali
Department of Computer Applications, CMR Institute of Technology, Bangalore (Affiliated to Visvesvaraya Technological University (VTU), Belagavi), Karnataka, India.
shiekhhaifa@gmail.com

Vakula Rani J
Department of Computer Applications, CMR Institute of Technology, Bangalore (Affiliated to Visvesvaraya Technological University (VTU), Belagavi), Karnataka, India.
vakula.r@cmrit.ac.in

Received: 19 February 2025 / Revised: 19 May 2025 / Accepted: 30 May 2025 / Published: 30 June 2025

Abstract – As the Internet of Things (IoT) continues to expand, smart homes have emerged as a prominent application area, characterized by a network of interconnected devices designed to improve daily life through automation, enhanced security, and energy management. Despite these benefits, the integration of numerous connected devices also brings about significant security challenges, particularly the threat of botnet attacks capable of disrupting entire device networks. The growing number and variety of smart home devices make it increasingly complex to establish effective security mechanisms. To address this concern, this study investigates the use of deep learning (DL) models in combination with Feature Engineering (FE) techniques to identify and classify multiple types of botnet attacks targeting smart home IoT systems. The research focuses on evaluating DL models, specifically Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), and Autoencoders (AE), using the BoT-IoT dataset. To manage the high dimensionality of the data, various features selection (FS) and reduction methods are applied, including Principal Component Analysis (PCA), Random Forest (RF) selection, and a hybrid PCA-RF method. These techniques help isolate the most relevant attributes, allowing the models to concentrate on key features that improve classification performance. A comparative analysis was performed to examine how each DL model responds to different FE strategies. The results demonstrated that both RNN-based models and AE showed strong performance in identifying botnet-related anomalies, reinforcing their suitability for smart home intrusion detection. The study also underscores the importance of employing advanced feature engineering methods to strengthen botnet detection capabilities within IoT environments.

Index Terms – IoT, IoT Security, Smart Home, DL Algorithms, Botnet Attack, Denial of Service (DoS), Distributed Denial of Service (DDoS).

1. INTRODUCTION

The IoT signifies a transformative technological development where everyday objects are networked over the internet, enabling them to gather, exchange, and act upon data [1]. This organization features a diverse range of gadgets, like sensors, healthcare monitors, household appliances, and personal gadgets, all interconnected through the IoT. While this connectivity enhances automation, efficiency, and convenience across industries, it raises significant security and privacy concerns. Among the diverse applications of IoT, smart homes stand out, providing enhanced comfort, energy management, and security through remotely controlled devices like thermostats, lighting, and security systems. These interconnected devices create a seamless living environment with automated and personalized functions for homeowners. Despite the numerous advantages of smart homes, assuring the security of these interconnected appliances remains paramount because of their role in handling sensitive information and controlling essential household operations [2]. The vulnerability of such systems poses risks of unauthorized access, data theft, and even physical breaches, underscoring the challenges of their networked nature, which creates multiple access points for potential assaulters. As the adoption of smart home technology continues to rise, the expanding attack surface further exposes users to cyber intrusions, emphasizing the urgent demand for strong security measures across the entire ecosystem. Among the growing threats to smart home devices, botnet attacks are particularly prevalent. A botnet, composed of compromised gadgets controlled by malicious actors, is often employed to execute coordinated operations like DoS attacks [3], DDoS attacks [4], data breaches, and unauthorized system access.

The detection and mitigation of botnet activities in IoT networks is a challenging task even though botnet threats are serious. It may also be particularly difficult for a conventional network intrusion detection system (IDS) to disclose botnet behavior in the smart home environment, due

RESEARCH ARTICLE

to the complexity and variety of IoT traffic. These systems generally depend on established criteria and signatures, which may prove inadequate for identifying changing botnet strategies that continuously evolve to circumvent conventional detection techniques. Additionally, the real-time nature of data generated by IoT gadgets, which are often vast in scale and high-dimensionality, further complicates matters. Botnet traffic can be complicated to distinguish from legitimate network activity, making it challenging for algorithmic detection methods to identify and stop the bots before the attack reaches its target, especially when traffic is obfuscated or disguised by the botnet.

1.1. Problem Statement

Traditional detection systems often rely on simple binary classification models that distinguish only between normal and malicious network traffic. While this may be sufficient in basic scenarios, it falls short in identifying and differentiating between various types of botnet attacks such as DDoS, DoS [5], reconnaissance activities, and data breaches. These attack types exhibit distinct patterns and therefore require more tailored detection techniques. Furthermore, existing systems are not well-equipped to manage the complexity of IoT traffic, which is typically characterized by high dimensionality, noise, and heterogeneity. Many conventional approaches suffer from limited accuracy, poor scalability, and weak generalization when applied to large and dynamic datasets. As botnets continue to evolve exploiting new vulnerabilities and deploying advanced attack strategies, there is a pressing need for intelligent, adaptable, and efficient solutions. This calls for the development of advanced ML models capable of accurately identifying and classifying multiple types of botnet threats within the ever-changing landscape of smart home networks.

Recent advancements in ML, especially DL, have demonstrated encouraging outcomes in addressing intricate classification challenges across many fields [6, 7]. One of the principal obstacles in implementing DL for IoT security is the large dimensionality of IoT data, which may result in overfitting and heightened processing demands. Consequently, it is imperative to devise techniques that might diminish data complexity while preserving critical information. In this context, dimensionality reduction and feature selection are vital for boosting the performance and computational efficacy of ML models.

To tackle this issue, we apply a new hybrid detection method that combines PCA based dimensionality reduction and attention-based RF for features selection for multi-class botnet attacks in smart home environments. through the application of these methodologies, we seek to enhance the accuracy, resilience, efficacy of botnet finding the drawbacks of conventional recognition strategies.

1.2. Smart Home Architecture

A smart home powered by IoT connects various household devices, enabling automated and remote control through internet-based networks. It includes systems for security, lighting, appliances, and temperature control, enhancing convenience and efficiency. Smart home [8] setups often integrate sensors [9] and controllers for real-time monitoring and management, aiming to optimize resource usage and improve the overall living experience. In smart home systems, the multitude of networked IoT devices produces extensive data, resulting in high-dimensional feature spaces. These devices, including sensors, cameras, and appliances, generate several data points, such as temperature measurements, motion detection, and consumption of energy. The architecture of a smart home involves smart devices linked to a gateway that connects these devices to the internet, as demonstrated in Figure 1. The diagram illustrates a scenario of IoT security in a smart home, emphasizing interactions among different components and the potential cyber threats they face. IoT smart homes, features interconnected devices like smart TVs, cameras, and locks, offer remote control and automation but are vulnerable to cyberattacks. The vast quantity and variety of this data provide considerable hurdles for conventional data analysis techniques, as high-dimensional data frequently encompasses redundant, noisy, and unwanted features. To address this issue, the only solution to effectively tackle that the Complexity of the smart home environment is the Management of the high-dimensionality of data created by these connected devices. Smart home appliances generate torrents of heterogeneous data via different sensors, appliances, and systems that can be overwhelming. However, this high-dimensional data is often noisy, redundant, and hard to interpret, posing challenges in deriving useful insights. As a result, it is imperative to prioritize dimensionality reduction by identifying the most significant features.

This paper introduces a hybrid machine learning methodology that integrates PCA for dimensionality reduction with RF for feature selection, with the objective of identifying multi-class botnet attacks in smart home networks. The main contributions of this work are as follows:

1. To develop hybrid feature engineering scheme. First, PCA is applied to decrease the dimensionality of the dataset. PCA achieves this by mapping the data to a lower-dimensional space—and removes noise and redundant information in the process. Then, RF is adopted to select features by determining the importance relevance score that calculates the relevant features of the feature set for the purpose of classification. This feature selection approach guarantees that the model prioritizes the most significant elements of the data, hence enhancing classification performance.

RESEARCH ARTICLE

2. To concentrate on multi-class botnet detection, which involves classifying various attacks of different kinds (DDoS, DoS, reconnaissance, and theft) separately. Using a multi-class approach gives more accurate detection, because the model can identify different botnet attack types from their features. They are vital for real world use cases, where botnets can have non-standard behaviors and attack vectors.
3. To combine the hybrid of PCA, RF with DL Algorithms These techniques are then used to feed deep learning models, which will analyze this processed data and classify it into their respective attack categories. This integration guarantees that the system is computationally efficient and proficient in accurately identifying multi-class botnet attacks.
4. To assess the effectiveness of several DL on the BoT-IoT dataset in smart home devices, and the performance of these algorithms compared with other new methods using metrics such as accuracy, precision, recall, F1-score, and FPR, and other parameters.

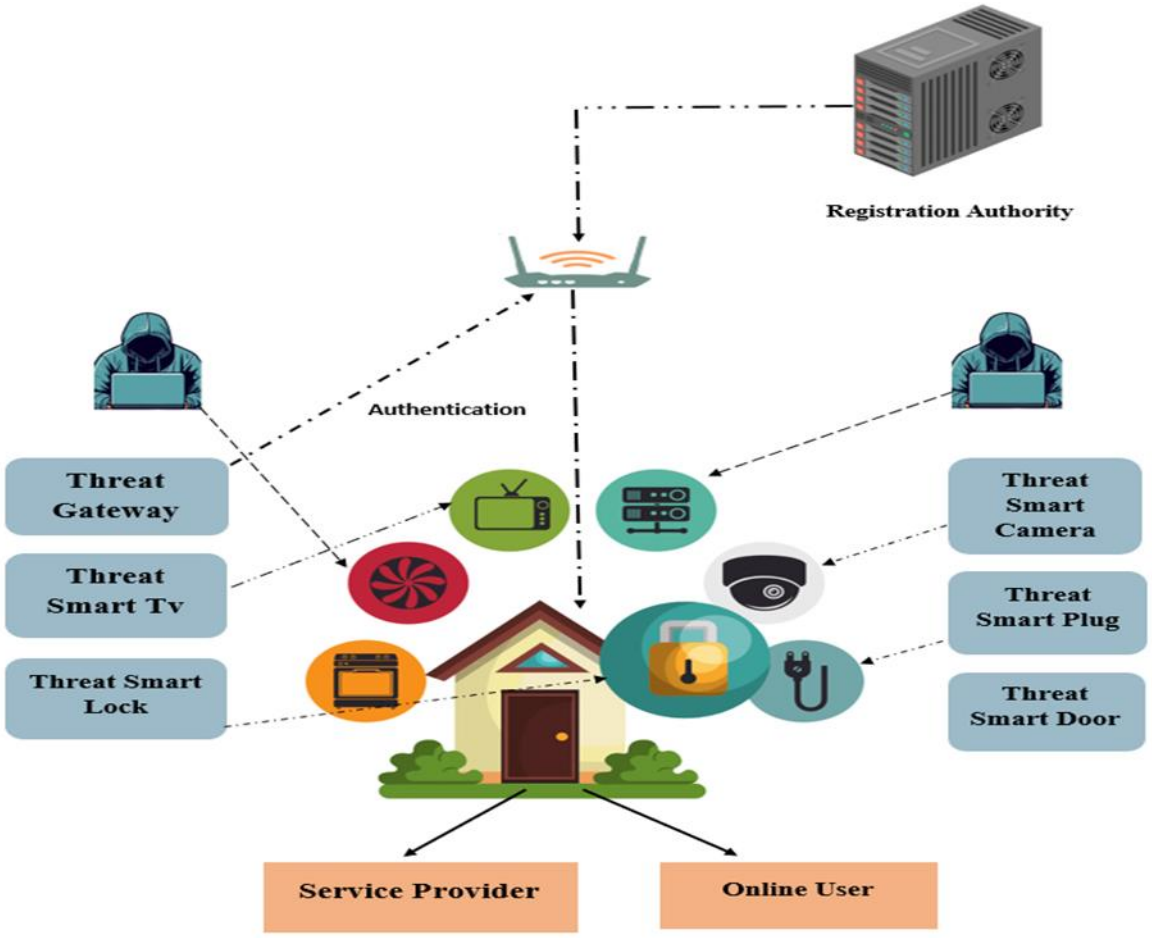


Figure 1 IoT Smart Home Architecture

The rest of the paper organize in section 2 discuss the literature review, section 3 the methodology of this paper, section 4 analysis the experiment and the results, and eventually, section 5 presents the conclusion.

2. LITERATURE REVIEW

The rise of IoT devices has significantly expanded the digital landscape, yet their limited computational power and often weak security mechanisms have made them prime targets for cyberattacks, particularly botnets. These botnets exploit device vulnerabilities to launch coordinated attacks, such as DDoS, posing serious threats to the stability of networks. The challenge in detecting such attacks is compounded by the heterogeneity of IoT devices, encrypted communications, and large-scale data transmission. Effective detection systems must also operate under resource constraints and deliver near real-time responses, highlighting the need for intelligent and adaptive solutions like ML and DL. Several studies have attempted to address these issues. In [10], the authors presented a technique using Fisher's score, AE, and DNN for

RESEARCH ARTICLE

both binary and multi-class classification. However, their model failed to analyse performance consistency across both classification types and overlooked critical evaluation metrics like FPR and prediction time factors essential for assessing model robustness in real-time environments.

In [11], a range of FS algorithms, including Fisher Score, Mutual Information (MI), Chi-square, PCA, and Recursive Feature Elimination (RFE) were applied to identify key attributes for botnet detection. While these methods help uncover relevant patterns, the study did not evaluate the classifiers in a multi-class setting nor did it examine parameters like FPR and sensitivity, which are necessary for measuring model stability and precision in real-world applications.

Addressing the issue of high-dimensional data in IoT environments, the authors of [12] proposed reducing the feature space for botnet detection using the N-BaIoT and Med-BIoT datasets. They implemented various filter and wrapper FS techniques across multiple ML algorithms, identifying optimal feature subsets for classification tasks. However, the study remained confined to classical ML approaches and lacked a comprehensive performance comparison, limiting insights into the generalizability of the proposed methods.

In [13], the authors introduced a FE-based framework using DMLP and CNN, enhanced by an unsupervised hierarchical graph clustering (HGC) method and a scatter search-based FS technique. While their system achieved 100% accuracy and zero FAR on binary classification tasks, it lacked integration with more advanced models and suffered from overfitting raising concerns about its scalability to multi-class problems and unseen datasets.

The study in [14] adopted a hybrid model that utilized RF for FS and employed LR, and KNN for classification on the UNSW_NB15 dataset. Although the approach showed promising results, it was limited to binary classification and failed to include domain-specific validation or evaluation metrics essential for real-world deployment.

In [15], a hybrid model combining PCA and RF with an ANN was designed to detect DDoS attacks utilising the BoT-IoT dataset. The study targeted multi-class classification and demonstrated high accuracy; however, it lacked assessment of important factors such as FPR and prediction latency, which are critical for practical cybersecurity applications.

Authors in [16] developed a hybrid ML-based botnet detection model using the CICIDS2017 dataset. They reduced features from 85 to 9 using Correlation Attribute Evaluation and PCA. The model was trained using six classifiers including RF, IBK, JRip, MLP, NB, and OneR. Though accuracy, precision, and recall were evaluated, the absence of metrics like prediction time and FPR weakened the model's practical applicability, especially since it only supported binary classification.

In [17], a novel approach was introduced using Information Gain and GA for FS, along with RF optimization, to analyse DNS query data from the TI-2016 DNS dataset. Despite employing optimization techniques, the model's accuracy dropped by 4% and remained focused only on binary classification without incorporating sufficient performance metrics.

The work in [18] presented two novel FS techniques based on dominant pattern recognition and frequent itemset mining, aided by Gini impurity and clustering to choose optimal features. Although the model achieved perfect TPR and 0% FPR, it suffered from overfitting, limiting its reliability on new data.

A comprehensive comparison of eight ML models including NB, DT, RF, SVM, LR, and deep models like CNN and MLP was undertaken in [19] using the N-BaIoT dataset. Feature reduction was done via PCA and Linear Discriminant Analysis (LDA). However, the models were only evaluated for binary classification, and essential parameters such as FPR and prediction time were not considered.

In [20], a DL-based model for botnet detection was introduced, which used quantile normalization and FS methods like Information Gain and City Block Distance, followed by oversampling via data augmentation. The proposed CNN-Fused Deep Stacked Autoencoder (CNN-FDSA) achieved high accuracy but was prone to high false positives, a critical flaw in real-world deployments where mislabelling normal activity as malicious can cause serious disruptions.

Lastly, the authors in [21] proposed three hybrid models integrating RF with classifiers like SVM, NB, KNN, and LR. While the traditional ML models achieved about 85% accuracy, they were restricted to binary classification and exhibited high error rates exceeding 20% pointing to room for significant improvement.

The survey of existing literature reveals that substantial progress has been made in the field of botnet detection through the application of both traditional ML and DL approaches. However, significant challenges remain particularly in handling high-dimensional IoT data, supporting multiclass classification, and optimizing feature selection through hybrid techniques. To bridge these gaps, this research builds on prior contributions and proposes several key innovations:

2.1. Hybrid Feature Selection Using PCA and RF

PCA is a common dimensionality reduction method that efficiently compresses data by preserving the most significant variance. However, it often overlooks the direct relevance of

RESEARCH ARTICLE

features to the prediction task. Conversely, RF assesses feature importance based on classification outcomes but becomes less effective in managing high-dimensional datasets [15]. This study proposes a hybrid method that integrates the strengths of both PCA and RF. Furthermore, this approach is integrated with deep learning models including AE, CNN, RNN, and LSTM to enhance detection performance in multiclass settings.

2.2. Enhanced Multiclass Botnet Classification

Many prior works, such as those in [15], primarily focus on binary classification or target specific botnet behaviors, limiting their practical use in complex environments. Our research addresses this limitation by implementing a multiclass classification strategy capable of identifying multiple types of botnet threats, including DDoS, DoS, reconnaissance, and information theft. This broader classification approach improves detection accuracy and reflects the diverse threat landscape in smart home networks.

2.3. Extensive Performance Evaluation

Conventional studies typically evaluate their models using basic metrics like accuracy, precision, and recall. However, these alone do not offer a complete picture of performance. This study broadens the evaluation criteria by incorporating additional indicators such as sensitivity, specificity, FPR, FNR, prediction latency, and detection rate. This holistic evaluation ensures a more accurate assessment of model effectiveness and operational readiness in real-time scenarios.

2.4. Targeted Application to Smart Home IoT Security

Unlike most existing research that focuses on generic IoT applications, our study is tailored specifically to smart home environments. These settings often involve devices with limited resources and heightened exposure to attacks due to constant connectivity and user interaction. By focusing on this niche, the research provides practical insights and solutions tailored to real-world smart home security needs using an integrated DL and hybrid feature selection framework.

3. METHODOLOGY

3.1. Overview of the Proposed Framework

The suggested methodology integrates PCA and RF as a hybrid feature engineering technique, subsequently employing four different DL models (AE, CNN, RNN, and LSTM) for the multiclass classification of botnet attacks. The main objective is to tackle the issues of dimensionality, feature redundancy, examine the performance of DL, and scalability in IoT-enabled smart home settings, while improving the robustness and precision of botnet identification. The proposed rely lies in Hybrid Feature Engineering by Integrating PCA's dimensionality reduction with RF's feature selection enables both efficient and effective representation of input data, Diverse DL Models by utilizing the advantages of them to capture diverse data patterns, including temporal, spatial, and anomalous behaviors, and Comprehensive Analysis by assessing the hybrid methodology and distinct DL models in comparison to raw data, PCA, and RF individually, emphasizing the advantages of the hybrid approach. The framework, illustrated in Figure 2, outlines a comprehensive process for detecting multiclass botnet attacks in IoT environments using advanced computational models.

3.2. Data Preparation

The Bot-IoT dataset [22] was selected as the main dataset used in the experiment for its relevance and suitability for IoT network traffic analysis. This dataset contains diverse traffic types that include benign IoT traffic and several botnet attacks, including DoS, DDoS, reconnaissance, and information theft. The sample size of this dataset, and the dataset structure in Figure 3 and Figure 4 respectively. The fact that it (the botnet) is comprehensive makes it a good candidate to be used when evaluating the proposed framework for ensuring the protection of IoT smart home devices against botnet. The dataset approximately represents the real-world IoT paradigms and challenges, providing a robust platform for testing the effectiveness of DL models in multiclass classification. Furthermore, the dataset's public availability improves validation and reproducibility, key components for scientific research. A network testbed was used to create the Bot-IoT dataset in a controlled environment. This required configuring servers, network infrastructure, and Internet of Things devices to replicate actual traffic patterns.

Several pretreatment processes were used to guarantee the dataset's quality and suitability for deep learning programs. To preserve the integrity of the dataset, the cleaning procedure included eliminating duplicates, incorrect entries, and missing values. Duplicate rows and logically inconsistent entries (e.g., negative packet counts) were found and fixed, and imputation techniques such as median replacement were applied to features with a large proportion of missing data. To translate categorical variables into numerical representations, feature encoding was used. To make sure that no artificial ordinal relationships were inferred, one-hot encoding was used for nominal categories like protocol types. Label encoding was used for ordinal characteristics in order to maintain their natural order.

The Synthetic Minority Oversampling Technique (SMOTE) was used to solve the imbalance in class distribution. SMOTE creates synthetic examples for minority classes by interpolation of existing examples. This procedure provided a balanced class distribution in the dataset, thus avoiding bias towards the dominant class in the training of DL models. Balancing the dataset was vital to enhancing classification performance, particularly in identifying uncommon attack types such as information theft and reconnaissance.

RESEARCH ARTICLE

This step is vital for DL models, which are sensitive to feature scaling, as it improves convergence during training and ensures fair weight distribution in optimization algorithms.

These preprocessing steps resulted in a clean, standardized, and feature-rich dataset ready for hybrid feature engineering using PCA and RF, followed by DL-based classification. By meticulously addressing data quality and preparation, this study ensures a robust foundation for exploring the efficacy of the proposed methodology in enhancing IoT security against botnet assaults.

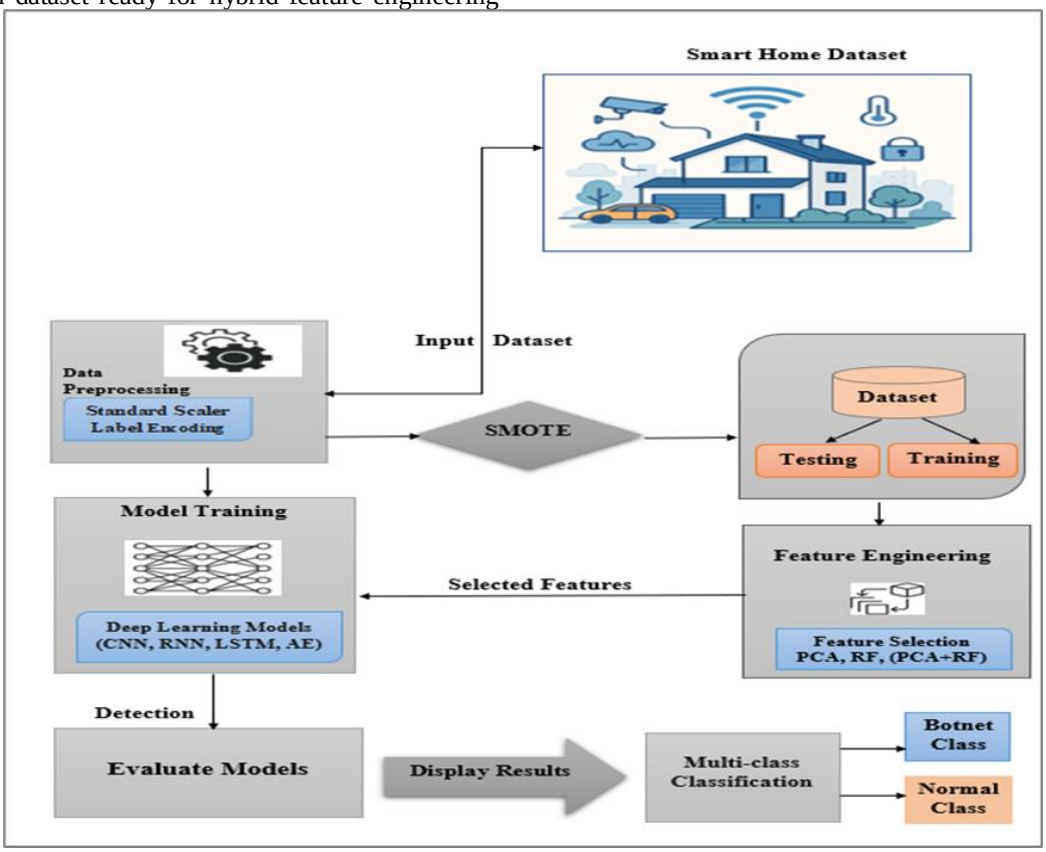


Figure 2 Model Framework

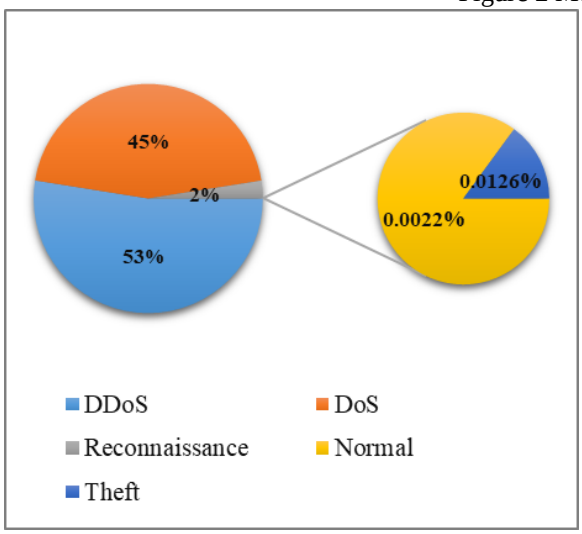


Figure 3 Dataset Sample Size

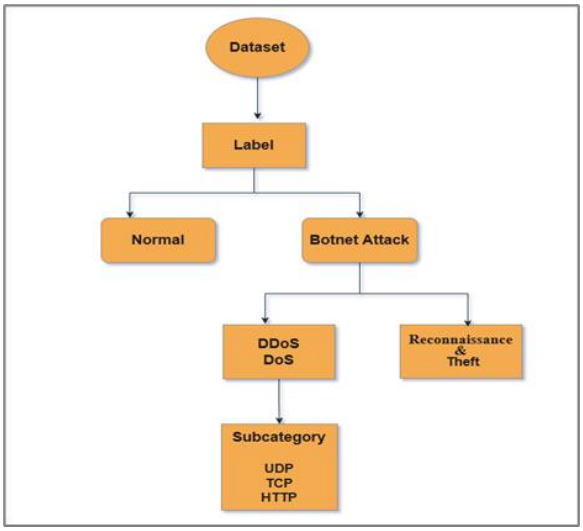


Figure 4 Dataset Sample Size

RESEARCH ARTICLE

3.3. Hybrid Feature Engineering (FE)

FE is a required step in ML workflows, especially when working with high-dimensional datasets like Bot-IoT. The proposed hybrid approach combines PCA and RF to address the limitations of using each technique independently. By leveraging the strengths of both methods, this approach enhances the robustness, relevance, and interpretability of the input data, ultimately improving classification performance in detecting multiclass botnet attacks.

3.3.1. Principal Component Analysis (PCA)

PCA is a dimensionality reduction technique that transforms the dataset into a set of orthogonal components, known as principal components, which maximize the variance within the data. This transformation reduces redundancy and noise while retaining the most significant information. PCA is particularly effective for large datasets, such as Bot-IoT, which contain many redundant and irrelevant features.

The selection of the number of principal components is guided by the explained variance ratio, ensuring that a majority of the dataset's information is preserved. By retaining components that account for a high percentage (e.g., 95%) of the variance, PCA achieves significant dimensionality reduction while minimizing information loss. This step also reduces computational burden, accelerates model training, and mitigates overfitting risks. PCA captures underlying patterns in the data, making it a powerful tool for enhancing the efficiency of subsequent model training. It accomplishes this by identifying orthogonal components that optimize variance in the data. The procedure entails calculating the covariance matrix and subsequently performing eigenvalue decomposition.

The covariance matrix C is computed as define in equation (1):

$$C = \frac{1}{n} \sum_{i=1}^{n} (x_i' - \mu)(x_i' - \mu^T) \quad (1)$$

Where,

x_i' is $i-$ th dataset sample,

μ is mean vactor,

n is the number of sample

To retain the majority of information, we select the top k components such that as define in equation (2):

$$\frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{m} \lambda_j} \geq T_{pca} \quad (2)$$

Where, λ_j is the eigenvalues of C, m is the total number of features, T_{pca} is the explained variance threshold.

The data is then transformed using the top k eigenvectors V_k as define in equation (3):

$$X_{pca} = X' . V_k \quad (3)$$

Where X' is the orginal dataset , and This dimensionality reduction phase guarantees the preservation of essential patterns in the dataset, facilitating efficient processing.

In Figure 5 demonstrates the Variance Ratio by PCA, this graph displays the variance proportion highlighted by the initial two principal components (PC1 and PC2) obtained by PCA. The X-axis illustrates the PCs, while the Y-axis shows the explained variance ratio, exhibiting the proportion of the dataset's total variability accounted for by each component. The graph shows PC1 includes around 25% of the entire variance, whereas PC2 represents about 10–15%. Collectively, these two elements account for around 35–40% of the variability in the dataset, giving them the most important axes for data summarization in reduced dimensions.

PCA plays a role in converting the original high-dimensional dataset into a reduced set of unrelated components that optimize variance. This figure illustrates PCA's capacity to diminish the dataset's dimensionality while keeping crucial details. PC1 and PC2 together reduce the data while preserving significant variability, facilitating interpretation, visualization, and analysis. This dimensionality reduction can be particularly beneficial for pattern recognition, noise reduction, and data preparation for ML applications, hence enhancing efficiency and interpretability in complicated datasets.

On the other hand, Figure 6 illustrates two scatter plots comparing overfitting behavior before and after PCA. Plot (a) shows overfitting for all features, with data points densely scattered between -1.5 and 1.5 along the mean values, indicating redundancy in the dataset. In contrast, Plot (b) shows overfitting after applying PCA, where dimensionality is minimized, and variance is better retained, showing improved feature representation. The comparison emphasizes the efficacy of PCA in tackling overfitting and enhancing model performance.

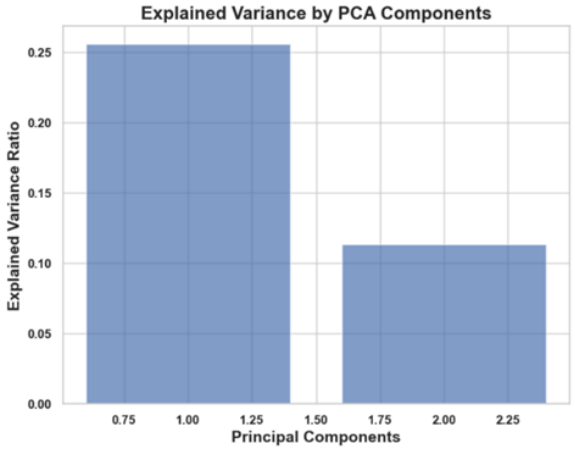
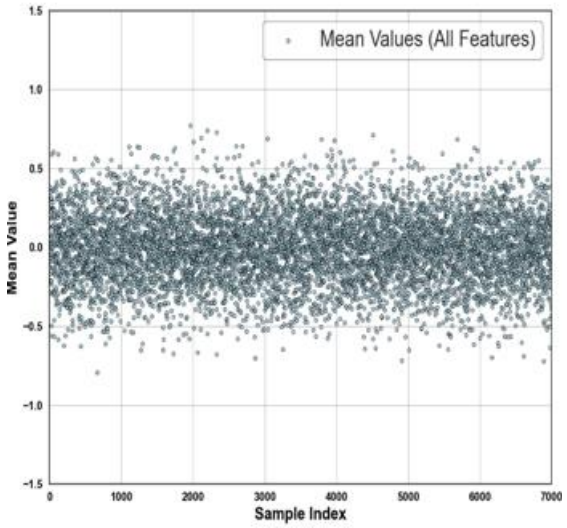
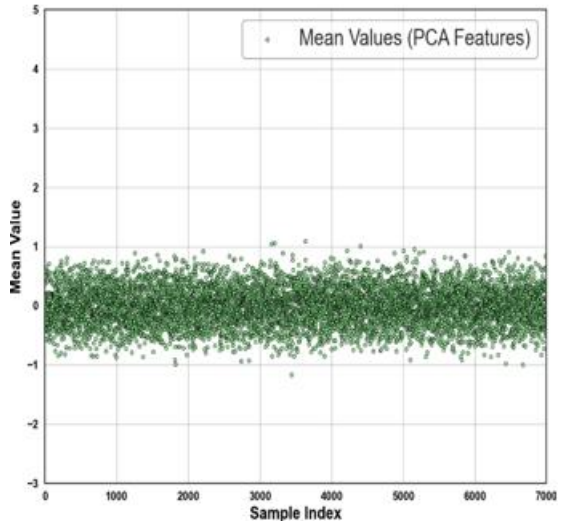


Figure 5 Variance Ratio by PCA

RESEARCH ARTICLE



(a) Overfitting Scatter Plot for all Features



(b) Overfitting Scatter Plot with PCA

Figure 6 Overfitting of all Features and with the Proposed

3.3.2. Random Forest (RF) for Feature Selection

While PCA emphasizes variance-based reduction, RF provides a complementary approach by evaluating the importance of individual features. RF, an EL method, builds multiple DTs and uses their combined outputs to determine the significance of each feature in predicting the target variable.

Features with high importance scores are selected using a predefined threshold, guaranteeing that only the most relevant features are kept. Unlike PCA, which transforms features into a new space, RF preserves the original feature representation, enabling direct interpretability. This is particularly useful for understanding the role of specific features, such as protocol type or traffic direction, in botnet detection. The importance score I_f for a feature f is calculated as define in equation (4):

$$I_f = \frac{1}{T} \sum_{t=1}^{T} Gaint_t(f) \quad (4)$$

Where T is the total number of trees in the forest, and $Gaint_t(f)$ is the contribution of feature f to the split quality in tree t. Features with significance scores over a certain threshold are preserved. RF offers a supplementary method to PCA by finding the most significant features according to their importance in classification. Table 1 discusses the seven major features with their significance scores, pkSeqID (47.95%), daddr (21.22%), and N_IN_Conn_P_DstIP (7.46%) are identified as the most significant factors for botnet detection. The graph in Figure 7 demonstrates the top seven features selected by RF, clearly displaying their impact on categorization accuracy. The RF algorithm systematically classifies features using their contributions to decision tree splits, eliminating redundancy while maintaining clarity. In contrast to PCA, RF preserves unique feature representations, facilitating domain-specific insights. The significance of pkSeqID underscores its vital role in detecting botnet activity.

Table 1 Top 7 Features Based on RF Importance

Rank	Feature	Importance
1	pkSeqID	0.479514
2	daddr	0.212239
3	N_IN_Conn_P_DstIP	0.074628
4	srate	0.062422
5	subcategory	0.036571
6	max	0.024498
7	saddr	0.017741

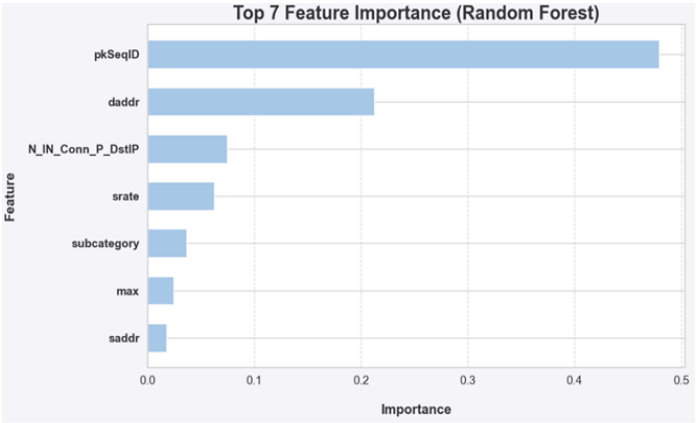


Figure 7 Feature Importance Selected by RF

RESEARCH ARTICLE

3.3.3. Hybridization of PCA and RF

The hybridization of PCA and RF addresses the limitations of using either technique in isolation. While PCA effectively reduces noise and redundancy, it may discard features with low variance that still hold predictive value. RF, on the other hand, focuses on feature relevance but can become computationally expensive when handling high-dimensional data. Combining these two methods balances their respective strengths.

The hybrid feature set is constructed by combining PCA-transformed components with features selected by RF. This guarantees that the ultimate input to the deep learning models is concise, informative, and exceptionally pertinent. The hybrid technique improves the model's ability to identify intricate patterns in multiclass botnet attacks by combining the variance-based reduction of PCA with the discriminative strength of RF.

The hybrid of PCA and RF is particularly suitable for the Bot-IoT dataset due to its high dimensionality and heterogeneous feature types. PCA reduces the dataset to its most significant components, focusing on underlying patterns, while RF ensures that discriminative features are not lost during this process. Together, these methods balance computational efficiency and feature relevance, creating a robust and scalable feature set for multiclass botnet detection. This approach highlights the study's novelty and establishes a foundation for enhancing IoT network security through DL.

There are several benefits to the hybrid feature building process like, Dimensionality Reduction PCA makes it possible to work with high-dimensional datasets like Bot-IoT by lowering the amount of work that required to be done and the amount of memory that is important, Relevance of Features (RF) makes sure that important features are kept, which works well with PCA's variance-based reduction. Accuracy by Using both variance-optimized and relevance-optimized features together creates a full set of features that enhances the accuracy of classification and the generalizability of the model. Noise Reduction which PCA gets rid of noise and duplicate data, and RF makes sure that features with specific prediction value are included, and Flexibility, where the hybrid method can be used with different IoT datasets, showing that it could be useful in real life. The hybrid feature set is defined as in equation (5):

$$X_{hybrid} = X_{pca} \cup F_{rf} \quad (5)$$

This hybrid methodology guarantees that the resultant feature set is both succinct (from PCA) and pertinent (from RF), hence augmenting the efficacy of the ensuing DL models.

PCA is used to decrease the dataset dimensionality. It begins by calculating the covariance matrix which captures the relationships between the features. These vectors are then used to derive eigenvalues and eigenvectors that represent the principal components of the data. The components are ordered according to their explained variance, reflecting the amount of information each component preserves from the original data. Top-k components having explained variance more than a given threshold (like 95 %) are chosen and the dataset is compressed to lower-dimensional space using these components. FS is a vital step in the data preprocessing stage, where we choose the most relevant features and eliminate any noise or redundancy in the data.

Then, FS is performed using RF. An RF model is trained utilizing PCA-reduced features, assessing the significance of each feature according to its contribution to the model's accuracy. Features with significance exceeding a specified threshold (e.g., 0.1) are chosen. The procedure proceeds iteratively, incorporating elements of greater significance until the selection criterion is no longer satisfied.

The features identified by both PCA and RF are amalgamated to form a hybrid feature set. The PCA-reduced features are incorporated initially, succeeded by the RF-selected features, guaranteeing the absence of duplication. This hybrid method integrates the dimensionality reduction capabilities of PCA with the feature selection efficacy of RF, resulting in a comprehensive set of features that encapsulate both the fundamental patterns in the data and the most significant variables for model prediction. This hybrid feature set is prepared for application in training machine learning models. The algorithm that displays all the aforementioned in Algorithm 1 pseudocode. This displays the hybrid method's ability to decrease redundancy and concentrate on the most relevant features, thus improving the model's generalization performance and lowering the potential of overfitting.

Input:

- Dataset: Original dataset
- $threshold_{PCA}$: Variance threshold for PCA (e.g., 95%)
- $threshold_{RF}$: Feature importance threshold for RF (e.g., 0.1)

Output:

- HybridFeatureSet: Final hybrid feature set

1. Apply PCA for Dimensionality Reduction

- Compute the covariance matrix Σ dataset
- Calculate eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and eigenvector $v_1, v_2, \ldots\ldots, v_n \; of \; \Sigma$
- Select the top k principal components:

If ExplainedVariance > $threshold_{PCA}$

Retain the top k components.

RESEARCH ARTICLE

End If

- transform the dataset into the PCA-reduced feature space:

PCAFeatures ← Dataset × $[v_1, v_2, \ldots\ldots, v_k]$

2. Apply Random Forest (RF) for Feature Selection

- Train an RF model on PCAFeatures.
- Calculate feature importance scores $Importance_i$ for each feature *i*

While $Importance_i > threshold_{RF}$

Add *i* to RFSelectedFeatures.

End While

3. Combine PCA-Reduced Features and RF-Selected Features

- Initialize HybridFeatureSet ← {}.

For each feature $f \in$ PCAFeatures:

Add f to HybridFeatureSet.

End For

For each feature $f \in$ RFSelectedFeatures:

If $f \notin$ HybridFeatureSet:

Add *f* to HybridFeatureSet.

End if

End For

4. Return Hybrid Feature Set

Output: HybridFeatureSet

Algorithm 1 Hybrid Feature Engineering (Hybrid PCA + RF)

3.4. Deep Learning Algorithms

To efficiently disclose and classify diverse forms of botnet attacks in IoT-based smart home environments, this work applies a hybrid feature set to four DL architectures: AE, CNN, RNN, and LSTM. By utilizing these models, we aim to capture both spatial and temporal aspects of network traffic, allowing for more precise identification of anomalies. Each model independently processes the optimized feature set, enabling a comprehensive performance comparison and highlighting the benefits of DL over conventional ML methods.

3.4.1. Autoencoder (AE)

AEs are unsupervised NN engineered to learn efficient representations of input data. They compress high-dimensional data into a lower-dimensional latent space and then reconstruct the original input. During this process, the model learns the normal behavior patterns of the data. When new input significantly deviates from learned patterns—such as botnet traffic not seen during training—the reconstruction error increases. This characteristic makes AEs particularly effective for identifying novel or stealthy attacks. Furthermore, AEs are advantageous in scenarios where labeled datasets are scarce, which is often the case in dynamic IoT ecosystems where new types of attacks emerge frequently.

3.4.2. Convolutional Neural Network (CNN)

Originally developed for image recognition tasks, CNNs have proven effective for analyzing structured data by extracting hierarchical spatial features through layers of convolutions and pooling. In the context of IoT security, CNNs can identify distinctive spatial patterns in network traffic, such as unusual packet flow behaviors or localized traffic anomalies indicative of botnet activity. Unlike traditional ML models that rely on handcrafted features, CNNs automatically learn relevant patterns, reducing the dependency on manual feature engineering and increasing detection robustness.

3.4.3. Recurrent Neural Network (RNN)

RNNs are explicitly engineered to process sequential data by preserving an internal state that evolves with each input. This capability enables them to model the temporal dependencies in network traffic, making them suitable for detecting patterns that unfold over time, such as those seen in coordinated botnet activities like reconnaissance scans or timed denial-of-service (DoS) attacks. Unlike conventional algorithms that treat each input independently, RNNs consider the order and timing of events, which is crucial for identifying sequential attack signatures.

3.4.4. Long Short-Term Memory (LSTM)

LSTMs build upon standard RNNs by addressing the vanishing gradient issue, allowing them to retain information over longer time intervals. Their architecture—featuring memory cells and gated mechanisms—makes them especially suited for recognizing extended patterns or multistage attacks, such as slow-paced or persistent botnet infiltrations. By preserving important contextual information across sequences, LSTMs enhance the ability to detect complex, temporally correlated threats that traditional ML models or even standard RNNs may fail to capture effectively. A detailed analysis of DL techniques, including LSTM and others, is provided in studies [23, 24], which emphasize their adaptability and performance in cybersecurity applications.

3.5. Performance Metrics

All four DL models are trained and evaluated utilizing the hybrid feature set obtained from PCA and RF. Their performance is assessed utilizing criteria like accuracy, precision, recall, F1-score, and detection rate. The work

RESEARCH ARTICLE

illustrates the adaptability and resilience of DL models in identifying multiclass botnet assaults through output comparison.

Numerous conventional requirements for performance parameters for DL classifiers are as follows from equation (6) to equation (14):

Accuracy: measures the ratio of correct forecasting over all forecasting with the equation (6):

$$\frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Precision: measures the ratio of the accuracy of positive forecasting as define in equation (7):

$$\frac{TP}{TP+FP} \quad (7)$$

Sensitivity: measures the ratio of positive instances correctly detected by the classifier as define in equation (8):

$$\frac{TP}{TP+FN} \quad (8)$$

$F1$ Score: measures the harmonic mean of precision and recall as define in equation (9):

$$\frac{recall \times precision}{recall + precision} \quad (9)$$

False Positive Rate (FPR): Measures the ratio of actual negatives incorrectly identified as define equation (10).

$$\frac{FP}{FP+TN} \quad (10)$$

False Negative Rate (FNR): Measures the ratio of actual positives incorrectly identified as negatives as define in equation (11):

$$\frac{FN}{FN+TP} \quad (11)$$

AUC-ROC: The area under the Receiver Operating Characteristic curve, summarizing the trade-off between sensitivity and specificity as define in equation (12):

$$TPR = \frac{TP}{TP+FN}, \; FPR = \frac{FP}{TN+FN} \quad (12)$$

Prediction Time (PT): The time taken by a model to render a prediction, typically measured in seconds or milliseconds.

Training Time: The time taken by a model to train on a dataset, usually measured in seconds, minutes, or milliseconds.

MCC (Matthews Correlation Coefficient): A measure of the quality of binary classification, considering true and false positives and negatives as define in equation (13):

$$\frac{TN * TP - FN * FP}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN}} \quad (13)$$

The Error Estimate a measure that quantifies the ratio of erroneous forecasting generated by the model. It is complementary to accuracy and is defined as in equation (14):

Error Estimate = 1- Accuracy (14)

This methodological framework highlights the superiority of DL compared to traditional methods in tackling IoT security concerns, especially in dynamic, high-dimensional, and heterogeneous settings such as smart homes. The integration of the hybrid feature set with cutting-edge DL architectures improves the reliability and efficiency of botnet detection, addressing major shortcomings in current research.

4. EXPERIMENTS AND RESULTS

This section provides an evaluation of the proposed hybrid detection framework, which combines PCA, RF for FS, the SMOTE for balancing class distribution, and various DL models for classifying multiclass botnet assaults utilizing the BoT-IoT dataset.

To evaluate the performance and robustness of the framework, different evaluation metrics were considered. These include standard classification measures such as accuracy, precision, recall, and F1-score, along with advanced metrics like AUC-ROC, sensitivity, specificity, prediction time, and error estimation. These metrics gives a comprehensive understanding of how effectively the model detects and classifies different botnet threats while also addressing challenges related to class imbalance and computational efficiency.

The experimental outcomes, summarized in Tables 2 through 5, demonstrate the notable impact of the PCA+RF hybrid FE approach. When classifiers were trained directly on raw, unprocessed data, they exhibited relatively low values in precision, AUC, and F1-score, suggesting limited capability in capturing the complex characteristics of the high-dimensional dataset. In contrast, applying the PCA+RF hybrid pre-processing substantially improved the performance of all DL models across the board.

Among the evaluated classifiers, the AE combined with PCA and RF consistently yielded the highest scores across all types of attacks, including DDoS, DoS, reconnaissance, and information theft. The RNN, also paired with the hybrid feature set, followed closely, particularly excelling in precision and AUC-ROC. These findings reinforce the significance of well-executed pre-processing—especially dimensionality reduction and feature relevance filtering—in enhancing classification accuracy and model efficiency in IoT security contexts.

The experimental results affirm the efficacy of the proposed framework in accurately disclosing a wide range of botnet threats. By integrating advanced DL models with hybrid feature engineering techniques, the system demonstrates

RESEARCH ARTICLE

improved generalizability, reduced error rates, and better real-time performance key factors for strengthening the resilience of smart home IoT networks

Figure 8 presents a comparative evaluation of the two top-performing DL models RNN and AE across four classes of botnet attacks: DDoS, DoS, reconnaissance, and information theft. Using the combined PCA and RF feature selection strategy, the models were assessed based on their sensitivity and specificity.

Table 2 PM Comparison of DL Models for DDoS Attack Detection

Attack	Metric	CNN + Raw	CNN & (PCA+RF)	RNN + Raw	RNN (PCA+RF)	LSTM + Raw	LSTM (PCA+RF)	AE + Raw	AE (PCA+RF)
DDoS	Precision	0.71	0.99	0.55	1.00	0.58	0.99	0.71	1.00
	AUC	0.75	0.99	0.77	1.00	0.67	1.00	0.78	1.00
	F1-Score	0.87	0.99	0.68	1.00	0.68	0.99	0.83	1.00

Table 3 PM Comparison of DL Models for DoS Attack Detection

Attack	Metric	CNN + Raw	CNN & (PCA+RF)	RNN + Raw	RNN (PCA+RF)	LSTM + Raw	LSTM (PCA+RF)	AE + Raw	AE (PCA+RF)
DoS	Precision	0.99	1.00	0.59	1.00	0.61	0.99	0.99	1.00
	AUC	0.85	0.99	0.80	1.00	0.66	0.99	0.79	1.00
	F1-Score	0.83	0.99	0.33	1.00	0.43	0.99	0.74	1.00

Table 4 PM Comparison of DL Models for Reconnaissance Attack Detection

Attack	Metric	CNN + Raw	CNN & (PCA+RF)	RNN + Raw	RNN (PCA+RF)	LSTM + Raw	LSTM (PCA+RF)	AE + Raw	AE (PCA+RF)
Reco.	Precision	0.00	0.99	0.06	1.00	0.00	1.00	0.00	1.00
	AUC	0.66	0.99	0.77	1.00	0.76	1.00	0.63	1.00
	F1-Score	0.00	0.99	0.00	1.00	0.00	1.00	0.00	1.00

Table 5 PM Comparison of DL Models for Theft Attack Detection

Attack	Metric	CNN +Raw	CNN & (PCA+RF)	RNN + Raw	RNN (PCA+RF)	LSTM + Raw	LSTM (PCA+RF)	AE + Raw	AE (PCA+RF)
Theft	Precision	0.00	0.99	0.00	1.00	0.00	0.99	0.00	1.00
	AUC	0.65	0.99	0.21	1.00	0.84	0.99	0.63	1.00
	F1-Score	0.00	0.99	0.00	1.00	0.00	0.99	0.00	0.99

The AE achieved perfect performance, with both sensitivity and specificity reaching 100%. In contrast, while the RNN also showed high performance, its sensitivity and specificity values were slightly lower, particularly in detecting DoS and Theft attacks where specificity fell below 0.99. The AE demonstrated superior capability in distinguishing between malicious and benign traffic, making it a more reliable model for accurate botnet detection in IoT-based smart homes. These results show that the AE model is highly effective for anomaly disclosure tasks where precision and recall are critical for maintaining security and minimizing false positives in real-world deployments.

In Figure 9, the performance of three distinct pre-processing strategies PCA-only, RF-only, and the integrated PCA+RF were analysed across key evaluation metrics: accuracy, MCC, and F1-meausre. The results clearly highlight that the hybrid approach (PCA+RF) yields the most consistent and superior performance across all deep learning models. While PCA and RF alone offered some improvements, their impact was limited compared to their combined usage.

RESEARCH ARTICLE

Further evidence is presented in Table 6, which outlines the performance metrics of the classifiers utilizing the PCA+RF method. Among them, AE demonstrated the fastest PT and recorded the lowest values for both the FPR and FNR, indicating a strong ability to differentiate both attacks and benign behaviour. RNN, on the other hand, achieved the lowest error estimate and similarly minimal FPR and FNR values.

Together, AE and RNN outperformed CNN and LSTM in both reliability and operational efficiency, making them particularly well-suited for real-time botnet detection in smart home environments. These results align with recent research highlighting the advantages of hybrid FE in optimizing DL model performance [23, 24].

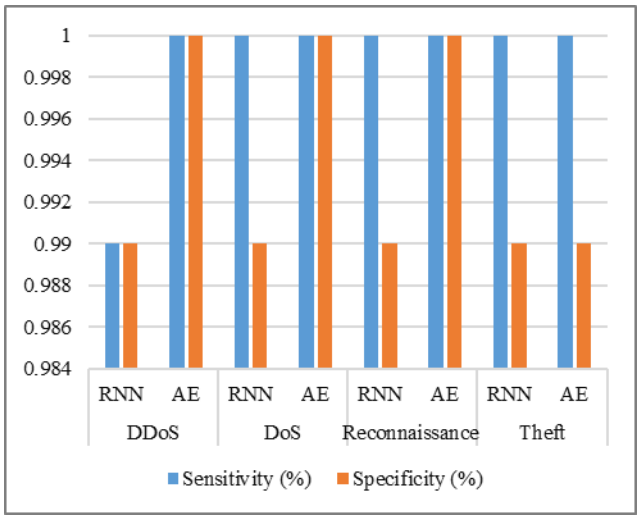


Figure 8 Performance of DL on Attacks Classes Across (PCA+RF) Approach

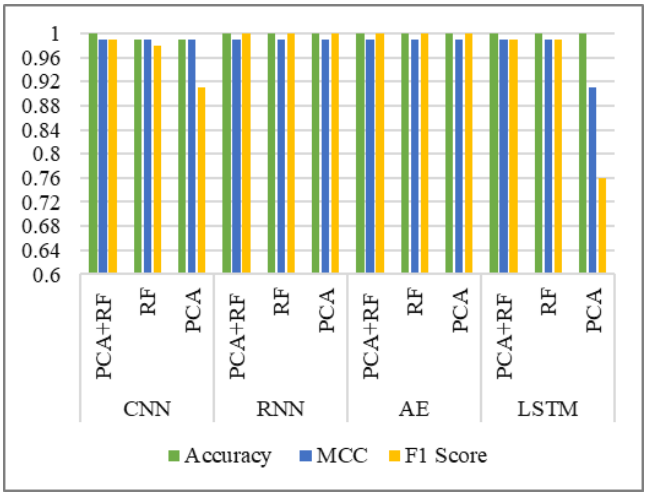


Figure 9 Performance Comparison of DL on Different Approaches

Figure 10 illustrates a comparative analysis of the prediction times (measured in seconds) for four DL models, CNN, RNN, LSTM, and AE, evaluated under four different pre-processing strategies: Raw Data, PCA, RF, and the hybrid PCA+RF. Among these models, RNN and LSTM consistently exhibit the longest prediction times, particularly when used in conjunction with the PCA+RF approach. This suggests that while these models are effective in terms of detection, they require more computational resources when processing high-dimensional, feature-engineered data.

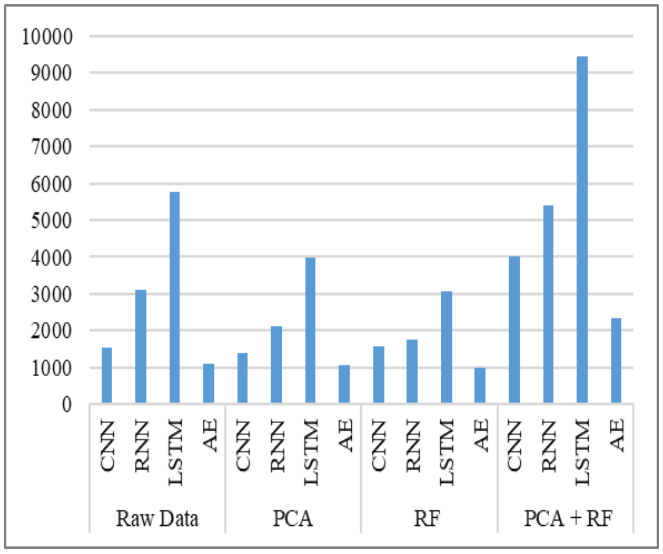


Figure 10 Prediction Time Comparison for Different Classifiers Across Different Approaches

Conversely, the AE model stands out for its efficiency, demonstrating the shortest prediction times across all pre-processing configurations. This indicates its suitability for scenarios requiring rapid detection and low-latency response. Both PCA and RF, when used individually, reduce prediction time in comparison to the combined PCA+RF approach.

However, models trained on raw, unprocessed data generally show moderate prediction times but perform less efficiently in terms of detection accuracy.

While the PCA+RF combination delivers the most accurate and reliable results, it also incurs the highest computational cost, especially for more complex models like RNN and LSTM. In contrast, using only RF offers a balanced trade-off, providing reduced computation time with respectable performance.

PCA, by itself, achieves notable time efficiency but may sacrifice some predictive power. Overall, the findings underscore that although pre-processing improves model performance, it also impacts computational demand a critical consideration for real-time smart home applications.

RESEARCH ARTICLE

Table 6 PM of Classifiers with PCA+RF Integration Approach

Classifier	FPR	FNR	Error Estimate	Prediction Time (s)
CNN	1.51	6.07	2.42	4015.11
RNN	0.003	0.0015	0.016	5387.11
LSTM	1.88	7.52	3.010	9461.45
AE	3.01	0.001	4.816	2354.88

Table 7 presents a comparative analysis of prediction times across various classifiers, considering both the complete feature set and a reduced set comprising the top seven features. Interestingly, for models such CNN and LSTM networks, a slight increase in prediction time (PT) was observed when using the reduced feature set. This suggests that, despite the lower number of features, additional computational steps may have been introduced during processing.

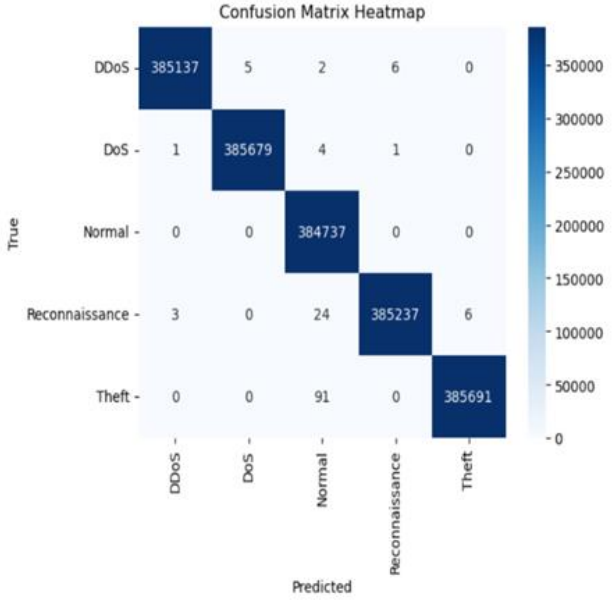
In contrast, RNN and AE showed noticeable reductions in PT with the selected top features, pointing toward a significant gain in efficiency due to the feature selection process. These results highlight that the effect of feature selection on prediction time is not uniform across all classifiers. Therefore, it is crucial to account for the unique characteristics and behaviors of each model when attempting to optimize for speed and efficiency through feature reduction strategies.

Table 7 Comparison of Prediction Time Between Top Selected Feature and all Features

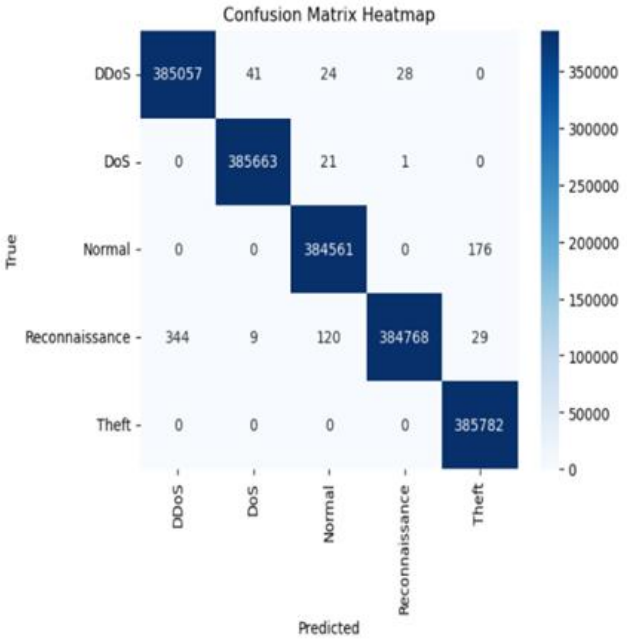
Classifiers	Prediction Time for All Features	Prediction Time for Top 7 Features
CNN	3810.2	4015.11
RNN	6601.12	5387.11
AE	4254.45	2354.88
LSTM	9426.98	9461.45

Figure 11 illustrates the confusion matrices for four machine learning models—CNN, RNN, LSTM, and AE—applied to a classification task involving five classes: DDoS, DoS, Normal, Reconnaissance, and Theft. These matrices visualize the relationship between actual and predicted class labels, where the diagonal entries represent correct predictions.

The concentration of high values along the diagonals suggests that each model achieved notable accuracy for certain classes. However, differences in model performance are evident through the varying distribution of off-diagonal values, which indicate instances of misclassification. These discrepancies reveal how each model handles different types of network traffic and emphasize the importance of selecting appropriate algorithms based on the specific classification challenges involved.

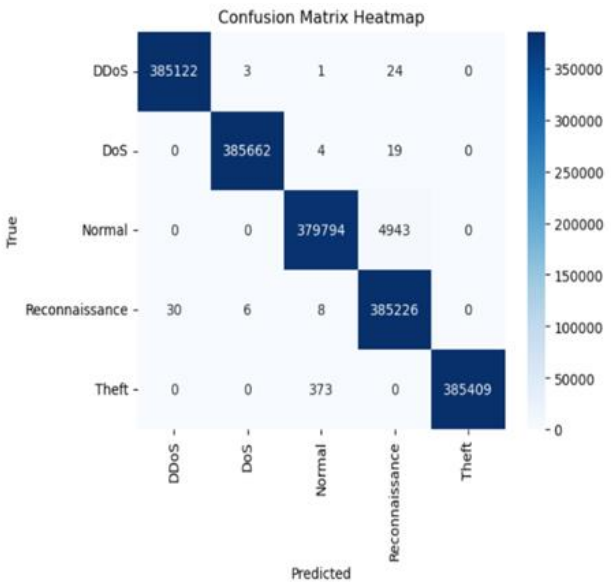


(a) CNN

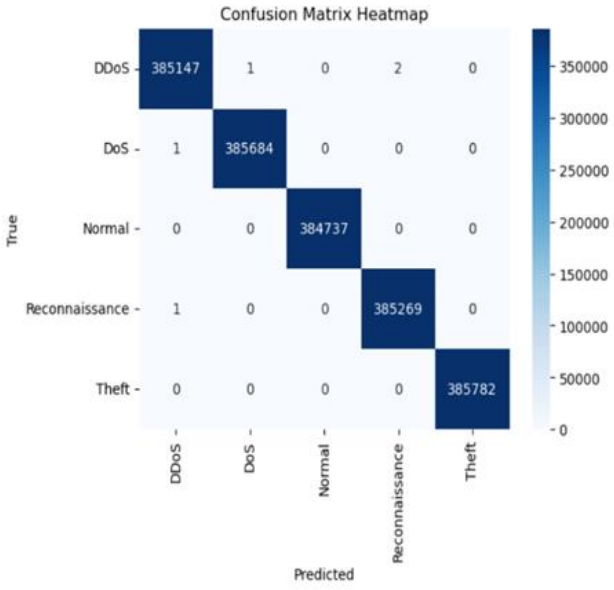


(b) RNN

RESEARCH ARTICLE



(c) LSTM



(d) AE

Figure 11 Confusion Matrices for Four Different Machine Learning Models

Table 8 Comparative Study Between the Proposed and Previous Work

Refence	Model	Dataset	Type of Classification	Features	Accuracy	Estimate Error
[25]	RF KNN	UNSW-NB15	Multi class	correlation feature selection	82.87	✕
[26]	RF SVM	UNSW-NB15	Binary class	correlation feature selection	97.50	✕
[27]	RF CNNLSTM	NSL-KDD, and UNSW-NB15	Binary class	29 Dynamic Feature Selector	92.76	✕
[28]	MPL	N-BaIoT	Multi class	PCA+XGB	0.99.93	✕
[29]	EL, RF, DT	UNSW-NB15	Binary Class	Mutual Information	0.92	✕
Proposed	AE, RNN, LSTM, CNN	BoT-IoT	Multiclass	Hybrid (PCA, RF)	0.99.99	1.8155

Table 8 presents a comparative evaluation of the proposed method against existing research efforts, highlighting the effectiveness and performance of various ML models in detecting botnet activity across different datasets. The suggested approach, which integrates AE, RNN, LSTM, and CNN, achieves an impressive accuracy of 99% on the BoT-IoT dataset. This performance surpasses that of earlier studies, such as those by Authors [13] and [14], which reported lower accuracy using models like RF and SVM. Notably, while many previous works focused primarily on binary classification tasks, the proposed system addresses multi-class classification, making it more adaptable to identifying a broader range of botnet attacks. An error rate of 1.8155 further demonstrates the model's robustness and practical applicability. The integration of hybrid FS techniques, including PCA and RF-based ranking, contributes to

RESEARCH ARTICLE

improved model performance by reducing feature complexity and enhancing relevant data representation. Overall, this comparative analysis underscores the superior accuracy and operational effectiveness of the proposed approach.

5. CONCLUSION

This research makes a meaningful contribution to enhancing smart home security by employing advanced deep learning techniques to detect and classify multiclass botnet attacks. By integrating PCA and RF for feature optimization, the proposed approach significantly improves the performance of RNN and AE. These models achieve outstanding results, with AUCS scores reaching 1.00 and F1-scores surpassing 0.98 across all attack categories, ensuring high-precision detection of malicious activity targeting diverse IoT devices. The models demonstrate strong capabilities in accurately distinguishing among various threat types including DDoS, DoS, Reconnaissance, and Theft. Thereby reducing false positives and enabling swift incident response. LSTM models also contribute positively to overall detection accuracy, particularly across multiple attack classes, while CNN models, though slightly less effective, offer valuable insight into specific attack patterns, especially for theft detection and identifying normal traffic. This work lays a strong foundation for integrating deep learning into Intrusion Detection and Prevention Systems (IDPS), promoting increased resilience and security in IoT-enabled smart home environments. Moving forward, future research will focus on optimizing model performance to reduce computational overhead, making these solutions more suitable for deployment in resource-constrained IoT environments. Additionally, broader validation using varied datasets and real-world IoT configurations will be essential to guarantee the generalizability and scalability of the proposed framework.

REFERENCES

[1] P. K. Sadhu, V. P. Yanambaka, and A. Abdelgawad, “Internet of Things: Security and Solutions Survey,” Sensors, vol. 22, no. 19, p. 7433, Sep. 2022, doi: 10.3390/s22197433.

[2] S. A. H. Ali and J. V. Rani, “Attack detection in IoT Using Machine Learning—A survey,” in Springer eBooks, 2023, pp. 211–228. doi: 10.1007/978-3-031-18497-0_16.

[3] M. M. Salim, S. Rathore, and J. H. Park, “Distributed denial of service attacks and its defenses in IoT: a survey,” The Journal of Supercomputing, vol. 76, no. 7, pp. 5320–5363, Jul. 2019, doi: 10.1007/s11227-019-02945-z.

[4] N. Abughazaleh, R. Bin, M. Btish, and H. M, “DOS attacks in IoT systems and proposed solutions,” International Journal of Computer Applications, vol. 176, no. 33, pp. 16–19, Jun. 2020, doi: 10.5120/ijca2020920397.

[5] P. Kumari and A. K. Jain, “A comprehensive study of DDoS attacks over IoT network and their countermeasures,” Computers & Security, vol. 127, p. 103096, Jan. 2023, doi: 10.1016/j.cose.2023.103096.

[6] A. A. Hezam, S. A. Mostafa, Z. Baharum, A. Alanda, and M. Z. Salikon, “Combining deep learning models for enhancing the detection of botnet attacks in multiple sensors internet of things networks,” JOIV International Journal on Informatics Visualization, vol. 5, no. 4, p. 380, Dec. 2021, doi: 10.30630/joiv.5.4.733.

[7] N. Butt et al., “Intelligent Deep Learning for Anomaly-Based Intrusion Detection in IoT smart home networks,” Mathematics, vol. 10, no. 23, p. 4598, Dec. 2022, doi: 10.3390/math10234598.

[8] N. U. Huda, I. Ahmed, M. Adnan, M. Ali, and F. Naeem, “Experts and intelligent systems for smart homes’ Transformation to Sustainable Smart Cities: A comprehensive review,” Expert Systems With Applications, vol. 238, p. 122380, Nov. 2023, doi: 10.1016/j.eswa.2023.122380.

[9] H. A. S. Ali and J. Vakula Rani, "Internet of Things and Digital Forensics: Recent Studies and Challenges," 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 959-966, doi: 10.1109/ICSCSS57650.2023.10169561.

[10] Abbasi, F., Naderan, M., & Alavi, S. E. (2024). Dimensionality reduction with deep learning classification for botnet detection in the Internet of Things. Expert Systems With Applications, 126149. https://doi.org/10.1016/j.eswa.2024.126149.

[11] Rathod, G., Sabnis, V., & Jain, J. K. (2024). Improving IoT Botnet Attack Detection using Machine Learning: Comparative Analysis of Feature Selection Methods and Classifiers in Intrusion Detection Systems. Ieee, 1–8. https://doi.org/10.1109/inocon60754.2024.10511883.

[12] Kalakoti, R., Nomm, S., & Bahsi, H. (2022). In-Depth feature selection for the statistical Machine Learning-Based botnet detection in IoT networks. IEEE Access, 10, 94518–94535. https://doi.org/10.1109/access.2022.3204001.

[13] Panda, M., Mousa, A. a. A., & Hassanien, A. E. (2021). Developing an efficient feature engineering and machine learning model for detecting IoT-Botnet cyber attacks. IEEE Access, 9, 91038–91052. https://doi.org/10.1109/access.2021.3092054.

[14] P. Saxena and R. B. Patel, "Efficient Hybrid Model for Botnet Detection Using Machine Learning," 2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM), Dubai, United Arab Emirates, 2023, pp. 1-7, doi: 10.1109/ICCAKM58659.2023.10449643.

[15] Jalo, N. H., & Heydarian, N. M. (2024b). A hybrid technique based on RF-PCA and ANN for detecting DDOS attacks IoT. Iraqi Journal of Data Science., 27–41. https://doi.org/10.51173/ijds.v1i1.9.

[16] Jabbar, A. F., & Mohammed, I. J. (2020). Development of an Optimized Botnet Detection Framework based on Filters of Features and Machine Learning Classifiers using CICIDS2017 Dataset. IOP Conference Series Materials Science and Engineering, 928(3), 032027. https://doi.org/10.1088/1757-899x/928/3/032027.

[17] Moubayed, A., Injadat, M., & Shami, A. (2020b). Optimized Random Forest Model for Botnet Detection Based on DNS Queries. Ieee, 1–4. https://doi.org/10.1109/icm50269.2020.9331819.

[18] Lefoane, M., Ghafir, I., Kabir, S., & Awan, I. (2022). Unsupervised Learning for feature Selection: A proposed solution for botnet detection in 5G networks. IEEE Transactions on Industrial Informatics, 19(1), 921–929. https://doi.org/10.1109/tii.2022.3192044.

[19] Saif, S., Yasmin, N., & Biswas, S. (2023). Feature engineering based performance analysis of ML and DL algorithms for Botnet attack detection in IoMT. International Journal of Systems Assurance Engineering and Management, 14(S1), 512–522. https://doi.org/10.1007/s13198-023-01883-7.

[20] Kalidindi, A., & Arrama, M. B. (2024). Feature selection and hybrid CNNF deep stacked autoencoder for botnet attack detection in IoT. Computers & Electrical Engineering, 122, 109984. https://doi.org/10.1016/j.compeleceng.2024.109984.

[21] Pandey, A., Thaseen, S., Kumar, C. A., & Li, G. (2020). Identification of botnet attacks using hybrid machine learning models. In Advances in intelligent systems and computing (pp. 249–257). https://doi.org/10.1007/978-3-030-49336-3_25.

[22] Koroniotis, N., Moustafa, N., Sitnikova, E., & Slay, J. (2018). Towards developing network forensic mechanism for botnet activities in the IoT based on machine learning techniques. In Mobile Networks and

RESEARCH ARTICLE

Management (pp. 30–44). https://doi.org/10.1007/978-3-319-90775-8_3.

[23] Haifa Ali Saeed Ali, Vakula Rani J, "Machine Learning for Internet of Things (IoT) Security: A Comprehensive Survey", International Journal of Computer Networks and Applications (IJCNA), 11(5), PP: 617-659, 2024, DOI: 10.22247/ijcna/2024/40.

[24] Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I., & Guizani, M. (2020). A survey of machine and deep learning methods for internet of things (IoT) security. IEEE Communications Surveys & Tutorials, 22(3), 1646–1685. https://doi.org/10.1109/comst.2020.2988293.

[25] Alshaeaa, H. Y., & Ghadhban, Z. M. (2024). Developing a hybrid feature selection method to detect botnet attacks in IoT devices. Kuwait Journal of Science, 51(3), 100222. https://doi.org/10.1016/j.kjs.2024.100222.

[26] Aldhyani, T. H. H., & Alkahtani, H. (2022). Artificial Intelligence Algorithm-Based Economic Denial of Sustainability Attack Detection Systems: Cloud computing environments. Sensors, 22(13), 4685. https://doi.org/10.3390/s22134685.

[27] Ahsan, M., Gomes, R., Chowdhury, M. M., & Nygard, K. E. (2021). Enhancing machine learning prediction in cybersecurity using dynamic feature selector. Journal of Cybersecurity and Privacy, 1(1), 199–218. https://doi.org/10.3390/jcp1010011.

[28] Danquah, L. K. G., Appiah, S. Y., Mantey, V. A., Danlard, I., & Akowuah, E. K. (2024). Computationally Efficient Deep Federated Learning with Optimized Feature Selection for IOT Botnet Detection. Intelligent Systems With Applications, 200462. https://doi.org/10.1016/j.iswa.2024.200462.

[29] Alghamdi, A., & Barsoum, A. (2024). A Comprehensive IDs to Detect Botnet Attacks Using Machine Learning Techniques. Ieee, 1–6. https://doi.org/10.1109/icmi60790.2024.10585846.

Authors

Haifa Ali Saeed Ali received the Engineering of Batcheler of Information Technology degree from the Engineering College, University of Aden, Yemen, in 2007, and MCA degree in Computer Application from Visvesvaraya Technological University (VTU), Belagavi, Karnataka, India, in 2017. She is currently a Ph.D. Research Scholar with the Department of Computer Application, CMR. Institute of Technology (affiliated to VTU), Bengaluru, India. She also working in Cyber Security, and IoT. Her area of interests includes Machine Learning, Deep Learning, Internet of Things (IoT), Cyber Security. She has around 6 publications in her credit which includes the journals, and conferences of international reputation.

Dr. Vakula Rani J is a Professor and President of the Institution Innovation Council at CMR Institute of Technology, Bangalore. With over 27 years of teaching and research experience, she specializes in Artificial Intelligence, Computer Vision, Machine Learning, Deep Learning, and Cloud Computing. Dr. Vakula has authored more than 25 research papers published in reputed international journals and conferences. She holds two Australian patents, two Indian design patents, and has filed & published eight Indian patents. A reviewer for renowned Q1 and Q2 journals, she is also certified as an Innovation Ambassador by the Ministry of Education, India. She severed as a member of the Board of Studies and Board of Examiner at various Universities. She is a recognized guide under VTU and guiding four Ph.D., research scholars. She is a Life member in leading National Professional Societies ISC and ISTE.

How to cite this article:

Haifa Ali Saeed Ali, Vakula Rani J, "Deep Learning Approach for Protecting IoT Smart Home Devices Against Multiclass Botnet Attacks", International Journal of Computer Networks and Applications (IJCNA), 12(3), PP: 307-323, 2025, DOI: 10.22247/ijcna/2025/20.