https://doi.org/10.33472/AFJBS.6.13.2024.2239-2353

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

INTRUSION DETECTION USING ML AND DL ALGORITHMS

Parimala Garnepudi[1], Mr R Kalyan Chakravarthy[2], Vara Lakshmi Bhavirisetti[3], Rushitha Sri Gogineni[4]

[1]Assistant Professor, Dept of CSE, VFSTR Deemed to be University
[2]TGT, Dr B R Ambedkar Gurukulam
[3]Department of CSE VFSTR Deemed to be University
[4]Department of CSE VFSTR Deemed to be University

garnepudi.parimala@gmail.com, parimalachakri@gmail.com, vara.b2002@gmail.com, rushithagogineni2002@gmail.com

Article Info

Volume 6, Issue 13, July 2024

Received: 04 June 2024

Accepted: 01 July 2024

Published: 26 July 2024

doi: 10.33472/AFJBS.6.13.2024.2239-2353

ABSTRACT:

The attacks in Vehicular Ad Hoc Networks have attracted many researchers in recent years for their potential in improving road safety, traffic efficiency and infotainment services. However, VANETs run in an open and dynamic environment, so they are subject to different security threats. Therefore, it is essential to design strong Intrusion Detection Systems. Intrusion detection systems have also been obtained within various procedures to detect intrusions based on machine learning and deep learning. In this paper, we design an Intrusion Detection System and apply Machine Learning and Deep Learning techniques to automobile networks. Machine Learning and Deep Learning systems have long been the basis for IP traffic and other uses, inspiring methods while providing concise expertise. Accuracy: The highest accuracy is obtained with the logistic regression method in machine learning and deep learning approaches with the accuracy of 99.38%. Several approaches like RNNs have been demonstrated to outperform with running accuracy and detective for each intrusion type in comparison to other methods and can further be applied. Therefore, investigating machine learning and deep learning approaches in terms of intrusion detection rate, accuracy, and false positives
needs further exploration.

Keywords: *Machine learning, Deep Learning, intrusions, attacks, network security, datasets, metrics.*

1. Introduction:

In recent years, there have been significant advancements in the field of network security, leading to the development of various new techniques and technologies. Vehicular ad hoc networks (VANETs) are a promising area of research, focusing on enabling communication between connected vehicles and infrastructure (V2X). One of the key challenges in VANETs is the secure and reliable exchange of information, given the dynamic and potentially hostile nature of the environment. Intrusion Detection Systems (IDS) have thus emerged as a crucial component in ensuring the security and integrity of VANETs. IDS analyze and monitor network traffic to detect and prevent any unauthorized access or malicious activities. By deploying IDS in VANETs, it is possible to identify and respond to potential security threats in real- time, safeguarding the communication and data exchange between vehicles and infrastructure.

- **Network-Based Intrusion Detection System:**

A network-based intrusion detection system (NIDS) is a security solution that monitors the entire network for suspicious traffic,
analyzing the protocol activity to detect potential security breaches.

- **Wireless Intrusion Detection System:**

A wireless intrusion detection system (WIDS) is an extension of NIDS that specifically focuses on monitoring and analyzing wireless networking protocols for abnormal or unauthorized activities within a wireless network.

- **Network Behavior Analysis:**

Network behavior analysis (NBA) is a security approach where the network is continuously monitored to detect and respond to suspicious behavior or potential threats, such as distributed denial-of-service (DDoS) attacks.

- **Host-Based Intrusion Detection System:**

A host-based intrusion detection system (HIDS) is a software solution that is installed on individual hosts or endpoints, monitoring the system's activities and logs for signs of unauthorized access or malicious activities. HIDS can detect attacks by identifying patterns or signatures and providing alerts and responses.

The anomaly-based IDS (AIDS) can detect the unknown malware and attacks by performing a deep analysis of the transmitted data [3]. The Machine Learning and Deep Learning algorithms evaluate the network condition by classifying the processed data into either normal or abnormal classes. They train and test AIDS for detecting attacks and use false alarm rate and accuracy to measure the ability of AIDS in using different datasets like NSL-KDD,KDD'99 and the UNSW-NB15.

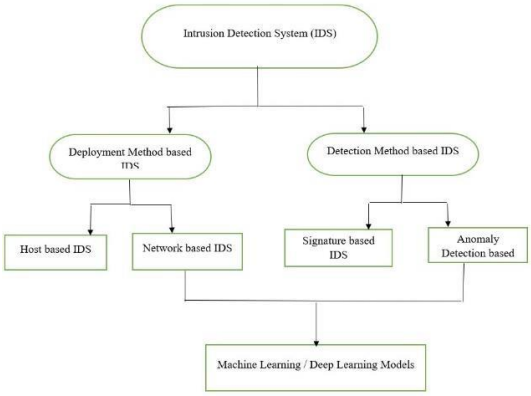


Fig 1: Intrusion detection system classification taxonomy.

2. Related work

The Discovery of knowledge on KDD cup conducted in the year of 1999 and featured the KDD'99 dataset[4]. And it is a subset of the Defence Advanced Research Projects Agency (DARPA) 1998 benchmark dataset. Since the contest, researchers had extensively used the dataset to train and test models for an accurate IDS.

This methodology used the integration approach for combining the various data models to identify the malicious nodes in the VANET system using the data-driven model. This methodology was tested on various environmental VANET systems to validate the proposed hybrid data-driven model. The modelling intrusion detection systems done using the machine learning(ML) and deep learning(DL). An example of this, the application of decision tree in the KDD'99, Pfanhringer. Sabhani and Serpen [5] applied the decision trees approach and obtained good accuracy but the approach did not perform well with U2R and R2L attacks that they are minor classes and include a large proportion of new attack types.

Mukkamala et al. [6] presented a neural network (NN) and a support vector machine (SVM) to perform whether it is an attack or normal classification. They also extracted thirteen important features, trained the two

models using these extracted features, and reportedd the results. They concluded that both SVM and neural networks provided accurate results, with the SVM performing slightly better. Bajaj and Arora [7] examined the contribution of all the 41 features in NSL KDD dataset and found that Naïve Bayes, SVM and simple cart methods were applied for classification. Three out of 41 features [urgent, num_outbound_cmds and is_hot_login] using the NSL-KDD training dataset which do not have any significant role in the ddetection of attacks. Five out of 41 features [su_attempted, num_file_creations, num_access_files, dst_host_count and dst_host_rerror_rate] had little significant role in ddetection of attacks.

Pervez et al. [8] also proposed an approach consisting of merging feature selection and classification for multiple class NSL-KDD intrusion detection dataset by using Support Vector Machine (SVM).

The proposed method achieved 91% classification accuracy using only three input features andd 99% classification accuracy using 36 input features, while 41 input features achieved 99% classification accuracy. It is important to mention that some of the researchers have been working on KDD'99 dataset samples rather than the complete training dataset ddue to the size of this dataset [9].

Ingre et al. [10] evaluated the performance of NSL-KDD dataset using ANN. Their work was basedd on the findings of Bajaj andd Arora [11]. Further, they found that features [land), wrong_fragment , num_failed_login and root_shell] have all zero values in the dataset. Thus, they reduced the number features is NSL KDD training and testing datasets to 29 features. For five class classification, the system had good capability to find the attack for the particular class in NSL-KDD dataset. In the year of 2015, Moustafa and Slay [12] pointed out that KDD'99 and NSL-KDD datasets did not mirror the current attacks a

system in IDS and thus derived a totally new network-based dataset called UNSWNB15, which is a comprehensive one. With this dataset, its features are different from the KDD'99 dataset but share some of them as well [13].

Moreover, they studied the properties of both datasets like UNSW NB15 and KDD99. The UNSW-NB15 algorithm has been replicated to the KDD-99 data set to demonstrate its efficiency and a rule-based feature selection approach was implemented on both datasets. However, it is not known how different features were compared in this work since the

features considered were different. The differences between the results were obtained, and it can be seen that the original KDD'99 features are less efficient than the replicated UNSW-NB15 hence, the UNSW-NB15
dataset had a higher accuracy than the KDDCUP'99 dataset. FPR of KDD'99 dataset is lower than FR of the UNSW- NB15 dataset [13].
On one hand, Hosseinzadeh and Kabiri [15] used ant colony optimisation model to tackle the famous KDD'99 dataset, whereas the outcomes of both algorithms were very close to each other suggesting that the algorithms were able to gather the clear-cut information from the dataset.A set of 5 best features [urgent (9), num_failed_logins
,count, rerror_rate and dst_host_srv_diff_host_rate] were selected under the category of Normal, a set of 4 best features [durations ,flag, root_shell and dst_host_srv_diff_host_rate] were selected under the category of DoS, a set of 4 best features [service, dst_bytes, count, serror] was selected under the category of U2R, a set of 3 best features [count, srv_count, diff_srv_rate] under R2L and a set of 8 best features [protocol_type, logged_in, flag, num_compromised num_access_files , hot,
, diff_srv_rate, dst_host_diff_srv_rate]

under the category of Probe. The evaluation of the method demonstrated its effectiveness, the number of features decreased on average by 88% and the detection error was reduced by up to 24% while based on KDD'99 dataset. Earlier, Zargari and Voorhis [16] encountered scrambled KDD-dataset which is processed by joining the voting system technique and Weka feature selection to acquire the best subset of features. The results confirmed that the selected subset of features was the best compared to the competing subsets that were calculated with data mining methods. The KDD dataset was implemented in the experimental setup and the evaluation of the developed features provided better detection rates.
In 2015, Mostafa and Slay [17] published the UNSW-NB15 dataset which had a few new features compared to KDD'99. NB15 and KDD'99 have only a few common features which make it hard to compare the two datasets. This research looks at the features of dataset UNSW-NB15 so as to decline the number of features and then suggest the subset of significant features that can be used for detection of intrusion in network traffics. Moreover, KDD'99 data set will be used to refine and analyse the results of Restrictions in terms of similarities and differences.

Algorithm:
i. Data Preparation
ii. Model Architecture
iii. Training
• Data Preprocessing
• Model Training
iv. Evaluation
v. Deployment

i. Data Preparation: Collect and pre-process real world data from VANETs, creating sequences of feature vectors.

ii. Model Architecture:
The VANET data is represented by the sequences of feature vectors which means that we can use LSTM to sequence the data. Introduce Recurrent Neural Networks (RNNs) to extract ordering relationships and patterns that are depicted over time in the data. Apply one or some RNN layers (e.g., Logical RNN, LSTM, GRU) to the input sequences for processing. Convolutional layer use can be one more extract specificity and representation of the input

sequences. By way of filters of assorted sizes, spatial data can be filtered out at whatever depth. Carefully choose such functions as activation ones (e.g., ReLU) and pooling layers (e.g., Max Pooling) to remove dimensionality and to isolate relevant features at the same time. Apply the flatting of the RNN and CNN layers into a one- dimensional vector. Proceed feeding the flattened vector through either multiple fully connected layers to perform feature transformation and extraction processes. Introduction of nonlinear relationships (e.g., ReLU activation functions) increases the model's expressive power, hence bringing it much closer to the real- world problems.

iii. Training:

- **Data Preprocessing:** Collect and preprocess real- world data from VANETs, creating sequences of feature vectors. Split the pre processed data into training and validation sets.
- **Model Training:** Train the model using supervised learning techniques, optimizing a suitable loss function (e.g., categorical cross-entropy, binary cross- entropy) and an

- optimization algorithm (e.g., Adam, RMSprop).

iv. Evaluation:

The validation set must be used for evaluation of the trained model performance basing on accuracy, precision, recall, and F1 scores.

v. Deployment:

Deploy the trained IDS model to work in the VANET framework for the immediate security monitoring and detection of threats.

- Monitoring and Updating: Evenly monitor the performance of the already deployed ICP system. Continuously maintain the effective IDS system by upgrading it according to the emerging threats and the fluctuations in the VANET environment.

3. Dataset Description:

The data set considered is NSL-KDD dataset [18], UNSW-NB15, KDD'99. The NSL-KDD dataset is one of the most popular which is used in the evaluation of IDS frame works. In addition to normal network traffic, the NSL-KDD contains these four types of intrusions: Denial of service(DoS), User to Root($U2R$), Remote to user($R2L$) and $Probe$. The NSL-KDD dataset is 52.3 MB in size and it includes the two subsets of the NSL-KDD dataset are considered namely, the NSL-KDD- Train and the NSL-KDD-Test. The (Table 3) NSL-KDD is made of 42 features of which 3 are non-numeric and 39 are numeric as depicted in Table 1. The dataset is divided into two partitions: the NSL- KDD-Train , which is 75 % of the NSL- KDD-Train and it will be used for training, the NSL-KDD-Test Evaluation that is 25 % the NSL-KDD-Test and it will be used for evaluation after the training process[19]. The table below shows the number of records in each dataset, and also the number of records associated with each attack type (Table1)

	Total	Normal	DoS	Probe	U2R	R2L
Train Test	1,25,973 22,544	67,343 9711	45,927 7458	11,656 2421	52 200	995 2654

Table 1: Details of Normal and Attack data.

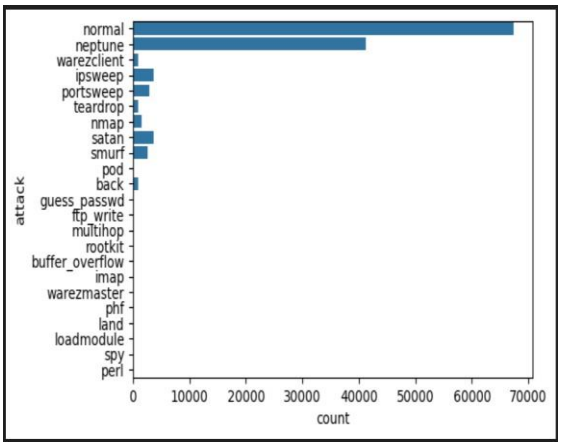
The UNSW-NB15 dataset contains 9 attacks, including DoS, worms, fumigations, and the Shellcode. The training set includes 175,341 records, while the test set contains 82,332 abnormal and normal records of several types[20]. And it consists of 45 features. The 9 attacks of UNSW-NB15 training dataset is represented in the form of graph(Fig 4).

Table with types of Attacks in the dataset.

Type	Test	Training
Worms	44	130
Shellcode	378	1133
Backdoor	583	1746
Analysis	677	2000
Reconnaissance	3496	10491
Dos	4089	12264
Fuzzers	6062	18184
Exploits	11132	33393
Generic	18871	40000
Normal	37000	56000
Total	**82,332**	**1,75,341**

Table 2. Details of UNSW-NB15

Fig 4: Attack categories on NSL – KDD data set



No.	Attribute	Format	No	Attribute	Format
f1	duration	numeric	f22	is_guest_login	numeric
f2	protocol_type	Non-numeric	f23	count	numeric
f3	service	Non-numeric	f24	srv_count	numeric
f4	flag	Non-numeric	f25	serror_rate	numeric
f5	src_bytes	numeric	f26	srv_serror_rate	numeric
f6	dst_bytes	numeric	f27	rerror_rate	numeric
f7	land	numeric	f28	srv_error_rate	numeric
f8	wrong_fragment	numeric	f29	same_srv_rate	numeric
f9	urgent	numeric	f30	diff_srv_rate	numeric
f10	hot	numeric	f31	srv_diff_host_rate	numeric
f11	num_failed_logins	numeric	f32	dst_host_count	numeric
f12	logged_in	numeric	f33	dst_host_srv_count	numeric
f13	num_compromised	numeric	f34	dst_host_same_srv_rate	numeric
f14	root_shell	numeric	f35	dst_host_diff_srv_rate	numeric
f15	su_attempted	numeric	f36	dst_host_same_src_port_host_rate	numeric
f16	num_root	numeric	f37	dst_host_srv_diff_host_rate	numeric
f17	num_file_creations	numeric	f38	dst_host_serror_rate	numeric
f18	num_shells	numeric	f39	dst_host_srv_serror_rate	numeric
f19	num_access_files	numeric	f40	dst_host_rerror_rate	numeric
f20	num_outbound_cmds	numeric	f41	dst_host_srv_rerror_rate	numeric
f21	is_host_login	numeric	f42	label	numeric

Table 3: NSL-KDD Attributes Description.

Table with Attributes, description and the type of the attribute used in the UNSW-NB15 Dataset.

Name	Description	Type
Proto	Transaction protocol	Nominal
State	The state and its protocol are specified e.g. ACC, CLO, CON, ECO, ECR, FIN, INT, MAS, PAR, etc.	Nominal
Sttl	Living value from source to destination	Integer
Swin	Advertising value of source TCP window	Integer

Dwin	Destination TCP window advertisement value	Integer
Tcprtt	TCP connection setup round-trip time, the sum of 'synack' and 'ackdat'.	Integer
Smeansz	Mean of the how packet size transmitted by the src	Integer
Response_body_len	Actual uncompressed content size of the data transferred from the server's http service.	Integer
Ct_state_ttl	No. for each state (13) according to the specific source/destination time value range for living .	Integer
St_dst_sport_ltm	No connections at 100 connections with the same destination address (14) and source port .	Integer
Is_sm_ips_ports	This variable is given value of other variable when the source (15) and the destination IP address is equal and port numbers (16) equal.	Binary

Table 4: UNSW-NB15 Dataset attributes.

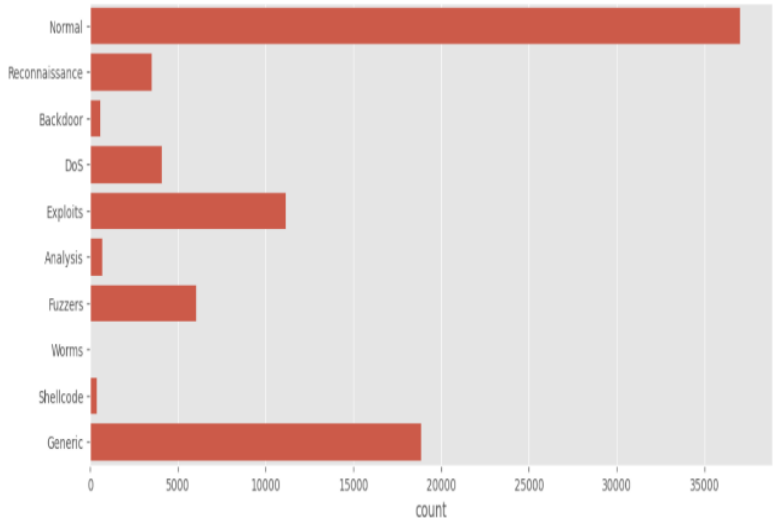


Fig 4: Attack categories in UNSW-NB15 training dataset.

The KDD-99 cup dataset is the subset of DARPA-98 dataset. KDD-99 dataset is a multi-variety dataset. The characteristics of the attributes used in the dataset are categorical and integer and It has fourty- two attributes.

Class	Training Set	Percentage	Test Set	Percentage
Normal	812,814	75.611%	60,593	19.481%
DoS	247,267	23.002%	229,853	73.901%
Probe	13,860	1.289%	4,166	1.339%
R2L	999	0.093%	16,189	5.205%
U2R	52	0.005%	228	0.073%
Total	1,074,992	100%	311,029	100%

Table 5: Details about the KDD'99 datasets.

2. Methodology:

The dataset NSL-KDD, UNSW-NB15 is chosen as the basis for the planned experiment. The NSL-KDD dataset consists of 42 features being labelled an attack or normal, and showing the identification of attack.

The NSL-KDD contains these four types of intrusions:

a) Denial of service(DoS): It is the Various forms of attacks, such as SYN Flood [21].
b) User to Root($U2R$): It does not contains allowed remote access [22].
c) Remote to user($R2L$): It identifies the unauthorized access to super user rights on the local system [23].
d) Probe: It is used for Monitoring and examination [24].

The NSL-KDD dataset has 42 attributes which are classified into three categories. They are basic, content, and the traffic features. Without the payload, the basic features can be derived from the packet headers. To calculate the traffic features the time interval is used.
The UNSW-NB15 dataset includes 45 features from which we can select the important features from the input it can
leads to simplification of the model processing and get the more accurate rates. Since the NSL-KDD and UNSW-NB15 datasets are reports the typical flow in network traffics, these are with the features with the same characteristics and behave similarly.

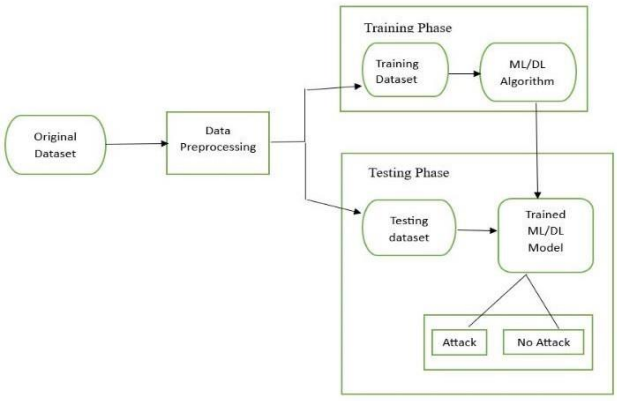


Fig 5: Generalized machine learning and deep learning network based intrusion detection system.

Evaluation metrics:
The Assessment of performance metrics for IDS based on certainty matrix values for ML and DL approaches [25]. The Evaluation metrics are different attributes used in the confusion

matrix.

1. True Positive (TP): The data instances correctly predicted as an Attack by the classifier.
2. False Negative (FN): The data instances wrongly predicted as Normal instances.
3. False Positive (FP): The data instances wrongly classified as an Attack.
4. True Negative (TN): The instances correctly classified as Normal instances.
5. The diagonal of confusion matrix denotes the correct predictions and the non-diagonal elements are the wrong predictions of a certain classifier (table 6).

- **False alarm rate:** It is also known as False-positive and is defined as percentage to all the normal samples with wrongly expected attack sample [26].

$$\text{Accuracy of False Rate} = \frac{\text{False Positive}}{\text{False Positive+True Negavtive}}$$

- **True negative rate:** It is the right number of normal samples which is divided by the overall number of normal samples [27]. Precision: It is the ratio of correctly expected Attacks to all Attacks Samples [26] [28].

$$\text{Accuracy of True Rate} = \frac{\text{True Negative}}{\text{Trie Negative+False Postive}}$$

- **Precision:** It is the ratio of correctly expected Attacks to all Attacks Samples [24] [25].

$$\text{Precision} = \frac{True\ Postive}{True\ Postive + False\ Postive}$$

- **Recall:** It is the proportion of all Attacks samples correctly listed to all attacks samples that are Attacks. It is also known as a Detection Rate [29] [30].

$$\text{Recall} = \frac{\text{True Postive}}{\text{True Postive+False Negative}}$$

- **F-Measure:** Precision and Recall are combined to form the harmonic mean. To put it another way, it is a mathematical method for evaluating a system's accuracy by taking into account both precision and recall [31] [32].

$$\text{F} - \text{Measure} = 2\frac{\text{Precision} * \text{Recall}}{\text{Precision+Recall}}$$

		Predicted class	
		Attack	**Normal**
Actual Class	**Attack**	True Positive	False Negative
	Normal	False Positive	True Negative

Table 6: Confusion Matrix
These are the evaluation metrics which are calculated by testing the proposed methodology by using the dataset.

3. Analysis and Results:

A. Experimental setup:

8GB of RAM, an AMD Ryzen 5 11th Gen GPU CPU, and Windows 11 were used to develop the suggested design. The model was trained and tested using the Python 3.8 version and the Visual Studio code. The Visual Studio Code was integrated with the seaborn, metrics and NumPy packages for pre-processing. Other libraries, like as Pandas, are used in the feature extraction process. The Kera's library was advised for ML and Dl models development.

4. Result

Performance evaluation of the proposed method is done by using the NSL-KDD dataset. IDS can be built only when there is an availability of an effective dataset. The performance of the proposed ML and DL algorithms are presented in this section. The main Advantages of using dataset are training set which doesn't consist of any
redundant records, and missing values. So that the classifier no longer produces the biased result.
The (table 7) results obtained by using the three datasets.

S. No	Datasets	SVM	DT	LR	KNN	CNN	RNN
1	KDD'99	0.79	0.78	0.79	0.78	0.82	0.82
2	NSL-KDD	0.79	0.78	0.80	0.79	0.82	0.81
3	UNSW-NB15	0.80	0.79	0.83	0.75	0.84	0.81

Table 7: Results using NSL-KDD dataset with ML and DL algorithms.

5. Conclusion:

The KDD'99 dataset is one of the earliest and widely used datasets for intrusion detection research. It contains a large number of features that are from network traffic, including normal and various types of attack instances. The NSL-KDD dataset is an improved version of the original KDD'99 dataset, addressing such as redundancy and ambiguity. It provides more balanced distribution of attack types and removes duplicate records. The NSL- KDD dataset serves as a valuable for evaluating IDS algorithms, offering a more realistic and representative dataset compared to KDD'99. But, it still has limitations, such as static attacks and lack of recent attack types. The UNSW-NB15 dataset is a more recent and comprehensive dataset for network intrusion detection, featuring diverse attack and modern network traffic patterns. It includes both packet-level and flow-level data, making it suitable for evaluating anomaly-based detection methods. The UNSW-NB15 dataset represents a significant in intrusion detection research, offering a challenging for evaluating IDS solutions. Its inclusion of real-world attack scenarios and diverse

network traffic characteristics make it highly relevant for effectiveness of modern intrusion detection techniques.

Future Work:

The categorization of classifiers is proposed in terms of lazy and eager learners. The experimental work has been carried out to evaluate the performance of the selected ML classifiers based on proposed categorization namely KNN, LR, DT, SVM and the DL algorithms are CNN, RNN for detection of intrusion. These classifiers are tested on UNSW-NB15, NSL- KDD, KDD'99 datasets. The classifiers are compared on the basis of precision, recall, F1-Score, accuracy. The results show that LR classifier in ML algorithms is better than other classifiers on UNSW-NB15, NSL KDD, KDD'99 using selected parameters. The accuracy of CNN classifier comes out to be best in the DL algorithms. In future, this work can be extended for selective attributes and multiclass classification for detection of intrusion using pothole detection.

6. References:

1. T. M. Mitchell, "Machine Learning", 1st ed. New York, NY, USA: McGraw-Hill, 1997.
2. S. Albrecht, J. Busch, M. Kloppenburg,
3. F. Metze, and P. Tavan, "Generalized radial basis function networks for classification and novelty detection: Self- organization of optimal Bayesian decision," Neural Netw., vol. 13, no. 10, pp. 1075–1093, 2000
4. A McCallum and K. Nigam, "A comparison of event models for naive Bayes text classification," in Proc. AAAI Workshop Learn. Text Categorization, vol. 752. Madison, WI, USA, 1998, pp. 41–48.
5. A. M. Kibriya, E. Frank, B. Pfahringer, and G. Holmes, "Multinomial naive Bayes for text categorization revisited," in Proc. Aust. Conf. Artif. Intell., Cairns, QLD, Australia, 2004, pp. 488–499.
6. Sabhnani M., and Serpen G., "Application of machine learning algorithms to KDD intrusion detection dataset within misuse detection context", International on Machine Learning, Models, Technologies and Applications, pp. 209-215, 2003.
7. Mukkamala S, Janoski G and Sung A 2002 proceedings of the IEEE International Joint Conference on Neural Networks (ANNIE), St. Louis, MO pp 1702–1707.
8. Bajaj and Arora, "Improving the Intrusion Detection using Discriminative Machine Learning Approach and Improve the Time Complexity by Data Mining Feature Selection Methods", International Journal of Computer Applications (0975- 8887), Volume 76-No.1, August 2013. [Onine] available:
9. Pervez M. S. and Farid D. M., "Feature selection and intrusion classification in NSL-KDD cup 99 dataset employing SVMs," The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014), Dhaka, 2014, pp. 1-6.
10. Zargari S. and Voorhis D., "Feature Selection in KDDdataset," International the 2012 Corrected on Emerging Intelligent
11. Data and Web Technologies, Bucharest, 2012, pp. 174-180.
12. Ingre B. and Yadav A., "Performance analysis of NSL-KDD dataset using ANN," 2015 International on Signal Processing and Communication Engineering Systems, Guntur, 2015, pp. 92-96.
13. Moustafa N. and Slay J., "The significant features of the UNSW- NB15 and the KDD99 sets for Network Intrusion Detection Systems", the 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS 2015), collocated with RAID 2015, 2016
14. Aghdam Hosseinzadeh M. and Kabiri, "Feature Selection for Intrusion Detection System Using Ant Colony Optimization," International Journal of Network Security,

Vol 18, No.3, May 2016, pp.420-432.
15. [13] G. H. John and P. Langley, "Estimating distributions continuous in Bayesian classifiers," in Proc. 11th Conf. Uncertainty Artif. Intell., Montreal, QC, Canada, 1995,
16. pp. 338–345.
17. Sara A. Althubiti, Eric Marcell Jones Jr, Kaushik Roy "LSTM for Anomaly- Based Network Intrusion Detection" 2018, 28th International Telecommunication networks and applications conference.
18. GravesA,MohamedA,HintonG.Speec hrecognitionwithdeeprecurrentneural
19. networks. Paper presented at: Proceedings of the IEEE on Acoustics, Speech and Signal Processing. Vancouver, BC, Canada: IEEE; 2013:6645-6649.
20. Lawrence S, Giles CL, Tsoi AC, Back AD. Face recognition: a convolutional neural-network approach. IEEE Trans Neural Netw.
21. Xiao Y, Xing C, Zhang T, Zhao Z. An intrusion detection model based on feature reduction and convolutional neural networks.
22. The NSL-KDD: NSL KDD dataset [Online].[19A_Deep_Learning
23. _Method_With_Filter_Based_Feature_Engineering_for_Wireless_Intrusi on_Detection_System].
24. O. V. Lee et al., "A malicious URLs detection system using optimization and machine learning classifiers," Indones. J. Electr. Eng. Comput. Sci., vol. 17, no. 3, pp. 1210–1214, 2020.
25. D. Warburton, "DDoS Attack Trends for 2020," F5 Application Threat INtelligence, 2021. Available:
26. [22] Y. Ti, M. Faizal, A. Razak, M. Fadli, T. Fui, and A. Firdaus, "Grasp on Next Generation Security Operation Centre (NGSOC): Comparative Study," Int. J. Nonlinear Anal. Appl., vol. 12, no. 2, pp. 867 896, 2021.
27. [23] S. R. T. Mat, M. F. A. Razak, M. N. M.
28. Kahar, J. M. Arif, S. Mohamad, and A. Firdaus, "Towards a systematic description of the field using bibliometric analysis: malware evolution," J. Sci., pp. 1–38, 2021. [24] K. A. Taher, B. Mohammed Yasin Jisan, and Md. M. Rahman, "Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection," International in Conference 2019 on Robotics, Electrical and Signal Processing Techniques (ICREST), Dhaka, Bangladesh, Jan. 2019, pp. 643–646. doi: 10.1109/ICREST.2019.8644161.
29. [25] S. M. Sohi, J.-P. Seifert, and F. Ganji, "RNNIDS: Enhancing network intrusion detection systems through deep learning," Computers & Security, vol. 102, p. 102151, Mar. 2021, doi: 10.1016/j.cose.2020.102151.
30. [26] C. Kalimuthan and J. Arokia Renjit, "Review on intrusion detection using feature selection with machine learning techniques," Materials Today: Proceedings, vol. 33, pp. 3794–3802, 2020, doi: 10.1016/j.matpr.2020.06.218.
31. [27] C. A. M. and R. K., "Performance evaluation of data clustering techniques using KDD Cup-99.
32. [28] N. Shone, T. N. Ngoc, V. D. Phai, and
33. Q. Shi, "A Deep Learning Approach to Network Intrusion Detection," IEEE Trans. Emerg. Top. Comput. Intell., vol. 2, no. 1,
34. pp. 41–50, Feb. 2018, doi: 10.1109/TETCI.2017.2772792.
35. Y. Xiao, C. Xing, T. Zhang, and Z. Zhao, "An Intrusion Detection Model Based on Feature Reduction and Convolutional Neural Networks," IEEE Access, vol. 7, pp. 42210– 42219, 2019, doi: 10.1109/ACCESS.2019.2904620.
36. [30] D. Agrawal, C. Agrawal, and H. Yadav, "A Machine Learning Based Intrusion Detection Framework Using KDDCUP 99 Dataset," vol. 4, no. 6, p. 11.

37. [31] G. Meena and R. R. Choudhary, “A review paper on IDS classification using KDD 99 and NSL KDD dataset in WEKA,” in 2017 on Computer, Communications and Electronics (Comptelix), Jaipur, India, Jul. 2017, pp. 553–558. doi:
38. 10.1109/COMPTELIX.2017.800403 2.
39. [32] P. S. Bayerl, R. Karlović, B. Akhgar, and G. Markarian, Eds., Community Policing - A European Perspective: Strategies, Best Practices and International Publishing, 2017. doi: 10.1007/978-3-319- 53396-4.