Information Security Management–CSE3502

ENSEMBLE LEARNING BASED CYBER-ATTACK DETECTION FOR IOT NETWORKS

FINAL REPORT

Submitted by

Manideep	20BCT0300
Atharv Mahesh Gote	20BCT0151
Akshita Langer	20BCE2872
Harshwardhan Singh Rajawat	20BCI0300

J COMPONENT CSE3502

Submitted to

Prof. Ruby D.

Assistant Professor Sr. Grade 1

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CONTENTS

SCOPE

The Internet of Medical Things (IoT), a type of Internet of Things (IoT) application, is a promising solution to several challenges in traditional healthcare systems, such as low-quality patient care, high healthcare costs, a shortage of medical personnel, and inadequate medical supplies. IoT systems offer many benefits, such as enhancing the effectiveness and efficiency of treatments and improving patient health outcomes.

However, recent cyber attacks have exposed significant vulnerabilities in IoT environments, and the frequency and severity of these attacks are increasing rapidly. To address these issues and make healthcare more effective and secure, we propose a novel method for detecting cyber attacks in IoT networks that combines ensemble learning and a fog-cloud architecture. This approach is particularly crucial given that Frost & Sullivan predicts that nearly 30 billion IoT devices will be in use, making security measures more critical than ever.

ABSTRACT

The Internet of Things (IoT) is a rapidly growing field in computing that serves as the foundation for future smart city infrastructure and has the potential to revolutionize traditional healthcare systems. Despite its benefits, IoT networks are highly susceptible to cyber-attacks due to privacy and security concerns. Detecting these attacks in IoT medical technology (IoT) requires a different approach than traditional security mechanisms, as IoT services demand specific requirements, such as low latency, battery life, and network bandwidth that cannot be fulfilled by traditional standalone cloud computing. This limitation leads to extended data recovery times, which can be detrimental in medical emergencies that require a quick response from healthcare professionals. Fortunately, the characteristics of fog computing, such as mobility support and position awareness, offer essential requirements for a wide range of sensors in IoT environments. Therefore, integrated solutions must be designed explicitly to provide fog-level security alarm facilities with advanced infrastructure-level attack detection capabilities, leveraging the computational power of fog nodes to perform granular traffic control and identify suspicious communication patterns. This paper proposes using an ensemble learning-based IDS with a fog-cloud architecture to address the security threats and challenges posed by dynamic IoT networks effectively.

To reduce or eliminate these vulnerabilities we have reviewed various methods.

KEYWORDS

KNN , Random forest , SVM , GNB , DNN , Decision tree , Voting , Ensemble learning , Fog-cloud architecture , Stacking , Bagging , Boosting , Deep learning ,

INTRODUCTION

- Cloud and fog computing have revolutionized the way we store and access data. However, one of the biggest challenges in these technologies is the security of the data. To address this problem, a team of technology enthusiasts embarked on a project to develop a solution using an ensemble machine learning model.

- The team decided to use three base learners and one meta classifier to build the ensemble model. The ensemble model is a combination of several models that work together to improve the accuracy of the predictions. By combining the predictions of the individual models, the ensemble model is more robust and can handle complex data sets.

- To train and test the model, the team used the NSL-KDD dataset. This dataset is commonly used in machine learning research and is designed to mimic real-world scenarios. The dataset consists of various types of attacks, such as DoS, Probe, and U2R, and normal traffic data.

- The team used the base learners to learn from the data and make predictions. The predictions were then combined using the meta classifier to generate the final prediction. The team tested the model using various metrics, such as accuracy, precision, recall, and F1-score, to evaluate its performance.

- Overall, the team's ensemble machine learning model showed promising results in detecting cyber attacks in cloud and fog computing environments. This model has the potential to improve the security of data and prevent cyber attacks from compromising sensitive information. By using machine learning, the team was able to create a more intelligent system that can learn and adapt to new attack patterns, making it a valuable tool for cybersecurity professionals.

- **Support Vector Machine (SVM):** Support Vector Machine (SVM) is a supervised learning model used for classification and regression problems

- **Autoencoder:** An autoencoder is a type of deep neural network that learns compact numerical representations for different sources of a signal through representational learning.

- **Bagging:** Bagging, also known as bootstrap aggregation, is an ensemble learning algorithm that reduces variance in a noisy dataset by randomly choosing data samples multiple times.

- **Bayesian Network:** Bayesian Network is a probabilistic graphical model that represents a set of variables and their conditional dependencies using a directed acyclic graph.

- **Boosting:** Boosting is an ensemble learning algorithm that combines weak base learners to create a strong learner by reducing training errors.

- **Cloud computing: c**loud computing is the on-demand delivery of IT resources over the Internet with pay-as-you-go pricing.

- **Decision Tree:** A decision tree is a non-parametric supervised machine learning method used for classification and regression

- **Deep Learning:** Deep learning is a machine learning algorithm that mimics the structure and function of the human brain to gain knowledge.

- **Deep Neural Network (DNN):** A Deep Neural Network (DNN) is an artificial neural network with multiple layers beween the input and output layers.

- **Distributed Denial-of-Service(DDoS):** Distributed Denial-of-Service (DDoS) is a malicious attempt to disrupt network traffic of a targeted server, service, or network by overwhelming it with a flood of internet traffic.

- **Ensemble learning :** Ensemble learning is the process of strategically generating and combining multiple models, such as classifiers or experts, to solve a particular computational intelligence problem.

- **Fog computing :** Fog computing is a decentralized computing infrastructure where data, compute, storage, and applications are located between the data source and the cloud

- **Internet of Things(IoT): The** Internet of Things (IoT) is a network of physical objects, called "things," embedded with sensors, software, and other technology to connect and exchange information with other devices or networks over the internet.

- **Intrusion detection system (IDS) :** An Intrusion Detection System (IDS) is a device or software application that monitors network traffic for malicious activity or policy violations.

- **Gradient Boosting:** Gradient Boosting is a machine learning technique used in regression and classification problems.

- **Naïve Bayes:** Naïve Bayes is a supervised machine learning algorithm based on applying Bayes' theorem with the assumption of independence among predictors for classification.

- **Probabilistic Neural Network**: Probabilistic Neural Network is a feedforward neural network commonly used in classification and pattern recognition problems.
- **Random Forest:** Random Forest is a supervised machine learning algorithm constructed from decision tree algorithms for classification and regression problems.

- **Ransomware:** It is a malicious software which employs encryption in order to prevent a victim from accessing their system unless they pay a ransom.

- **Recurrent Neural Networks:** Recurrent Neural Networks are a class of artificial neural networks used for sequential or time series data.

- **Rotation Forest:** Rotation Forest is a tree-based ensemble method used for generating classifier ensembles based on feature extraction. .

- **Stacking:** It is an ensemble learning algorithm where the output of one set of base learners is fed as input into a meta classifier for training.

LITERATURE SURVEY

Papers close to our project

Sl.No.	Name of the author with year	Major technologies used	Results/Outcomes of their Research	Drawbacks if any
1	Shilan S. Hameed, Wan Haslina Hassan, and Liza Abdul Latiff,2020.	Naïve Bayes, Hoeffding Tree and Ensemble of their various forms using Weighted Majority Algorithm (WMA)	The results demonstrated that the proposed model is effective for attack detection at the fog layer, while it gives better accuracy, higher detection rate and lower false positive rate with average detection time.	The accuracy of the proposed model was 98.89% but the time taken was very high (10.81 seconds). Various feature reduction algorithms or optimizers can be added to reduce the time taken to process the dataset and increase efficiency.
2	Swarna Priya R.M., Praveen Kumar Reddy Maddikunta, Parimala M., Srinivas Koppu, Thippa Reddy Gadekallu, Chiranji Lal Chowdhary,	Deep Neural Networks, K-Nearest Neighbor (KNN), Naïve Bayes (NV), Random Forest (RF), Support Vector Machine (SVM) and Deep Neural	The proposed model has a faster convergence rate in finding global minima.	Limited evaluation: The proposed approach was only evaluated using a small number of datasets, which may not be enough to demonstrate its efficacy in all circumstances. Limited comparison: Only a few existing methods were used to compare the proposed approach, making it difficult to assess its performance in full. Lack of explanation: It may be challenging for other researchers to duplicate and expand on the work because the authors did not provide a thorough explanation of the features used in the proposed approach.

	Mamoun Alazab,2020.	Networks (DNN)		
3	A. Verma en V. Ranga,2019.	Boosted Trees, Bagged Trees, Subspace Discriminant and RUSBoosted Trees	ELNIDS can help protect RPL based networks from various routing attacks by detecting Sinkhole,Blackhole, Sybil,Clone ID, Selective Formatting, Hello Flooding and Local Repair attacks.	While the highest accuracy rate is 94.5% of the Boosted Trees, the Subspace Discriminant method gives the lowest accuracy of 78.7%. Ensemble methods are computationally expensive.
4	Prabhat Kumar, Govind P. Gupta, Rakesh Tripathi,2021.	Decision tree, Naive Bayes and Random Forest	The proposed method is able to detect malicious behavior in a highly dynamic and heterogeneous network of IoT. and as the parameters used to design the system are easy to update in real-time, the overall performance of the system is improved.	The detection system cannot identify DDoS and Ransomwares.
5	M. Saharkhizan, A. Azmoodeh, A. Dehghantanha, K.-K. R. Choo, en R. M. Parizi,2020.	LSTM , Decision tree	On evaluating the approach using real-world Modbus network traffic dataset,precision of 95.1% is obtained against cyber attack detection in IoT devices.	The approach using LSTM requires longer training as well as consume huge amounts of memory.

Main Paper: Prabhat Kumar, Govind P. Gupta, Rakesh Tripathi, “An ensemble learning and fog-cloud architecture-driven cyber-attack detection framework for IoT networks”, *Computer Communications*, Volume 166, 2021, Pages 110-124, ISSN 0140-3664, https://doi.org/10.1016/j.comcom.2020.12.003.

Proposed method is a cyber-attack detection system for IoT environment that utilizes fog-architecture and ensemble-based machine learning. The fog-architecture helps in decentralizing the security mechanism that is typically based on cloud.

Working: The training dataset undergoes preprocessing, including feature mapping, replacing missing values with feature means, identifying optimized features using correlation coefficient method, and feature normalization. Subsequently, 10-fold cross-validation is applied to train three machine learning algorithms, namely Decision tree, Naive Bayes, and Random Forest. The prediction outputs are then used to construct a predictive model. During the incoming IoT traffic, it is first preprocessed and then fed into the predictive model to detect any malicious attacks. If any malicious behavior is detected, the system administrator is notified; otherwise, a response is generated for the request.

Outcome: The proposed approach demonstrates the capability to detect malicious behavior in a dynamic and heterogeneous IoT network. Moreover, the system's parameters can be easily updated in real-time, resulting in improved overall performance.

Drawback : The detection system cannot identify DDoS and Ransomwares.

1. S. Manimurugan, “IoT-Fog-Cloud model for anomaly detection using improved Naïve Bayes and principal component analysis”, Journal of Ambient Intelligence and Humanized Computing, bll 1–10, 2021.

   The proposed method consists of three layers, namely the fog layer, the cloud layer, and the Internet of Things (IoT) layer. The fog layer facilitates data processing and aggregation, which results in reduced latency between IoT sensors and the cloud. The cloud layer manages large volumes of data and provides storage services. The IoT layer comprises devices, and this integration of IoT with fog and cloud computing creates a state-of-the-art platform that supports IoT applications for smart cities and addresses cyber security concerns. The proposed method employs an advanced Naive Bayes classifier that is based on the Principal Component Analysis method to extract features, which enables the classification of attacks. The results indicate that the suggested technique improves the effectiveness of anomaly detection and enhances the accuracy, detection rate, and precision of performance analysis. However, the method may encounter false positive rates, which require resolution in the future.

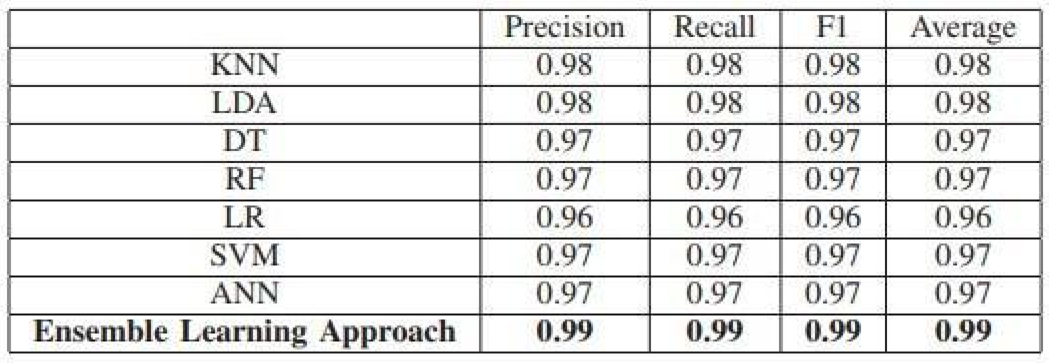
2. M. Saharkhizan, A. Azmoodeh, A. Dehghantanha, K.-K. R. Choo, en R. M. Parizi, “An ensemble of deep recurrent neural networks for detecting IoT cyber attacks using network traffic”, *IEEE Internet of Things Journal*, vol 7, no 9, bll 8852–8859, 2020.

   A sophisticated deep learning technique has been proposed, which involves the incorporation of clusters of Long-Short-Term-Memory (LSTM) modules into a detector ensemble. By utilizing LSTM, the approach is capable of recognizing long-term data patterns and identifying dependency patterns within incoming data sequences. A decision tree is then used to combine these components and generate an aggregate output. The effectiveness of this approach was evaluated using a real-world dataset of Modbus network traffic, and the results indicate that it achieved a 99% detection accuracy rate for cyber attacks on IoT devices.

Drawback: The method using LSTM has the drawback of taking a long time to train and using up a lot of memory.

3. S. Khare en M. Totaro, “Ensemble learning for detecting attacks and anomalies in iot smart home”, in *2020 3rd International Conference on Data Intelligence and Security (ICDIS)*, 2020, bll 56–63.

   This article compares the performance of conventional machine learning techniques and an ensemble machine learning model for detecting anomalies in IoT smart home environments. The raw sensor data is analyzed and prepared to be fed into any machine learning algorithm. The data is then divided into training and test data sets after analysis. The training data undergoes feature scaling and is taught using various machine learning algorithms and ensemble learning techniques, which enhances the effectiveness of the final model. Finally, the performance of the model is evaluated by comparing the assessment data to the training data. The results indicate that the ensemble learning approach outperforms conventional machine learning techniques in terms of precision, recall, F1, and accuracy.



	Precision	Recall	F1	Average
KNN	0.98	0.98	0.98	0.98
LDA	0.98	0.98	0.98	0.98
DT	0.97	0.97	0.97	0.97
RF	0.97	0.97	0.97	0.97
LR	0.96	0.96	0.96	0.96
SVM	0.97	0.97	0.97	0.97
ANN	0.97	0.97	0.97	0.97
Ensemble Learning Approach	**0.99**	**0.99**	**0.99**	**0.99**

Drawbacks: It is very computationally complex and has not been tried in a real-world IoT environment.

4. A. Verma en V. Ranga, “ELNIDS: Ensemble learning based network intrusion detection system for RPL based Internet of Things”, in *2019 4th International conference on Internet of Things: Smart innovation and usages (IoT-SIU)*, 2019, bll 1–6

   Ensemble learning models are recommended to enhance model dependability and prediction. To detect routing attacks in RPL-based 6LoWPAN networks, four ensemble techniques, including Boosted Trees, Bagged Trees, Subspace Discriminant, and RUSBoosted Trees, are employed. The ELNIDS approach's ability to identify various routing attacks such as Sinkhole, Blackhole, Sybil, Clone ID, Selective Formatting, Hello Flooding, and Local Repair attacks can aid in safeguarding RPL-based networks. However, the accuracy rates vary among the ensemble techniques, with Boosted Trees achieving the highest accuracy at 94.5%, and the Subspace Discriminant method having the lowest accuracy rate at 78.7%.

5. R. Maharaja, P. Iyer, en Z. Ye, “A hybrid fog-cloud approach for securing the Internet of Things”, *Cluster Computing*, vol 23, no 2, bll 451–459, 2020.

A security mechanism utilizing fog computing is proposed as a means of safeguarding IoT devices from malicious attacks. The system is comprised of three layers of security: a VPN server, a network analysis unit, and a challenge-response unit. All incoming communication is first encrypted by the VPN server before being unencrypted upon exiting the tunnel, providing an added layer of protection. A traffic analysis component that employs machine learning and decision tree classification to distinguish between legitimate and fraudulent requests is the second layer. Finally, the challenge-response unit is activated to authenticate the source if the traffic analysis unit detects unusual behavior from a trusted node.

The proposed method is shown to be effective in filtering malicious activity in a hybrid fog-cloud architecture. Moreover, it offers low response latency and network costs due to the flexible task sharing between the fog and cloud. However, to improve classification accuracy, it would have been preferable to employ other machine learning algorithms. Additionally, the effectiveness of the proposed system has not been assessed in real-world IoT settings or applications.

6. E. Tsogbaatar *et al.*, “DeL-IoT: A deep ensemble learning approach to uncover anomalies in IoT”, *Internet of Things*, vol 14, bl 100391, 2021.

Suggested: DeL-IoT is an approach to deep ensemble learning that utilizes principles of deep autoencoders and probabilistic neural networks (PNN) to detect dynamic assaults in IoT anomalies. This method is more effective as it also employs Software-Defined Networking (SDN). The outcome of this approach is that it is capable of identifying abnormalities and managing dynamic flow during an attack, with a higher detection rate than current techniques. However, its drawback is that it cannot predict or identify very brief, intermittent, or multiscale DDoS attacks.

7. Bin Jia, Xiaohong Huang, Rujun Liu, Yan Ma, "A DDoS Attack Detection Method Based on Hybrid Heterogeneous Multiclassifier Ensemble Learning", *Journal of Electrical and Computer Engineering*, vol. 2017, Article ID 4975343, 9 pages, 2017.

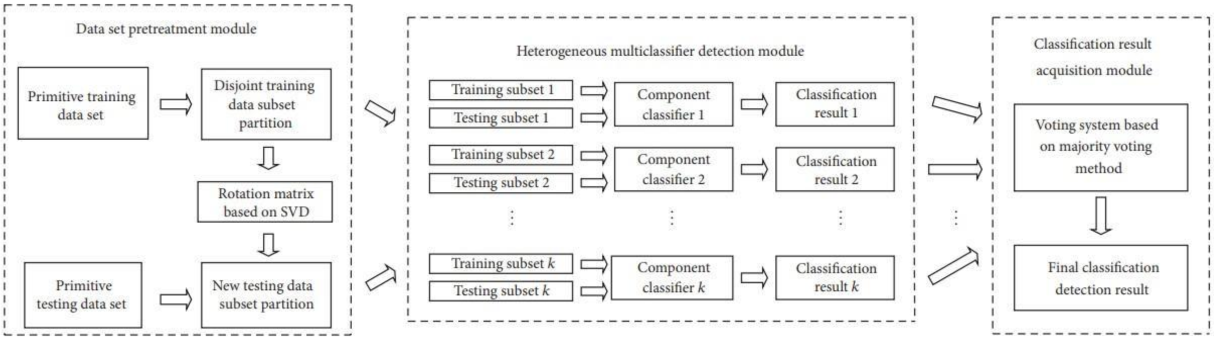

Data set pretreatment module
Primitive training data set
Disjoint training data subset partition
Rotation matrix based on SVD
Primitive testing data set
New testing data subset partition
Heterogeneous multiclassifier detection module
Training subset 1
Testing subset 1
Component classifier 1
Classification result 1
Training subset 2
Testing subset 2
Component classifier 2
Classification result 2
Training subset k
Testing subset k
Component classifier k
Classification result k
Classification result acquisition module
Voting system based on majority voting method
Final classification detection result

FIGURE 1: Hybrid heterogeneous multiclassifier ensemble classification model.

The hybrid heterogeneous multi-classifier ensemble classification model consists of three modules: the Data Set Pretreatment Module, Heterogeneous Multiclassifier Detection Module, and Classification Result Acquisition Module. In the Data Set Pretreatment Module, the primitive training data set is divided into k distinct training data subsets using SVD. The new training data subset is created through linearly independent base transformation. The primitive testing data sets are also split into k data subsets corresponding to the features of the new training data subsets, and the new testing data subsets are created using Rotation Forest. These k fresh training data subsets and testing data subsets are then sent to the k component classifiers in the heterogeneous multi-classifier detection module. The voting system uses the majority voting method to vote on k outcomes in the Classification Result Acquisition Module to obtain the final classification detection outcome. The proposed model is tested using the 1999 KDD Cup for Knowledge Discovery and Data Mining and is shown to be a stable and effective detection technique, as demonstrated by the results of the True Negative Rates (TNR), accuracies, and precisions of the Random Forest, k-NN, and Bagging algorithms after SVD and un-SVD processing. However, the KDD-99 dataset is unsuitable for contemporary settings due to its age, skewed objectives, non-stationarity across training and test datasets, pattern repetition, and irrelevant characteristics, as noted in a detailed analysis of the dataset.

8. Hariharan Rajadurai, Usha Devi Gandhi, “A stacked ensemble learning model for intrusion detection in wireless network”, *Neural Comput & Applic (2020).*

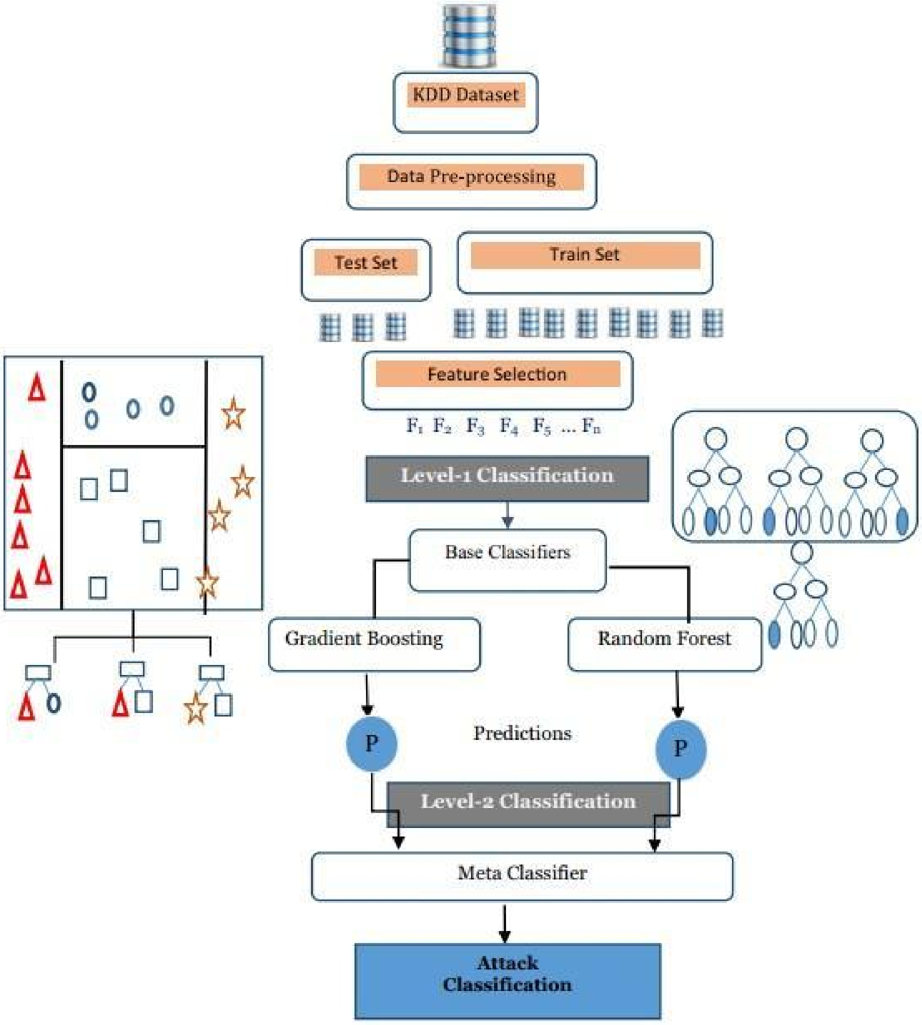


Fig. 3 Stacked ensemble-based IDS architecture

Proposed: The proposed ensemble learning model for network intrusion detection uses gradient boosting machine (GBM) and random forest (RF) algorithms. The suggested approach integrates the characteristics of both classifiers as basis classifiers. A meta-classifier is employed in the second level of classification after the first level. The dataset used is NSL-KDD, a recent version of KDD-99.
Outcome: The suggested stacked ensemble learning method outperformed traditional machine learning approaches in terms of detection rate, recall, and accuracy for each attack type. The results show that the RNN and ANN approaches have significantly lower precision and recall values compared to the proposed method for each attack type.

9. Yun Zhou, Peichao Wang, "An ensemble learning approach for XSS attack detection with domain knowledge and threat intelligence", *Computers & Security,* Volume 82, 2019, Pages 261-269, ISSN 0167-4048.

The proposed approach for detecting XSS attacks involves creating an ontology for the attack process and extracting features from it to characterize the attack. Bayesian learning is used to set the nodes in each Bayesian network for the attack characteristics, with a scoring and searching learning algorithm. An ensemble learner is generated by grouping individual models using a voting mechanism, and threat intelligence is developed to enhance the performance of the model and combat attack concealment. Complement rules are also created using gathered intelligence. Node priority sorting is carried out to determine the impact of each node on the final detection outcome. The proposed approach is based on a whitebox model, which can be used to augment rules derived from threat intelligence, enabling security staff to respond to incidents quickly and identify concealed attacks. The training dataset consists of 151,658 records, including 135,507 regular records and 16,151 XSS payloads from GitHub, while the testing dataset comprises 3,497 XSS payloads and 6,503 normal records collected from various GitHub sources and security forums.

Outcome:

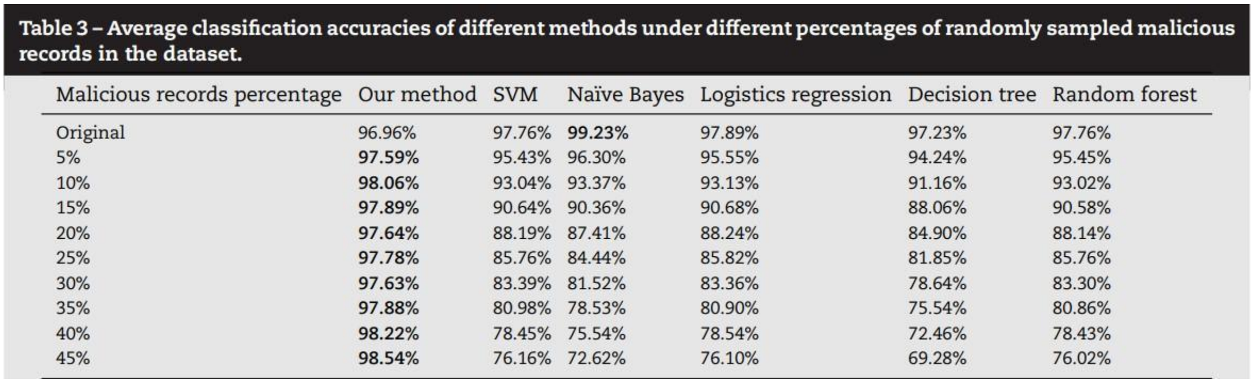


Table 3 – Average classification accuracies of different methods under different percentages of randomly sampled malicious records in the dataset.

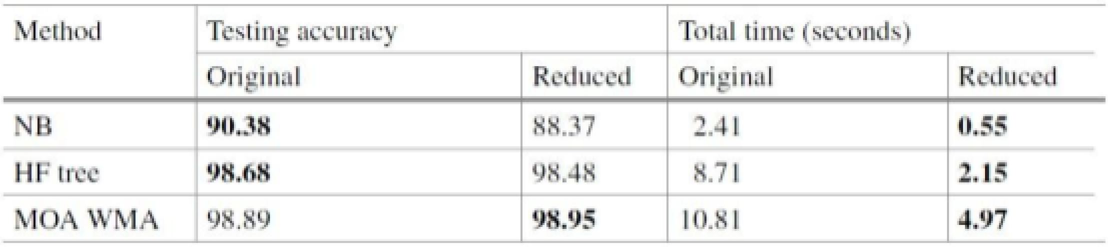
Malicious records percentage	Our method	SVM	Naïve Bayes	Logistics regression	Decision tree	Random forest
Original	96.96%	97.76%	**99.23%**	97.89%	97.23%	97.76%
5%	**97.59%**	95.43%	96.30%	95.55%	94.24%	95.45%
10%	**98.06%**	93.04%	93.37%	93.13%	91.16%	93.02%
15%	**97.89%**	90.64%	90.36%	90.68%	88.06%	90.58%
20%	**97.64%**	88.19%	87.41%	88.24%	84.90%	88.14%
25%	**97.78%**	85.76%	84.44%	85.82%	81.85%	85.76%
30%	**97.63%**	83.39%	81.52%	83.36%	78.64%	83.30%
35%	**97.88%**	80.98%	78.53%	80.90%	75.54%	80.86%
40%	**98.22%**	78.45%	75.54%	78.54%	72.46%	78.43%
45%	**98.54%**	76.16%	72.62%	76.10%	69.28%	76.02%

10. Poulmanogo Illy, Georges Kaddoum, Christian Miranda Moreira, Kuljeet Kaur, and Sahil Garg, "Securing Fog-to-Things Environment Using Intrusion Detection System Based On Ensemble Learning", *2019 IEEE Wireless Communications and Networking Conference (WCNC),* 2019, pp. 1-7.

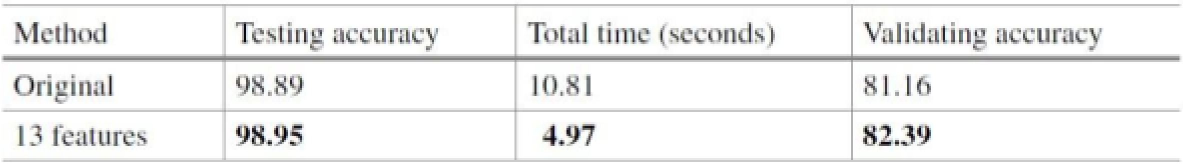
This research stands out because it employs ensemble learners to enhance the accuracy of intrusion detection by utilizing the most realistic intrusion dataset currently available. The proposed deployment architecture involves two layers of classifications in a fog-to-things environment. The first layer focuses on anomaly detection to significantly reduce classification latency, while the second layer carries out attack classifications to enable accurate preventive measures. The base learners used in this study include deep neural networks, parametric and non-parametric models, and decision trees. Bagging and Random Forest classifiers were also implemented to execute learners, along with boosting algorithm-based multistage ensembles. The NSL-KDD dataset was used for the study, and the results showed that ensemble learners generate better classification models, with an overall accuracy of 85.81% and 84.25% for binary classification and attack classification, respectively. Moreover, a deployment architecture was proposed, which combines attack classification in the cloud with anomaly detection in fog nodes to direct intrusion prevention tasks more efficiently.

11. Shilan S. Hameed, Wan Haslina Hassan, and Liza Abdul Latiff, “An Efficient Fog-Based Attack Detection Using Ensemble of MOA-WMA for Internet of Medical Things”, *Innovative Systems for Intelligent Health Informatics,* IRICT 2020. Lecture Notes on Data Engineering and Communications Technologies, vol 72. Springer, Cham.

An ensemble attack detection method is proposed to detect stream data attacks at the fog layer. The method employs fundamental classifiers that use stream- and incremental-based algorithms to work with IoT fog and nature devices. A weighted majority approach is used to achieve the best accuracy with the least latency. The classifiers used include Nave Bayes, Hoeffding Tree, and Ensemble of their many variants employing Weighted Majority Algorithm (WMA). The classifiers are used in their incremental form to make them compatible with IoT streaming data, which grows over time. The proposed approach, Online Stream Data Analysis using Weighted Majority Algorithm (MOA WMA), employs a collection of four learners, including Hoeffding Tree, Hoeffding Tree with Naive Bayes on leaves, Hoeffding Tree with Naive Bayes Adaptive on leaves, and Nave Bayes classifier. The NSL-KDD dataset was used for the study, and feature selection was performed using the Information Gain (IG) filter, with multiple feature selection thresholds being applied. The outcomes show that the proposed model is efficient for attack detection at the fog layer, providing superior accuracy, a higher detection rate, and a lower false positive rate with an average detection time. The results also suggest that choosing 13 features is the best selection criteria in terms of giving average performance/accuracy (testing accuracy and unseen data accuracy) and total running time of the MOA-WMA model.



Method	Testing accuracy		Total time (seconds)	
	Original	Reduced	Original	Reduced
NB	**90.38**	88.37	2.41	**0.55**
HF tree	**98.68**	98.48	8.71	**2.15**
MOA WMA	98.89	**98.95**	10.81	**4.97**



Method	Testing accuracy	Total time (seconds)	Validating accuracy
Original	98.89	10.81	81.16
13 features	**98.95**	**4.97**	**82.39**

12. Swarna Priya R.M., Praveen Kumar Reddy Maddikunta, Parimala M., Srinivas Koppu, Thippa Reddy Gadekallu, Chiranji Lal Chowdhary, Mamoun Alazab, “An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoT architecture”, Computer Communications, Volume 160, 2020, Pages 139-149.

he proposed method for constructing an efficient and effective Intrusion Detection System (IDS) in the IoT environment involves the use of a Deep Neural Network (DNN). The network parameters are preprocessed, improved upon, and tweaked using hyperparameter selection methods. A thorough examination of experiments in DNN with other machine learning techniques is compared on the benchmark intrusion detection dataset. The proposed DNN model outperforms current machine learning approaches, resulting in a 15% accuracy increase and a 32% reduction in time complexity, allowing for faster alerts to prevent the aftereffects of intrusion in sensitive cloud data storage. The suggested methodology involves four primary processes: pre-processing, dimensionality reduction, classification, and the proposed IDS model for the IoT environment. The performance of the hybrid PCA-GWO for dimensionality reduction over classification utilizing well-known classifiers is examined. The IDS's performance was enhanced by using the GWO approach for feature extraction, which finds global minima with a higher rate of convergence. The proposed IDS model offers safe data transfer, avoiding fatal and major concerns among end-users. The results were obtained using Kaggle's globally benchmarked intrusion detection dataset.

13. Al-Abassi, Abdulrahman, et al. "An ensemble deep learning-based cyber-attack detection in industrial control system." IEEE Access 8 (2020): 83965-83973.

The proposed approach uses multiple unsupervised Stacked Autoencoders (SAE) to learn new representations from unbalanced datasets, followed by feeding these representations into a Deep Neural Network (DNN) and using them as a binary classifier to detect attacks using a Decision Tree (DT). The main contributions of the paper are the construction of new, balanced representations using deep representation learning, the use of ensemble deep learning methods to detect cyber-attacks, and the creation of a generalized model that requires minimal system modification.

The study uses accurate ICS datasets collected in 2015 and 2018 for evaluation and shows that the proposed method outperforms existing strategies in all four measures, particularly in the f-measure. Future research will focus on improving the accuracy of the proposed method and creating a new model to recognize different types of attacks and their locations to strengthen ICS network security against similar cyberattacks.

14. Aman, Azana Hafizah Mohd, Wan Haslina Hassan, Shilan Sameen, Zainab Senan Attarbashi, Mojtaba Alizadeh, and Liza Abdul Latiff. "IoT amid COVID-19 pandemic: Application, architecture, technology, and security." *Journal of Network and Computer Applications* 174 (2021): 102886.

his paper provides a comprehensive overview of the potential uses and challenges of implementing IoT in the context of the COVID-19 pandemic. The paper highlights the benefits of IoT-based systems in detecting, tracking, and controlling the spread of infectious diseases. It also provides insights into how various countries have implemented IoT-based solutions to mitigate the effects of the pandemic.

The paper also addresses the potential security concerns associated with the implementation of IoT-based systems, such as data privacy, confidentiality, and integrity. The authors highlight the importance of ensuring that IoT systems are designed with security in mind and suggest various security measures that can be implemented to mitigate these risks.

One notable aspect of the paper is the focus on historical epidemics, which provides useful insights into how technology, particularly IoT, has been used in the past to control the spread of infectious diseases. The paper also emphasizes the importance of balancing the need for effective disease control with the need to protect individual privacy and data security.

Overall, this paper provides a useful overview of the potential benefits and challenges of implementing IoT-based solutions in the context of the COVID-19 pandemic. The paper highlights the importance of designing secure and privacy-preserving IoT systems to mitigate the risks associated with the implementation of such systems.

15. Hameed, Shilan S., Ali Selamat, Liza Abdul Latiff, Shukor A. Razak, Ondrej Krejcar, Hamido Fujita, Mohammad Nazir Ahmad Sharif, and Sigeru Omatu. "A Hybrid Lightweight System for Early Attack Detection in the IoT Fog." *Sensors* 21, no. 24 (2021): 8289.

The paper presents a novel hybrid approach for early attack detection in the IoT fog. The approach uses six different incremental classifiers that are implemented in an adaptive online system. The classifiers are designed to work on lightweight fog devices, with minimal memory usage of less than 6 MiB. The system uses seven different types of sensors and NetFlow data to detect nine different types of recent attacks. The outcomes show that the proposed system operates with 100% accuracy on the lightweight fog devices, has a quick detection time, and is less sensitive to concept drift. The results of the single-criteria comparative analysis demonstrate that the WHTE ensemble is more precise and less sensitive to concept drift than the other classifiers.

Overall, the paper addresses an important issue in the IoT fog environment and presents a promising solution that is designed to work with lightweight devices and has a high level of accuracy. The use of multiple sensors and the incorporation of various types of attacks in the evaluation is a strength of the study. However, it would have been useful to see a comparison with existing approaches to attack detection in the IoT fog. Additionally, the paper does not provide a detailed discussion of the potential limitations of the proposed approach or the challenges associated with its implementation. Nevertheless, the study is a valuable contribution to the field of IoT security and provides a foundation for future research in this area.

16. Yu, Zengchen, Syed Umar Amin, Musaed Alhussein, and Zhihan Lv. "Research on disease prediction based on improved DeepFM and IoT." *IEEE Access* 9 (2021): 39043-39054.

This work proposes the use of an improved version of DeepFM for predicting the occurrence of hepatitis in patients using structured illness prediction data from the 2020 Artificial Intelligence Challenge Preliminary Competition. The improved DeepFM model showed superior performance in terms of AUC when compared to other models. The goal of this research is to reduce the workload of physicians by applying this model to electronic medical data and focusing on samples with higher predicted occurrence rates. To ensure that the disease can be predicted in advance, IoT sensors can be used to collect additional data such as blood pressure, height, weight, cholesterol, etc. After integrating IoT, a healthcare system that excels at forecasting and timely performance can be created. Overall, this work has the potential to improve the accuracy and efficiency of disease prediction and prevention in the healthcare industry.

17. Ashfaq, Zarlish, Abdur Rafay, Rafia Mumtaz, Syed Mohammad Hassan Zaidi, Hadia Saleem, Syed Ali Raza Zaidi, Sadaf Mumtaz, and Ayesha Haque. "A review of enabling technologies for Internet of Medical Things (IoT) Ecosystem." *Ain Shams Engineering Journal* 13, no. 4 (2022): 101660.

Additionally, this study highlights the importance of data privacy and security in IoT, as well as the need for standardization in terms of data exchange and interoperability. The authors also discuss the potential ethical concerns surrounding the use of IoT, such as issues of informed consent and the responsible use of sensitive medical data. Overall, the study emphasizes the potential benefits of IoT in healthcare and the need for continued research and development to address the challenges and concerns associated with its implementation.

18. Ihnaini, Baha, M. A. Khan, Tahir Abbas Khan, Sagheer Abbas, Mohammad Sh Daoud, Munir Ahmad, and Muhammad Adnan Khan. "A smart healthcare recommendation system for multidisciplinary diabetes patients with data fusion based on deep ensemble learning." *Computational Intelligence and Neuroscience* 2021 (2021).

The proposed smart healthcare recommendation system for diabetes, based on deep machine learning and data fusion, has achieved an impressive accuracy of 99.6%, according to the study. The use of data fusion has helped to reduce unnecessary processing resources and improve the system's overall performance in predicting and recommending the disease. The ensemble machine learning model was used to predict diabetes and was tested on a well-known diabetic dataset. The results showed that the proposed system was more effective in predicting and recommending diabetes compared to existing deep machine learning techniques. The study suggests that the proposed approach can be used in automated diagnostic and recommendation systems for diabetic patients due to its increased illness diagnosis performance. Overall, this work showcases the potential of using advanced machine learning techniques to improve healthcare outcomes and suggests that such systems can be effective in predicting and preventing chronic diseases such as diabetes.

19. Reshiwaran, A., L. Jegatheswaran, Isaac Joshua Sakira, and Nor Azlina Abd Rahman. "A Review on IoT device Vulnerabilities and Countermeasures." In *Journal of Physics: Conference Series*, vol. 1712, no. 1, p. 012020. IOP Publishing, 2020.

It's important to note that while there may be challenges associated with implementing IoT in healthcare, there are also potential benefits that can outweigh these challenges. For example, IoT devices can provide real-time patient data to healthcare providers, allowing for more personalized and effective treatment plans. They can also automate certain tasks and reduce the workload of healthcare professionals, leading to improved efficiency and cost savings in the long run.

Regarding the three challenges mentioned, it's true that implementing IoT devices will require investment in supporting infrastructure and network security measures. However, it's important to weigh these costs against the potential benefits of IoT devices. Additionally, advancements in technology are making these devices more affordable and accessible.

Regarding security concerns, it's important to note that all connected devices, including IoT devices, can be vulnerable to cyber attacks. However, with proper security measures in place, such as strong authentication protocols and data encryption, these risks can be minimized.

Finally, regarding network strain, this is a valid concern as IoT devices generate a large amount of data that needs to be transmitted securely and efficiently. However, advancements in networking technology, such as 5G and edge computing, are making it easier to manage this data and ensure secure transmission.

Overall, while there are challenges associated with implementing IoT in healthcare, the potential benefits, such as improved patient outcomes and increased efficiency, make it a promising area for future development.

20. Abbas, Adeel, Muazzam A. Khan, Shahid Latif, Maria Ajaz, Awais Aziz Shah, and Jawad Ahmad. "A New Ensemble-Based Intrusion Detection System for Internet of

Things." *Arabian Journal for Science and Engineering* (2021): 1-15.

The proposed IDS model showed promising results in terms of accuracy, false alarm rate, and processing power. The use of multiple feature selection methods helped in identifying the most relevant features for intrusion detection. Moreover, the ensemble-based approach helped in improving the overall accuracy of the model, as compared to traditional ANN and DL approaches. The performance of the proposed model was evaluated in both binary and multi-class scenarios, which adds to its versatility and practical applicability in real-world scenarios. Overall, the study contributes to the development of efficient and effective intrusion detection models for network security using ensemble-based learning techniques.

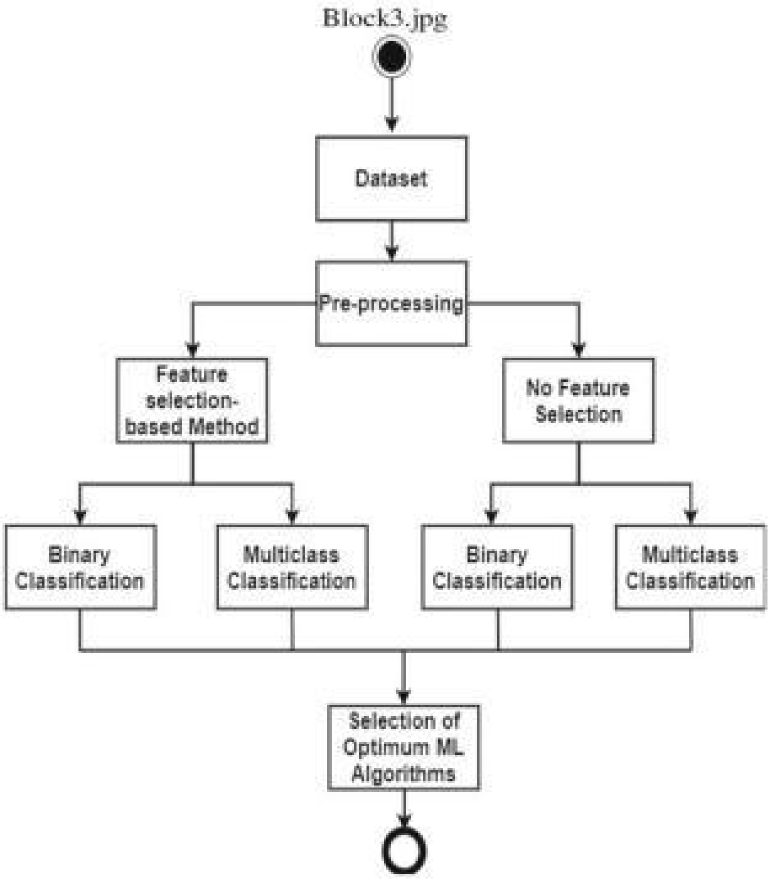

Block3.jpg
Dataset
Pre-processing
Feature selection-based Method
No Feature Selection
Binary Classification
Multiclass Classification
Binary Classification
Multiclass Classification
Selection of Optimum ML Algorithms

Fig. 4 Flowchart of intrusion detection process

21. Smys, S., Abul Basar, and Haoxiang Wang. "Hybrid intrusion detection system for internet of things (IoT)." *Journal of ISMAC* 2, no. 04 (2020): 190-199.

It seems there is a mistake in the description. The sentence "Results from experiments show that the suggested hybrid paradigm is more vulnerable to attacks on the IoT network." contradicts the previous sentence that states "The suggested paradigm can identify various forms of attacks". It is likely that the intended meaning was "Results from experiments show that the suggested hybrid paradigm is more effective in detecting attacks on the IoT network."

Assuming this is the case, the corrected description would be:

This research paper presented a hybrid convolutional neural network-based intrusion detection system for Internet of Things networks that can identify various forms of attacks. The suggested paradigm can be used for many different IoT applications. The proposed research is examined, validated, and contrasted with deep learning and traditional machine learning models.

Results from experiments show that the suggested hybrid paradigm is more effective in detecting attacks on the IoT network. It achieves a greater detection accuracy of 98% when compared to traditional recurrent neural networks through experimental verification, making the application ideal for various IoT scenarios.

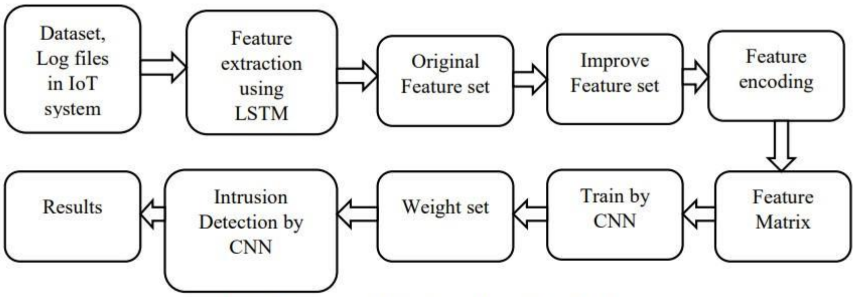


Fig. 5 Proposed Intrusion Detection System

22. Begli, MohammadReza, Farnaz Derakhshan, and Hadis Karimipour. "A layered intrusion detection system for critical infrastructure using machine learning." In *2019 IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE)*, pp. 120-124. IEEE, 2019.

In this paper, the authors propose a secure framework for remote healthcare systems to protect against network threats such as DoS and U2R attacks. They propose an intrusion detection system (IDS) using Support Vector Machine (SVM), a machine learning method. The effectiveness of their proposed framework is evaluated using various metrics, which demonstrate the efficacy of their layered architecture for IDS. This approach can be useful in ensuring the security of healthcare data in remote healthcare systems, where data is transmitted over networks and is vulnerable to cyber-attacks.

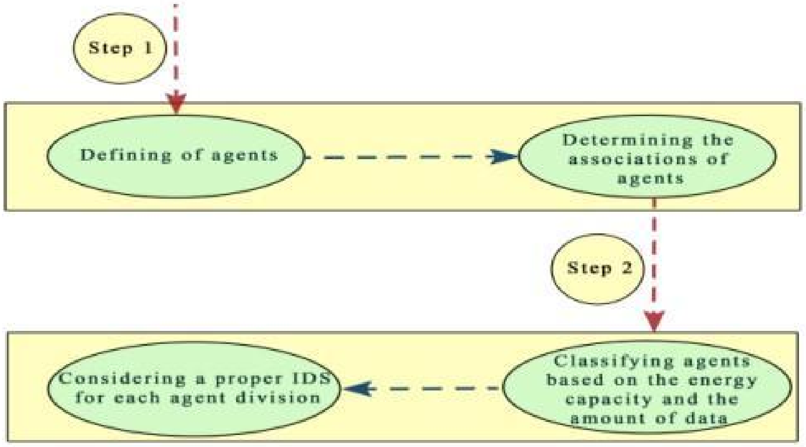

Step 1
Defining of agents
Determining the associations of agents
Step 2
Classifying agents based on the energy capacity and the amount of data
Considering a proper IDS for each agent division

Figure 1. Guideline of our proposed framework.

Outcome: The proposed framework achieves a high accuracy rate and low false positive rate.

Drawbacks: The evaluation is not done with real medical data.

23. Hatzivasilis, George, Othonas Soultatos, Sotiris Ioannidis, Christos Verikoukis, Giorgos Demetriou, and Christos Tsatsoulis. "Review of security and privacy for the Internet of Medical Things (IoT)." In *2019 15th international conference on distributed computing in sensor systems (DCOSS)*, pp. 457-464. IEEE, 2019.

This paper emphasizes the importance of security and privacy controls in IoT settings to protect users and stakeholders. It provides an overview of fundamental defence mechanisms that should be implemented to ensure complete security and privacy. The healthcare sector is protected by all-by-design methods that shield the user/patient from a wide range of attacks and dangers. The paper also discusses the potential for secure functionality at each layer and explores contemporary solutions. The study can serve as a best-practices manual for general IoT or specific IoT applications, while also taking into account the CE viewpoint. Overall, this paper provides valuable insights for ensuring security and privacy in IoT systems.

24. Engineer, Margi, Razma Tusha, Ankit Shah, and Kinjal Adhvaryu. "Insight into the importance of fog computing in Internet of Medical Things (IoT)." In *2019 International*

Conference on Recent Advances in Energy-efficient Computing and Communication (ICRAECC), pp. 1-7. IEEE, 2019.

The paper highlights the importance of the fog layer in IoT applications and its ability to provide low latency, quick decision-making, and high-quality service. It suggests that the fog-based architecture offers faster response times compared to previous cloud-based IoT architecture. The research also emphasizes that the fog layer's distributed storage approach is better than centralized storage in situations like network breakdown or task overload since it brings data closer to the end-users. This ability to accurately and quickly predict significant healthcare events can help save lives and facilitate timely decision-making.

25. Saba, Tanzila. "Intrusion detection in smart city hospitals using ensemble classifiers." In *2020 13th International Conference on Developments in eSystems Engineering (DeSE)*, pp. 418-422. IEEE, 2020.

The presented approach for identifying malicious attacks through an intrusion detection system involves reducing features and training ensemble classifiers. The KDD-99 dataset is used to evaluate this approach, and is preprocessed before feeding it to the classifier. The behavior of IoMT devices in the environment is then analyzed using ensemble methodologies to detect any malicious activity. The results show that this approach can be used to establish a reliable and consistent network for smart healthcare, with a high level of accuracy (93.2%) in identifying malicious activity.

26. Saranya, T., S. Sridevi, C. Deisy, Tran Duc Chung, and MKA Ahamed Khan. "Performance analysis of machine learning algorithms in intrusion detection system: A review." *Procedia Computer Science* 171 (2020): 1251-1260.

The suggestion is to compare various machine learning techniques for intrusion detection systems, including Linear Discriminant Analysis (LDA), classification and regression trees (CART), and Random Forest, and to use them for categorizing intrusions. To evaluate the effectiveness, the KDD-CUP dataset is employed in the study. The results reveal that the accuracy, false positive rate, and detection rate are dependent on both the algorithm used and the domain of application.

Drawbacks: A real-time dataset was not used for the evaluation.

27. Aladaileh, Mohammad A., Mohammed Anbar, Iznan H. Hasbullah, Yung-Wey Chong, and Yousef K. Sanjalawe. "Detection techniques of distributed denial of service attacks on software-defined networking controller–a review." *IEEE Access* 8 (2020): 143985-143995.

This paper provides an introduction to SDN and emphasizes the importance of SDN features in network management, monitoring, and programming using the SDN controller. The controller also plays a vital role in protecting the network from various attacks, as described in section II. In the third section, the paper examines the security issues that the SDN controller faces, particularly with respect to DDoS attacks, and delves into various types of DDoS attacks. The fourth section of the paper thoroughly analyzes existing DDoS detection methods and compares them based on predetermined standards. The article also categorizes DDoS attack detection methods currently in use based on their technique, threshold nature, and deployment location in the SDN environment. The study highlights the drawbacks of several detection methods and suggests using a more effective method to improve detection accuracy and lower false positive rates. Combining the strengths of existing systems can lead to a more comprehensive detection strategy against DDoS attacks, which researchers could pursue.

28. Ijaz, Muhammad, Gang Li, Ling Lin, Omar Cheikhrouhou, Habib Hamam, and Alam Noor. "Integration and applications of fog computing and cloud computing based on the internet of things for provision of healthcare services at home." *Electronics* 10, no. 9 (2021): 1077.

This paper focuses on the application of cloud computing and fog computing in the healthcare industry to process, compute, store, and share resources. These technologies enable easier access to IoT-based applications, and the paper highlights their effectiveness in developing healthcare models and frameworks. The study proposes a framework for home hospitalization that utilizes both cloud computing and fog computing approaches, including an environment sensing system and Android applications. The framework's usability for patients and healthcare professionals was evaluated using the system suitability scale (SUS).

29. Raju, K. Butchi, Suresh Dara, Ankit Vidyarthi, V. MNSSVKR Gupta, and Baseem Khan. "Smart Heart Disease Prediction System with IoT and Fog Computing Sectors Enabled

by Cascaded Deep Learning Model." *Computational Intelligence and Neuroscience* 2022 (2022).

This paper focuses on the application of cloud computing and fog computing in the healthcare industry to process, compute, store, and share resources. These technologies enable easier access to IoT-based applications, and the paper highlights their effectiveness in developing healthcare models and frameworks. The study proposes a framework for home hospitalization that utilizes both cloud computing and fog computing approaches, including an environment sensing system and Android applications. The framework's usability for patients and healthcare professionals was evaluated using the system suitability scale (SUS).

PROPOSED MODEL (Attack Detection)

- Datasets: NSL-KDD dataset
- Ensemble model with learners: Decision tree, random forest, naïve bayes
- Reduction process through feature selection on the dataset

ARCHITECTURE DIAGRAM

The NSL-KDD dataset was used as input and preprocessed through feature selection and encoding, with any null values removed. Three base classifiers (KNN, Decision Tree, and Random Forest) were implemented simultaneously, and their outputs were fed into the meta classifier AdaBoost Classifier. This architecture was then stacked into an ensemble model.

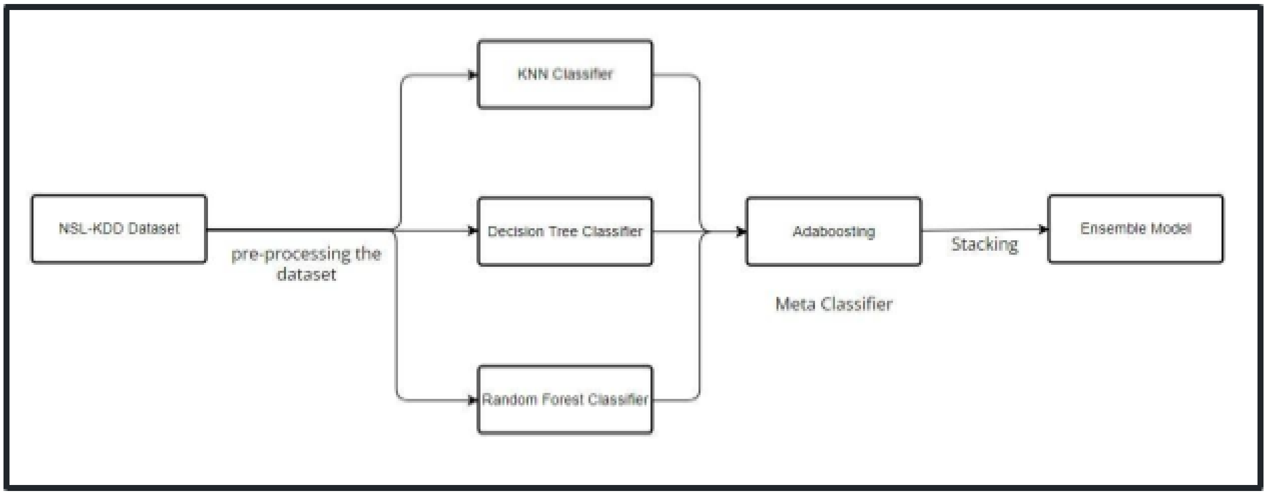
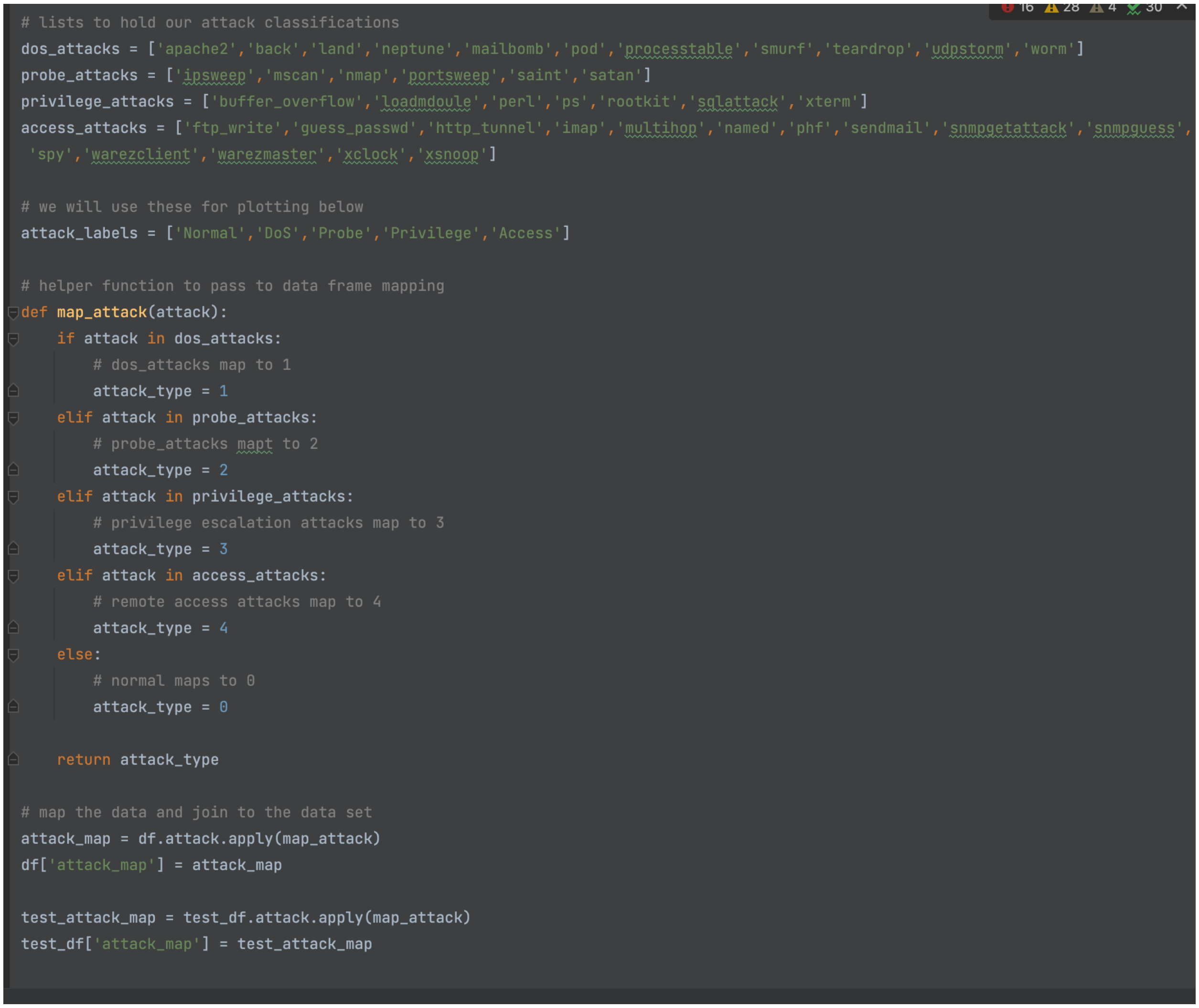


Figure 1. Architecture Diagram

SCREENSHOTS

Preprocessing:



```
# lists to hold our attack classifications
dos_attacks = ['apache2','back','land','neptune','mailbomb','pod','processtable','smurf','teardrop','udpstorm','worm']
probe_attacks = ['ipsweep','mscan','nmap','portsweep','saint','satan']
privilege_attacks = ['buffer_overflow','loadmdoule','perl','ps','rootkit','sqlattack','xterm']
access_attacks = ['ftp_write','guess_passwd','http_tunnel','imap','multihop','named','phf','sendmail','snmpgetattack','snmpguess',
 'spy','warezclient','warezmaster','xclock','xsnoop']

# we will use these for plotting below
attack_labels = ['Normal','DoS','Probe','Privilege','Access']

# helper function to pass to data frame mapping
def map_attack(attack):
    if attack in dos_attacks:
        # dos_attacks map to 1
        attack_type = 1
    elif attack in probe_attacks:
        # probe_attacks mapt to 2
        attack_type = 2
    elif attack in privilege_attacks:
        # privilege escalation attacks map to 3
        attack_type = 3
    elif attack in access_attacks:
        # remote access attacks map to 4
        attack_type = 4
    else:
        # normal maps to 0
        attack_type = 0

    return attack_type

# map the data and join to the data set
attack_map = df.attack.apply(map_attack)
df['attack_map'] = attack_map

test_attack_map = test_df.attack.apply(map_attack)
test_df['attack_map'] = test_attack_map
```

FIG A

Attacks and normal traffic

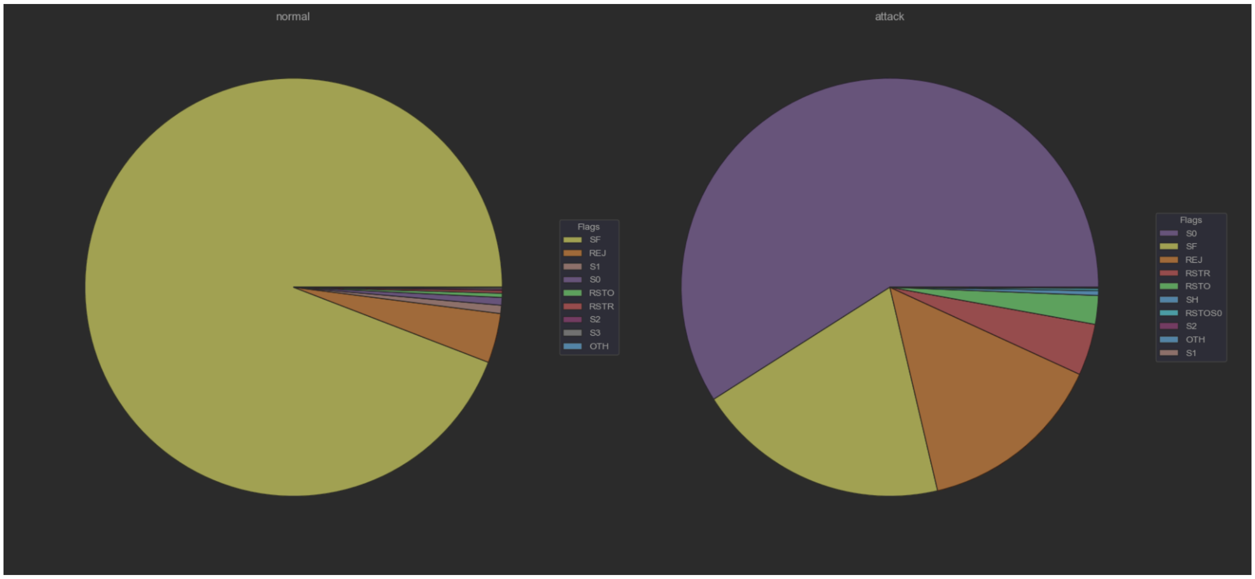
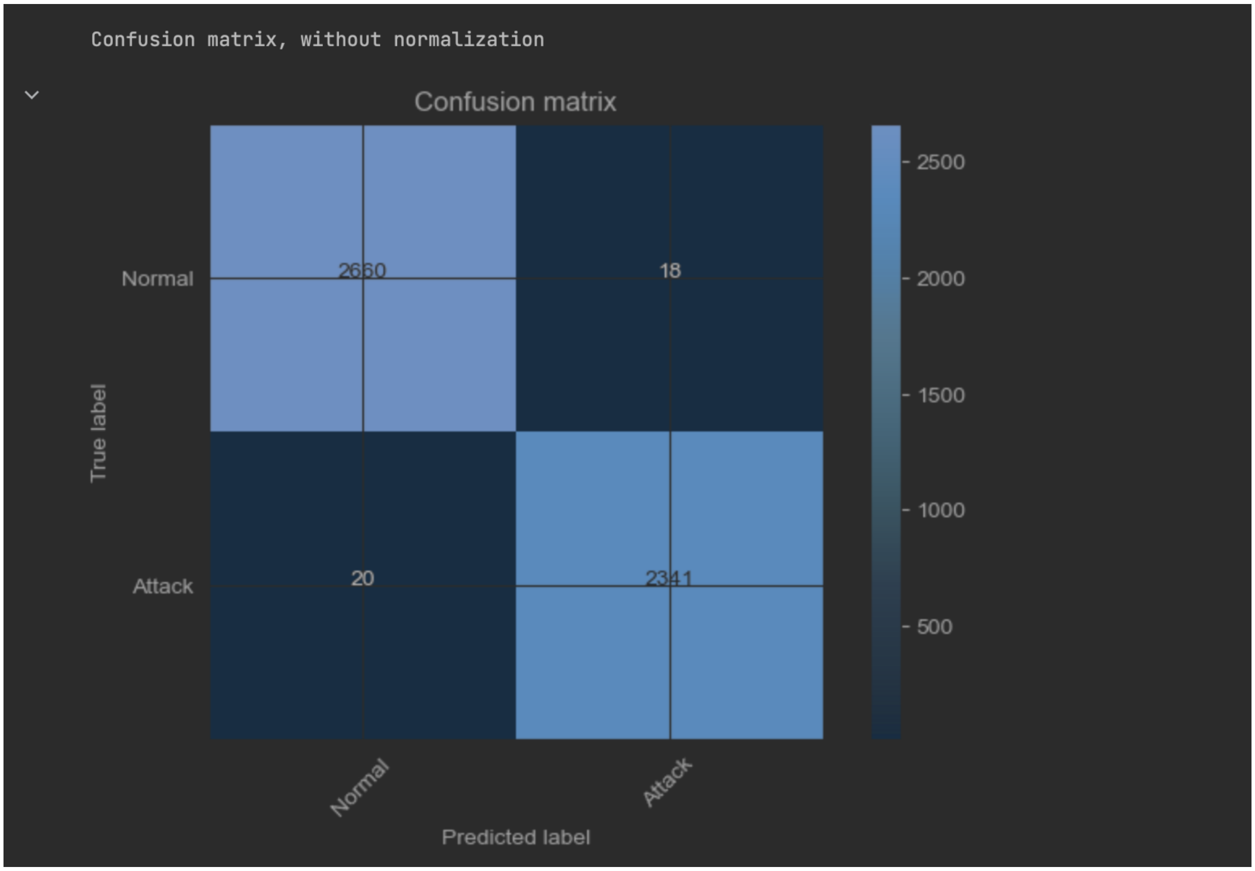


FIG B

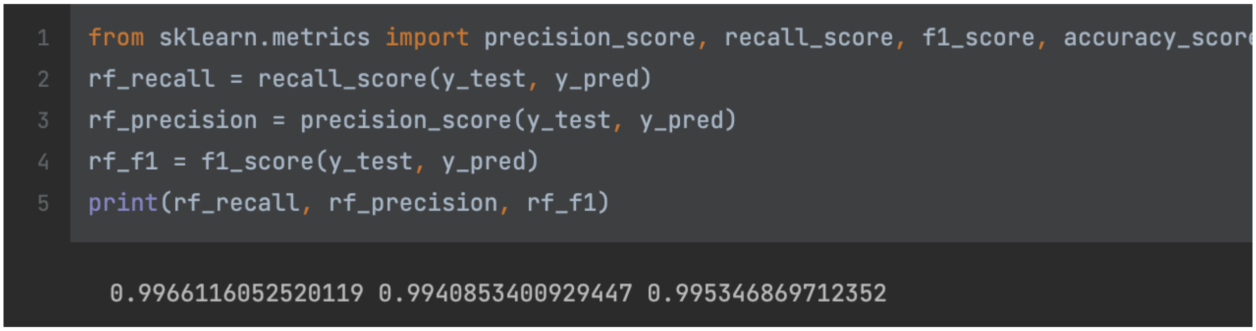
Adaboost



```
Confusion matrix, without normalization
```

FIG C

AdaBoost recall, precision, F1 Score.

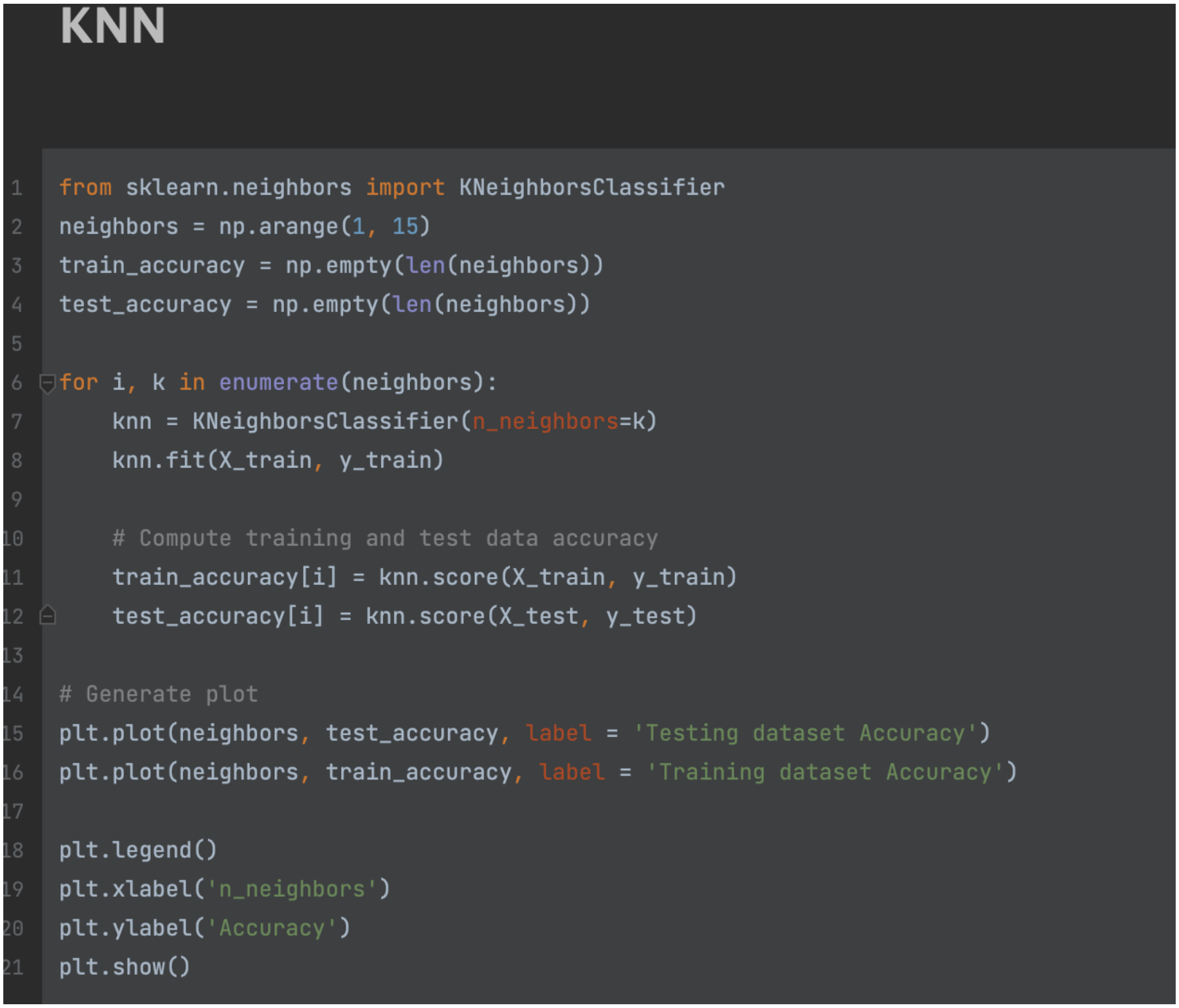


```
from sklearn.metrics import precision_score, recall_score, f1_score, accuracy_scor
rf_recall = recall_score(y_test, y_pred)
rf_precision = precision_score(y_test, y_pred)
rf_f1 = f1_score(y_test, y_pred)
print(rf_recall, rf_precision, rf_f1)
```

```
0.9966116052520119 0.9940853400929447 0.995346869712352
```

FIG D

KNN



KNN

```
from sklearn.neighbors import KNeighborsClassifier
neighbors = np.arange(1, 15)
train_accuracy = np.empty(len(neighbors))
test_accuracy = np.empty(len(neighbors))

for i, k in enumerate(neighbors):
    knn = KNeighborsClassifier(n_neighbors=k)
    knn.fit(X_train, y_train)

    # Compute training and test data accuracy
    train_accuracy[i] = knn.score(X_train, y_train)
    test_accuracy[i] = knn.score(X_test, y_test)

# Generate plot
plt.plot(neighbors, test_accuracy, label = 'Testing dataset Accuracy')
plt.plot(neighbors, train_accuracy, label = 'Training dataset Accuracy')

plt.legend()
plt.xlabel('n_neighbors')
plt.ylabel('Accuracy')
plt.show()
```

FIG E

Accuracy Over Dataset

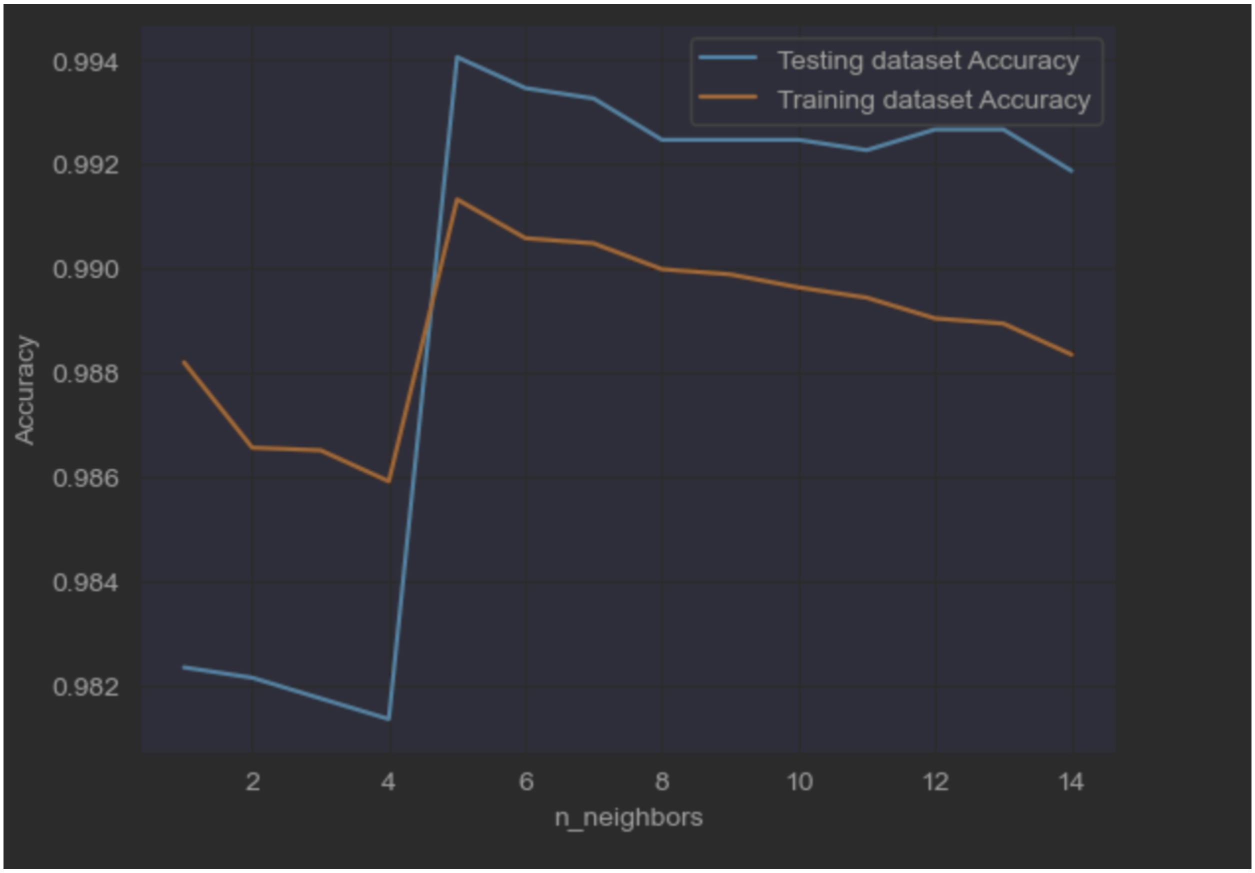
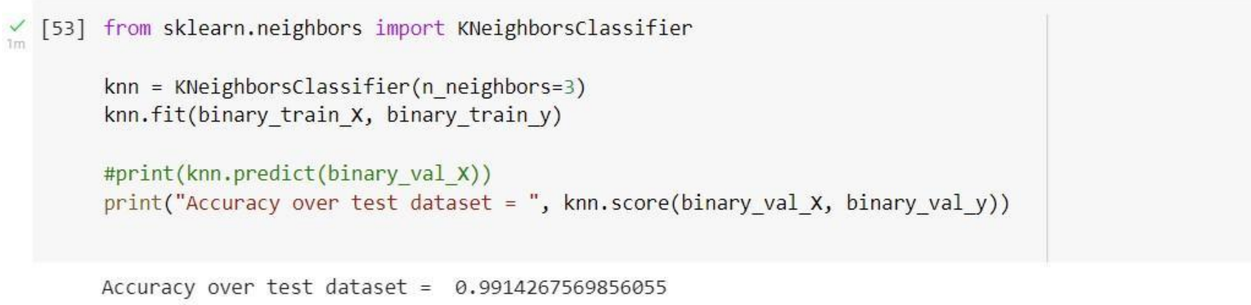


FIG F

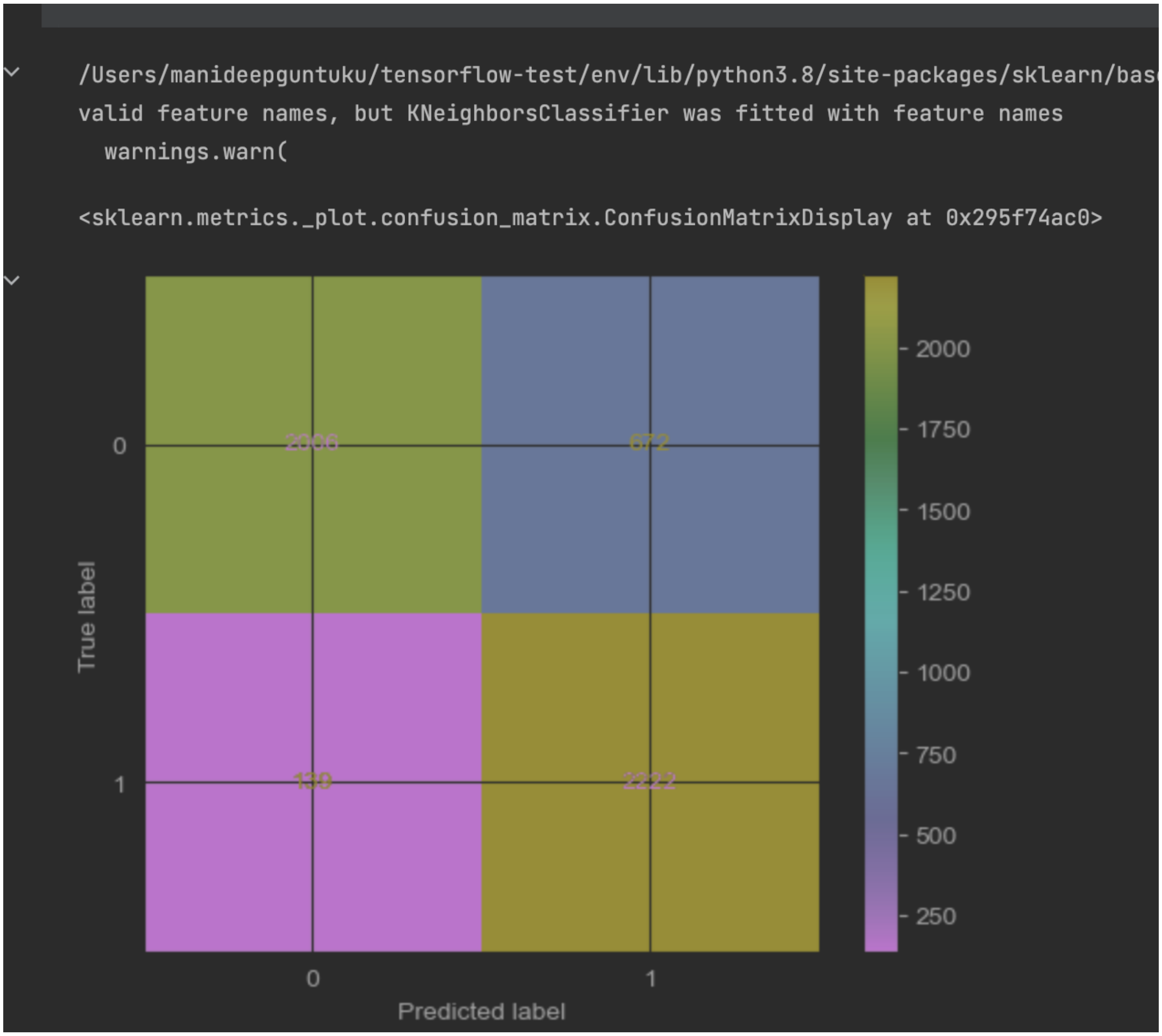


```
[53] from sklearn.neighbors import KNeighborsClassifier

     knn = KNeighborsClassifier(n_neighbors=3)
     knn.fit(binary_train_X, binary_train_y)

     #print(knn.predict(binary_val_X))
     print("Accuracy over test dataset = ", knn.score(binary_val_X, binary_val_y))
```

```
Accuracy over test dataset =  0.9914267569856055
```

From the above graph, we can conclude that **k=3** will give the highest accuracy for knn model over test dataset.

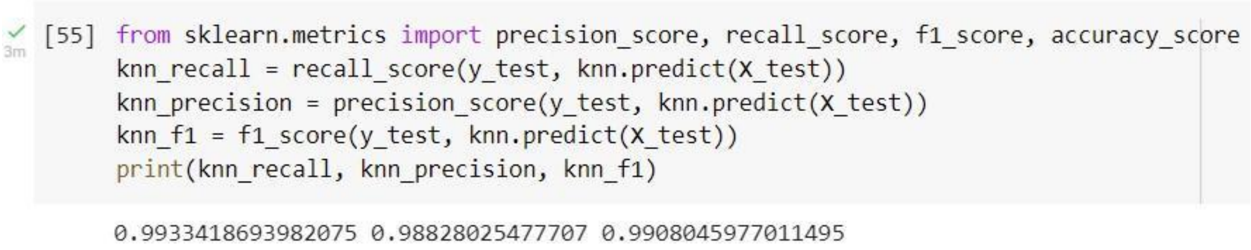
KNN confusion matrix



```
/Users/manideepguntuku/tensorflow-test/env/lib/python3.8/site-packages/sklearn/bas
valid feature names, but KNeighborsClassifier was fitted with feature names
  warnings.warn(

<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x295f74ac0>
```

FIG G

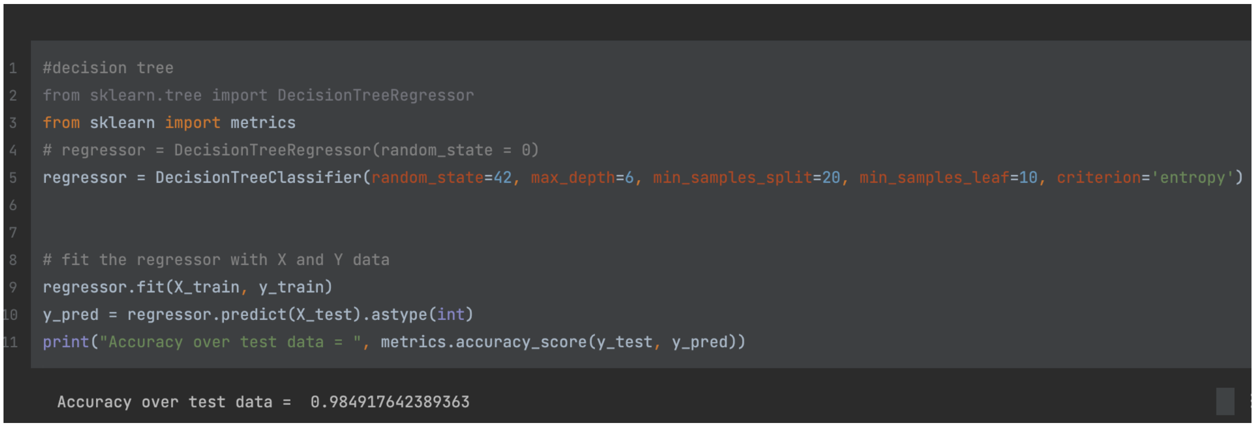
KNN recall, precision, F1 score.



```
[55] from sklearn.metrics import precision_score, recall_score, f1_score, accuracy_score
     knn_recall = recall_score(y_test, knn.predict(X_test))
     knn_precision = precision_score(y_test, knn.predict(X_test))
     knn_f1 = f1_score(y_test, knn.predict(X_test))
     print(knn_recall, knn_precision, knn_f1)

     0.9933418693982075 0.98828025477707 0.9908045977011495
```

Decision tree

Decision tree accuracy



```
#decision tree
from sklearn.tree import DecisionTreeRegressor
from sklearn import metrics
# regressor = DecisionTreeRegressor(random_state = 0)
regressor = DecisionTreeClassifier(random_state=42, max_depth=6, min_samples_split=20, min_samples_leaf=10, criterion='entropy')


# fit the regressor with X and Y data
regressor.fit(X_train, y_train)
y_pred = regressor.predict(X_test).astype(int)
print("Accuracy over test data = ", metrics.accuracy_score(y_test, y_pred))
```

```
Accuracy over test data =  0.984917642389363
```

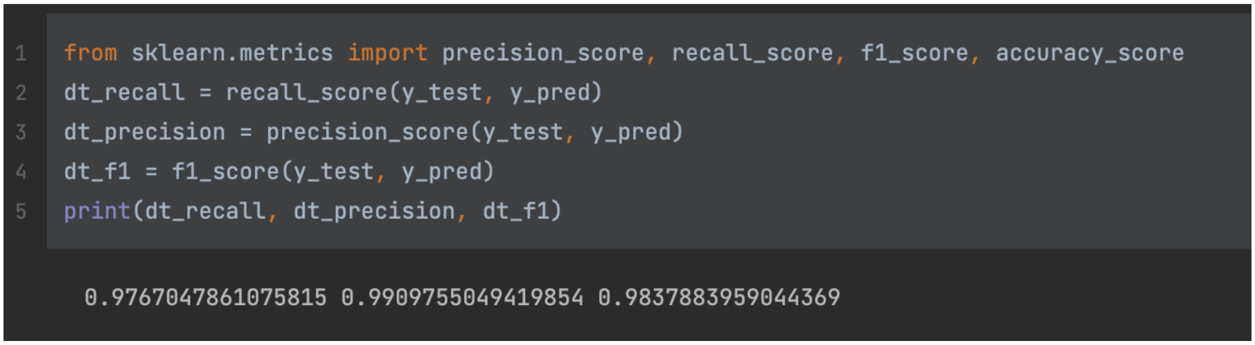
FIG H

Decision tree confusion matrix

	precision	recall	f1-score	support
0	8.98	0.99	0.99	2678
1	0.99	0.98	0.98	2361
accuracy			0.98	5039
macro avg	8.99	0.98	0.98	5839
weighted avg	0.98	0.98	0.98	5839

FIG I

Decision tree recall, precision, F1 score.



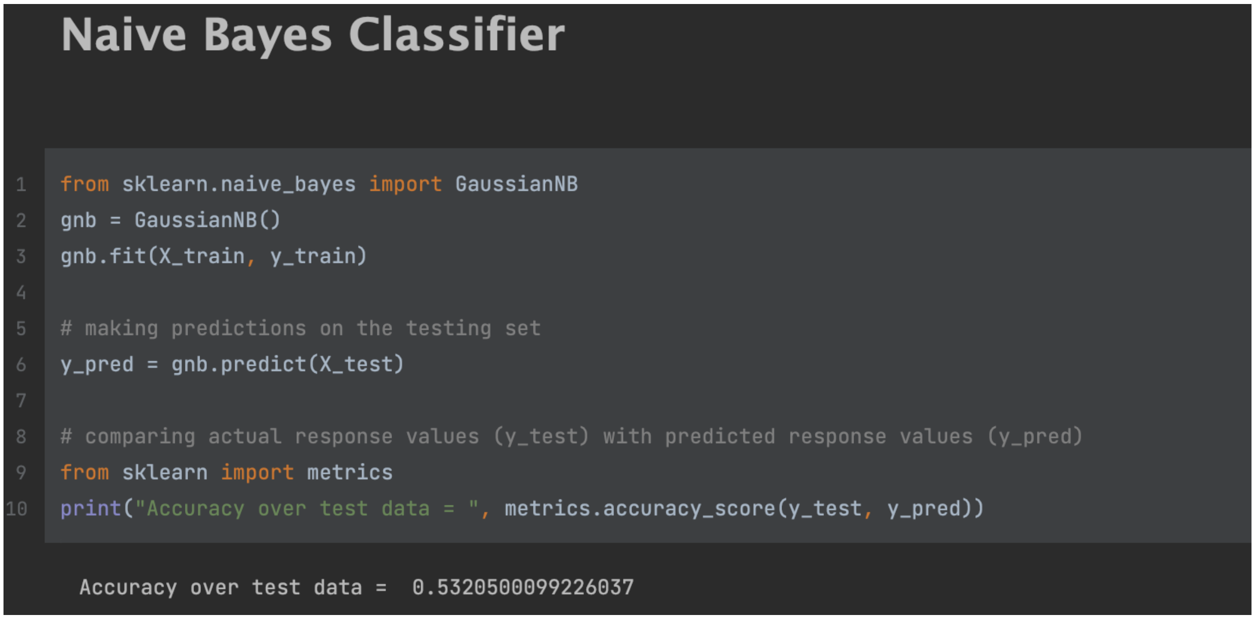
```
from sklearn.metrics import precision_score, recall_score, f1_score, accuracy_score
dt_recall = recall_score(y_test, y_pred)
dt_precision = precision_score(y_test, y_pred)
dt_f1 = f1_score(y_test, y_pred)
print(dt_recall, dt_precision, dt_f1)
```

```
0.9767047861075815 0.9909755049419854 0.9837883959044369
```

FIG J

Naive Bayes classifier

Naive Bayes accuracy



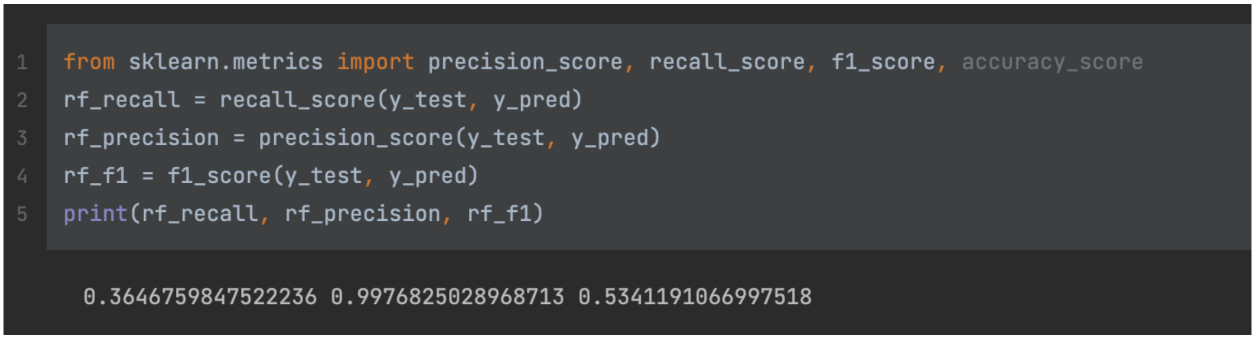
Naive Bayes Classifier

```
from sklearn.naive_bayes import GaussianNB
gnb = GaussianNB()
gnb.fit(X_train, y_train)

# making predictions on the testing set
y_pred = gnb.predict(X_test)

# comparing actual response values (y_test) with predicted response values (y_pred)
from sklearn import metrics
print("Accuracy over test data = ", metrics.accuracy_score(y_test, y_pred))
```

```
Accuracy over test data =  0.5320500099226037
```

FIG K

Naive bayes recall, precision, F1 score.

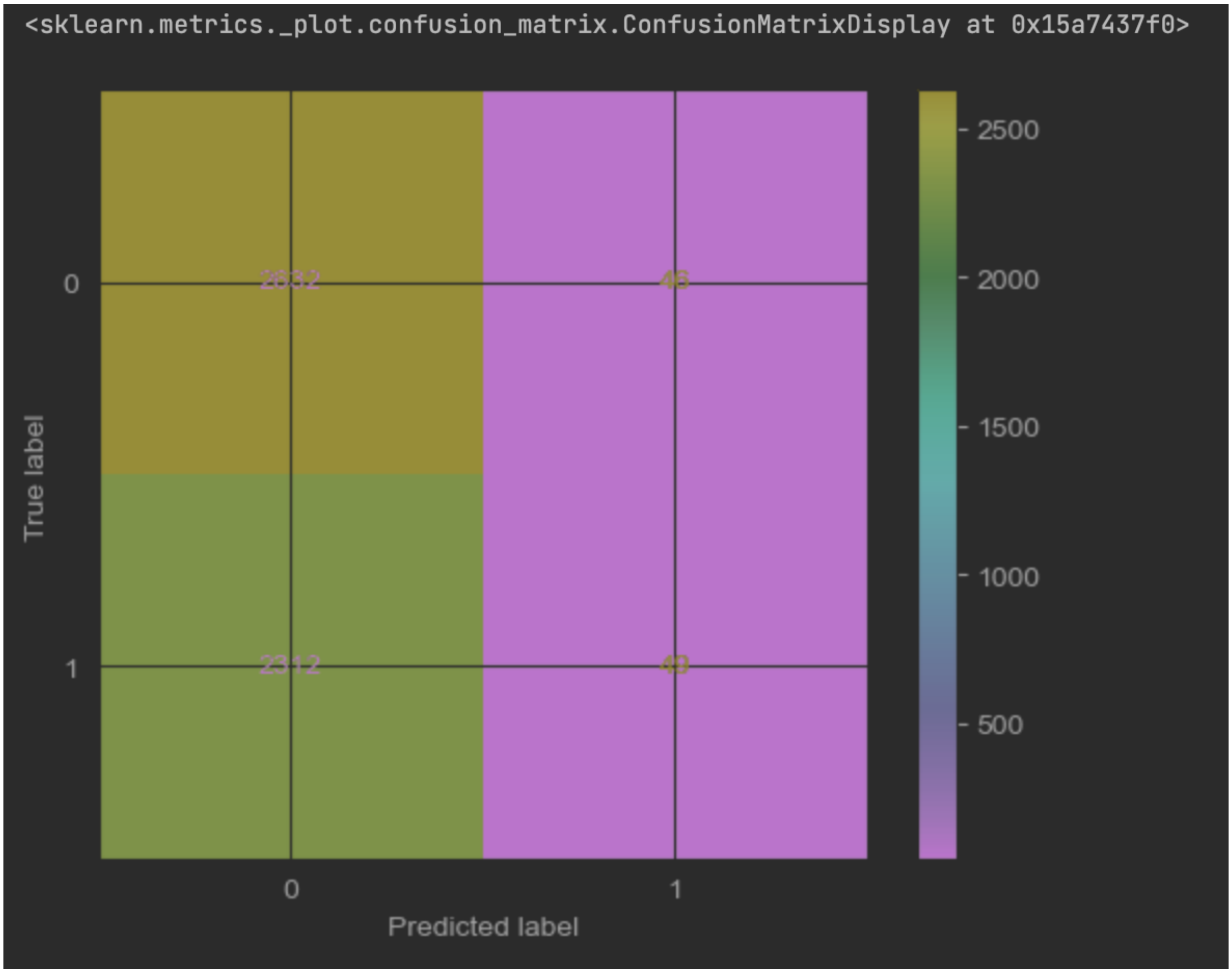


```
from sklearn.metrics import precision_score, recall_score, f1_score, accuracy_score
rf_recall = recall_score(y_test, y_pred)
rf_precision = precision_score(y_test, y_pred)
rf_f1 = f1_score(y_test, y_pred)
print(rf_recall, rf_precision, rf_f1)
```

```
0.3646759847522236 0.9976825028968713 0.5341191066997518
```

FIG L

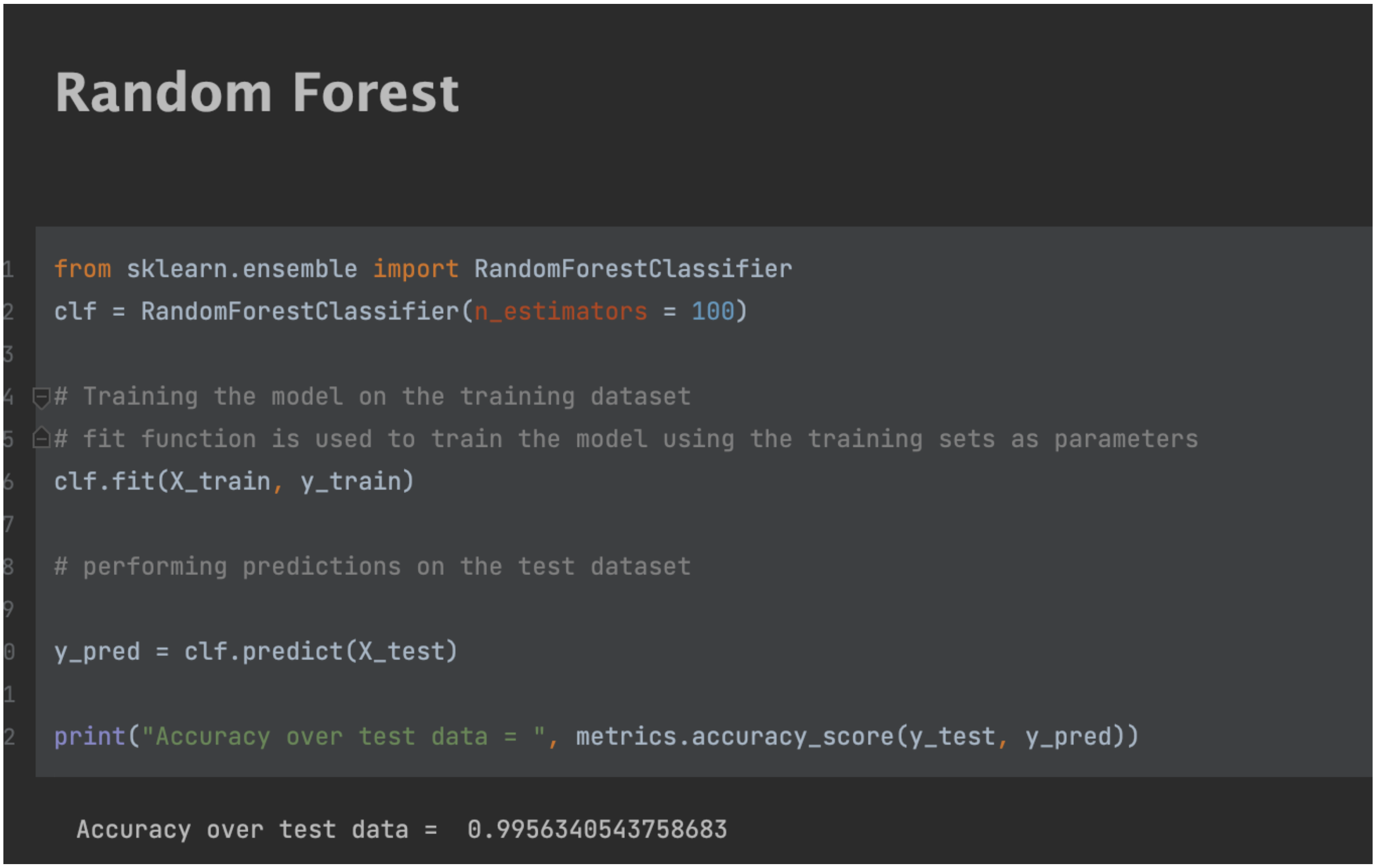
Naive Bayes Confusion matrix



```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x15a7437f0>
```

FIG M

Random forest Accuracy



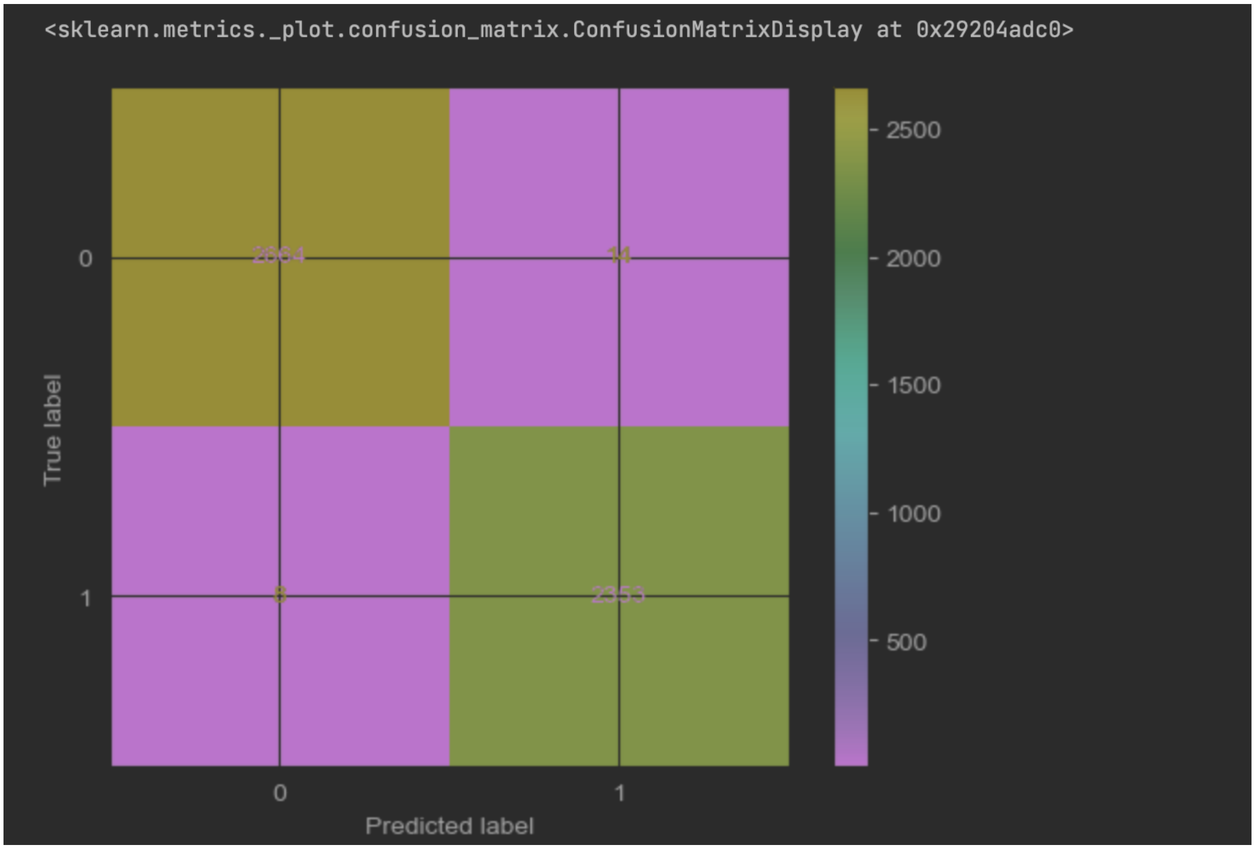
Random Forest

```
from sklearn.ensemble import RandomForestClassifier
clf = RandomForestClassifier(n_estimators = 100)

# Training the model on the training dataset
# fit function is used to train the model using the training sets as parameters
clf.fit(X_train, y_train)

# performing predictions on the test dataset

y_pred = clf.predict(X_test)

print("Accuracy over test data = ", metrics.accuracy_score(y_test, y_pred))
```

```
Accuracy over test data =  0.9956340543758683
```

FIG N

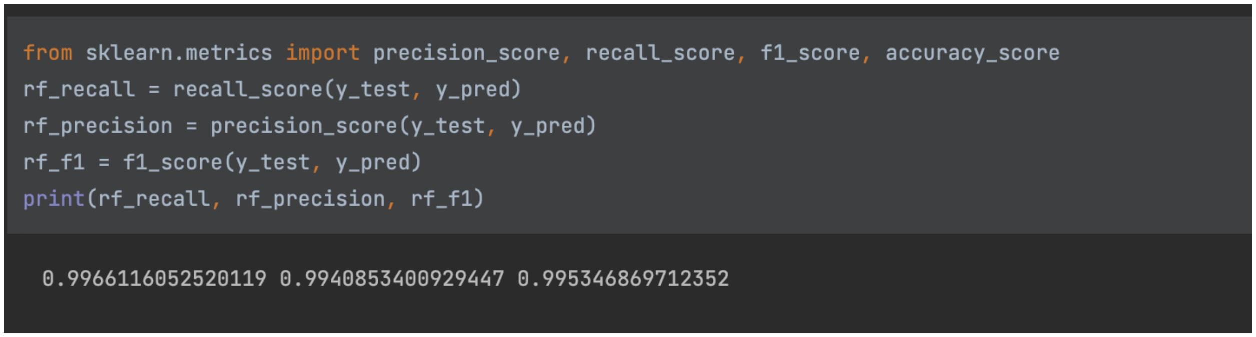
Random forest confusion matrix



```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x29204adc0>
```

FIG O

Random forest recall, precision, F1 score

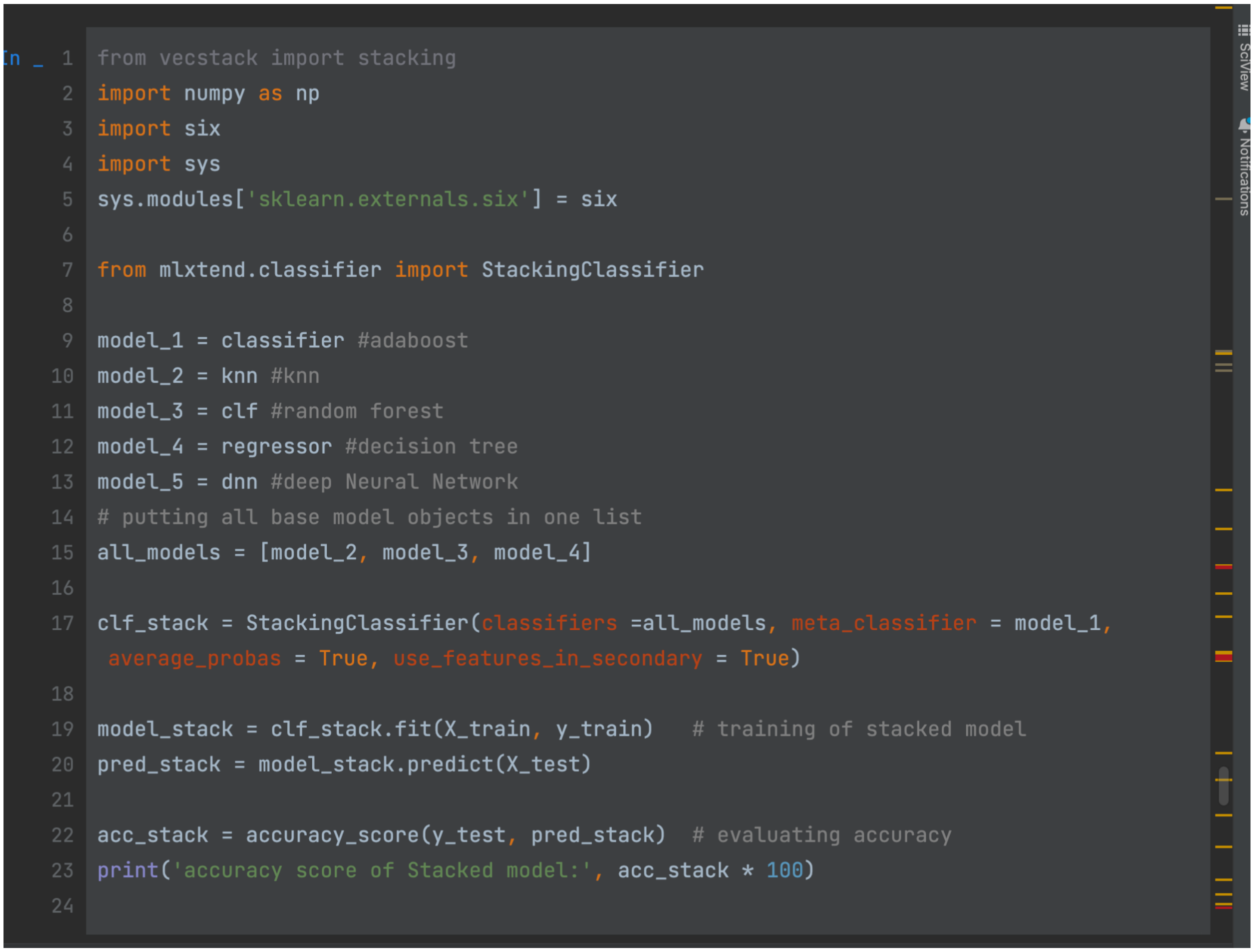


```
from sklearn.metrics import precision_score, recall_score, f1_score, accuracy_score
rf_recall = recall_score(y_test, y_pred)
rf_precision = precision_score(y_test, y_pred)
rf_f1 = f1_score(y_test, y_pred)
print(rf_recall, rf_precision, rf_f1)
```

```
0.9966116052520119 0.9940853400929447 0.995346869712352
```

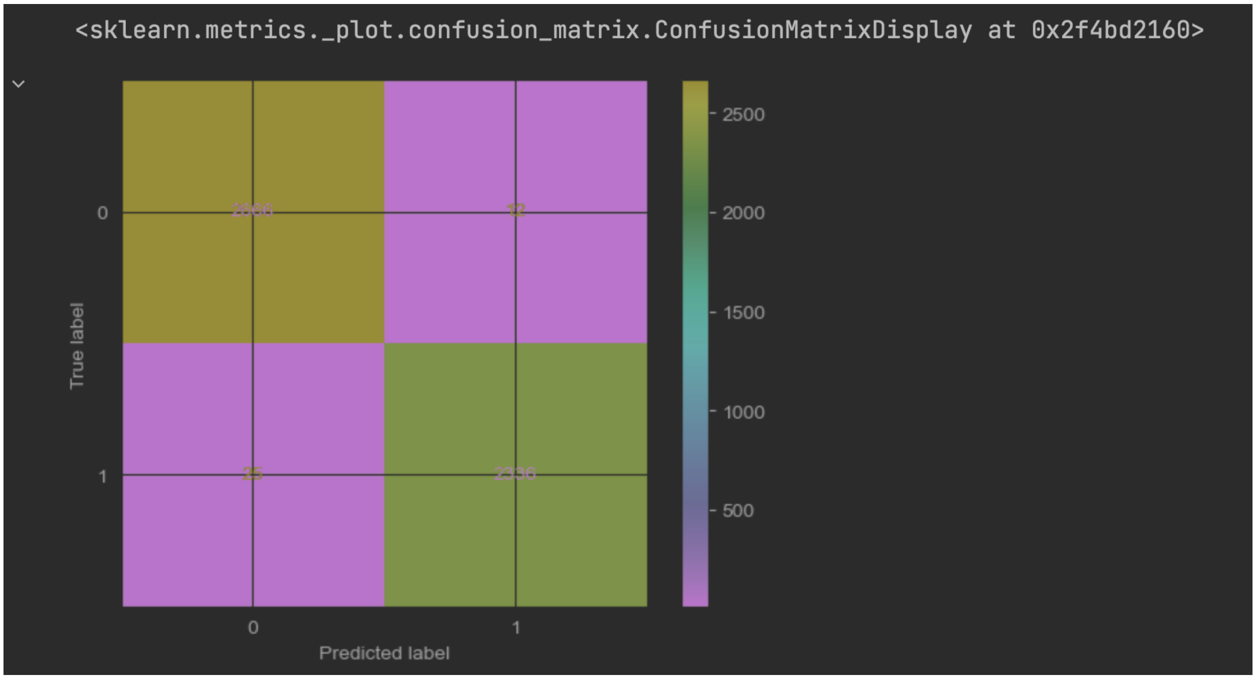
FIG P

Ensemble



```
from vecstack import stacking
import numpy as np
import six
import sys
sys.modules['sklearn.externals.six'] = six

from mlxtend.classifier import StackingClassifier

model_1 = classifier #adaboost
model_2 = knn #knn
model_3 = clf #random forest
model_4 = regressor #decision tree
model_5 = dnn #deep Neural Network
# putting all base model objects in one list
all_models = [model_2, model_3, model_4]

clf_stack = StackingClassifier(classifiers =all_models, meta_classifier = model_1,
 average_probas = True, use_features_in_secondary = True)

model_stack = clf_stack.fit(X_train, y_train)   # training of stacked model
pred_stack = model_stack.predict(X_test)

acc_stack = accuracy_score(y_test, pred_stack)  # evaluating accuracy
print('accuracy score of Stacked model:', acc_stack * 100)

```

FIG Q



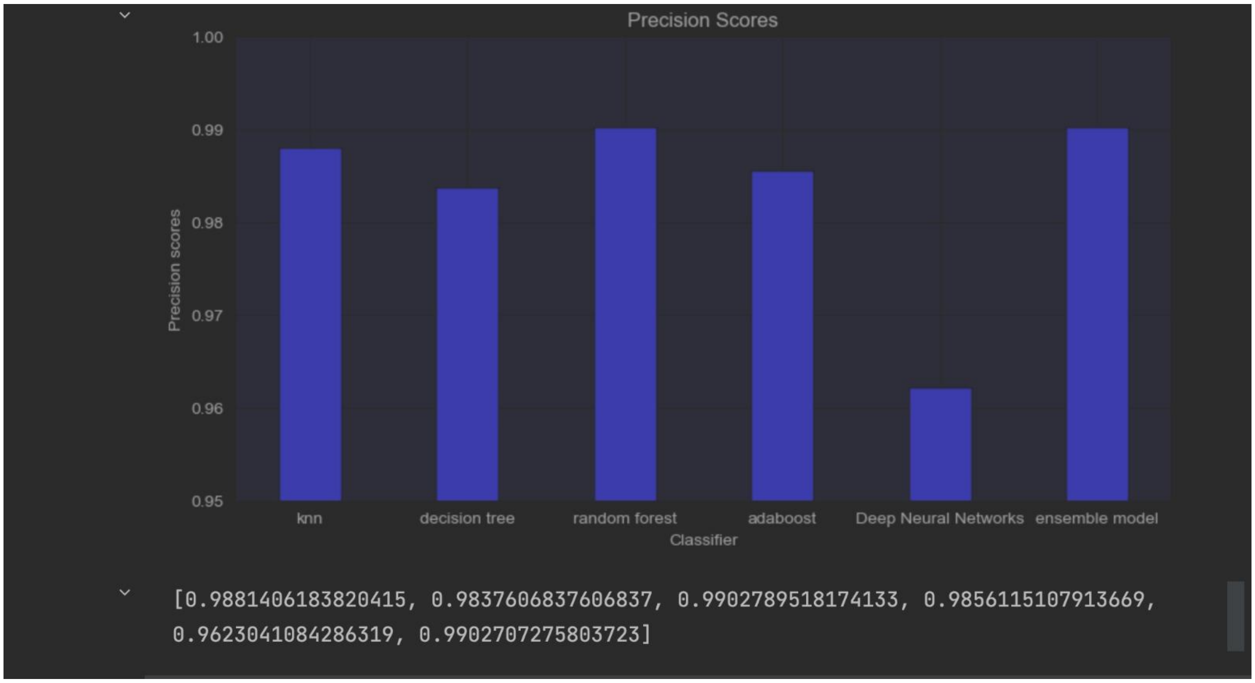
```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x2f4bd2160>
```

FIG R

	precision	recall	f1-score	support
0	0.99	1.00	0.99	2678
1	0.99	0.99	0.99	2361
accuracy			0.99	5039
macro avg	0.99	0.99	0.99	5039
weighted avg	0.99	0.99	0.99	5039

FIG S

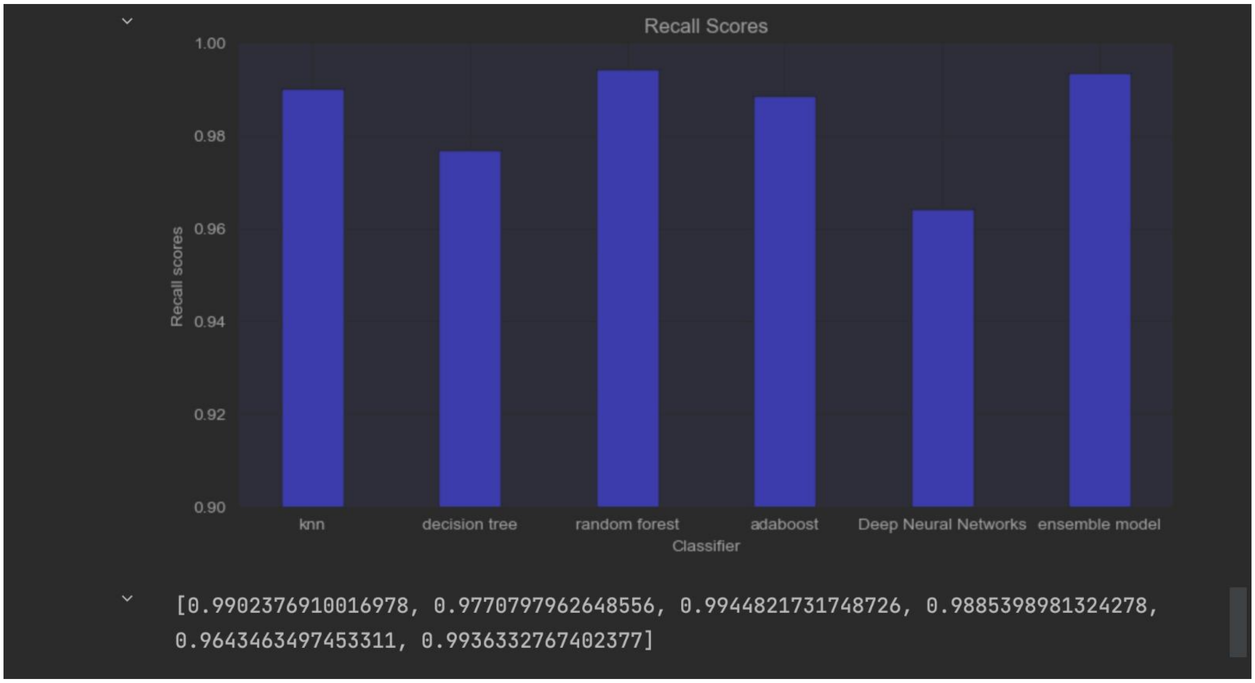
Visualization



```
[0.9881406183820415, 0.9837606837606837, 0.9902789518174133, 0.9856115107913669,
0.9623041084286319, 0.9902707275803723]
```

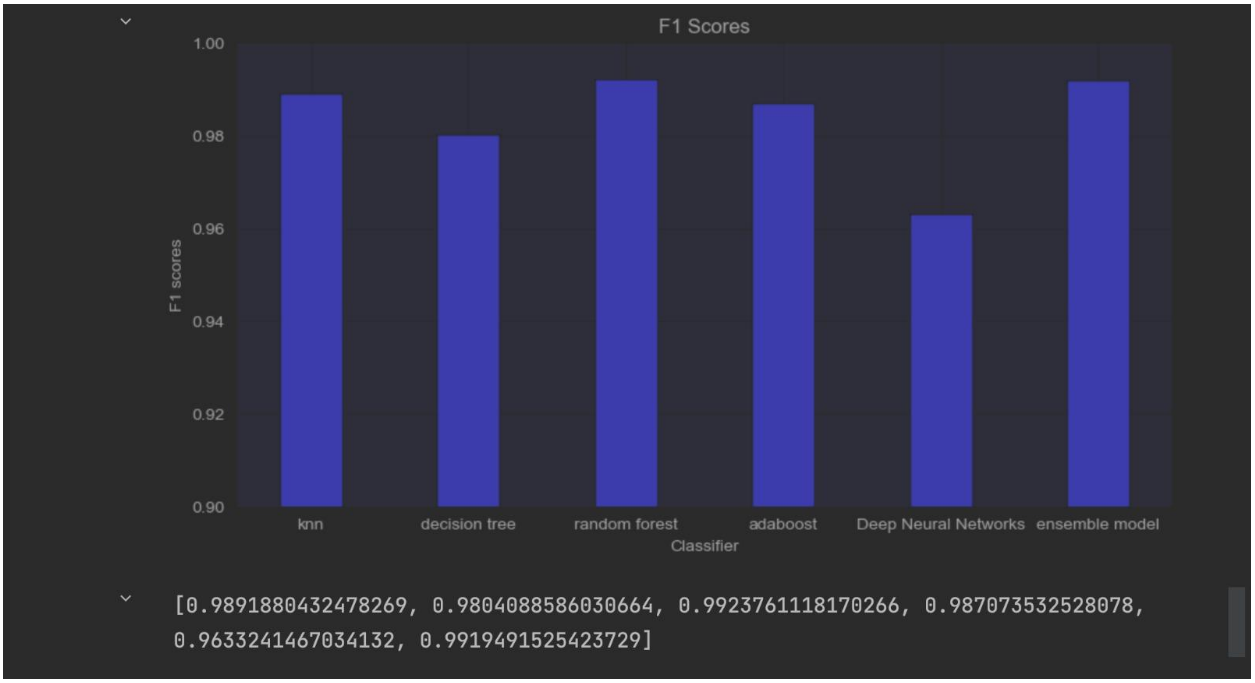
FIG T

(We have compared the Precision scores & Recall scores of the three classifiers (i.e. Random forest, KNN and Decision tree) , one meta classifier (i.e. AdaBoost),DNN with the ensemble model)



```
[0.9902376910016978, 0.9770797962648556, 0.9944821731748726, 0.9885398981324278,
0.9643463497453311, 0.9936332767402377]
```

FIG U

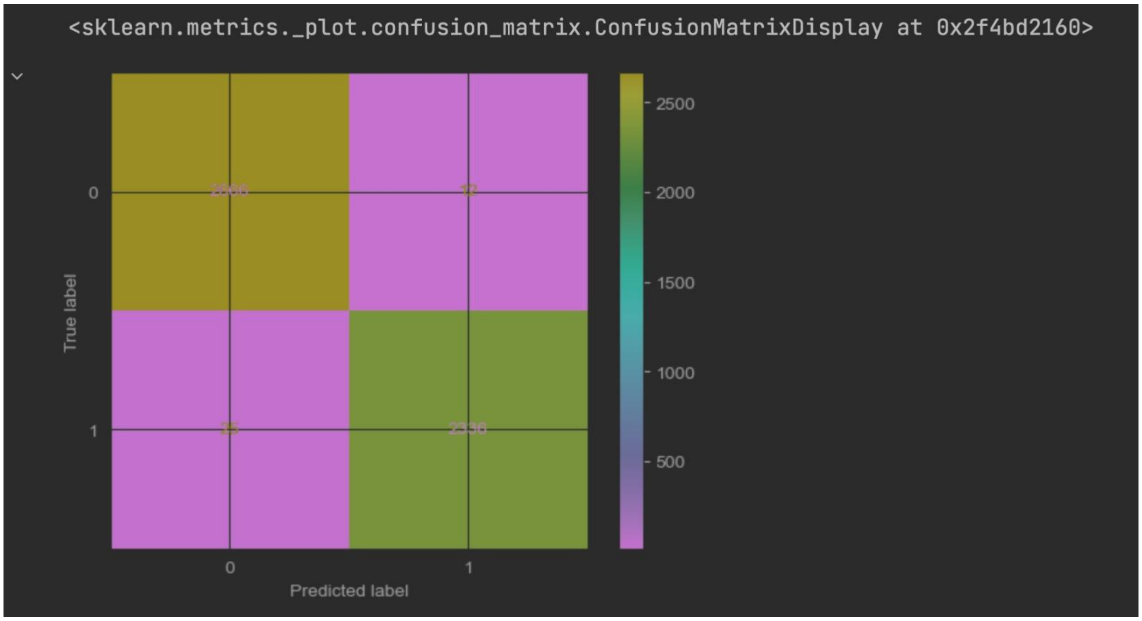


```
[0.9891880432478269, 0.9804088586030664, 0.9923761118170266, 0.987073532528078,
0.9633241467034132, 0.9919491525423729]
```

FIG V

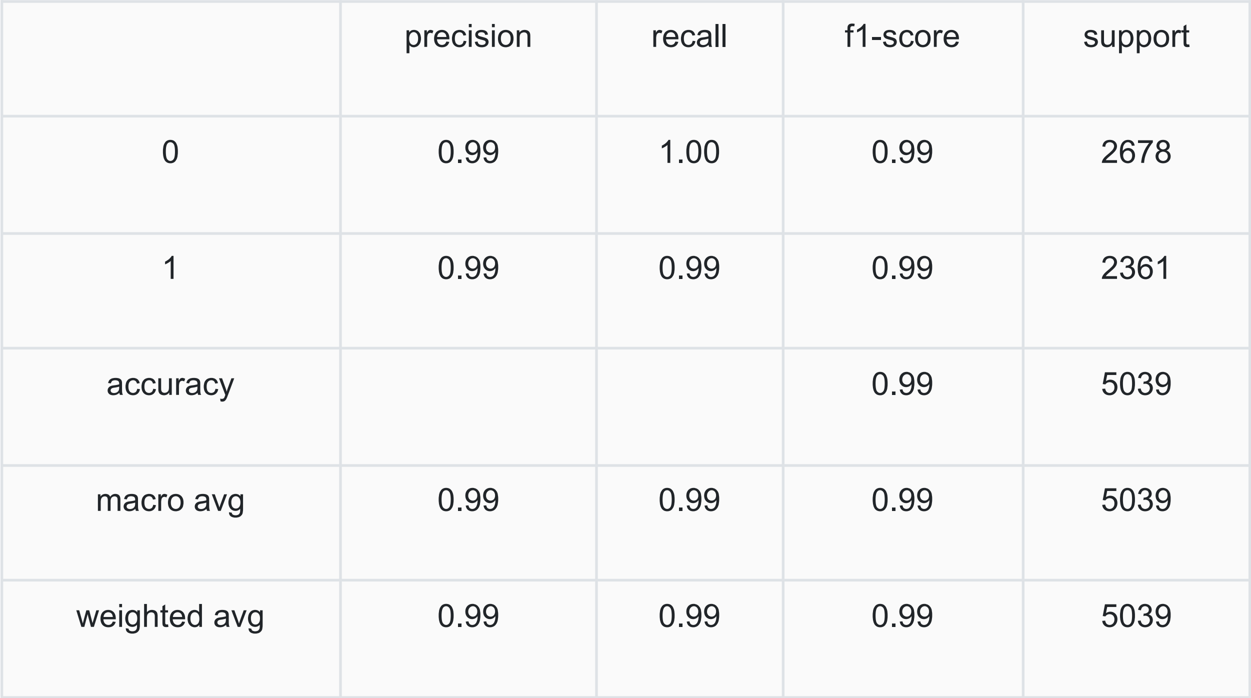
(We have compared the F1 scores of the three classifiers (i.e. Random forest, KNN and Decision tree) , one meta classifier (i.e. AdaBoost) with the ensemble model.)

RESULTS



```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x2f4bd2160>
```

FIG W



	precision	recall	f1-score	support
0	0.99	1.00	0.99	2678
1	0.99	0.99	0.99	2361
accuracy			0.99	5039
macro avg	0.99	0.99	0.99	5039
weighted avg	0.99	0.99	0.99	5039

FIG X

CONCLUSION

We undertook a project to develop a model for detecting intrusion attacks in IoT networks by evaluating the performance of four classifiers: KNN, Decision Tree, Random Forest, and AdaBoost on the widely used NSL-KDD dataset. After analyzing the accuracy, precision, recall, and F1-score of each classifier, we decided to implement ensemble learning to enhance the accuracy and robustness of the intrusion detection system.

Our research demonstrates that the ensemble learning approach is an effective and practical method for detecting intrusion attacks in IoT networks. By combining the strengths of multiple models, our ensemble learning model outperformed individual classifiers, resulting in higher accuracy and detection rates.

The results of our research provide valuable insights for future researchers and practitioners interested in creating more advanced and precise intrusion detection systems for IoT networks.

REFERENCES

[0] Prabhat Kumar, Govind P. Gupta, Rakesh Tripathi, "An ensemble learning and fog-cloud architecture-driven cyber-attack detection framework for IoMT networks", *Computer Communications*, Volume 166, 2021, Pages 110-124, ISSN 0140-3664, https://doi.org/10.1016/j.comcom.2020.12.003.

[1] S. Manimurugan, "IoT-Fog-Cloud model for anomaly detection using improved Naïve Bayes and principal component analysis", *Journal of Ambient Intelligence and Humanized Computing*, bll 1–10, 2021.

[2] M. Saharkhizan, A. Azmoodeh, A. Dehghantanha, K.-K. R. Choo, en R. M. Parizi, "An ensemble of deep recurrent neural networks for detecting IoT cyber attacks using network traffic", *IEEE Internet of Things Journal*, vol 7, no 9, bll 8852–8859, 2020.

[3] S. Khare en M. Totaro, “Ensemble learning for detecting attacks and anomalies in iot smart home”, in *2020 3rd International Conference on Data Intelligence and Security (ICDIS)*, 2020, bll 56–63.

[4] A. Verma en V. Ranga, “ELNIDS: Ensemble learning based network intrusion detection system for RPL based Internet of Things”, in *2019 4th International conference on Internet of Things: Smart innovation and usages (IoT-SIU)*, 2019, bll 1–6.

[5] R. Maharaja, P. Iyer, en Z. Ye, “A hybrid fog-cloud approach for securing the Internet of Things”, *Cluster Computing*, vol 23, no 2, bll 451–459, 2020.

[6] E. Tsogbaatar *et al.*, “DeL-IoT: A deep ensemble learning approach to uncover anomalies in IoT”, *Internet of Things*, vol 14, bl 100391, 2021.

[7] Bin Jia, Xiaohong Huang, Rujun Liu, Yan Ma, "A DDoS Attack Detection Method Based on Hybrid Heterogeneous Multiclassifier Ensemble Learning", *Journal of Electrical and Computer Engineering*, vol. 2017, Article ID 4975343, 9 pages, 2017. https://doi.org/10.1155/2017/4975343

[8] Hariharan Rajadurai, Usha Devi Gandhi, “A stacked ensemble learning model for intrusion detection in wireless network”, *Neural Comput & Applic (2020).* https://doi.org/10.1007/s00521-020-04986-5

[9] Yun Zhou, Peichao Wang, “An ensemble learning approach for XSS attack detection with domain knowledge and threat intelligence”, *Computers & Security,* Volume 82, 2019, Pages 261-269, ISSN 0167-4048, https://doi.org/10.1016/j.cose.2018.12.016.

[10] Poulmanogo Illy, Georges Kaddoum, Christian Miranda Moreira, Kuljeet Kaur, and Sahil Garg, “Securing Fog-to-Things Environment Using Intrusion Detection System Based On Ensemble Learning”, *2019 IEEE Wireless Communications and networking Conference (WCNC),* 2019, pp. 1-7, doi: 10.1109/WCNC.2019.8885534.

[11] Shilan S. Hameed, Wan Haslina Hassan, and Liza Abdul Latiff, “An Efficient Fog-Based Attack Detection Using Ensemble of MOA-WMA for Internet of Medical Things”, *Innovative Systems for Intelligent Health Informatics,* IRICT 2020. Lecture Notes on Data Engineering and Communications Technologies, vol 72. Springer, Cham. https://doi.org/10.1007/978-3-030-70713-2_70

[12] **Swarna Priya R.M., Praveen Kumar Reddy Maddikunta, Parimala M., Srinivas Koppu, Thippa Reddy Gadekallu, Chiranji Lal Chowdhary, Mamoun Alazab, “An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture”, *Computer Communications*, Volume 160, 2020, Pages 139-149, ISSN 0140-3664, https://doi.org/10.1016/j.comcom.2020.05.048.**

[13] **Al-Abassi, Abdulrahman, et al. "An ensemble deep learning-based cyber-attack detection in industrial control system." *IEEE Access 8 (2020): 83965-83973.***

[14] **Aman, Azana Hafizah Mohd, Wan Haslina Hassan, Shilan Sameen, Zainab Senan Attarbashi, Mojtaba Alizadeh, and Liza Abdul Latiff. "IoMT amid COVID-19 pandemic: Application, architecture, technology, and security." *Journal of Network and Computer Applications* 174 (2021): 102886.**

[15] **Hameed, Shilan S., Ali Selamat, Liza Abdul Latiff, Shukor A. Razak, Ondrej Krejcar, Hamido Fujita, Mohammad Nazir Ahmad Sharif, and Sigeru Omatu. "A Hybrid Lightweight System for Early Attack Detection in the IoMT Fog." *Sensors* 21, no. 24 (2021): 8289.**

[16] **Yu, Zengchen, Syed Umar Amin, Musaed Alhussein, and Zhihan Lv. "Research on disease prediction based on improved DeepFM and IoMT." *IEEE Access* 9 (2021): 39043-39054.**

[17] **Ashfaq, Zarlish, Abdur Rafay, Rafia Mumtaz, Syed Mohammad Hassan Zaidi, Hadia Saleem, Syed Ali Raza Zaidi, Sadaf Mumtaz, and Ayesha Haque. "A review of enabling technologies for Internet of Medical Things (IoMT) Ecosystem." *Ain Shams Engineering Journal* 13, no. 4 (2022): 101660.**

[18] **Ihnaini, Baha, M. A. Khan, Tahir Abbas Khan, Sagheer Abbas, Mohammad Sh Daoud, Munir Ahmad, and Muhammad Adnan Khan. "A smart healthcare recommendation system for multidisciplinary diabetes patients with data fusion based on deep ensemble learning." *Computational Intelligence and Neuroscience* 2021 (2021).**

[19] **Reshiwaran, A., L. Jegatheswaran, Isaac Joshua Sakira, and Nor Azlina Abd Rahman. "A Review on IoMT device Vulnerabilities and Countermeasures." In *Journal of Physics: Conference Series*, vol. 1712, no. 1, p. 012020. IOP Publishing, 2020.**

[20] Abbas, Adeel, Muazzam A. Khan, Shahid Latif, Maria Ajaz, Awais Aziz Shah, and Jawad Ahmad. "A New Ensemble-Based Intrusion Detection System for Internet of Things." *Arabian Journal for Science and Engineering* (2021): 1-15.

[21] Smys, S., Abul Basar, and Haoxiang Wang. "Hybrid intrusion detection system for internet of things (IoT)." *Journal of ISMAC* 2, no. 04 (2020): 190-199.

[22] Begli, MohammadReza, Farnaz Derakhshan, and Hadis Karimipour. "A layered intrusion detection system for critical infrastructure using machine learning." In *2019 IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE)*, pp. 120-124. IEEE, 2019.

[23] Hatzivasilis, George, Othonas Soultatos, Sotiris Ioannidis, Christos Verikoukis, Giorgos Demetriou, and Christos Tsatsoulis. "Review of security and privacy for the Internet of Medical Things (IoMT)." In *2019 15th international conference on distributed computing in sensor systems (DCOSS)*, pp. 457-464. IEEE, 2019.

[24] Engineer, Margi, Razma Tusha, Ankit Shah, and Kinjal Adhvaryu. "Insight into the importance of fog computing in Internet of Medical Things (IoMT)." In *2019 International Conference on Recent Advances in Energy-efficient Computing and Communication (ICRAECC)*, pp. 1-7. IEEE, 2019.

[25] Saba, Tanzila. "Intrusion detection in smart city hospitals using ensemble classifiers." In *2020 13th International Conference on Developments in eSystems Engineering (DeSE)*, pp. 418-422. IEEE, 2020.

[26] Saranya, T., S. Sridevi, C. Deisy, Tran Duc Chung, and MKA Ahamed Khan. "Performance analysis of machine learning algorithms in intrusion detection system: A review." *Procedia Computer Science* 171 (2020): 1251-1260.

[27] Aladaileh, Mohammad A., Mohammed Anbar, Iznan H. Hasbullah, Yung-Wey Chong, and Yousef K. Sanjalawe. "Detection techniques of distributed denial of service attacks on software-defined networking controller–a review." *IEEE Access* 8 (2020): 143985-143995.

[29] Ijaz, Muhammad, Gang Li, Ling Lin, Omar Cheikhrouhou, Habib Hamam, and Alam Noor. "Integration and applications of fog computing and cloud computing based on the internet of things for provision of healthcare services at home." *Electronics* 10, no. 9 (2021): 1077.

[30] Raju, K. Butchi, Suresh Dara, Ankit Vidyarthi, V. MNSSVKR Gupta, and Baseem Khan. "Smart Heart Disease Prediction System with IoT and Fog Computing Sectors Enabled by Cascaded Deep Learning Model." ***Computational Intelligence and Neuroscience*** **2022 (2022).**