Dynamic intrusion detection system for improved cloud security

MSc Practicum Part 2
MSc in Cybersecurity

Anusha Palakkattu East Madom Ramadas
Student ID: 23124903

School of Computing
National College of Ireland

Supervisor: Vikas Sahni

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Anusha Palakkattu East Madom Ramadas

Student ID: 23124903

Programme: Master of Science in Cybersecurity **Year:** 2024

Module: MSc Practicum part 2

Supervisor: Vikas Sahni

Submission Due Date: 12/12/2024

Project Title: Dynamic intrusion detection system for improved cloud security

Word Count: 5989 **Page Count:** 20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Anusha Palakkattu East Madom Ramadas

Date: 10/12/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Dynamic intrusion detection system for improved cloud security

Anusha Palakkattu East Madom Ramadas
23124903

Abstract

The adoption of cloud computing is accelerating and securing the cloud environments against advancing cyber threats has become a necessity. Traditional intrusion detection systems (IDS) lack the capability of real time detection and have delayed response time. A machine learning (ML) based IDS which detect unknown attacks can help organisations to identify evolving cyber-attacks. This work investigates the use of extreme learning machine (ELM) algorithm in enhancing cloud-based IDS. Dataset leveraged for this work is CSE-CIS-CID2018. The performance of ELM model was evaluated and compared with other ML models like random forest (RF), decision tree (DT), Naive Bayes (NB), Artificial Neural Networks (ANN), and Deep Neural Networks (DNN). The results revealed that ELM model achieved an accuracy of 96.42%, whereas RF achieved an accuracy of 97.12%. ELM consumed less training time but consumed more time to predict compared to other ML models. The key findings from this project is that ML based IDS can enhance cloud security and while ELM may be efficient in certain scenarios, RF can be useful in another set of scenarios.

1 Introduction

Cloud computing is a technology that provides on-demand resources for computing, in a shared pool, that is scalable and available. By the definition provided by National Institute of Standards and Technology (NIST), cloud computing is a model that may enable convenient, ubiquitous and on-demand network access to shared pool of configurable computing resources, such as servers and storage, which can be easily provisioned and released with minimal management efforts (Mell & Grance, 2011). Cloud computing service models consists of Software as a Service (SaaS), Platform as Service (PaaS) and Infrastructure as a Service (IaaS). There are different cloud deployment models such as public, private, hybrid and community cloud.

As cloud helps organisations to broaden their networks, scale projects and pool resources, operate with optimal resource usages, and other increased advantages, many companies have adopted cloud. Moreover, pandemic has accelerated the adoption of cloud, including hybrid work environment opportunities and low maintainability cost. However, increase in demand of cloud has made it a target for cyber-attacks because the cloud resources are available through the internet and internet is always vulnerable. The resources accessed through cloud infrastructure can have security threats that can occur to local computing resources like, malware and ransomware attacks. Additionally, they can also have cyber threats through cloud, such as secure storage issues. From a survey conducted by (Gupta, Laxmy and Sharma, 2014), cloud security issues can include, but not limited to, multi-tenancy, insider/outsider attacks and network security attacks. Multi-tenancy can cause issue of confidentiality due to the data sharing between cloud platforms and between the customers. Network security threats including DDoS, man in the middle attack, and malware injections can also occur in a cloud environment. In a recent data breach at a cloud storage company Snowflake, datasets stored in the servers of their cloud platform were compromised. This attack led to data breaches at companies like Santard and Ticketmaster (Whittaker, 2024).

Cloud computing is being used widely since past few years, but all the devices linked to the cloud do not have high security. There are IoT devices, that are vulnerable to attacks and may be used as bots for botnet attacks for disrupting the cloud servers. Usage of a good intrusion detection system (IDS) can help detect malicious behaviours and activities in the network. To mitigate against potential cyber-attacks in a cloud infrastructure, organisations should employ mitigation strategies like intrusion detection systems (IDS). An IDS can be helpful for detecting abnormal behaviours or malicious activities by monitoring networks and systems. Implementing IDS in cloud environment is a necessity, as it can protect from outsider and insider threats, inspect inbound and outbound traffic, and prevent unauthorised access to computing resources. Setting and IDS for both on-premises and cloud infrastructures of an organisation will contribute to the good security posture of the organisation, as it will alert the security team of any unusual behaviour, and the incident response team can perform accordingly, to stop the attack.

There are different types of intrusion detection systems such as signature-based and anomaly-based. In a signature-based intrusion detection system, known attack patterns, known as signatures, are stored in a database and the current system or network activity is compared with the patterns available. This type of IDS helps in mitigating against attacks whose patterns are already known. Whereas in anomaly-based IDS, it detects anomalous behaviours, that deflect from standard behaviour. This makes it capable of detecting both known and unknown attacks. It can even detect zero-day attacks and internal malicious attacks. With the evolution of technology, the method of cyberattacks also evolves. Combining the capabilities of ML with IDS can help in detecting evolving cyber threat patterns in cloud environments.

A comprehensive review of academic research papers, case studies and industry reports were conducted to understand the integration of IDS in cloud infrastructure. Reviewing and analysing various academic articles, eBooks, journals and conference papers provided insights of cyber threats in the cloud environment and the need for an IDS. Incorporation of ML to IDS can ensure mitigation from known and unknown cyber threats. The research highlighted that even though many conventional ML methods such as random forest (RF), decision tree (DT), and support vector machine (SVM) has been utilised to build IDS, extreme machine learning algorithm (ELM) has not been leveraged for developing IDS for securing cloud environments.

Even though, ML can provide dynamic intrusion detection capability, and enhanced cloud security posture for an organisation, training the model can consume computing resources. Additionally, choosing of dataset can be challenging as biased and outdated dataset can affect the results predicted by the IDS model.

1.1 Research question

The identified research question is: can using Extreme Learning Machine (ELM) algorithm to build an IDS for cloud environments improve the detection speed and accuracy of the IDS as compared to ML algorithms such as DT, RF, NB, ANN and DNN?

1.2 Report structure

The structure of the report is as follows:

- Literature review (Section 2): A comprehensive review of different conference papers, academic journals and articles was performed to understand the cyber threats occurring in the cloud environment, current intrusion detection systems employed, and their merits and demerits. Additionally, emerging dynamic intrusion detection systems were analysed and the advantages and disadvantages of the methodologies leveraged were identified.
- Research Methodology (Section 3): A detailed description of proposed methodology.
- Design Specification (Section 4): A detailed description of design specification of proposed model and process.
- Implementation (Section 5): A detailed explanation of implementation process of the ML models.
- Evaluation (Section 6): A detailed evaluation of the results obtained.
- Conclusion and future work (Section 7): The conclusion drawn and proposed future work.

2 Related Work

A comprehensive review of various conference papers, academic articles and journals was conducted to gain insights on the cyber threats on cloud infrastructure and to develop the project. Additionally, industry reports and case studies were also reviewed to learn more about recent cyber-attacks on cloud, their impacts on individuals and organisations, and mitigation strategies adopted by various individuals and industries. The following is the critical analysis of the previous related works conducted.

2.1 Analysis of cyber threats in cloud infrastructure

In recent years, as the usage of cloud increased, the risk of cyber-attacks in cloud environments also increased. Chauhan and Shiaeles (2023), conducted an analysis of security issues of cloud environment, and it was found out that data breaches, lack of secure storage, unauthorised access, insider attacks and unsecured APIs are few major security issues. To mitigate issues, cloud users should implement intrusion detection systems, that can monitor and analyse user and network activities of the cloud infrastructure, and alert when any unusual behaviour is detected.

Muhammad et al. (2023), performed an analysis of security vulnerabilities in cloud environments including public, private, hybrid and community clouds. The authors identified cyber threats including data breaches, authentication, privacy preservability, data access controllability, cloud based IoT application vulnerabilities, phishing, and key exposure. They also proposed countermeasures specifically for each cloud models. Hamed and Marjan (2020), conducted a survey on security challenges that can occur in cloud environments which included insecure APIs, service hijacking and data breaches, and categorised the cyber threats into infrastructure related, network related, and data related. The authors proposed countermeasures such as access control, encryption and multi-factor authentications.

In a study conducted by Gupta and Vashisth (2023), it was found that various cyberattacks can occur in cloud including account hacking, API hijacking, service abuse, trojan attacks, DDoS attacks and credential tampering. Authors also discussed about various case studies of cyber-attacks on organisations such as Amazon, Apple and Google. Along with setting strong passwords, multi-factor authentication and encryptions, intrusion detection systems should be implemented across organisations to mitigate attacks that can occur in cloud infrastructure. These highlights the requirement for integrating an IDS to monitor and analyse activities in the cloud environment.

2.2 Analysis of securing cloud infrastructure using intrusion detection systems

Lata and Singh (2022) conducted an analysis of IDS for cloud security by reviewing 43 articles and concluded that implementation of IDS for cloud will enhance security. However, the paper shows that there is existing scope for enhancing the security techniques as the threats in cloud also advances. In a review conducted by Razdan, Gupta and Seth (2021), performance analysis of network intrusion detection systems based on detection rate, false positive rate and accuracy, were performed. It depicts that the IDS using ANN based hybrid approach shows better performance. Qi et al. (2023) performed an investigation and analysis on IDS in cloud environments and summarised the approaches to build IDS for cloud security. The authors concluded that utilising AI could improve intrusion detection. The evaluation included performing a comparison of advantages and disadvantages of available datasets.

In a review of IDS in cloud environments, performed by Prabhakaran and Kulandasamy (2021), the IDS are classified into Host-based (HIDS), Network-based (NIDS), Distributed (DIDS), and Hypervisor-based IDS, and the effect of each is discussed. Different novel IDS techniques such as signature-based, anomaly-based and hybrid, and advanced techniques like genetic algorithms, fuzzy logic and neural networks were analysed. The paper gave insights about the importance of incorporating IDS to improve cloud security of an organisation and necessity of continuous innovations for continuously evolving cyber threats.

Chang, et al. (2022) conducted a survey of intrusion detection systems on various cloud computing technologies utilised in enterprise environment. The paper throws light on the need for good incident response plans along with good IDS in organisations. To setup a suitable IDS for an organisation, the company should conduct risk assessments and identify the resources to protect from cyber-attacks. Authors are also concerned about high false positive rate of IDS, as it might trigger alerts and countermeasures that are not required and can even disrupt business operations. Additionally, attackers can trigger false positive rates by exploiting known traffic patterns and flood alerts.

2.3 Analysis of securing cloud infrastructure using dynamic intrusion detection systems

The cloud infrastructure can be susceptible to attacks like distributed denial-of-service (DDoS). In the paper presented by Kiranmai, Vasantha and Jyoshna (2023), an anomaly detection is proposed by studying any abnormal behaviour in the network by sniffing the traffic using tools like Wireshark. In the proposed system, the authors built a system that dynamically process the network traffic data and creates association rules. These rules can be leveraged to detect abnormality in the cloud environment. This approach helps in dynamic detection of anomalies that can occur in cloud and reduces cost, resource and time consumption required for training a system.

Gill, Saxena and Sharma (2024), utilised a hybrid IDS that uses both signature-based and anomaly-based models for detections. They proposed a noncooperative game theoretic model that utilises NE based probabilistic monitoring strategy. The advantage of this system is that it is suitable for detecting attacks whose signature is already known and is in the database, and any new attack that is different from the pattern stored in the database. The disadvantage is that it requires good amount of computing resources, time and cost, as the minimum requirement of the hardware to employ the proposed system is 8GB RAM with an i5 processor, Python 3 and RStudio. This requirement may increase on large scale enterprise cloud infrastructures. For anomaly detection of cloud infrastructure, Sanagana and Tummalachervu (2024), proposed a deep learning-based IDS, which leverages LSTM classification model and Adam optimiser for training with effective convergence and optimisation. This provided accurate detections with false positive rates.

Sowmya and Mary (2023) analysed AI based methods to detect cyber-attacks. This showed that ML, DL and ensemble-based methods provided good performance in attack detection. A survey on DL approaches for IDS was conducted by Ferrag, Marglaras, Moschoyiannis and Janicke. Additionally, authors also classified 35 datasets for the analysis. 7 DL models were analysed using CSE-CIC-IDS2018 and the Bot-IoT real traffic datasets with performance indicators, namely, false alarm rate, accuracy, and detection rate. An IDS for cloud security was build using random forest (RF) ML method using Bot-IoT and NSL-KDD datasets and accuracy of 98.3% and 99.99% was achieved respectively (Attou et al., 2023). Authors implemented feature reduction by reducing features to 2, demonstrating the potential of using a smaller number of features by contrasting the results with those of other classifiers.

In another research, Almusallam et al. (2023) used Radial Basis Function Neural Network (RBFNN) and RF ML models for building IDS in cloud computing environment. RF classifier was used for feature selection, and the RBFNN algorithm was used to detect intrusion. They used Bot-IoT and NSL-KDD datasets and achieved an accuracy higher than 93%. Ibrahimi, Jouhari and Jakout (2024) evaluated the efficiency of ML algorithms in improving IDS for new complex and dynamic threats. They utilised RF, DT, naive bayes (NB), and gradient boost (GB) with CSE-CIC-IDS2018 dataset. RF, DT and GB demonstrated a high-performance matrix close to 99% but NB delivered lower accuracy.

In the paper published by Tuan-Hong and Iftekar (2023), long term performance evaluation of six ML models, namely, RF, DT, NB, support vector machine (SVM), artificial neural network (ANN), and deep neural network (DNN) were conducted with two datasets, CIC dataset and the LUFlow dataset. They used separate datasets for training and testing. The study concluded that ANN has highest performance compared to others and additionally, identified overfitting risks. Tut and Mohsenabad conducted a study using bio-inspired optimisation ML algorithms for improving IDS such as Artificial Bee Colony (ABC), Flower Pollination Algorithm (FPA), and Ant Colony Optimization (ACO). They utilised CSE-CIC-IDS2018 dataset. The models demonstrated detection accuracies of 99% with ACO, 98.7% with FPA, and 98.6% with ABC, with minimal model-building time.

Tiwari and Jain (2022) published a firewall mechanism utilising ML for traffic classification to enhance IDS by combining decisions of past node with current outputs of the algorithm. They used UNSW-NB-15 dataset. The system demonstrated a detection accuracy of 97.68%. Authors proposed a hybrid ML approach utilising Convolutional Neural Networks (CNN), SVM, and K-Nearest Neighbours (KNN) for improving performance and detection of IDS (Shukla and Sharma, 2023). Dataset used was CIC-IDS-2017 and NSL-KDD. It showed that CNN with Long Short-Term Memory (LSTM) increased classification accuracy from 95% to up to 97.68%.

Vinolia, Kanya and Rajavarman (2023) reviewed the application of ML and DL for enhancing IDS in cloud environments. They explored methods such as data mining and soft computing. The results demonstrated that DL models have higher accuracy of 99.95% and offer better generalisation than ML models. Babu et al. (2023), proposed an IDS using deep neural networks (DNN) in cloud environments. Network Intrusion dataset was used for training and evaluating. DNN model was evaluated using performance metrics, such as accuracy, true positive rate, precision, false positive rate, and F1-score, and demonstrated high results.

Reji, Joseph, Nancy, and Lourdes (2023) proposed a hybrid ML model using Seagull Optimization Algorithm (SOA) for feature selection with Extreme Learning Machine (ELM) classifier to classify attacks in Internet of Things (IoT) networks. CIC-IDS-2018 dataset. The model SOA-ELM achieved 94.22% accuracy, detection rate of 93.45% and an F1-score of 91.26%. Authors proposed an IDS for cloud environments using ensemble model weighted by the Crow Search Algorithm (CSA) (Bakro et al., 2023). The classifier combines ML and DL model which includes, LSTM, SVM, XGBoost, and Fast Learning Network (FLN), with experiments on the NSL-KDD, Kyoto, and CSE-CIC-IDS-2018 datasets. The ensemble method demonstrated higher performance than traditional ML models. However, combining multiple models can be challenging in cloud environments with resource constraints.

2.4 Identified research gap

Based on the research through these various related works, it was found that though various research works have been published on intrusion detection systems based on ML models, most of them uses the traditional ML models like RF, SVM and KNN. Due to these reflections, it was concluded that the approach of ELM has not been utilised yet for building the IDS for cloud environments. The ELM has fast learning capabilities and other features which are easy to implement, thus making it suitable for intrusion detections. Below is a detailed description of the work carried out.

3 Research Methodology

After comprehensive research of related works, it was found that ELM has not been used to analyse ML based IDS employed in cloud environments. Additionally, ELM has the capacity to handle high dimensional data with high speed and efficiency. The research was conducted in a structured manner to evaluate performance of ELM compared to traditional ML models such as RF, DT, and ANN. Data cleaning and preprocessing was performed on the dataset to handle outliers and missing values. Dataset was divided into training and testing subsets in the ratio of seventy and thirty respectively, to improve performance and better evaluation of the machine learning models. Hyperparameter tuning was performed to improve performance of the models. The evaluation of the models was performed using the calculation of accuracy, precision, recall, F1-score. Confusion matrix was plotted to understand the model performance. A 5-fold cross validation was conducted for accuracy evaluation of the models.

3.1 Dataset description

To conduct the experiment, CSE-CIC-IDS2018[1] dataset was chosen as it has real time network traffic and different attack types. This dataset comprises of various cyberattack scenarios including DoS, DDoS, botnet, brute-force, web attacks, network infiltration and heartbleed. The dataset contains 80 extracted features of the real time network traffic captured from 420 machines and 30 servers.

3.2 Data preprocessing

It is important to clean and process the data provided in the dataset as it may contain noises. The format of CSE-CIC-IDS2018 dataset was analysed to understand the essence of the attributes and standard deviation was analysed to understand the dataset statistically. This helped to preprocess the data efficiently. On analysis, it was found that the CSE-CIC-IDS2018 dataset has class imbalances, where there were big portion of benign samples while some malicious samples like web attack contributed to very small portion. To reduce the class imbalance, the major classes were down sampled. Attack samples were combined so that dataset contained only two labels, namely benign and malicious. Dataset was cleaned by removing null or NaN (Not a Number) values, infinite values, missing values, outliers and duplicates. This aids in any problem that arise due to data inconsistencies. In one of the dataset files, there were 84 columns instead of 80 columns, therefore the 4 extra columns were removed to maintain consistency and performance.

3.3 Feature selection and classification

After data processing, feature selection was performed by implementing random forest classifier. The features were ranked by the importance and top 20 features were selected. After feature selection, the number of features were further reduced using brute force method, where

[1] https://registry.opendata.aws/cse-cic-ids2018/

ML models such as DT, RF, NB, ANN and DNN were trained using different number of features. The feature set that gave good accuracy with least number of features was chosen as final feature set.

Hyperparameter tuning for each model was performed using the final feature set. It is a process of choosing optimal set of hyperparameters for training the ML model so that the model achieves best performance. A five-fold cross validation was performed to endorse the accuracy and check overfitting of the models.

3.4 Evaluation

During evaluation, the results obtained from the ML models were analysed by employing several assessment criteria such as accuracy, precision, recall and F1-score. The models' predictions were assessed by test data subset, which had been isolated from training subset. The prediction efficiency of ELM model was compared with other ML models. Accuracy validation of the models was performed using 5-fold cross validation to prevent overfitting. The performance of the models was analysed using confusion matrix. The evaluation and comparison of the ELM model with other ML models was essential as it provided the understanding of abilities and drawbacks of the model. Additionally, it helped in determining whether the model has the acceptable levels of precision and robustness for intrusion detection.

4 Design Specification

To analyse the performance of ELM model-based IDS for improving cloud security, it was compared with ML models such as DT, RF, NB, ANN and DNN. The frameworks of the ELM and the assessment criteria utilised is described in this section. Figure 1 depicts the design architecture of the project.

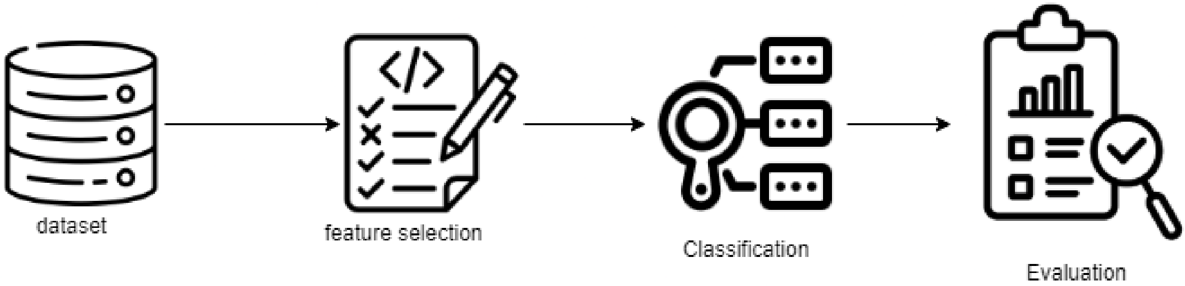


Figure 1: Design architecture.

4.1 Extreme learning machine

Extreme learning machine is a Single-Layer Feedforward Neural Network (SLFN), that has fast and efficient training capabilities. In this algorithm, input weights and hidden biases are randomly assigned. This will remain constant while training and testing. It has only one hidden layer and output weights are calculated by simple matrix operation rather than backpropagation. This aids in faster training process, which results in quick learning from large datasets accurately with limited computational resources.

4.2 Performance evaluation

Following are the assessment criteria utilised for assessing the model performance:

- **Accuracy:** It is the ratio of the proportion of correctly classified instances from the total number of instances. From checking accuracy of a model, we can draw the conclusion of the correctness of the classification model. However, solely relying on this evaluation method can mislead in case of imbalanced datasets, in which some classes may be less present or not present.
- **Precision:** It depicts the measure of proportion of the true positive instances from the total number of positive predictions. This performance measure becomes crucial in a setting where cost of false positives is potentially high.
- **Recall:** It is the measure of proportion of the true positive instances from the total number of actually classified positive instances. A model having high recall indicates that in detecting an intrusion in the network, most of the instances gets detected and a few benign traffic might be misclassified.
- **F1-Score:** It is the mean of precision and recall. As precision and recall aids in the understanding of the trade-offs between false positives and false negatives, the F1-score can be useful in class imbalances.
- **Confusion matrix:** It is a matrix that illustrates a detailed breakdown of result of classification. It depicts how many true positives, false positives, true negatives and false negatives were predicted by the model.

5 Implementation

A 64-bit Windows based system with 16 GB RAM was utilised to build the models. To implement this work jupyter notebook in the Anaconda, an open-source platform of version 2.6.0 was used. Python of version 3.12.4 was used to program the models. Various python libraries were used such as pandas, numpy, scikit-learn and matplotlib. To build ELM model scikit-elm library was used.

5.1 Data cleaning and preprocessing

To develop the project, CSE-CIC-IDS2018 dataset was initially analysed to understand the data formats and attributes. Additionally, statical analysis was performed by plotting standard deviation. The dataset contained null values, NaN values, infinite and missing values. It also contained outliers and duplicates. Therefore, dataset was cleaned by removing such values. In one of the dataset files, there were 84 columns present instead of 80 columns, hence the 4 extra columns were removed to maintain data consistency. During analysis of the dataset, it was understood there were class imbalances, in which benign samples were present in huge portions while some malicious samples such as web attack was only present in small proportion. To reduce the class imbalance, the major classes were down sampled. Attack samples were combined so that dataset contained only two labels, namely benign and malicious.

5.2 Feature selection

After data cleaning and preprocessing, feature selection was performed using RF classifier. The features were ranked by the importance score provided by the RF and top 20 features were selected. The feature importance rank is depicted in Figure 2.

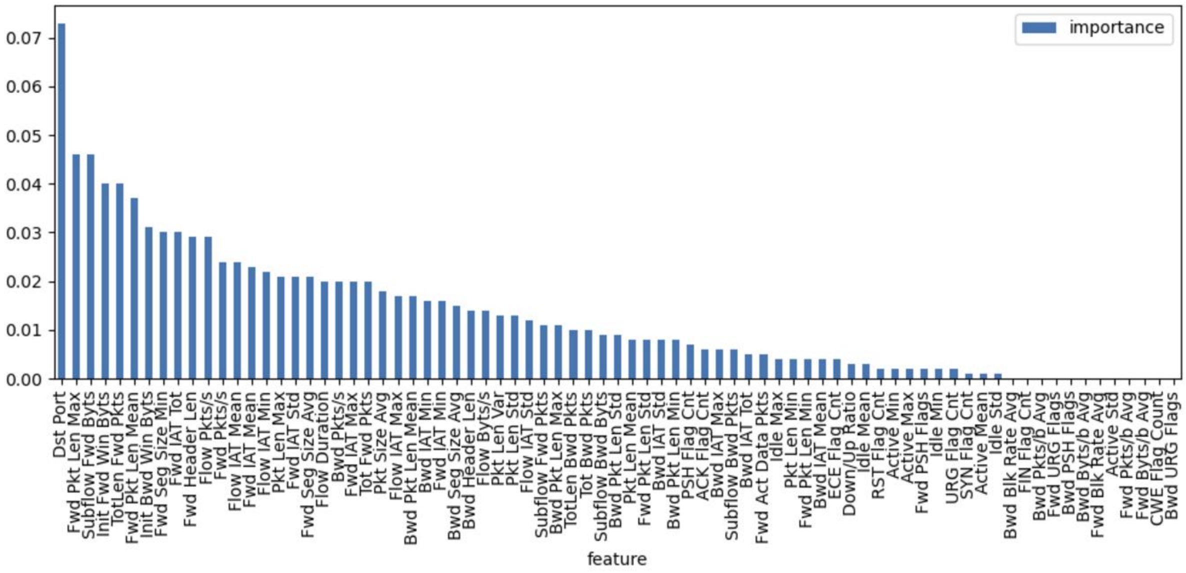


Figure 2: Feature importance rank by RF classifier.

After feature selection, the number of features were further reduced using brute force method, where ML models such as DT, RF, NB, ANN, DNN and ELM were trained using different number of features. The feature set that gave good accuracy with least number of features was chosen as final feature set. Figure 3 depicts the accuracy of the models with respect to features. Hyperparameter tuning for each model was performed using the final feature set.

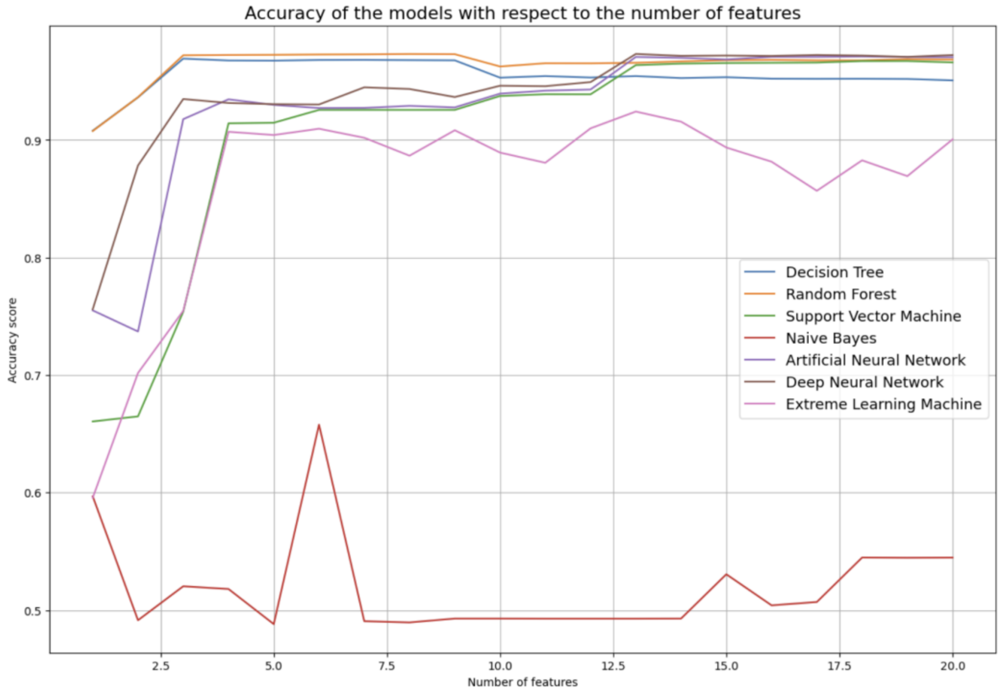


Figure 3: Accuracy of models with respect to number of features.

5.3 Model building

The main focus was to build an IDS model based on ELM algorithm. The dataset was split into 2 subsets, 70% for training and 30% for testing. The ELM model was optimised by hyperparameter tuning using GridSearchCV. Number of neurons, activation function and regularisation parameter were exhaustively searched in the parameter space. The optimal hyperparameters found were as follows:

- Number of Neurons (n_neurons): 7000
- Activation Function (ufunc): ReLU (Rectified Linear Unit)
- Regularization Parameter (alpha): 1e-05

A five-fold cross validation was performed on each model to validate the accuracy and the accuracies obtained for each model is illustrated in Figure 4.

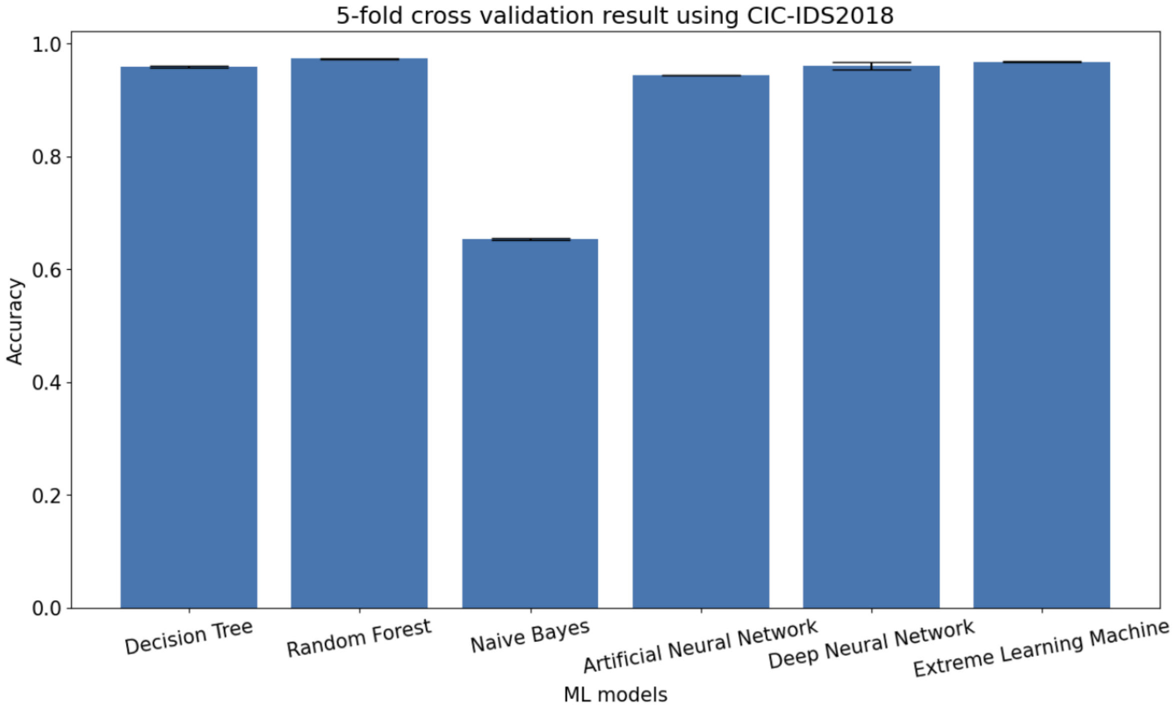


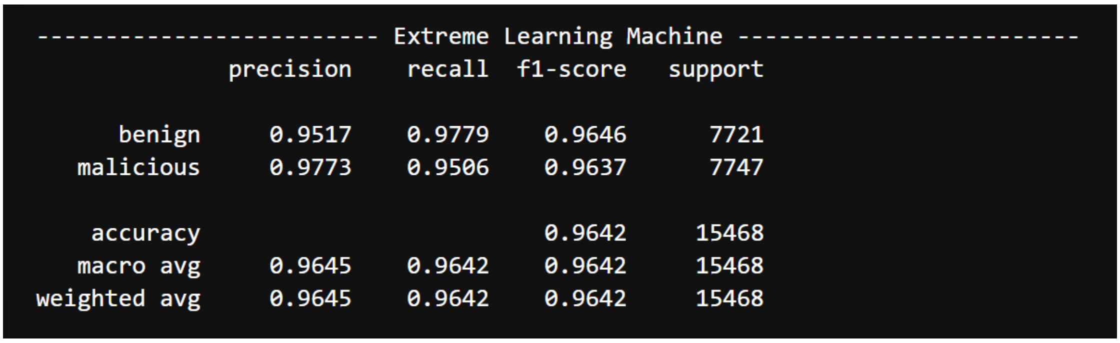
Figure 4: Five-fold cross validation results for all models.

6 Evaluation

Performance evaluation of ELM model for IDS in securing cloud environments utilising CSE-CIC-IDS2018 dataset was conducted and compared with other traditional ML models such as DT, RF, NB, ANN and DNN. Accuracy, precision, recall, F1-score and time efficiency of the models were considered for the evaluation.

6.1 Performance of ELM

The performance of ELM with respect to accuracy, precision, recall and f1-score is illustrated in Figure 5.



```
------------------------- Extreme Learning Machine -------------------------
              precision    recall  f1-score   support

      benign     0.9517    0.9779    0.9646      7721
   malicious     0.9773    0.9506    0.9637      7747

    accuracy                         0.9642     15468
   macro avg     0.9645    0.9642    0.9642     15468
weighted avg     0.9645    0.9642    0.9642     15468
```

Figure 5: Evaluation result of ELM model.

6.2 Performance of other ML models

The performance of DT, RF, NB, ANN and DNN was compared to ELM. The confusion matrix plotted is illustrated in following Figure 6.

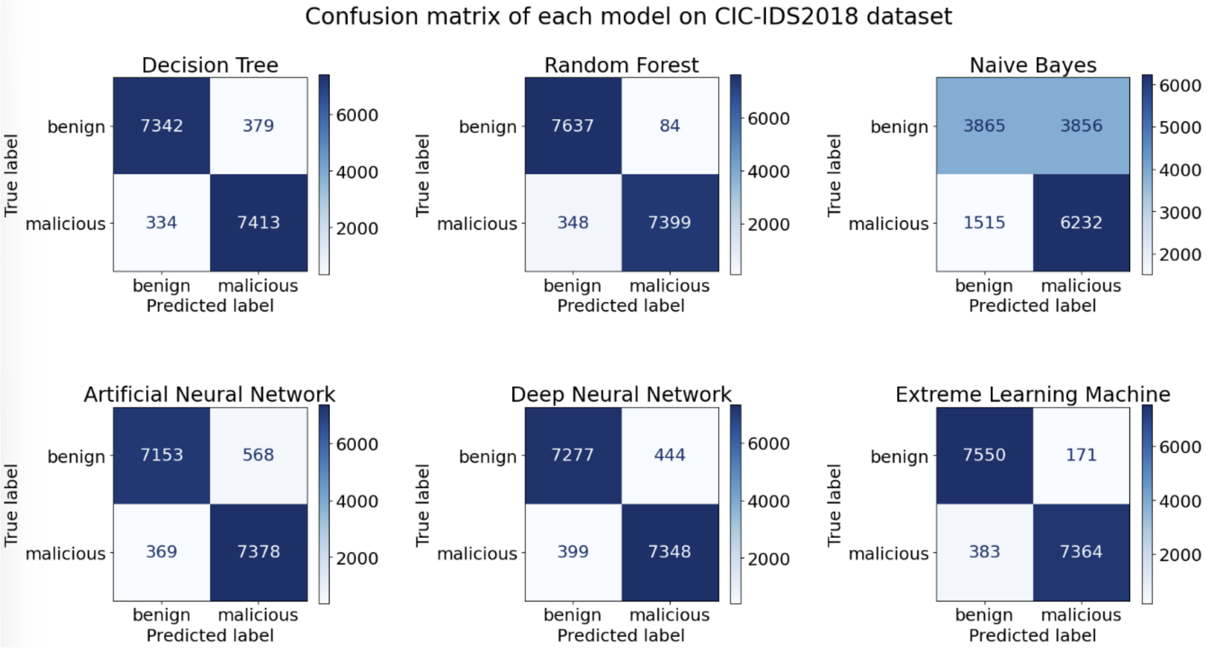


Figure 6: Confusion matrix of all models.

6.3 Discussion

On analysing the performance of ELM and comparing to other ML models, ELM possessed with 96.42% accuracy as illustrated in Figure 7. Even though this is an impressive accuracy, ELM falls behind slightly when compared to RF where accuracy was 97.12%. As seen in Figure 8, F1-score of ELM is 0.96. On comparing F1-scores of all models, RF has the highest of 0.97. The time consumed by each model to predict is depicted in Figure 9. From this we can understand that ELM took 1.75 seconds to predict. This underlines that ELM is slower than RF, where the prediction time was only 0.27 seconds. This concludes that RF is better than ELM with respect to prediction accuracy and time consumption on CSE-CIC-IDS2018 dataset.

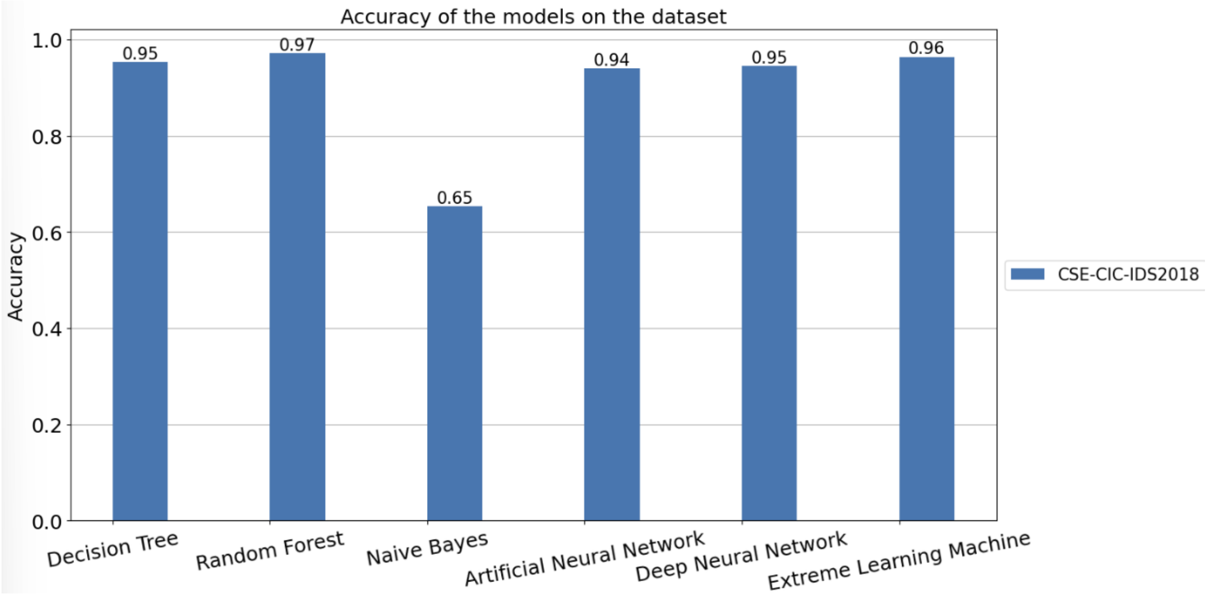


Figure 7: Accuracy of all the models.

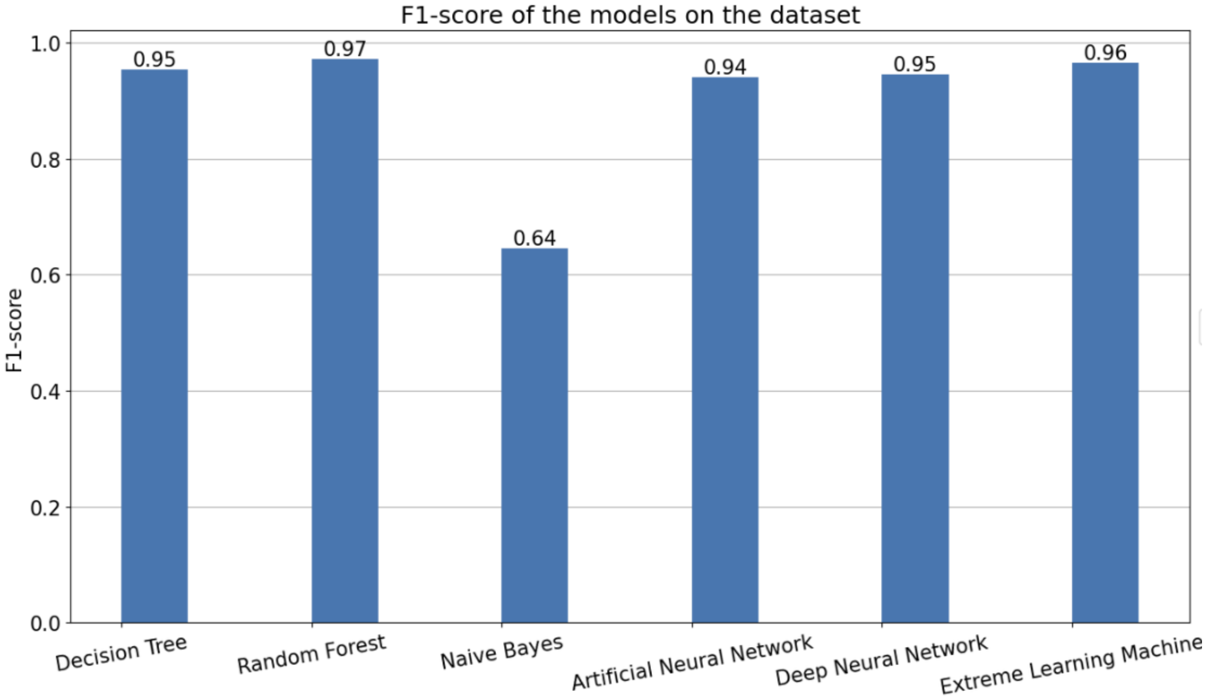


Figure 8: F1-score of all models.

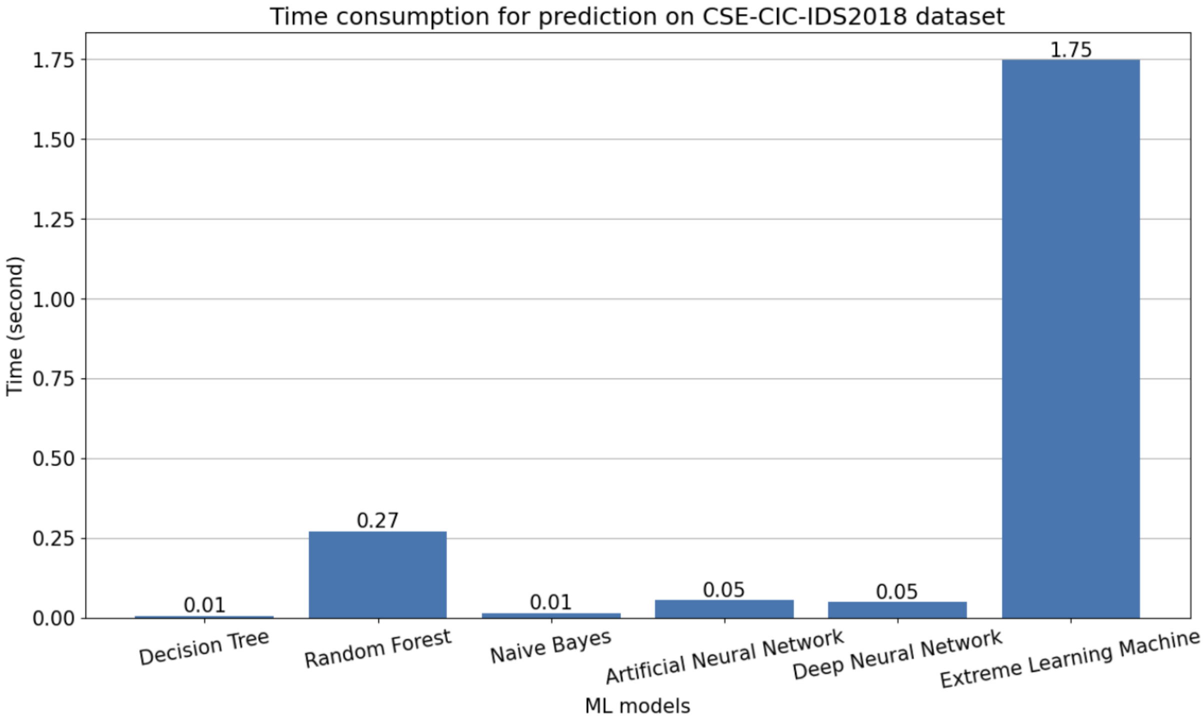


Figure 9: Time taken for prediction by each model.

7 Conclusion and Future Work

The main aim of this project was to experiment and evaluate the performance of ELM model compared to other ML models such as DT, RF, NB, ANN and DNN, for developing IDS that secure cloud environments. ELM model was successfully developed and evaluated using the real-time network traffic dataset, CSE-CIC-IDS2018. The results revealed that ELM achieved a commendable accuracy of 96.42%, but RF achieved a better accuracy of 97.12% compared to ELM. Analysis of the evaluation exposed that while ELM takes less training time and shows promising performance to be integrated with IDS applications, RF might be effective in particular scenarios.

From the results of the experiment, it can be concluded that ELM is slower than RF, as ELM took 1.75 seconds to predict, whereas the prediction time for RF was only 0.27 seconds. This concludes that RF is better than ELM in aspects of prediction accuracy and time consumption on CSE-CIC-IDS2018 dataset. Furthermore, the key findings from this research project signifies that ELM can achieve high performance while competing with other models such as DNN and ANN. Even though these insights indicate the efficiency of integrating ML based IDS to enhance cloud security, there is a necessity for model optimisation and evaluation with various datasets. Due to resource and time constraints, a publicly available IDS dataset was chosen, as generating a unique dataset was not possible. Additionally, training the model using single dataset can potentially bring limitations because providing varying network conditions or attack types can increase performance. As a future work the model can be evaluated using additional datasets. Furthermore, the model can be tested in real-world stimulated scenarios to evaluate its performance in various network conditions for securing cloud.

References

Chang, V., Golightly, L., Modesti, P., Xu, Q., A., Doan, L., M., T., Hall, K., Boddu, S., Kobusińska, A., 2022. A Survey on Intrusion Detection Systems for Fog and Cloud Computing. Future Internet, 14(3), pp. 89-89. doi: 10.3390/fi14030089

Hamed, T., & Marjan, K., R., 2020. A survey on security challenges in cloud computing: issues, threats, and solutions, Journal of Supercomputing, vol. 76 Issue 12, pp. 9493-9532. 40p. doi: 10.1007/s11227-020-03213-1

Gupta, G., Laxmi, P.R., Sharma, S., A Survey on Cloud Security Issues and Techniques. Available at: https://arxiv.org/pdf/1403.5627 [Accessed on: Nov 01, 2024].

Chauhan, M. & Shiaeles, S., 2023. An Analysis of Cloud Security Frameworks, Problems and Proposed Solutions, Network, MDPI, pp. 422-450. vol. 3(3). doi: 10.3390/network3030018.

Gill, K. S., Saxena, S. & Sharma, A., 2024. NCGTM: A Noncooperative Game-Theoretic Model to Assist IDS in Cloud Environment, IEEE Transactions on Industrial Informatics, pp. 3124-3132. doi: 10.1109/TII.2023.3300452.

Muhammad, D., Shanshan, T., Chuangbai, X., Hisham, A., Muhammad, W., & Sadaqat, U., R., 2023. Cyberattacks and Security of Cloud Computing: A Complete Guideline, Symmetry, Vol. 15 Issue 11, p1981. 33p. doi: 10.3390/sym15111981.

Gupta, Y. & Vashisth, R., 2023. Cyber Threats in Cloud Computing Environment. 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, IEEE. pp. 548-555. doi: 10.1109/ICESC57686.2023.10193701.

Kiranmai, B., Vasantha, S. V. & Jyoshna, B., 2023. Abnormal Behavior Detection to Avoid Attacks in Cloud based on Association Rules, 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, IEEE, pp. 429-433. doi: 10.1109/ICACRS58579.2023.10404118.

Razdan, S., Gupta, H. & Seth, A., 2021. Performance of Network Intrusion Detection Systems in Cloud Computing: A Review, 2nd Global Conference for Advancement in Technology (GCAT), Bangalore, India, IEEE, pp. 1-7. doi: 10.1109/GCAT52182.2021.9587481.

Lata, S., & Singh, D., 2022. Intrusion detection system in cloud environment: Literature survey & future research directions, International Journal of Information Management Data Insights, Vol. 2, Iss. 2, pp. 100134. doi: 10.1016/j.jjimei.2022.100134.

Qi, W., Wu, W., Wang, H., Ou, L., Hu, N., & Tian, Z., 2023. Intrusion Detection Techniques Analysis in Cloud Computing, IEEE 12th International Conference on Cloud Networking (CloudNet), Hoboken, NJ, USA, pp. 360-363. doi: 10.1109/CloudNet59005.2023.10490069.

Prabhakaran, V., & Kulandasamy, A., 2021. Detailed Review on IDS Techniques in Cloud Computing, 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, pp. 769-774. doi: 10.1109/ICCES51350.2021.9489091.

Tummalachervu, C., K. & D. P. R. Sanagana, 2024. Securing Cloud Computing Environment via Optimal Deep Learning-based Intrusion Detection Systems, Second International Conference on Data Science and Information System (ICDSIS), Hassan, India, IEEE, pp. 1-6. doi: 10.1109/ICDSIS61070.2024.10594404.

Sowmya, T. & Mary Anita E.A., 2023. A comprehensive review of AI based intrusion detection system, Measurement: Sensors, vol. 28. doi: 10.1016/j.measen.2023.100827.

Ferrag, M., A., Maglaras, L., Moschoyiannis, S., & Janicke, H., 2020. Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study, Journal of Information Security and Applications, vol. 50. doi: 10.1016/j.jisa.2019.102419.

Attou, H., Guezzaz, A., Benkirane, S., Azrour, M., & Farhaoui, Y., 2023. Cloud-Based Intrusion Detection Approach Using Machine Learning Techniques, Big Data Mining and Analytics, vol. 6, no. 3, pp. 311-320., doi: 10.26599/BDMA.2022.9020038.

Almusallam, N., Attou, H., Alabdultif, A., Guezzaz, A., Mohy-eddine, M., Benkirane, S., & Azrour, M., 2023. Towards an Intelligent Intrusion Detection System to Detect Malicious Activities in Cloud Computing, Applied Sciences, vol. 13(17). doi: 10.3390/app13179588.

Ibrahimi, K., Jouhari, M., & Jakout, Z., 2024. Enhancing Intrusion Detection Systems Using Machine Learning Classifiers on the CSE-CIC-IDS2018 Dataset, International Conference on Wireless Networks and Mobile Communications (WINCOM), Leeds, United Kingdom, pp. 1-6, doi: 10.1109/WINCOM62286.2024.10655131.

Tuan-Hong, C., & Iftekhar, S., 2023. Evaluation of Machine Learning Algorithms in Network-Based Intrusion Detection Using Progressive Dataset, Symmetry (20738994), Vol. 15, issue 6, p1251. 31p. doi: 10.3390/sym15061251.

Tut, M., A., & Mohsenabad, H., N., 2024. Optimizing Cybersecurity Attack Detection in Computer Networks: A Comparative Analysis of Bio-Inspired Optimization Algorithms Using the CSE-CIC-IDS 2018 Dataset, Applied Science, vol. 14, issue 3. doi: 10.3390/app14031044.

Tiwari, G., & Jain, R., 2022. A Novel Framework for Secure Cloud Computing Based IDS Using Machine Learning Techniques, International Conference on Soft Computing & Machine Intelligence (ISCMI), pp. 258-263. doi: 10.1109/ISCMI56532.2022.10068437.

Shukla, A., K., and Sharma, A., 2023. Cloud Base Intrusion Detection System using Convolutional and Supervised Machine Learning, International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, pp. 1-5. doi: 10.1109/ISCON57294.2023.10112007.

Vinolia, A., Kanya, N., and Rajavarman, V., N., 2023. Machine Learning and Deep Learning based Intrusion Detection in Cloud Environment: A Review, 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, pp. 952-960. doi: 10.1109/ICSSIT55814.2023.10060868.

Babu, G., R., Chintalapati, P., V., Majji, T., Hasini, S., S., L., HemaSri, M., and Prasanthi, M., L., 2023. A Design of an Integrated Intrusion Detection System on Cloud Using DNN, International Conference on Advanced Computing & Communication Technologies (ICACCTech), Banur, India, pp. 72-80, doi: 10.1109/ICACCTech61146.2023.00021.

Reji, M., Joseph, C., Nancy, P., Lourdes M., A., 2023. An intrusion detection system based on hybrid machine learning classifier, Journal of Intelligent & Fuzzy Systems, Business Source Ultimate. doi: 10.3233/JIFS-222427.

Bakro, M., Kumar, R., R., Alabrah, A., A., Ashraf, Z., Bisoy, S., K., Parveen, N., Khawatmi, S., & Abdelsalam, A., 2023. Efficient Intrusion Detection System in the Cloud Using Fusion Feature Selection Approaches and an Ensemble Classifier, Electronics 2023, 12(11), 2427. doi: 10.3390/electronics12112427.