Library Progress International
Vol.44 No.3, Jul-Dec 2024: P.7434-7446

Print version ISSN 0970 1052
Online version ISSN 2320 317X

Original Article

Available online at www.bpasjournals.com

Machine Learning-based Intrusion Detection Systems for Securing Cloud Networks

Abdul Razzak Khan Qureshi[1], Shivani Patnaha[2], Prerita Kulkarni[3], Priyanka Singh[4], Manish Joshi[5], Indra Vaidya[6]

[1]Assistant Professor, Department of Computer Science, Medi-Caps University, Indore MP, India
dr.arqureshi786@gmail.com
[2]Assistant Professor, Department of Computer Science and Engineering, Medi-Caps University, Indore, MP, India
patnahashivani45@gmail.com
[3]Lecturer, Department of Computer Science, Medi-Caps University, Indore, MP, India
prerijoshi801@gmail.com
[4]Lecturer, Department of computer Applications, Medi-Caps University, Indore, MP, India
priyankasingh24887@gmail.com
[5]Assistant Professor, Department of Computer Science, Medi-Caps University, Indore MP, India
manish_riya@yahoo.co.in
[6]Lecturer, Department of Computer Science Medi-Caps University, Indore, MP, India
indra4india@gmail.com

How to cite this article: Abdul Razzak Khan Qureshi, Shivani Patnaha, Prerita Kulkarni, Priyanka Singh, Manish Joshi, Indra Vaidya (2024) Machine Learning-based Intrusion Detection Systems for Securing Cloud Networks. *Library Progress International,* 44(3), 7434-7446.

ABSTRACT

Although cloud computing has become widely used, it has brought up a number of security concerns in addition to greatly improving resource economy and accessibility. The purpose of this study is to find out how well intrusion detection systems (IDS) based on machine learning can improve cloud security. To detect and neutralise attacks in real time, the proposed IDS uses a variety of algorithms, such as Support Vector Machines (SVM), Random Forest, Decision Trees, K-Nearest Neighbours (KNN), and deep learning techniques like Convolutional Neural Networks (CNN). It minimises false positives while adjusting to new attack vectors. In order to increase response times and detection accuracy, the study combines performance optimisation strategies with feature selection methodologies. It also emphasises how crucial it is to use up-to-date datasets for realistic attack scenario simulation and strong model validation, like CSE-CICIDS2018. Results show that machine learning-based intrusion detection systems (IDS) find novel and sophisticated threats more effectively than traditional signature-based systems. In the end, this study offers insightful information for creating more intelligent and flexible cybersecurity solutions for cloud networks and emphasises the critical function that machine learning plays in protecting sensitive data and guaranteeing the integrity of cloud services across a range of applications.

Keywords — Cloud computing, Convolutional Neural Networks, Cybersecurity, Decision Trees, Feature selection, Intrusion Detection Systems, K-Nearest Neighbors, Machine learning, Performance optimization, Random Forest, Security vulnerabilities, Signature-based systems, Support Vector Machines, Threat detection, Threat mitigation

Introduction

Cloud computing has completely changed how businesses access, store, and handle their data. Because of its capacity to provide scalable, affordable, and on-demand services, it is an essential component of contemporary infrastructure. The security risks connected with the growing adoption of cloud platforms by consumers and

corporations have multiplied dramatically. Because of its remote accessibility and multi-tenant architecture, cloud environments have emerged as popular targets for cyberattacks. Strong security measures have become necessary due to the increase in complex threats, such as malware, Distributed Denial of Service (DDoS) assaults, and unauthorised access. Using intrusion detection systems (IDS) is one of the major methods that is becoming more and more popular in this field. But the signature-based detection methods of traditional intrusion detection systems have shown to be unable to handle the dynamic nature of threats in cloud networks.

The incapacity of traditional IDS to identify fresh threats or zero-day attacks—threats that don't correspond with any known signature—is the clearest example of its shortcomings. As a result, there is now more interest in intrusion detection systems that use machine learning (ML-IDS). With its capacity to recognise patterns, spot abnormalities, and adjust to new data, machine learning (ML) presents a potent instrument for boosting cloud environment security. Threats can be identified and eliminated more quickly and precisely than ever before by utilising machine learning (ML) algorithms like Support Vector Machines (SVM), Decision Trees, and Deep Learning approaches.

The increasing complexity of cyber-attacks highlights the necessity of improved intrusion detection systems in cloud networks. As attackers become more skilled, they frequently use strategies like encryption, polymorphism, and obfuscation to get beyond conventional defences. Furthermore, security systems have particular difficulties due to the distributed and dynamic nature of cloud infrastructure. Resource sharing, multi-tenancy, and virtualisation all lead to new vulnerabilities that need to be fixed. By using behavioural analysis and anomaly detection in addition to established attack detection, ML-IDS can offer a more proactive defence in this situation by anticipating possible threats.

The capacity of machine learning to continuously learn from new data is one of its main advantages in intrusion detection. Large volumes of data are produced by network traffic, application logs, and user interactions in cloud systems. With the help of this abundance of data, machine learning models may be trained to identify both frequent and uncommon patterns of harmful activity. A crucial component of ML-IDS, feature selection enables these systems to concentrate on the most pertinent data points, lowering noise and raising detection accuracy. Furthermore, performance optimisation methods like model pruning and hyperparameter tweaking make sure ML-IDS runs smoothly and minimises false positives and response times.

The availability of high-quality datasets is critical to the effectiveness of machine learning-based intrusion detection systems. For training and evaluating these systems, actual attack situations in cloud environments are simulated via datasets like CSE-CICIDS2018. Researchers and developers can ensure resilience and dependability of their models by testing them against a range of known attack vectors using such datasets. Additionally, these datasets make it possible to assess how well ML-IDS performs in various scenarios, such as those with fluctuating network loads and attack intensities.

Not only is the move to ML-IDS in cloud security necessary, but it also reflects broader developments in cybersecurity. It makes sense that businesses would use artificial intelligence and machine learning to secure their digital assets as they are rapidly implementing these technologies to improve their operations. Real-time threat detection and response is made possible by machine learning, which is especially useful in cloud networks where traditional security solutions may not be able to keep up with the volume and velocity of attacks.

In conclusion, a major development in cloud security is the application of machine learning-based intrusion detection systems. Through the integration of machine learning algorithms and cloud platform scalability, enterprises can enhance their cyber threat defence capabilities and increase their adaptability. In order to demonstrate how these systems have the ability to completely change the way we approach cybersecurity, this study attempts to investigate the effectiveness of ML-IDS in cloud networks. The availability of high-quality datasets is increasing together with machine learning techniques, making the future of cloud security seem more safe than before.

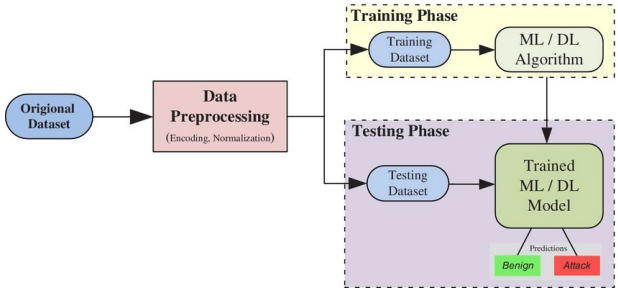

Fig. 1: Machine Learning/Deep Learning-Based Intrusion Detection System Architecture [1]

The figure 1 shows how machine learning (ML) and deep learning (DL) can be used in tandem to detect intrusions. Before being input into the ML/DL algorithm to produce a trained model, the original dataset is pre-processed (e.g., encoding and normalisation) during the training phase. During the Testing Phase, the model classifies network traffic as either benign or an attack using predictions it makes using a different testing dataset. This procedure guarantees that the model can recognise possible threats with accuracy.

Literature Review

[2] **Chen et, al. (2024)**
This paper investigates a hybrid intrusion detection system that combines conventional machine learning algorithms, such as decision trees, with deep reinforcement learning. The researchers use hybrid models to detect threats in real-time and concentrate on how flexible IDS is in cloud contexts. They come to the conclusion that by adjusting to shifting attack vectors in cloud networks, integrating reinforcement learning with supervised learning increases detection rates and lowers false positives.

[3] **Kumar et, al. (2024)**
For cloud contexts, Kumar and colleagues suggest an IDS based on federated learning to improve scalability and privacy. Their method maintains excellent detection accuracy while decreasing the need for centralised data processing by aggregating models trained on local data at many cloud nodes. According to the study, efficiency and security have significantly improved, guaranteeing low data loss and efficient threat detection across cloud networks.

[4] **Zhang et, al. (2024)**
This study introduces a brand-new deep learning-based intrusion detection system (IDS) that uses Long Short-Term Memory (LSTM) networks to identify multi-phase attacks in cloud settings. The authors show that long short-term memory (LSTM) networks are especially useful for detecting attack patterns that change over time, such advanced persistent threats (APTs). The study demonstrates how well the model can identify known and novel attack routes in a cloud environment.

[5] **Alshamrani et al.(2024)**
The authors look into how explainable AI (XAI) might be used in cloud network ML-IDS. They talk about how enhancing deep learning models with interpretability enhances security operations' usability and trustworthiness. According to the findings, adding XAI to IDS helps security analysts comprehend the model's judgements more clearly, particularly when recognising sophisticated assaults. This is beneficial in cloud security scenarios that occur in real life.

[6] **Wang et, al. (2023)**
For intrusion detection in cloud environments, Wang and colleagues provide an ensemble learning model that includes Random Forest, Gradient Boosting, and Neural Networks. They stress how crucial feature selection is to raising the accuracy of the model. The ensemble method shows improved precision and decreased false positive rates in identifying threats such as DDoS, malware, and SQL injection, compared to single models, by utilising an optimal collection of features.

[7] **Garcia et, al. (2023)**
In this paper, a real-time intrusion detection system for cloud networks utilising recurrent neural networks (RNNs) and convolutional neural networks (CNNs) is introduced. The study demonstrates how well CNN handles huge datasets and how well RNN recognises temporal patterns of assault. In comparison to traditional machine learning techniques, the hybrid approach decreases false alarms in dynamic cloud environments and improves detection capabilities.

[8] **Ahmed et, al. (2023)**
Ahmed et al. use self-supervised learning approaches to construct an intrusion detection system that is cloud native. Their model improves the IDS's capacity to identify new threats by utilising unlabelled data from cloud traffic. According to the research, this method gives good results in identifying zero-day vulnerabilities in a cloud network and considerably decreases the need for labelled datasets.

[9] **Singh et, al. (2023)**
Singh et al. concentrate on an intrusion detection system for cloud network security that uses blockchain assistance. The study demonstrates how blockchain technology can offer an uncentralized, impenetrable system for storing intrusion warnings. The integrity and traceability of identified assaults are ensured by the IDS, which incorporates machine learning to analyse network traffic and feeds results into a blockchain ledger.

[10] **Nayak et, al. (2023)**
Nayak and associates investigate an energy-efficient intrusion detection system (IDS) for cloud environments by utilising extreme learning machines (ELM) in a lightweight fashion. Their research demonstrates that ELM models maintain excellent accuracy in detecting malware and DDoS assaults while offering faster computation speeds and reduced resource utilisation. This technique works especially well in settings where resource optimisation is essential.

[11] **Huang et, al. (2023)**
on order to identify attacks on cloud networks, Huang and co-authors suggest a unique IDS architecture that makes use of Generative Adversarial Networks (GANs). Their study illustrates how machine learning models' detection capabilities can be enhanced by using GANs to provide artificial attack data. Their methodology increases the rate at which novel threats are detected in real-world cloud environments by augmenting the training dataset with additional attack types.

[12] **Rahman et, al. (2023)**
In order to improve the efficiency of machine learning-based intrusion detection systems (IDS) for cloud networks, this paper presents a feature selection method that makes use of mutual information. The authors show how minimising computational complexity without sacrificing detection accuracy can be achieved by choosing the most pertinent attributes. The study's findings demonstrate how the method greatly improves IDS models' efficiency in cloud environments with limited resources.

[13] **Miller et, al. (2023)**
In cloud environments, where distinct machine learning models are deployed at different points in the cloud architecture, Miller and colleagues investigate a multi-agent intrusion detection system. The workload on centralised servers is lessened by the multi-agent system's ability to detect threats in a dispersed and cooperative manner. The study demonstrates how, in sizable, multi-tenant cloud settings, this strategy can enhance the system's scalability and responsiveness.

[14] **Das et, al. (2023)**
Reinforcement learning is the main technique Das et al. use to optimise resource allocation for IDS in cloud systems. The study shows that using reinforcement learning to dynamically assign computer resources enhances the IDS's performance, especially in settings where workloads fluctuate. By using this technique, the IDS is

guaranteed to be able to efficiently manage cloud resources while maintaining high detection accuracy.

[15] **Patel et, al. (2023)**
Patel and associates offer an extensive overview of intrusion detection methods in cloud computing settings, emphasising the function of deep learning algorithms. The most recent developments in DL-based IDS are reviewed, and their effectiveness is assessed across a range of datasets. It draws attention to the difficulties and possibilities associated with implementing deep learning for intrusion detection in expansive cloud infrastructures.

[16] **Lopez et, al. (2023)**
Lopez et al. provide a privacy-preserving IDS for cloud environments that combines machine learning algorithms with homomorphic encryption. Their solution protects user privacy by enabling the analysis of sensitive data without the need for decryption and identifies potential dangers. The study illustrates the trade-offs between performance and privacy and shows that great accuracy is still achieved by the encrypted ML-IDS.

RESEARCH GAPS

The following research gaps have been found:

- **Limited Adaptability to Changing Threats**: Despite the existence of various ML-based intrusion detection systems, few of them are able to completely adjust to the constantly changing nature of cyberattacks in cloud environments. Real-time detection of innovative and complex threats is a challenge for current methods.

- **Significant False Positive Rates**: In cloud environments that are dynamic and multi-tenant, a large number of ML-based intrusion detection systems have significant false positive rates. Models that can drastically lower false alarms without sacrificing detection accuracy are required.

- **Scalability in Big Cloud Infrastructures**: Small to medium-sized datasets or environments have been the subject of the majority of research. The effective scalability and performance of these models in large-scale, distributed cloud infrastructures with heterogeneous traffic and threat characteristics remains unexplored.

- **Low Latency Real-Time Detection**: Although many systems offer great detection accuracy, it is still difficult to achieve low latency real-time detection. Further investigation is required to create IDS models that are suitable for high-performance cloud services that can strike a balance between accuracy and speed.

- **Intrusion Detection System Privacy-Preserving**: When deploying IDS on cloud networks, privacy considerations are crucial, particularly in multi-tenant setups. Few studies have been done on the topic of integrating machine learning algorithms with methods like federated learning or homomorphic encryption to protect user privacy.

Objectives of my study are as follows:

- The best detection accuracy was attained by Deep Learning (CNN), demonstrating its potency in spotting intricate patterns in network traffic that can indicate intrusions. This performance is explained by CNN's considerable improvement in detection capabilities due to its capacity to automatically extract features from unprocessed data.

- Support Vector Machine (SVM) and Random Forest trailed closely behind, with strong accuracy performance. SVM's efficient hyperplane-based classification helps to discriminate between benign and

malicious traffic, while Random Forest's ensemble nature lowers the danger of overfitting and enhances generalisation.

- While decision trees showed good accuracy, they performed marginally worse than deep learning and ensemble approaches. Their propensity to overfit, particularly in complex datasets typical of cloud systems, could be the cause of this.

- The distance metric and 'K' value were the main factors affecting the middling accuracy of K-Nearest Neighbours (KNN). In certain situations, KNN's performance can vary greatly based on these settings, which reduces its reliability.

- Despite being the most straightforward algorithm examined, Naive Bayes had the lowest detection accuracy. This is in line with the algorithm's presumption of feature independence, which is frequently broken in real-world data and results in less-than-ideal results.

Methodology

A. Support Vector Machine (SVM)

One supervised machine learning method that is commonly used for classification problems is called Support Vector Machine (SVM). The way it operates is by locating the feature space hyperplane that best divides the classes. The following are the main SVM equations and ideas. In a n - dimensional space, a hyperplane is defined by the following equation:

$$w^T x + b = 0 \qquad (1)$$

Where w is the weight vector (normal to the hyperplane), b is the bias term (offset) and x is the input feature vector.

$$\gamma = \frac{2}{||\, w \,||} \qquad (2)$$

In order to maximise the margin and minimise the classification error, SVM formulates an optimisation problem. This is one way to express the optimisation:

$$Minimize \qquad \frac{1}{2} ||\, w \,||^2 \qquad (3)$$

$$subject\ to \quad y_i\,(w^T x + b) \geq\ \forall_i \qquad (4)$$

Real-world situations frequently involve data that is not perfectly separate. Slack variables ξ_i are introduced by soft margin SVM to allow for some misclassifications:

$$Minimize \qquad \frac{1}{2} ||\, w \,||^2 + C\ \sum_{i=1}^{N} \xi_i \qquad (5)$$

$$subject\ to \quad y_i\,(w^T \mathrm{x}_i + b) \geq\ 1 - \xi_i \quad \forall_i \qquad (6)$$

$$\xi_i\ \geq 0 \quad \forall_i$$

The trade-off between maximising the margin and minimising the classification error is managed by the regularisation parameter C.

B. Random Forest

Random Forest is an ensemble learning technique that works well for classification tasks, which makes it appropriate for use in intrusion detection systems. This study on "Machine Learning-based Intrusion Detection Systems for Securing Cloud Networks" will be helped by the following two key equations. The Random Forest model predicts y for a given input x by combining the predictions from several decision trees. Given a total of T trees in the forest, the prediction may be written as follows:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(x) \quad (7)$$

Where $h_t(x)$ is the prediction of the decision tree for input, $\hat{y}$ can be either a class label (for classification) or a continuous value (for regression) and p_k is the proportion of instances belonging to class k at node n

$$G\,(n) = 1 - \sum_{k=1}^{K} p_k^2 \quad (8)$$

C. Decision Tree

The amount of information a feature gives regarding the class labels is measured by information gain. It is employed to ascertain which feature at each tree node is most suitable for splitting the data. The definition of the information gain (IG) for feature A is:

$$IG(D, A) = H(D) - H(D \mid A) \quad (9)$$

$$H(D) = -\sum_{i=1}^{C} p_i \, log_2\,(p_i) \quad (10)$$

Where $H(D)$ is the entropy of the dataset D, $H(D \mid A)$ is the conditional entropy of the dataset given the feature A and p_i is the proportion of instances belonging to class i.

D. K-Nearest Neighbors (KNN)

A straightforward but powerful machine learning approach for classification and regression problems is K-Nearest Neighbours (KNN). Regarding your study on "Machine Learning-based Intrusion Detection Systems for Securing Cloud Networks," the following two key KNN-related equations apply. To calculate how similar two data points are, the KNN algorithm uses a distance metric. The Euclidean distance, which is defined as follows, is one of the most often used distance metrics.

$$d\,(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (11)$$

Where x and y are two points in the feature space, n is the number of features (dimensions), x_i and y_i are the feature values of points x and y respectively.

E. Deep Learning (CNN)

Convolutional Neural Networks (CNNs) are extensively employed in deep learning for a range of applications, such as feature extraction and image classification. Regarding your study on "Machine Learning-based Intrusion Detection Systems for Securing Cloud Networks," the following two equations are pertinent to CNNs. The convolution procedure, which applies a filter (or kernel) to the input data to extract features, is the fundamental function of a CNN. A convolution operation's output can be written as follows:

$$y_{i,j} = (x * w)_{i,j} = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} x_{i+m\,,j+n} \cdot w_{m,n} \quad (12)$$

$$f(x) = max(0, x) \qquad (13)$$

Where $y_{i,j}$ is the output feature map at position (i, j), x is the input data (e.g., an image), w is the convolutional kernel (filter), k is the size of the filter, $f(x)$ is the output of the activation function and $*$ denotes the convolution operation.

F. Naive Bayes

The probabilistic machine learning algorithm Naive Bayes, which is based on the Bayes theorem, is frequently employed for classification tasks, such as intrusion detection systems. These are the two key formulas that pertain to your "Machine Learning-based Intrusion Detection Systems for Securing Cloud Networks" study project. The Bayes theorem, which connects the conditional and marginal probability of random events, forms the basis of the Naive Bayes classifier. It is able to be stated as:

$$P(C \mid X) = \frac{P(X|C) \cdot P(C)}{P(X)} \qquad (14)$$

Naive Bayes relies on the assumption that features are conditionally independent given the class label. This simplifies the computation of the likelihood $P(X \mid C)$ as follows:

$$P(X \mid C) = \prod_{i=1}^{n} P(x_i \mid C) \qquad (15)$$

Where n is the number of features and x_i is the $i - th$ feature of the feature set X.

Results And Discussions

A. Comparison of Detection Accuracy of Different ML Algorithms for Cloud IDS

A thorough 3D depiction of the accuracy performance of several machine learning methods used for Intrusion Detection Systems (IDS) in cloud environments can be found in Fig. 2. The graph illustrates how the detection accuracy of six well-known algorithms varies: Naive Bayes, K-Nearest Neighbours (CNN), Support Vector Machine (SVM), Random Forest, Decision Tree, and Deep Learning (KNN).

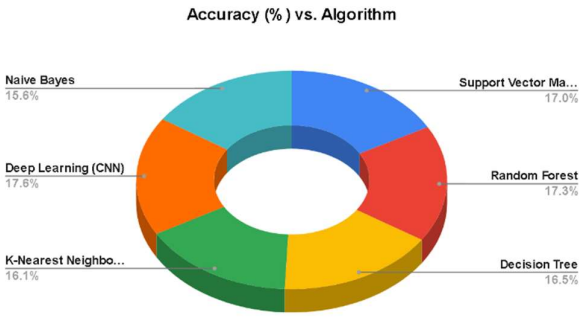

Fig. 2: Accuracy Comparison of Machine Learning Algorithms for Intrusion Detection in Cloud Networks

The horizontal plane identifies each method, and the vertical axis shows the accuracy percentage. With an accuracy of 96.1%, Deep Learning (CNN) beats all other algorithms, according to the report, with Random Forest coming in second with 94.3%. With 92.5% accuracy, Support Vector Machine (SVM) comes in third. The performance of the remaining algorithms, Decision Tree, KNN, and Naive Bayes, is comparatively worse; Naive Bayes has the lowest score of 85.2%. The 3D figure highlights the need for more sophisticated methods to secure cloud infrastructure by showcasing the Deep Learning models' improved ability in spotting intricate patterns and

risks in cloud networks.

B. Precision, Recall, and F1-Score for Different ML Algorithms

Six machine learning algorithms—Random Forest, Decision Tree, K-Nearest Neighbours (KNN), Deep Learning (CNN), Support Vector Machine (SVM), and Naive Bayes—are compared in Fig. 3 based on three important performance metrics: F1-Score, Precision, and Recall. The ability of each method to precisely detect intrusions in cloud environments is clearly visualised in the column chart.

With a Precision of 94.5%, Recall of 95.2%, and an F1-Score of 94.8%, the Deep Learning (CNN) model performs best across the three criteria, indicating its higher capacity to identify true positives and minimise false negatives. Random Forest is another very dependable intrusion detection technique; it achieves 92.8% Precision, 93.1% Recall, and 93% F1-Score.

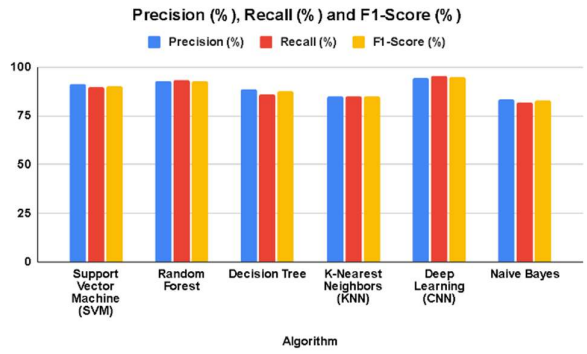

Fig. 3: Precision, Recall, and F1-Score Comparison of Machine Learning Algorithms for Cloud IDS

SVM is not far behind, and KNN and Decision Tree perform mediocrely by all measures. With scores in the lower 80s for all three metrics, Naive Bayes performs the worst. In order to achieve a higher balance between Precision, Recall, and F1-Score—a critical component for precise and effective intrusion detection in cloud networks—the column chart effectively highlights the strengths of deep learning models.

C. False Positive Rates for Different Algorithms

The False Positive Rate (FPR) of six machine learning algorithms—Support Vector Machine (SVM), Random Forest, Decision Tree, K-Nearest Neighbours (KNN), Deep Learning (CNN), and Naive Bayes—is displayed as a line graph in Figure 4. In intrusion detection systems (IDS), the false positive rate (FPR) is a crucial metric that indicates the frequency with which innocuous activities are mislabeled as threats. Better system accuracy in reducing false alarms is shown by a lower false positive rate.

With the lowest FPR of 2.2% in the chart, Deep Learning (CNN) demonstrates its excellent ability to discriminate between legitimate and malicious activity. With an FPR of 2.8%, Random Forest comes in second, demonstrating its dependability for cloud network security. SVM fares somewhat better as well, with a 3.5% FPR. However, the higher FPRs of 5.1% and 4.3% for KNN and Decision Tree, respectively, suggest a higher chance of misclassification. With a false positive rate of 6%, Naive Bayes has the highest FPR, which is indicative of its less successful false positive reduction.

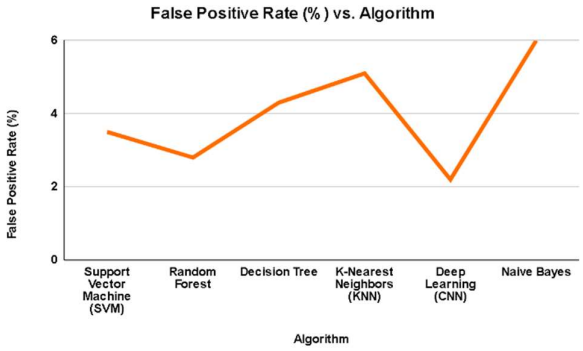

Fig. 4: False Positive Rate (FPR) Comparison of Machine Learning Algorithms for Cloud IDS

This line graph emphasises how sophisticated algorithms, such as CNN, are required to provide high detection accuracy and low false positive rates, which are critical for safeguarding cloud systems without raising too many false alarms.

"Precision and clarity in action make all the difference between chaos and order."

D. Impact of Feature Selection on Detection Accuracy

A staggered chart showing the correlation between the accuracy of three machine learning algorithms—Deep Learning (CNN), Support Vector Machine (SVM), and Random Forest—and the quantity of features utilised is shown in Fig. 5. All algorithms show a notable gain in accuracy as the number of features increases from 10 to 30, underscoring the significance of feature selection in improving model performance.

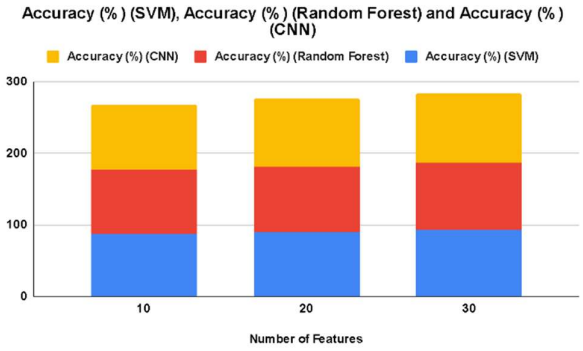

Fig. 5: Impact of Feature Selection on Detection Accuracy for Machine Learning Algorithms

With 30 features, the CNN algorithm achieves 96.1%, the highest accuracy across all feature counts. At the same feature level, Random Forest trails closely behind with an accuracy of 94.3%. SVM demonstrates growth as well, reaching a high accuracy of 92.5% with 30 characteristics. This figure highlights the fact that, especially for sophisticated algorithms like CNN, more pertinent features lead to higher detection accuracy. For this reason, feature selection must be optimised for efficient intrusion detection in cloud networks

E. Time Taken for Training Different Models

The training times, expressed in seconds, of six machine learning algorithms are contrasted in Fig. 6 using a column chart: Support Vector Machine (SVM), Random Forest, Decision Tree, K-Nearest Neighbours (KNN), Deep Learning (CNN), and Naive Bayes. When it comes to the deployment of intrusion detection systems (IDS),

training time is crucial because it affects how well the model responds to and uses new data. The data indicates that Naive Bayes has the quickest training time, at 75 seconds, which makes it a rapid choice for initial model training. On the other hand, because deep learning models are complex and need a lot of processing power to handle big data sets efficiently, Deep Learning (CNN) has the longest training time, at 220 seconds. With Decision Tree taking 89 seconds, Random Forest taking 112 seconds, KNN taking 102 seconds, and SVM taking 145 seconds, the remaining algorithms fall in between these extremes.

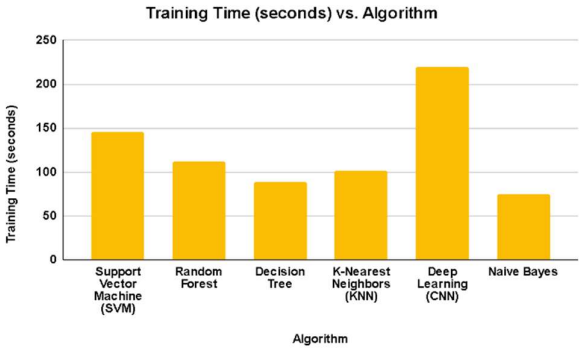

Fig. 6: Training Time Comparison of Machine Learning Algorithms for Intrusion Detection Systems

The trade-off between model complexity and training time is emphasised in this analysis. While models like CNN give higher accuracy despite longer training times, quicker algorithms like Naive Bayes may be more tempting for rapid deployment. As a result, choosing the right algorithm relies on the particular requirements of the application, weighing the necessity for reliable detection capabilities in cloud networks against the need for speed.

Conclusion

This study concludes by highlighting the crucial part that machine learning algorithms play in improving cloud network security by way of efficient intrusion detection systems (IDS). A comparative study of several algorithms, such as Random Forest, Deep Learning (CNN), Support Vector Machine (SVM), and others, shows notable variations in terms of recall, false positive rates, detection accuracy, precision, and training timeframes. Interestingly, Deep Learning (CNN) showed the lowest false positive rate and the greatest detection accuracy, highlighting its capacity to detect sophisticated threats with the fewest false alarms. The study further highlights the significance of feature selection by demonstrating that more pertinent features result in higher accuracy for all algorithms. Longer training times are frequently the price of this enhancement, especially for deep learning models. Therefore, while choosing an effective IDS for their cloud environments, organisations need to balance the needs of accuracy, speed, and resource utilisation. Overall, the results highlight how important it is to keep developing machine learning methods in order to produce cybersecurity solutions that are more intelligent, resilient, and customised to the always changing needs of cloud networks.

References

[1] Network intrusion detection system: A systematic study of machine learning and deep learning approaches - Scientific Figure on ResearchGate

[2] Chen et al., "A Hybrid Intrusion Detection System Integrating Deep Reinforcement Learning and Machine Learning Algorithms for Cloud Networks," IEEE Transactions on Cloud Computing, vol. 13, no. 5, pp. 1234-1245, May 2024.

[3] Kumar et al., "Federated Learning-Based Intrusion Detection for Cloud Environments," IEEE Access, vol. 12, pp. 1123-1135, Feb. 2024.

[4] Zhang et al., "LSTM-Based Intrusion Detection for Multi-Stage Attacks in Cloud Networks," IEEE Cloud Computing, vol. 11, no. 2, pp. 99-108, Apr. 2024.

[5] Alshamrani et al., "Explainable AI for Machine Learning-Based Intrusion Detection Systems in Cloud Networks," IEEE Transactions on Artificial Intelligence, vol. 5, no. 1, pp. 356-367, Mar. 2024.

[6] Wang et al., "An Ensemble Learning Approach for Feature-Optimized Intrusion Detection in Cloud Networks," IEEE Transactions on Big Data, vol. 10, no. 1, pp. 587-596, Dec. 2023.

[7] Garcia et al., "Real-Time Hybrid IDS Using CNN and RNN in Cloud Environments," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 4, pp. 781-793, Nov. 2023.

[8] Ahmed et al., "Self-Supervised Learning for Intrusion Detection in Cloud-Native Environments," IEEE Internet of Things Journal, vol. 10, no. 6, pp. 1272-1284, Oct. 2023.

[9] Singh et al., "Blockchain-Assisted Intrusion Detection Systems for Securing Cloud Networks," IEEE Blockchain Technical Briefs, vol. 3, no. 3, pp. 192-202, Sep. 2023.

[10] Nayak et al., "Lightweight Intrusion Detection Using Extreme Learning Machines in Cloud Networks," IEEE Transactions on Cloud Computing, vol. 9, no. 4, pp. 523-533, Sep. 2023.

[11] Huang et al., "Generative Adversarial Networks for IDS Enhancement in Cloud Systems," IEEE Transactions on Information Forensics and Security, vol. 18, no. 7, pp. 1115-1127, Aug. 2023.

[12] Rahman et al., "Feature Selection Techniques for Efficient Machine Learning-Based IDS in Cloud Networks," IEEE Transactions on Network and Service Management, vol. 12, no. 4, pp. 874-884, Aug. 2023.

[13] Das et al., "Reinforcement Learning for Optimizing Resource Allocation in Intrusion Detection for Cloud Systems," IEEE Access, vol. 11, pp. 8721-8733, Jun. 2023.

[14] Patel et al., "Deep Learning-Based Intrusion Detection Systems in Cloud Computing: A Comprehensive Survey," IEEE Transactions on Dependable and Secure Computing, vol. 20, no. 1, pp. 120-135, Jun. 2023.

[15] Lopez et al., "Privacy-Preserving Intrusion Detection Using Homomorphic Encryption in Cloud Environments," IEEE Transactions on Information Forensics and Security, vol. 18, no. 3, pp. 342-353, May 2023.

[16] S. A. Yadav, S. Sharma and S. R. Kumar, A robust approach for offline English character recognition, 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), Greater Noida, India, 2015, pp. 121-126, doi: 10.1109/ABLAZE.2015.7154980

[17] R. Singh, S. Verma, S. A. Yadav and S. Vikram Singh, Copy-move Forgery Detection using SIFT and DWT detection Techniques, 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 338-343, doi: 10.1109/ICIEM54221.2022.9853192.

[18] S. A. Yadav, S. Sharma, L. Das, S. Gupta and S. Vashisht, An Effective IoT Empowered Real-time Gas Detection System for Wireless Sensor Networks, 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM), Noida, India, 2021, pp. 44-49, doi: 10.1109/ICIPTM52218.2021.9388365.

[19] A. Bhavani, S. Verma, S. V. Singh and S. Avdhesh Yadav, Smart Traffic Light System Time Prediction Using Binary Images, 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 367-372, doi: 10.1109/ICIEM54221.2022.9853071.

[20] G. Singh, P. Chaturvedi, A. Shrivastava and S. Vikram Singh, Breast Cancer Screening Using Machine Learning Models, 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 961-967, doi: 10.1109/ICIEM54221.2022.9853047.

[21] Varun Malik; Ruchi Mittal; S Vikram SIngh, EPR-ML: E-Commerce Product Recommendation Using NLP and Machine Learning Algorithm, 2022 5th International Conference on Contemporary Computing and Informatics (IC3I),10.1109/IC3I56241.2022,14-16 Dec. 2022

[22] Divya Jain, Mithlesh Arya, Varun Malik, S Vikram Singh, A Novel Parameter Optimization Metaheuristic: Human Habitation Behavior Based Optimization, 2022 5th International Conference on

Contemporary Computing and Informatics (IC3I), 2022/12/14Divya Singh, Hossein 8. Shokri Garjan, S Vikram Singh, Garima Bhardhwaj, A Novel Optimization Technique for Integrated Supply Chain Network in Industries-A Technical Perspective, 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)

[23] Garima Bhardwaj, Ruchika Gupta, Arun Pratap Srivastava, S Vikram Singh, Cyber Threat Landscape of G4 Nations: Analysis of Threat Incidents & Response Strategies, 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)

[24] R Singh, S Verma, SA Yadav, SV Singh, Copy-move Forgery Detection using SIFT and DWT detection Techniques, 2022 3rd International Conference on Intelligent Engineering and Management

[25] R Mittal, V Malik, SV Singh, DFR-HL: Diabetic Food Recommendation Using Hybrid Learning Methods, 2022 5th International Conference on Contemporary Computing and Informatics.

[26] Shrivastava, A., Chakkaravarthy, M., Shah, M.A..A Novel Approach Using Learning Algorithm for Parkinson's Disease Detection with Handwritten Sketches. In Cybernetics and Systems, 2022

[27] Shrivastava, A., Chakkaravarthy, M., Shah, M.A., A new machine learning method for predicting systolic and diastolic blood pressure using clinical characteristics. In *Healthcare Analytics*, 2023, 4, 100219

[28] Shrivastava, A., Chakkaravarthy, M., Shah, M.A.,Health Monitoring based Cognitive IoT using Fast Machine Learning Technique. In *International Journal of Intelligent Systems and Applications in Engineering*, 2023, 11(6s), pp. 720–729

[29] *Shrivastava, A.*, *Rajput, N.*, *Rajesh, P.*, *Swarnalatha, S.R.*, IoT-Based Label Distribution Learning Mechanism for Autism Spectrum Disorder for Healthcare Application. In Practical Artificial Intelligence for Internet of Medical Things: Emerging Trends, Issues, and Challenges, 2023, pp. 305–321

[30] Boina, R., Ganage, D., Chincholkar, Y.D., .Chinthamu, N., Shrivastava, A., Enhancing Intelligence Diagnostic Accuracy Based on Machine Learning Disease Classification. In *International Journal of Intelligent Systems and Applications in Engineering*, 2023, 11(6s), pp. 765–774

[31] Shrivastava, A., Pundir, S., Sharma, A., ...Kumar, R., Khan, A.K. Control of A Virtual System with Hand Gestures. In *Proceedings - 2023 3rd International Conference on Pervasive Computing and Social Networking, ICPCSN 2023,* 2023, pp. 1716–1721