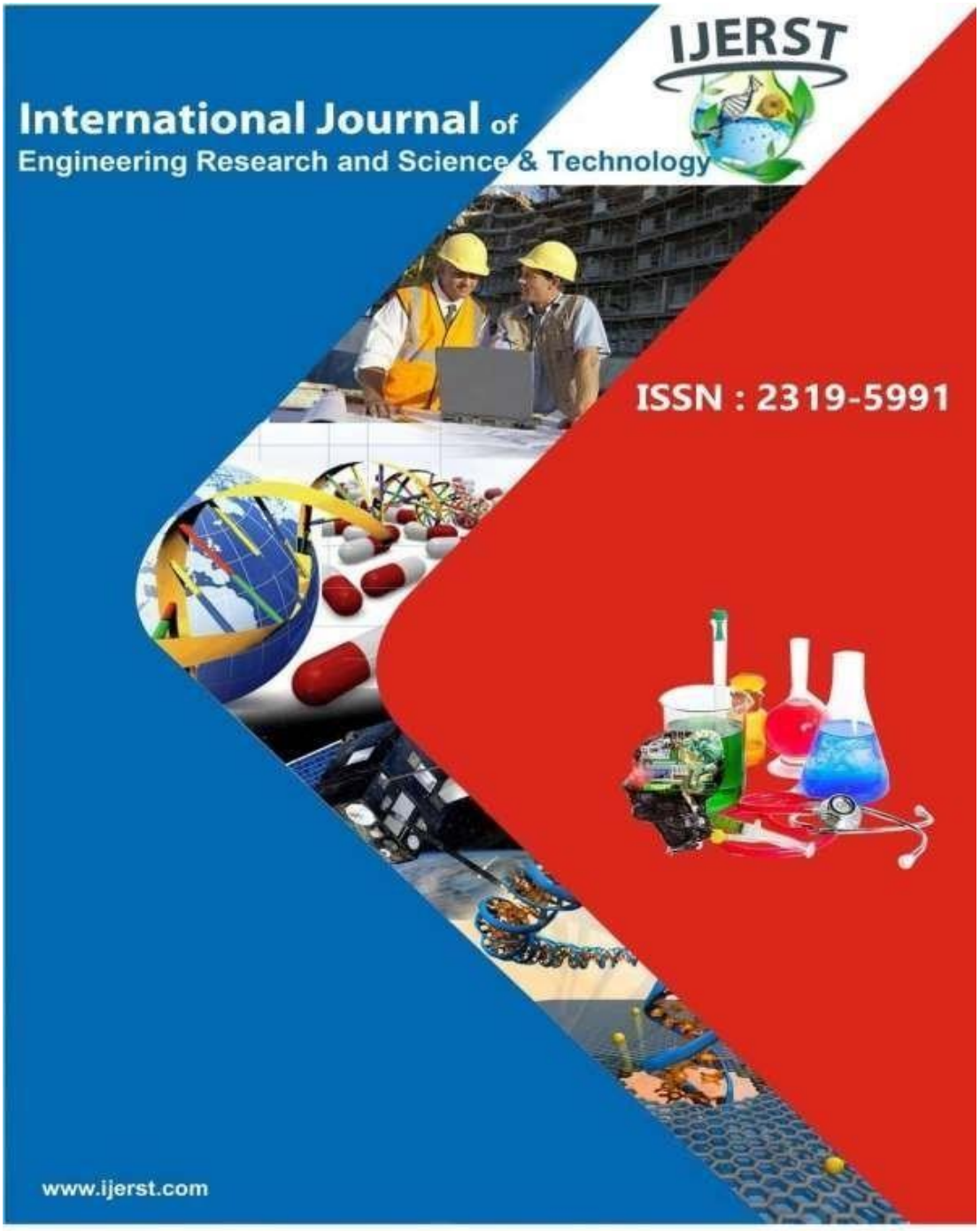


Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 18, Issuse.4,Oct 2022

CLASSIFYING PREDICTING DDOS ATTACKS USING MACHINE LEARNING

Mr K.Yakhoob[1],K. Vaishnavi[2],K.Sandhya Rani[3],B.Kavya[4]

[1]Assistant professor, Department of CSE, Princeton College of engineering and technology for women Narapally vijayapuri colony ghatkesar mandal, Pin code-500088

[2,3,4]UG Students,Department of CSE, Princeton College of engineering and technology for women Narapally vijayapuri colony ghatkesar mandal, Pin code-500088

Article Info

Received: 09-08-2022 *Revised: 10 -09-2022* *Accepted: 22-10-2022*

ABSTRACT

Distributed Denial of Service (DDoS) attacks pose a significant threat to network security, targeting critical infrastructure and disrupting services. Traditional methods of mitigating DDoS attacks rely on signature-based detection and traffic filtering, which may not be effective against sophisticated and evolving attack techniques. In this project, we propose a novel approach for classifying and predicting DDoS attacks using machine learning algorithms. By analyzing network traffic patterns and extracting relevant features, such as packet headers, flow characteristics, and payload content, we train machine learning models to differentiate between normal traffic and DDoS attacks. Additionally, we develop predictive models to forecast potential DDoS attacks based on historical data and real-time network monitoring. Our experimental results demonstrate the effectiveness of the proposed approach in accurately classifying and predicting DDoS attacks, thereby enhancing network security and enabling proactive defense strategies.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 18, Issuse.4,Oct 2022

.INTRODUCTION

The proliferation of Distributed Denial of Service (DDoS) attacks presents a formidable challenge to network security, posing a significant threat to the availability and integrity of online services. These attacks leverage a multitude compromised devices to flood target systems with malicious traffic, resulting in service disruptions and financial losses for organizations. Traditional methods of mitigating DDoSdetect and prevent DDoS attacks. In response to this challenge, this project proposes a novel approach for classifying and predicting DDoS attacks using machine learning techniques. By harnessing the power of machine learning algorithms, we aim to analyze network traffic patterns and extract relevant features that can distinguish between normal traffic and DDoS attack traffic. Moreover, we seek to develop predictive models that can anticipate potential DDoS attacks based on historical data and real-time network monitoring. This proactive approach to DDoS defense holds the promise of enhancing network security and enabling organizations to mitigate the impact of DDoS attacks more effectively. Through this project, we aim to contribute to the advancement of DDoS defense strategies by leveraging the capabilities of machine learning to detect and predict DDoS attacks with greater accuracy and efficiency. By developing robust classification and prediction models, we endeavor to empower organizations to safeguard their network infrastructure against the growing threat posed by DDoS attacks.

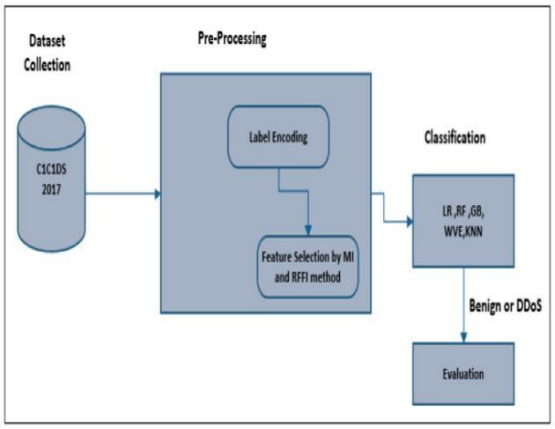


Fig :Block diagram

II.EXISTING PROBLEM

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 18, Issuse.4,Oct 2022

Traditional methods of mitigating Distributed Denial of Service (DDoS) attacks, such as signature-based detection and traffic filtering, are increasingly ineffective against the sophisticated tactics employed by attackers. These methods often rely on predefined patterns or known attack signatures, making them susceptible to evasion techniques and zero-day attacks. Moreover, the reactive nature of these approaches means that organizations are often unable to detect and respond to DDoS attacks until after they have already caused significant damage, resulting in service disruptions and financial losses.

III. PROPOSED SOLUTION

To address the limitations of existing DDoS mitigation methods, our proposedsolution leverages machine learning techniques for proactive and accurate detection of DDoS attacks. By training machine learning algorithms on large volumes of network traffic data, we can develop robust models capable of identifying anomalous patterns associated with DDoS attacks in real time. These models can learn to differentiate between normal traffic patterns and malicious DDoS attack traffic, enabling organizations to detect and respond to attacks more effectively.

Furthermore, our solution includes the implementation of predictive analytics algorithms that can anticipate potential DDoS attacks based on historical data and emerging trends. By analyzing patterns and trends in network traffic behavior, these predictive models can provide early warnings of impending DDoS attacks, allowing organizations to take proactive measures to mitigate their impact. Additionally, our solution incorporates adaptive learning capabilities, enabling the models to continuously evolve and adapt to new attack techniques and variations.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

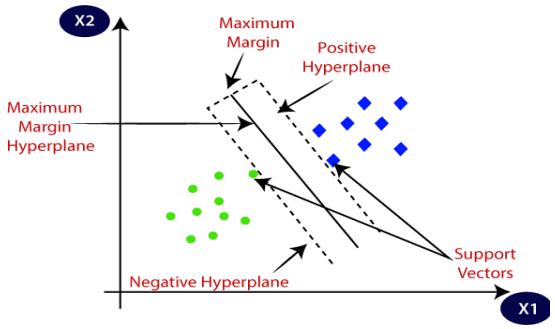
Vol. 18, Issuse.4,Oct 2022

IV. ALGORITHMS

Random ForestGiven that the random forest uses a combination of trees to make its classification predictions, it is likely that some decision trees will provide the right result while others would not. However, after combining all of the trees, the proper result is predicted. Two such assumptions for an improved Random forest classifier are shown below. So that the classifier can make a reliable prediction, rather than a guess, there should be some real values in the feature variable of the dataset. There can't be much overlap between the trees' forecasts.

Applications of Random Forest

Random forest was largely used in the following four places:The financial sector relies heavily on this method to identify potential funding risks.In the field of medicine, this method may be used to detect disease trends and assess potential dangers associated with such trends.Using this method, we may pinpoint areas with a similar land use pattern.Advertising and marketing: This formula may be used to identify common trends in both fields.Gains from Unplanned Forestation Category and Regression tasks are both feasible for Random Forest to complete.It is capable of handling large, dimensional datasets.It improves the layout's precision and eliminates the overfitting issue. Challenges posed by Unplanned Forestry Even though arbitrary forest may be used for both classification and regression tasks, it performs poorly for the latter. SVM Support Popular solutions for Monitored Knowing include the Support Vector Machine (SVM), which may be utilised to solve both classification and regression issues. However, in Machine Learning, it is often used for Category issues.



The SVM formula's goal is to find the optimal line or decision boundary that can divide the n-dimensional space into classes, making it easier to later assign the new data element to the proper category. A hyperplane defines the narrowest feasible set of options.

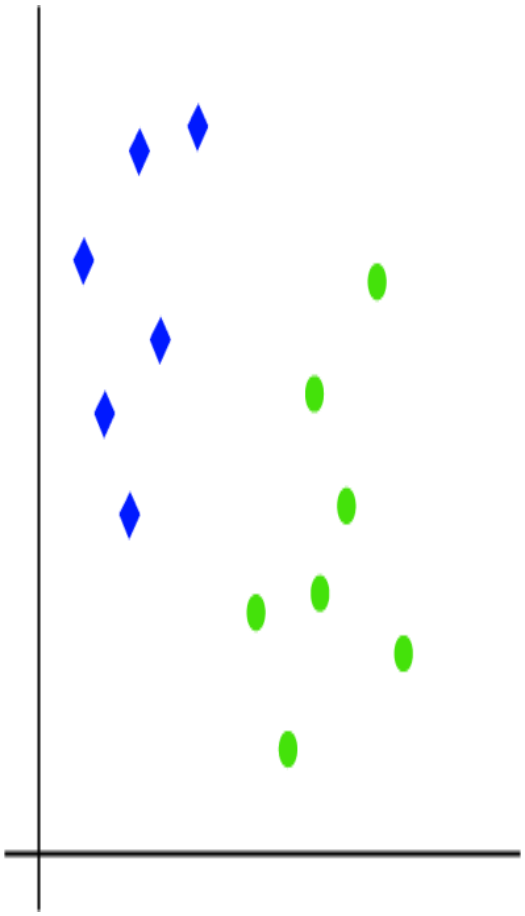
Face recognition, picture classification, and text classification are just few of the uses for the SVM algorithm.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 18, Issuse.4,Oct 2022

There are two distinct types of SVM:



Linear Support Vector Machines: If a dataset can be split into two groups using a single straight line, we say that the data is linearly separable, and we use a classifier known as Direct SVM for this kind of information.

When a dataset cannot be categorised along a straight line, it is said to be non-linear, and the classifier used to categorise it is known as a non-linear support vector machine (SVM).

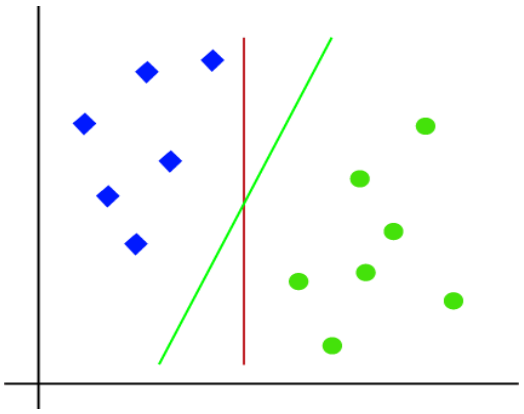
Direct SVM: The SVM formula's operation may be grasped by an illustration. Consider a data collection labelled "green" and "blue," with corresponding functions labelled "x1" and "x2." We need a classifier that can decide if a pair of collaborators (x1, x2) should be labelled as green or blue. Think about the diagram below:

So as it is 2-d space so by just using a straight line, we can easily separate these two classes. But there can be multiple lines that can separate these classes. Consider the below image:

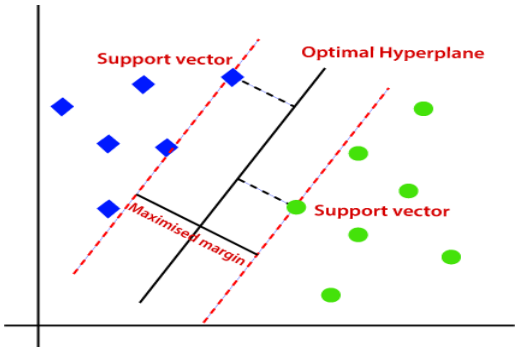
International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 18, Issuse.4,Oct 2022

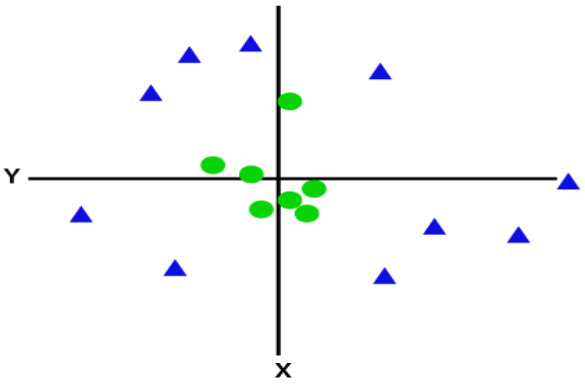


This optimum line or boundary for making a choice is known as a hyperplane, and it may be located with the help of the SVM formula. The SVM method locates the intersection of the two paths that is most closely aligned. We refer to these points as "support vectors." Margin refers to the space outside of the hyperplane that the vectors occupy. SVM's goal is to maximise this profit margin. The ideal hyperplane is the one that has the greatest possible margin.



Non-Linear SVM:

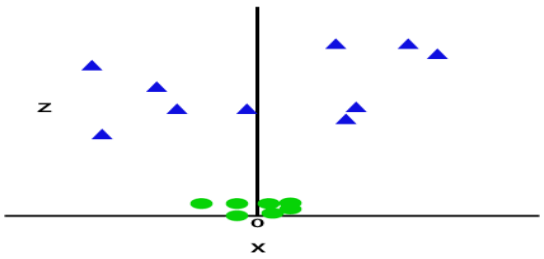
If data is linearly arranged, then we can separate it by using a straight line, but for non-linear data, we cannot draw a single straight line. Consider the below image:



So to separate these data points, we need to add one more dimension. For linear data, we have used two dimensions x and y, so for non-linear data, we will add a third dimension z. It can be calculated as:

$z=x^2+y^2$

By adding the third dimension, the sample space will become as below image:

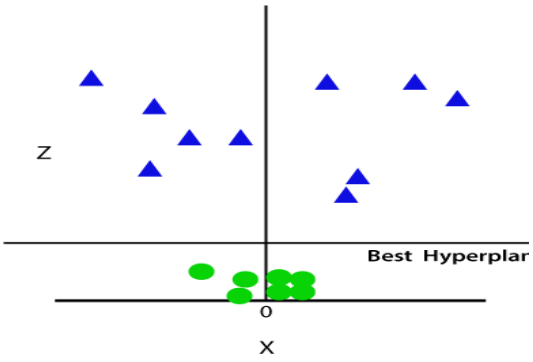


So now, SVM will divide the datasets into classes in the following way. Consider the below image:

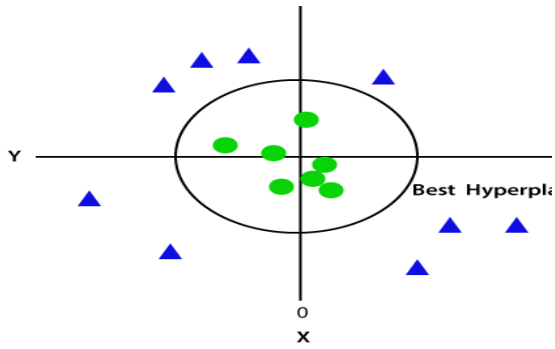
International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 18, Issuse.4,Oct 2022



Since we are in 3-d Space, hence it is looking like a plane parallel to the x-axis. If we convert it in 2d space with z=1, then it will become as:



Hence we get a circumference of radius 1 in case of non-linear data.

LR

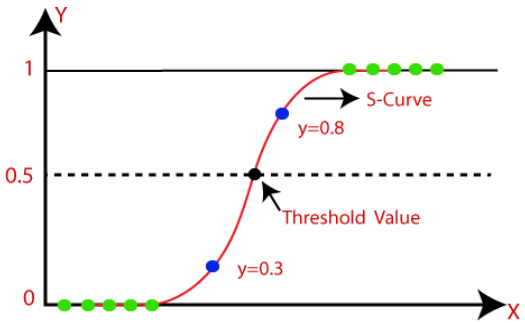
One of the most well-known AI algorithms, logistic regression is a part of the Managed Discovering methodology. Using a given collection of independent factors, it may make predictions about the categorical dependent variable.

The outcome of a dependent variable of interest may be predicted using logistic regression. Therefore, the final output must be a single, unambiguous number.

and 1, it gives the probabilistic values that fall between those two extremes, such as "Yes" and "No" or "0" and "1," etc.

Outside of their respective applications, Linear Regression and Logistic Regression are quite similar. Regression problems may be solved with direct regression, whereas category problems using logistic regression.

Instead of a straight line, the "S" shaped logistic function is fitted in logistic regression. This function expects two possible values, either 0 or 1.



Note: Logistic regression uses the concept of predictive modeling as regression; therefore, it is called logistic regression, but is used to classify samples; Therefore, it falls under the classification algorithm.

Assumptions for Logistic Regression:

The dependent variable must be categorical in nature.

The independent variable should not have multi-collinearity.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 18, Issuse.4,Oct 2022

It is simple to understand as it follows the same process which a human follow while making any decision in real-life.

It can be very useful for solving decision-related problems.

It helps to think about all the possible outcomes for a problem.

There is less requirement of data cleaning compared to other algorithms.

Disadvantages of the Decision Tree

The decision tree contains lots of layers, which makes it complex.

It may have an overfitting issue, which can be resolved using the Random Forest algorithm.

For more class labels, the computational complexity of the decision tree may increase.

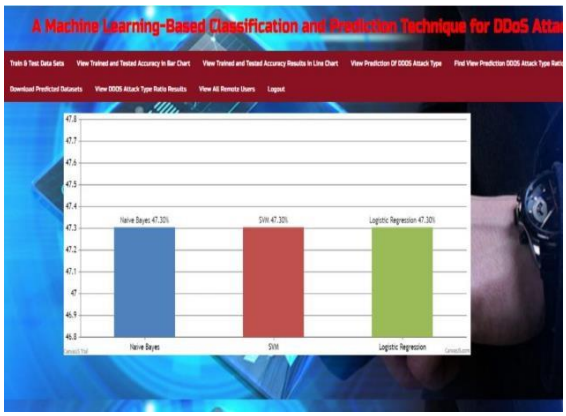


Fig.1. Output results.

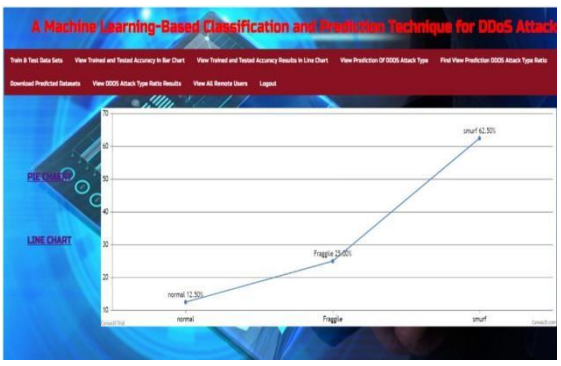


Fig.2. Output graphs.

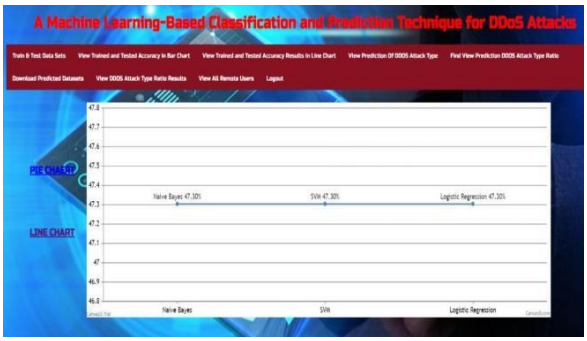


Fig.3. Accuracy levels.

Fig.4. DDos Attack detection.

V.CONCLUSIONS

Finding evidence of a distributed denial of service attack is a common challenge. Lack of cloud solution is triggered by this kind of attack, hence detection is essential. This kind of attack may be identified using a machine learning model. The purpose of this research is to locate a DDoS attack that is very effective. The CICDDoS 2019 and CICIDS 2017 datasets were used in this study. Experiments used a variety of information from both databases linked to DDoS attacks. We use both the MI and RFFI methods to determine which functions are most important. Machine

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 18, Issuse.4,Oct 2022

KNN, LR) are then fed the selected functions. When compared to other methods, RF's overall prediction accuracy of 0.99993 when using 16 functions and 0.999977 when using 19 features is superior. Using MI and RFFI as attribute choosing strategies, we find that RF, GB, WVE, KNN, and LR all perform quite well. Future work on DDoS and other strike detection could use semantic networks and wrapper function choosing techniques like sequential function selection.

VI. REFERENCES

1. Malik, N.; Sardaraz, M.; Tahir, M.; Shah, B.; Ali, G.; Moreira, F. Energy-efficient load balancing algorithm for workflow scheduling in cloud data centers using queuing and thresholds. Appl. Sci. 2021, 11, 5849.

2. Yan, Q.; Yu, F.R. Distributed denial of service attacks in software-defined networking with cloud computing. IEEE Commun. Mag. 2015, 53, 52–59.

3. Lau, F.; Rubin, S.H.; Smith, M.H.; Trajkovic, L. Distributed denial of service attacks. In Proceedings of the SMC 2000 Conference Proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics.'Cybernetics Evolving to Systems, Humans, Organizations, and Their Complex Interactions'(Cat. No. 0), Nashville, TN, USA, 8–11 October 2000;

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 18, Issuse.4,Oct 2022

IEEE: Piscataway, NJ, USA, 2000; Volume 3, pp. 2275–2280.

4. Sambangi, S.; Gondi, L. A Machine Learning Approach for DDoS (Distributed Denial of Service) Attack Detection Using Multiple Linear Regression. Proceedings 2020, 63, 51.

5. Erickson, B.J.; Korfiatis, P.; Akkus, Z.; Kline, T.L. Machine learning for medical imaging. Radiographics 2017, 37, 505–515.

6. Hasan, A.; Moin, S.; Karim, A.; Shamshirband, S. Machine learning-based sentiment analysis for twitter accounts. Math. Comput. Appl. 2018, 23, 11.

7. Malik, S.; Tahir, M.; Sardaraz, M.; Alourani, A. A Resource Utilization Prediction Model for Cloud Data Centers Using Evolutionary Algorithms and Machine Learning Techniques. Appl. Sci. 2022, 12, 2160.

8. Aljamal, I.; Tekeoğlu, A.; Bekiroglu, K.; Sengupta, S. Hybrid intrusion detection system using machine learning techniques in cloud computing environments. In Proceedings of the 2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA), Honolulu, HI, USA, 29–31 May 2019; IEEE:

9. Piscataway, NJ, USA, 2019; pp. 84–89. Kushwah, G.S.; Ranga, V. Optimized extreme learning machine for detecting

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 18, Issuse.4,Oct 2022

DDoS attacks in cloud computing. Comput. Secur. 2021, 105, 102260.

10. Makuvaza, A.; Jat, D.S.; Gamundani, A.M. Deep Neural Network (DNN) Solution for Real-time Detection of Distributed Denial of Service (DDoS) Attacks in Software Defined Networks (SDNs). SN Comput. Sci. 2021, 2, 1–10.

11. Manimurugan, S.; Al-Mutairi, S.; Aborokbah, M.M.; Chilamkurti, N.; Ganesan, S.; Patan, R. Effective attack detection in internet of medical things smart environment using a deep belief neural network. IEEE Access 2020, 8, 77396–77404.

12. Intrusion Detection Evaluation Dataset (CIC-IDS2017). Available online: https://www.unb.ca/cic/datasets/ids-2017.html (accessed on 30 September 2021).

DDoS Evaluation Dataset (CIC-DDoS2019). Available online: https://www.unb.ca/cic/datasets/ddos-2019.html (accessed on 27 April 2022).

13.Khan, S.; Kifayat, K.; Kashif Bashir, A.; Gurtov, A.; Hassan, M. Intelligent intrusion detection system in smart grid using computational intelligence and machine learning. Trans. Emerg. Telecommun. Technol. 2021, 32, e4062.

14. Sandhu, R.S.; Samarati, P. Access control: Principle and practice. IEEE Commun. Mag. 1994, 32, 40–48.

15. Khan, M.S.; Khan, N.M.; Khan, A.; Aadil, F.; Tahir, M.; Sardaraz, M. A low-complexity, energy-efficient data securing model for wireless sensor network based on linearly complex voice encryption mechanism of GSM technology. Int. J. Distrib. Sens. Netw. 2021, 17, 15501477211018623.

16. Sardaraz, M.; Tahir, M. SCA-NGS: Secure compression algorithm for next generation sequencing data using genetic operators and block sorting. Sci. Prog. 2021, 104, 00368504211023276.

17. Zhong, Z.; Xu, M.; Rodriguez, M.A.; Xu, C.; Buyya, R. Machine Learning-based Orchestration of Containers: A Taxonomy and Future Directions. ACM Comput. Surv. (CSUR) 2021.

18. Bindra, N.; Sood, M. Detecting DDoS attacks using machine learning techniques and contemporary intrusion detection dataset. Autom. Control. Comput. Sci. 2019, 53, 419–428.

19. Kshirsagar, D.; Kumar, S. An efficient feature reduction method for the detection of DoS attack. ICT Express 2021, 7, 371–375.