

## Evaluation of Bernoulli Naive Bayes model for detection of distributed denial of service attacks

Ayodeji Olalekan Salau<sup>1,5</sup>, Tsehay Admassu Assegie<sup>2</sup>, Adedeji Tomide Akindadelo<sup>3</sup>, Joy Nnenna Eneh<sup>4</sup>

<sup>1</sup>Department of Electrical/Electronics and Computer Engineering, Afe Babalola University, Ado-Ekiti, Nigeria

<sup>2</sup>Department of Computer Science, College of Natural and Computational Science, Injibara University, Injibara, Ethiopia

<sup>3</sup>Department of Basic Sciences, Babcock University, Ilishan Remo, Nigeria

<sup>4</sup>Department of Electronic Engineering, University of Nigeria, Nsukka, Nigeria

<sup>5</sup>Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India

### Article Info

#### Article history:

Received Apr 30, 2022

Revised Aug 5, 2022

Accepted Nov 9, 2022

#### Keywords:

Bernoulli Naive Bayes

Machine learning

Malware

Malware detection

SYN-flood attack

### ABSTRACT

Distributed denial of service is a form of cyber-attack that involves sending several network traffic to a target system such as DHCP, domain name server (DNS), and HTTP server. The attack aims to exhaust computing resources such as memory and the processor of a target system by blocking the legitimate users from getting access to the service provided by the server. Network intrusion prevention ensures the security of a network and protects the server from such attacks. Thus, this paper presents a predictive model that identifies distributed denial of service attacks (DDSA) using Bernoulli-Naive Bayes. The developed model is evaluated on the publicly available Kaggle dataset. The method is tested with a confusion matrix, receiver operating characteristics (ROC) curve, and accuracy to measure its performance. The experimental results show an 85.99% accuracy in detecting DDSA with the proposed method. Hence, Bernoulli-Naive Bayes-based method was found to be effective and significant for the protection of network servers from malicious attacks.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



### Corresponding Author:

Ayodeji Olalekan Salau

Department of Electrical/Electronics and Computer Engineering, Afe Babalola University

Ado-Ekiti, Nigeria

Email: ayodejisalau98@gmail.com

## 1. INTRODUCTION

Distributed denial of service attack (DDSA) is a form of security attack where the attacker attempts to exhaust resources, such as bandwidth, memory space, or processor time of the target system. The legitimate users are then blocked due to the unavailability of service in the network. Hence, DDSA has become a serious issue for network security [1], [2].

To overcome the DDSA challenge, several research works have been conducted. Although there have been lots of studies undertaken recently on the countermeasures and methods to mitigate the DDSA, there is little work on the application of machine learning (ML) in combatting the severity of this form of attack [2]–[23]. Network attack detection using a ML model is an automated system where a ML algorithm determines the class of network traffic into either “normal” or “malicious” class based on the observations in the training set. There are two network traffic classes used in training and testing the ML models which are the “normal” traffic class and the “malicious” traffic class. The objective of this research is to propose a ML model by employing a Bernoulli Naive Bayes algorithm to effectively detect DDSA. This study aims to answer the following key questions:

- a. How to develop a ML model by employing a Bernoulli Naive Bayes algorithm to detect a distributed denial-of-service attack with an acceptable level of accuracy?
- b. What is the performance of Bernoulli Naive Bayes on DDSA detection?
- c. Which network traffic feature has a strong relationship to DDSA?

The remaining sections of this paper are arranged as follows: the second section presents the related works on denial-of-service attack detection. The method is presented in section 3. The experimental results are presented and discussed in section 4. Section 5 concludes the paper with the addition of recommendations for future research.

## 2. RELATED WORK

Much research works has investigated different ways of automation of the DDSA detection using supervised ML algorithms in literature. A thorough analysis of the literature reveals that research still needs to be conducted to build intelligent systems using ML approaches to mitigate the DDSA. Some of the previous research works on the automation of DDSA are discussed in this section. Some of the ML algorithms employed on different DDSA data repositories such as KDD 99 and Kaggle include support vector machine (SVM), decision tree (DT), random forest, artificial neural network (ANN), and Naive Bayes. According to Kim and Cho [3], a neural network based DDSA identification model was proposed. The proposed model is supposed to identify DDSA in web traffic such as HTTP-post traffic.

In another study by Su [4], K-nearest neighbor (KNN) based DDSA identification model was developed. The model aims to detect flooding attacks, such as synchronize (SYN)-flood. The flooding attack is a type of network intrusion where the attacker consumes the resources of a system in a network by sending frequent network traffic, resulting in a denial of the resources to the legitimate users. ML approaches have many applications in building intelligent systems that can automate network intrusion mitigation such as distributed DDSA [5]. The applications of ML systems are vital to knowledge discovery (KD) from intrusion data records. They are used to develop predictive models to automatically monitor network traffic in the incidence of intrusion into the network system. Different ML algorithms such as ANN, DT, Naive Bayes, and SVM have been used by several researchers to develop predictive models that can automatically identify network traffic classes. According to Salunkhe and Mali [6], an ensemble based DDSA identification model was developed. The model was implemented using ensemble classifiers, J48 and logistic regression.

In the work by Kurniabudi *et al.* [7], the Naive Bayes algorithm was used to develop a model for the identification of DDSA. The authors compared the performance of different algorithms for the identification of DDSA. The comparative result shows that KNN is a better algorithm for identification of the attack. But, the limitation of supervised ML algorithms is that they can be implemented to detect known attacks [8]. The algorithms are being trained using a pre-defined label as either normal and malicious network traffic feature. The models usually detect an anomaly or network intrusion based on the known features in the network traffic. According to Farooqi and Munir [9], a SVM-based DDSA identification model was proposed. Another research by Abusitta *et al.* [10] was conducted to mitigate DDSA with a KNN-based DDSA detection model. The model was proposed to mitigate UDP-flooding and IP spoofing detection.

A DT based DDSA detection model was proposed in [11] to effectively identify DDSA. The authors used the KDD99 intrusion detection data repository. The proposed model is effective in DDSA detection using the classification algorithm. Although, the accuracy can be improved to a better value. In another study by Haq *et al.* [12], DT and random forest algorithms were employed to classify simple network management (SNMP) datasets to detect DDSA targeted at an SNMP server. The models have different prediction accuracy. The DT algorithm performed well as compared to the random forest algorithm.

A Naive Bayes-based predictive model was proposed for the automated detection of DDSA [13], [14]. Feature selection is applied to extract relevant features in the dataset. As showcased in the study, the presented feature selection approach improved the prediction accuracy of the model. Furthermore, the feature selection approach helped in the reduction of irrelevant features from the data repository and improved the training complexity. In Alsariera [15], a study on DDSA against HTTP flooding detection was presented. The authors proposed a model which used a ML approach, specifically, a Naive Bayesian algorithm. The proposed model has a high predictive performance and can be used on HTTP servers for DDSA detection.

ML algorithms have different accuracies in DDSA detection. A comparative analysis presented in [16], [17] on the performance of ML algorithms such as ANN, DT, and Naive Bayes on DDSA detection shows that ANN algorithm has better performance. The accuracy of the ANN algorithm on DDSA detection is 84.5%. Another method for DDSA detection using transport control protocol (TCP) connection parameters, such as SYN, SYN-acknowledge (ACK), and ACK bit was proposed in [18]. The proposed

method was proposed to detect DDSA based on the TCP connection parameters using KNN to classify the malicious and normal TCP traffic.

According to Tally and Amintoosi [19], a new framework for the detection of a distributed denial-of-service attack on a domain name server (DNS) was proposed which employed the random forest algorithm. As showcased in the results and analysis section of the study, an acceptable level of accuracy was achieved. But feature selection was not applied to the dataset and the performance of the framework could be improved by applying feature selection to the dataset. According to Ramasamy and Eric [20], an ML system for the detection of a distributed denial-of-service attack was proposed which employed deep learning. The proposed model is effective in DDSA detection, especially in software-defined networks and cloud computing environments. A SVM-based DDSA detection model was proposed in [21]. The proposed model is designed to be applied in the cloud computing environments. The model is effective, and the performance is acceptable but still requires improvement.

Among the various methods to strengthen network security, ML approaches plays a vital role in automating network attack detection [22]. The ML approaches help in developing self-learning and knowledge-based or intelligent systems which are used in mitigating the major security risks related to distributed denial of service and other types of attacks. The study by Assegie and Nair [23], proposed a convolutional neural network (CNN) method for DDSA detection. The other ML algorithms, namely: logistic regressions, ANN, SVM, and CNN have better performance than the other presented algorithms.

### 3. METHOD

The Kaggle distributed denial of the service data repository was used in this work to create a predictive model to detect DDSA in network traffic. The dataset used consists of 450 observations of which 213 observations are normal network traffic and 237 of the observations are malicious network traffic. Each observation has 14 features. The dataset was divided into training and testing sets. The training set consists of 80% of the observations in the dataset and the testing set consists of 20% of the observations in the dataset. The data repository used in this work is summarized in Table 1. For the implementation and experimental testing, Python programming language is used for implementing the pearson's correlation matrix, while the test on the proposed model was implemented using the python language. The python programming language provides many scientific libraries for handling large dataset processing and ML algorithm implementation which is widely used by numerous researchers worldwide. The Kaggle application layer distributed denial of service dataset was used in this research, and it consists of 10 features with 450 observations. The features of the dataset are presented in Table 1.

Table 1. The distributed denial of service dataset description

No.	Feature	Description
1	Destination port	The port number of the destination
2	Flow duration	The duration of traffic flow
3	Total_Backward_Packets	Total number of packets forwarded back
4	Total_Length_of_Fwd_Packets	Total number of packets forwarded
5	Flow_Packets_Sec	Number of packets transmitted per second
6	FIN_Flag_Count	Number of FIN bits in the network traffic
7	SYN_Flag_Count	Number of SYN bits in the network traffic
8	RST_Flag_Count	Number of RST bits in the network traffic
9	ACK_Flag_Count	Number of ACK bits in the network traffic
10	Class label	The target feature (0=Normal traffic, 1=Dos attack)

#### 3.1. Feature correlation analysis

Different types of DDSA are surveyed to determine the most important features of DDSA. There are thirteen features extracted from the dataset using a person's correlation analysis. The correlation analysis enabled us to determine the relationship between each feature of the dataset. Based on the relationship between features and the target feature, those features having a strong correlation with the target feature were selected during the training. The correlation matrix of each extracted feature in the dataset is shown in Figure 1. The feature extraction helps in the identification of the features highly correlated to the target. The features strongly correlated to the target are important for achieving better accuracy in prediction [24]–[26].

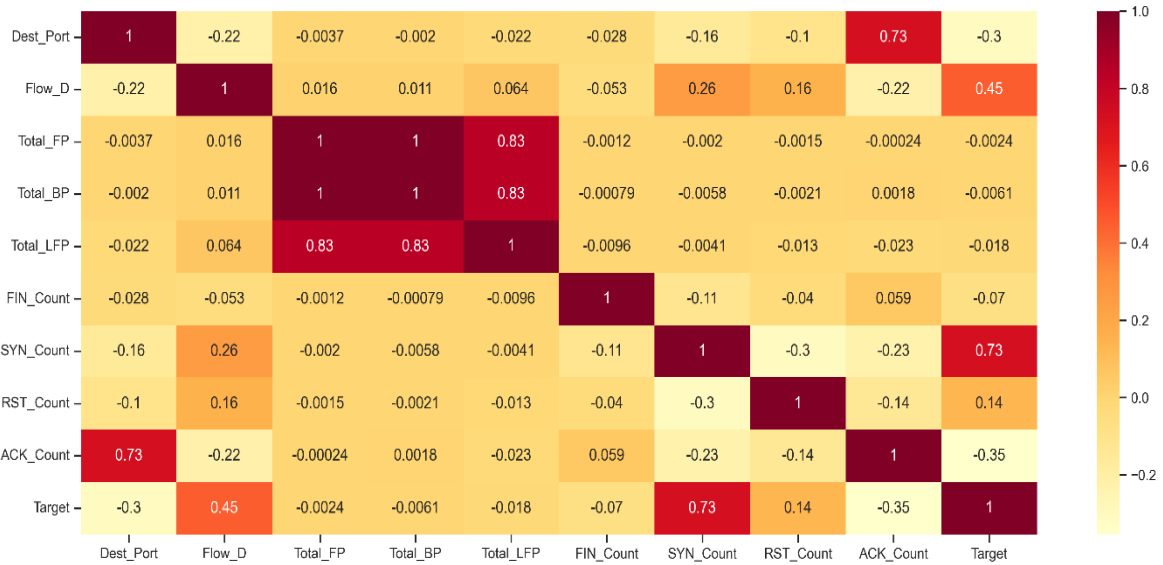


Figure 1. Distributed denial of service dataset feature correlation

4. RESULTS AND DISCUSSION

The results of this research are described in this section. The predictive effectiveness of the Bernoulli Naive Bayes model was analyzed on the testing set. Apart from the accuracy-test, other performance measures such as confusion matrix and ROC curve were used to test the proposed model’s efficiency for DDSA attack detection.

4.1. Accuracy analysis

The predictive accuracy of the proposed model is evaluated using an accuracy score for random tests of the model. The experimental test results of the accuracy scores of the model for each experiment are summarized in Table 2. The predictive accuracy of the SVM model for distributed denial of service detection is shown in Figure 2. As shown in Figure 3, the proposed model is effective in the detection of distributed denial of service with an accuracy score of above 84% for five random tests conducted on the model.

Table 2. Accuracy of the proposed model

Experimental test	Models accuracy (%)
1	87.77
2	90.00
3	81.11
4	86.66
5	84.44

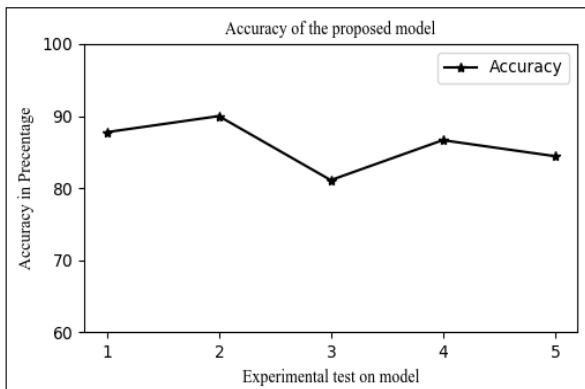


Figure 2. Accuracy of the developed model

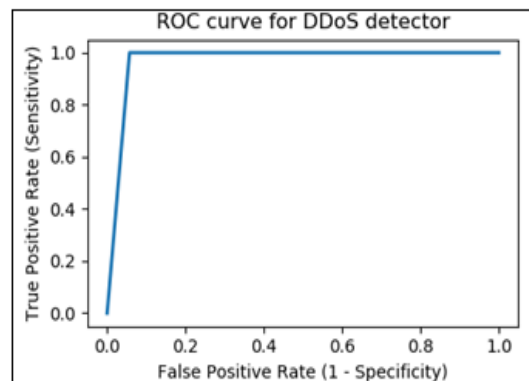


Figure 3. ROC curve of the developed model

## 4.2. Receiver operating characteristics curve analysis

The ROC shows the true positive rate or the number of correct predictions of the model when the given observation is a DDSA feature. The false positive rate or the number of correct predictions by the model has DDSA when the given observation is a normal network traffic. The ROC curve of the proposed distributed denial of service detection model is shown in Figure 3. The results show that model's sensitivity is high initially but reaches a steady state from 0.1 to 1.0.

## 5. CONCLUSION

In this study, a method for the DDSA detection was developed with the Bernoulli Naïve method. The method was found to be effective in the prediction of SYN-flood attacks in a network. The performance of the developed method was evaluated with accuracy as a metric to measure the predictive performance of SYN-flood attacks on the test set. The experimental result shows that the developed method performs well in predicting DDSA with 85.99% accuracy. Hence, the Bernoulli Naive Bayes-based method is shown to be effective and significant for the protection of the server from attacks. In the future, the authors would test the Bernoulli Naive Bayes based on other datasets and develop more effective methods for the automation of DDSA.





## REFERENCES

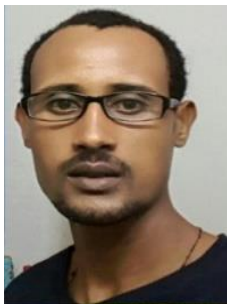
- [1] D. Lin, "Network intrusion detection and mitigation against denial of service attack," Penn Libraries University of Pennsylvania, 2013.
- [2] A. O. Salau, N. Marriwala, and M. Athae, "Data security in wireless sensor networks: attacks and countermeasures," in *Mobile Radio Communications and 5G Networks*, Singapore: Springer, 2021, pp. 173–186, doi: 10.1007/978-981-15-7130-5\_13.
- [3] T.-Y. Kim and S.-B. Cho, "Web traffic anomaly detection using C-LSTM neural networks," *Expert Syst. Appl.*, vol. 106, pp. 66–76, Sep. 2018, doi: 10.1016/j.eswa.2018.04.004.
- [4] M.-Y. Su, "Using clustering to improve the KNN-based classifiers for online anomaly network traffic identification," *J. Netw. Comput. Appl.*, vol. 34, no. 2, pp. 722–730, Mar. 2011, doi: 10.1016/j.jnca.2010.10.009.
- [5] B. Ahmad, W. Jian, and Z. A. Ali, "Role of machine learning and data mining in internet security: standing state with future directions," *J. Comput. Networks Commun.*, vol. 1, pp. 1–10, Jul. 2018, doi: 10.1155/2018/6383145.
- [6] U. R. Salunkhe and S. N. Mali, "Security enrichment in intrusion detection system using classifier ensemble," *J. Electr. Comput. Eng.*, pp. 1–6, 2017, doi: 10.1155/2017/1794849.
- [7] K. Kurniabudi *et al.*, "Network anomaly detection research: a survey," *Indones. J. Electr. Eng. Informatics*, vol. 7, no. 1, Mar. 2019, doi: 10.52549/ijeeci.v7i1.773.
- [8] Z. Rui, Z. Shaoyan, L. Yang, and J. Jianmin, "Network anomaly detection using one-class support vector machine," in *Proceedings of the International MultiConference of Engineers and Computer Scientists 2008*, 2008.
- [9] A. H. Farooqi and A. Munir, "Intrusion detection system for IP multimedia subsystem using k-Nearest neighbor classifier," in *2008 IEEE International MultiTopic Conference*, Dec. 2008, pp. 423–428, doi: 10.1109/INMIC.2008.4777775.
- [10] A. Abusitta, M. Bellaiche, and M. Dagenais, "An SVM-based framework for detecting DoS attacks in virtualized clouds under changing environment," *J. Cloud Comput.*, vol. 7, no. 9, pp. 1–18, 2018, doi: 10.1186/s13677-018-0109-4.
- [11] A. Sanmorino, "A study for DDOS attack classification method," *J. Phys. Conf. Ser.*, vol. 1175, no. 1, pp. 1–6, 2019, doi: 10.1088/1742-6596/1175/1/012025.
- [12] N. F. Haq, M. Avishek, F. Muhammad, A. Rahman, M. Rafni, and D. Md., "Application of machine learning approaches in intrusion detection system: a survey," *Int. J. Adv. Res. Artif. Intell.*, vol. 4, no. 3, pp. 9–18, 2015, doi: 10.14569/IJARAI.2015.040302.
- [13] N. T. Lam, "Detecting unauthorized network intrusion based on network traffic using behavior analysis techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 4, pp. 46–51, 2021, doi: 10.14569/IJACSA.2021.0120407.
- [14] A. Das, Pramod, and B. S. Sunitha, "Anomaly-based network intrusion detection using ensemble machine learning approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 2, pp. 635–645, 2022, doi: 10.14569/IJACSA.2022.0130275.
- [15] Y. A. Alsariera, "Detecting generic network intrusion attacks using tree-based machine learning methods," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 2, pp. 597–603, 2021, doi: 10.14569/IJACSA.2021.0120275.
- [16] A. Alsaedi and M. Z. Khan, "Performance analysis of network intrusion detection system using machine learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 12, pp. 671–678, 2019, doi: 10.14569/IJACSA.2019.0101286.
- [17] Slamet and I. I. M. Abdelaziz, "An enhanced classification framework for intrusions detection system using intelligent exoplanet atmospheric retrieval algorithm," *Bull. Electr. Eng. Informatics*, vol. 11, no. 2, pp. 1018–1025, Apr. 2022, doi: 10.11591/eei.v11i2.3308.
- [18] N. P. Shetty, J. Shetty, R. Narula, and K. Tandona, "Comparison study of machine learning classifiers to detect anomalies," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 5, pp. 5445–5452, Oct. 2020, doi: 10.11591/ijece.v10i5.pp5445-5452.
- [19] M. T. Tally and H. Amintoosi, "A hybrid method of genetic algorithm and support vector machine for intrusion detection," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 1, pp. 900–908, Feb. 2021, doi: 10.11591/ijece.v11i1.pp900-908.
- [20] M. Ramasamy and P. V. Eric, "An improved deep bagging convolutional neural network classifier for efficient intrusion detection system," *Bull. Electr. Eng. Informatics*, vol. 11, no. 1, pp. 405–413, Feb. 2022, doi: 10.11591/eei.v11i1.3252.
- [21] A. Boukhalifa, N. Hmina, and H. Chaoni, "Parallel processing using big data and machine learning techniques for intrusion detection," *IAES Int. J. Artif. Intell.*, vol. 9, no. 3, pp. 553–560, Sep. 2020, doi: 10.11591/ijai.v9.i3.pp553-560.
- [22] F. A. Vadhill, M. F. Nanne, and M. L. Salihi, "Importance of machine learning techniques to improve the open source intrusion detection systems," *Indones. J. Electr. Eng. Informatics*, vol. 9, no. 3, pp. 774–783, Sep. 2021, doi: 10.52549/ijeeci.v9i3.3219.
- [23] T. A. Assegie and P. S. Nair, "Comparative study on methods used in prevention and detection against address resolution protocol spoofing attack," *J. Theor. Appl. Inf. Technol.*, vol. 97, no. 16, pp. 4259–4269, 2019.
- [24] A. O. Salau and S. Jain, "Feature extraction: a survey of the types, techniques, applications," in *2019 International Conference on Signal Processing and Communication (ICSC)*, Mar. 2019, pp. 158–164, doi: 10.1109/ICSC45622.2019.8938371.





- [25] M. Panda and M. R. Patra, "Network intrusion detection using naive bayes," *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 7, no. 12, pp. 258–263, 2007.
- [26] M. Zaman and C.-H. Lung, "Evaluation of machine learning techniques for network intrusion detection," in *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, Apr. 2018, pp. 1–5, doi: 10.1109/NOMS.2018.8406212.

## BIOGRAPHIES OF AUTHORS







**Ayodeji Olalekan Salau**     received a B.Eng. degree in Electrical/Computer Engineering from the Federal University of Technology, Minna, Nigeria. He received his M.Sc and Ph.D degrees in Electronic and Electrical Engineering from the Obafemi Awolowo University, Ile-Ife, Nigeria. His research interests include research in the fields of computer vision, image processing, signal processing, machine learning, power systems engineering, and nuclear engineering. Dr. Salau serves as a reviewer for numerous reputable international journals. His research has been published in a number of reputable international conferences, books, and major international journals. He is a registered Engineer with the Council for the Regulation of Engineering in Nigeria (COREN), a member of the International Association of Engineers (IAENG), and a recipient of the Quarterly Franklin Membership with ID number CR32878 given by the Editorial Board of London Journals Press in 2020 for top quality research output. In addition, Dr. Salau's paper was awarded the best paper of the year 2019 in Cogent Engineering. Furthermore, he is the recipient of the International Research Award on New Science Inventions (NESIN) under the category of "Best Researcher Award" given by ScienceFather in 2020. He is the recipient of the International Best Researcher Award given by the ISSN International Science and Technology Awarding body (IISTAC-2022) in 2022. Presently, Dr. Salau works at Afe Babalola University in the Department of Electrical/Electronics and Computer Engineering. He can be contacted at email: ayodejisalau98@gmail.com.







**Tsehay Admassu Assegie**     received a B.Sc., degree in Computer Science from Dilla University, Dilla Ethiopia in 2013. He received a Master's degree in Computer Science from the Faculty of Science at Andhra University, India in 2016. Currently, he is working as Lecturer in the Department of Computer Science, College of Natural and Computational Science, Injibara University, Injibara, Ethiopia. His research interests include machine learning, medical image analysis, and pattern recognition. He has published 38 research articles in internationally reputed and peer-reviewed journals. He can be contacted at email: tsehayjournal@gmail.com.



**Adedeji Tomide Akindadelo**     an experimentalist with computational skills received the B.Sc. (Edu) in Physics and M.Sc. Materials science degrees from the Obafemi Awolowo University, Ile-Ife, Nigeria. His research interests include research in the fields of Nano technology, materials synthesis and characterization, machine learning, control systems engineering and power systems technology. Mr. Adedeji's research has been published in many reputable international conferences, books, and major international journals. He is a trained physicist with family life. Married Adebimpe Temitope Akindadelo and blessed with Ayowade Ademobo Akindadelo. Currently, Mr. Adedeji Akindadelo works at Babcock University in the Department of Basic Sciences. He can be contacted at email: akindadeload@babcock.edu.ng.



**Joy Nnenna Eneh**     is currently a Senior Lecturer in the Department of Electronic Engineering University of Nigeria Nsukka. Dr. Eneh's research interests are in the areas of intelligent control, artificial intelligence, robotics, and machine learning. She can be contacted at email: Nnenna.eneh@unn.edu.ng.