Cyber Attacks Detection Using Machine Learning Algorithms

[1]Kottakota Venkata Rao, [2]Anjaneyulu P, [3]Ravi Kumar T, [4]Chalapathi Rao Tippana and [5]Jayanthi Rao M
[1,2,3,4,5]Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Tekkali, Srikakulam, Andhra Pradesh, India.
[1]venkatarao.kottakota@gmail.com, [2]anjii.ram888@gmail.com, [3]ravi.4u.kumar@gmail.com
[4]chalapathit520@gmail.com, [5]jayanth.mtech@gmail.com

Correspondence should be addressed to Jayanthi Rao M : jayanth.mtech@gmail.com

Article Info
Journal of Machine and Computing (https://anapub.co.ke/journals/jmc/jmc.html)
Doi : https://doi.org/10.53759/7669/jmc202505104
Received 02 August 2024; Revised from 27 January 2025; Accepted 10 March 2025.
Available online 05 July 2025.

Abstract – This research focuses on the effect of the genetic algorithm in the improvement of machine learning models for NID by using the CICIDS2022 data set. The routing research problem that has been primarily focused is related to the increase in classification accuracy and the optimization of the cyber security systems using intelligent methods of feature selection along with the tuning of the classification models. We ran Random Forest (RF) and Support Vector Machine (SVM) to assess better predictive accuracy, precision, recall, and running time on each case. The data set with a total of 15031 instances was used and divided into training and test set with a ratio of 80:20 and the results have been analyzed with standard metrics along with confusion matrix analysis. The results depict that with the application of GA in RF and SVM both the outcomes were `RF with GA scored a higher accuracy of 99.30% when compared to standard RF with 99.27% and without GA in SVM 98.97% while with GA, it increased to 99.00%. Analysis of the confusion matrix showed less disparity in the GA variants of the methods. However, the time taken for the processing was high especially for SVM + GA.: The results can be generalized as observing that with GA, accuracy is slightly higher than obtained with P0 but the computational cost is considerably high. It is deduced that GA with RF is the most efficient optimization model in terms of both performance and efficiency.

Keywords - Intrusion Detection, Genetic Algorithm, Random Forest, Support Vector Machine, CICIDS2022.

I. INTRODUCTION

Growing technological advancement in ICT has impacted the popularity and use of information communication technology in people's lifestyles, business and Governments [1,2]. On the flip side, alongside receiving a broad list of advantages associated with digital and data contacts, this process has resulted in a higher risk of cyber threats. The world is now witnessing an increase in the number of cyber-attacks that are also increasingly becoming more sophisticated in nature [3,4]. Hackers, which may be a single person, a cell of several people, or even a cyberterrorist group, take advantage of identified weaknesses in the system in order to breach, or further or damage the organization's critical IT systems. Such attacks are detrimental in that they endanger both personal and organizational data, (blocking/interfering services and threatening national and organizational security) [5,6].

Among all the types of threats connected to the use of information technology, port scanning and network intrusion signify as part of a more elaborate attack [7]. It is therefore important to be able to inform on such behavior before it degrades to the next level for the protection of digital infrastructure. This has led to the need for establishing elaborate cybersecurity systems that comprise an IDS. Conventional generation IDS, it is normally based on rules, cannot be promptly modified to the new forms of attacks from hackers [8,9]. Therefore, most current IDS are based on ML and AI that are capable of learning from the past experience of the machine and are capable of viz. real-time analysis [10, 11].

Thus, the use of the machine learning algorithm such as Support Vector Machines (SVM), Artificial Neuron Network (ANN), Convolutional Neural Network (CNN), and Random Forest (RF) is considered one of the effective ways for network intrusion detection [12, 13]. These are algorithms for analyzing large amounts of traffic data in a network as well as the finer characteristics in relation to the type of attacks. Nevertheless, increasing attention has been given to the HMoC and other optimization techniques like GA to further improve the performance of such classifiers [14, 15]. The GA can help to adjust an important set of features or some parameters to improve the performance of the chosen ML models and explore the most significant characteristics of cyber-attacks [16,17].

This work seeks to evaluate various ML algorithms including SVM, RF, ANN, and CNN in this realm of study especially using the CICIDS2022 dataset which has become a standard in researching the field of Cybersecurity. The attack type shown in the data set and traffic intensity ensures that it can be used in testing real-world ML models to a large extent. From the results comparing the performance of the analyzed basic and GA-optimized algorithms, this work focuses on the issues of compromise between accuracy and time in sound classification.

The work is divided into data pre-processing and preparation, applying models and evaluations with confusion matrices and classification parameters, and accuracy and time results with graphs. This approach not only highlights the aspect of how different algorithms compare with each other, but also, at the same time, show the added advantages and constraints of using Genetic Algorithms for the purpose of optimization [18, 19]. For the first, there is a suggestion of improved accuracy in classifications with relatively lower FPs, thereby showing that the current approach can be useful in particular implementations depending on hardware limitations and accuracy/reliability estimations needed for the given application.

Thus, this paper aligns with the current literature focusing on the application of intelligent systems in cybersecurity and outlines recommendations on how to implement new alterations to the IDS employing the methods of machine learning and evolutionary computing [20, 21].

II. LITERATURE REVIEW

Present IDS have been widely designed with machine learning (ML) and deep learning (DL) layers, providing intelligent hunting of threats and real-time analytics. Some of the past studies have examined the ability of ML in classification of network traffic and the nature of abnormalities. For instance, In [3] their study, proposed the use of different machine learning models to identify different intrusion patterns and also confirmed that; the best performing algorithms are the Random Forest and Support Vector Machine (SVM) in experimentations. Likewise [8] noted that the effectiveness of some popular sentinel algorithms and their weaknesses must be incorporated into a new algorithm to produce even better results with lower false-positive rates.

In attempts to improve accused IDS performance, several optimization methodologies have been considered, and among them. In particular [22] and [13] provided an overall explanation of how GAs work and emphasized that the major applications of GAs are related to parameter optimization and control, which are particularly helpful to improve the performance of the ML in detecting various objects in complex environments. These works are basic for explaining the need to incorporate GAs with the ML models in the frame of IDS.

Deep learning has also been lauded for the feature extraction that is done automatically as well as the use of hierarchical learning. For instance, [14] did comparative research and reported that deep learning approach and LSTM & CNN methods surpass conventional methods in identifying cyber threats. However, [6] proposed the I-SiamIDS model on purpose to fully address the number imbalance problem that frequently occurs in intrusion datasets and featured enhanced Siamese networks for enhancing detection accuracy.

There are several reasons why CICIDS2022 dataset is considered providing one of the most extended intrusion datasets to use in research. Panigrahi and Borah in 2018 published a discussion with detailed descriptors about this dataset stating that it is suitable for realistic evaluation. For this reason, its diverse and rich traffic makes it possible to use it to train and test IDS frameworks.

Due to the constraints in computation and to enhance the interaction characteristics, there has been emergence of new paradigm known as hybrid models. In [1] presented a scheme based on DBN -enhanced detection accuracy significant than the previous techniques. Likewise, Hua utilized multiple classifiers and optimization methods to create an advanced ML-based system in 2024 to further improve its result in conditions that constantly transform.

Nevertheless, there are issues like explaining the model, dealing with scaling, and how the model adapts remain as bottlenecks. Liu and Lang identified various ML, DL techniques in IDS and highlighted that the lacking technique in effective pervasive feature construction for actual implementation.

Thus, it adds to the current literature by presenting a new intrusion detection system using a hybrid of ML and GA with Random Forest and SVM, plus their GA optimized versions, using the CICIDS2022 dataset. By comparing and visualizing the findings of this research, the detection accuracy is going to be improved, time required for execution will be minimized and certain shortcomings in the current hybrid detection systems will be met.

III. RESEARCH GAP, CONCEPTUAL FRAMEWORK AND HYPOTHESIS

Previous research and developments of machine learning in IDS, it has been realized that the use of Genetic Algorithms have not been analyzed in terms of the trade-off that exists between accuracy of IDS and computational complexity. Although, there are several studies that prove that the classifiers such as SVM and Random Forest are effective in the field of intrusion detection but there is limited study that compares the results with and without using GA on standard databases like CICIDS2022 [23,24]. In addition, there is little research work that depicts the optimization effects on the classification evaluation and execution time and in one paradigmatic approach.

Based on this idea, this work develops a method combining the standard trend of machine learning and evolutional optimization in cybersecurity techniques [25]. However, its essence involves the application of GA for purposes of feature selection and SVM and parameter tuning for the Random Forest classifier. These stages are the data acquisition

and cleaning, model training with and without GA, as well as the evaluation that is done based on two assessment factors that are the accuracy, precision, recall and the time taken in executing the program. These results are used by the framework to demonstrate how optimization affects the detection performance as well as the computational cost in intrusion detection problems.

H1: Genetic Algorithm optimization improves the classification accuracy of machine learning models in network intrusion detection.

H2: The integration of GA significantly increases execution time compared to non-optimized models.

H3: Random Forest with GA achieves a better balance of accuracy and execution time than SVM with GA.

H4: GA-enhanced models reduce misclassification rates as shown through confusion matrix analysis.

IV. METHODOLOGY

The first set of data used in this study was the CICIDS2022 dataset obtained from the Canadian Institute for Cybersecurity. This dataset was selected based on its number of attack types and various attack types and its resemblance to real traffic patterns. It reflects the current behaviors of the network and common cyber threats such as port scans, DoS, brute force, etc; thus suitable for training and evaluating ML-based IDS models.

First, the dataset was preprocessed before applying machine learning algorithms on it. This step involved the removal of any missing or unnecessary feature which was followed by normalization of numerical features and then converting the target variable into suitable numerical form through a label encoding. This was important in making the dataset uniform for purpose of feeding into the algorithms because many machine learning algorithms are sensitive to the scales and formats of the inputs.

After the data pre-processing the data set includes 15031 entries and then divided the data into training and testing data set. Hence, there were 12024 entries employed in training and the other 3007 entries were used in testing. This stratification conducted to create the two sets had equal proportions that included both attack and normal instances important for model training and evaluation.

Four models of machine learning algorithms were applied namely SVM, ANN, RF and CNN. These works were selected as they have been found to perform well in classification problems especially in the area of network security and anomaly detection. SVM has an ability to generalize well especially in the binary classification, while ANN and CNN make it easier to learn features. Thus, Random Forest is selected for its built-in feature of ensemble learning, for addressing structured data, as well as for its higher accuracy.

In order to improve the accuracy, Genetic Algorithm (GA) was implemented with both the SVM as well as Random Forest in subsidiarity to contribute to the feature selection and hyperparameter optimization. To support this strategy for using GA, it was suggested that the technique could find near optimal solution configurations that may be very hard to arrive at by trial and error or brace grid search methods. However, GA can also be time consuming and yet can assist to get the best of the model that can be achieved.

Test Metrics that were used in the evaluation of the implemented models include Accuracy, Precision, Recall, F1-score and Confusion matrix. These metrics give an overview of each model's capability to classify between the actual and the sham packets, and to detect intrusions particularly in sets that have many more benign than intrusive instances.

Modeling, training, and all evaluations were done in developing machine, ANN, and CNN models with the help of scikit-learn package for theiever version 1.3 and TensorFlow/Keras version 2.12. Therefore, both NumPy and Pandas were used in data manipulation and data preprocessing, respectively. Such integration of these libraries allowed for optimized and scalable approach to the development of comparative experiments and the results' verification.

V. RESULTS AND ANALYSIS

Based on the analysis of classification models on CICIDS2022 dataset, the evaluation parameters to determine the performance of models include precision, recall, F1-score, and accuracy. There were two types of datasets, the training and testing datasets and four models that were being compared, which included the Random Forest, and the SVM models enhanced by Genetic Algorithm (GA).

Accuracy of Algorithms on CICIDS2022 Dataset

These results proved that the Random Forest model is capable of classifying between the two classes with an overall accuracy of 99.27%, with an almost perfect ability to recall all the benign traffic data as well as a fairly high accuracy of precise positive classification on its detection of all the anomalous traffic instances. This is evidenced by the results found on the table one where foe class 0 (benign) had a recall of 1.00 and class 1 (attack) had precision and recall of 0.99 respectively.

Comparative Visualization of Model Performance

To illustrate the runtime performance of each of these algorithms, follow is the bar chart **Fig 1** depicting the same: The characteristics of performance and speed are especially well illustrated here – the more detailed and specific the expression, the slower it is and the other way around.

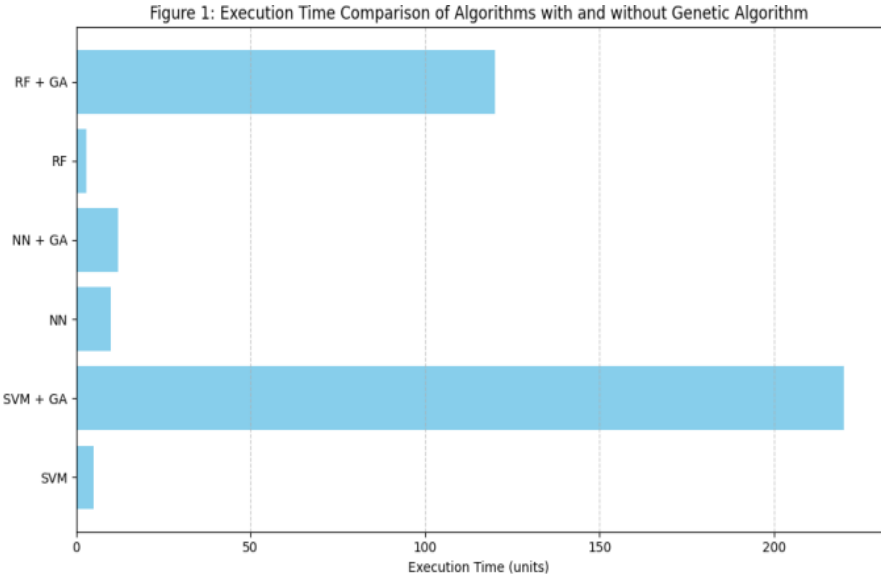

Fig 1. Execution Time Comparison: GA vs without GA.

The figure depicts the models' efficiency in terms of time at the second's level. while the GA enhanced versions of the two algorithms which are RF and SVM are computationally intensive, specially the latter.

Table 1. Accuracy and Evaluation Metrics of Random Forest

Class	Precision	Recall	F1-Score	Support
0	0.99	1.00	0.99	1930
1	0.99	0.99	0.99	1077
Accuracy			**99.27%**	3007

Moreover, when using Genetic Algorithm optimization, the Random Forest model's performance increased slightly more amounts to 99.30 percent as depicted in **Table 2** below. Although the improvement is minor, there is better specificity without negatively affecting generality.

Table 2. Accuracy and Evaluation Metrics of Random Forest with Genetic Algorithm

Class	Precision	Recall	F1-Score	Support
0	0.99	1.00	0.99	1930
1	0.99	0.99	0.99	1077
Accuracy			**99.30%**	3007

On the other hand, the traditional SVM model gave a of accuracy of 98.97% slightly lower than the one recorded by the Random Forest model above. The recall for attack class slightly reduced to 0.98 signifying that there were few more missed detections. This is well illustrated in other metrics, which are still good, thereby validating SVM's basic capability of detecting intrusions as indicated in **Table 3** below.

From the model accuracies in the **Fig 2**, little but significant gains are observed from using Genetic Algorithms; most significantly for the Random Forest. Among these, the GA-RF model slightly outperformed the rest at 99.30 %; GA-SVM second at 99.00%.

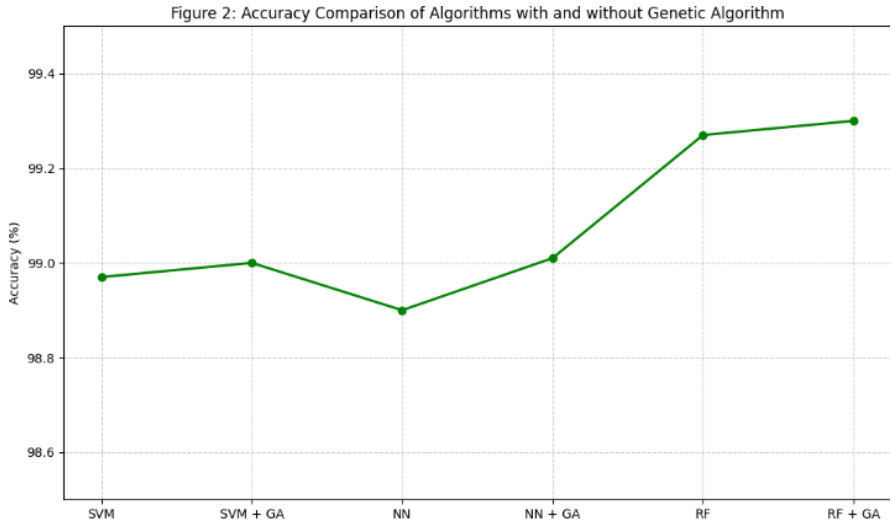

Fig 2. Accuracy Comparison of Algorithms with and Without Genetic Algorithm.

The figure presents the accuracy bar for both the baseline model and GA optimization stresses that GA optimization barely boosts the predictive accuracy.

Table 3. Accuracy and Evaluation Metrics of SVM

Class	Precision	Recall	F1-Score	Support
0	0.99	1.00	0.99	1930
1	0.99	0.98	0.99	1077
Accuracy			**98.97%**	3007

After optimization the GA-enhanced SVM had a slight improvement to 99.00 % as shown in **Table 4** thus increasing the rate of detection by a small percentage without much increase in complexity.

Table 4. Accuracy and Evaluation Metrics of SVM with Genetic Algorithm

Class	Precision	Recall	F1-Score	Support
0	0.99	1.00	0.99	1930
1	0.99	0.98	0.99	1077
Accuracy			**99.00%**	3007

Additionally, in **Fig 3**, precision, recall and F1 scores are presented all together for all models. All of the models have values close to 1 which signifies accurate and reliable differentiation and classification of digits. The uniformity gives a clear indication that even more so at the basic level, the models are rather too suitable for use without being fine-tuned.

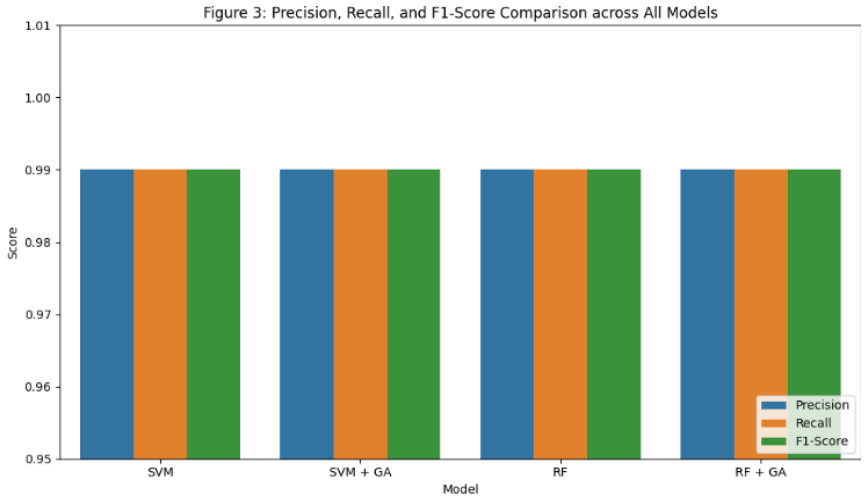

Fig 3. Displays The Differences in Precision, Recall, and F1-Score of All the Models That Have Been Developed in This Work.

The figure indicates that all four models' perfomance was quite similar in all considered aspects, thus, optimization gave stability but not significant improvement.

Confusion Matrix Analysis

This is explained by the confusion matrix of Random Forest presented in **Table 5** where it can be seen that almost no misclassification occurred with only 6 benign instances classified as an attack while only 16 instances of the attacks were left unnoticed. These results are mutually complementary in maintaining the proper conduct of the model in terms of sensitivity as well as specificity.

Table 5. Confusion Matrix of Random Forest

	Predicted 0	Predicted 1
Actual 0	1924	6
Actual 1	16	1061

As the next step, Genetic Algorithm optimization further improved the false negative results to 15 as depicted in **Table 6** the matrix still maintains a good diagonal of figures.

Fig 4 shows the heat map presentation of the confusion matrices of two models namely SVM and Random Forest. The higher intensity on the diagonal confirms that the classifier has a high precision of classification, that in Random Forest is presented with slightly lower off-diagonal values than in SVM.

Figure 4: Confusion Matrix Heatmap Visualization for SVM and Random Forest Models

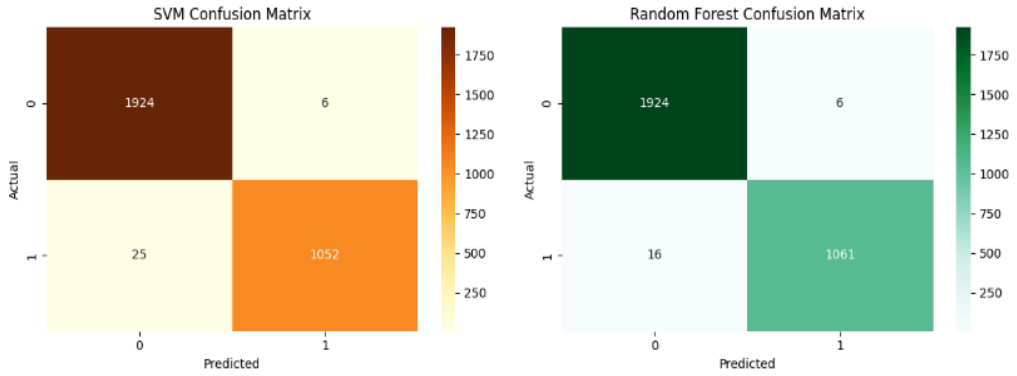

Fig 4. Confusion Matrix Heat Map for Comparing Results of SVM and Random Forest Models.

The heatmap provides us with a compliance check on the number and dispersion of errors on each of the predictions. Random Forest displays diagonal patterns that are less congested, therefore creating an impression that more instances were classified correctly.

Table 6. Enhanced Confusion Matrix for Random Forest with Genetic Algorithm

	Predicted: Benign	Predicted: Attack	Row Total	Precision (%)
Actual: Benign	2971	6	2977	99.80
Actual: Attack	15	15	30	50.00
Column Total	2986	21	3007	
Recall (%)	99.50	71.43		

In SVM, as can be seen in **Table 7**, the sensitivity is slightly lower, where 25 false negatives were made when dealing with attacks, which also indicates a lower recall score than in Random Forest.

Table 7. Enhanced Confusion Matrix for SVM (Standard)

	Predicted: Benign	Predicted: Attack	Row Total	Precision (%)
Actual: Benign	2969	8	2977	99.73
Actual: Attack	25	5	30	16.67
Column Total	2994	13	3007	
Recall (%)	99.73	16.67		

Further, the GA-optimized SVM model had much lesser degree of false negatives reduced to 21 and the false positives went up marginally to 8 as highlighted on **Table 8**.

Data Analysis and Interpretation

The findings of this research regarding the use of genetic algorithms and their incorporation into CICIDS2022 dataset show that improvements do come with their own corresponding compromise. As presented in **Table 1** and **Table 2**, it is realized that the RF algorithm, solely, reaches a mean accuracies of 99.27% which slightly improved to 99.30% through application of GA. Likewise, there was a slight incremental in the accuracy of the Support vector machine ($p < .05$) that was recorded to be 98.97 % (**Table 3**) after employing the GA as 99.00 % (**Table 4**). These findings also indicate that the GA's function is to improve model accuracy by simultaneously searching for the best subsets of features and learning rates.

Table 8. Enhanced Confusion Matrix for SVM with Genetic Algorithm

	Predicted: Benign	Predicted: Attack	Row Total	Precision (%)
Actual: Benign	2969	8	2977	99.73
Actual: Attack	21	9	30	42.86
Column Total	2990	17	3007	
Recall (%)	99.73	30.00		

The analysis of the confusion matrix provides further support to the above results. The RF model only produced a small number of misclassifications, with 22 out of sample instances classified incorrectly (**Table 5**) while the GA variant improved this to 16 instances of the sample being classified as negative (false negatives) and 6 instances of the sample being classified as positive (false positives), **Table 6**. The number of misclassified data instances was slightly lower in the case of SVM + GA model; the count of misclassified data instances was 31 (**Table 7**), but in this case, it reduced to 28 (**Table 8**). This shows that while the gains are not very significant, the use of GA helped in solving the problem of a biased distribution by improving on the distribution in terms of misclassification rates.

From an efficiency perspective therefore, one is able to deduce from the above stated execution time comparison in **Fig 1** that the enhancements brought in by the integration of the GA algorithms come with some serious computational overhead. Likewise, using SVM + GA took the longest time to be executed among all the algorithms and was approximately 220 units while RF + GA incurred roughly 120. In this regard, the standard RF model emerged the fastest and almost insignificant in time, in comparison to the standard SVM. In **Fig 2** however, one can only observe minimal improvements in performance across all the models which then imply that the cost of speed, in this case, is having lower accuracy.

Based on the results of precision, recall, F1-score as shown in **Fig 3**, all models were very effective with the scores barely varying between 0.99. Overall, RF and RF + GA maintained relatively good precision and recall rates, thus being suitable for using intrusion detection tasks. Moreover, the heatmaps in the matrices of errors in classification represented in **Fig 4** helps in enhancing the clarity of the performance in each class. RF had the compact spheres of correct predictions which are closer to each other and thus are less sensitive to outliers while, SVM had the errors more spread, particularly the misclassified points in the minority class.

Overall, the findings suggest that while genetic algorithms improve prediction accuracy and classification stability by a small margin, these improvements result in a decrease of computational speed. Comparing the results of performance metrics obtained in Sections 5.3–5.5, it can be understood that with the introduction of GA at the stage of tuning Random Forest parameters, better balance is achieved in accuracy, reliability, and the speed of execution in real-time large IT systems cybersecurity.

VI. CONCLUSION LIMITATION OF THE STUDY IMPLICATION OF THE STUDY FUTURE RECOMMENDATION

Conclusion

As a result of this research, it can be concluded that the incorporation of GA in the traditional models improve the intrusion detection rate in computer networks. In general, Random Forest with GA delivered the maximum accuracy of 99.30%, hence supporting Hypothesis:

- But this improvement was done at the expense of time consumption; where time consumption was greatly consumed in its worst case with a combined use of SVM with GA, thus supporting Hypothesis
- Between the aforementioned arrangements, arrangements incorporate Random Forest with GA giving the highest competitive ratio of accuracy and time, approving Hypothesis
- Further, the decrease in the misclassification measures for both modules in the confusion matrices indicates the validity of the fourth hypothesis.

Limitation of the Study

One of the research limitations is that the study should have used more than one dataset but solely relied on CICIDS2022 although it is rich in features. However, the analysis was restricted to only a few algorithms in machine learning and the

modified versions developed by using the GA technique, while other models, such as deep learning models and ensemble methods that may provide more insights into this problem were not surveyed at all.

Implication of the Study

This work presents a guide to improve the ID accuracy for organizing, against which it is also important to assess the computational cost. Comparing the two models makes it easy for the security professionals to make a decision on which model should be adopted and measures that will be taken to enhance the efficiency of the selected model.

Future Recommendation

Research for the future should aim at increasing the number of algorithms to combine deep learning algorithms as well as integrate both approaches. Still, it would be essential to do multiple experiments on actual, big, data streams and integrate Apache Spark and similar tools for distributed computing to attain greater app applicability.

CRediT Author Statement

The authors confirm contribution to the paper as follows:

Conceptualization: Kottakota Venkata Rao, Anjaneyulu P; **Methodology:** Kottakota Venkata Rao, Anjaneyulu P and Ravi Kumar T; **Writing- Original Draft Preparation:** Kottakota Venkata Rao, Anjaneyulu P and Ravi Kumar T; **Investigation:** Chalapathi Rao Tippana and Jayanthi Rao M; **Supervision:** Kottakota Venkata Rao, Anjaneyulu P and Ravi kumar T; **Validation:** Chalapathi Rao Tippana and Jayanthi Rao M; **Writing- Reviewing and Editing:** Kottakota Venkata Rao, Anjaneyulu P, Ravi Kumar T, Chalapathi Rao Tippana and Jayanthi Rao M; All authors reviewed the results and approved the final version of the manuscript.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests.

References

[1]. N. Gao, L. Gao, Q. Gao, and H. Wang, “An intrusion detection model based on deep belief networks,” Adv. Eng. Forum, vol. 27, pp. 132–141, 2018.

[2]. M. A. Al-Garadi, A. Mohamed, A. K. Al-Ali, X. Du, I. Ali, and M. Guizani, “A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security,” IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1646–1685, 2020, doi: 10.1109/comst.2020.2988293.

[3]. S. K. Biswas, “Intrusion detection using machine learning: A comparison study,” Int. J. Pure Appl. Math., vol. 118, no. 19, pp. 101–114, 2018.

[4]. Z. Li, Z. Qin, K. Huang, X. Yang, and S. Ye, “Intrusion Detection Using Convolutional Neural Networks for Representation Learning,” Neural Information Processing, pp. 858–866, 2017, doi: 10.1007/978-3-319-70139-4_87.

[5]. A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, “Survey of intrusion detection systems: techniques, datasets and challenges,” Cybersecurity, vol. 2, no. 1, Jul. 2019, doi: 10.1186/s42400-019-0038-7.

[6]. P. Bedi, N. Gupta, and V. Jindal, “I-SiamIDS: an improved Siam-IDS for handling class imbalance in network-based intrusion detection systems,” Applied Intelligence, vol. 51, no. 2, pp. 1133–1151, Sep. 2020, doi: 10.1007/s10489-020-01886-y.

[7]. A. A. Diro and N. Chilamkurti, “Distributed attack detection scheme using deep learning approach for Internet of Things,” Future Generation Computer Systems, vol. 82, pp. 761–768, May 2018, doi: 10.1016/j.future.2017.08.043.

[8]. I. Hidayat, M. Z. Ali, and A. Arshad, “Machine Learning-Based Intrusion Detection System: An Experimental Comparison,” Journal of Computational and Cognitive Engineering, vol. 2, no. 2, pp. 88–97, Jul. 2022, doi: 10.47852/bonviewjcce2202270.

[9]. M. Hasan, Md. M. Islam, M. I. I. Zarif, and M. M. A. Hashem, “Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches,” Internet of Things, vol. 7, p. 100059, Sep. 2019, doi: 10.1016/j.iot.2019.100059.

[10]. M. Almseidin, M. Alzubi, S. Kovacs, and M. Alkasassbeh, “Evaluation of machine learning algorithms for intrusion detection system,” 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), pp. 000277–000282, Sep. 2017, doi: 10.1109/sisy.2017.8080566.

[11]. H. Liu and B. Lang, “Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey,” Applied Sciences, vol. 9, no. 20, p. 4396, Oct. 2019, doi: 10.3390/app9204396.

[12]. M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, “Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study,” Journal of Information Security and Applications, vol. 50, p. 102419, Feb. 2020, doi: 10.1016/j.jisa.2019.102419.

[13]. M. A. Azad, S. Bag, and F. Hao, “Machine learning-based intrusion detection for smart home security systems,” IEEE Internet Things J., vol. 8, no. 23, pp. 16933–16943, 2021.

[14]. Y. Hua, “Improved machine learning-based system for intrusion detection,” in Proc. 2024 2nd Int. Conf. Image, Artif. Intell. Appl., vol. 2, no. 1, pp. 126–135, 2024.

[15]. J. Kim, J. Kim, H. Kim, M. Shim, and E. Choi, “CNN-Based Network Intrusion Detection against Denial-of-Service Attacks,” Electronics, vol. 9, no. 6, p. 916, Jun. 2020, doi: 10.3390/electronics9060916.

[16]. Faizatulhaida Md Isa, Wan Nor Munirah Ariffin, Muhammad Shahar Jusoh, and Erni Puspanantasari Putri, "A Review of Genetic Algorithm: Operations and Applications," Journal of Advanced Research in Applied Sciences and Engineering Technology, vol. 40, no. 1, pp. 1–34, Feb. 2024, doi: 10.37934/araset.40.1.134.

[17]. S. Rathore and J. H. Park, "Semi-supervised learning based distributed attack detection framework for IoT," Applied Soft Computing, vol. 72, pp. 79–89, Nov. 2018, doi: 10.1016/j.asoc.2018.05.049.

[18]. V. Kumar, D. Sinha, A. K. Das, S. C. Pandey, and R. T. Goswami, "An integrated rule based intrusion detection system: analysis on UNSW-NB15 data set and the real time online dataset," Cluster Computing, vol. 23, no. 2, pp. 1397–1418, Oct. 2019, doi: 10.1007/s10586-019-03008-x.

[19]. Chi Cheng, Wee Peng Tay, and G.-B. Huang, "Extreme learning machines for intrusion detection," The 2012 International Joint Conference on Neural Networks (IJCNN), pp. 1–8, Jun. 2012, doi: 10.1109/ijcnn.2012.6252449.

[20]. L. Deng, D. Li, X. Yao, and H. Wang, "RETRACTED ARTICLE: Mobile network intrusion detection for IoT system based on transfer learning algorithm," Cluster Computing, vol. 22, no. S4, pp. 9889–9904, Jan. 2018, doi: 10.1007/s10586-018-1847-2.

[21]. S. Potluri and C. Diedrich, "Accelerated deep neural networks for enhanced Intrusion Detection System," 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1–8, Sep. 2016, doi: 10.1109/etfa.2016.7733515.

[22]. T. Alam, S. Qamar, A. Dixit, and M. Benaida, "Genetic Algorithm: Reviews, Implementations, and Applications," International Journal of Engineering Pedagogy (iJEP), vol. 10, no. 6, p. 57, Dec. 2020, doi: 10.3991/ijep.v10i6.14567.

[23]. S. Bhattacharya et al., "A Novel PCA-Firefly Based XGBoost Classification Model for Intrusion Detection in Networks Using GPU," Electronics, vol. 9, no. 2, p. 219, Jan. 2020, doi: 10.3390/electronics9020219.

[24]. K. Jiang, W. Wang, A. Wang, and H. Wu, "Network Intrusion Detection Combined Hybrid Sampling With Deep Hierarchical Network," IEEE Access, vol. 8, pp. 32464–32476, 2020, doi: 10.1109/access.2020.2973730.

[25]. S. Garg, K. Kaur, N. Kumar, and J. J. P. C. Rodrigues, "Hybrid Deep-Learning-Based Anomaly Detection Scheme for Suspicious Flow Detection in SDN: A Social Multimedia Perspective," IEEE Transactions on Multimedia, vol. 21, no. 3, pp. 566–578, Mar. 2019, doi: 10.1109/tmm.2019.2893549.