Cyber Security and Applications 2 (2024) 100033

Contents lists available at ScienceDirect

Cyber Security and Applications

journal homepage: http://www.keaipublishing.com/en/journals/cyber-security-and-applications/

Enhancing intrusion detection systems through dimensionality reduction: A comparative study of machine learning techniques for cyber security

Faisal Nabi [a,*], Xujuan Zhou [b]

[a] Muhammad Ali Jinnah University, Karachi Pakistan
[b] University of Southern Queensland, Toowoomba, Australia 4350

ARTICLE INFO

Keywords:
Cyber security
Intrusion detection system
Supervised machine learning
Anomaly detection
PCA
Random projection

ABSTRACT

Our research aims to improve automated intrusion detection by developing a highly accurate classifier with minimal false alarms. The motivation behind our work is to tackle the challenges of high dimensionality in intrusion detection and enhance the classification performance of classifiers, ultimately leading to more accurate and efficient detection of intrusions. To achieve this, we conduct experiments using the NSL-KDD data set, a widely used benchmark in this domain. This data set comprises approximately 126,000 samples of normal and abnormal network traffic for training and 23,000 samples for testing. Initially, we employ the entire feature set to train classifiers, and the outcomes are promising. Among the classifiers tested, the J48 tree achieves the highest reported accuracy of 79.1 percent. To enhance classifier performance, we explore two projection approaches: Random Projection and PCA. Random Projection yields notable improvements, with the PART algorithm achieving the best-reported accuracy of 82.0 %, outperforming the original feature set. Moreover, random projection proves to be more time-efficient than PCA across most classifiers. Our findings demonstrate the effectiveness of random projection in improving intrusion detection accuracy while reducing training time. This research contributes valuable insights to the cybersecurity field and fosters potential advancements in intrusion detection systems.

Introduction

Due to the increasing frequency and sophistication of cyber-attacks across various domains, network security has become a critical area of research garnering global attention. Cybercriminals employ diverse techniques to breach users' security, gaining unauthorized access to sensitive data and profiting from activities like eavesdropping [1]. Conventional firewalls and anti-virus software, unfortunately, fall short in detecting zero-day attacks, denial of service attacks, data theft, and other sophisticated attack types. As a result, cyber security crimes continue to rise due to vulnerabilities in computer systems, ineffective security policies, and a lack of awareness about cybercrime [2]. In 2016 alone, over three billion zero-day attacks were reported, necessitating urgent and effective solutions to combat these threats [3].

In response to these challenges, intrusion detection systems (IDSs) have garnered significant attention from cyber security researchers. IDSs are software products designed to automate the process of monitoring and analyzing intrusions. An intrusion is defined as any attempt to compromise the confidentiality, integrity, availability, or bypass the security mechanisms of a network or a computer system [4]. Unlike traditional firewalls, the primary objective of an intrusion detection system is to detect various signs of attacks as early as possible.

By proactively identifying and responding to potential intrusions, IDSs play a crucial role in enhancing network security and safeguarding against evolving cyber threats. Their ability to detect and mitigate attacks in real-time is vital in maintaining the integrity and confidentiality of sensitive data, thereby making them an indispensable component in modern cybersecurity strategies. As cyber-attacks continue to evolve, ongoing research and advancements in intrusion detection systems will remain essential in ensuring the resilience and security of our interconnected digital world.

In the realm of intrusion detection systems (IDSs), there are two main types: signature-based IDSs and anomaly-based IDSs. Signature-based IDSs analyze incoming traffic by comparing it to predefined patterns representing known attacks. They are effective at detecting attacks with high accuracy and low false alarms but are limited to recognizing only attacks stored in their database, necessitating constant updates with new attack signatures [4]. On the other hand, anomaly-based IDSs continuously monitor incoming traffic, raising an alarm if any deviation from normal behavior exceeds a certain threshold. These systems can detect novel attack types but may generate a larger number of false alarms [5].

Peer review under responsibility of KeAi Communications Co., Ltd.
* Corresponding author.
E-mail address: faisal.nabi@yahoo.com (F. Nabi).

https://doi.org/10.1016/j.csa.2023.100033
Received 4 August 2023; Received in revised form 28 November 2023; Accepted 14 December 2023
Available online 11 January 2024

Anomaly-based IDSs learn normal behavior using machine learning (ML) algorithms with various data instances characterizing network traffic. ML techniques are divided into unsupervised (no labeled classes) and supervised (labeled classes) learning, with the latter being common in anomaly detection. Supervised algorithms utilize labeled data sets representing normal and anomaly behaviors as features, training a classification model to detect new attack patterns and raise alarms [1].

Many anomaly-based IDSs employing ML algorithms have been proposed (e.g., [6–14]). However, a key challenge is the high dimensionality of data sets used for training the classification models, leading to increased training time. This is crucial for the effectiveness of online IDSs. Additionally, redundant information may exist, reducing classification accuracy and increasing false alarms. To address this, dimensionality reduction approaches are used, transforming the high-dimensional feature space into a lower-dimensional space. Techniques like Principal Component Analysis (PCA) preserve variance between data instances, while faster solutions like random projection (RP) use a random matrix based on a certain distribution, such as Gaussian, to reduce dimensionality [9].

This paper aims to investigate the performance of common supervised machine learning algorithms for anomaly-based intrusion detection (IDs). Additionally, the impact of two dimensionality reduction techniques, namely Principal Component Analysis (PCA) and Random Projection (RP), on classification performance is explored. While PCA is a well-known method in this domain, its time-consuming nature prompts us to assess the performance of RP, which offers a faster alternative. As a result, the main contributions of this work are as follows:

- Analyzing the performance of commonly used supervised machine learning algorithms for anomaly-based intrusion detection. This analysis involves training a classification model using approximately 126,000 samples of normal and anomaly patterns from the NSL-KDD data set.
- Examining the effect of applying PCA dimensionality reduction algorithm on classification performance.
- Examining the effect of applying RP dimensionality reduction algorithm on classification performance.
- Comparing the classification performance achieved by PCA and RP to identify potential advantages and trade-offs of each approach.

By addressing these aspects, this study seeks to provide valuable insights into the effectiveness of supervised machine learning algorithms for anomaly-based intrusion detection and the impact of dimensionality reduction techniques on classification performance. The findings will contribute to a better understanding of which methods are more suitable for efficient and accurate intrusion detection in practical applications.

The remainder of this paper is structured as follows: Section 2 reviews related work in this field. Section 3 discusses the methodology employed in this study. Section 4 delves into the experimental findings. Finally, Section 5 concludes the paper and makes recommendations for future research.

Related work

The area of supervised learning and intrusion detection has garnered significant attention among cyber security researchers. Numerous studies focus on applying common supervised ML techniques and evaluating their performance on popular intrusion datasets. Examples of these techniques include decision trees, random forests, Bayes methods, support vector machines (SVM), neural networks, ensemble classifiers, and more.

Supervised ML for intrusion detection

Recently, the authors in [6] experimented with four supervised machine-learning algorithms for intrusion detection: logistic regression, SVM, naïve Bayes, and random forest. Training was conducted on the NSL-KDD dataset, covering four attack types (DOS, Probe, user to root, root to local). Reported accuracy results are 84 % (logistic regression), 79 % (naïve Bayes), 75 % (SVM), and 99 % (random forest). Random forest's near-perfect accuracy raises overfitting concerns. In [7], the same problem was addressed with cross-validation as the validation method and feature selection applied before feeding the data to three classifiers: j48, naïve Bayes, and REPTREE. Feature selection proved effective in enhancing classification performance. In [8], SVM and k-nearest neighbor were tested on the KDD CUP99 dataset (32,000 samples) for normal and four attack types. Two experiments were conducted: one using the full feature set and the other with PCA for dimensionality reduction. PCA improved accuracy to around 90 % in both cases. Similarly, in [9], SVM with different kernels was experimented for intrusion detection. PCA was effective in enhancing classification performance, with the RBF kernel SVM achieving over 99 % accuracy, though overfitting concerns remain. A similar approach was applied in [10], yielding improved classification performance with PCA.

In [11], the authors focused on detecting distributed DOS attacks (DDOS) using machine learning algorithms on the CICIDS2017 dataset. Feature selection reduced the feature set from 85 to 12 features, and random forest achieved the best results with around 96 % accuracy. High training time raised concerns. In [12], SVM and artificial neural networks were experimented for intrusion detection on the UNSW-NB-15 dataset. Feature reduction methods (categorization, univariate feature selection, PCA) were employed, and categorization yielded the best results with over 90 % accuracy, outperforming PCA. In [13], k-means clustering with feature selection was proposed for intrusion prediction on the KYOTO dataset. Clustering significantly improved classification performance, achieving very high accuracy rates.

In [14], a different approach using random projection for intrusion detection based on Apache web server log data was explored. The approach showed potential for effective intrusion identification through visualization. Lastly, in [15], an end-to-end system was proposed for intrusion detection using novel data sets simulating intrusion in LAN and cloud environments. Decision tree and regression showed good results in LAN and cloud environments, respectively.

In [19], the authors used the KDD'99 and the NSL-KDD datasets to train decision tree (DT), multi-layer perceptron (MLP), random forest (RF), and a stacked autoencoder (SAE) model for detecting network intrusion. In their comparative study, they claimed that the random forest classifier showed the most consistent and accurate results. Similarly, the authors of [21] also used the benchmarking dataset NSL-KDD to conduct a comparative study for intrusion detection using four ML techniques including Random Forest, J48, ZeroR, and Naïve Bayes. However, they did not involve the data dimensions reduction techniques in their study.

Summary for identifying the research gaps

The problem of intrusion detection and supervised learning has garnered global attention, leading to numerous studies using various ML algorithms and validation methods. Notably, the choice of validation method can significantly impact classification performance, with cross-validation often providing better results than independent testing data sets. Additionally, the size of the testing data set has a bearing on the classification outcomes. While some works report very high classification results, concerns arise about potential overfitting issues. Moreover, it is observed that normal behavior yields better accuracy measures compared to intrusion behaviors, an aspect often overlooked in overall accuracy reporting across all classes.

Feature selection and PCA are frequently utilized to reduce dimensionality and generally enhance classification performance. However, PCA comes with a substantial training time cost due to matrix calculations. In contrast, our work proposes a novel approach by employing random projection combined with machine learning for intrusion detection, a highly efficient and rapid method in comparison to PCA. The results demonstrate the superiority of the random projection approach

Table 1
Number of instances per class.

Class	Training data # Instances	Testing data # Instances
Normal	67,343	9711
Anomaly	58,630	12,833

over both PCA and the full feature set, as illustrated in subsequent sections. To assess its performance, we evaluate the proposed approach on an independent data set from NSL-KDD comprising over 22,000 samples representing normal and anomaly behaviors.

Methodology

In this section, the methodology for building the classifier and feature selection analysis is presented.

Study dataset

The data set used in this study is NSL-KDD full dataset available from the UNB data sets repository [16]. The dataset is collected from diverse sources, such as network traffic flows, and contain valuable information about user behavior, host configurations, and system settings. Analyzing this information is crucial for studying attack patterns and identifying abnormal behaviors. It consists of a diverse range of intrusions simulated in a military network environment. It simulated a typical US Air Force LAN to create an environment to acquire raw TCP dump data for a network. The LAN was simulated like a real environment and breached with multiple attacks. A connection is a sequence of TCP packets starting and ending at some time duration between which data flows to and from a source IP address to a target IP address under some defined protocol. There are 125, 973 TCP/IP connections (instances) that are characterized with 41 features extracted from normal and anomaly data for training the model, and 22,544 for testing the model as illustrated in Table 1. Moreover, a sample of the feature set is illustrated in Table 2. According to [16], it does not contain any redundancy in the training records. Moreover, there are no duplicates in the testing data. We normalize the data before feeding it to the dimensionality reduction algorithms.

Some well-known datasets in this domain include DARPA, KDD CUP99, NSL-KDD, KYOTO, CICIDS2017, UNSW-NB-15, among others. For comprehensive details regarding these datasets, including feature sets, classes, and other relevant information, the authors in [2] provide a detailed discussion. The NSL-KDD dataset was chosen as it's a widely recognized benchmark for intrusion detection, providing diverse attack samples. Using a consistent dataset allows fair comparison of ML algorithms. The research focused on dimensionality reduction's impact on classification using NSL-KDD. Evaluating multiple algorithms on this dataset ensures reliable conclusions. Future work can explore different datasets to assess algorithm performance in various scenarios.

Dimensionality reduction

As the dimensionality of the feature set is relatively high (41 features), we experiment with two projection approaches for reducing the dimensionality of the feature set: 1) principal component analysis (PCA) and random projection (RP).

In the first approach, PCA, the high dimensional feature space is reduced into a lower-dimensional feature space using an orthogonal projection that maximizes the variance and separation between data and can lead to better classification performance. Given a P- dimensional observed data vector, y. PCA transforms the data observation into a lower-dimensional space of dimension D, where each observation x, in this lower dimensionality space can be expressed as

$$x = W\ (y - \mu) \tag{1}$$

Where W is a $P \times D$ matrix achieving the desired linear transformation of the data and μ is the mean of the data. The P-dimensional vectors of the matrix W are given by the D dominant eigenvectors (υ), associated with the highest Eigenvalues (λ), of the sample covariance matrix

$$S = \frac{\sum_{i=1}^{N} (y_i - \mu)(y_i - \mu)^T}{N} \tag{2}$$

Such that $S_\upsilon = \lambda_\upsilon$ and N is the number of observations. The data in the reduced space are uncorrelated such that their covariance.

Matrix $S_x = \sum_{i=1}^{n} \frac{xx^T}{N}$ is diagonal and its elements are the Eigenvalues, (λ) [17].

On the other hand, there is a simple yet efficient method for dimensionality reduction based on random projections.

In this method, the original data Y in a higher dimensional space, is transformed into a lower dimensional space X via: $X = WY$, where W is a DXP random matrix where D is of a very small dimensionality compared to P and its columns are realizations of independent and identically distributed zero-mean-normal variables that are scaled to have a unit length. This idea is motivated by the Johnson- Lindenstrauss lemma which states that if points in high dimensional feature space of dimension P, are projected onto a randomly selected lower-dimensional space of suitable dimension D, then the distances between points are approximately preserved if D is large enough.

$$\left\langle \left(\|\phi(y_i) - \phi(y_j)\|_D^2 - \|y_i - y_j\|_D^2 \right) \right\rangle_\emptyset \leq_D^2 \|(y_i) - (y_j)\|_P^4 \tag{3}$$

where $\| \|_P$ and $\| \|_D$ denote the Euclidian distances norms in V_P and V_D, respectively and $< >_\emptyset$ is the average overall possible is tropic random choices for the unit vectors defining the random mapping $\emptyset$ [17].

In our experiments, we evaluate the performance of random projection based on two different choices for the elements of the matrix W:

- The first choice is generated using a Gaussian distribution with the satisfaction of two main properties: orthogonality and normality.
- The Gaussian distribution can be replaced by a simpler distribution, we refer to it as Sparse, such as:

$$W = \sqrt{3} \times \begin{cases} -1\ with\ probablity \frac{1}{2} \\ +1\ with\ probablity \frac{1}{2} \end{cases} \tag{4}$$

Classification

Once the data set is prepared, it is fed to the chosen supervised machine learning algorithms, to experiment with their classification performance on the ability to differentiate between the two classes: normal and anomalous. We experiment before and after dimensionality reduction and use five well-known classification algorithms: BayesNet, Naïve Bayes, J48, PART, and Random Forest. BayesNet is a classification technique that probabilistic graphical model that uses a directed acyclic graph for representing the feature set and their conditional dependencies. Naïve Bayes is a classification technique based on the Bayes' theorem [17], this theorem can describe the probability of an event based on the previous knowledge of conditions related to that event. Naïve Bayes classifier task to classify a new object to a specific class assumes that the feature in classes is not directly related. J48 algorithm is the java implementation of the C4.5 algorithm which builds decision trees based on the training data. PART algorithm iterates for several iterations build a partial decision tree using the C4.5 algorithm at each iteration and makes the best leaf into a rule. Finally, the random forest algorithm is a classification technique that constructs multiple decision trees and outputs the class that represents the average prediction across the multiple trees [18].

Performance evaluation

To evaluate the performance of the classifiers, we build the classifier using the training data set and test its performance on the supplied

Table 2
Description of the features.

Feature	Description
Duration	length (number of seconds) of the connection
protocol type	type of the protocol, e.g. tcp, udp, etc.
Service	network service on the destination, e.g., http, telnet, etc.
src_bytes	number of data bytes from source to destination
dst_bytes	number of data bytes from destination to source
Flag	normal or error status of the connection
Land	1 if connection is from/to the same host/port; 0 otherwise
wrong_fragment	number of "wrong" fragments
Urgent	number of urgent packets
Content features	
Hot	number of "hot" indicators
num_failed_logins	number of failed login attempts
logged_in	1 if successfully logged in; 0 otherwise
num_compromised	number of "'compromised" conditions
root_shell	1 if root shell is obtained; 0 otherwise
su_attempted	1 if "su root" command attempted; 0 otherwise
num_root	number of "root" accesses
num_file_creations	number of file creation operations
num_shells	number of shell prompts
num_access_files	number of operations on access control files
num_outbound_cmds	number of outbound commands in an ftp session
is_hot_login	1 if the login belongs to the "hot" list; 0 otherwise
is_guest_login	1 if the login is a "guest" login; 0 otherwise
Traffic features using a 2-second time window	
count	number of connections to the same host as the current connection in the past two seconds
serror_rate	% of connections that have "SYN" errors
rerror_rate	% of connections that have "REJ" errors
same_srv_rate	% of connections to the same service
diff_srv_rate	% of connections to different services
srv_count	number of connections to the same service as the current connection in the past two seconds
srv_serror_rate	% of connections that have "SYN" errors
srv_rerror_rate	% of connections that have "REJ" errors
srv_diff_host_rate	% of connections to different hosts

testing data set (separate data set from training data set). Classification accuracy is calculated on the tested data as the ratio between correctly classified samples divided by the total number of tested samples. Another performance evaluation measure is the false positive rate (FPR) which calculates the rate of false-positive and is calculated as the number of false positives divided by the total number of true negatives and false positives.

Experiments results and discussions

This section presents the experiment results in three parts. Firstly, we display the classification outcomes across the five classifiers without dimensionality reduction. Next, we present the classification results after applying PCA. Lastly, we analyze the impact of random projection on the classification performance.

Experiment 1: experiments using the full training data set

In this experiment, the five supervised learning algorithms discussed earlier are utilized to build the classifier using the full feature set of 41 features from the training data set. Subsequently, the model is tested on the testing data set. Table 3 displays the accuracy and FPR results obtained from the five supervised machine learning algorithms before any dimensionality reduction is applied.

It is clear from Table 4 that the highest accuracy and the lowest FPR are obtained using the J48 classification algorithm with an accuracy of (79.1 %) and a false positive rate of (18.5 %). Generally, accuracy and FPR results are stable across the five algorithms with no dramatic changes. For further analysis of the performance results of the best classifier (J48), Table 4 presents the confusion matrix for this classifier. The rows represent the ground truth classes and the columns represent the predicted classes.

Table 3
Algorithm's classification results before dimensionality reduction.

Classification Algorithm	Accuracy (%)	FPR (%)
Bayes-Net	71.4	25.5
Naïve Bayes	73.1	23.5
J48	79.1	18.5
PART	73.9	24.0
Random Forest	77.8	20.1

Table 4
Confusion matrix for J48 classifier.

a	B	← classified as
9240	471	*a* = normal
34,231	8602	*b* = anomaly

It is noted from each of these tables that the normal instances are classified correctly with a higher percentage than that of anomaly instances, which we refer to as true positive rate. As it is clear from the above table, the true positive rate for the normal class is (=9240/9711) that is 94.6 % compared to (=8602/12833) is 67.0 % true positive rate for the anomaly class, which indicates the difficulty in predicting new intrusions.

Experiment 2: experiments using classifiers after PCA

In this experiment, we applied PCA to the data set to reduce its dimensionality, and then we assessed the performance of the five classi-

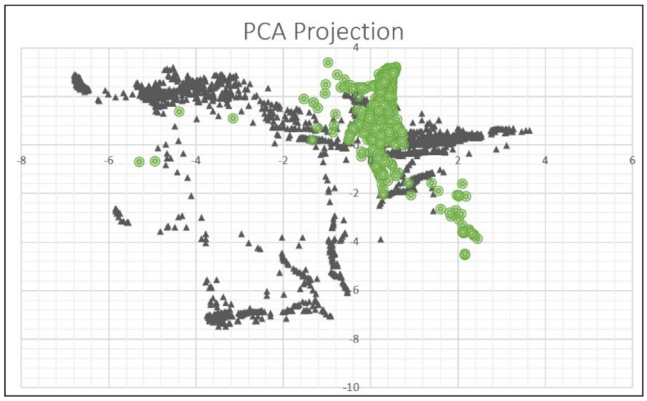

Fig. 1. The PCA projection of 3000 samples from the Normal class (circles) and 3000 samples from the Anomaly class (triangles).

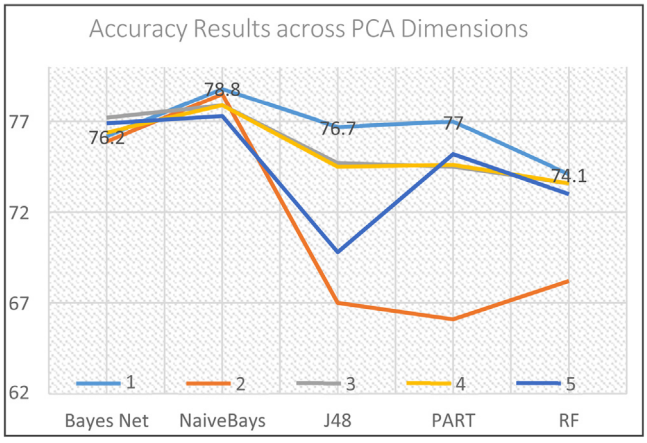

Fig. 2. Classification accuracy results across the first 5 Eigenvectors (dimensions).

Table 5
Algorithm's classification results after projecting the data with PCA.

Classification Algorithm	Accuracy (%)	FPR (%)
Bayes-Net	76.2	21.5
Naïve Bayes	78.8	19.0
J48	76.7	21.0
PART	77.0	20.5
Random Forest	74.1	24.0

fication algorithms in the reduced feature space. Our experiments revealed that approximately 92 % of the data variance could be explained by the first eigenvector of the covariance matrix (first principal component). Fig. 1 illustrates the PCA projection of 6000 samples, while Fig. 2 displays accuracy results across the first five eigenvectors (reduced dimensions). Notably, the highest accuracy was achieved when projecting the data into a one-dimensional feature space. Consequently, we projected the data into a one-dimensional feature space and evaluated its performance using the five aforementioned classification algorithms. The results are summarized in Table 5.

Table 5 clearly indicates that the Naïve Bayes algorithm achieved the highest reported accuracy of 78.8 %. The corresponding confusion matrix, presented in Table 6, reveals a true positive rate of 95.6 % for the normal class, while the anomaly class has a true positive rate of 66.2 %. In general, PCA proved to be effective in enhancing the performance of three out of the five tested classifiers. However, it is worth noting that the best reported accuracy result obtained with the full training data set (J48 algorithm) surpassed that of the reduced data set (Naïve Bayes algorithm). For a visual representation of the effect of applying PCA across the five classifiers, refer to Fig. 3.

Table 6
Confusion matrix for Naïve Bayes classifier.

a	B	← classified as
9281	430	*a* = normal
4343	8490	*b* = anomaly

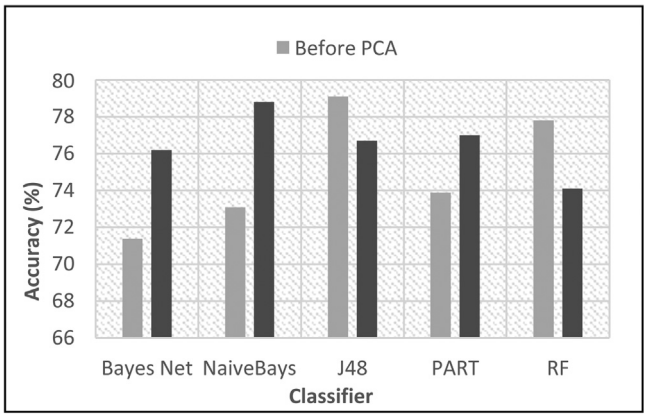

Fig. 3. Classification accuracy results before and after PCA.

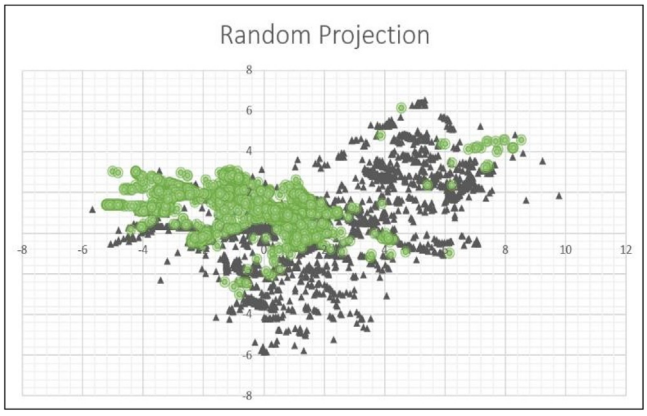

Fig. 4. A random projection of 3000 samples from the Normal class, represented by circles, and 3000 samples from the Anomaly class, represented by triangles.

Experiment 3: experiments using classifiers after random projection

In this experiment, our initial focus is on evaluating the performance of the Gaussian random matrix across the five classification algorithms while varying the reduced dimensions from one to up to five dimensions. Fig. 4 visually presents a random projection of 6000 samples, offering valuable insights. As demonstrated in Fig. 5, the optimal classification performance is achieved when the data is reduced into a 5-dimensional feature space. Consequently, we proceed to transform the data into this reduced space using both Gaussian and Sparse matrices, applying them across the five classification algorithms. The comprehensive results of this process are illustrated in Table 7.

The results presented in Table 7 highlight the superiority of the Gaussian matrix over the Sparse matrix in terms of providing better accuracy and false-positive rates. This is attributed to the Gaussian matrix's ability to achieve a more effective dimensionality reduction, preserving the underlying data structure and relationships more efficiently. The even spread of projected data points in the lower-dimensional space contributes to higher accuracy levels for classifiers trained on the Gaus-

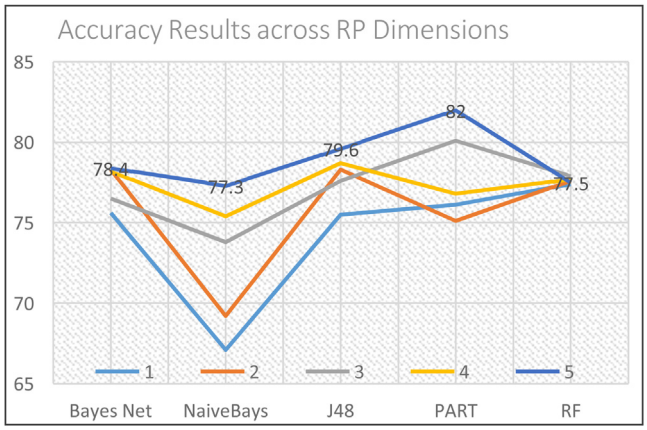

Fig. 5. Classification accuracy results across five dimensions with random projection.

Table 7
Algorithm's classification results after random projection.

Matrix	Gaussian		Sparse	
Classification Algorithm	Accuracy (%)	FPR (%)	Accuracy (%)	FPR (%)
Bayes-Net	78.4	16.0	75.5	22.0
Naïve Bayes	77.3	20.5	71.6	25.0
J48	79.6	17.0	77.2	20.5
PART	82.0	16.2	75.3	21.5
Random Forest	77.5	20.0	77.3	20.5

sian matrix compared to the Sparse matrix. Additionally, the Gaussian matrix outperforms the Sparse matrix in terms of false-positive rates, a crucial metric for intrusion detection systems, ensuring fewer normal instances are misclassified as anomalies. In the absence of the Gaussian matrix, the alternate option is to use the Sparse matrix for dimensionality reduction. However, this choice may come with drawbacks. The Sparse matrix might not perform as effectively as the Gaussian matrix, leading to less accurate classification and reduced ability to discriminate between normal and anomalous instances. The clustering of data points in the lower-dimensional space could result in a loss of relevant information, hindering effective data representation.

Furthermore, using the Sparse matrix may increase the risk of overfitting, particularly with high-dimensional data. The Gaussian matrix's capacity to provide a more generalized representation helps mitigate this risk, while the Sparse matrix might struggle to maintain generalization capability.

In conclusion, the Gaussian matrix emerges as the preferred option for enhancing intrusion detection systems and cyber security due to its ability to retain essential data characteristics, improve accuracy, and reduce false-positive rates. On the other hand, using the Sparse matrix might result in decreased classification performance and increased risk of overfitting. The selection of the Gaussian matrix ensures a more robust and reliable intrusion detection system, making it a valuable dimensionality reduction technique for practical implementation. Therefore, for further comparison with the original high-dimensional data set and PCA results, we will consider the outcomes associated with the Gaussian matrix projection.

Table 7 highlights that the PART algorithm achieved the highest reported accuracy of 82.0 %, making it the best-performing approach in this study. The associated confusion matrix in Table 8 allows us to calculate precision, recall, and F1 measures. Precision is calculated by dividing true positives by the sum of true positives and false positives, while recall is the ratio of true positives to the sum of true positives and false negatives. The F1 measure is calculated as the harmonic mean of precision and recall. Table 9 presents these measures for both classes.

Table 8
Confusion matrix for PART classifier.

a	B	← classified as
9438	273	*a* = normal
3784	9049	*b* = anomaly

Table 9
Precision, recall, and F1 measures for PART classifier.

Measurement Class	Precision (%)	Recall (%)	F1 (%)
Normal	71.4	97.2	82.3
Anomaly	97.1	70.5	81.7

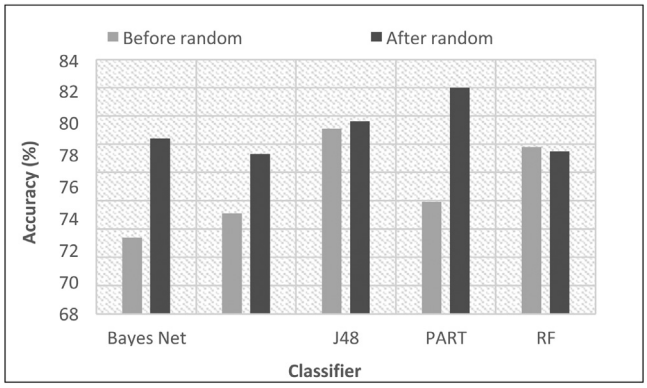

Fig. 6. Classification accuracy results before and after random projection.

Table 10
Comparison between PCA and Random projection accuracy results.

Classification Algorithm	PCA (%)	Random projection (%)
Bayes-Net	76.2	78.4
Naïve Bayes	78.8	77.3
J48	76.7	79.6
PART	77.0	82.0
Random Forest	74.1	77.5

The high precision rate for the anomaly class (97.1 %) indicates that among all predicted instances classified as intrusions, 97.1 % are genuine intrusions. Conversely, the high recall measure for the normal class (97.2 %) indicates that 97.2 % of the instances in the normal class are correctly classified as normal. The overall F1 measure shows a balanced performance for both classes.

Comparing the results after random projection (Gaussian) with the original data set, we observe that it has effectively enhanced the majority of classifiers with improvements ranging from 0.5 % to 8.1 %. These enhancements are illustrated in Fig. 6, demonstrating the efficacy of random projection in improving classification performance across various classifiers.

Experiment 4: comparison between PCA and random projection

Table 10 presents a comparison of accuracy results after applying two projection techniques, PCA and random projection (Gaussian), on the NSL-KDD data set. The table showcases the impact of these dimensionality reduction methods on the classification performance of various supervised machine learning algorithms. The accuracy values for each algorithm are reported, allowing for a direct comparison between the two projection approaches.

From the table, it can be observed how PCA and random projection (Gaussian) influence the performance of the classifiers. The comparison provides insights into the effectiveness of each technique in enhancing the accuracy of intrusion detection systems. The results shed light on

which dimensionality reduction approach yields better performance for each specific classifier, enabling the selection of the most suitable technique based on the desired classification outcome. Overall, this comparison aids in understanding the trade-offs and benefits of using PCA and random projection (Gaussian) for intrusion detection tasks, offering valuable guidance for building robust and efficient cyber security systems.

From the table, several noteworthy observations can be made:

- Random projection (Gaussian) outperforms PCA: Across the majority of the classification algorithms, random projection yields better accuracy results compared to PCA. This suggests that random projection is more effective in preserving the essential data characteristics and improving classification performance for intrusion detection.
- PART classifier with random projection achieves the highest accuracy: Among all the experiments conducted, the best-reported accuracy on the data set is achieved by the PART classifier after applying random projection. This highlights the effectiveness of random projection in enhancing the performance of this specific classifier for intrusion detection.
- Encouraging results for random projection: The results indicate that random projection is a promising dimensionality reduction technique for intrusion detection. Its simplicity, power, and faster implementation make it a viable alternative to PCA in enhancing the accuracy of classifiers for cyber security tasks.

Overall, the comparison demonstrates that random projection is a valuable technique for improving the performance of intrusion detection systems. Its advantages over PCA in terms of accuracy and computational efficiency make it an appealing choice for real-world applications. The encouraging results from these experiments further motivate researchers and practitioners to explore and leverage random projection as an effective tool in the field of cyber security and intrusion detection.

Experiments summary

In summary, the experiments reveal the following key points:

- The full training data set demonstrates effectiveness for classification, achieving 79.1 % accuracy and 18.5 % false-positive rate (FPR) on the testing data set using the J48 algorithm. However, due to its large size (around 126,000 instances), training with the full data set requires a significant amount of time.
- PCA has been effective in enhancing the performance of three classifiers, showing promise in reducing dimensionality and improving classification results. However, the best-reported accuracy achieved using the full data set surpasses the accuracy attained with PCA.
- Random projection is highly effective in enhancing the performance of the majority of classifiers, with accuracy improvements of more than 8.0 % observed with the PART algorithm.
- Applying random projection to the data set provides better accuracy results when compared to using the full training data set, offering a more efficient dimensionality reduction technique.
- Random projection outperforms PCA with the majority of classifiers and requires much less time for computation, making it a more favorable option in terms of both accuracy and efficiency.

These findings suggest that while the full training data set demonstrates strong classification performance, its large size poses computational challenges. PCA and random projection provide effective dimensionality reduction techniques, with random projection showing particular promise in achieving improved accuracy and efficiency across various classifiers. As a result, random projection emerges as a viable and valuable approach for intrusion detection systems and cyber security applications, offering a powerful alternative to the traditional methods for enhancing classification performance.

Conclusion and future work

In this paper, we addressed the problem of automated intrusion detection and utilized the widely used NSL-KDD data set, which contains approximately 126,000 instances for training and 23,000 samples for testing. We applied five popular classification algorithms to the full training data set, namely Bayes Net, Naïve Bayes, J48, PART, and Random Forest. The best-reported results were achieved with the J48 algorithm, attaining a relatively good accuracy of 79.1 %.

To tackle the high dimensionality issue of the 41-dimensional feature vector, we experimented with two projection approaches: PCA and random projection. PCA demonstrated effectiveness in enhancing the performance of three out of the five tested classifiers, resulting in improvements ranging from 3.1 % to 5.7 %. The success of PCA can be attributed to its ability to transform the feature space into a lower-dimensional subspace while retaining crucial information through feature selection. This led to a more efficient and informative data representation, thereby improving classifier performance. Moreover, PCA's noise reduction capability contributed to more accurate and robust classifiers by emphasizing essential data patterns while reducing noise and irrelevant information. Additionally, PCA's ability to prevent overfitting was valuable for high-dimensional datasets, as it provided a more generalized representation of the data. Furthermore, the computational efficiency gained through PCA was beneficial, as it reduced the computational burden on classifiers, making them suitable for real-time or large-scale applications. Additionally, we found that random projection was also effective, improving the performance of the majority of classifiers compared to the original data set. The best-reported accuracy after applying random projection was 82.0 %, outperforming the accuracy achieved before using this technique. Moreover, random projection proved to be more efficient than PCA, requiring less training time for most classifiers.

For future work, we intend to explore other dimensionality reduction techniques, such as LDA and Kernel PCA and other start-of-the-art methods, such as the new method developed in [20], to assess their impact on classification performance. Conducting experiments on various data sets will allow us to identify the most effective approach for enhancing intrusion detection systems' accuracy and efficiency. A combination of supervised, unsupervised, and semi-supervised techniques can be employed to enhance the overall effectiveness of the intrusion detection system in a dynamic and evolving cyber threat landscape in the future work as well. Recently, deep learning methods are also applied to intrusion detection [22,23]. In the future, deep learning will be explored to facilitate intrusion detection systems. By gaining insights into the strengths and limitations of different techniques, we aim to develop more robust and reliable intrusion detection systems capable of effectively countering evolving cyber threats in practical scenarios.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests

Faisal Nabi reports financial support and writing assistance were provided by University of Southern Queensland. Faisal nabi reports a relationship with usq that includes: non-financial support. Faisal nabi has patent pending to n/a. n/a

CRediT authorship contribution statement

Faisal Nabi: Writing – original draft. **Xujuan Zhou:** Formal analysis, Supervision.

References

[1] A. Verma, V. Ranga, Machine learning based intrusion detection systems for IoT applications, Wirel. Person. Commun. 111 (4) (2020) 2287–2310.
[2] A. Thakkar, R. Lohiya, A review of the advancement in intrusion detection datasets, Procedia Comput. Sci. 167 (2020) 636–645.

[3] A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman, Survey of intrusion detection systems: techniques, datasets and challenges, Cyber Secur. 2 (1) (2019) 1–22.
[4] R. Bace, P Mell, NIST Special Publication On Intrusion Detection Systems, Booz-Allen And Hamilton Inc, Mclean VA, 2001.
[5] H. Liu, B. Lang, Machine learning and deep learning methods for intrusion detection systems: a survey, Appl. Sci. 9 (20) (2019) 4396.
[6] M.C. Belavagi, B. Muniyal, Performance evaluation of supervised machine learning algorithms for intrusion detection, Procedia Comput. Sci. 89 (2016) 117–123.
[7] K. Kumar, J.S. Batth, Network intrusion detection with feature selection techniques using machine-learning algorithms, Int. J. Comput. Appl. 150 (12) (2016).
[8] I. Kumar, N. Mohd, C. Bhatt, S.K. Sharma, Development of IDS using supervised machine learning, in: Soft computing: Theories and Applications, Springer, Singapore, 2020, pp. 565–577.
[9] P. Nskh, M.N. Varma, R.R. Naik, Principle component analysis based intrusion detection system using support vector machine, in: 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE, 2016, pp. 1344–1350.
[10] S. Waskle, L. Parashar, U. Singh, Intrusion detection system using PCA with random forest approach, in: 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), IEEE, 2020, pp. 803–808.
[11] N. Bindra, M. Sood, Detecting DDoS attacks using machine learning techniques and contemporary intrusion detection dataset, Autom. Control Comput. Sci. 53 (5) (2019) 419–428.
[12] N. Aboueata, S. Alrasbi, A. Erbad, A. Kassler, D. Bhamare, Supervised machine learning techniques for efficient network intrusion detection, in: 2019 28th International Conference on Computer Communication and Networks (ICCCN), IEEE, 2019, pp. 1–8.
[13] F. Salo, M. Injadat, A. Moubayed, A.B. Nassif, A. Essex, Clustering enabled classification using ensemble feature selection for intrusion detection, in: 2019 International Conference on Computing, Networking and Communications (ICNC), IEEE, 2019, pp. 276–281.
[14] A. Juvonen, T. Hamalainen, An efficient network log anomaly detection system using random projection dimensionality reduction, in: 2014 6th international conference on new technologies, mobility and security (NTMS), IEEE, 2014, pp. 1–5.
[15] G.D.C. Bertoli, L.A.P. Júnior, O. Saotome, A.L. Dos Santos, F.A.N. Verri, C.A.C. Marcondes, ... J.M.P De Oliveira, An end-to-end framework for machine learning-based network intrusion detection system, IEEE Access 9 (2021) 106790–106805.
[16] https://www.unb.ca/cic/datasets/nsl.html, Accessed 29-4-2021
[17] C.M. Bishop, Pattern Recognition and Machine Learning, springer, 2006.
[18] Witten, I.H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with Java implementations.
[19] A. Devarakonda, N. Sharma, P. Saha, S. Ramya, Network intrusion detection: a comparative study of four classifiers using the NSL-KDD and KDD'99 datasets, Journal of Physics: Conference Series, 2161, IOP Publishing, 2022.
[20] S. Anita, S.M. Hadi, N.H. Nosrati, Network intrusion detection using data dimensions reduction techniques, J. Big Data 10 (1) (2023).
[21] K. Arunesh, M. Manoj Kumar, A comparative study of classification techniques for intrusion detection using Nsl-Kdd data sets, Int. J. Adv. Technol. Eng. Sci. 5 (2) (2017).
[22] L. Ashiku, C. Dagli, Network intrusion detection system using deep learning, Procedia Comput. Sci. 185 (2021) 239–247.
[23] Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah, F. Ahmad, Network intrusion detection system: a systematic study of machine learning and deep learning approaches, Transact. Emerg. Telecommun. Technolog. 32 (1) (2021) e4150.