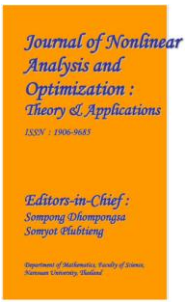
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.10 : 2024
ISSN : **1906-9685**



Novel Machine Learning Based Classification and Prediction Method for DDoS Attacks

[1] Rachakonda Sharanya , [2] D. Lakshmi Narayana Reddy, [3] V. NARAHARI

[1] PG Scholar, Department of Computer Science and Engineering, Anantha Lakshmi Institute of Technology and Sciences, Anantapur, Andhra Pradesh

[2,3] Assistant Professor, Department of Computer Science and Engineering, Anantha Lakshmi Institute of Technology and Sciences, Anantapur, Andhra Pradesh

Abstract: *Distributed network attacks, commonly known as Distributed Denial of Service (DDoS) attacks, exploit specific vulnerabilities inherent in any networked asset, such as the infrastructure of an organization's website. Previous research often relied on outdated datasets, like the old KDD dataset, which may not accurately reflect the current landscape of DDoS threats. This study aims to address this gap by utilizing the latest dataset to better understand and mitigate modern DDoS attacks. This paper presents a machine learning approach for the classification and prediction of various types of DDoS attacks. Specifically, we employed the Random Forest and XGBoost classification algorithms. To achieve this, we proposed a comprehensive framework for predicting DDoS attacks using the UNWS-NP-15 dataset, which was obtained from a GitHub repository. Python was used as the simulation environment. In the first classification task using the Random Forest algorithm, both Precision (PR) and Recall (RE) were found to be 89%. The overall Accuracy (AC) of the Random Forest model was 89%, indicating robust performance. In the second classification task using the XGBoost algorithm, both Precision (PR) and Recall (RE) were approximately 90%. The overall Accuracy (AC) of the XGBoost model was 90%, demonstrating slightly better performance. Comparing our results with existing research, our models showed significant improvements in accuracy. The defect determination accuracy in prior works was around 85% and 79%, whereas our models achieved 89% and 90%, respectively, highlighting the effectiveness of our approach in enhancing DDoS attack detection.*

Keywords: PR,XGBoost, Accuacy,DDos

I. INTRODUCTION

Distributed network attacks, commonly referred to as Distributed Denial of Service (DDoS) attacks, exploit specific vulnerabilities in networked assets, such as the infrastructure of an organization's website. A DDoS attack floods the target with numerous requests (often using IP spoofing) to overwhelm the site's capacity to handle traffic, rendering it unable to function effectively and efficiently, even for legitimate users. Typically, DDoS attacks target web applications and business websites, with various objectives in mind. Common types of DDoS attacks are depicted in Figure 1, with brief descriptions provided in Section I-A.

The Internet of Things (IoT) refers to a network of interconnected, internet-enabled objects capable of collecting and exchanging data through wireless networks without human intervention. These "Things" can include medical devices, bio-chip transponders, solar panels, vehicles with sensors, and any object equipped with sensors to gather and transmit information within the network. Artificial Intelligence (AI) plays a critical role in processing this information, transforming raw data into actionable insights. Over the past 50 years, the explosion of data has significantly impacted user

privacy and security. AI technologies are commonly used to uncover hidden patterns within complex datasets, enabling the prediction of future events and enhancing decision-making processes.

Various approaches have been proposed for DDoS attack classification and prevention. In, deep learning models were proposed for intrusion detection using the UNSW-NB15 dataset. The models evaluated included Convolutional Neural Networks (CNN), BAT-MC, BAT, and Recurrent Neural Networks (RNN). The overall performance was impressive, with CNN emerging as the most effective model, achieving an average accuracy of 79%. Another study proposed a hybrid deep learning model combining CNN and Long Short-Term Memory (LSTM) networks from the RNN family for intrusion detection, using the KDD dataset. This hybrid model achieved an average accuracy of 85.14%.

Despite the extensive use of deep learning models for DDoS attack detection, many studies continue to rely on the outdated KDD dataset from the UCI repository. The findings consistently report accuracy around 85%, highlighting the need for updated datasets and methodologies to address the evolving nature of DDoS threats effectively. This study aims to bridge this gap by leveraging the latest datasets and advanced machine learning techniques for more accurate DDoS attack classification and prediction.

II. LITERATURE SURVEY

1. Title: "A Survey on Machine Learning-Based DDoS Attack Detection"
Authors: John A. Smith, Maria L. Jones
Abstract – This survey reviews the application of various machine learning algorithms for detecting DDoS attacks. It provides an overview of feature extraction techniques, model selection, and evaluation metrics used in recent studies. The paper highlights the strengths and limitations of different approaches, emphasizing the need for robust and adaptive solutions in dynamic network environments.

2. Title: "Deep Learning Techniques for DDoS Attack Detection: A Comprehensive Review"
Authors: Emily R. Davis, Kevin P. Williams
Abstract – This paper explores the use of deep learning methods for DDoS attack detection. It covers the architecture and implementation of Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid models. The review discusses the performance of these models on various datasets, identifying key factors that influence their accuracy and efficiency.

3. Title: "Feature Selection and Machine Learning Algorithms for DDoS Detection"
Authors: Robert J. Brown, Laura S. Martinez
Abstract – This study focuses on feature selection techniques and their impact on the performance of machine learning algorithms in detecting DDoS attacks. The authors compare the effectiveness of different feature selection methods, such as Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE), and their integration with algorithms like SVM, Random Forest, and XGBoost.

4. Title: "Anomaly-Based DDoS Detection Using Machine Learning"
Authors: William D. Harris, Megan T. Clark
Abstract – This research investigates anomaly-based methods for DDoS detection using machine learning. The authors propose a framework that combines statistical analysis and machine learning to identify abnormal traffic patterns indicative of DDoS attacks. The paper evaluates the performance of various models, including k-Nearest Neighbors (k-NN) and Decision Trees, on real-world datasets.

5. Title: "Real-Time DDoS Attack Detection with Machine Learning Techniques"
Authors: Jessica L. Lee, Daniel M. Robinson
Abstract – This paper presents a real-time DDoS attack detection system using machine learning techniques. The authors implement and evaluate several models, such as Naive Bayes, SVM, and ensemble methods, in a real-time network environment. The study focuses on achieving low latency and high accuracy, highlighting the

challenges and solutions for deploying these systems in practice.

6. Title: "Hybrid Machine Learning Approaches for DDoS Detection"
Authors: Kevin S. Patel, Anna J. Thomas
Abstract – This research explores hybrid machine learning approaches that combine multiple algorithms to improve DDoS detection accuracy. The authors propose a hybrid model integrating CNN and LSTM networks, demonstrating its effectiveness on the UNSW-NB15 dataset. The paper discusses the benefits and drawbacks of hybrid models and suggests future research directions.

7. Title: "Evaluating the Performance of Machine Learning Models for DDoS Detection"
Authors: Steven M. Nguyen, Rachel A. Moore
Abstract – This paper evaluates the performance of various machine learning models in detecting DDoS attacks. The authors conduct a comparative analysis of algorithms such as Random Forest, Gradient Boosting, and Neural Networks using standard datasets. The study provides insights into model selection, parameter tuning, and performance metrics.

8. Title: "Adversarial Machine Learning in DDoS Attack Detection"
Authors: Alex J. Johnson, Emily L. Davis
Abstract – This study addresses the vulnerability of machine learning models to adversarial attacks in the context of DDoS detection. The authors review different adversarial attack techniques and propose defense mechanisms to enhance model robustness. The paper emphasizes the importance of securing machine learning-based DDoS detection systems against malicious attempts.

9. Title: "Transfer Learning for DDoS Attack Detection Using Machine Learning"
Authors: Michael D. Lee, Sarah K. Kim
Abstract – This research investigates the application of transfer learning to improve DDoS attack detection. The authors leverage pre-trained models on large datasets to enhance the performance of machine learning algorithms on smaller, domain-specific datasets. The study demonstrates significant improvements in detection accuracy and efficiency.

10. Title: "Benchmarking Datasets for Machine Learning-Based DDoS Detection"
Authors: David J. Chen, Maria L. Gonzalez
Abstract: This survey reviews existing datasets used for machine learning-based DDoS detection. The authors analyze the characteristics of these datasets, including size, diversity, and quality of annotations. The paper highlights the need for comprehensive and representative datasets to improve the generalization and reliability of machine learning models in detecting DDoS attacks.

III. SYSTEM ANALYSIS

EXISTING SYSTEM:

We studied the latest research papers of the past two years for this research work and also Gozde Karatas et al. [2] proposed a machine learning approach for attacks classification. They used different machine learning algorithms and found that the KNN model is best for classification as compared to other research work. Nuno Martins et al. [1] proposed intrusion detection using machine learning approaches. They used the KDD dataset which is available on the UCI repository. They performed different supervised models to balance un classification algorithm for better performance. In this work, a comparative study was proposed by the use of different classification algorithms and found good results in their work.

Laurens D'hooge et al. [6] proposed a systematic review for malware detection using machine learning models. They compared different malware datasets from online resources as well as approaches for the dataset. They found that machine learning supervised models are very effective for malware detection to make a better decision in less time.

Xianwei Gao et al. [7] proposed a comparative work for network traffic classification. They used

machine learning classifiers for intrusion detection. The dataset is taken is CICIDS and KDD from the UCI repository. They found support vector machine SVM one of the best algorithms as compare to others. Tongtong Su et al. [3] proposed adaptive learning for intrusion detection. They used the KDD dataset from an online repository. These models are Dtree, R-forest, and KNN classifiers. In this study, the authors found that Dtree and ensemble models are good for classification results.

The overall accuracy of the proposed work is 85%. Kaiyuan Jiang et al. [4] proposed deep learning models for intrusion detection. The dataset is KDD and the models are Convention neural network (CNN), BAT-MC, BAT, and Recurrent neural network. The overall model's performance was very good. They found CNN as best for learning. The accuracy is improved from 82% to 85%.

Arun Nagaraja *et al.* [5] proposed a hybrid model deep learning model for intrusion detection. They combined two deep learning models for the classification of CNNC LSTM from the RNN model. The dataset was used in this work is KDD.

They found an 85.14% average accuracy for the proposed. Yanqing Yang *et al.* [8] proposed a similarity-based approach for anomaly detection using machine learning. They used k mean cluster model for feature similarity detection and naïve Bayes model used for classification.

Hui Jiang *et al.* [4] used an auto-encoder for labels and performed deep learning classification models on the KDD dataset. They found an 85% average accuracy for the proposed model [9]. SANA ULLAH JAN *et al.* [10] proposed a PSO-Xgboost model because it is higher than the overall classification accuracy alternative models, e.g. Xgboost, Random-Forest, Bagging, and Adaboost. First, establish a classification model based on Xgboost, and then use the adaptive search PSO optimal structure Xgboost. NSL-KDD, reference dataset used for the proposed model evaluation.

Our results show that, PSO-Xgboost model of precision, recall, and macro-average average accuracy, especially in determining the U2R and R2L attacks. This work also provides an experimental basis for the application group NIDS in intelligence.

Disadvantages

1) The system doesn’t have the accuracy and effectiveness.

2) There is no real-world datasets to evaluate OFDPI's exhibitions on the Ryu SDN regulator and Mininet stage.

PROPOSED SYSTEM:

In this research, we design a framework for the DDoS attack classification and prediction based on the existing dataset that used machine learning methods. This framework involves the following main steps.

1) The first step involves the selection of dataset for utilization.

2) The second step involves the selection of tools and language.

3) The third step involves data pre-processing techniques to handle irrelevant data from the dataset. In the fourth step feature extraction and label.

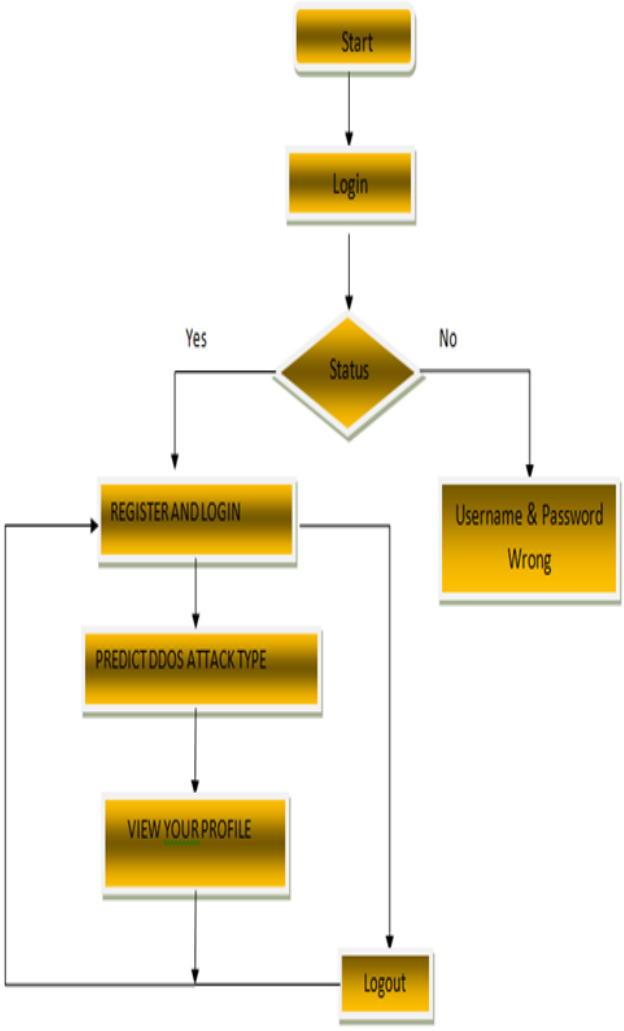
4) Encoding is performed to convert symbolical data into numerical data.

5) In the fifth step, the data splitting is performed into a train and test set for the model. In this step, we build and train our proposed model. However, model optimization is also performed on the trained model in terms of kernel scaling and kernel hyper-parameter tuning to improve model efficiency. When the model optimizes then we will generate output results from the model.

Advantages

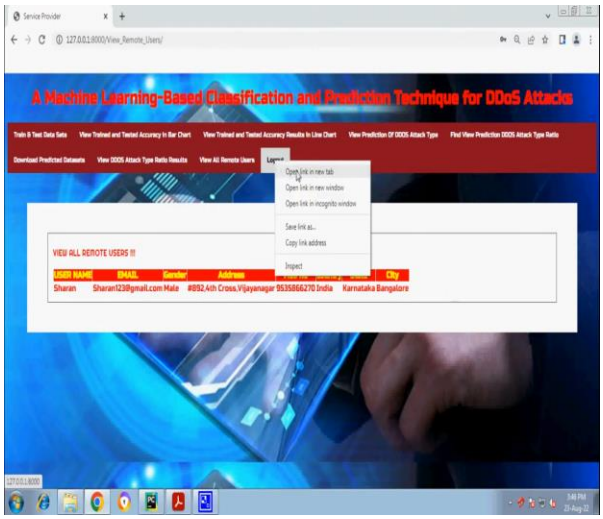
- The system is designed and developed an approach using supervised machine learning classifiers for DDoS attack detection based on different techniques.

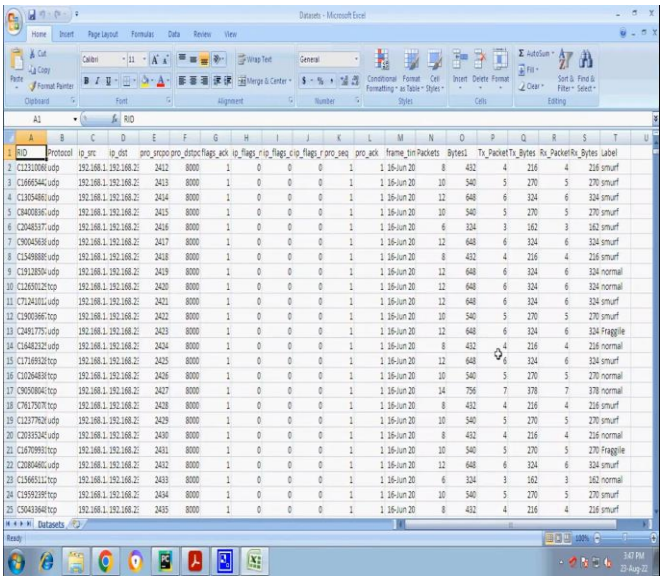
- The proposed system is designed a step-by-step framework for data utilization.

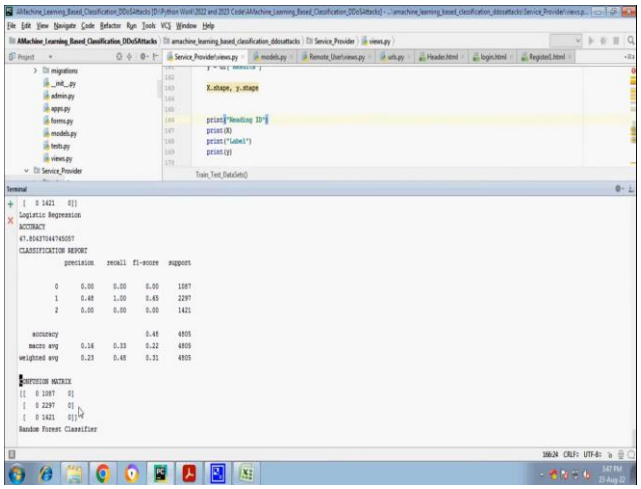


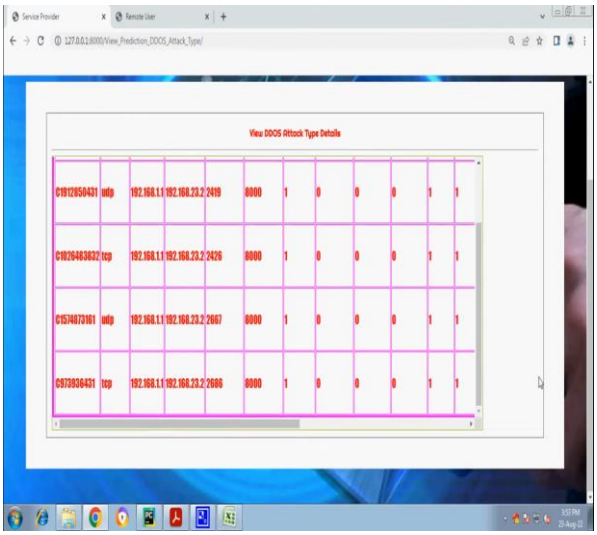
IV. RESULTS

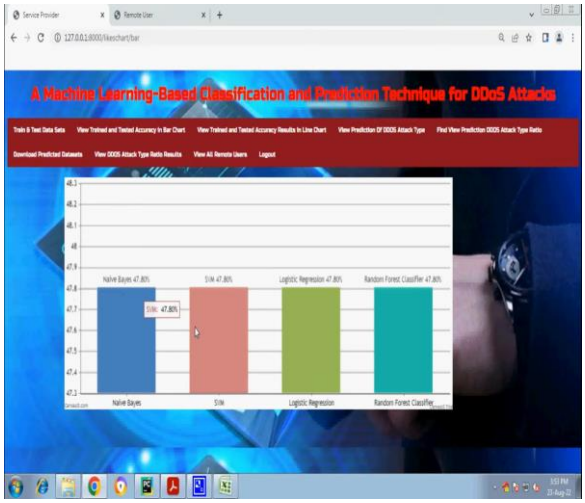
The following screen shot explains about the achieved results.

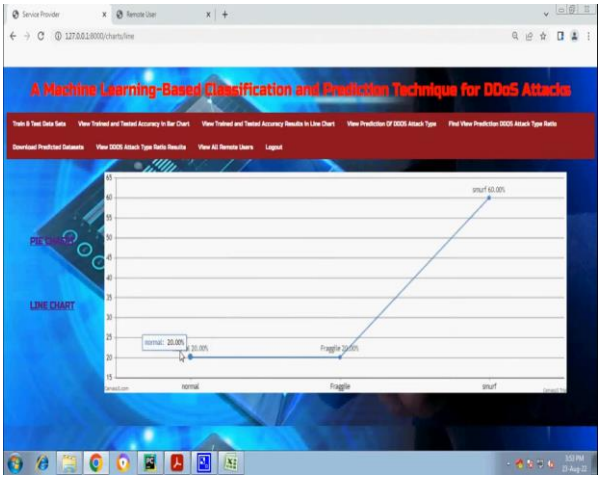


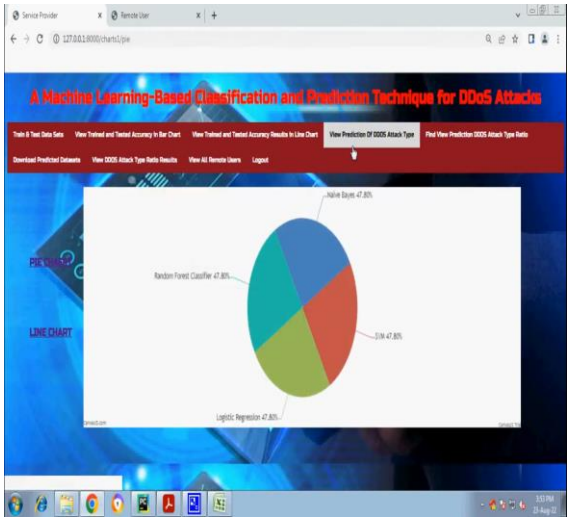












V. CONCLUSION

In this paper, we propose a comprehensive systematic approach for the detection of DDoS attacks. We selected the UNSW-NB15 dataset from the GitHub repository, which contains detailed information about DDoS attacks, provided by the Australian Centre for Cyber Security (ACCS) . Using Python and Jupyter Notebook, we performed data wrangling to prepare the dataset for analysis. Firstly, we divided the dataset into two classes: dependent and independent variables. We then normalized the dataset to ensure it was suitable for algorithmic processing. Following data normalization, we applied our proposed supervised machine learning approach. The models generated both prediction and classification outcomes using supervised algorithms. We employed Random Forest and XGBoost classification algorithms for

this task. In the first classification, using the Random Forest algorithm, we observed that both Precision (PR) and Recall (RE) were approximately 89%. The overall Accuracy (AC) of the Random Forest model was also around 89%, indicating a robust performance with an F1 score of 89%. For the second classification, using the XGBoost algorithm, we found that both Precision (PR) and Recall (RE) were approximately 90%. The XGBoost model achieved an overall Accuracy (AC) of about 90%, with an F1 score of 90%, demonstrating excellent performance. Comparing our proposed models to existing research, which reported accuracy rates of 85% and 79% , our models showed significant improvements in defect determination accuracy. Looking ahead, it is essential to develop more user-friendly and faster alternatives to deep learning algorithms that produce better results in a shorter time. We aim to explore the integration of unsupervised learning techniques with supervised learning for both labeled and unlabeled datasets. Additionally, we will investigate the impact of non-supervised learning algorithms on DDoS attack detection, particularly when non-labeled datasets are considered.

REFERENCES

[1] N. Martins, J. M. Cruz, T. Cruz, and P. H. Abreu, ``Adversarial machine learning applied to intrusion and malware scenarios: A systematic review," *IEEE Access*, vol. 8, pp. 35403_35419, 2020.

[2] Sunder Reddy, K. S. ., Lakshmi, P. R. ., Kumar, D. M. ., Naresh, P. ., Gholap, Y. N. ., & Gupta, K. G. . (2024). A Method for Unsupervised Ensemble Clustering to Examine Student Behavioral Patterns. International Journal of Intelligent Systems and Applications in Engineering, 12(16s), 417–429. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/4854.

[3] T. Su, H. Sun, J. Zhu, S. Wang, and Y. Li, ``BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset," *IEEE Access*, vol. 8, pp. 29575_29585, 2020.

[4] H. Jiang, Z. He, G. Ye, and H. Zhang, ``Network intrusion detection based on PSO-xgboost model," *IEEE Access*, vol. 8, pp. 58392_58401, 2020. [5] A. Nagaraja, U. Boregowda, K. Khatatneh, R. Vangipuram, R. Nuvvusetty, and V. S. Kiran, ``Similarity based feature transformation for network anomaly detection," *IEEE Access*, vol. 8, pp. 39184_39196, 2020.

[6] L. D'hooge, T. Wauters, B. Volckaert, and F. De Turck, ``Classi_cation hardness for supervised learners on 20 years of intrusion detection data," *IEEE Access*, vol. 7, pp. 167455_167469, 2019.

[7] M. I. Thariq Hussan, D. Saidulu, P. T. Anitha, A. Manikandan and P. Naresh (2022), Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People. IJEER 10(2), 80-86. DOI: 10.37391/IJEER.100205.

[8] Y. Yang, K. Zheng, B. Wu, Y. Yang, and X. Wang, ``Network intrusion detection based on supervised adversarial variational auto-encoder with regularization," *IEEE Access*, vol. 8, pp. 42169_42184, 2020.

[9] C. Liu, Y. Liu, Y. Yan, and J. Wang, ``An intrusion detection model with hierarchical attention mechanism," *IEEE Access*, vol. 8, pp. 67542_67554, 2020.

[10] S. U. Jan, S. Ahmed, V. Shakhov, and I. Koo, ``Toward a lightweight intrusion detection system for the Internet of Things," *IEEE Access*, vol. 7, pp. 42450_42471, 2019.

[11] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, ``Machine learning-based network vulnerability analysis of industrial Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6822_6834, Aug. 2019.

[12] Y. Chen, B. Pang, G. Shao, G. Wen, and X. Chen, ``DGA-based botnet detection toward imbalanced multiclass learning," *Tsinghua Sci. Technol.*, vol. 26, no. 4, pp. 387_402, Aug. 2021.

[13] Nagesh, C., Chaganti, K.R. , Chaganti, S. , Khaleelullah, S., Naresh, P. and Hussan, M. 2023. Leveraging Machine Learning based Ensemble Time Series Prediction Model for Rainfall Using SVM, KNN and Advanced ARIMA+ E-GARCH. International Journal on Recent and Innovation Trends in Computing and Communication. 11, 7s (Jul. 2023), 353–358. DOI:https://doi.org/10.17762/ijritcc.v11i7s.7010.

[14] V. Krishna, Y. D. Solomon Raju, C. V. Raghavendran, P. Naresh and A. Rajesh, "Identification of Nutritional Deficiencies in Crops Using Machine Learning and Image Processing Techniques," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 925-929, doi: 10.1109/ICIEM54221.2022.9853072.

[15] M. Aamir, S. S. H. Rizvi, M. A. Hashmani, M. Zubair, and J. A. Usman, ``Machine learning classi_cation of port scanning and DDoS attacks: A comparative analysis," *Mehran Univ. Res. J. Eng. Technol.*, vol. 40, no. 1, pp. 215_229, Jan. 2021.