Predictive Cyber Attack Detection in Cloud Environments with Machine Learning from the CICIDS 2018 Dataset

Mani Gopalsamy
Senior Cyber Security Specialist, Louisville, KY, USA- 40220,

Abstract- *The exponential rise of computer network assaults and network applications throughout the world has occurred with an increase in cyberattacks. Datasets like CSE-CIC-IDS2018 were developed in order to train prediction models for network-based intrusion detection precisely because of this. The primary goal of these datasets is to promote research on anomaly-based identification using various ML techniques, rather than to store signature-based detection structures. With data acquired over 10 days, CSE-CIC-IDS2018 comprises around 16,000,000 occurrences. Three phases make up the investigation: preliminary data cleaning, exploratory data analysis, and data normalisation techniques. This study evaluates the performance of the Multi-Layer Perceptron Backpropagation (MLP-BP) model in classification tasks. The MLP-BP model demonstrates exceptional accuracy at 98.97%, underscoring its high overall correctness in classifying instances. Precision reaches 99%, reflecting the model's reliability in positive predictions, while recall is 98%, indicating its effectiveness in identifying actual positives. An excellent balance between recall and precision is shown by the model's F1-score of 99.38%, which confirms its remarkable overall performance.*

Keywords- Predictive Cyber Attack Detection, Cloud Environments, Machine Learning, CICIDS 2018 Dataset, Cybersecurity, Intrusion Detection, Anomaly Detection.

I. INTRODUCTION

The IoT and cloud computing have become a revolutionary paradigm that has impacted many industries, including education, healthcare, and military applications. Its inherent cost-effectiveness and outstanding dependability, which gave businesses the freedom to grow their operations with never-before-seen flexibility, were the source of its attractiveness. However, as cloud infrastructure became more and more reliant upon it, the menacing and constant danger of cyberattacks surfaced [1]. These evil attacks on digital infrastructure cause disruptions to regular system functions and carry out harmful actions that jeopardise the availability, integrity, confidentiality, and privacy of data. It is now crucial to strengthen cloud network security in response to this expanding threat[2][3]. The detection system's design was developed in recognition of how urgent it was to protect these systems against intrusions [4][5][6].

The exponential expansion and quick development of IT technology have made cyber-attacks a serious danger to network telecommunications. Most cyberattacks are comprised of software designed to breach network security in order to undermine security[7]. Malware attacks often include the introduction of a dangerous external element into a protected network, meaning that the attack starts outside a network's perimeter protection. Viruses, worms, and trojan horses are all examples of malware attack tools [8]. The victim machine suffers damage as a result of the security breach, which can take many forms, including data corruption, phishing for sensitive information, and denial of service (DoS) attacks, which involve overwhelming network resources with traffic in order to prevent users from accessing the network's services[9].

Conversely, cyber-attacks might originate from an inside network resource; this kind of assault is also called an insider threat [10]. Cyberattacks perpetrated by authorised users are the most prevalent kind of insider attack. Alternately said, a legitimate user is adept at breaking security protocols, which gives them the ability to swiftly conduct criminality and access network resources[9]. Cyberattacks and the need for efficient detection methods to discover them are the primary concerns of the study. The article delves into several machine learning algorithm-based detection methods for distinct cyber-attack kinds. Some of the parameters used to compare the algorithms were accuracy, FPR, FNR, performance, and datasets[11].

Researchers studying cloud environments have shown a growing interest in using ML methods. Using the tagged traces from the training models, the ML-based security system classifies the aberrant and healthy behaviours[2]. ML in cybersecurity is a formidable technique that enhances systems' capacity to comprehend a variety of patterns and

anticipate possible data risks by extracting distinct feature sets[12][13]. Models that effectively defend systems from suspicious and spyware-related behaviour are generated by optimising processing and training methods[14]. This revolutionary technique enables systems to learn from and adjust to data, enabling them to make wise judgements without explicit programming [15]. ML algorithms make use of both real-time and historical data to find patterns in typical behaviour as well as anomalies that can point to security risks. These algorithms learn to identify novel and emerging attack vectors by training on a variety of datasets. ML improves IDS by enabling quicker and more accurate threat detection, decreasing false positives, and adjusting to changing threats[16][17].

The abundance of samples found in incursion databases is one driving force behind this effort. As an example, there are about 1 million samples in the CICIDS 2018 dataset. There is a finite amount of attack sites that can be effectively masked, therefore most datasets consist of benign traffic. Attack samples, on the other hand, are in the minority. The sole exception to this rule is DoS assaults. Therefore, the IDS become more complicated as a consequence of the high quantity of benign samples, since the training overhead of the ML models is raised.

A. Contribution of the study

There are numerous contributions of this study is multifaceted and impactful in the domain of cyber-attack detection. Here's a detailed summary of the key contributions:

- To enhances the model's efficiency and effectiveness. This analysis allows for the simplification of the model, potentially improving its performance and interpretability.
- Multilayer Perceptron with Backpropagation (MLP-BP), LSTM, AdaBoost, and DAE-DNN. An advantages and disadvantages of each strategy are highlighted in this comparison study, which offers insightful information about how successful each is in detecting cyberattacks.
- To evaluation metrics are F1-score, recall, accuracy, and precision to assess model performance. clear understanding of how well each model performs and ensures that the findings are reliable and actionable.
- To achieves an exceptional accuracy of 98.97%. This high accuracy is indicative of the robustness and effectiveness of proposed approach in distinguishing between benign and malicious network traffic.
- Advanced techniques like AdaBoost and DAE-DNN demonstrates the study's commitment to leveraging advance methods for improving cyber-attack detection. These techniques contribute to the accuracy and robustness of the detection models.

B. Structure of paper

This is the outline for the remaining sections of the paper. Section II requires a literature study on the topic of cyber-attack detection in cloud environments using ML. Analysis and discussion of the findings follow in Section IV, while Section III explains the procedures and methodologies used. The study's conclusion and plans for further research are detailed in Section V.

II. LITERATURE REVIEW

In this section, provide some previous work on cyber-attack prediction based on machine learning. Some authors have attempted to forecast cyberattacks in cloud environments by combining DL and ML methods. It was explained how to employ in order to identify cyberattacks.

In, Nocera et al., (2022) The suggested methodology for this work, which is based on ML and anomaly detection techniques, emphasises how the use of network IDS has grown to be a crucial area of cyber security and how DL approaches outperform more conventional methods of pattern recognition and anomaly detection, such as SVM or DT, in terms of precision and accuracy in identifying and isolating this kind of malicious attacks[18].

In, A et al. (2024) An IDS examines the patterns of data transmission on a network. Autoencoder features are used to assess classifiers from a variety of ML algorithms, like K-NN, RF, GB, DT, LR, and SVM. The K-NN classifier predicts cyberattacks with an astounding 99.28% accuracy rate, outperforming other models in the process. In addition, the suggested method assigns performance ratings to each classifier and assesses how well various kinds of ML classifiers identify network assaults in the NSL_KDD dataset[19].

In, Kharlamova, Traeholt and Hashemi, (2023) offers a ML-based cyberattack detection technology that helps lessen a BESS's susceptibility to online assaults. It forecasts the condition of charge and identifies possibly damaged data via Adaptive Boosting. The example of the simulated dataset illustrates the advantages of using the unique cyberattack detection technique. Cyberattacks have the potential to impair BESSs' capacity to respond to system demands in a sufficient and dependable manner[20].

In, Kannan et al. (2024) demonstrates the multiple DL methods to predict the anomalies and other potential threats with more accuracy in real time. Nowadays, cyber-attacks are growing predominantly due to the development of technologies. The AI, ML and DL techniques leverage enormous amounts of data to identify the cyber-attacks[21].

In, Madala et al. (2023) suggest that automated AI research has great potential to enhance cyber-security and prevent cyber-attacks by predicting cyberattacks and preparing security measures. Our results demonstrate how accurate ML techniques are in predicting cyberattacks, particularly when it comes to RF and NN models[22].

In, Swaminathan et al. (2022) aid in the proactive prevention of risks and the timely response to harmful activity by cyber security personnel. This study employs supervision of ML techniques to examine cybercrime in four different models and forecast the impact of the stated characteristics only on identifying the threat method and the offender. They will examine the effectiveness of three ML algorithms: KNN, RF, and LR[23].

In, Deepak et al. (2023) Cyberattack detection and prevention are difficult tasks, but new developments in AI-based security models and prediction tools have made these problems easier to address. ML approaches to use two models to analyse cybercrime and forecast how the qualities may help identify the criminal and the cyberattack strategy. Results from testing several cyber-attack strategies showed that the SVM linear model had the best accuracy rate. An important takeaway from the first model is the variety of assaults that victims may expect to encounter. There is a positive correlation between education and money and the probability of becoming a victim of cyberattacks, according to our research[24].

According to the above-mentioned literature evaluations, substantial research has been conducted on autonomous cyber-attack detection, as indicated in Table 1.

Table. 1 Comparative Table for Literature Review

Reference	**Methodology**	**Performance**	**Limitations & Future Work**
[18]	Machine Learning & Anomaly Detection (Deep Learning vs. SVM/DT)	Deep Learning outperforms SVM/DT	No detailed limitations are provided; further evaluation of other models needed.
[19]	Autoencoder Deep Learning Model; Evaluates classifiers: K-NN, Random Forest, Gradient Boosting, etc.	K-NN: 99.28% Accuracy	Need to explore scalability and real-time performance.
[20]	Adaptive Boosting for Cyberattack Detection in Battery Energy Storage Systems (BESS)	Effective for BESS vulnerability	Applicability to real-world scenarios; integration with IoT technologies.
[21]	Deep Learning methods for anomaly detection and prediction of threats	High accuracy in real-time	Exploration of specific DL methods and their comparative performance.
[22]	AI research for cyber-attack prediction; Random Forest & Neural Networks highlighted	High accuracy with RF & Neural Nets	Further model comparisons and fine-tuning are required.
[23]	Supervised ML methods (Logistic Regression, Random Forest, K-NN) for cybercrime investigation	Comparative efficacy of ML models	Need for more detailed evaluation and real-world applicability.
[24]	AI-based models for detecting and preventing cyberattacks; SVM linear model highlighted	Highest accuracy with SVM	Further testing on diverse attack methods; explore socio-economic factors.

III. METHODOLOGY

The methodology for cyber-attack detection using the CICIDS 2018 dataset follows a structured process to achieve high accuracy. First, the data is collected from a CICIDS 2018 dataset and undergoes data preprocessing, which includes feature reduction, zero variance feature removal, and timestamp removal to eliminate unnecessary data and enhance the model's performance. Important features are then identified to focus on those most relevant to the detection process. Next, data normalisation is applied using Z-Score normalisation techniques to ensure consistency in data scaling. Data splitting into train and test with 80:20 ratio. The pre-processed and normalised data is then used to train model: Multilayer Perceptron with Backpropagation (MLP-BP) a training, the models are evaluated based on metrics like Precision, accuracy, F1-score, and recall. Also, MLP-BP model compare with LSTM, AdaBoost, DEA-DNN models. In terms of cyberattack detection, both models show excellent predictive performance. The final result reflects the robustness of the approach in distinguishing between benign and malicious network traffic. Figure 1 below depicts the flowchart of the methodological system used to identify cyberattacks.

Detailed descriptions of the following phases are provided in the flowchart diagram:

A. Data Collection

Data collection is a process of gathering information from many sources and then evaluating it to uncover patterns and viable strategies for investigating problem areas[25]. This comparison analysis used the CIC-IDS2018 dataset, which includes the attack classifications, individual outcomes for various symptoms, and fundamental cyberattack information. A collaborative endeavour between CSE and CIC was suggested in 2018, and the result is the CSE-CIC-IDS2018 dataset. You may get the dataset via AWS.

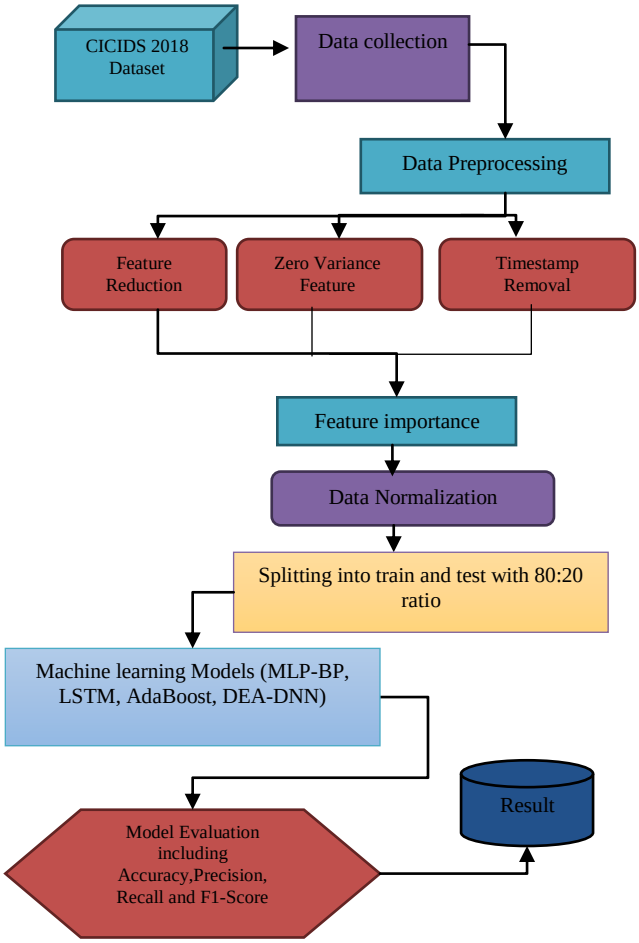

Fig. 1 Methodology flow chart for cyber-attack detection

B. Data Preprocessing

Removing unwanted data from a dataset is referred to as data preprocessing. This meant deleting the first four features—"Src IP," "Src Port," "Flow ID," and "Dst IP"—to reduce it to 80 features. Selected for their significance were both datasets with different feature counts. The dataset was homogenised to 80 features and then stripped of the specified baseline characteristics. Then, the eliminated features from both days were aggregated to create a homogeneous dataset for further analysis. For all 80 attributes—including the labels—the study included finding the lowest, maximum, standard deviation, and mean values of the data. For every instance, we eliminated ten fields that had constant zero values. Furthermore, in order to ensure that learners cannot differentiate between the two tasks of attack prediction and attack detection, we eliminated the "Timestamp" columns. Once the extraneous features have been removed. The results of the tests will include 69 attributes and 8,997,323 rows of data. Data quality was ensured using a series of procedures. After these cleaning processes were finished, the dataset was narrowed to 6,634,943 rows, which is suitable for further study and usage.

C. Feature importance

The effectiveness of the ML model on any particular task may be improved by a variety of inputs. A family of methods known as "Feature Importance" is used to rate the importance of various cyber-attack factors by giving values to the characteristics that are fed into a prediction model. To get a feel for the model, look at the feature significance ratings. where less important features can be eliminated to simplify the model and potentially improve its performance shown in Figure 2.

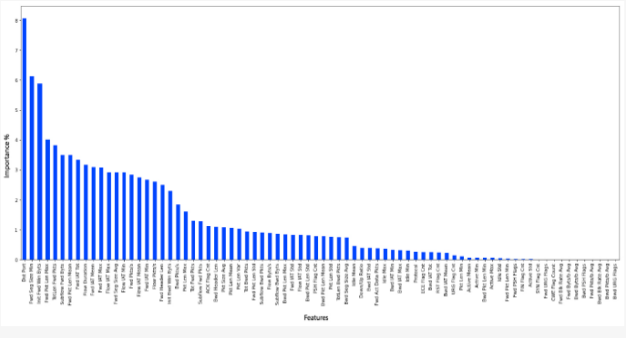

Fig. 2 Importance of feature

D. Data Normalization

Data preparation in ML includes normalisation, which ensures that numerical column values are uniform and on a constant scale. Standardising data to a specific scale, commonly ranging from 0 to 1, is an essential step in data preparation for NIDS. Cybersecurity dataset quirks dictate the Z-score normalisation used in the CSE-CIC-IDS-2018 analysis.

$$x' = \frac{x - mean}{std} \qquad (1)$$

Z-score Normalization: This technique sets the mean and standard deviation of a feature's values to zero and one, respectively. This is achieved by taking a mean of a feature out of each value and splitting the result by the standard deviation. The strategy's mathematical formula, (1), denotes the original value X and the normalised value X′.

E. Data splitting

The pre-processed and normalised dataset is divided into training and testing sets. The training dataset contains 80% of data, while the testing dataset has 20%.

F. Classification Models

Some categorisation models for the prediction of cyber-attack detection are described in this section. These models are used for comparative analysis.

1) MLP-BP

Multi-Layer Perceptron (MLP-BP) is a widely used classification method for cyber-attack detection, especially with datasets like CIC-IDS2018. It works by processing input data through multiple hidden layers, where each layer's neurons compute weighted sums of the inputs, followed by an activation function, typically ReLU. This helps a model learn complex patterns in network traffic that distinguish normal traffic from malicious attacks. A single neuron in the output layer, the last layer, outputs a probability between 0 and 1 using the sigmoid activation function. A threshold of 0.5 is applied, where values above 0.5 indicate a potential attack, and values below 0.5 classify the traffic as normal. Binary cross-entropy, which measures a discrepancy between actual labels and the projected probability (0 for normal, 1 for assault), is used as the loss function during model training. This process allows the MLP model to effectively classify and detect cyber-attacks based on network traffic patterns. Figure 3 shows the architecture of MLP.

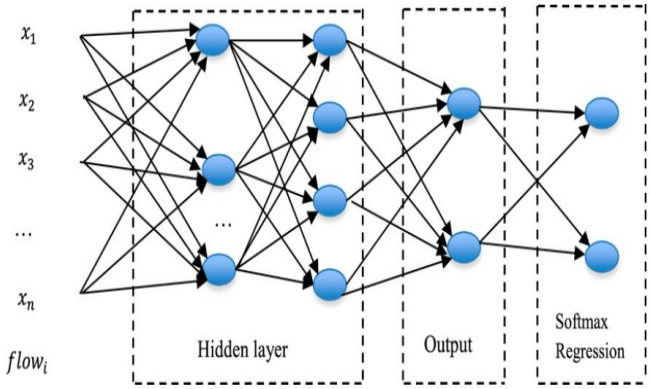

Fig. 3 The architecture of MLP

2) LSTM

The LSTM structure is a recently proposed RNN that aims to solve the problem of long-term reliance. With the addition of the forget gate, input gate, and output gate, it enhances the standard RNN. While the input gate adds new data to the neural network, the forget gate removes irrelevant data, and the output gate determines the current node's output. A single LSTM cell's construction is shown in Figure 4[26].

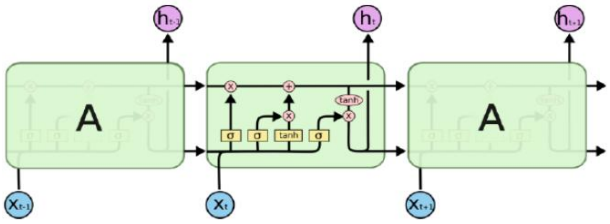

Fig. 4 Structure Diagram of a Single LSTM Cell

A most crucial component of a system is the classification engine, which we built using LSTM [27][28]. LSTM is capable of learning not only the features of the present network traffic but also the characteristics of the network traffic from the past. In most cases, while attacking a network, the perpetrators will launch a succession of simultaneous activities. That being the case, the present network traffic is either unrelated to or otherwise typical of the past[29].

3) AdaBoost

A key component of ada boosting methods is constructing a model using the training dataset, and then constructing a second model to address any faults introduced by the first model [30]. This process will be repeated until the mistakes are reduced and the dataset can be accurately forecasted. The key strength of AdaBoost lies in combining weak classifiers, such as decision stumps, Naive Bayes, or logistic regression, which individually perform only slightly better than random guessing[31]. This enables the model to learn from difficult cases. To create a strong model for attack detection, AdaBoost finally uses a weighted majority vote to combine the predictions of these classifiers. To guarantee effective performance, AdaBoost models are assessed using metrics including recall, accuracy, precision, and FPR.

4) DAE-DNN

The DAE-DNN (Denoising Autoencoder - Deep Neural Network) model is highly effective for cyber-attack detection by combining robust feature extraction and classification capabilities[32]. In this architecture, the Denoising Autoencoder (DAE) first processes raw, noisy input data, like network traffic or system logs, and compresses it into a lower-dimensional representation while filtering out irrelevant noise[33]. This step helps in isolating key features related to attack patterns. The compressed, clean data is then fed into the DNN, which, through multiple hidden layers, learns complex relationships among these features to classify various types of cyber-attacks or detect anomalies. The model is trained in two phases: the DAE is pre-trained to denoise and compress the data, followed by DNN training for attack classification[34][35]. This approach is particularly effective for detecting sophisticated attacks in real-time, improving accuracy while minimising false positives and negatives.

G. Performance matrix Model Evaluation

To choose the assessment meter and evaluate the model more effectively, it is important to comprehend how each metric measures. Comparing the effectiveness of ML algorithms was the goal, and all of these performance metrics—such as accuracy score, recall, F1score, and precision—were assessed.

1) Accuracy

Accuracy is the proportion of cases in a model that is correctly categorised and the total error in class prediction. A classifier that mostly predicts the majority class may be accurate yet misclassify minority class occurrences, calculated as Equ.2.

$$Acuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (2)$$

2) Precision

When all the data is classified into a class, precision is the percentage of cases that are correctly assigned to that class. It is calculated using the one-vs-all method for each class, calculated as Equ.3:

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

3) Recall

Number of instances properly sorted into a class is assessed by recall or sensitivity. Recall is calculated using the one-vs-all approach, like precision calculate as Equ.4

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

4) F1-score

The F1-score or F-measure is a weighted harmonic mean of recall and precision. Taking a broader perspective allows for a more comprehensive assessment, calculate as Equ.5.

$$F1-score = 2 * \frac{precision * recall}{precision + recall} \quad (5)$$

Where,

- **TP (True Positive):** The number of cases in which the model forecasted the positive class with accuracy.
- **TN (True Negative):** The figure of occurrences when a model accurately forecasted a negative class.
- **FP (False Positive):** Count of positive class predictions that were wrongly made by the model
- **FN (False Negative):** The total number of times the model got the negative class prediction wrong.

IV. RESULTS AND DISCUSSION

The simulated results of cyber-attack detection prediction based on ML techniques in a Python environment are discussed in this section. Results, dataset description, and classifier statistics are all part of this section, which displays the outcomes of the dataset assessment that was conducted for this research.

A. Dataset Description

The research uses an open-source tool that includes 10 CSV files, each representing a day of the collected network activity, and over 16.2 million data. In addition, the CIC Flow Meter tool was able to successfully extract over 80 characteristics. The six main forms of intrusion assaults included in this dataset are online attacks, bot, brute force, DDoS, and infiltration (Figure 5).

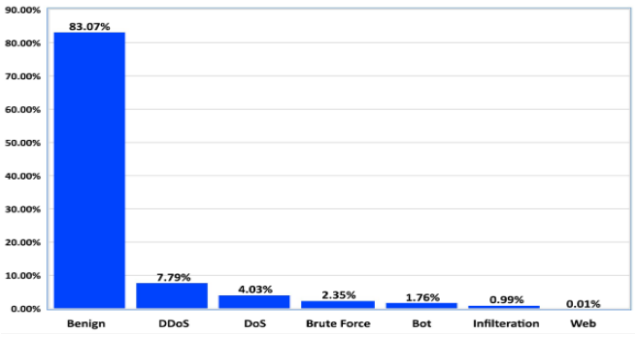

Fig. 5 Distribution of attack classes in CSE-CIC-IDS2018.

Figure 5 shows a percentage distribution of different forms of network traffic and assaults. The majority of traffic, 83.07%, is classified as benign. The second highest percentage is DDoS attacks, accounting for 7.79%. DoS attacks make up 4.03%, followed by Brute Force attacks at 2.35%, Bot attacks at 1.76%, and Infiltration at 0.99%. Web attacks are the least common, representing only 0.01% of the traffic. This chart highlights that benign traffic significantly outweighs all other types of malicious activities.

B. Experiment results

This section poses the findings of the MLP-BP model applied to a large dataset for predicting cyber-attack detection using machine learning.

Table. 2 MLP-BP model performance matrices for cyber selection detection prediction

Model	Accuracy	Precision	Recall	F1-score
MLP-BP	98.97	99	98	99.38

There are many performance measures that are shown in Table 2, the MLP-BP model's performance metrics, with an accuracy of 98.97%, indicating high overall correctness. Its precision rating is 99% (meaning that most positive predictions are correct) and recall rate is 98% (indicating that it effectively identifies real positives). An F1-score of 99.38% indicates that the model has an exceptional balance between recall and precision, which is indicative of its outstanding classification performance. The bar graph of this performance is shown in Figure 6.

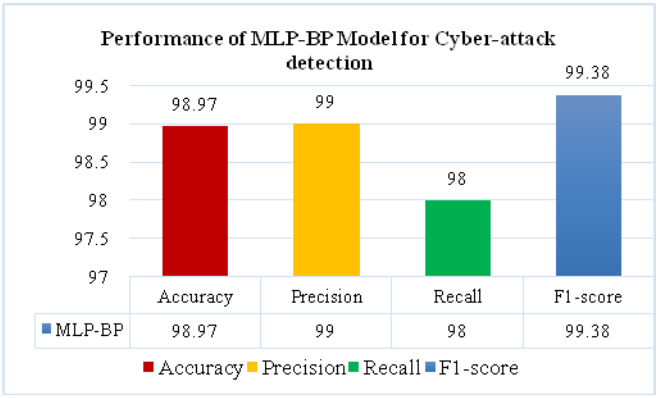

Fig. 6 Bar graph of parameters Performance of MLP-BP model cyber-attack detection prediction.

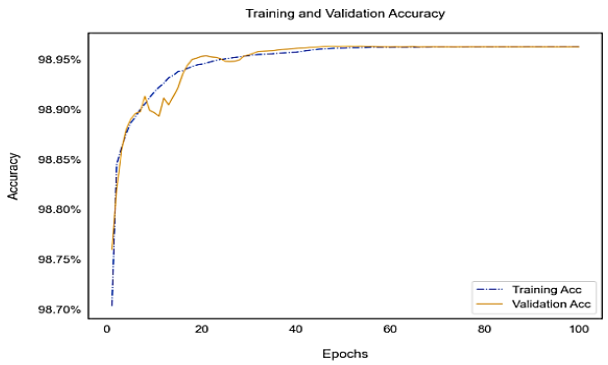

Fig. 7 Accuracy of MLP-BP for binary classification

The accuracy of an ML model's training and validation across 100 epochs is shown in Figure 7. Both the training (blue dashed line) and validation accuracy (orange line) steadily increase, converging around **98.97%**. The model

shows strong performance early on, with accuracy stabilising after 20 epochs and no signs of overfitting, as both training and validation curves remain closely aligned throughout. The final accuracy for both training and validation is **98.97%**, indicating excellent generalisation to unseen data.

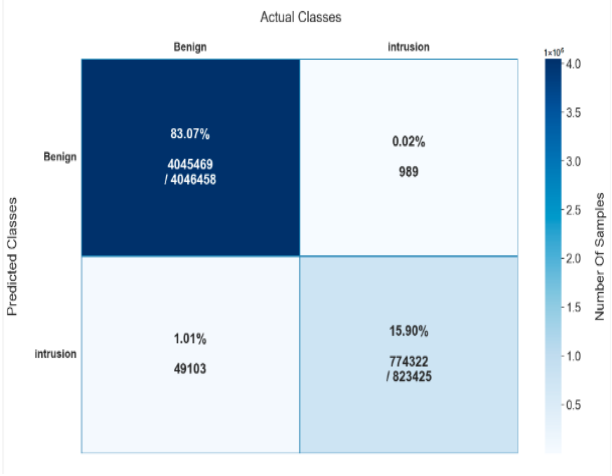

Fig. 8 Confusion matrix of MLP-BP for binary classification.

Figure 8, shows a model's performance in classifying **benign** and **intrusion** samples. It correctly classified 83.07% of benign and 15.90% of intrusion samples, with very few misclassifications (1.01% FN and 0.02% FP). The model achieved a high accuracy **98.97%**, a **precision 99.98%**, and an **F1-score 99.38%**, demonstrating strong performance in identifying both benign and malicious network activity with minimal errors.

C. Comparative analysis and Discussion

A comparative comparison using different models for cyber-attack detection. In terms of performance measures, the following Table 3 compares and contrasts different DL and ML models utilised for prediction cyber-attack detection.

Table. 3 Comparison between various model for cyber-attack detection

Models	**MLP-BP**	**LSTM**[29]	**AdaBoost**[36]	**DEA-DNN** [34]
Accuracy	98.97	96.20	86.20	95.79
precision	99	96	96	95
Recall	98	96	88	95
F1-Score	99.38	-	-	95

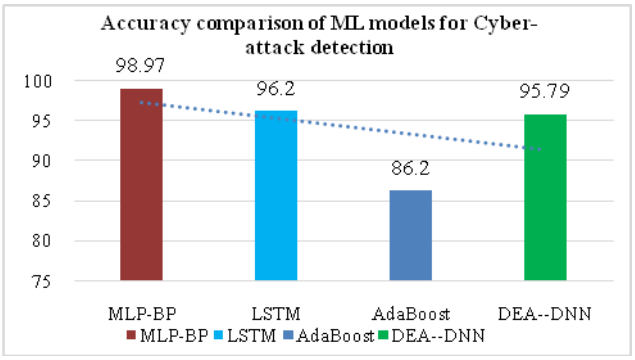

Fig. 9 Accuracy graph of comparative models for cyber-attack detection.

Figure 9, displayed a compares the accuracy of four machine learning models (MLP-BP, LSTM, AdaBoost, and DEA-DNN) in detecting cyber-attacks. MLP-BP achieved the highest accuracy at 98.97%, followed by LSTM and AdaBoost. DEA-DNN had the lowest accuracy at 95.79%.

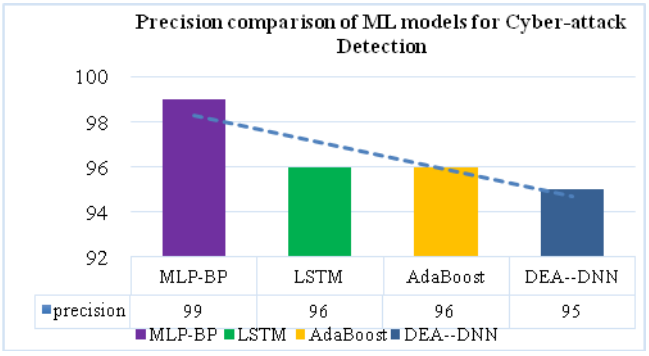

Fig. 10 Comparative models graph of precision for cyber-attack detection.

Figure 10 shows the precision of four machine learning models (MLP-BP, LSTM, AdaBoost, and DEA-DNN) in detecting cyber-attacks. LSTM and AdaBoost both reached 96% precision, while MLP-BP reached 99%. DEA-DNN's 95% precision rate was the lowest. It seems that MLP-BP outperforms the other models when it comes to properly recognising cyber-attacks. On the other hand, DEA-DNN might perhaps make some inaccurate predictions.

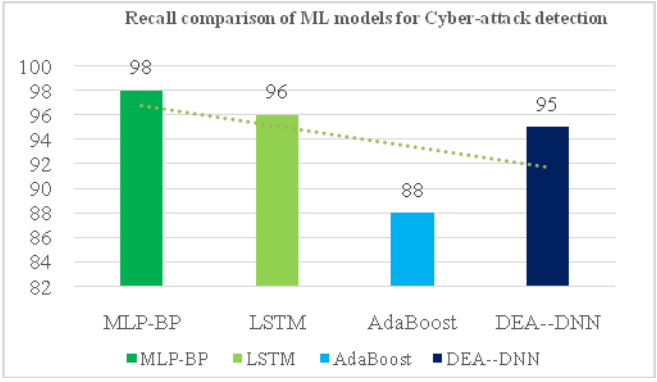

Fig. 11 Comparative models graph of Recall for cyber-attack detection.

Figure 11, compares the recall of four machine learning models (MLP-BP, LSTM, AdaBoost, and DEA-DNN) in detecting cyber-attacks. MLP-BP achieved the highest recall at 98%, followed by LSTM at 96%. AdaBoost and DEA-DNN had lower recalls at 88% and 95%, respectively. This suggests that MLP-BP is the most effective model in identifying all instances of cyber-attacks, while DEA-DNN may miss some attacks.

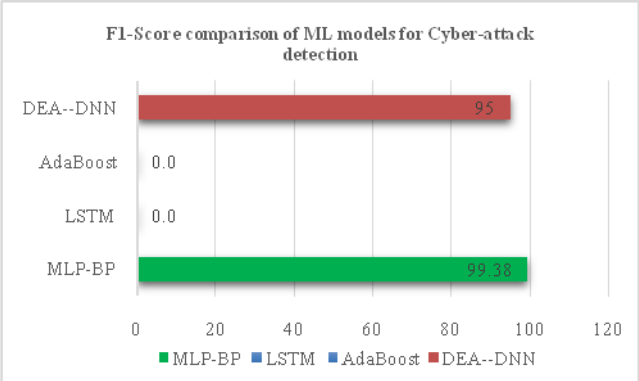

Fig. 12 Comparative Model graph of F1-Score

Figure 12, F1-score of four machine learning models (MLP-BP, LSTM, AdaBoost, and DEA-DNN) in predicting COVID-19 cases. MLP-BP achieved a highest F1-score at 99.38%, followed by DEA-DNN at 95%. LSTM and AdaBoost had very low F1scores of 0.0. This suggests that MLP-BP is a most effective model for cyber-attack detection.

V. CONCLUSION AND FUTURE SCOPE

This paper uses the CICIDS 2018 dataset to propose a robust approach for predicting cyber-attack detection in cloud systems. Various machine learning models are applied to obtain high predicted accuracy. The best-performing model was the Multilayer Perceptron with Backpropagation (MLP-BP) model, which had an F1-score of 99.38%, 99% precision, 98.97% accuracy, and 98% recall. MLP-BP outperformed other models, including LSTM, AdaBoost, and DEA-DNN, in comparison, especially when it came to correctly categorising both benign and harmful network data. The suggested strategy's efficacy is substantiated by the superb recall-to-precision ratio and steady performance across several criteria. Anoutcomeimplies that an MLP-BP model is an effective instrument for immediately identifying cyberattacks, enhancing cloud environment security. Although there are still some lingering problems with data processing and labelling, the use of ML has resulted in noticeable advancements for all of these tasks. Attacks against computer networks and the apps that run on them have been on the rise in recent years. As a result, many intrusion detection datasets, such CICIDS2018, have been created to train classification algorithms. Due to the many SCADA system vulnerabilities and the wide range of assaults, the mission of traditional IDSs will grow more difficult in the future. Inadequate cyberattack detection in vital systems may have detrimental effects on the economy and public safety. The detecting potential of DL and ML to help IDS is explored in this paper.

REFERENCES

[1] N. Moustafa, N. Koroniotis, M. Keshk, A. Y. Zomaya, and Z. Tari, “Explainable Intrusion Detection for Cyber Defences in the Internet of Things: Opportunities and Solutions,” *IEEE Commun. Surv. Tutorials*, 2023, doi: 10.1109/COMST.2023.3280465.

[2] G. Sreelatha, A. V. Babu, and D. Midhunchakkarvarthy, “A Survey on Cloud Attack Detection using Machine Learning Techniques,” *Int. J. Comput. Appl.*, 2020, doi: 10.5120/ijca2020920887.

[3] V. V. Kumar, A. Sahoo, and F. W. Liou, “Cyber-enabled product lifecycle management: A multi-agent framework,” in *Procedia Manufacturing*, 2019. doi: 10.1016/j.promfg.2020.01.247.

[4] D. Ramachandran, M. Albathan, A. Hussain, and Q. Abbas, “Enhancing Cloud-Based Security: A Novel Approach for Efficient Cyber-Threat Detection Using GSCSO-IHNN Model,” *Systems*, 2023, doi: 10.3390/systems11100518.

[5] B. H. Awaji *et al.*, “Novel multiple access protocols against Q-learning-based tunnel monitoring using flying ad hoc networks,” *Wirel. Networks*, vol. 30, no. 2, pp. 987–1011, Feb. 2024, doi: 10.1007/s11276-023-03534-y.

[6] H. Sinha, “ANALYZING MOVIE REVIEW SENTIMENTS ADVANCED MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING METHODS,” *Int. Res. J. Mod. Eng. Technol. Sci. (*, vol. 06, no. 08, pp. 1326–1337, 2024.

[7] J. Jang-Jaccard and S. Nepal, “A survey of emerging threats in cybersecurity,” in *Journal of Computer and System Sciences*, 2014. doi: 10.1016/j.jcss.2014.02.005.

[8] E. Mbunge, B. Muchemwa, J. Batani, and N. Mbuyisa, “A review of deep learning models to detect malware in Android applications,” *Cyber Security and Applications*. 2023. doi: 10.1016/j.csa.2023.100014.

[9] A. Alshehri, N. Khan, A. Alowayr, and M. Y. Alghamdi, “Cyberattack Detection Framework Using Machine Learning and User Behavior Analytics,” *Comput. Syst. Sci. Eng.*, 2023, doi: 10.32604/csse.2023.026526.

[10] F. Yuan, Y. Cao, Y. Shang, Y. Liu, J. Tan, and B. Fang, “Insider threat detection with deep neural network,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes*

in Bioinformatics), 2018. doi: 10.1007/978-3-319-93698-7_4.

[11] S. Saini and D. A. Kalia, "Detection of Cyber Attacks using Machine Learning," *Int. J. Res. Appl. Sci. Eng. Technol.*, 2023, doi: 10.22214/ijraset.2023.55918.

[12] I. H. Sarker, A. S. M. Kayes, S. Badsha, H. Alqahtani, P. Watters, and A. Ng, "Cybersecurity data science: an overview from machine learning perspective," *J. Big Data*, 2020, doi: 10.1186/s40537-020-00318-5.

[13] R. P. Vamsi Krishna Yarlagadda, "Secure Programming with SAS: Mitigating Risks and Protecting Data Integrity," *Eng. Int.*, vol. 6, no. 2, pp. 211–222, 2018.

[14] F. Hussain, R. Hussain, S. A. Hassan, and E. Hossain, "Machine Learning in IoT Security: Current Solutions and Future Challenges," *IEEE Commun. Surv. Tutorials*, 2020, doi: 10.1109/COMST.2020.2986444.

[15] S. Mishra and A. K. Tyagi, "The Role of Machine Learning Techniques in Internet of Things-Based Cloud Applications," in *Internet of Things*, 2022. doi: 10.1007/978-3-030-87059-1_4.

[16] P. L. S. Jayalaxmi, R. Saha, G. Kumar, M. Conti, and T. H. Kim, "Machine and Deep Learning Solutions for Intrusion Detection and Prevention in IoTs: A Survey," *IEEE Access*, 2022, doi: 10.1109/ACCESS.2022.3220622.

[17] Z. Chen, M. Simsek, B. Kantarci, M. Bagheri, and P. Djukic, "Machine learning-enabled hybrid intrusion detection system with host data transformation and an advanced two-stage classifier," *Comput. Networks*, vol. 250, no. November 2023, p. 110576, 2024, doi: 10.1016/j.comnet.2024.110576.

[18] F. Nocera *et al.*, "Cyber-Attack Mitigation in Cloud-Fog Environment Using an Ensemble Machine Learning Model," in *2022 7th International Conference on Smart and Sustainable Technologies, SpliTech 2022*, 2022. doi: 10.23919/SpliTech55088.2022.9854372.

[19] S. N. A, B. K, G. E, and M. I. C, "Deep Learning For Enhanced Cyber-Attack Detection," in *2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS)*, 2024, pp. 874–878. doi: 10.1109/ICC-ROBINS60238.2024.10533961.

[20] N. Kharlamova, C. Traeholt, and S. Hashemi, "AdaBoost-Based Cyberattack Detection Algorithm for Battery Systems Providing Frequency Regulation," in *2023 IEEE 3rd International Conference on Industrial Electronics for Sustainable Energy Systems, IESES 2023*, 2023. doi: 10.1109/IESES53571.2023.10253743.

[21] B. Kannan, M. Sakthivanitha, S. Jayashree, and R. Maruthi, "Prediction of Cyber Attacks Utilizing Deep Learning Model using Network/Web Traffic Data," *Proc. 3rd Int. Conf. Appl. Artif. Intell. Comput. ICAAIC 2024*, pp. 363–367, 2024, doi: 10.1109/ICAAIC60222.2024.10575032.

[22] R. Madala, N. Vijayakumar, N. Nandini, S. Verma, S. D. Chandvekar, and D. P. Singh, "Automated AI Research on Cyber Attack Prediction and Security Design," in *Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2023*, 2023. doi: 10.1109/IC3I59117.2023.10397798.

[23] A. Swaminathan, B. Ramakrishnan, M. Kanishka, and R. Surendran, "Prediction of Cyber-attacks and Criminality Using Machine Learning Algorithms," in *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2022*, 2022. doi: 10.1109/3ICT56508.2022.9990652.

[24] N. S. Deepak, T. Hanitha, K. Tanniru, L. R. Kiran, N. R. Sai, and M. Jogendra Kumar, "Analyze and Forecast the Cyber Attack Detection Process using Machine Learning Techniques," in *2023 4th International Conference on Electronics and Sustainable Communication Systems, ICESC 2023 - Proceedings*, 2023. doi: 10.1109/ICESC57686.2023.10193289.

[25] A. P. A. Singh, "STRATEGIC APPROACHES TO MATERIALS DATA COLLECTION AND INVENTORY MANAGEMENT," *Int. J. Bus. Quant. Econ. Appl. Manag. Res.*, vol. 7, no. 5, 2022.

[26] J. L. Leevy and T. M. Khoshgoftaar, "A survey and analysis of intrusion detection models based on CSE-CIC-IDS2018 Big Data," *J. Big Data*, 2020, doi: 10.1186/s40537-020-00382-x.

[27] V. Rohilla, S. Chakraborty, and R. Kumar, "Deep learning based feature extraction and a bidirectional hybrid optimised model for location based advertising," *Multimed. Tools Appl.*, 2022, doi: 10.1007/s11042-022-12457-3.

[28] V. Rohilla, M. Kaur, and S. Chakraborty, "An Empirical Framework for Recommendation-based Location Services Using Deep Learning," *Eng. Technol. Appl. Sci. Res.*, 2022, doi: 10.48084/etasr.5126.

[29] P. Lin, K. Ye, and C. Z. Xu, "Dynamic network anomaly detection system by using deep learning techniques," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019. doi: 10.1007/978-3-030-23502-4_12.

[30] S. Mathur., "Supervised Machine Learning-Based Classification and Prediction of Breast Cancer," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12(3), 2024.

[31] R. K. Vinita Rohilla, Sudeshna Chakraborty, "Random Forest with harmony search optimisation for location based advertising," *Int J Innov Technol Explor Eng*, vol. 8, no. 9, pp. 1092–1097, 2019.

[32] R. Tandon, "The Machine Learning Based Regression Models Analysis For House Price Prediction," *Int. J. Res. Anal. Rev.*, vol. 11, no. 3, pp. 296–305, 2024.

[33] S. Mathur and S. Gupta, "Classification and Detection of Automated Facial Mask to COVID-19 based on Deep CNN Model," in *2023 IEEE 7th Conference on Information and Communication Technology, CICT 2023*, 2023. doi: 10.1109/CICT59886.2023.10455699.

[34] Y. N. Kunang, S. Nurmaini, D. Stiawan, and B. Y. Suprapto, "Attack classification of an intrusion detection system using deep learning and hyperparameter optimisation," *J. Inf. Secur. Appl.*, 2021, doi: 10.1016/j.jisa.2021.102804.

[35] R. Tandon, "Face mask detection model based on deep CNN techniques using AWS," *Int. J. Eng. Res. Appl.*, vol. 13, no. 5, pp. 12–19, 2023.

[36] H. Najafi Mohsenabad and M. A. Tut, "Optimizing Cybersecurity Attack Detection in Computer Networks: A Comparative Analysis of Bio-Inspired Optimization Algorithms Using the CSE-CIC-IDS 2018 Dataset," *Appl. Sci.*, vol. 14, no. 3, p. 1044, Jan. 2024, doi: 10.3390/app14031044.