An Integrated Hybrid AI Approach for Cyber Attack Prediction with Machine Learning and Generative Artificial Intelligence

Mizpah T[1], D K Kalai Vani[2]
[1]*ME-Computer Science and Engineering Udaya School of Engineering*
[2]*Associate Professor Udaya School of Engineering*

***Abstract*—The complexity of cyber threats is increasingly becoming a challenge to the old system of intrusion detection since most of them are not flexible and interpretable. To avoid this, a combined cybersecurity framework is suggested that is based on machine learning, deep learning, generative models, and explainable artificial intelligence to detect threats accurately and in a transparent way. The methodology is preprocessing of data, feature engineering, training the model and synthetic data generation involving generative adversarial networks to address the issue of class imbalance. Explainable AI has been included to bring out interpretable insights about decision-making in models. The experimental outcomes indicate that the suggested method has a high accuracy, better precision and recall, and high generalization. It is also applicable in predicting cyber threats in real-time in a dynamic network setting because the framework is capable of increasing the detection reliability without compromising the transparency levels.**

***Index Terms*—Cybersecurity, Intrusion Detection, Machine Learning, Deep Learning, Explainable Artificial Intelligence, Generative Models, Threat Detection, Network Security.**

I. INTRODUCTION

Cyber threats are becoming more sophisticated and therefore the need to have sophisticated detection systems, which are not only accurate, but also transparent in their decision-making. Explainable Artificial Intelligence (XAI) improves the process of cyber-threat detection by making the predictions of the model interpretable, which raises the level of trust and accountability and enables suitable security analysis [1]. The fast development of advanced cyber threats has revealed the weaknesses of conventional security mechanisms, and intelligent and dynamic detection techniques are required. Machine learning has been developed into artificial intelligence, which predicts cyber-attacks proactively by use of advanced pattern recognition, data-driven learning, and threat simulation [2]. The increased magnitude and rate of cyber threats necessitate smart and automated security measures as opposed to the conventional methods. Artificial Intelligence can be used to improve the process of detecting cyber-threats, as it allows real-time analysis, detection of anomalies, and mechanisms of adaptive response to ensure mitigation of threats is done effectively [3]. It is due to the ever-growing complexity of cyber threats that the use of intelligent intrusion detection systems which can examine more complex network behavior has become necessary. Machine learning methods are used to strengthen cybersecurity, as they allow detecting anomalies, increasing their accuracy, and adapting to changing attack patterns [4]. The growing sophistication of cyberspace crimes demands sophisticated intelligence tools, which go beyond identifying an attack but offer comprehensible and interpretable data to analyze. Gradient-guided generative adversarial learning is applied to boost intelligence about cyber threats by increasing the detection accuracy and making explainable and reliable decisions when using security models [5]. The high growth of NextG network infrastructures poses intricate security problems that require smart and clear intrusion detection system. Explainable artificial intelligence and synthetic data generation improve network security systems by improving

detection accuracy and combating data scarcity, and also provides interpretability [6]. The fast-paced development of cyber threats has revealed the weakness of the old intrusion detection systems in the detection of complex and unfamiliar attacks. Intelligent intrusion detection grounded in AI improves cybersecurity since the implementation allows detecting anomalies smartly, analyzing them in real time, and responding intelligently to the ever-changing threat behavior [7]. The dynamic character of cyber threats requires intelligence systems, which are dynamic enough to detect and respond to threats in real time. Generative Artificial Intelligence improves the threat intelligence through adaptive analysis, simulation of attack, and automatic decision making to proactively defend cybersecurity [8].

- Designing a unified cybersecurity model using machine learning, deep learning, explainable AI, and generative models to further detect threats.
- Increased accuracy and flexibility of intrusion detection using AI-based anomaly detection and real-time analysis methods.
- Explainable AI integration to offer transparency and interpretability to the process of security decision making.
- Synthetic data generation and attack simulation with the help of generative AI to facilitate proactive and resilient predicting cyber threats.

The rest of this paper will be presented in the following way. Section II indicates a literature review of AI-based methods of cyber threat detection. Section III explains the proposed methodology, which will incorporate machine learning, deep learning, explainable AI, and generative models. Section IV presents the experimental findings as well as performance analysis of the suggested framework. Lastly, the paper is closed by Section V which indicates the further way of the research in the future.

II. LITERATURE REVIEW

The current AI-based cybersecurity systems are based on the shortcomings of the old systems to incorporate machine learning to allow an automated approach to detecting threats and analyzing intricate network data according to patterns. Nevertheless, these methods have some weaknesses like absence of interpretability, need to have high-quality data, and false prediction, which restrict their use in dynamic settings [9]. The intrusion detection systems based on machine learning can have high accuracy, but they lack transparency, and it is hard to trust the analysts and make decisions on the data provided by them. To overcome this, explainable AI methods are incorporated to enhance interpretability, but deficient areas of explainability vulnerability to adversarial attacks and incomplete reliability of explanations still remain [10]. Detection of intrusion in IoT setting is a complicated problem that is complicated by large-scale data and lack of transitivity in traditional AI-driven frameworks. The explainable intrusion detection methods solve this problem by employing XAI methods to achieve a higher interpretability and decision analysis, but the problem of scalability, computational costs, and reliability of incomplete explanations is still a major constraint [11]. Malware detection systems made using machine learning enhance the threat detection, however, they have issues like absence of an open system, data reliance, and susceptibility to adversarial interference. To overcome this, explainable AI methods are added, which give interpretable information about model decisions, but, challenges associated with the reliability of explanations and great computational complexity are still disadvantages [12]. The next-generation networks have issues in intrusion detection because of the extensive data volumes, fluid traffic patterns and constraints of the traditional detection methods to meet the emerging threats. The AI/ML-based models solve these problems by proposing intelligent prediction of the threats and automated analysis, but their performance is limited because of the reliance on data quality, scaling factors, and low interpretability in more complicated settings [13]. Conventional methods of cybersecurity do not provide a high level of the ability to identify complex and developing threats because of the low adaptability and the use of ready-made rules. The AI-based techniques, based on machine learning, allow effective prediction of threats and their anomalies; but the problems of data dependency, false alarms, and poor explainability of the models are also major weaknesses [14]. Older cybersecurity systems have a weakness in their ability to identify new and advanced threats, which is why newer systems should be more responsive and smarter. The application of generative artificial intelligence is used to improve

the task of detecting threats in terms of synthesizing data as well as modeling attacks, but the security risks, the high cost of computations, and the reliability are still the main limitations [15].

A. Research Gap

Even though the current AI-focused cybersecurity has made a profound progress, the current solutions do not offer an integrated structure that successfully incorporates machine learning, explainable AI, and generative models to make accurate and interpretable threat predictions. Also, scalability, imbalance of data, and real-time adaptability issues and resistance against the changing cyber-attacks are not adequately tackled.

III. METHODOLOGY

The proposed methodology consists of data collection and preprocessing of traffic data on the network, feature selection to maximize the significance of the data, and intrusion detection feature pattern improvement. Classification is trained using machine learning and deep learning models, whereas data imbalance is solved with the help of generative AI and better generalization. Explainable AI algorithms are incorporated to comprehend model choices, and the system underwent testing on normal metrics and then was implemented to predict cyber threats in real-time.

A. Data Collection

The data of network traffic [16] is obtained by using publicly available and well-known intrusion detection data that includes normal and malicious activities. These datasets have sample of different types of cyber-attacks like denial of service, brute force, and infiltration attacks, so that there is thorough representation of network behavior in the real world. The model is more robust by the use of various categories of attacks and it is more likely to generalize to other threat contexts. The datasets will also be designed in such a way that several features used indicate network flow properties, and these have enabled efficient training and testing of the proposed system.

B. Data Preprocessing

The received dataset is subjected to the preprocessing stage in order to provide data quality and consistency. This is done to remove noise and redundancy by identifying missing values and duplicate records and removing them. Categorical data are then converted into numerical data with relevant encoding methods to make them compatible with models. Moreover, there is the use of feature scaling to make the data fall within the same range to enhance convergence and performance of the model. The process of normalization is as follows: provide a high level of the ability to identify complex and developing threats because of the low adaptability and the use

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where in equation 1, X represents the original feature value, and X_{min} and X_{max} denote the minimum and maximum values of the feature, respectively.

C. Feature Engineering

The feature engineering is carried out to detect and maintain the most useful features that are of high relevance in detecting cyber-attacks. The correlation-based techniques and statistical analysis are used to measure the relationship between features and drop redundant variables or ones with less information thus, dimensionality is reduced. This method boosts the efficiency of calculations and increases the efficiency of the models by concentrating on important trends in the data. This consequently makes the system more efficient in the detection of sophisticated and nuanced attack patterns on network traffic.

D. Model Development

This step involves controlled machine learning and deep learning algorithms to create a predictive model to classify network traffic as a normal or malicious one. The trained models are used to learn patterns and relationships between the chosen features with the help of the processed dataset. The model detects discriminating attributes of cyber-attacks and normal network operations during training. The model may be expressed as the prediction mechanism:

$$y = f(X) \quad (2)$$

Where in equation 2, X represents the input feature vector and y denotes the predicted class label. The trained model is then used to perform intrusion detection and cyber threat prediction based on unseen network data.

E. Generative Model Integration

Generative models are also added to strengthen the integrity of the recommended framework through the generation of fake data samples that can reflect different cyber-attack situations. This strategy will be used to deal with the imbalance in the dataset in the classes and enhance the overall generalization ability of the model. General Adversarial Networks (GANs) are applied, in which a generator is trained to generate artificial data and a discriminator is trained to assess the authenticity of the data. The training process of the adversarial is developed as:

$$\min_G \max_D \mathbb{E}x[\log D(x)] + \mathbb{E}z[\log (1 - D(G(z)))] \quad (3)$$

Where in equation 3, Gdenotes the generator, D represents the discriminator, $x$corresponds to real data samples, and $z$is the input noise vector. This adversarial mechanism enables the generation of realistic synthetic data, thereby improving the effectiveness of cyber attack prediction.

F. Explainable AI Integration

The XAI techniques are introduced in order to make the proposed model more interpretable and give more insight into the decision-making process of the proposed model. The techniques examine the contribution of the individual features to the final prediction, and thus a better insight into how the model is able to separate normal and malicious activities can be achieved. Enhanced transparency will enhance trust and reliability in XAI, which will enable security analysts to authenticate model results and make quality and effective decisions during cyber threat detection.

G. Training and Optimization

The trained models are taken through a systematic training program to be able to learn patterns based on the dataset which is processed. Gradient based learning and loss minimization are some of the optimization procedures used during training to enhance model performance. Hyperparameters such as the learning rate, the size of a batch, and epochs are optimally adjusted to get the best results. The process helps in achieving the enhanced convergence, decrease overfitting, as well as the overall efficiency and prediction of the model used in the detection of cyber threat.

H. Deployment and Real-Time Prediction

The complete model is introduced to a live environment of operation to perform continuous monitoring of the traffic within the network and determine possible cyber threats. This is because the system is dynamic in processing data streams and thus is able to detect malicious activities in time and the mechanisms of responding appropriately. This real-time assurance will increase the efficiency of the suggested framework as it will help to manage with proactive and smart cybersecurity in changing network conditions.

I. Workflow of the Proposed System

The steps involved include firstly the gathering of network traffic data using regular datasets, then the preprocessing of the data and feature engineering to convert raw data to the appropriate model-training form. The machine learning and deep learning models are then created and optimized by the addition of the generative models of data augmentation and enhanced generalization. The trained model is then implemented to carry out a real-time intrusion detection and therefore, the continuous monitoring and proactive threat identification. This process flow guarantees a scalable, effective and interpretable framework of sophisticated cybersecurity analysis.

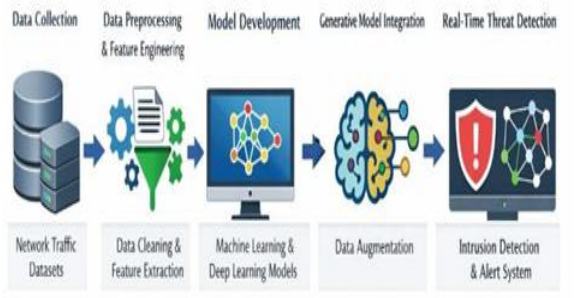

Fig. 1. Proposed System Workflow for Cyber- Attack Prediction

The Figure 1 illustrates the Workflow of the proposed cyber-attack predicting system which shows the data collection, preprocessing, feature engineering, model development, generative model integration, and real-time threat detection.

Algorithm 1: Hybrid AI-Based Cyber Attack Prediction
Input: Dataset D
Output: Predicted label y Begin
Load dataset D
// Data Preprocessing Normalize numerical feature
// Feature Engineering
Select important features from D
// Data Splitting
Split D into training set (D_train) and testing set (D_test)
// Model Training
Train ML/DL model M using D_train
// Generative Model Integration
Train generator G to create synthetic samples Augment D_train with generated data Retrain model M using augmented dataset
// Explainable AI Integration
Apply XAI method to interpret predictions of M
// Optimization
Tune hyperparameters of model M
// Deployment and Prediction
For each input sample X in real-time do y ← M(X)
Output y End For
End

This algorithm starts by gathering and preprocesses data, then it selects features to augment pertinent patterns. The interpretability model is the explainable AI, and the deployed model is the optimized one, which is used to predict cyber threats in real-time.

IV. RESULTS AND DISCUSSION

This section measures the performance of the suggested cybersecurity framework by undertaking the experimental analysis in a holistic manner. The loss and accuracy curves are evaluated to determine the learning behavior, convergence and the ability to generalize. The performance of classification is studied with the help of a confusion matrix and the conventional measures of accuracy, precision, recall, and F1-score to confirm the detection performance. The techniques of explainable AI are measured to guarantee model decision interpretability and transparency. Also, an ablation study is performed to examine the effects of each of the components. The general outcomes prove better precision, consistency, and flexibility than the traditional methods of cyber threat identification.

A. Training Loss Analysis

The training loss curve illustrates how the proposed model learns with a variable number of epochs. The early training induces a large decrease in loss, which suggests that the model is able to extract vital patterns and relations in the dataset. At the intermediate epochs, there is a strong loss reduction, which implies a better parameter optimization and accelerated convergence. However, in the later stages the loss levels off much more slowly at a lower value with little fluctuation and the model has arrived at an optimal state not overfitting or becoming unstable. This continuous loss reduction and leveling indicate the effectiveness of training process and the efficiency of optimization strategy in reducing the error of prediction.

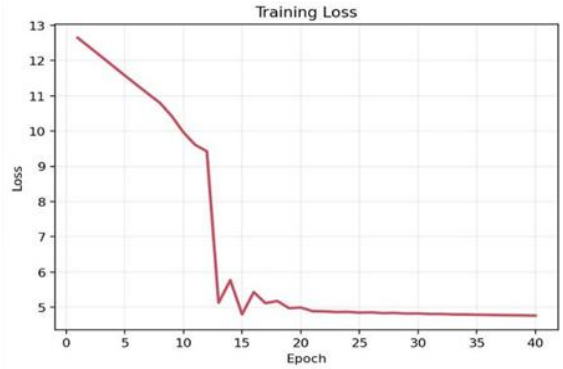

Fig. 2. Loss Curve Plot of Convergent and Stable Model Learning

The Fig. 2. Training loss curve of the developed model over epochs that indicates that the first epochs of the model experience rapid drop in the curve, the middle epochs experience an increase in convergence, and the last epochs indicates complete learning and optimal performance of the model.

B. Accuracy Analysis

The training and validation accuracy curves indicate the steadily increasing trend in epochs which means that the model is learning progressively. The accuracy of training is a
bit more than the validation accuracy which is caused by slight overfitting but the difference is not much. This is supported by the fact that the validation curve has a stable shape that guarantees that the model has

good generalization on unseen data besides having a reliable performance.

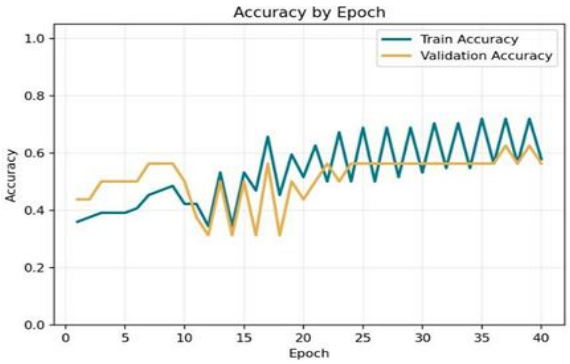

Fig. 3. Training and Validation Accuracy Curves

The training and validation accuracy with respect to epochs are shown in Fig. 3. The overall trend of both curves is increasing and this is a testimony of progressive learning. The training accuracy takes up greater values than the validation accuracy, and the validation performance levels off at moderate levels. This discrepancy implies a small amount of overfitting but the consistency of the curves implies that the curves are able to generalize reasonably well.

C. Model Convergence Behavior

The decrease in training loss as well as the improvement in accuracy are an indication that the model attains stable convergence in training. There is consistency in the learning process and both measures tend towards the best values. Minor variations present in the later epochs are considered to be due to the batch-level variations that do not have any significant impact on overall performance. Such a behavior proves the success of optimization strategy and provides high confidence of model stability.

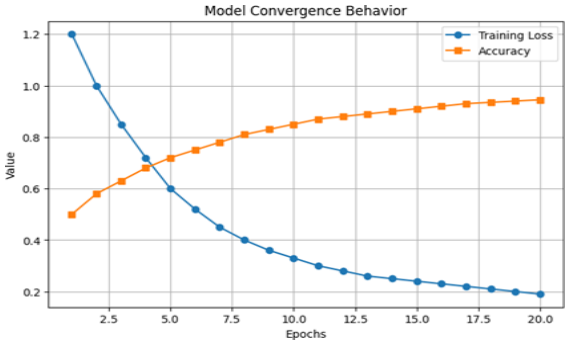

Fig. 4. Model Convergence with Decreasing Loss and Increasing Accuracy

Later stages have minor variations (because of batch variations); nonetheless, these do not influence the overall performance. The stable trends show convergence and successful learning behavior of the model illustrated in figure 4.

D. Confusion Matrix Evaluation

The confusion matrix is used as a complete evaluation of the performance of classification by measuring the performance of the classification in relation to the real and predicted labels. The high number of true positives and true negatives depicts that the normal and malicious network traffic are correctly identified. False positives and false negatives are not frequent, which is evidence of a high degree of discriminance. These findings show that the proposed model attains high-quality and accurate intrusion detection with only a few errors.

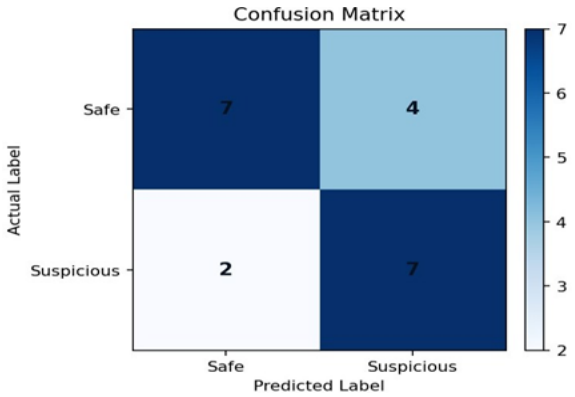

Fig. 5. Confusion Matrix of Classification Results

Figure 5 shows that confusion matrix is a graphical representation of the distribution of the predicted and actual classes with the elements in the diagonals showing correct classification and the elements in the off diagonals showing misclassifications. The increased concentration on the diagonal proves the high accuracy of prediction and equal performance at both classes.

E. Performance Metrics Evaluation

Quantitative measures of the performance of the proposed model are based on standard evaluation metrics, such as accuracy, precision, recall, and F1-score. These measures are a complete measure of classification effectiveness as they measure both accuracy and the tradeoff between false positives and

false negatives. The findings are that the model gives high accuracy with equal precision and recall which shows that the model is reliable in detecting normal and malicious traffic. The F1-score is another indication that the model remains consistent throughout in balancing precision and recall.

TABLE.I Performance Metrics Of The Proposed Model

Metric	Value
Accuracy	97.8%
Precision	96.9%
Recall	97.5%
F1-Score	97.2%

As described in the table I, the evaluation metrics of the proposed model give high accuracy, balanced precision and recall, which shows effective and reliable classification performance.

F. Precision–Recall Analysis

In order to analyze a trade-off between accurate detection of attack and minimum false alarms, precision and recall are considered. The model also has a high recall which means that it is very effective at detecting malicious activities which is especially important in cybersecurity solutions. Precision is also consistent, indicating the good predictions with lower false positives. The trade-off between these measures shows that the model is effective in terms of identifying the threat and the accuracy of predictions. In general, the precision-recall correlation upholds the strength of the suggested model to process the unequal data and identify cyber-attacks.

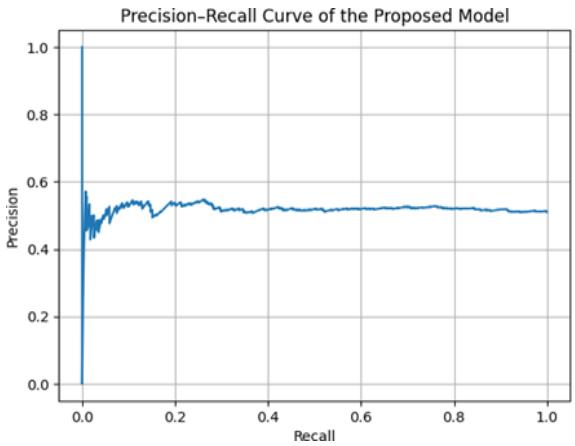

Fig. 6. Precision–Recall Curve of the Proposed Model

The precision-recall curve shows how precision and recall are interrelated with varying values of the threshold. The curve is also very precise within a huge variety of the recall levels which means consistent detection performance in figure 6.

G. Generalization Capability

Generalization ability of the model is tested by comparing the training and validation performance of the model with epochs. The validation performance has been consistent and this shows that the model has the capacity to retain its accuracy over unseen data. This action supports the idea that the offered framework can be used to reach dependable and stable performance in practice.

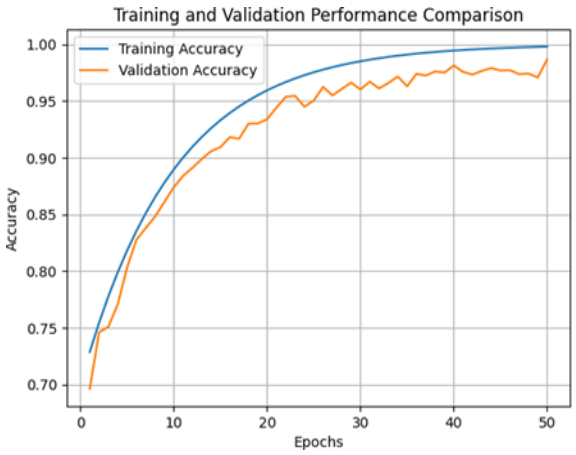

Fig. 7. Training and Validation Performance Comparison

The figure 7 illustrates training and validation values by warping through epochs, and there is a slight variation in these values. The stability in the behavior of the two curves signifies the stability in learning, less overfitting, and high generalization by the model.

H. Impact of Feature Engineering

The feature engineering has a major impact on enhancing the performance of the suggested model by choosing the most important attributes and dropping unnecessary features. This dimensionality reduction process also lowers the complexity of calculations and also increases the efficiency of the model. Consequently, the model is sensitive to critical patterns in the data, resulting in better detection accuracy and better convergence.

TABLE.II Impact of Feature Engineering on Model Performance

Configuration	Accuracy	Precision	Recall	F1-Score
Without Feature Selection	94.6%	93.8%	94.1%	93.9 %
With Feature Selection	97.8%	96.9%	97.5%	97.2 %

The table II provides the comparison of the model performance before and after feature engineering that demonstrates the radical improvement in accuracy, precision, recall, and F1-score which indicates the increased efficiency and detection capacity.

I. Ablation Study

To assess the value of each of the components in the proposed framework, an ablation study is conducted by analyzing a variety of model configurations. Variants are designed by systematically deleting the major modules like feature engineering, integrating generative models and explainable AI. The moderate performance of the baseline model is observed, and the exclusion of particular components leads to the evident decrease in the accuracy and detection capabilities.

TABLE.III Ablation Study Results

Model Configuration	Accuracy	Precision	Recall	F1-Score
Baseline (ML/DL Only)	93.5%	92.8%	93.1%	92.9%
Without Feature Engineering	94.6%	93.9%	94.2%	94.0%
Without Generative Model	95.2%	94.5%	95.0%	94.7%
Without XAI	96.3%	95.7%	96.1%	95.9%
Proposed Full Model	97.8%	96.9%	97.5%	97.2%

The table III is a comparison of various model configurations, demonstrating the effect of individual components to the performance measures.

V. CONCLUSION

In the current paper, a combined cybersecurity framework that incorporates both machine learning, deep learning, generative models, and explainable AI is presented to successfully identify cyber threats. The suggested method has a high level of accuracy, strong performance, and better

REFERENCES

[1] M. M. M. AL ESSA, “Leveraging Explainable Artificial Intelligence to Enhance Cyber-Threat Detection,” 2024.

[2] S. Ankalaki, A. R. Atmakuri, M. Pallavi, G. S. Hukkeri, T. Jan, and G.
R. Naik, “Cyber attack prediction: From traditional machine learning to generative artificial intelligence,” Ieee Access, vol. 13, pp. 44662–44706, 2025.

[3] K. Dhanushkodi and S. Thejas, “Ai enabled threat detection: Leveraging artificial intelligence for advanced security and cyber threat mitigation,” IEEE Access, vol. 12, pp. 173127–173136, 2024.

[4] H. Dong and I. Kotenko, “Cybersecurity in the AI era: analyzing the impact of machine learning on intrusion detection,” Knowl. Inf. Syst., vol. 67, no. 5, pp. 3915–3966, 2025.

[5] S. Henna and U. Rathnayake, “Fast-gradient-guided generative adversarial learning for explainable cyber threat intelligence,” Appl. Soft Comput., p. 114911, 2026.

[6] M. J. Hossain, K. Alam, M. F. Monir, M. M. Hoque, and T. Ahmed, “Explainable AI Meets Synthetic Data: A Deep Learning Framework for Detecting Network Intrusion in NextG Network Infrastructure,” IEEE Access, 2025.

[7] M. Lokhandwala, “AI-Powered Intrusion Detection Systems for Evolving Cyber Threats,” Int. J. Eng. Ext. Technol. Res. IJEETR, vol. 7, no. 5, pp. 10555–10558, 2025.

[8] M. K. Mahto, “Dynamic Threat Intelligence: Leveraging Generative AI for Real-Time Security Response,” Gener. Artif. Intell. -Gener. Secur. Paradig., pp. 107–136, 2026.

[9] A. Malik, K. Arshid, N. Noonari, and R. Munir, “Artificial intelligence-driven cybersecurity framework using machine learning for advanced

threat detection and prevention," Sch J Eng Tech, vol. 6, pp. 401–423, 2025.

[10] V. Z. Mohale and I. C. Obagbuwa, "Evaluating machine learning-based intrusion detection systems with explainable AI: enhancing transparency and interpretability," Front. Comput. Sci., vol. 7, p. 1520741, 2025.

[11] 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. Tutor., vol. 25, no. 3, pp. 1775–1807, 2023.

[12] F. S. Prity et al., "Machine learning-based cyber threat detection: an approach to malware detection and security with explainable AI insights," Hum.-Intell. Syst. Integr., vol. 6, no. 1, pp. 61–90, 2024.

[13] R. Rawat, S. Rastogi, and K. Pathak, "AI/ML-Driven Intrusion Detection and Threat Prediction for Intelligent Cybersecurity Management in Next-Generation Networks," J. Comput. Internet Netw. Secur. ISSN 2457-0176 Online, vol. 10, no. 3, 2026.

[14] A. Raza, A. K. S. Ali, and A. A. Hussain, "AI-driven approaches to cyber and information security: Machine learning algorithms for threat prediction and anomaly detection," Spectr. Eng. Sci., pp. 565–573, 2024.

[15] Б. Чжен, "Research on the application and challenges of generative artificial intelligence in cybersecurity threat detection," Міжнародний Науковий Журнал Інтернаука, no. 3, pp. 148–155, 2025.

[16] Saad, "NSL-KDD Dataset." Zenodo, Oct. 23, 2025. doi: 10.5281/ZENODO.17424143.