

Cloud Computing Security Using Intelligent Intrusion Detection Mechanisms

Hamza Nasir¹, Azeem Ayaz², Shahzmaan Nizamani³, Saima Siraj⁴, Shahid Iqbal⁵, Muhammad Kamran Abid⁶

^{1,6}Department of Computer Science, NFC, IET, Multan, Pakistan

²Information Technology, SBBU, Shaheed Benazirabad, Pakistan.

^{3,4}Department of Information Technology, QUEST Nawab Shah, Pakistan.

⁵Department Computer Engineering, B. Z. University, Multan, Pakistan.

Corresponding Author: Muhammad Kamran Abid: (kamranabidhiraj@gmail.com)

ABSTRACT

New challenges of data breaches, unauthorized access, and insider threats have emerged with the rise of Cloud Computing. Attackers find cloud systems attractive due to their common infrastructure. To combat these security threats, strong security procedures must be integrated. An intrusion detection system (IDS) is a vital instrument for safeguarding networks and cloud environments. It tracks system activity and network traffic. This study compares the efficacy of various machine learning models for intrusion detection. The Random-Forest-Classifer model performed the best with a test accuracy of 99.88% in recognizing known and unknown threats. Furthermore, various benefits have been identified when contrasting this intelligent intrusion detection system with conventional rule-based and signature-based methods. The use of machine learning in intrusion detection systems offers enhanced security measures and improved threat detection capabilities, highlighting the importance of leveraging advanced technologies to secure cloud environments. This research illuminates the necessity to employ cutting-edge techniques to protect cloud computing systems from malicious actions in the era of cybersecurity. More than that, it stresses on the necessity to invest in enrichment mechanisms in machine learning-based intrusion detection systems so as to stay abreast with the evolving security problems.

Keywords: Cloud Security, Intrusion Detection, PCA, Machine Learning, Cloud Threat Detection

1. INTRODUCTION

The Cloud Computing network access paradigm is used to give resources like a network, data centers, hardware, software, and utilities when needed. For this reason, Cloud Computing is a technology that has a lot of potential and provides a number of advantages, such as accessibility, distant data collection, and storage. Because of its various features, including self-services, scalability, and availability, it drastically lowers expenses. The National Institute of Standard Technologies states that this paradigm is made up of three cloud service models. This architecture does have certain drawbacks, though, such as lower security and maintenance concerns. Because the private cloud is on-site and offers better security, businesses prefer it. Generally speaking, Cloud Computing has a number of security problems that impede the rapid adoption of cloud infrastructure, including privacy difficulties, regulatory compliance, and the potential for data loss. The sensitivity of users and organizations is one of these problems. Even though several tactics, like firewalls and antivirus software, have been created and put into use to protect cloud environments, data, and apps against assaults, more work has to be done. Next, intrusion detection, a collection of cutting-edge technologies that identify unwanted activity, is used to reinforce cloud security. However, intrusions are discovered through the inappropriate use of signature databases by intrusion detection systems [1]. It is not immune to a fresh attack with a distinct signature, but it also does not generate false alerts. Furthermore, a number of limitations that affect enormous data volumes, immediate detection, data quality, and other aspects of the system can hinder the efficacy of intrusion detection systems (IDSs).

Along with the benefits that come with Cloud Computing, the risk environment however, has been stretched with incidents of hacking and other activities that impinge on data security. This feature makes cloud network more attractive for criminals - the joint infrastructure, which requires implementing a range of programs to secure the system. There is an imperative to probe the efficiency of these IDS and build a three-machine learning model comparison model for intrusion detection across networks and cloud environment. In this research, the goal is to bridge this gap through an assessment of the performance of machine learning models in either predicting known or unknown threats in cloud infrastructures. It focuses particularly on the head-to-head document outlining the relative effectiveness of Random Forest Classifier approach versus outdated rule-based and signature-based methods. This

study shall sieve out the most efficient model to ensure that the upcoming intelligent intrusion detection systems are secure and solid enough to tackle the intrusions in cloud environments.

2. LITERATURE REVIEW

PaaS builds on IaaS by offering technical tiers and management software instances, freeing developers to focus on creating apps rather than managing the underlying infrastructure. SaaS, on the other hand, does away with the need for local program installation by giving users access to fully functional software that can be accessed over the cloud [2-7].

While each service model has its benefits, each one also has its drawbacks. Virtualization is the foundation for IaaS infrastructure provisioning; nonetheless, there are significant drawbacks, and IaaS services may someday become obsolete. Among the problems PaaS encounters are interoperability, privacy, authorization, host sensitivity, extensibility, and reliability. However, SaaS security issues include dependability, data security, authorization, authentication, and network monitoring [8-13]. These security issues need to be taken seriously by cloud providers. Players that try to exploit vulnerabilities in cloud environments are constantly switching up the tools and tactics they use. From time to time, it may be difficult for conventional intrusion detection systems to accurately identify variations in network traffic characteristics. Because of this, scientists emphasize how important it is to use ML and DL methods to increase IDS capabilities [1, 14-16]. A wide range of businesses, including finance, government, science, and security, are starting to apply ML and DL. For cybersecurity applications, machine learning's ability to classify and cluster data is crucial [17-20].

In this study, we conducted a complete comparison of various machine learning techniques, including logistic regression, KNeighborsClassifier, GaussianNB, Linear SVC(LBasedImpl), DecisionTreeClassifier, and Random Forest Classifier. These algorithms have demonstrated good performance in CC situations, effectively handling the complex and constantly evolving nature of intrusion detection.

3. METHODOLOGY

Data Collection and Preprocessing: The system gathers diverse data sources like system logs, network traffic, and SIEM data. This raw data is then pre-processed through several crucial stages:

- **Cleaning and filtering:** Eliminates noise, duplicate entries, and broken packets to ensure data integrity.
- **Feature engineering:** Transforms categorical features into numerical ones and extracts relevant features for efficient analysis.
- **Dimensionality reduction:** Utilizes PCA to compress the high-dimensional data into a lower-dimensional space without sacrificing crucial information. This improves training efficiency and reduces computational overhead.

For the final system, the model that performs best on unseen data in terms of accuracy and generalizability is selected. As shown in Figure 1, this study methodology combines feature engineering, dimensionality reduction, and machine learning algorithms in an effort to create a reliable and effective intrusion detection system that can successfully defend cloud systems against a variety of dynamic threats.

4. RESULT AND DISCUSSION

The systematic process guarantees that the dataset is appropriately adjusted for efficient utilization in intrusion detection systems. High-dimensional data processing is made possible by PCA's ability to drastically cut down on algorithm training times when dealing with a large number of features. This is just one advantage of PCA. Without PCA, processing 100 features, for instance, would require 4950 scatter plots, which would probably make in-depth analysis difficult. The accuracy and performance of the intelligent intrusion detection system (IDS), which are enhanced by the cautious use of the pretreatment pipeline, are critical to effective intrusion detection in the cloud computing environment.

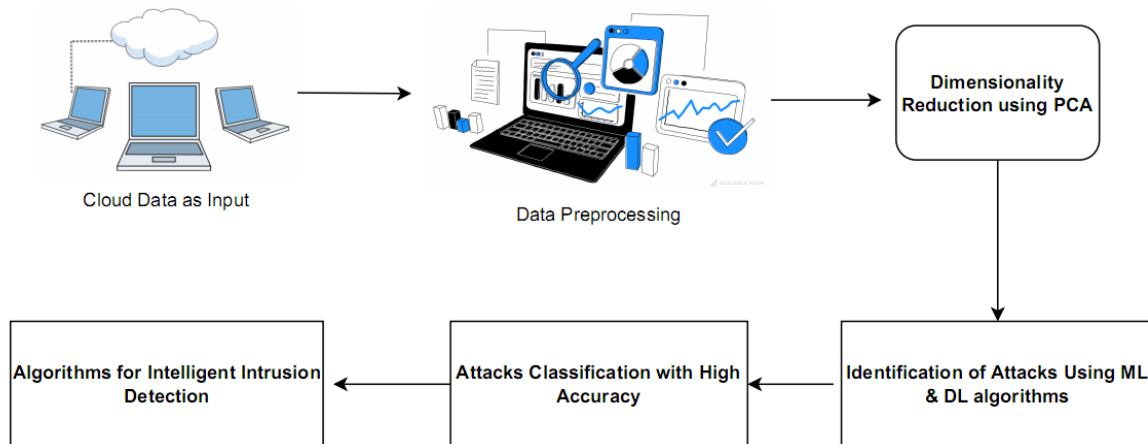


Figure 1. Methodology for intelligent intrusion detection system

4.1. Modeling

During the modeling process, a machine learning algorithm is trained to predict labels from features in order to produce a trained model for additional inference. Consider creating an application that uses a model trained on tagged photos of faces to interpret facial expressions and anticipate emotions. A crucial component of this process is the statistical model known as logistic regression. It calculates the probability of an event and is commonly used in classification tasks. In logistic regression, the greatest likelihood technique is used to estimate coefficients after probabilities are transformed using a logit function. Specifically, binary logistic regression is suitable for binary outcomes (spam or not), but multinomial logistic regression performs well when dealing with three or more unique outcomes in a series. It can be used to predict preferences in a range of contexts, including movie genres, because to its versatility.

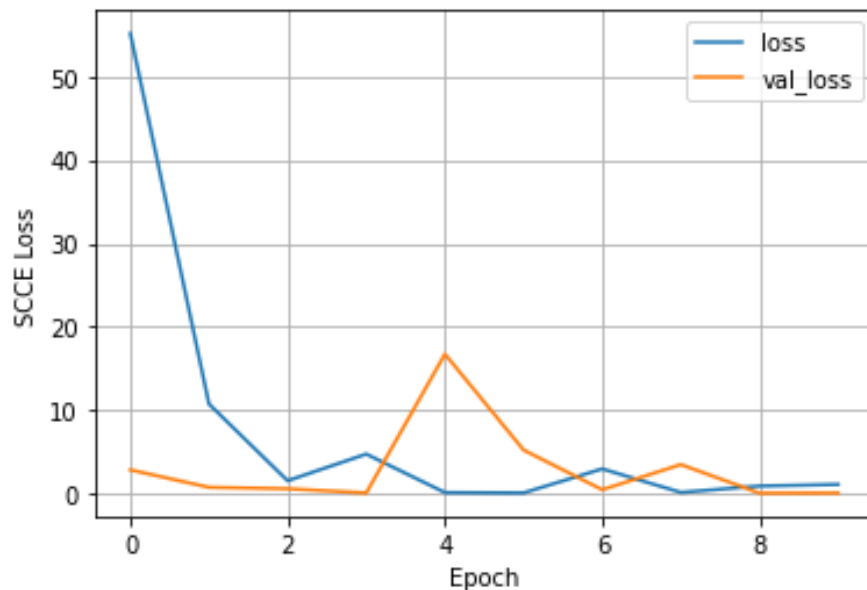


Figure 2. Model training vs testing accuracy

Figure 2 further illustrates that all of the models have higher training accuracy than test accuracy. This suggests that the models might be overfitting the training data by a small amount. The smallest gap between test and training accuracy indicates that the ANN model, in contrast to the other models, appears to be less overfitting.

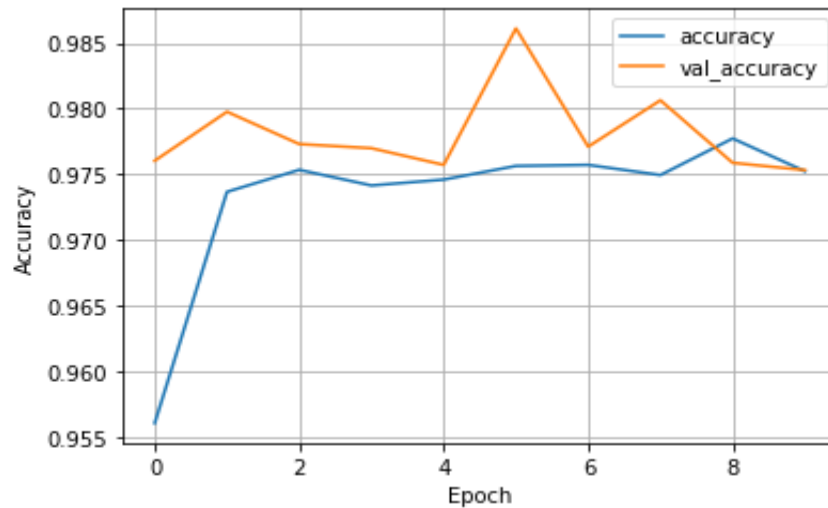


Figure 3. Training accuracy improve with No. of Epoch

Accuracy is calculated using data from the test and training sets. As Figure 3 shows, the approach has a little lower test accuracy but a higher training accuracy. This suggests that the algorithm might have over fitted the training set. Furthermore, the algorithm's evolution is demonstrated by Figure 3's test accuracy, which increases from 95.60% in epoch 1 to 97.53% in epoch 10. This suggests that when the algorithm learns from the data, it will eventually perform more effectively. Principal component analysis, or PCA, is a statistical technique used to extract maximum information from high-dimensional data by identifying the most important elements within a dataset. Depending on how much diversity each feature adds to the final result, certain traits are chosen. The attribute that most significantly contributes to variance is the first primary component. The feature that explains the second largest variation is the second principal component, and so on. One striking feature is that there is no relationship at all between the fundamental elements.

Model Creation and Instruction: A range of intelligence techniques, including machine learning algorithms, are employed in the construction of the intrusion detection models. Feature selection strategies are used to extract the most relevant features from the dataset. To train the models, a tagged dataset with both normal and attack events is utilized.

Table 1. Models Comparison

Model	Training Accuracy	Test Accuracy	Training Precision	Test Precision	Training Recall	Test Recall
Logistic Regression	87.97%	87.62%	83.81%	83.57%	91.86%	91.61%
KNeighbors Classifier	99.05%	98.94%	99.23%	99.06%	98.73%	98.67%
Gaussian NB	91.80%	91.61%	92.63%	92.53%	89.48%	89.30%
Linear SVC(LBasedImpl)	97.48%	97.29%	97.85%	97.55%	96.71%	96.65%
Decision Tree Classifier	99.99%	99.87%	100.00%	99.85%	99.99%	99.87%
RandomForest Classifier	99.99%	99.88%	99.99%	99.93%	99.99%	99.81%
PCA Random Forest	99.99%	99.82%	99.99%	99.89%	99.99%	99.72%

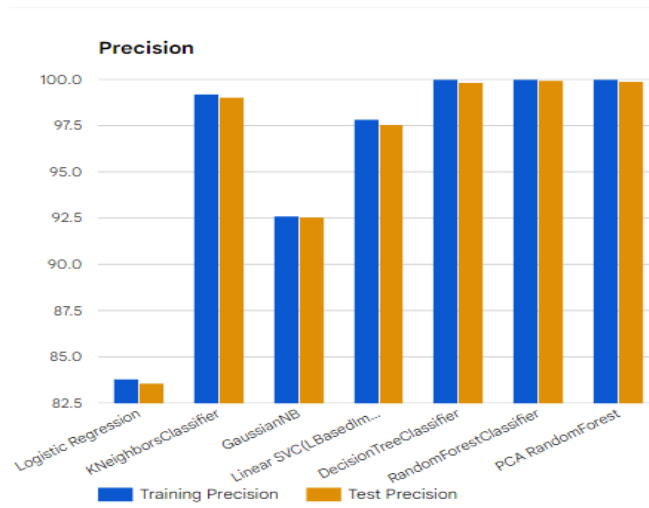


Figure 4. Training and Test Precision of Different algorithms

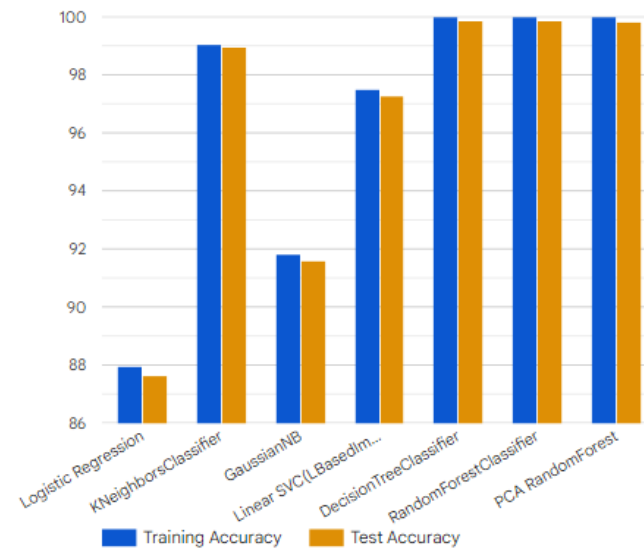


Figure 5. Training and text Accuracy of Different algorithms

Table 2. Comparison of deployed models

Model	Training Accuracy	Test Accuracy
Logistic Regression	87.97%	87.62%
KNeighborsClassifier	99.05%	98.94%
GaussianNB	91.80%	91.61%
Linear SVC(LBasedImpl)	97.48%	97.29%
DecisionTreeClassifier	99.99%	99.87%
RandomForestClassifier	99.99%	99.88%

PCA RandomForest	99.99%	99.82%
Epoch 1/10	95.60%	97.60%
Epoch 2/10	97.37%	97.98%
Epoch 3/10	97.53%	97.73%
Epoch 4/10	97.41%	97.70%
Epoch 5/10	97.46%	97.57%
Epoch 6/10	97.56%	98.61%
Epoch 7/10	97.57%	97.71%
Epoch 8/10	97.49%	98.06%
Epoch 9/10	97.77%	97.59%
Epoch 10/10	97.52%	97.53%

Table 2 shows the training and test accuracy of several machine learning models, which models have the best test accuracy: Random Forest Classifier, PCA Random Forest, and Decision Tree Classifier, with a combined test accuracy of more than 99.8% as shown in the Figure 4 and Figure 5 accordingly. The test accuracies of the K-Neighbors Classifier, Logistic Regression, and Gaussian-NB models all continue to be quite high despite their decline. Furthermore, test accuracy is generally lower than training accuracy, as Table 1 illustrates. This suggests that the models might be overfitting the training data by a small amount. The overfitting is not too terrible because the test accuracies are still good.

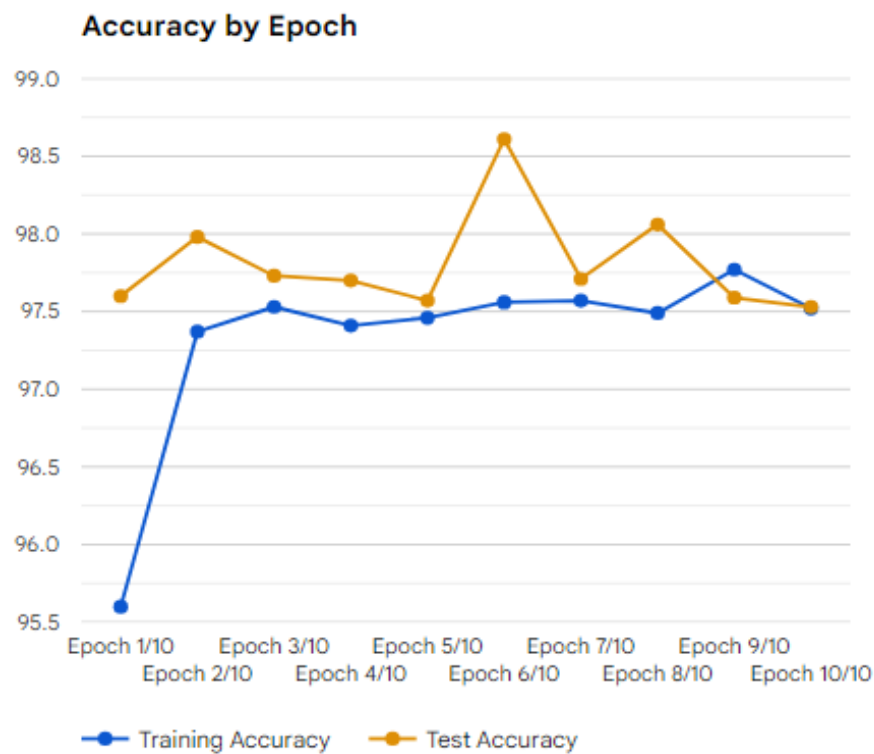


Figure 6. Training and test accuracy over 10 Epoch

Figure 6 shows the training and test accuracy of a machine learning algorithm over ten epochs. Accuracy is computed using the test and training data (blue line). The graph shows that the method has a high training accuracy but a relatively poor test accuracy. This suggests that the algorithm might have over fitted the training set. Figure 6, where test accuracy increases from 95.60% in epoch 1 to 97.53% in epoch 10. This suggests that when the algorithm learns from these experiences, it will eventually perform better. A look at the graph's strong test accuracy and low overfitting rate indicates that the machine learning system is performing well overall. In just 10 epochs, the system achieves a test accuracy of over 97%, proving its capacity to quickly identify the underlying patterns in the data.

In comparison to other models, the Random Forest Classifier's accuracy is high, as seen in figure 7, with its accuracy being close to 99%, whereas other models' accuracy ranges from 87% to 98%. Additionally, the model's training accuracy is higher than that of other models. In terms of test accuracy and overfitting, Figure 7 shows that the Random Forest model outperforms the other models. Despite having more parameters than the others, the Random Forest Classifier outperforms the others in detecting the underlying patterns in the data. In terms of overfitting, the Random Forest model outperforms the Decision Tree approach. The excellent test accuracy of the Random Forest Classifier model indicates that it may effectively generalize to new data. The proposed intelligent intrusion detection system outperforms conventional methods in terms of flexibility, detection accuracy, and threat recognition.

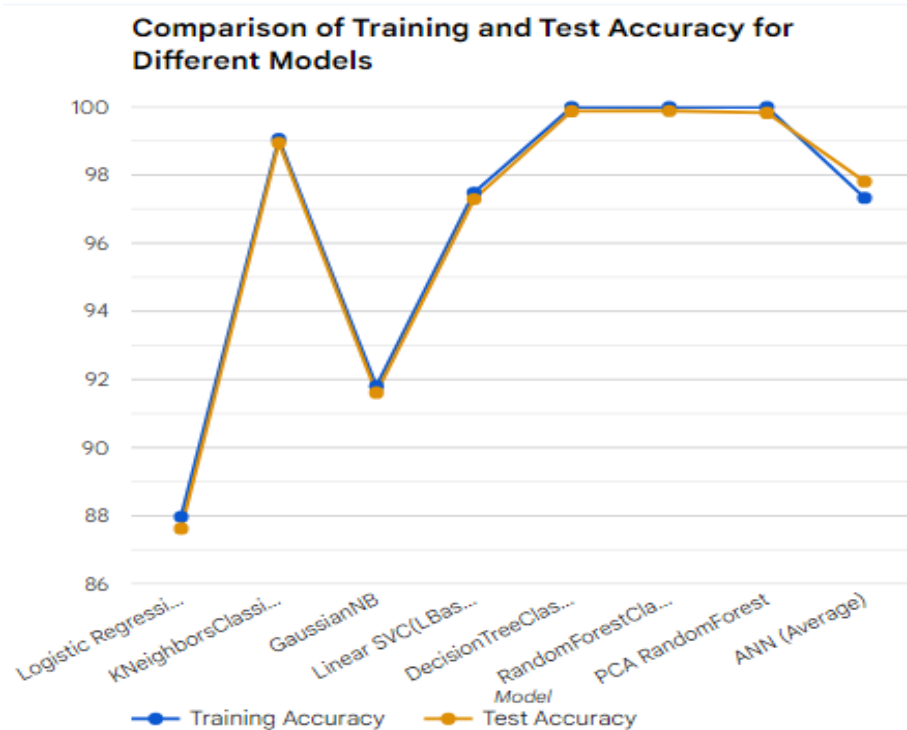


Figure 7. Comparison of training and test accuracy

5. LIMITATIONS

The paper assesses performance of mob all machine learning systems for intrusion detection cloud environment, which signals the risk of being hacked, suffering unauthorized access, or an employee's wrongdoing in the cloud computing age. In the study, there is recognition of the critical role that strong security measures can play. Nonetheless, the algorithm-based model called Random-Forest-Classifier is highlighted which demonstrated accuracy of 99.88% for both known as well as unknown threats. It outdid the various conventional methods of rule-based and signature-based types of recognition. Systematically development of machine learning algorithm will be a prerequisite to match the changing security challenges.

6. CONCLUSION AND FUTURE DIRECTIONS

Many algorithms have been tried for high accuracy, with the main focus of the research being on the design of intelligent intrusion detection systems with high accuracy for cloud computing security. The models have been chosen

based on a survey of the literature, and the random forest model performs better than the other models that were chosen. The algorithms that are primarily evaluated in various scenarios are chosen for accuracy comparison. The accuracy of the random forest model is almost 99%, with higher accuracy previously discussed in the literature. Critical procedures like feature engineering and cleaning are included in the data preparation step to verify data appropriateness and eliminate extraneous information. Models are evaluated, trained, and their performance adjusted based on evaluation results using labeled datasets. Using the adaptability of machine learning, it successfully tests the intrusion detection system in actual or simulated cloud settings by comparing the experimental validation technique to rule-based and signature-based systems. In the Results and Discussion chapter, several machine learning models are compared, with an emphasis on the Random Forest Classifier's higher test accuracy. In summary, the study offers important insights into cloud security and emphasizes the effectiveness of the recommended intelligent intrusion detection system.

The future work directions comprise the expanding of the machine learning models, connecting of deep learning methods including CNNs and RNNs, modification of intrusion detection systems according to the previously mentioned cloud environments, and integrating anomaly detection technique for improved threat detection and reaction. Among the limitations of the study, we can point to dataset bias, a difference in model performance based on the size and complexity of the network, and the study's narrow framework that encompasses known as well as unknown threats. However, excluding zero-day attacks and a computationally complicated deployment of machine learning in cloud systems is another limitation that needs to be addressed.

REFERENCES

- [1] S. Kavitha, N. Uma Maheswari, and R. Venkatesh, "Intelligent Intrusion Detection System using Enhanced Arithmetic Optimization Algorithm with Deep Learning Model," *Teh. Vjesn.*, vol. 30, no. 4, pp. 1217–1224, 2023, doi: 10.17559/TV-20221128071759.
- [2] A. Kumar *et al.*, "An intrusion identification and prevention for cloud computing: From the perspective of deep learning," *Optik (Stuttg.)*, vol. 270, p. 170044, 2022, doi: <https://doi.org/10.1016/j.ijleo.2022.170044>.
- [3] S. Lata and D. Singh, "Intrusion detection system in cloud environment: Literature survey & future research directions," *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 2, p. 100134, 2022, doi: <https://doi.org/10.1016/j.ijime.2022.100134>.
- [4] U. K. Lilhore *et al.*, "HIDM: Hybrid Intrusion Detection Model for Industry 4.0 Networks Using an Optimized CNN-LSTM with Transfer Learning," *Sensors*, vol. 23, no. 18, 2023, doi: 10.3390/s23187856.
- [5] H. Attou *et al.*, "Towards an Intelligent Intrusion Detection System to Detect Malicious Activities in Cloud Computing," *Appl. Sci.*, vol. 13, no. 17, 2023, doi: 10.3390/app13179588.
- [6] B. Ali and A. I. Awad, "Cyber and physical security vulnerability assessment for IoT-based smart homes," *Sensors*, vol. 18, p. 817, 2018.
- [7] I. Ali *et al.*, "Systematic literature review on IoT-based botnet attack," *IEEE Access*, vol. 8, pp. 212220–212232, 2020.
- [8] S. Ahmed *et al.*, "Towards supply chain visibility using Internet of things: A dyadic analysis review," *Sensors*, vol. 21, p. 4158, 2021.
- [9] M. Burhan, R. A. Rehman, B. Khan, and B.-S. Kim, "IoT elements, layered architectures and security issues: A comprehensive survey," *Sensors*, vol. 18, p. 2796, 2018.
- [10] W. H. Hassan, "Current research on Internet of Things (IoT) security: A survey," *Comput. Netw.*, vol. 148, pp. 283–294, 2019.
- [11] H. Ibrahim, "A Review on the Mechanism Mitigating and Eliminating Internet Crimes using Modern Technologies: Mitigating Internet crimes using modern technologies," *Wasit J. Comput. Math. Sci.*, vol. 1, pp. 76–108, 2022.
- [12] T. Wang, M. Z. A. Bhuiyan, G. Wang, L. Qi, J. Wu, and T. Hayajneh, "Preserving balance between privacy and data integrity in edge-assisted Internet of Things," *IEEE Internet Things J.*, vol. 7, pp. 2679–2689, 2019.
- [13] J. Qin *et al.*, "Deep learning-based software and hardware framework for a noncontact inspection platform for aggregate grading," *Measurement*, vol. 211, p. 112634, 2023.
- [14] R. Kumar, P. Kumar, A. Jolfaci, and A. K. M. N. Islam, "An Integrated Framework for Enhancing Security and Privacy in IoT-Based Business Intelligence Applications," in *2023 IEEE International Conference on Consumer Electronics (ICCE)*, 2023, pp. 1–6. doi: 10.1109/ICCE56470.2023.10043450.
- [15] S. Subramani and M. Selvi, "Multi-objective PSO based feature selection for intrusion detection in IoT based wireless sensor networks," *Optik (Stuttg.)*, vol. 273, p. 170419, 2023, doi: <https://doi.org/10.1016/j.ijleo.2022.170419>.

- [16] M. Mayuranathan, S. K. Saravanan, B. Muthusenthil, and A. Samydurui, "An efficient optimal security system for intrusion detection in cloud computing environment using hybrid deep learning technique," *Adv. Eng. Softw.*, vol. 173, p. 103236, 2022, doi: <https://doi.org/10.1016/j.advengsoft.2022.103236>.
- [17] M. Saied, S. Guirguis, and M. Madbouly, "Review of artificial intelligence for enhancing intrusion detection in the internet of things," *Eng. Appl. Artif. Intell.*, vol. 127, p. 107231, 2024, doi: <https://doi.org/10.1016/j.engappai.2023.107231>.
- [18] S. Rizvi, A. Kurtz, J. Pfeffer, and M. Rizvi, "Securing the Internet of things (IoT): A security taxonomy for IoT," in *Proceedings of the 2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE)*, 2018, pp. 1–3.
- [19] K. Shaukat, T. M. Alam, I. A. Hameed, W. A. Khan, N. Abbas, and S. Luo, "A review on security challenges in Internet of things (IoT)," in *Proceedings of the 2021 26th International Conference on Automation and Computing (ICAC)*, 2021, pp. 2–4.
- [20] A. Odeh and A. Abu Taleb, "Ensemble-Based Deep Learning Models for Enhancing IoT Intrusion Detection," *Appl. Sci.*, vol. 13, no. 21, p. 11985, 2023, doi: 10.3390/app132111985.