Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, August : 2024

A Comparative Analysis of Various Classifiers for the Attack Detection of DOH Traffic

Vaka Padmavathi[1], K. Swathi[2], B. Basaveswara Rao[3]
[1,] Research Scholar, Dept. of CSE, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur, padmavaka527@gmail.com
[2]Associate Professor, Dept. of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram
[3] Professor, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur

Abstract: This paper explores the capabilities of different classifiers for detecting the malicious activities of DNS over HTTPS (DoH) traffic. Several research studies are conducted and presented the classification power of various classifiers for NIDS or HIDS through experimental evaluation using benchmark datasets. How ever, very few researchers have studied to explore the effect of various classifiers for attack detection of DOH traffic. It is necessary to investigate which particular classical classifier will fit to achieve high detection rate with a minimal computational overhead because certain security flaws of DNS and it is mostly targeted by the attackers. This paper address this problem by focusing on the main classification families related to distance, probability, tree, parameter and statistical models. To adopt the classical classification models for each family, kNN, Naïve Bayes, J48, SVM, Random Forest, SLR, and Logic Boost are considered respectively. The main objective of this paper is to provide a prior knowledge, about the capabilities of these classifiers using a benchmark dataset CIRA-CIC-DoHBrw-2020 by conducting experiments. For this purpose, a two-layered classification model is adopted. Finally a comparative analysis is carried out based on various performance measures and building time of various classifiers.

Keywords—classification, Machine Learning, DNS over HTTPS, DoH, malicious DoH.

I. Introduction

Domain Name System (DNS) helps to map the given domain name to corresponding physical IP addresses. DNS suffers from security as well as privacy issues since it uses User Datagram Protocol (UDP) which happens to be unreliable delivery protocol. Instead of traditional DNS, the DoH and DoT techniques have been introduced to strengthen the user's security and privacy on the internet. Due to the drastic increase in the usage of DNS over HTTP, the malicious activities also faced by the users. To improve the security and privacy of these services the concept of DNS over HTTPS (DoH) is introduced by providing the required immunity from data manipulation of DNS with the Man-in-the-Middle attacks. Because of the characteristics of the DNS operation process, attackers often exploit DNS to attack the system. DNS cache poisoning, Fast flux DNS and phishing are some of the DNS attacks.

According to the research on latest network attacks (**Vekshin, D., et al. (2020)**, the techniques of attacks on users by spreading malicious domains are predicted to have sophisticated transformations and serious consequences. Therefore, the problem of identifying and proposing a proper detection mechanism which helps to give an early warning about the functioning of the malicious play crucial role. These malicious domains also classified according to their nature. Several research studies have suggested different detection mechanisms for these malicious activates.. In particular, the method of detecting malicious domains based on behaviour analysis techniques using machine learning and deep learning techniques is highly effective because it has the

ability to detect new malicious domains(Vani, R. (2017), S. Al-Emadi et al. (2020)).There are various types of machine learning algorithms among them classification is widely used and accepted for identifying malicious activities. Recently several researchers have adopted ML algorithms to build an efficient Intrusion Detection System (IDS) because, these algorithms provide better detection rates with a minimal computational overhead (Rao, B. B., et al. (2017,Jafar, M.T., et al. (2021), Biswas, S. K. (2018).).

Currently ML algorithms playing a vital role for the detecting malicious activities of DNS over HTTPS traffic. has developed a systematic two-layer approach for detecting DoH traffic and distinguishing Benign traffic from Malicious in DoH traffic with the help of a few machine learning models. Time-series classifiers are also used to detect and characterize the DoH traffic Jafar, M.T., et al. (2021).The other works also established systematic approaches for detecting DoH traffic(Banadaki, Yaser M. (2020) Vekshin, D., et al. (2020)). Most of these works are adopted benchmark DoH traffic data CIRA-CIC-DoHBrw-2020 dataset for conducting the experiments.

From the last decade some of the research papers have been published to detect the DoH tunnels with numerous types of classification algorithms. In this scenario there are two challenges faced by the researchers, which is the most suitable classifier among them and how much does it provide better performance over others for detecting the tunnels of DoH. This work is intended to give an idea to the defenders about addressing the aforementioned challenges regarding the implementation of defending mechanisms through classical classification for detecting the DoH tunnels. For this comparative study the capabilities of seven classical classification algorithms related to five different families are being carried. The main contributions are as follows.

i) To classify the statistical features of the HTTPS traffic flows, the publicly available benchmark dataset CIRA-CIC-DoHBrw-2020 is adopted because it contains contemporary cyber-attacks related DNS tunnelling.
ii) The experiments are conducted with a two-layered model for seven classification algorithms and computed performance measures along with building time of the each model.

iii)Based on the experimental results discuss the capabilities of the detection of DNS tunnels.

The remaining of this article is organized as follows. In section II the review of literature is presented in which, various authors work related to the current work is discussed. In section III preliminaries related to the current work is provided. Section IV discusses the methodology related to the experimentation. Results are discussed in section V and finally conclusions are provided in section VI.

II Review of Literature

This section briefly discusses about the related work done by several researchers in the area DoH detection using various Machine Learning approaches particularly classification techniques. Recently Vekshin, D., et al. (2020) have provided an insight for DoH data Detection mechanism based on five popular classification techniques. Experimental results have been proved to have 99.9% of accuracy. Banadaki, Yaser M. et al. (2020) have proposed a two-layered model for DoH traffic detection. All the experiments were drawn on CIRA-CIC-DoHBrw-2020 dataset. They have adopted

several classification techniques for the attack detection among them XG Boost and LGBM classifiers have exhibited good detection rates. Montazeri Shatoori, et al. (2020) have discussed one of the security issues, misuse of DNS protocol to create hidden channels by tunnelling data using DNS packets. In their work they have proposed a two-layered model intended to classify various activities on tunneling simultaneously specify whether the traffic is malicious or not.. Jafar, M. T., et al. (2020)have introduced a systematic approach for identifying malicious and encrypted DNS queries by examining the network traffic and deriving statistical characteristics. Afterwards they have implemented several ML methods: (RF: Random Forest, DT: Decision Tree Classifier, GNB Gaussian Naive Bayes, KNN: k-nearest neighbor, Logistic regression, SVC: Support Vector Classifier, QDA: Quadratic Discriminant Analysis, SGD)". These models were employed to evaluate their ability to detect malicious DNS traffic using the CIRA-CIC-DoHBrw2020 data set. The Experiments revealed a good accuracy score where DT and RF models have achieved the highest accuracy, 99.99 % relative to other detection methods. Do Xuan, C., et al. (2020) proposed a machine learning model for the detection of malicious domain Their proposed methods differed in looking for and extracting features that accurately specify the behavior of malicious domains and normal domains. To achieve this, they have adopted Random Forest (RF) classifier. By adjusting the parameters of the RF, they have achieved the optimal parameters, this model has been proved to achieve good detection rates. **Tally, M. T., &Amintoosi, H. (2021)** have proposed a hybrid model that improves the performance of IDS. The authors adopted Genetic Algorithm (GA) for feature selection purpose and for the intrusion detection they used SVM. The proposed hybrid model is implemented on NSL-KDD dataset. The model has achieved 92% f-measure.

III. Background

This section explains the CIRA-CIC-DoHBrw-2020 dataset structure, various Machine Learning algorithms and different performance metrics that are adopted for this work. The following table presents the list of all 34 features and their descriptions of the adopted dataset.

Table 1: List of 34 statistical features and their description of the dataset

Feature #	Feature Name	Description
1	SourceIP	IP address of the source machine
2	DestinationIP	IP address of the destination machine
3	SourcePort	Port number of source machine
4	DestinationPort	Port number of destination machine
5	TimeStamp	Time stamp of the traffic packet
6	Duration	Duration of the communication
7	FlowBytesSent	Total number of bytes transmitted from source to destination
8	FlowSentRate	% of bytes transmitted from source to destination
9	FlowBytesReceived	Total number of bytes transmitted from destination to the source
10	FlowReceivedRate	% of bytes transmitted from destination to source
11	PacketLengthVariance	Variance value for the length of the packet
12	PacketLengthStandardDeviation	Standard Deviation for the length of

		the packet
13	PacketLengthMean	Mean of packet length
14	PacketLengthMedian	Median of packet length
15	PacketLengthMode	Mode value of packet Length
16	PacketLengthSkewFromMedian	Skewness from median for the packet length
17	PacketLengthSkewFromMode	Skewness from mode for the packet length
18	PacketLengthCoefficientofVariation	Coefficient of the variation value for the packet length
19	PacketTimeVariance	Variance of the time of packet received
20	PacketTimeStandardDeviation	Standard deviation value for the packet time
21	PacketTimeMean	Mean value for the packet time
22	PacketTimeMedian	Median value for the Packet time
23	PacketTimeMode	Mode value for the packet time
24	PacketTimeSkewFromMedian	Skewness of median for the packet time
25	PacketTimeSkewFromMode	Skewness of mode for the packet time
26	PacketTimeCoefficientofVariation	Coefficient of variation values for packet time of the server
27	ResponseTimeTimeVariance	Time variance of response time
28	ResponseTimeTimeStandardDeviation	Standard deviation values for response time of the server
29	ResponseTimeTimeMean	Mean value of response time of the server
30	ResponseTimeTimeMedian	Median time of response time of the server
31	ResponseTimeTimeMode	Mode of the response time of the server
32	ResponseTimeTimeSkewFromMedian	Skewness from median for response time of the server
33	ResponseTimeTimeSkewFromMode	Skewness from mode of the response time
34	ResponseTimeTimeCoefficientofVariation	Coefficient of variation for response time

3.1 Classifiers

In the Machine learning techniques, the classification plays a dominant role in networking intrusion detection. They are categorised according to the algorithm applied in model building. Distance, parameter, probabilistic, conditional and neural network based algorithms are prominent among the state-of-art algorithms. In this work six classification algorithms were selected each from one category to detect the malicious activities over DoH traffic. Further to evaluate the performance for assessing the capabilities of classifiers. A brief discussion of these algorithms are given below.

3.2 Distance based Classifiers:

These classifiers are used to identify each data sample as an object and finds distance between unknown sample to the known samples to predict the unknown sample class label. There are good number of distance metrics like Euclidean, Mahanobalies, Manhattan, city block, and Minkovski etc. In this work, kNN classifier (Basaveswara Rao B & Swathi, K. (2017). is used for distance based classifier for the prediction which is used to find k number of nearest neighbour based on Euclidian distance metric.

3.3 Parameter based classifiers:

Parameter based models uses hyperplanes to separate positive samples and negative samples. Linear regression, SVM are some of parametric models which can be used as binary classification. SVM locates a hyper plane that will maximize the distance from the members of each class to the optimal hyper plane. When the data space is not linearly separable, kernel tricks are adopted to convert the non-linear space to liner.

3.4 Probability based Classifiers:

It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. C4.5 (J48) is an algorithm used to generate a decision tree developed by Ross Quinlan mentioned earlier. C4.5 is an extension of Quinlan's earlier ID3 algorithm. Decision tree algorithm falls under the category of supervised learning. They can be used to solve both regression and classification problems. Decision tree uses the tree representation to solve the problem in which each leaf node corresponds to a class label and attributes are represented on the internal node of the tree. It can be representing any Boolean function on discrete attributes using the decision tree.

3.5 Random forest: Random forest is a supervised model used for both classification and regression problems. bagging technique that uses

3.6 ADA boost : Logit Boost is a boosting classification algorithm. Logit Boost and Ada Boost are close to each other in the sense that both perform an additive logistic regression. The difference is that AdaBoost minimizes the exponential loss, whereas LogitBoost minimizes the logistic loss.

3.7 Performance Metrics:

During this study to compare these classifiers the traditional performance measures Accuracy, Precision, Recall, F-Measure and area under curve (AUC) are considered. Computational time is also taken as one of the criterion or comparison. A highest accurate classifier is a trustable one. Along with accuracy, false alarm rate is also one of the metrics to findout how many normal users are detected as attackers. All these metrics are evaluated through conducting the experiments and constructing the following confusion matrix.

Table 2: Confusion Matrix

Number of samples		Predicted	
		Attack	Normal
Actual	Attack	**TP**	**FN**
	Normal	**FP**	**TN**

True Positive (TP): Test samples predicted as Attack and their class labels are actually Attack.

True Negative (TN): Test samples predicted as Normal and their class labels are actually Normal

False Positive (FP): Test samples predicted as Attack and their class labels are actually Normal

False Negative (FN): Test samples predicted as Normal and their class labels are actually Attack

Based on the entries of the confusion matrix, the following measures are calculated.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} (3)$$

$$Recall = \frac{TP}{TP+FN} (4)$$

$$\text{F1 Score} = = \frac{\text{2 x Recall X Precision}}{Recall+precision} (5)$$

$$\textbf{Area Under the Curve (AUC)} = \frac{\text{TP}}{2\times(\text{TP+FN})} + \frac{\text{TN}}{2\times(\text{TN+FP})} (6)$$

IV. Methodology

In this section, a two layered architecture is adopted on the lines of **Montazeri Shatoori, et al. (2020**) to classify the statistical representation of flows. To detect whether the traffic data is DoH or Non-DoHin layer-one and in layer-two to detect whether it is an attack or a benign one.In pre-processing module, the raw dataset is converted into a suitable format for processing various classification techniques. All these features arein numeric type so there is no need of data transformation. To eliminate the feature influence on the classification results min-max normalization is carried out.

In the Layer-I the pre-processed dataset provided for the training to classify whether the traffic is DoH or non-DoH. For this six different classification techniques are used and results were compared. Later in Layer II DoH traffic is further classified to predict attacks and benign samples. Figure 1 depicts the evolutionary process of two-layered classification approach.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, August : 2024

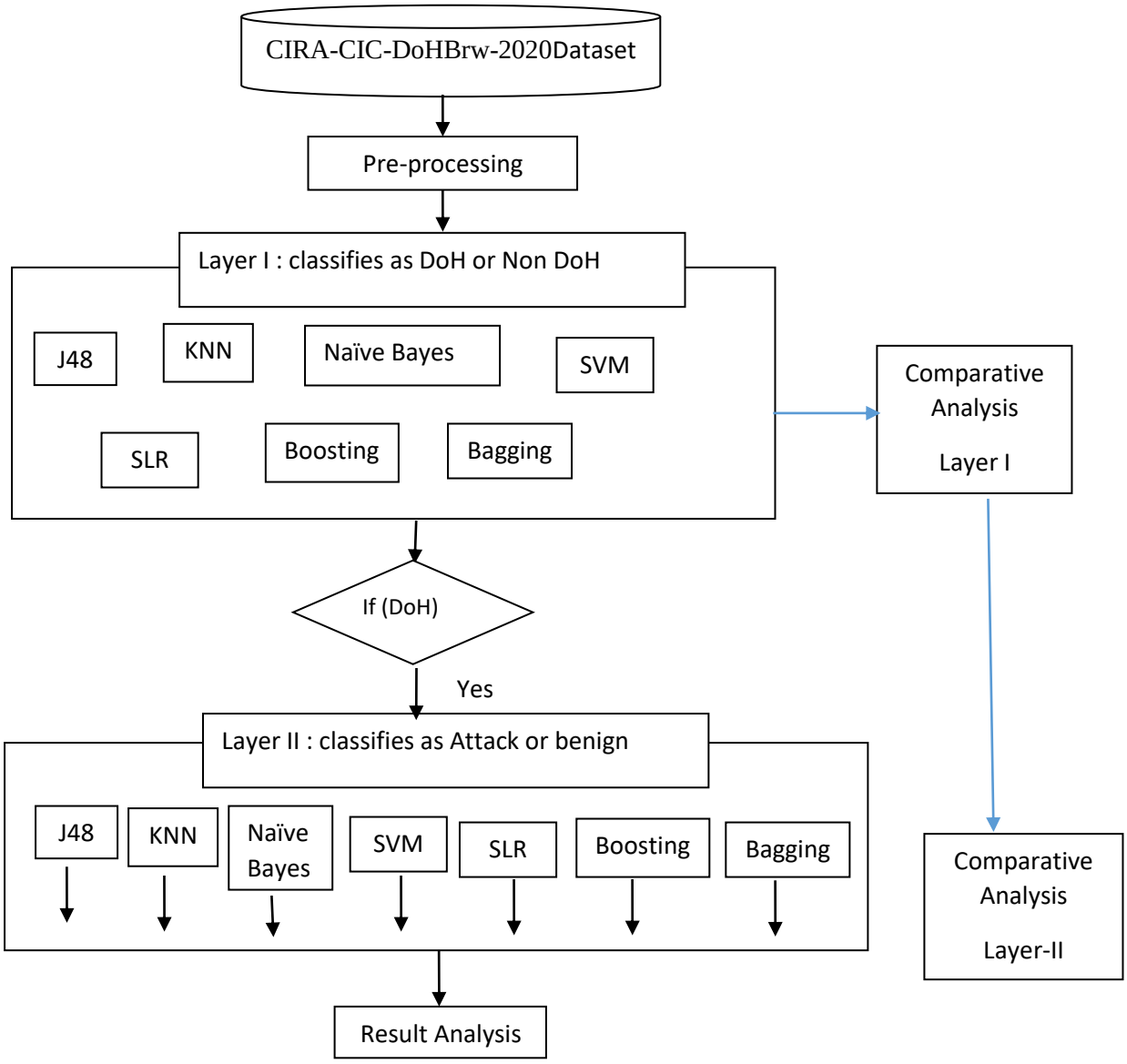


Figure 1: evolutionary process of two-layered classification approach.

V. Experimental Results and Discussions.

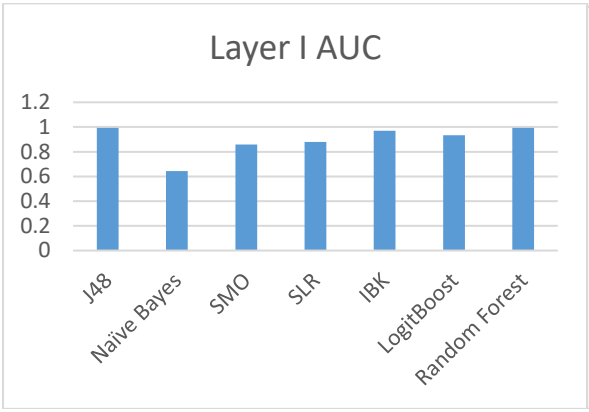
This section discusses about the effect of seven classification models of various types of learning families through experiments. These experiments were conducted in Weka 3.9 under windows 10 environment with a 10 fold cross validation. All tests are experimented based on the 10% of the CIRA-CIC-DoHBrw-2020 dataset i.e., 1,16,714 instances with both DoH and NON-DoH patterns. For this evolutionary process accuracy, recall, precision, f-measure and AUC are considered along with the computational time. The classifiers J48, Naïve Bayes, SVM, Simple Logistic Regression,

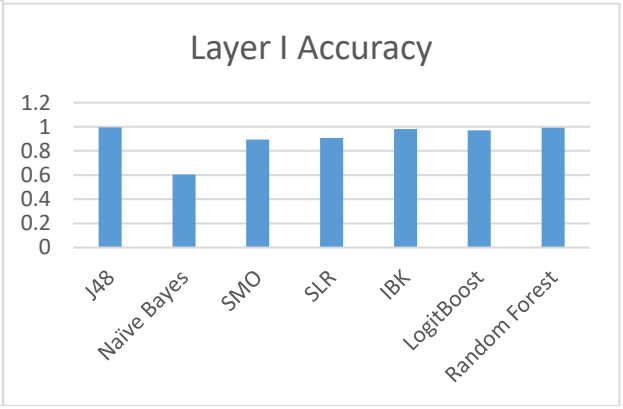
Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, August : 2024

IBK, LogiBoost and Random forest are selected for this two layered classification model. The results are presented in the following tables as well as graphs.

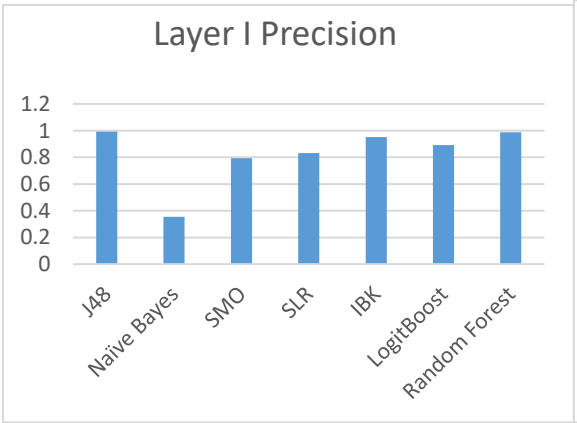
Table 3. Performance of seven classifiers forLayer I

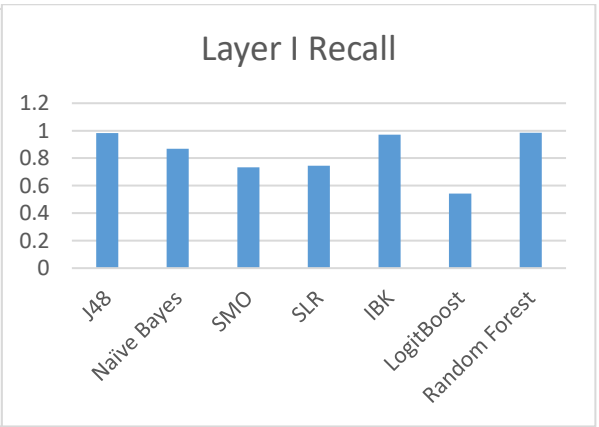
LAYER I						
Type of Learning Model	Algorithm	**Accuracy**	**Recall**	**Precision**	**F Measure**	**AUC**
Decision Tree	**J48**	**0.994362**	**0.983757**	**0.992146**	**0.9879334**	**0.994362**
Probabilistic	**Naïve Bayes**	**0.489016**	**0.868607**	**0.354766**	**0.5037748**	**0.489016**
Parametric	**SVM**	**0.890013**	**0.732876**	**0.793049**	**0.7617763**	**0.890013**
Distance Based	**IBK**	**0.97996**	**0.970146**	**0.950858**	**0.9604053**	**0.97996**
	SLR	**0.903644**	**0.745077**	**0.831994**	**0.786141**	**0.878867**
Boosting	**LogiBoost**	**0.903516**	**0.64936**	**0.922647**	**0.7622489**	**0.903516**
Bagging	**Random Forest**	**0.992854**	**0.984276**	**0.986838**	**0.9855551**	**0.992854**





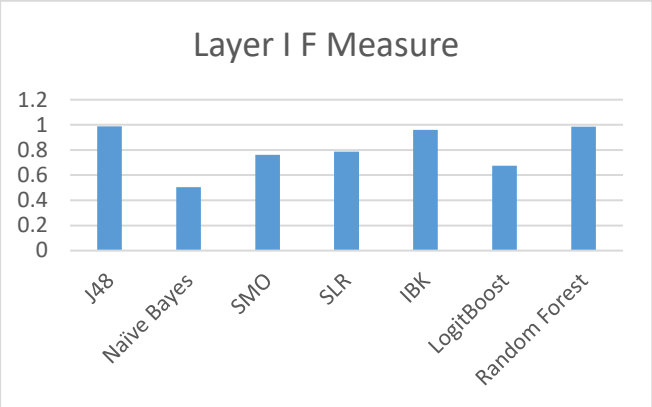
(a) **(b)**





(c) (d)

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, August : 2024



(e)

Figure 2: Effect of various Metricsfor different classifiers of LayerI

After observing the different performance metrics presented in Table 3 and Figure II (a,b,c and d) , the following observations are given below for Layer I.

The Accuracy of J48 yields highest value when compared to other classifiers, the next near highest is Random Forest followed by IBK with a minimum difference. Whereas Naïve Bayes Accuracy is very low, and it is less than 50%.This type of similar trend followed by the reaming all metrics. But the recall exhibits differently with higher value when compared to SVM and logicBoost which gets minimum value i.e., 65%. The Naïve Bayes performed with a minimum value for Precision, F Measure and AUC when compared to others.

For further investigation and to quantify the relative position of these metrics w.r.t to the performance, through a ranking procedure with a uniform metric called as a Combined Metrics Rank(CMR). To evaluate the CMR, the following numerical ranks are assigned from 7 to 1 to the classifiers based on their individual performance(Table 1) and are presented in Table 4.

Table 4. Ranks of the various classifiers based on their Performance in Table 1

Metric/ Classifier	Accuracy	Recall	Precision	F Measure	AUC	CMR
J48	7	6	7	7	7	6.8
Naïve Bayes	1	4	1	1	1	1.6
SMO	2	2	2	2	2	2
SLR	4	3	3	3	3	3.2
IBK	5	5	5	5	5	5
LogitBoost	3	1	4	3	3	2.8
Random Forest	6	7	6	6	7	6.4

From the above table it is observed that the J48 is in top rank with an CMR value of 6.8 and outperformed than other classifiers. And the second nearest outperformed classifier is Random

Forest classifiers with a marginal difference of CMR is of 0.4.The third highest performer is IBK with a mean rank of 5.The Naive Bayes classier is exhibited with lowest performance among all the classifiers and the second lowest classifier is SMO with an CMR difference of 0.4 followed by LogitBoost and SLR.

From the above table it is observed that the top three ranked classifiers all performance metrics follows same type of order of Layer I with marginal difference. The accuracy of the J48 is higher than random forest and IBK with a very in significant differences of 0.002 and 0.006 respectively. The ROC curves of J48 and random forest occupies equally more area than IBK, but it is negligible. Precision, Recall and F Measure metrics are also followed in this criterion.
This is a clear indication about the false positive rate of these classifiers is minimum. It is a necessary condition that the classifier is a better classifier w.r.t network security perspective, if the classifier false negative rate is less than false positive rate. This condition full fill, if the recall values are higher than Precision values or equal. Except Naïve Bayes the remaining all classifiers recall values are higher than the precision values with a marginal difference.

Table 5. Performances of seven classifiers for Layer II

Algorithm	Accuracy	Recall	Precision	F Measure	AUC
J48	0.999623	0.99992	0.999679	0.999799	0.99921
Naïve Bayes	0.423747	0.377877	0.98919	0.546852	0.553223
SMO	0.975863	1	0.975859	0.987782	0.98793
SLR	0.981285	0.99934	0.981892	0.990539	0.846501
IBK	0.993043	0.998004	0.994602	0.9963	0.981204
LogitBoost	0.955309	0.999175	0.95167	0.974844	0.971867
Random Forest	0.99789	0.999116	0.998634	0.998875	0.99267

Algorithm	Accuracy	Recall	Precision	F Measure	AUC	CMR
J48	7	6	7	7	7	6.8
Naïve Bayes	1	1	4	1	1	1.6
SMO	3	7	2	3	5	4
SLR	4	5	3	4	2	3.6
IBK	5	2	5	5	4	4.2
LogitBoost	2	4	1	2	3	2.4
Random Forest	6	3	6	6	6	5

The following observations are noted for Layer II when consider the different performance metric results of Table 4 and Figure 5.

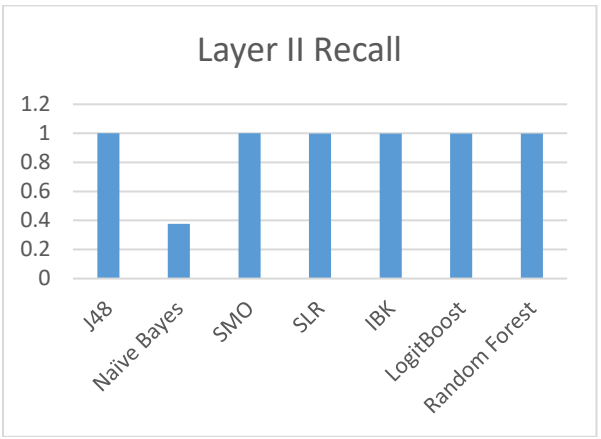
The Accuracy of J48 yields highest value when compared to other models, the next highest is Random Forest followed by IBK with a minimum difference. Whereas Naïve Bayes Accuracy is very low less than 50%.The above similar trend followed by the reaming all metrics. But the recall

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, August : 2024

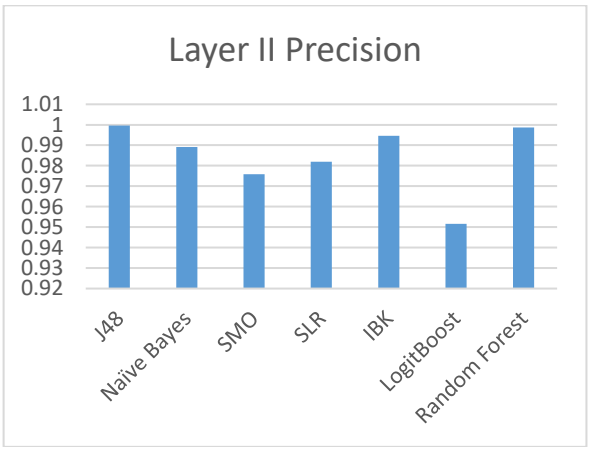
exhibits differently with higher value when compared to SVM and logiBoost which gets minimum value i.e 65%. The Naïve Bayes performed with a minimum value for metrics Precision, F Measure and AUC when compared to others.

From the above table it is observed that the J48, Random Forest, and IBK are in top position with differences of 1.8 and 2.6 when compared to J48. The SMO yields better CMR value when compared to Layer I with a difference of 2 and with a marginal difference of 0.2 of IBK The Naive Bayes classier is exhibited with lowest performance among all the classifiers and the second lowest classifier is LogiBoost with a CMR difference of 0.4 followed by SMO and SLR.

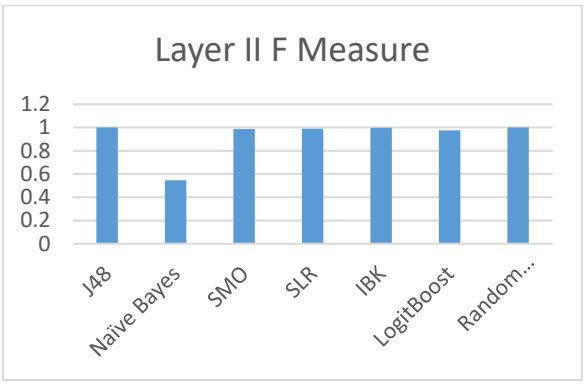
By observing these top three ranked classifiers, the accuracy of the J48 is higher than random forest and IBK with a very in significant differences of 0.002 and 0.006 respectively. The ROC curves of J48 and random forest occupies equally more area than IBK, but it is negligible. Precision, Recall and F Measure metrics are also followed this criterion. The recall values are higher than the precision values with a insignificant differences. The Naïve Bayes recall value is very less than the precision value. From these results the Naïve Bayes performs very poorly when compared to others.



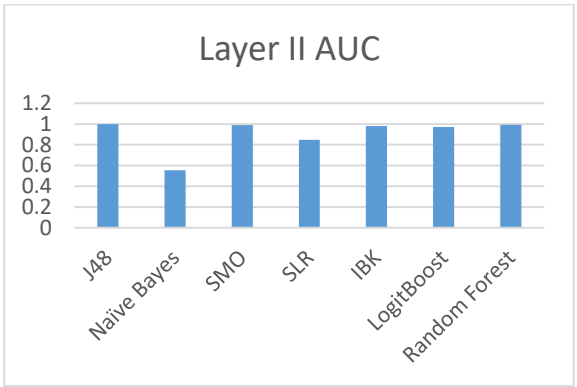
(a)



(b)

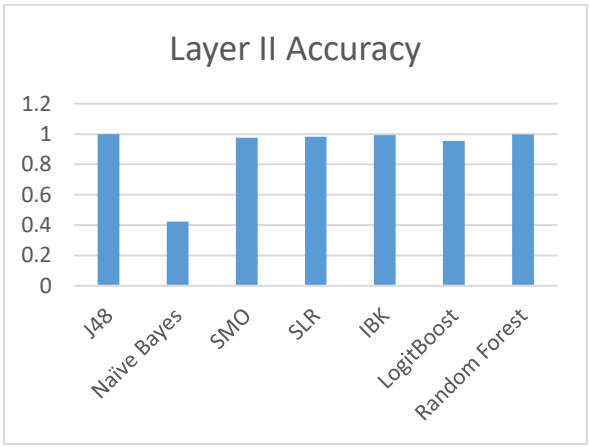


(c)



(d)

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, August : 2024



(e)

Figure 3: Comparison of various Performance Metrics for different classifiers of Layer II

Table 4: Overall Computational time for layer I and Layer II of different Classification Models

Type of Learning Model	Algorithm	Computational time
Decision Tree	J48	00:02:19
Probabilistic	Naïve Bayes	00:08:51
Parametric	SVM	01:53:03
Distance Based	IBK	00:02:19
Boosting	LogiBoost	00:04:32
Bagging	Random Forest	00:27:34

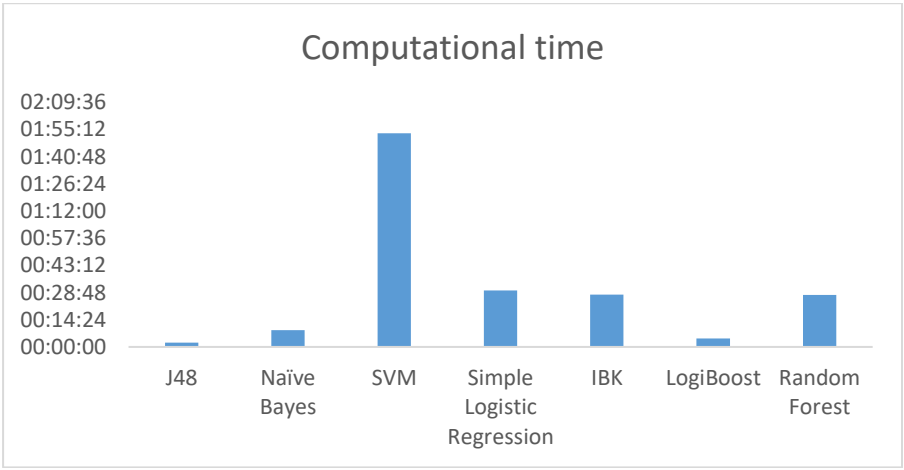


Figure 4: Comparison of building time for different classifiers

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, August : 2024

The following table and graph illustrates the effect of various performance measures for layer 1 and Layer II along with the total computational time of the layer I and II.

Algorithms	CMR of Layer I	CMR of Layer II	Rank of Computational Time
J48	6.8	6.8	**7**
Naïve Bayes	1.6	1.6	**3**
SMO	2	4	**1**
SLR	3.2	3.6	**5**
IBK	5	4.2	**7**
LogitBoost	2.8	2.4	**4**
Random Forest	6.4	5	**2**

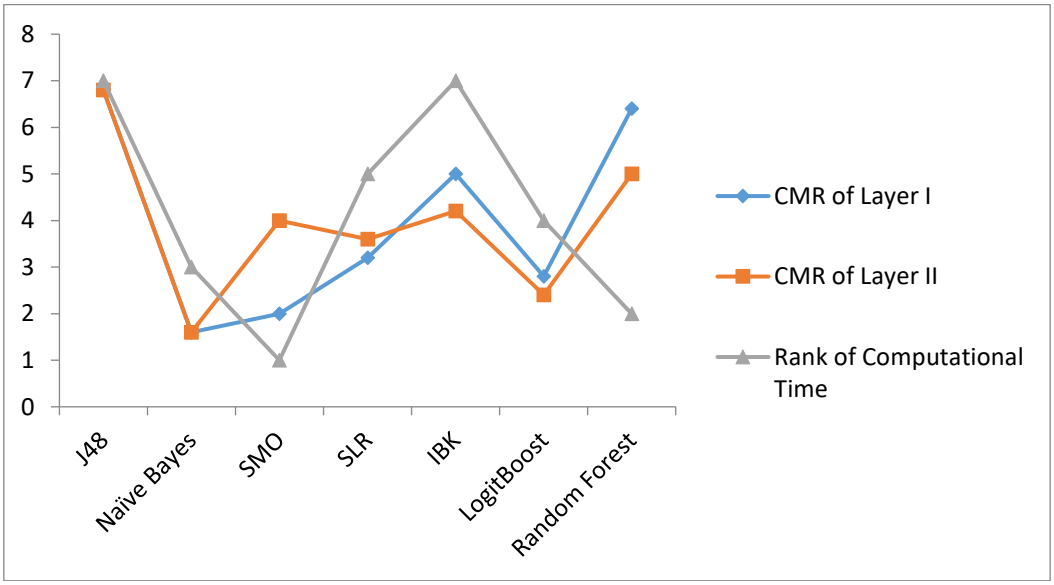


Figure 5: Comparison of computational time rank , layer I and layer II ranks for different classifiers

The Layer II of SMO and SLR is performed better than Layer I, whereas j48 and Naïve Bayes equally performs, and others are performed with a less perform ace of marginal difference. Among the top three ranked classifiers of layer 1 and layer II RF and IBK yields lowest values of layer II when compared to Layer I whereas J48 maintains equal performance with a highest CMR. Computational time point of view the J48 gain in computational time is high followed by IBK and SVM gives the lowest gain in computational time. Among the top three ranked classifiers, Random Forest computational time is high but if all metric values are higher than IBK. So, to consider the computational time as a one of the crucial parameter to decide whether the classifier is superior or not? Then the order of these three classifiers is J48, IBK and Random

Forest because the various metrics of differences are very nominal in the interval from0.01 to 0.004 for layer II and it is from 0.035 to 0.012.

Based on the above discussions it is suggested that J48(tree based classifier) and IBK(distance based classifier)are superior classical classifiers. Random Forest other tree-based classifier is also very good w.r.t to performance metrics wise but their building time is very costly.

IV Conclusion

In this paper presented an evaluation of seven classical ML models on up-to-date datasetCIRA-CIC-DoHBrw-2020 , which is publicly available and it consists of HTTPS traffic flows**.** To assess the efficiency of these models through two layered binary classification model, i) classify DoH traffic from Non DoH and ii)characterize the benign and malicious DoH flows. Based on the experimental results of the both layers the decision tree model J48 performs better than others and the building time also minimum when compare to others. Regarding to the performance the Random Forest is in second position but the building time is very high with a difference of 25 sec. The distance based model IBK is in next position when compare with performance metrics but the building time is equal to J48. From this evidence of these experimental results J48 is outperformer classifier among these classical models. As a network security perspective the building time is also a crucial component to achieve better detection rate, IBK is also considered as a one of the best classifier with a insignificant differences of the different metrics of layer I and II.

References

Zhang, C., (2021) Zhang, C., Chen, Y., Meng, Y., Ruan, F., Chen, R., Li, Y., & Yang, Y. (2021). A Novel Framework Design of Network Intrusion Detection Based on Machine Learning Techniques. *Security and Communication Networks*, *2021*.

Vekshin, D., et al. (2020) Vekshin, D., Hynek, K., &Cejka, T. (2020, August). Doh insight: Detecting dns over https by machine learning. In *Proceedings of the 15th International Conference on Availability, Reliability and Security* (pp. 1-8).

Banadaki, Yaser M. (2020) Banadaki, Yaser M. "Detecting Malicious DNS over HTTPS Traffic in Domain Name System using Machine Learning Classifiers." *Journal of Computer Sciences and Applications* 8.2 (2020): 46-55.

MontazeriShatoori, et al. (2020) MontazeriShatoori, Mohammadreza, et al. "Detection of DoH Tunnels using Time-series Classification of Encrypted Traffic." *2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*. IEEE, 2020.

Jafar, M.T., et al. (2021) Jafar, M. T., Al-Fawa'reh, M., Al-Hrahsheh, Z., &Jafar, S. T. "Analysis and Investigation of Malicious DNS Queries Using CIRA-CIC-DoHBrw-2020 Dataset*" 1st International Conference on Computing and Machine Intelligence (ICMI 2021)*

A. Nadler, et al. (2019) A. Nadler, A. Aminov, and A. Shabtai, "Detection of malicious and low throughput data exfiltration over the DNS protocol," Comput. Secur., vol. 80, pp. 36–53, 2019.

Tally, M. T., &Amintoosi, H. (2021). A hybrid method of genetic algorithm and support vector machine for intrusion detection. *International Journal of Electrical & Computer Engineering (2088-8708)*, *11*(1).

Hrushak, S., &Pavlenko, C. (2020, April). Advantages of DNS-over-HTTPS over DNS. In *COMPUTER AND INFORMATION SYSTEMS AND TECHNOLOGIES*.

Hao, M., Corral-Rivas, J. J., González-Elizondo, M. S., Ganeshaiah, K. N., Nava-Miranda, M. G., Zhang, C., ... & Von Gadow, K. (2019). Assessing biological dissimilarities between five forest communities. *Forest Ecosystems*, *6*(1), 1-8.

Vani, R. (2017). Towards efficient intrusion detection using deep learning techniques: a review. *Int J Adv Res ComputCommunEng ISO*, *3297*, 2007.

S. Al-Emadi, A. Al-Mohannadi and F. Al-Senaid, "Using Deep Learning Techniques for Network Intrusion Detection," *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, 2020, pp. 171-176, doi: 10.1109/ICIoT48696.2020.9089524.

Thaseen, I. S., Poorva, B., &Ushasree, P. S. (2020, February). Network intrusion detection using machine learning techniques. In *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)* (pp. 1-7). IEEE.

Rao, B. B., & Swathi, K. (2017). Fast kNN classifiers for network intrusion detection system. *Indian Journal of Science and Technology*, *10*(14), 1-10.
DOI10.17485/ijst/2017/v10i14/93690

Biswas, S. K. (2018). Intrusion detection using machine learning: A comparison study. *International Journal of pure and applied mathematics*, *118*(19), 101-114.