[1]Mohammed M. Salih

[2]Basma M. Khaleel

Intrusion Detection System using Neural Network

Journal of Electrical Systems

Abstract: - Ensuring network security has consistently been a matter of utmost importance, especially concerning the confidentiality of enterprises and individuals' privacy. This significance is accentuated as valuable and sensitive information is continually transferred across networks. Furthermore, as various systems depend more on web services like e-government platforms, online banking, email, and e-commerce, the role of intrusion detection systems (IDS) has grown significantly. IDS systems are extensively employed to safeguard data and mitigate the impact of network intrusions and security breaches.

The research community seeks to develop an intrusion detection system to meet security challenges. However, a significant limitation of most existing systems is their lack of capability to identify novel forms of threats. Therefore, the present study suggests the utilization of machine learning methodologies, specifically neural networks, inside the framework of a network intrusion detection system (IDS).

The objective of this research is to construct (IDS) utilizing a neural network framework in order to identify and mitigate instances of unauthorized network access and hostile behavior. The system conducts an analysis of network traffic patterns and detects any intrusions and anomalies. The model is trained with the KDD-99 dataset, and the evaluation of the network's performance is conducted through the utilization of various metrics in addition to the confusion matrix. The experimental findings demonstrated that the proposed model attained a level of accuracy equivalent to 99.7%. Consequently, it may be deemed a viable solution for effectively tackling intrusion detection issues.

Keywords: neural networks, machine learning, intrusion detection, preprocessing**.**

I. INTRODUCTION

An intrusion detection system (IDS) is a mechanism designed to examine network traffic for any signs of unusual or suspicious behavior and issues alerts upon detecting such activities. It operates in conjunction with an Information and Event Management (SIEM) system. SIEM, in turn, compiles data from various origins and deploys alert filtering methods to differentiate between genuine security alerts and false alarms. Intrusion prevention systems also monitor incoming network packets to see if they include any malicious activity and send alerts if this occurs. While intrusion detection systems monitor networks for suspicious activity, they are also vulnerable to false alarms. There are several types of intrusion detection systems, as shown in Figure 1[1-2].

1. Network Attack Detection System (NIDS): Installed on the network to scan traffic from various devices connected to it, detect known attacks, and issue alerts [3].
2. Host Intrusion Detection Systems (HIDS): Installed on individual devices, monitoring incoming and outgoing data packets from these devices, and issuing alerts if suspicious behavior is detected on these devices[4].
3. A protocol-based intrusion detection system (PIDS): Controls and interprets the protocol between users and servers and aims to secure web servers [5].
4. Application Protocol-Based Intrusion Detection System (APIDS): Monitors and interprets communications at the application protocol level to detect intrusion into applications and their communications [6].
5. Hybrid IDS: Combines multiple types of intrusion detection systems to increase their efficiency in detecting threats [7].

[1] *Corresponding author: Al-Rafidain University College, Baghdad, Iraq

[2] Al-Rafidain University College, Baghdad, Iraq

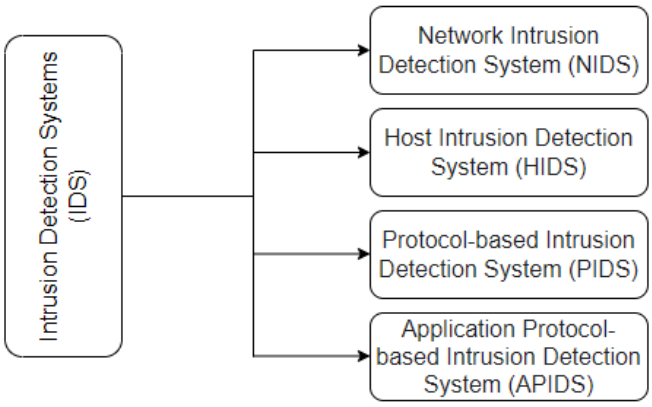

Figure 1 : Categories of Intrusion Detection Systems .

Intrusion detection is an important field in security and network protection. It is among the broadest topics explored in the world of information security and is essential to protecting organizations and individuals from cyber threats. However, no system is perfect without flaws.

Therefore, there is an urgent need to continue research and development in the field of intrusion detection systems to achieve the highest level of security. The use of artificial intelligence techniques has become an urgent necessity in a strong intrusion detection system capable of analyzing traffic in electronic networks, accurately detecting anomalous behavior, and achieving the best results in terms of system accuracy, intrusion detection rate, and reducing false positives.

Incorrect detection can cause the system to be overridden and not approved. Artificial neural networks can be defined as a technology based on the working principle of the human brain as shown in Figure 2 [8]. Artificial neural networks essentially mimic a biological neural network but use a limited set of concepts. Processing elements (also known as neural or cognitive) are linked to other processing elements. Neurons are usually arranged in a layer or vector, where the output of one layer serves as the input to the next layer and so on through the other layers until the output layer [9].

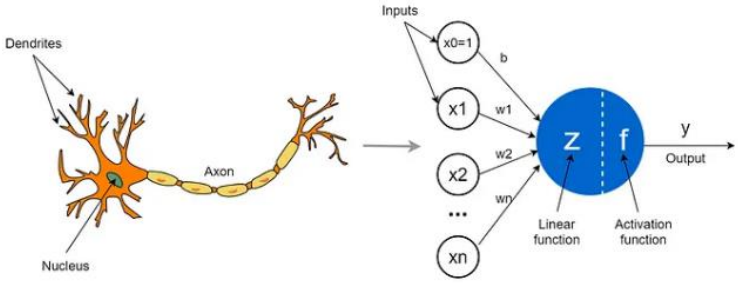

Figure 2 : The neural networks[8]

This research relies on three main components: intrusion detection systems (IDS), neural networks (NN), and databases. Standard network traffic data was used to evaluate intrusion detection systems and was preprocessed. The (NN) algorithm was used to build the proposed system, as these networks are useful in detecting intrusions and achieving high accuracy in distinguishing between anomalous activities in networks.

The performance of the proposed system was evaluated and compared using the confusion matrix as one of the tools for evaluating the performance of intrusion detection systems. Experiments were conducted and neural networks were trained on the dataset KDD-99 dataset. In our research, we seek to improve the efficiency of intrusion detection systems and provide higher security for organizations and individuals in the face of cyber threats.

II. RELATED WORK

We suggest training machine learning algorithms on the KDD-99 dataset with the aim of developing a reliable system for detecting network intrusions. This section discusses notable research in the field of designing systems that use machine learning (ML) techniques in intrusion detection and proposes the use of the KDD-99 dataset for experimental testing.

1. [10]: This research effort involves comparing different fusion methods for distinguishing normal network traffic from specific Denial of Service (DoS) attacks. The methods compared include majority voting, Dempster-Shafer, Naïve Bayes association, neural network (NN), and averaging. The researchers found that the halftone network with integrated radial basis performed the best. Fusion methods are used to combine the output of multiple classifiers to improve classification accuracy.
2. [9]: In this study, the authors introduced a neural network (NN) and a kernel Support Vector Machine (SVM) design for classifying attacks and normal network traffic. Thirteen features were extracted for training the model. The results showed that both algorithms were effective, with SVM slightly outperforming the NN. However, using only the thirteen selected features led to a slight drop in accuracy.
3. [12]: This research proposed a deep belief network (DBN) model for an Intrusion Detection System (IDS) using the KDD-99 dataset. The DBN was found to outperform Support Vector Machines (SVMs) and artificial neural networks (ANNs). DBNs are a type of deep learning model that has been successful in various machine learning tasks.
4. [13]: In this study, researchers presented an intelligent proposal for a hybrid deep network designed for binary and five-factor classification. Hybrid deep networks typically combine traditional machine learning techniques with deep learning methods to leverage the strengths of both. This approach is useful in various classification tasks.
5. [15]: The researchers in this study contributed to developing an intrusion detection model using the Random Forest (RF) algorithm with non-deep asymmetric encryption (NDAE). They used a hybrid dataset consisting of both the 10% KDD-99 and the entire NSL-KDD dataset. The researchers reported results for both five-class and thirteen-class classification, indicating that their solution not only provided accurate results but also significantly reduced training time. Random Forest is an ensemble learning method that can handle various types of data and is known for its accuracy and robustness.

These studies demonstrate the diversity of machine learning and deep learning techniques applied to intrusion detection and network security. Researchers are continually exploring and refining methods to improve the accuracy and efficiency of these systems, which are crucial for identifying and mitigating network attacks and anomalies.

III. PROPOSED SYSTEM

The proposed methodology for building an Intrusion Detection System using Neural consists of several main stages, which will be elucidated in the following section:

A. Dataset

The KDD-99 dataset is obtained from the California University , Irvine's Machine Learning Repository, which is commonly known as the UCI ML Repository. This online collection is renowned for providing datasets extensively utilized in machine learning and data mining research. The KDD-99 dataset was initially employed in the Third International Competition for Knowledge Discovery and Data Mining Tools. The primary aim of this competition was to address data mining and knowledge discovery challenges within the field of computer science. The KDD-99 dataset is substantial, comprising 4,000,000 cases or records, and it encompasses a diverse range of categorical variables, often of a non-numeric nature. Additionally, the dataset includes numerical features. Importantly, the KDD-99 dataset contains no missing data [20].

Attacks		Attacks	
Ftp_write	R2L	Smurf	Dos
Multihop		Neptune	
Phf		Back	
Spy		Teardrop	
httptunnel		Pod	
Worm		Land	
Xlock		apache2	
Xsnoop		Mailbomb	
Named		Processtable	
Sendmail		Udpstorm	
Loadmodule	U2R	Satan	Probe
Buffer_overflow		Ipsweep	
Perl		Portsweep	
Rootkit		Nmap	
snmpgetattak		Mscan	
snmpguess		Saint	
Sqlattack		Warzclient	R2L
Xterm		Guess_passwd	
ps		Warzemaster	
Normal		Imap	

B. Preprocessing

The KDD-99 dataset stands out for its complete absence of missing values. The preprocessing stage encompasses a set of primary steps:

1. The first step involves defining attack types. In this stage, the various attack types present in the dataset are identified, contributing to the creation of a dedicated dictionary specific to attack types. Table 1 provides an attack types in the KDD-99 dataset. This categorization and definition of attack types serve as a foundational component for subsequent analysis and modeling in the context of intrusion detection.

Table 1: types of attacks

Attack Name	Attack Type
'normal'	'normal',
'back'	'dos',
'buffer_overflow'	'u2r',
'ftp_write'	'r2l',
'guess_passwd'	'r2l',
'imap'	'r2l',
'ipsweep'	'probe',
'land'	'dos',
'loadmodule'	'u2r',
'multihop'	'r2l',
'neptune'	'dos',
'nmap'	'probe',
'perl'	'u2r',
'phf'	'r2l',
'pod'	'dos',
'portsweep'	'probe',
'rootkit'	'u2r',
'satan'	'probe',
'smurf'	'dos',
'spy'	'r2l',
'teardrop'	'dos',
'warezclient'	'r2l',
'warezmaster'	'r2l'

.

In Table 1, the denial of service attack represents the most common type in the data set, as shown in Figure 3.

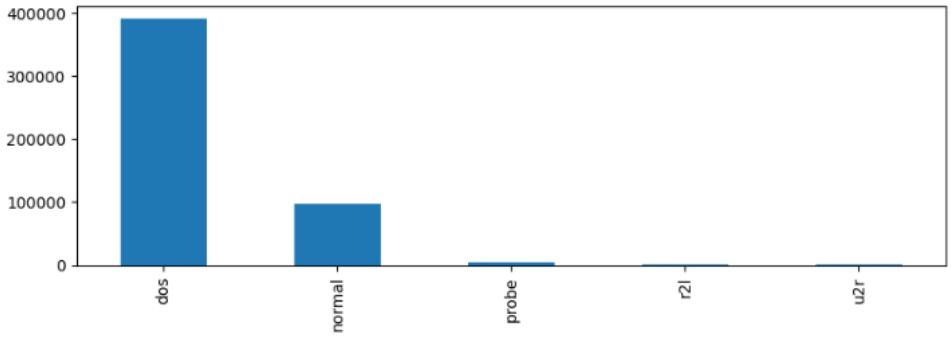

Figure 3: Distribution of attacks in the dataset.

2. Feature Reduction

Feature Reduction in machine learning using the Pearson correlation coefficient entails picking the most informative features by evaluating their association with the target variable. The Pearson correlation coefficient, symbolized by the letter "r," measures the relationship between two variables (magnitude and direction), where -1 represents a strong negative correlation, while 1 indicates a high positive correlation, in addition to a value of 0, which represents no correlation. The coefficient plays a vital role in feature selection and other machine learning tasks, facilitating the identification of significant features. Figure 4 illustrates the process of reducing features.

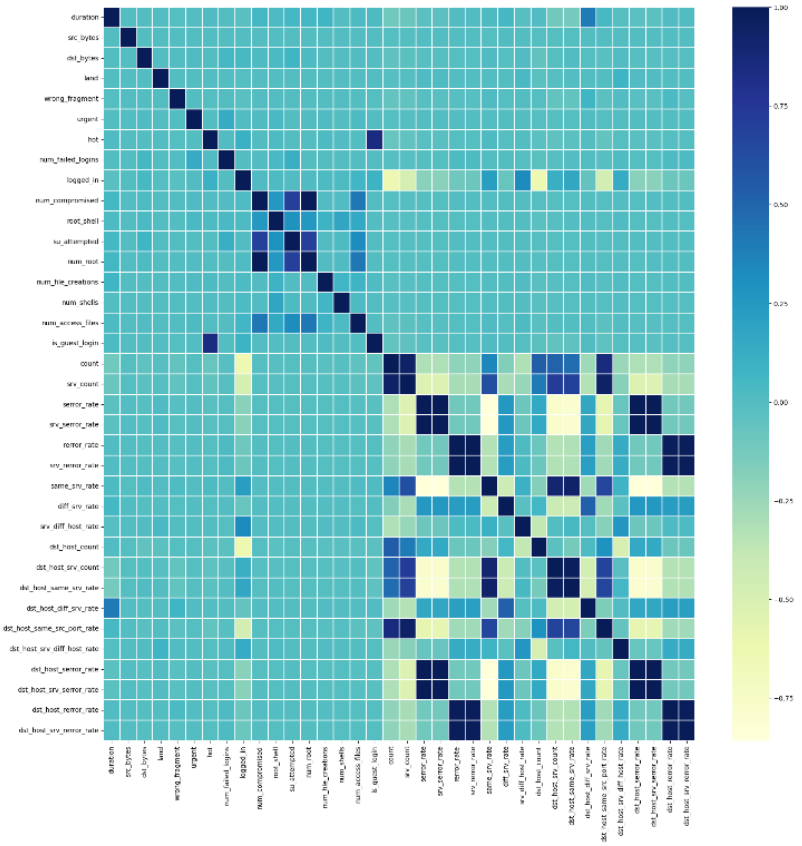

Figure 3: Pearson Correlation Coefficient

3. Protocol type encoding

In this preprocessing step, categorical data, undergo encoding using a “Label Encoder” in order to convert them into numerical representations. This step enables the use of these variables in machine learning algorithms. Tables 2,3 and figure 4 display encoded “protocol types” and ‘flag’.

Table 2

Protocol Type	Code
ICMP	0
TCP	1
UDP	2

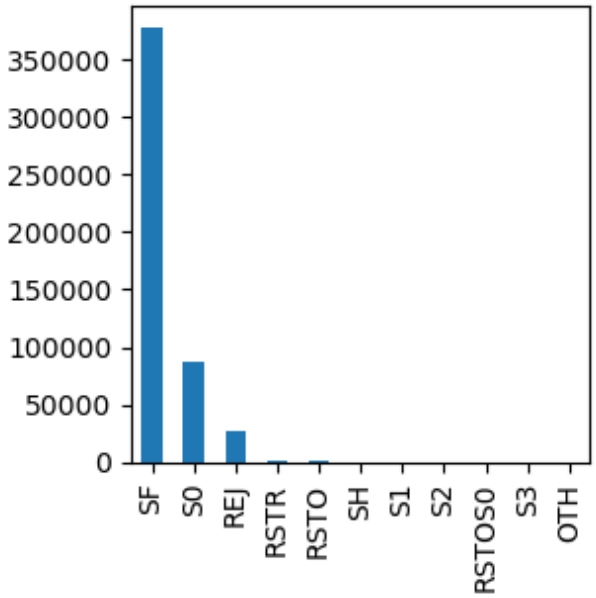

Figure 4: Flags Histogram

Table 3: the count of feature flag after encoding processing

Flag	Count	Code
SF	378440	0
S0	87007	1
REJ	26875	2
RSTR	903	3
RSTO	579	4
SH	107	5
S1	57	6
S2	24	7
RSTOS0	11	8
S3	10	9
OTH	8	10

After the preprocessing step, we note the remaining features are 32 out of 42.

C. Normalization

An additional preprocessing procedure involves min-max normalization, often referred to as feature scaling. This method entails applying a linear transformation to the data, effectively rescaling it within a range of (0, 1) [21]. The normalization process is carried out in accordance with equation (1):

$$x\ new = \frac{x - min\ (x)}{(x)\ - min\ (x)} \quad (1)$$

Where x new represent normalized x.

D. Splitting dataset

In the next step, we Convert time series data to a dataset divided into two main parts: the testing set and the training set. In this procedure, 20% of the data was allocated to the testing set, while the training set received 80%.

E. Machine learning

Machine learning, a fundamental part of AI, employs algorithms to learn from data and make informed decisions. It encompasses supervised learning, where models are trained on labeled data for predictions; unsupervised learning, which uncovers patterns in unlabeled data; and reinforcement learning, involving agent-environment interactions for optimal choices. Supervised learning includes techniques like support vector machines, decision trees, and logistic regression, while regression predicts continuous outcomes. This framework offers powerful tools for data analysis, prediction, and decision-making across various fields. Figure 6 shows the ML types [22-23].

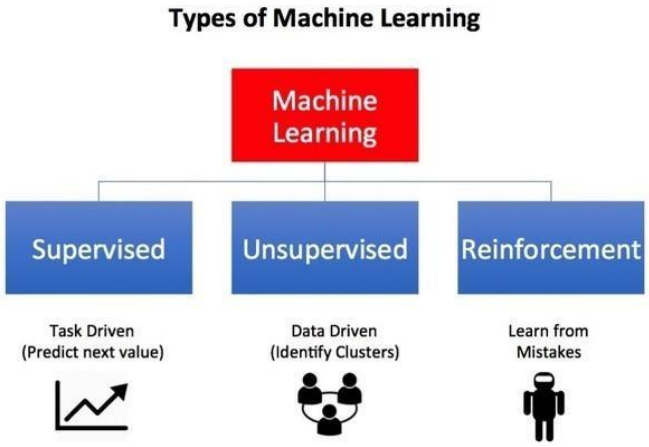

Figure 5: Type of Machine learning [22].

To perform the task of financial fraud detection and user behavior analysis, four machine learning algorithms have been selected, as elaborated in the following section.

1. Neural Network

A neural network is a formal structure consisting of nodes that represent neurons in their biological dimensions and are interconnected by arcs. Each node is assigned a weight. An artificial neural network has interconnected processing units that collaborate to process information and produce significant outcomes called layers [24]. Each layer performs a set of specific tasks, as shown below:[24], and neural network architecture is shown in figure 6.

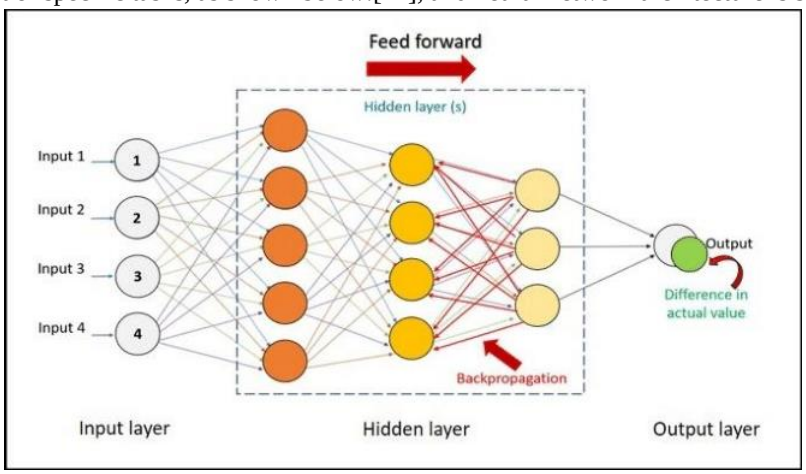

Figure 7: neural network algorithm [24]

A. The input layer receives data representing explanatory attributes for each observation. It has nodes equal to the number of input features and serves as the network's initial data receiver, duplicating and transmitting the values to the hidden layers for processing.

B. Hidden layers apply transformations to the input data using weighted connections. They can be stacked in multiple layers to capture complex patterns. The primary operation involves multiplying input values by weights, summing them, and often applying activation functions to introduce non-linearity.

C. The output layer produces final network predictions or responses. In classification, it typically has one node. The output values are derived from the processed information in the hidden layers. The network's effectiveness relies on weight selection and adjustment learned during training.

After thorough training, it can be utilized to efficiently extract patterns and classify data. Research indicates that the neural network software sector is projected to experience significant growth at a compound annual growth rate (CAGR) of 33.5% from 2019 to 2026 [25]..

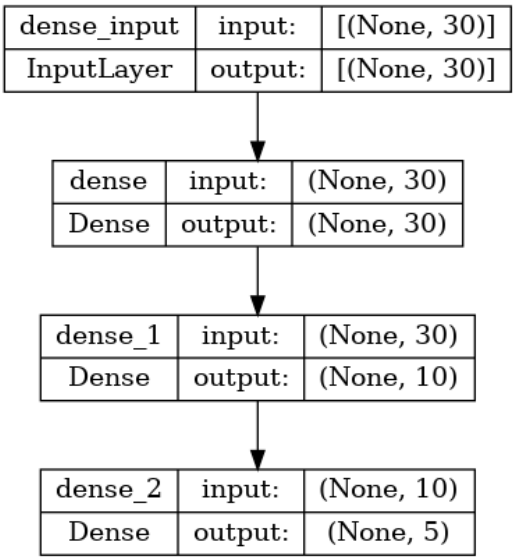

Figure 8: the proposed structure

The proposed model consists of two input and output layers with three hidden layers to perform the search task of creating an intrusion detection system using a neural network as shown in figure 8

IV. METHODOLOGY

In this section, we present the proposed methodology. In this section, we present the proposed methodology for building an artificial intelligence model that focuses on designing an intrusion detection system using a neural network. The main stages include problem definition, data collection, pre-processing, selection of relevant features for the AI model, selection of an appropriate machine learning algorithm based on data characteristics, and training and evaluation of the selected models. The research uses a neural network on the KDD-99 dataset. Model performance is evaluated using metrics of precision, precision, recall, and F1 score for a comprehensive evaluation. The proposed approach is shown in Figure 11. This methodology provides a well-organized strategy for developing an artificial intelligence model for intrusion detection using a neural network, which contributes to mitigating illegal traffic.

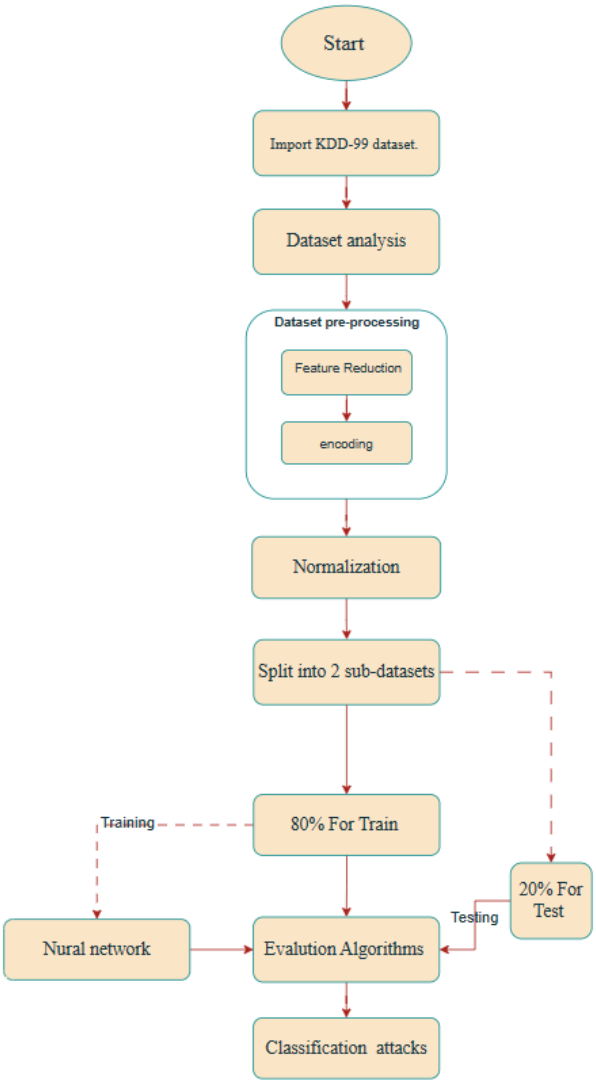

Figure 10: Block diagram of main stages for the proposed model

V. EXPERIMENTAL RESULTS

In this section, we present the implementation results of the neural network algorithm used in the proposed system. Furthermore, we apply classification metrics based on the distributions of the confusion matrix.

1. Neural network

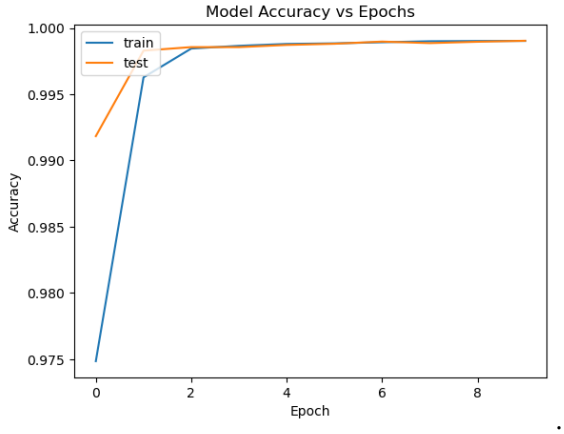

Figure 11 : The train model curve .

Figure 11 shows the training phase of the neural network, where the neural model recorded a training accuracy of 99.7% in epoch 8 of training. The results indicate the success of the proposed approach in analyzing traffic and distinguishing between normal traffic and abnormal traffic.

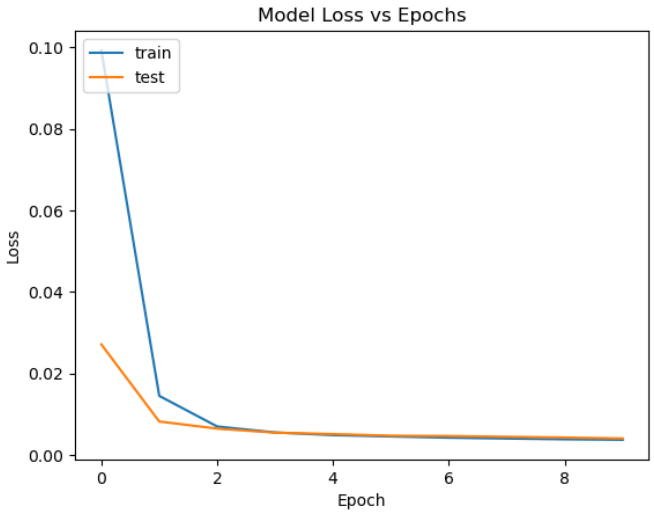

Figure 12 : The loss curve.

The loss curve shown in Figure 12 also indicates low percentages. Which indicates the detection accuracy of the proposed model.

VI. DISCUSSION

The experimental results of the proposed model indicate the accuracy of the model in analyzing traffic and accurately detecting anomalous behavior in Internet networks. The proposed model is characterized by generalization and lack of overfitting. The loss rate was very small..

VII. CONCLUSION

Network security remains a paramount concern in safeguarding sensitive information and individual privacy, especially with the increasing reliance on web services and data transmission. Intrusion Detection Systems (IDS) play a pivotal role in detecting and mitigating network attacks and breaches. In this research, we applied machine learning (ML) techniques to the KDD-99 dataset to assess their effectiveness in detecting attacks on the internet network. The proposed methodology for building an IDS using Neural Networks, based on the KDD-99 dataset offers a comprehensive and well-organized strategy. Through data preprocessing, feature reduction, normalization, and machine learning, particularly neural networks, the system can effectively analyze and distinguish between normal and abnormal network traffic. The promising results obtained during the experimental phase underscore the system's potential for enhancing network security. In a world where network attacks are ever-evolving, this approach represents a valuable tool for identifying and responding to malicious activity, contributing to the overall cybersecurity landscape.

REFERENCES

[1] R. Heady, G.F. Luger, A. Maccabe and M. Servilla, “The architecture of a Network Level Intrusion Detection System,” Department of Computer Science, College of Engineering, University of New Mexico, 1990, pp. 1-17.

[2] R. Bace and P. Mell, “NIST Special Publication on Intrusion Detection Systems,” Booz-Allen and Hamilton inc, Mclean VA, 2001, pp. 5-22. [10] R.A. Kemmerer and G. Vigna, “Intrusion Detection : A brief History and Overview,” Comp

[3] B. R. Raghunath and S. N. Mahadeo, "Network Intrusion Detection System (NIDS)," 2008 First International Conference on Emerging Trends in Engineering and Technology, Nagpur, India, 2008, pp. 1272-1277, doi: 10.1109/ICETET.2008.252..

[4] Saleh, A.I., Talaat, F.M. & Labib, L.M. A hybrid intrusion detection system (HIDS) based on prioritized k-nearest neighbors and optimized SVM classifiers. Artif Intell Rev 51, 403–443 (2019). https://doi.org/10.1007/s10462-017-9567-1

[5] Valerio, D., & da Costa, J. S. (2010). A review of tuning methods for fractional PIDs. In 4th IFAC Workshop on Fractional.

[6] Rehman, S. U., Thang, K. F., & Lai, N. S. (2019). Automated PCB identification and defect-detection system (APIDS). International Journal of Electrical, Vol. 9, No. 1, February 2019, pp. 297~306,.ISSN: 2088-8708, DOI: 10.11591/ijece.v9i1.pp297-306

[7] A. Garg and P. Maheshwari, "A hybrid intrusion detection system: A review," 2016 10th International Conference on Intelligent Systems and Control (ISCO), Coimbatore, India, 2016, pp. 1-5, doi: 10.1109/ISCO.2016.7726909.

[8] Islabudeen, M., & Kavitha Devi, M. K. (2020). A Smart Approach for Intrusion Detection and Prevention System in Mobile Ad Hoc Networks Against Security Attacks. Wireless Personal Communications, 112(1). https://doi.org/10.1007/s11277-019-07022-5.

[9] verfasst von, Giuseppe Petrosino and Federico Bergenti "An introduction to the major features of a scripting language for JADE agents," in AI* IA 2018–Advances in Artificial Intelligence: XVIIth International Conference of the Italian Association for Artificial Intelligence, Trento, Italy, November 20–23, 2018, Proceedings 17, Springer, 2018, pp. 3–14.

[10] 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

[11] Chan A, Ng W W, Yeung D S, Tsang E C et al. 2005 Proceedings of 2005 international conference on machine learning and cybernetics vol 6 pp 18–21

[12] Heba F E, Darwish A, Hassanien A E and Abraham A 2010 2010 10th international conference on intelligent systems design and applications (IEEE) pp 363–367

[13] Gao N, Gao L, Gao Q and Wang H 2014 2014 Second International Conference on Advanced Cloud and Big Data (IEEE) pp 247–252

[14] Vinayakumar R, Alazab M, Soman K, Poornachandran P, Al-Nemrat A and Venkatraman S 2019 IEEE Access 7 41525–41550

[15] Liu C, Gu Z and Wang J 2021 IEEE Access 9 75729–75740 ISSN 2169-3536

[16] Wang Z, Liu Y, He D and Chan S 2021 Computers & Security 103 102177 ISSN 0167-4048 URL https://www.sciencedirect.com/science/article/pii/S0167404821000018

[17] Shone N, Ngoc T N, Phai V D and Shi Q 2018 IEEE Transactions on Emerging Topics in Computational Intelligence 2 41–50

[18] Singh K, Kaur L and Maini R 2021 Computational Methods and Data Engineering ed Singh V, Asari V K, Kumar S and Patel R B (Singapore: Springer Singapore) pp 223–241 ISBN 978-981-15-6876-3

[19] Niyaz Q, Sun W and Javaid A Y 2016 CoRR abs/1611.07400 (Preprint 1611.07400) URL http: //arxiv.org/abs/1611.07400

[20] Hindy H, Atkinson R, Tachtatzis C, Colin J N, Bayne E and Bellekens X 2020 Electronics 9 ISSN 2079-9292 URL https://www.mdpi.com/2079-9292/9/10/1684

[21] Wu Z, Wang J, Hu L, Zhang Z and Wu H 2020 Journal of Network and Computer Applications 164 102688 ISSN 1084-8045 URL https://www.sciencedirect.com/science/article/pii/S1084804520301624

[22] Ahmad I, Basheri M, Iqbal M J and Rahim A 2018 IEEE access 6 33789–33795

[23] Negandhi P, Trivedi Y and Mangrulkar R 2019 Emerging Research in Computing, Information, Communication and Applications ed Shetty N R, Patnaik L M, Nagaraj H C, Hamsavath P N and Nalini N (Singapore: Springer Singapore) pp 519–531 ISBN 978-981-13-6001-5

[24] Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F., Abdullah, H., & Saidur, R. (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renewable and Sustainable Energy Reviews, 33, 102-109..

[25] Levin, I. (2000). KDD-99 classifier learning contest LLSoft's results overview. ACM SIGKDD Explorations Newsletter, 1(2), 67–75. https://doi.org/10.1145/846183.846201.

[26] Ileberi .E., Sun. Y., and Wang .Z, "A machine learning based credit card fraud detection using the GA algorithm for feature selection," J. Big Data, vol. 9, no. 1, pp. 1–17, 2022.