Original Article

Accurate Deauthentication Attack Detection using Linear Discriminant Analysis in Comparison with Multilayer Perceptron.

B Janardhan[1], P Jagadeesh[2]

[1]Research Scholar, Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105
[2]Project Guide, Corresponding Author, Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105

Abstract

Aim: This study aims to compare the accuracy of Linear Discriminant Analysis to Multilayer Perceptron in detecting deauthentication attack. **Materials and Methods:** 61,000 samples were taken for this analysis with two data sets in the ratio of 80% and 20% samples each. 80% data set is used for training the model and 20% data is used for testing. By the workflow of this research, the data set has been imported, LDA and MLP code has been implemented with the help of jupyter notebooks in the google colab platform. The sample size is calculated from the values obtained from the previous studies with help of the online statistical analysis tool with confidence level of 95% and the margin error of 1%. **Results:** The accuracy of LDA was 82.25% $P<0.05$, whereas the accuracy of the MLP was 86.48% $P<0.05$. **Conclusion:** For the given data set MLP (Multilayer Perceptron) performs significantly better than LDA (Linear Discriminant Analysis) in detecting the deauthentication attack.

Keywords: Linear Discriminant Analysis, Multilayer Perceptron, Machine Learning, Deauthentication attack, Novel Features Selection, Mean

DOI: 10.47750/pnr.2022.13.S04.213

INTRODUCTION

A deauthentication attack is a type of attack which targets the communication between router and the device. These attacks are carried out by means of a deauthentication frame, which is a management frame defined in IEEE 802.11. This frame sent from a router to a device forces the device to disconnect.In technical terms it's called sanctioned technique to inform a rogue station that they have been disconnected from the network. This deauthentication frame is neither encrypted nor authenticated. This is done to overcome the extra resources required for decrypting the frames (Bahaweres, Mokoginta, and Alaydrus 2017). Due to this the attacker can spoof the mac address in the deauthentication frames and disconnect any device from the wifi network. The Deauth-DoS detection is performed here with the aid of a machine learning-based intrusion detection system (IDS) (Agarwal, Biswas, and Nandi 2015). DOS is an attack that consumes all the resources of the network, therefore halting the communication (Zhang, Zheng 2008). It is crucial to detect deauthentication attack in order to overcome all other attacks that rely on the deauthentication attack (Cheema, Bansal, and Sofat 2011b). Therefore, detecting Deauth-DoS attacks and protecting the network from them is a good way to get rid of other attacks that use Deauth-DoS as the base.

A total of 35 conference papers and 3 journal papers are included in IEEE Xplore. (Liu, Sung, and Yao 2014) demonstrates that Deauth-DoS attacks against solely Wi-Fi-based ad-hoc devices, like health care lab machines, can cause serious damage. (Cheema, Bansal, and Sofat 2011b) performed a real deauth-DOS attack on a wifi network and found that the attack reduced the network's connection speed and bandwidth. (Kristiyanto and Ernastuti 2020) analyzed the level of security of WiFi connectivity against deauthentication using Arduino ESP8266 NodeMCU WiFi and Lua programming. (Yu and Tsai 2011) performs intrusion detection using machine learning algorithms and compares the performance of different algorithms and concludes SVM works. (Amudha, Karthik, and Sivakumari 2013) considered attack detection to be a classification problem because the goal is to

determine if the packet is regular or an attack packet. As a result, the approved intrusion detection system model can be implemented using major machine learning algorithms. (Alkasassbeh et al. 2016) used several machine learning algorithms to detect DOS intrusions, MLP algorithm demonstrated the highest accuracy rate of 98.36%. Our institution is passionate about high quality evidence based research and has excelled in various fields (Devarajan et al., 2021; Dhanraj & Rajeshkumar, 2021; Kamath et al., 2020; Nandhini et al., 2020; Parakh et al., 2020; Perumal et al., 2021; Pham et al., 2021; Sathiyamoorthi et al., 2021; Tesfaye Jule et al., 2021; Uganya et al., 2021). Deauthentication attack can be prevented through modification of protocol, but such modifications may contravene existing systems that adhere to specific standards. This proposed method implements an intrusion detection system using machine learning to detect the Deauth-DOS. MLP and LDA are used in the detection of deauthentication attack. Their performances are compared based on the accuracy values obtained in detecting deauthentication attack.

Materials and Methods

The proposed research is conducted in the Signal and Image Processing Lab at Saveetha School of Engineering. Two study groups were identified, which were normal and attack groups. Using G power, 9604 samples are calculated with a 95% confidence level on accuracy and 1% margin of error. The data sets were taken from the University of New Brunswick, Canada. The NSL-KDD dataset used in this project consists of 50,000 samples (Tavallaee et al. 2009). Each sample consists of 42 attributes/features as shown in the table1.

It is necessary to process the NLS-KDD dataset from the University of New Brunswick, Canada before it can be used in the machine learning model. The processed dataset is given for training and testing. During initial data processing, missing data are removed and null values are replaced. As the next step descriptive statistics is done on the features available in the data set followed by features selection. The process of features selection can be done manually or automatically. Here features selection is done using Randomforest classifier, all attributes that are significant in differentiating attacks are plotted in descending order, and the 10 most significant ones are selected for classification of data as normal and attack. The preprocessed dataset with features are given as input to LDA and MLP. From the total sample size 80% of the data is given for training and the remaining 20% is given for testing. Finally the models are trained and tested against the data sets and the accuracy of the models in detection of deauthentication attack is obtained. The learning process of LDA and MLP is given below.

Linear Discriminant Analysis is the most commonly used dimensionality reduction technique in supervised learning. Basically, it is a preprocessing step before applying machine learning and pattern classification (Yang 2019). By projecting the dataset into a moderate-dimensional space, it maximizes overfitting and computational costs while preserving separable features. The initial step in LDA is to compute the separate ability amid various classes,i.e, the distance between the mean of different classes, that is also known as a between-class variance (Mohamad and Bouchachia 2020). The second step in the process is To compute the distance among the mean and sample of each class, that is also known as the within class variance. The final step in the LDA is to create the lower dimensional space that maximizes the between class variance and minimizes the within class variance.

A multilayer perceptron (MLP) is a type of artificial neural network that is feedforward (ANN). MLPs are the most fundamental deep neural network models, consisting of a succession of fully linked layers.A multilayer perceptron (MLP) is a sophisticated artificial neural network. It is made up of many perceptrons. They are made up of an input layer that receives the signal, an output layer that makes a decision or prediction about the input, and an arbitrary number of hidden layers that serve as the MLP's true computational engine. MLPs with a single hidden layer can approximate any continuous function.

SPSS Analysis

The LDA and MLP are tested on the Google Colab (Bisong 2019) cloud platform with 12.67GB RAM and 107.27 GB of disk space. SPSS tools are used for all analyses. SPSS is used to calculate the independent sample t-test, and Python is used to calculate group statistics (Aldrich 2018). In SPSS, the mean, standard deviation, and significant difference between the two groups are measured. The simulated mean values and standard deviation are shown in table 2. In this study, independent variables are the input features dependent variable is accuracy.

Results

LDA and MLP are compared according to their accuracy in detecting deauthentication attack. Table 1 consists of a number of features and classes present in the data set. Table 2 is the statistical analysis of the data set and consists of minimum, maximum, mean and standard deviation of all the features available. Using the same data set, each model was trailed ten times and the results are presented in Table 3. T-test analysis is done using statistical packages of social sciences (IBM-SPSS v21) and the results are tabulated in Table 4. From the mean accuracy graph in Fig 2, it can be observed that MLP detects deauthentication attack accurately compared to LDA. From

the results obtained can be observed that the accuracy of MLP is greater than the LDA. The accuracy values of LDA have not much deviation from its mean value compared to MLP.
Fig. 1 shows that src_bytes is the most significant feature of the data sets that are used to detect deauthentication attack, followed by flags, same_srv_rate, diff_srv_rate, and dst_host_same_srv_rate. From Fig. 2 it can be observed that the accuracy of LDA for the first iteration is 82.2% and it continues to remain at 82% till the fifth iteration. During the sixth iteration the accuracy value increases to 82.7%. After the sixth iteration the accuracy value decreases and settles at 82.2% till tenth iteration. From Fig. 3 it can be observed that the accuracy of MLP for the first iteration is 84.66% and it increases gradually as the number of iterations increases and reaches a maximum of 88.55% at the tenth iteration. Fig. 4 shows that MLPs are more accurate than LDAs. The standard deviation in accuracy for MLP is lesser compared to LDA. Therefore, MLP detects deauthentication attacks better than LDA
.

Discussion

From the results obtained it can be concluded that MLP have a higher accuracy than LDA in this analysis. MLP accuracy is 86.48 %, whereas LDA accuracy is 82.25 percent. There are 42 features in the data set collection. All the available features are plotted based on their significance level in detection of attack in the process of features selection. In Fig. 1, the relevance of each attribute in determining whether a datapoint is of normal or attack class is displayed. The model is trained using the 10 most important features available from all the features provided during features selection process. It is determined that MLP detects the attack more accurately than LDA after training and testing the model against the data set.
(Halimaa A. and Sundarakantham 2019) uses SVM to develop intrusion detection, which is a superclass of deauthentication attack detection, and achieves a 97.27 percent accuracy rate. The prior model had a higher accuracy because the data set had less data samples of 17000 for . Because the proposed work uses additional data samples, the accuracy is reduced. (Almseidin et al. 2017) uses various machine learning algorithms to perform intrusion detection, which is a subclass of deauthentication attack detection, with the highest accuracy of 93.77 percent obtained using Random Fores. The accuracy values of the proposed study and the prior work are similar. (B. Zhang et al. 2018) uses a Gaussian Naive Bayes algorithm to detect attacks and achieves an accuracy of 83.28 percent. The proposed work has an accuracy value of 86.48, which is higher than the existing work. (Agarwal et al. 2016) studied the detection of flooded DoS attacks using SVM and found an accuracy of 98.7%, which is way greater than the accuracy found in the proposed work. The earlier effort had a greater accuracy since it employed a smaller dataset and used a different feature for the analysis due to the smaller dataset confidence level on the accuracy is less.
Because the dataset employed in the proposed work is of finite size, the confidence in the accuracy attained is also restricted. The trust in the accuracy can be raised by increasing the datasets. A combination of several Machine Learning models will be used in the future to increase model accuracy using larger data sets and hybrid multilevel models. The future model may be able to categorize attacks more successfully by building well-organized classifiers.

Conclusion

Accurately detecting a deauthentication intention is to reduce the chance of communication bandwidth being restricted. MLP produced an accuracy of 86.48% compared to the LDA whose accuracy is of 82.25%.When LDA and MLP were compared, the current study discovered that MLP performed much better in identifying deauthentication attack.

REFERENCES

1. Agarwal, Mayank, Santosh Biswas, and Sukumar Nandi. 2015. “Detection of De-Authentication DoS Attacks in Wi-Fi Networks: A Machine Learning Approach.” 2015 IEEE International Conference on Systems, Man, and Cybernetics. https://doi.org/10.1109/smc.2015.55.
2. Agarwal, Mayank, Dileep Pasumarthi, Santosh Biswas, and Sukumar Nandi. 2016. “Machine Learning Approach for Detection of Flooding DoS Attacks in 802.11 Networks and Attacker Localization.” International Journal of Machine Learning and Cybernetics. https://doi.org/10.1007/s13042-014-0309-2.
3. Almseidin, Mohammad, Maen Alzubi, Szilveszter Kovacs, and Mouhammd Alkasassbeh. 2017. “Evaluation of Machine Learning Algorithms for Intrusion Detection System.” 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY). https://doi.org/10.1109/sisy.2017.8080566.
4. Bahaweres, R. B., S. Mokoginta, and M. Alaydrus. 2017. “Comparison of Methods for Localizing the Source Position of Deauthentication Attacks on WAP 802.11n Using Chanalyzer and Wi-Spy 2.4x.” Journal of Physics: Conference Series. https://doi.org/10.1088/1742-6596/801/1/012056.
5. Cheema, Rupinder, Divya Bansal, and Sanjeev Sofat. 2011b. “Deauthentication/Disassociation Attack: Implementation and Security in Wireless Mesh Networks.” International Journal of Computer Applications. https://doi.org/10.5120/2901-3801.

6. Liu, Hsiang-Chuan, Wen-Pei Sung, and Wenli Yao. 2014. Computer, Intelligent Computing and Education Technology. Taylor & Francis.
7. Kristiyanto, Yogi, and E. Ernastuti. 2020. "Analysis of Deauthentication Attack on IEEE 802.11 Connectivity Based on IoT Technology Using External Penetration Test." CommIT (Communication and Information Technology) Journal. https://doi.org/10.21512/commit.v14i1.6337.
8. Mohamad, Saad, and Abdelhamid Bouchachia. 2020. "Online Gaussian LDA for Unsupervised Pattern Mining from Utility Usage Data." 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA). https://doi.org/10.1109/icmla51294.2020.00016.
9. "The Multilayer Perceptron." n.d. Neural Computing: An Introduction. https://doi.org/10.1887/0852742622/b335c4.
10. Bisong, Ekaba. 2019. Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners. 1st ed. Berlin, Germany: APress.
11. Halimaa A., Anish, and K. Sundarakantham. 2019. "Machine Learning Based Intrusion Detection System." In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). IEEE. https://doi.org/10.1109/icoei.2019.8862784.
12. Yang, Xin-She. 2019. "Logistic Regression, PCA, LDA, and ICA." Introduction to Algorithms for Data Mining and Machine Learning. https://doi.org/10.1016/b978-0-12-817216-2.00012-0.
13. Yu, Zhenwei, and Jeffrey J. P. Tsai. 2011. Intrusion Detection: A Machine Learning Approach. World Scientific.
14. Zhang, Yan, Jun Zheng, and Honglin Hu. 2008. Security in Wireless Mesh Networks. CRC Press.
15. Aldrich, James O. 2018. Using IBM SPSS Statistics: An Interactive Hands-On Approach. SAGE Publications.
16. Tavallaee, Mahbod, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani. 2009. "A Detailed Analysis of the KDD CUP 99 Data Set." In 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications. IEEE. https://doi.org/10.1109/cisda.2009.5356528.

Tables and Figures

Table 1. Samples, features, and classes from Datasets - The data sets consist of two classes attack class and normal class with 42 features.

Data Sets	Features	Classes
Test Dataset	42	2
Train Dataset	42	2

Table 2. Statistical features of the NSL-KDD dataset - The mean, standard deviation, minimum and maximum values are tabulated

	count	mean	std	min	25%	50%	75%	max
duration	50000	86.9285	927.6287847	0	0	0	0	29053
src_bytes	50000	6515.08484	142678.4508	0	0	0	244	18828976
dst_bytes	50000	2309.65104	43058.86775	0	0	0	478	5131424
land	50000	0.0004	0.0199962	0	0	0	0	1
urgent	50000	6.00E-05	0.013416408	0	0	0	0	3
hot	50000	0.1452	1.633122868	0	0	0	0	77
num_failed_logins	50000	0.00074	0.038203162	0	0	0	0	4
logged_in	50000	0.37382	0.483821547	0	0	0	1	1
num_compromised	50000	0.19534	9.873936002	0	0	0	0	884
root_shell	50000	0.00102	0.031921466	0	0	0	0	1
su_attempted	50000	0.001	0.042415044	0	0	0	0	2
num_root	50000	0.20306	10.92727847	0	0	0	0	975
num_file_creations	50000	0.01268	0.501362389	0	0	0	0	40
num_shells	50000	0.00026	0.01612258	0	0	0	0	1

num_access_files	50000	0.00372	0.091358243	0	0	0	0	8
is_guest_login	50000	0.00612	0.077991456	0	0	0	0	1
count	50000	94.4	101.5810589	0	4	39	185	511
srv_count	50000	19.27404	43.67500993	0	3	8	17	502
serror_rate	50000	0.4076464	0.488829594	0	0	0	1	1
srv_serror_rate	50000	0.4065064	0.488561702	0	0	0	1	1
rerror_rate	50000	0.103159	0.301654323	0	0	0	0	1
srv_rerror_rate	50000	0.104535	0.302062167	0	0	0	0	1
same_srv_rate	50000	0.5528708	0.456239461	0	0.07	0.99	1	1
diff_srv_rate	50000	0.0494412	0.114967439	0	0	0.03	0.06	1
srv_diff_host_rate	50000	0.0661986	0.205012136	0	0	0	0	1
dst_host_count	50000	196.03142	91.55295447	0	128	255	255	255
dst_host_srv_count	50000	104.53792	110.3305853	0	10	25	255	255
dst_host_same_srv_rate	50000	0.454722	0.446524815	0	0.04	0.17	1	1
dst_host_diff_srv_rate	50000	0.0521972	0.094205699	0	0	0.05	0.07	1
dst_host_same_src_port_rate	50000	0.0628646	0.192257523	0	0	0	0.01	1
dst_host_srv_diff_host_rate	50000	0.0133572	0.051442688	0	0	0	0.01	1
dst_host_serror_rate	50000	0.4070912	0.487968758	0	0	0	1	1
dst_host_srv_serror_rate	50000	0.4022024	0.488018533	0	0	0	1	1
dst_host_rerror_rate	50000	0.103876	0.298456104	0	0	0	0	1
dst_host_srv_rerror_rate	50000	0.1028422	0.297259306	0	0	0	0	1

Table 3. Accuracy of LDA and MLP in the detection of deauthentication attack over 10 iterations. Accuracy of LDA is constant and does not change as the iterations increase. Accuracy of MLP increases arbitrarily as iterations increase.

Trial	LDA	MLP
1	0.82203	0.84667
2	0.82214	0.85500
3	0.82209	0.85698

4	0.82133	0.85034
5	0.82197	0.86199
6	0.82797	0.84836
7	0.82197	0.88051
8	0.82214	0.88453
9	0.82133	0.87543
10	0.82203	0.88855

Table 4. Independent sample test - Independent sample T- test is performed for the dataset with a 95% confidence interval and a significance level P < 0.05 (MLP appears to perform significantly better than LDA)

		Levene's Test for Equality of Variances		T-test for Equality of Means						
		F	Sig.	t	df	sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Accuracy	Equal variances assumed	2.878	.107	6.932	18	.000	.02540	.00366	.01770	.03310
	Equal variances not assumed			6.932	9.927	.000	.02540	.00366	.01770	.03357

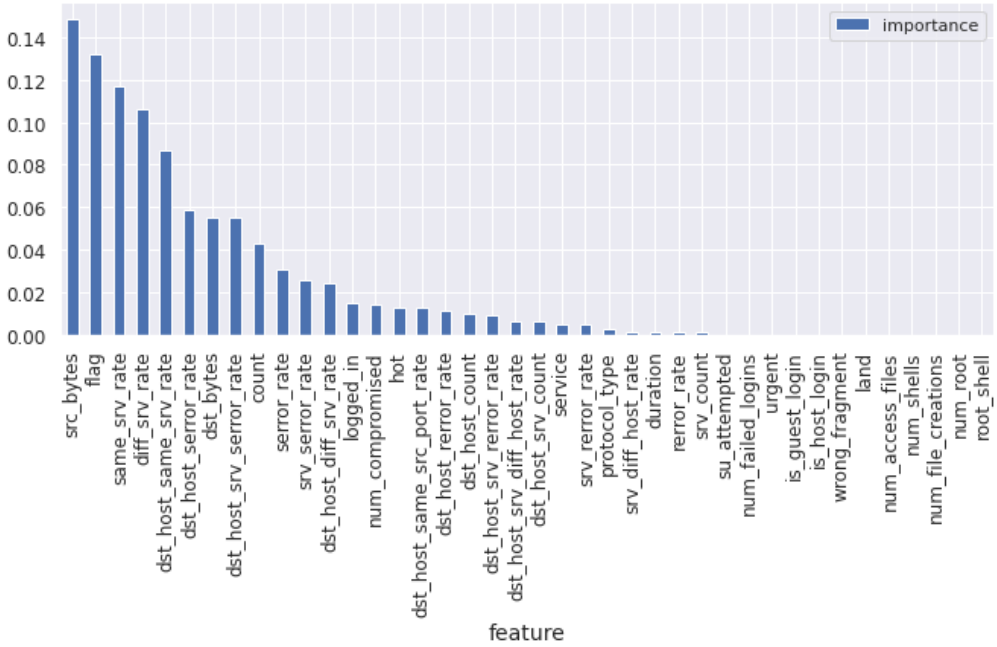

Fig. 1. Features selection - The features are plotted based on their significance in detecting the attack.

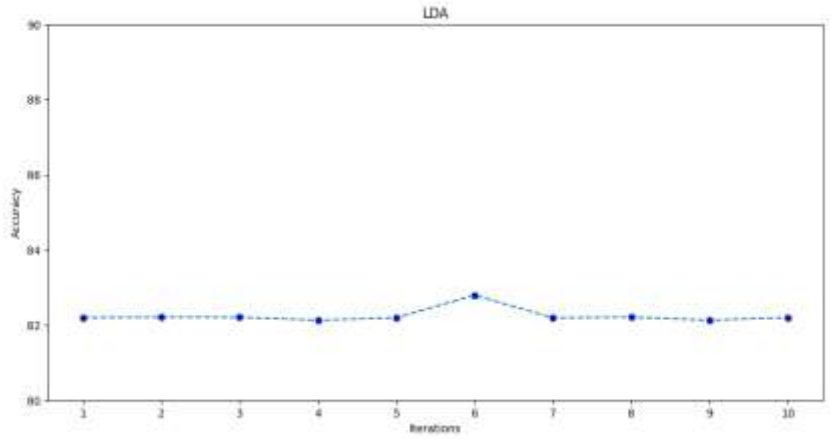

Fig. 2. Accuracy of LDA over 10 iterations - Accuracy of LDA remains at 82% throughout ten iterations therefore standard deviation is less.

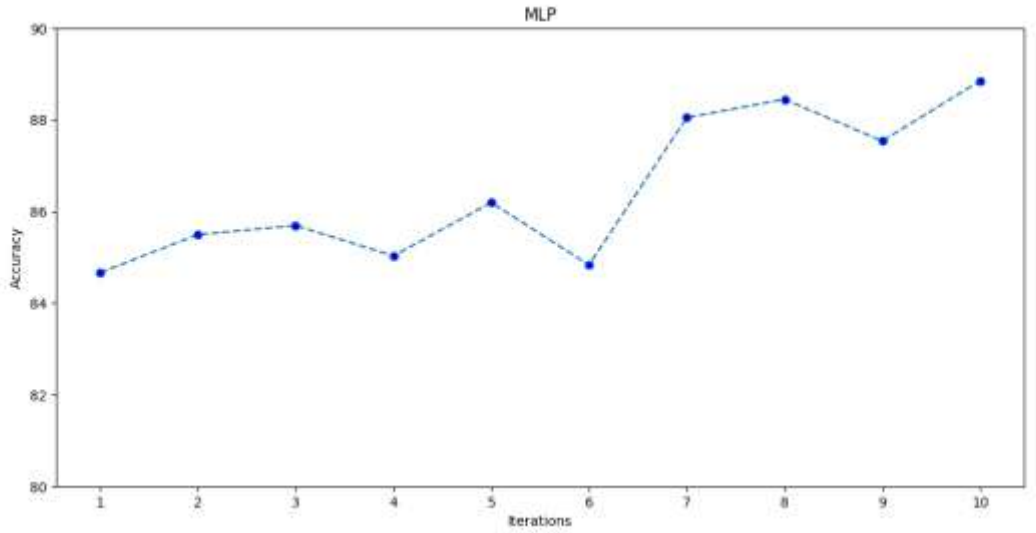

Fig. 3. Accuracy of MLP over 10 iterations - Accuracy of MLP is initially 84% and increases arbitrarily over 10 iterations and reaches a maximum value of 88% at tenth iteration.

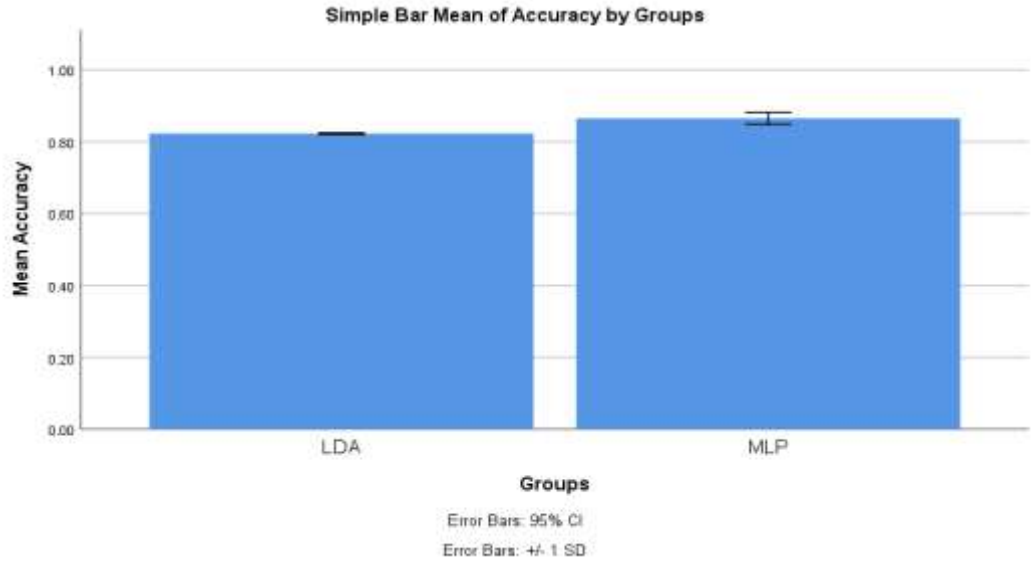

Fig. 4. Mean accuracy graph of LDA and MLP - Accuracy of MLP is high compared to LDA in detection of deauthentication attack. Standard Deviation (SD) of Decision Tree classifier is high compared to Standard Deviation (SD) of LDS. X Axis: LDS vs MLP,Y Axis: Mean accuracy of detection ± 1 SD.