Published Online First: December, 2024
https://doi.org/10.21123/bsj.2024.11067
P-ISSN: 2078-8665 - E-ISSN: 2411-7986

Baghdad Science Journal

Towards An Efficient Internet of Things Intrusion Detection by Using Support Vector Machine

***Rawan Abo Zidan**[*1], George Karraz[2]*

[1]PHD Program, Syrian Virtual University, Damascus, Syria.
[2]Department of Artificial Intelligence and Natural Languages Processing, Faculty of Information Technology Engineering, Damascus University, Damascus, Syria.
*Corresponding Author.

Received 29/02/2024, Revised 06/07/2024, Accepted 08/07/2024, Published Online First 20/12/2024

Abstract

Intrusion Detection Systems (IDS) are crucial for safeguarding Internet of Things (IoT) networks against security threats. The integration of Support Vector Machine (SVM) with smart IDs has been a significant advancement in detecting anomalous activities. This research contributes to this field by implementing the Gaussian Pyramid (GP) algorithm, significantly reducing the processing amount and storage space required for large IoT network traffic datasets. This approach enables the GP model to classify thousands of data points in large-scale problems with high-dimensional input space. Notably, the GP model, with various kernel sizes, outperforms traditional nonlinear SVM and Artificial Neural Networks (ANN) in terms of efficiency and accuracy. For instance, with kernel sizes of 5, 7, and 9, the GP model demonstrated superior performance on the NSL-KDD dataset, achieving accuracy and AUC (Area Under the Curve) values higher than both nonlinear SVM and ANN. In kernel size 9, the GP model achieved the highest overall accuracy of 0.96% on the CIC-DDoS2019 dataset. The experimental results confirm that applying the GP model to IoT data traffic significantly reduces time complexity and enhances the performance of binary and multi class SVM, marking a substantial advancement in IoT intrusion detection.

Keywords: Gaussian Pyramid, GP Model, IDS, IoT, SVM.

Introduction

IoT is a networking model that contains devices connected to each other, forming their own network that can be accessed remotely, giving us an enormous amount of valuable data. IoT has wide applications including health, activity recognition, sports, tracking, localization, safety, education, and many other applications[1].

Every IoT device is fortified with integral sensors and wireless communication abilities, the sensors can collect data about the atmosphere and communicate with one another, along with the owner of the house and a principal monitoring system[2]. These imperatives which also handle challenging issues like similarity, efficacy, full utility, and accessibility, limit and impede the development of IoT frameworks[3].

With the development of network technologies, network information security is becoming more and more important. Attacks on IoT, if they remain undetected for an extended period, cause severe service interruption resulting in financial loss[4]. It is vitally important to find a fast and effective intrusion detection method, such a method helps the IoT network from intrusions and attacks, and this is what motivates researchers to use machine learning techniques.

Machine learning (ML) is a new and powerful solution for accurate and high-performance models in computer-based solutions. There are many good classification techniques in the literature including artificial neural networks, k-nearest neighbors' classifier, decision trees, Bayesian classifier, and SVM algorithm. From these techniques, SVM is one of the best-known techniques to optimize the expected solution[5].

SVM has proved to be a robust, efficient classification method and has been used in many IoT applications. SVM works best when the dataset is small, but it has $O(m^3)$ time and $O(m^2)$ space complexities, where m is the training sample. SVM training requires solving (large-scale) convex programming problems[6]. As the dataset size increases, the training time and memory required for SVMs tend to grow, especially in network data traffic.

SVM has been used as part of IDS for a long but encounters the problem of long training and testing times, high error rates, and low true positive rates, which limit the use of SVM in network intrusion detection[7]. Thus, there is still a need for new ways to improve SVM in large network traffic data to detect anomalous attacks.

This paper aims to provide an efficient approach to IoT intrusion detection by using GP to solve nonlinear SVM problems in terms of time and space. The combination of the GP technique and SVM enables faster training model building of the dataset, which needs less memory spacing to create a reliable IDS and remove noise in the dataset. Due to the exposure of malicious activity, provide the best IoT intrusion detection using SVM to ensure the infrastructure’s security. Reducing the dimension of large-scale problems (like in IoT traffic data) is difficult and computationally expensive[8], to solve this problem the model uses the GP algorithm. Reduce training and testing time by reducing the size of the dataset to ¼ of the original dataset size. Classification IoT traffic data using nonlinear SVM needs high space, it has $O(m^2)$ space complexities, using the GP algorithm smooths the dataset and reduces the degree of nonlinearity. The performance of SVMs sensitive to noise[9] , using GP removes the noise in the dataset. Below is the key contribution of the study.

- Developing a machine learning-based model capable of accurately detecting and classifying ransomware in IoT environments. This should address the limitations of existing detection systems.
- Implementing a feature extraction mechanism that effectively distinguishes between benign and malicious IoT traffic, enhancing the model's predictive accuracy.

Evaluating the performance of the model using a comprehensive dataset, such as the UNSW BoT IoT dataset, to ensure scalability and effectiveness across diverse IoT scenarios. The key novel contributions of the research are threefold: Firstly, the model introduces a novel application of the Gaussian Pyramid (GP) algorithm for processing IoT network traffic data, which significantly reduces the amount of data and space required for large datasets. Secondly, the methodology utilizes a variety of smoothing kernels proposed by the GP model to resize or sub-sample the dataset, leading to faster model training and testing phases. Lastly, the paper provides an extensive review of recent advancements in SVM for Intrusion Detection. The experimental results demonstrate that the GP-based approach, when compared with traditional SVM algorithms, significantly reduces the time complexity and enhances the performance of binary SVM in terms of accuracy and efficiency. These contributions mark a substantial advancement in the field of IoT intrusion detection, offering promising implications for the security and efficiency of IoT networks.

The rest of the article is organized as such; Section 2 presents related work of applying SVM on IoT intrusion detection, Section 3 presents motivation and objective, section 4 shows the importance of security in IoT, section 5 exhibits the proposed methodology, and Section 6 presents the experimental results and analysis. Finally, Section 7 concludes the research.

Related Work

In the burgeoning field of cybersecurity, particularly in the context of the Internet of Things (IoT), the exploration and understanding of related works is crucial. The landscape of IoT security is rapidly evolving, with new threats and challenges emerging constantly. This section delves into the existing body of research, focusing on various methodologies and technologies that have been developed to safeguard IoT environments against a wide range of cyber threats, including ransomware, malware, and other forms of cyberattacks. The paper critically analyzes the strengths and limitations of current intrusion detection systems (IDS), machine learning algorithms, and anomaly detection techniques that have been applied in IoT contexts. Several studies are performed on IoT intrusion detection as follows.

Jiang et al.[10] proposed a model based on a hybrid sampling approach. To handle the data imbalance problem, researchers proposed a network intrusion detection algorithm that combines two sampling techniques: one-sided selection (OSS) to reduce noise in the majority (normal) category. Synthetic Minority Oversampling Technique (SMOTE) to augment minority samples (intrusive samples). This results in a balanced dataset that enables the model to better learn features from a small number of samples and reduce training time. The algorithm uses a deep neural network architecture, consisting of a convolutional neural network (CNN) for extracting spatial features and a bidirectional long short-term memory (BiLSTM) for extracting temporal features.

Su et al.[11] proposed a BAT model that used bidirectional long short-term memory (BLSTM) to solve the problems of low accuracy and feature engineering in intrusion detection and combined it with attention mechanisms. An attention mechanism is used to examine network flow vectors, which consist of packet vectors generated by the BLSTM model so that the model can obtain key features for network traffic classification. Use multiple convolutional layers to capture local features of traffic data. This model is called BAT-MC. Softmax classifier is used to classify network traffic. Advantages of the proposed method: The end-to-end BAT-MC model does not require any feature engineering knowledge as it automatically learns the most important features from the hierarchy. It can well describe network traffic behavior and effectively improve anomaly detection capabilities.

Fu et al.[12] proposed a DLNID Model to solve the problem of low detection accuracy, this paper proposed a deep learning model for network intrusion detection (DLNID). DLNID combines an attention mechanism and a bidirectional long-term memory (Bi-LSTM) network. It first extracts the sequence features of traffic through a convolutional neural network (CNN), then redistributes the weight of each channel through an attention mechanism, and finally uses Bi-LSTM to learn the sequence features of the network.

Wisanwanichthan et al.[13] developed Dual Layer Hybrid Approach (DLHA) to address the limitations of a single classifier, this paper proposes a dual-layer hybrid approach (DLHA) for network intrusion detection. DLHA employs a Naive Bayes classifier as the first layer to detect DoS and detect attacks. Then, an SVM classifier is used as the second layer to distinguish R2L and U2R attacks from normal instances. The researchers compared the performance of DLHA with other published IDS techniques using the NSL-KDD dataset.

Gao et al.[14] took the NSL-KDD dataset as a research object and proposed an adaptive ensemble learning model for intrusion detection. The researchers constructed a Multitree algorithm by adjusting the proportion of training data and setting up multiple decision trees. To improve the overall detection effect, the researchers choose several base classifiers, including decision tree, random forest, kNN, and DNN, and design an ensemble adaptive voting algorithm.

Alamri et al.[15] proposed using an adaptive bandwidth profile-based threshold to accurately detect attacks and minimize packet drop rates. utilizing a bandwidth control algorithm alongside the adaptive threshold to penalize flows that exceed the threshold, thereby preventing bandwidth depletion and maintaining network functionality even during attacks.

Boonchai et al.[16] aimed to improve the detection and classification of DDoS attacks using deep learning techniques to prevent network failures. The study proposed models based on deep neural networks (DNNs) for multiclass classification of DDoS attacks. Two models were implemented: a simple DNN structure and a Convolutional Autoencoder.

Salih et al.[17] introduced novel lightweight Deep learning models, termed Cybernet models, designed for detecting and recognizing various Distributed Denial of Service (DDoS) attacks. These models were characterized by having fewer than 225,000 learnable parameters, making them computationally efficient with faster training and inference times.

Song et al.[18] addressed the issue of network attackers evolving and presenting new patterns, using deep-learning techniques. The model determines the optimal model architecture and hyperparameter settings for autoencoders.

Importance of Security in IoT

IoT devices are used for various purposes through an open network which makes the devices, therefore, more accessible to the users. On one hand, IoT makes human life technologically advanced, easygoing, and conformable[19].

IDSs are one promising avenue for monitoring IoT environments and are mainly effective at the network level. IDSs deployed in IoT environments analyze network data packets and generate real-time responses[20], the utilization of IDS is required to detect attacks and identify anomalies in networks, alerting security systems to potential threats, which can use this information to implement more effective controls. There are many types of IDS, the most common are.

- A system that analyzes incoming network traffic.
- A system that monitors the computer infrastructure and operating system files.

There is also a subset of IDS types, the most common are**.**

- Signature-based technique: which scans system files and detects anomalies by looking for specific patterns, such as byte sequences that may be found within network packet headers or known malicious instruction sequences used by malware.
- Anomaly-based: method uses machine learning to classify attacks as either normal or anomalous.

Most networks use Signature-based IDS, but it has limits when it comes to uncovering unknown attacks. Security is a very important issue that could confront IoT development[21]. In this paper, the goal is to improve the performance of Anomaly-based IDS using SVM.

Materials and Method

This section outlines the innovative approach, which combines advanced machine learning techniques with real-time data analysis to accurately identify and classify ransomware threats. The model leverages a unique combination of feature extraction and classification strategies, setting it apart from existing methods. Fig. 1 shows the flowchart of the intrusion detection method based on Gaussian Pyramid (GP) and linear SVM for faster and more accurate training model building of a dataset.

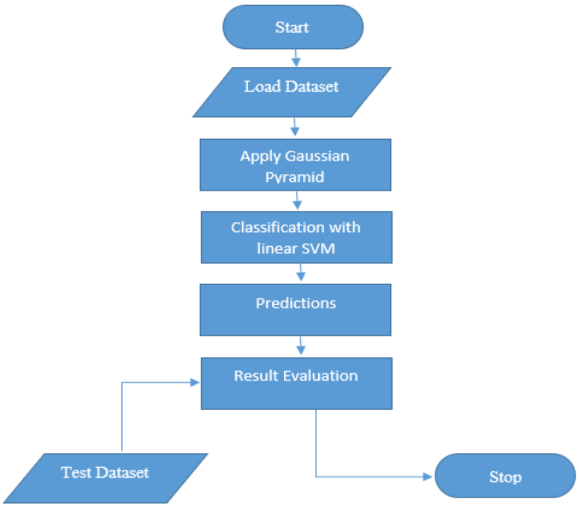

Figure 1. Flowchart of Gaussian Pyramid Model

The system consists of three main components:

Dataset Collection

The model uses different sources of datasets which are considered benchmarks for the analysis of IDS and have been used in many kinds of research, it normalizes the values of the feature to the range of [-1,1] by applying the rule $(X - X_{Mean})/X_{Max}$, the datasets are as follows:

- **NSL-KDD dataset[22]:** A benchmark dataset for network intrusion detection systems, known for its balance of normal and malicious network traffic samples, which addresses certain inherent problems in the original KDD'99 dataset. Generated by Canadian Institute for Cybersecurity unit based at the University of New Brunswick, for anomalous activity detection in IoT networks, this data set is effective in comparing different types of IDS methods, Furthermore, the number of records in the NSL-KDD set is reasonable.
- Table **1** presents the categorical features that were deleted.

Table 1. Details of the NSL-KDD dataset

Dataset	Normal	Abnormal	Total	Num of Features
Training	67.343	58.630	125.973	39
Testing	9711	12.833	22. 544	39

- **CIC-DDoS2019 dataset[23]:** The CIC-DDoS2019 dataset introduced by Canadian Institute for Cybersecurity includes 50,063,112 records, including 50,006,249 rows for Distributed Denial of Service DDoS attacks and 56,863 rows for benign traffic. Each row has 86 features. The statistics for attacks in training and testing for the dataset are summarized in Table 2. The training dataset contains 12 DDoS attacks including Network Time Protocol (NTP), Domain Name System (DNS), Lightweight Directory Access Protocol (LDAP), Microsoft SQL Server (MSSQL), Network Basic Input Output System (NetBIOS), Simple Network Management Protocol (SNMP), Simple Service Discovery Protocol (SSDP), User Datagram Protocol (UDP), UDP-Lag, Web DDoS, SYN and TFTP, while the test dataset contains seven attacks, namely, MSSQL, NetBIOS, PortScan, LDAP, UDP, UDPLag and SYN in the testing day. as it reported in Table 2 below.

Table 2. Details of the CIC-DDoS2019 dataset

Attack Type (DDoS)	Flow Count
DNS	5071011
LDAP	2179930
MSSQL	4522492
NetBIOS	4093279
NTP	1202642
SNMP	5159870
SSDP	2610611
SYN	1582289
TFTP	20082580
UDP	3134645
UDP-Lag	366461
Web DDoS	439
Normal	56863

- **IoTID20 dataset[24]**: provides a reference point to identify anomalous activity across the IoT networks. Specifically designed for IoT networks, this dataset includes a wide range of IoT-specific attack vectors and normal traffic patterns, providing a comprehensive platform for testing intrusion detection methodologies in IoT-specific scenarios. The new IoT botnet dataset has a more comprehensive network and flow-based features, contains various types of IoT attacks and families, a large number of general features, and high-rank features as presented in Table 3.

Table 3. Details of the IoTD20 dataset

Normal	Abnormal	Total	Num of Features
40.073	585.710	625. 783	80

Gaussian Pyramid (GP)

GP is a technique in image processing, developed by the computer vision, image processing, and signal processing communities. The pyramid is constructed by calculating the weighted average of the neighboring pixels and scaling the image down. There are two kinds of operations in the Gaussian Pyramid: Reduce and Expand.

- Reduce Operation: reduce half of the width and height of the image.
- Expand Operation: increase two times the width and height of the image.

In the model, performing a reduce operation on the entered dataset to reduce half of the width and height, each new value contains a local average that corresponds to a weighted neighborhood's previous value on a lower level of the pyramid, and the elements of a Gaussian Pyramids are a smoothed copy of the dataset at one scale.

In the proposed GP model, there are two levels on the pyramid, the first one is the entered dataset, and the second level is the new dataset after passing the Gaussian kernel on it. The reduced operation in GP is done according to the relation given below:

$$g_l(i.j) = \sum_m \sum_n w(m.n) g_{l-1}(2i + m.2j + n)..1$$

Where l represents the level, $w(m.n)$ is the generated Gaussian pyramid.

The main purpose of GP is smoothing the dataset to remove the noise and reduce the degree of nonlinearity in IoT network traffic data.

Support Vector Machine (SVM)

The SVM algorithm is one of the supervised machine learning algorithms, developed by Cortes and Vapnik, for binary classification problems. SVM is an implementation of the Structural Risk Minimization (SRM) principle. it has been used to solve different classification problems, for example, signal processing, recognition of fonts, intrusion detection, classification of facial expression, recognition of images, speech recognition, detection of genes, text classification diagnosis of faults, chemical analysis, and other fields.

SVM differs from the other classification methods significantly. It intends to find a decision boundary (hyperplane) that maximizes the margin between different classes.

There are two types of SVM: linear and nonlinear, the model will use the linear SVM to train the dataset after applying the GP, and the combination of the GP technique and linear SVM will smooth the dataset to reduce the degree of nonlinearity.

Linear SVM: is utilized for linearly distinguishable information, which implies if a dataset can be arranged into two classes by utilizing a solitary straight line, at that point such information is named linearly separable data, and the classifier is utilized as Linear SVM classifier[25].

Given a training dataset of n points of the form:

$$(x_1, y_1), \dots\dots, (x_n, y_n) \quad \dots 2$$

Where $y_i \in \{1, -1\}$, SVM wants to find the "maximum-margin hyperplane", which is defined so that the distance between the hyperplane and the nearest point from either group is maximized, SVM requires the solution of the following unconstrained optimization problem:

$$\underset{w}{sin}\ \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi(w; x_i, y_i) \quad \dots 3$$

Where $\xi(w; x_i, y_i)$ is a loss function, and C > 0 is a penalty parameter.

This paper classifies the data using linear SVM after applying the GP algorithm, by finding a separating line (or hyperplane) that separates the normal class from the anomalous class.

Results and Discussion

The implementation of the study involves a series of steps. First, a comprehensive dataset comprising both benign and ransomware-related IoT traffic is collated and preprocessed. Advanced feature smoothing techniques are then applied to identify ransomware-specific characteristics. These features are integrated into the SVM linear model. The model undergoes extensive training and validation to enhance accuracy and efficiency. The final phase involves deploying these models on novel datasets to assess their effectiveness in real-world scenarios, aiming to detect and classify ransomware in IoT environments.

Performance Analysis

The paper studied important performance metrics to evaluate the performance of the detection model classified.

Precision: refers to the quality of a positive prediction made by the model in Eq. 4.

$$Precision = TP/(TP + FP) \dots 4$$

Recall: it is calculated as the ratio between the number of Positive samples correctly classified as Positive to the total number of Positive samples in Eq. 5.

$$Recall = TP/(TP + FN) \dots \ 5$$

F-measure: it is calculated as the harmonic mean of precision and recall, giving each the same weighting in Eq .6.

$$F1 - score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \dots 6$$

Accuracy: it indicates the proportion of correct classifications of the total records in the testing set in Eq. 7.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \dots .7$$

Where:

- True positive (TP): anomaly instances correctly classified as an anomaly.
- False positive (FP): normal instances wrongly classified as an anomaly.
- True negative (TN): normal instances correctly classified as normal.
- False negative (FN): anomaly instances wrongly classified as normal.

Implementing the algorithm described in the previous section using Python. The code uses sklearn and matplotlib. All experiments are executed on a machine with an Intel(R) Core (TM) i5-5200U 2.20 GHz processor and 8GB of RAM.

The performance of the proposed model is evaluated using 70% of the data for training the SVM classifier, and 30% was used for testing on CIC-DDoS2019 and

IoTID20 datasets, the NSL-KDD dataset contains separate files for training and testing.

Regarding the aim of providing the best IoT intrusion detection using SVM that could detect threats and anomalous activity, the paper shows the result of applying GP on NSL-KDD, CIC-DDoS2019, and IoTID20 datasets with different kernel sizes, in which the model started with kernel size 3 then continue with odd numbers till 9

Comparing the experiment results with traditional SVM and Resilient Backpropagation artificial neural networks (ANN) which is a feedforward ANN, it is used to overcome the weakness of ANN presented in Table **4**.

Table 4. Comparative analysis of the proposed model, nonlinear SVM, and Resilient Backpropagation ANN on the NSL-KDD dataset.

	Precision	**Recall**	**F1-score**	**Accuracy**	**Training Time(s)**	**Testing Time(s)**	**AUC**
Nonlinear SVM	0.80	0.78	0.78	0.78	850.45	90.55	0.82
ANN	0.97	0.92	0.93	0.94	752. 55	80.40	0.87
GP model, kernel size=3	0.70	0.67	0.66	0.67	733. 15	120.84	0.82
GP model, kernel size=5	0.85	0.82	0.82	0.82	633. 17	93. 51	0.91
GP model, kernel size=7	0.85	0.82	0.82	0.82	557. 30	72. 48	0.90
GP model, kernel size=9	0.95	0.95	0.95	0.95	483.77	68.45	0.99

Table 4 shows the performance results of the GP-proposed model with traditional nonlinear SVM and ANN. In the GP model, different kernel sizes from 3 to 9 to evaluate the model. Regarding accuracy and AUC, in kernel size equal to 5,7,9 the GP model gets better results than nonlinear SVM and ANN on the NSL-KDD dataset. Conclusions can be made that, the training time for the GP model is less in all kernels than in traditional SVM which makes GP a lightweight model.

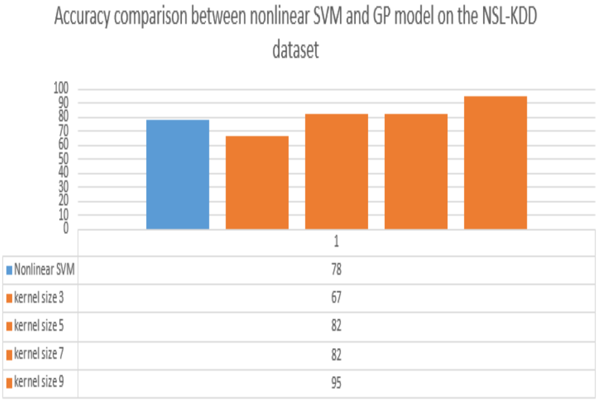

Figure 2. Accuracy comparison between nonlinear SVM and GP model on the NSL-KDD dataset

The results on accuracy are shown in Figure 2, as showing improved classification accuracy in kernel sizes 5,7, and 9 but the accuracy in kernel size 3 is less than traditional nonlinear SVM.

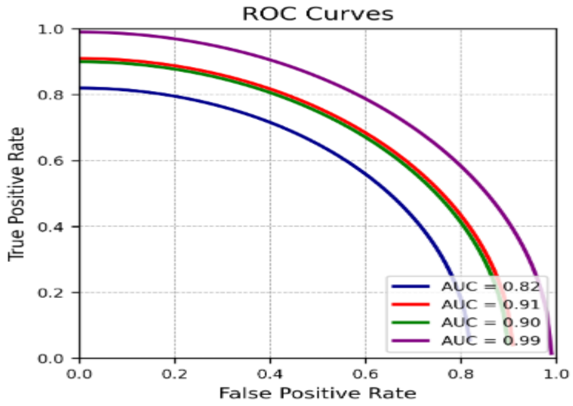

Figure 3.AUC of all kernel sizes of the GP model on the NSL-KDD dataset

As Figure 3 shows the AUC of the GP model on the NSL-KDD dataset has a good performance in all kernels. From a computational viewpoint, GP can train the model on the whole training set with better AUC than classical SVM on large datasets.

Table 5. Comparative analysis of the proposed model, nonlinear SVM, and Resilient Backpropagation ANN on the CIC-DDoS2019 dataset.

	Precision	Recall	F1-score	Accuracy	Training Time(s)	Testing Time(s)	AUC
SVM	0.87	0.87	0.88	0.88	8900.45	390.55	0.89
ANN	0.96	0.96	0.96	0.96	7505.50	355.15	0.96
GP model, kernel size=3	0.92	0.92	0.92	0.92	4172.92	97.84	0.92
GP model, kernel size=5	0.93	0.93	0.93	0.94	5972.7	81.37	0.94
GP model, kernel size=7	0.94	0.94	0.94	0.94	4739.48	71.24	0.95
GP model, kernel size=9	0.95	0.95	0.95	0.96	3509.08	50.55	0.97

According to the Table **5** results, the GP model is preferred especially for large-scale problems, because getting a faster testing time because using GP shrinks the dataset size to ¼ of the original size before the training phase.

In general, AUC is a better measure than accuracy, applying the GP model to the CIC-DDoS2019 dataset gets the best AUC in kernel size 9, in overall it results better than traditional nonlinear SVM with AUC 0.89% and ANN 0.96%.

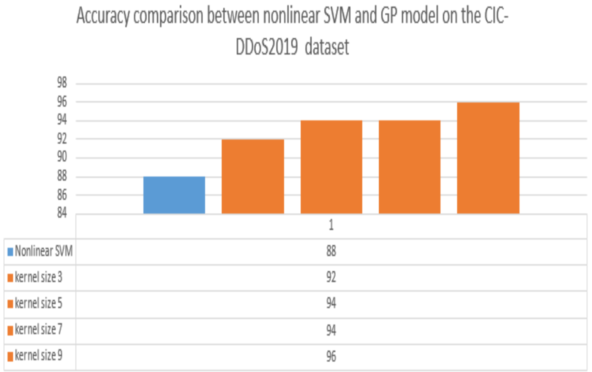

Figure 4. Accuracy comparison between nonlinear SVM and GP model on the CIC-DDoS2019 dataset

Figure 4 shows the highest overall accuracy (0.96%) was obtained with kernel size 9, which proves that the GP model improves intrusion detection on the CIC-DDoS2019 dataset.

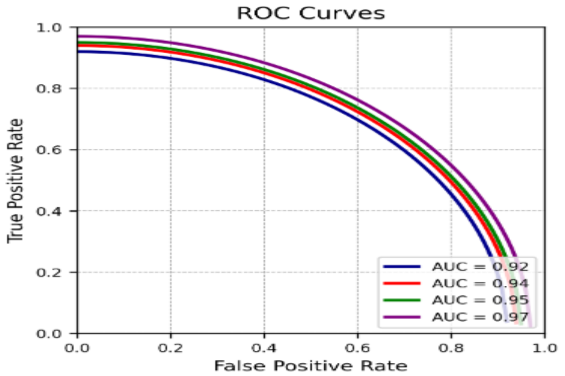

Figure 5. AUC of all kernel sizes of the GP model on the CIC-DDoS2019 dataset

It is noticed in Figure 5 that, the AUC in small kernels is approximate to traditional SVM, as getting bigger kernels, the model gets improved AUC. As the kernel size increases, there is a noticeable improvement in the performance of AUC. This gradual improvement indicates the ability of the GP model to capture more fine-grained relationships within the data as it operates on larger kernel spaces.

Table 6. Comparative analysis of the proposed model, nonlinear SVM, and Resilient Backpropagation ANN on the IoTID20 dataset.

	Precision	Recall	F1-score	Accuracy	Training Time(s)	Testing Time(s)	AUC
Nonlinear SVM	0.91	0.33	0.43	0.43	4400.00	500.00	0.56
ANN	0.80	0.82	0.80	0.78	3430.30	300.55	0.75
GP model, kernel size=3	0.92	0.91	0.91	0.91	2833. 15	120.84	0.85
GP model, kernel size=5	0.93	0.89	0.90	0.89	2123. 86	93. 51	0.88
GP model, kernel size=7	0.93	0.93	0.93	0.93	1406. 79	72. 48	0.90
GP model, kernel size=9	0.93	0.88	0.90	0.88	992.86	68.45	0.90

As

Table 6 shows that traditional nonlinear SVM has a 0.56% AUC and 0.43% accuracy, but the GP model performs better in all kernels. GP classification results show that the best AUC with kernel sizes 7 and 9 proves the effectiveness of the GP model over traditional nonlinear SVM. The training and testing time is less than nonlinear SVM in all kernels because the GP model shrinks the dataset size to ¼ of the original dataset size. This proposed technique helps classifications of large-scale datasets like IoTID20 while maintaining the variance of all kinds of features.

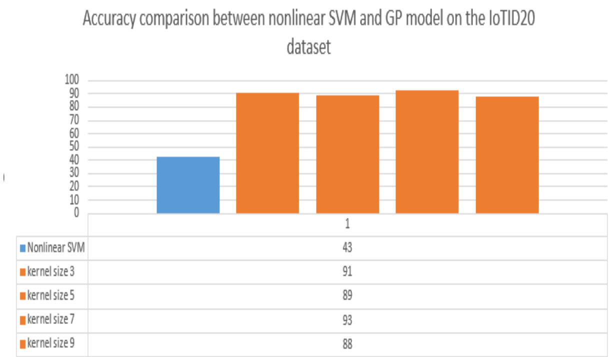

Figure 6. Accuracy comparison between nonlinear SVM and GP model on IoTID20 dataset

As showing in Figure 6 there is a lot of improved performance regarding overall accuracy on the IoTID20 dataset.

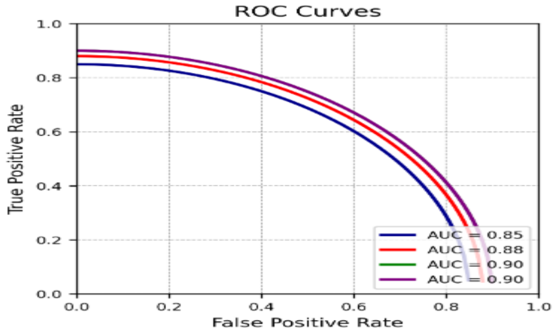

Figure 7. AUC of all kernel sizes of the GP model on the IoTID20 dataset

Figure 1 proves that the GP model improves classification in large datasets, and it exhibits good performance by comparing kernels in the IoTID20 dataset. Providing validation of the ability of the Gaussian process (GP) model to enhance classification accuracy, especially in the context of large datasets. By examining performance metrics across different kernel sizes in the IoTID20 dataset, it becomes clear that the GP model shows remarkable efficiency in distinguishing complex patterns and anomalies as the kernel size expands.

The limitations of the study are:

- Using the Gaussian pyramid is inherently designed for 2D data, such as images. Numeric datasets, particularly in fields like finance, biomedical data, or any multi-dimensional data, do not inherently conform to this 2D structure. Applying Gaussian pyramids to such datasets often requires reshaping or reinterpreting the data, which may not always be meaningful or appropriate.

- The GP model is preferred especially for large-scale problems.

Comparison with Related Previous Work

Table 7. Comparative Analysis of IoT Malware Detection Techniques

Study	Used Technique	No. of Classes	Dataset	Precision	Recall	F1-score	Accuracy
Proposed Technique	Gaussian Pyramid+SVM	Binary	NSL-KDD	95%	95%	95%	95%
Jiang et al. [10]	CNN+BiLSTM	Binary	NSL-KDD	85.82%	84.49%	85.14%	83.58%
Su et al. [11]	BAT-MC	Binary	NSL-KDD	-	-	-	84.25%
Fu et al. [12]	DLNID	Binary	NSL-KDD	86.38%	93.17%	89.65%	90.73%
Wisanwanichthan et al. [13]	DLHA	Binary	NSL-KDD	88.16%	90.14%	89.19%	87.55%
Gao et al. [14]	Adaptive Ensemble	Binary	NSL-KDD	86.50%	86.50%	85.20%	85.20%
Proposed Technique	Gaussian Pyramid+SVM	Multi	CICDDoS2019	95%	95%	95%	96%
Alamri et al. [15]	XGBoost	Multi	CICDDoS2019	93%	92%	92%	91.26%
Boonchai et al.[16]	DNN, CNN-AE	Multi	CICDDoS2019	-	-	-	91.9%
Salih et al.[17]	3 Dense layers	Multi	CICDDoS2019	94.21%	94%	94%	94.21%
Proposed Technique	Gaussian Pyramid+SVM	Binary	IoTID20	93%	93%	93%	93%
Goutte et al.[24]	SVM	Binary	IoTID20	-	-	-	84%
Song et al[18]	DL	Binary	IoTID20	-	-	97%	94%

Table ***7*** offers a comparative analysis of various IoT malware detection techniques on the NSL-KDD, CICDDoS2019, and IoTID20 datasets. It evaluates the proposed GP model on the NSL-KDD dataset and compares it with methods developed by Jiang et al[10], Su et al[11], Fu et al[12], Wisanwanichthan et al[13], and Gao et al[14]. Adaptive Ensemble. The comparison focuses on key performance indicators like Precision, Recall, F1-score, and Accuracy. The proposed GP model shows superior performance with 95% precision, 95% recall, a 95% F1 score, and Table ***7*** while Goutte et al[24] achieved 84% using SVM, and the accuracy of Song et al[18] is 94% compared to 93% in this study.

95% accuracy. The GP technique shows varying degrees of effectiveness, with accuracy ranging from 67% to 95%, indicating a competitive landscape in IoT malware detection methodologies on the NSL-KDD dataset.

Achieving 96% on multi-class CICDDoS2019 which indicates the effectiveness of the GP model on multi-class cases. The proposed GP model in this study on the IoTID20 dataset achieved 93% as shown in

Conclusion

The proposed model helps IDs to get rid of online threats and detect cybercriminals to protect sensitive information in IoT networks using machine learning technologies.

The integration of Gaussian pyramids with linear SVM offers remarkable improvements in prediction accuracy, speed, and computational efficiency. This model proves especially effective in managing large-scale problems, outperforming traditional nonlinear SVM in terms of both spatial and temporal efficiency. Looking ahead, the aim is to further test the GP model in parallel computing environments and on a broader range of real-world scenarios to comprehensively evaluate its performance.

Authors' Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.
- No animal studies are present in the manuscript.
- No human studies are present in the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee at Syrian Virtual University.

Authors Contribution statement

G.K. devised the model, R.A.Z. worked out almost all of the technical details, and performed the numerical calculations for the suggested experiment.

References

1. John Dian F, Vahidnia R, Rahmati A. Wearables and the Internet of Things (IoT), applications, opportunities, and challenges: A Survey. IEEE Access. 2020; (8): 69200-69211. https://doi.org/10.1109/access.2020.2986329
2. Meghana S, Srinath R. A novel mechanism for clone attack detection in hybrid IoT. Int Res J Eng Technol. 2019; 7(5):264-268.
3. Abdulhadi HM, Aldeen YAAS, Yousif MA, Jaseem M Jalal, Madni SHH. Enhancing Smart Cities with IoT and Cloud Computing: A Study on Integrating Wireless Ad Hoc Networks for Efficient Communication. Baghdad Sci J. 2023; 20(6 Suppl): 2672-2672. https://doi.org/10.21123/bsj.2023.9277
4. Awajan A. A novel deep learning-based intrusion detection system for IOT networks. Computers. 2023; 12(2): 34-51. https://doi.org/10.3390/computers12020034
5. Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning. J Appl Sci Technol Trends. 2021; 2(01): 20-28. https://doi.org/10.38094/jastt20165
6. Piccialli V, Sciandrone M. Nonlinear optimization and support vector machines. Ann Oper Res. 2022; 314(1): 15-47. https://doi.org/10.1007/s10288-018-0378-2
7. Prakruthi ST, Muralidharan A, Dhanalakshmi B, Dubey A. A Survey on the Various UAV Landing Sign Detection Techniques. 2018; 6(3):1417-1420.
8. Tavara S. Parallel computing of support vector machines: a survey. ACM Comput Surv. 2019; (6): 1-38. https://doi.org/10.1145/3280989
9. Lou C, Xie X. Multi-view universum support vector machines with insensitive pinball loss. Expert Syst Appl. 2024; 248: 123480. https://doi.org/10.1016/j.eswa.2024.123480
10. Jiang K, Wang W, Wang A, Wu H. Network intrusion detection combined hybrid sampling with deep hierarchical network. IEEE Access. 2020; 8: 32464-32476. https://doi.org/10.1109/access.2020.2973730
11. Su T, Sun H, Zhu J, Wang S, Li Y. BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset. IEEE Access. 2020. https://doi.org/10.1109/access.2020.2972627
12. Fu Y, Du Y, Cao Z, Li Q, Xiang W. A deep learning model for network intrusion detection with imbalanced data. Electronics. 2022; 11(6): 898-900. https://doi.org/10.3390/electronics11060898
13. Wisanwanichthan T, Thammawichai M. A double-layered hybrid approach for network intrusion detection system using combined naive bayes and SVM. IEEE Access. 2021; 9: 138432-138450. https://doi.org/10.1109/access.2021.3118573
14. Al-Qatf M, Lasheng Y, Al-Habib M, Al-Sabahi K. Deep learning approach combining sparse autoencoder with SVM for network intrusion detection. IEEE Access. 2018; 8: 194269-194288. https://doi.org/10.1109/access.2018.2869577
15. Alamri HA, Thayananthan V. Bandwidth control mechanism and extreme gradient boosting algorithm for protecting software-defined networks against DDoS attacks. IEEE Access. 2020. https://doi.org/10.1109/access.2020.3033942
16. Boonchai J, Kitchat K, Nonsiri S. The classification of DDoS attacks using deep learning techniques. In: 2022 7th International Conference on Business and Industrial Research (ICBIR); 2022. https://doi.org/10.1109/icbir54589.2022.9786394
17. Salih AA, Abdulrazaq MB. Cybernet Model: A New Deep Learning Model for Cyber DDoS Attacks Detection and Recognition. Comput Mater Contin. 2024; 78: 1275-1295. https://doi.org/10.32604/cmc.2023.046101
18. Song Y, Hyun S, Cheong YG. Analysis of autoencoders for network intrusion detection. 2021. https://doi.org/10.3390/s21134294
19. Kurani A, Doshi P, Vakharia A, Shah M. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. Ann Data Sci. 2023; 10(1): 183-208. https://doi.org/10.1007/s40745-021-00344-x
20. Huang J, Lu J, Ling CX. Comparing naive Bayes, decision trees, and SVM with AUC and accuracy. In:

Third IEEE International Conference on Data Mining; 2023. https://doi.org/10.1109/icdm.2003.1250975

21. Salim KG, Al-alak SMK, Jawad MJ. Improved image security in Internet of Things (IoT) using multiple key AES. Baghdad Sci J. 2021; 18(2): 0417-0417. https://doi.org/10.21123/bsj.2021.18.2.0417
22. NSL-KDD dataset. Canadian Institute for Cybersecurity.
23. DDoS evaluation dataset (CIC-DDoS2019) dataset. Canadian Institute for Cybersecurity.
24. Goutte C, Zhu X. Advances in Artificial Intelligence: 33rd Canadian Conference on Artificial Intelligence; 2020. https://doi.org/10.1007/978-3-030-47358-7
25. Abo Zidan R, Karraz G. Gaussian Pyramid for Nonlinear Support Vector Machine. Appl Comput Intell Soft Comput. 2022; 2022(1): 5255346. https://doi.org/10.1155/2022/5255346

نحو انترنت فعال لكشف التسلل بشبكة إنترنت الأشياء باستخدام آلات المتجهات الداعمة

روان أبوزيدان[1]، جورج كراز[2]

[1] برنامج الدكتوراه، كلية هندسة المعلوماتية، الجامعة الافتراضية السورية، دمشق، سوريا.
[2] قسم الذكاء الاصطناعي ومعالجة اللغات الطبيعية، كلية هندسة تكنولوجيا المعلومات، جامعة دمشق، دمشق، سوريا.

الخلاصة

تعد أنظمة كشف التسلل ضرورية لحماية شبكات إنترنت الأشياء من التهديدات الأمنية، لقد كان تكامل آلات المتجهات الداعمة مع أنظمة كشف التسلسل الذكية بمثابة تقدم كبير في اكتشاف الأنشطة الشاذة. يساهم البحث في هذا المجال من خلال استخدام خوارزمية الهرم الغاوسي، مما يقلل بشكل كبير من كمية المعالجة ومساحة التخزين المطلوبة لمجموعات بيانات حركة مرور شبكة إنترنت الأشياء الكبيرة. يمكّن نموذج الهرم الغاوسي من تصنيف آلاف نقاط البيانات في المشكلات الواسعة النطاق مع مساحات إدخال عالية الابعاد. ومن الجدير بالذكر تفوق الهرم الغاوسي بمختلف احجام النواة يتفوق على آلات المتجهات الداعمة غير الخطية التقليدية وعلى الشبكات العصبونية الاصطناعية من حيث الكفاءة والدقة. على سبيل المثال مع أحجام النواة 5، 7و9 أظهر نموذج الهرم الغاوسي أداءاً فائقاً على مجموعة بيانات NSL-KDD، محققاً الدقة والمنطقة أسفل المنحني أعلى من آلات المتجهات الداعمة التقليدية والشبكات العصبونية. في النواة ذات الحجم 9 حقق النموذج دقة قدرها 0.96 % على مجموعة بيانات CIC-DDoS2019. تؤكد نتائجنا التجريبية أن تطبيق نموذج الهرم الغاوسي على حركة بيانات انترنت الأشياء يقلل بشكل كبير من تعقيد الوقت ويعزز أداء آلات المتجهات الداعمة الثنائية ومتعددة الصفوف، مما يمثل تقدماً كبيراً في اكتشاف التسلسل في إنترنت الأشياء.

الكلمات المفتاحية: الهرم الغاوسي، نموذج GP، أنظمة كشف التسلل، إنترنت الأشياء، آلات المتجهات الداعمة.