Deep Learning-Driven Cardiometric Analysis on Chest X-Ray Images for Early Cardiac Diagnosis

K Ansha[1], Brintha C[2]
[1,2]*Department of Computer science and Engineering, Udaya School of Engineering*

***Abstract*—The proliferation of advanced cyber-attacks has been triggered by the active construction of digital infrastructure, whereby the efficient intrusion detection systems are the determinant of providing the network security. The paper under consideration offers a mixed method of machine learning-based cyber-attacks prediction with the help of GANs to enhance the detectability. The real world of the normal and evil network traffic is rendered with the help of the UGR16 information. Preprocessing of data and feature extraction arrangements are installed in such a manner that quality of data and the complexity are increased. The machine learning schemes used in the classification include: Random Forest, Support Vector Machine and Logistic Regression, and GAN is employed to generate false attack samples to overcome the issue of imbalanced classes. The system accuracy of the proposed system is 95% which represents the effectiveness of the system to detect the normal and attack traffic. The findings of the experiments demonstrate the improvement of the precision, recall and F1-score, i.e. the larger the detection ability the smaller the false positives. The model also has good generalization behavior besides stable training behavior and validation. In general, the problem of generative and predictive strategies is an effective and scaled solution to the current issues of cybersecurity.**

***Index Terms*—Cyber Attack Detection, Intrusion Detection System, Machine Learning, Generative Adversarial Network, UGR16 Dataset.**

I. INTRODUCTION

The high rate of development of the digital infrastructure has made it difficult to detect cyber-attacks, which have become more advanced and widespread. The networked systems are very crucial to organizations within any given industry and therefore they are susceptible to any form of computer malware, phishing, as well as denial-of-service attacks[1]. The IDS is significant in terms of tracking network traffic and detection of suspicious activities on the network. Nevertheless, monitoring network data is not effective as the volume of network data and its complexity continues to grow. Thus, the modern-day cybersecurity requires the use of automated and smart detection systems. These systems aid in integrity, confidentiality and accessibility of data [2].

Scenario of signature-based detection that is predominantly used by the traditional Intrusion Detection Systems seek to recognize patterns of attack that are familiar. They have been effective in identifying the threats, which were already established; they cannot identify emerging threats or variations of the threat, which is often known as the zero-day attacks. Besides that, signature based systems need to be constantly updated and this may be resource exhausting and time consuming too [3]. This is their weakness that constrains their productivity under dynamic and large-scale network environments. This leads to increased requirement of more adaptable and smart methods of detection. This has been limited to the implementation of machine learning strategy[4].

The machine learning methods have provided a significant boost to the intrusion detection system to learn the patterns on data and identify the aberrations. The examples of such algorithms that are typically applied to the problem of cybersecurity classification are SVM, RF, and LR to mention only several of them[5]. Such models also have challenges even though they have advantages like unbalanced datasets, the quantity of attack cases is very small in comparison to normal traffic. False models and low detection rate are the result of the imbalance. Moreover, the classical models of machine learning might fail to generalize those patterns of attack that are not trained. As a result, their strength and capacity ought to be fortified in the future by enhancements[6].

The UGR16 modeling information is a real and large-sized model of the network traffic comprising of the beneficial and harmful activities. In the current paper, the author suggests a combination of machine learning and generative methods, including GANs[7]. The generative model can be useful in combating the issue of data imbalance through the synthesis of attack examples. This also maximizes training process, as well as in detection. The given framework will offer more viable and efficient means of predicting cyber-attacks[8].

The key contribution of this study was outlined below:

- The paper recommends developing a hybrid approach for hacking attack prediction by integrating traditional machine learning models GANs to enhance detection accuracy and overall performance.
- It employs the UGR 16 data to emulate an experience of an assessment situation that involves a heterogeneous network traffic and attack situation.
- The article solves the issue of asymmetry of the classes, which generates fake samples of the attacks, being produced by generative models and raising the resistance of models.
- The suggested solution has fewer false positive and more accurate than the traditional intrusion detection protocol and mechanisms.

The following format will be followed in the rest of this paper. The description of data set and its nature is located in Section II. The second (III) deals with problem statement and research problems. Section IV presents the proposed methodology which consists of preprocessing, prediction and modeling. Section V and conclusion and future work are also presented in the final sections as the results and performance analysis are given.

II. LITERATURE REVIEW

Intrusion detection systems based on DL have shown a great increase in the recognition of extremely complex and unknown attack pattern through the automatic learning of features on the network traffic data. These models are not associated with manual feature engineering, and are more accurate on the feature finding than the conventional methods. They however, need much data which has been labeled to be trained properly and face the challenge of the issue of data imbalance frequently. Moreover, it might be constrained when it comes to its real time use since it possesses a high level of computation complexity and training time. Nevertheless, these shortcomings do not hinder the future evolution of contemporary systems of cybersecurity to the utmost degree on the basis of deep learning [9]. The data of KDD Cup 99 have been used as a guideline in the comparison of methods of intrusion detection system. It has a huge amount of network traffic data that are divided into normal and attack cases that allow building the classification models. Serious problems also lie within the dataset including duplicity and disproportion of some types of attacks. This renders discrimination learning and overvalued scores of performances in majority of models. Then, the information might not reflect the actual network environment and the current attacks scenarios in life situations [10].

The plan aims at enhancing the host-based intrusion detection that compares the system call patterns that the applications produce. It incorporates discontinuous and contiguous sequences to derive other semantic relationships between system calls and therefore achieve a better result of abnormal behavior. The method is more effective in the process of detecting the presence of the unknown and complex attacks that would otherwise go unnoticed by simple signature systems. It is however computationally high and can produce false positives because of variations in normal system behaviour. Nevertheless, it possesses a stronger and multi-faceted mechanism of intrusion detecting [11]. It is a maximization optimization of the intrusion detection systems which are executed by Evolutionary algorithm application with the aim of detecting anomalies in a network traffic flow. The model is augmented with detection rules and parameters, and thus, it increases its detection capabilities between the normal and malicious activity. It can however be very computation time consuming and convergence time consuming to optimize. Nevertheless, in spite of these shortcomings, a solution can be found to the problem of identifying anomalies in the sphere of cybersecurity with evolutionary methods being flexible and adaptable to the issue at hand [12].

GANs are an example of deep learning that is self-possessed of dual elements, an originator, and a discriminator, and is trained competitively. The generator creates the counterfeit versions of data and the discriminator determines the genuine data of the

data. This confrontation procedure permits educating the model on a complex data distribution and produce valid outputs. Nevertheless, GANs are hard to train and they can experience mode collapse and instability. They however prove useful both within augmenting data and imbalanced data in other applications [13]. The system offers hierarchical space and time-based founded intrusion detection system using deep learning to get to know about network traffic information features. Through the use of deep neural networks, the model can extract complicated patterns and dependencies which the other traditional methods might not be at a position to extract. It improves the degree of awareness particularly when it comes to advanced and dynamic cyber-attack. The model is however very expensive in terms of computation as well as huge datasets to be trained. In spite of these shortcomings, it greatly increases the intrusion detecting system performance[14].

The specified work is aimed at the control of unbalanced datasets in the classification, which is also an issue with intrusion detection systems. It draws especially close attention to the fact that the right evaluation measures rather than accuracy because of the possibility to evaluate the model performance. Some of the methods that are employed to address the problem of class imbalance include resampling, synthetic data generation and cost-sensitive learning. Nonetheless, even though it is not addressed in the right way unbalanced situation can result in biased models and overfitting. All in all, the above approaches offer enhanced accuracy and classification algorithms efficiency to recognize rare occurrences such as the occurrence of cyber-attacks [15]. The LSTM networks will inform this as it will identify the sequence patterns of the network traffic to determine intrusion. The model is able to obtain time dependencies and hence detect complex and time-related attack patterns. It enhances the functionality of detection particularly on the dynamic and time-based advanced threats. Nonetheless, it is very computationally costly and needs extensive training data. Nonetheless, they offer an effective resolution of the issue of sequence-based intrusion detector despite such disappointments [16].

A. Research Gap

Despite the fact that the level of development of the intrusion detection systems has reached considerable levels, some challenges remain that are yet to be addressed. Lack of imbalance in faced data is a weakness of conventional machine learning models, which lead to biased predictions and low detection of infrequent cases of attacks. Deep learning methods are more accurate and demand lots of computation power and high number of labels that limit the application of deep learning in practice. Failure to identify unfamiliar as well as dynamic attack patterns on real time environments has also been a challenge to the extant procedures. That is why such a hybrid and effective approach is mandatory to have a possibility to fix the problem of data imbalance and enhance the performance of generalization and the overall detection.

III. RESEARCH METHODOLOGY

The suggested methodology will involve a hybrid design, which will involve pre-processing of the information, feature extraction, machine learning models, and generative techniques to enhance the results of data-mining in order to detect cyber-attacks. It optimizes the system with big network traffic data and resolves the incompatibility of data, and low detection accuracy. The processing of the data is preprocessed to assure the quality of data and the selection of features is made to make the processing less complex. This is then succeeded by running the machine learning algorithms and a GAN that is introduced in order to come up with fake samples of attack. Lastly, the trained model generates a prediction of the network traffic that is either normal or malicious, which is more accurate and stronger.

B. Data Collection

The data employed in this research is obtained by utilizing UGR16 data that is realistic and large-scale network traffic data that can be used in intrusion detection study. Such data set and other types of cyber-attacks are addressed in reality world situations as regular and malicious traffic are part of such data set. The data is clear and is applied usually to assess cybersecurity models. This information will be convenient to work with because the proposed system will be trained and tested on sound and diverse data, which will enhance the effectiveness of implementing it in the detection of cyber-attacks [17].

C. Data Preprocessing

The former is the preparation and cleaning of the data to analyze them. The absent values of the data set are either dropped or substituted to eliminate inconsistency when training the model. Some of the techniques used in normalizing the features include min-max scaling and they are expressed as (1).

$$\text{Xnorm} = \text{Xmax} - \text{XminX} - \text{Xmin} \quad (1)$$

This is used to ensure that all the features fall within the range of common values that ensure that models are stable. The method of encoding the categorical variables into numbers is called label encoding, which is needed to enable their incorporation into machine learning algorithms.

D. Feature Extraction

It entails feature extraction process, which is chosen to make a good prediction based on the best attributes in the dataset. Dimensionality reduction and cost reduction are achieved by dimensionality reduction and elimination of redundant and insignificant features. The feature selection may be stated as a simple form represented on (2).

$$\text{Fselected} = \{\text{f1}, \text{f2}, \text{f3}, \ldots, \text{fn}\} \quad (2)$$

where $F_{selected}$represents is used to denote the sub set of important features. This enhances the efficiency of the models and lowers the overfitting. The system runs faster and has a higher generalization because it has been designed to pay attention to notable aspects.

E. Machine Learning Model

This step entails the training of the models of classifications in the perspective of distinguishing between normal and abnormal network traffic. They include RF algorithm, SVM and LR algorithms. It is possible to perform the probabilistic prediction of the attack with the aid of the LR in the following equation (3).

$$\text{P}(\text{y} = 1 \mid \text{x}) = 1 + \text{e} - (\text{w} \cdot \text{x} + \text{b})1 \quad (3)$$

where wrepresents weights and bis bias. These models are trained on training data and are evaluated to find out the best performing classifier. The fact that several algorithms are used guarantees the results of the prediction are powerful and precise.

F. Generative Adversarial Network

The unbalanced data sets are handled with the help of GAN to generate counterfeit attack samples to the minority population. It is composed of two parts where the generator provides artificial samples of data and the discriminator is the one that estimates whether the data is natural or artificial. The discriminator attempts to induce classification of the data it is and the producer efforts to deceive the discriminator with the fake data. It is a type of adversarial learning which may be used in the effective estimation of the underlying data distribution.

In this case, D(x) represents here the likelihood that the input information is real as opposed to G(z) which is the output data depending on random noise z. This is the role that is supposed to be played by the generator and the role that is supposed to be played by the discriminator. This leads to an increased classification accuracy and less biasness towards majority classes.

G. Prediction Model

The trained classifiers are used to construct the final prediction model which utilizes balanced data. The input features feed the model which is used to determine whether the traffic is of normal character or an attack. The projection will be put as (4).

$$\text{y}\hat{} = \text{f(x)} \quad (4)$$

where f(x)is the trained model. The reliability and efficiency of intrusion detection system is guaranteed by the preprocessing and feature extraction, machine learning and GAN combination.

The imbalance of classes is addressed with the help of GAN and creates fake attack samples, which increases the training data. And lastly, the trained model is very accurate and powerful in predicting the nature of network traffic. The result of such combination will be better detection performance, bias reduction, and their generalization in real cybersecurity scenarios.

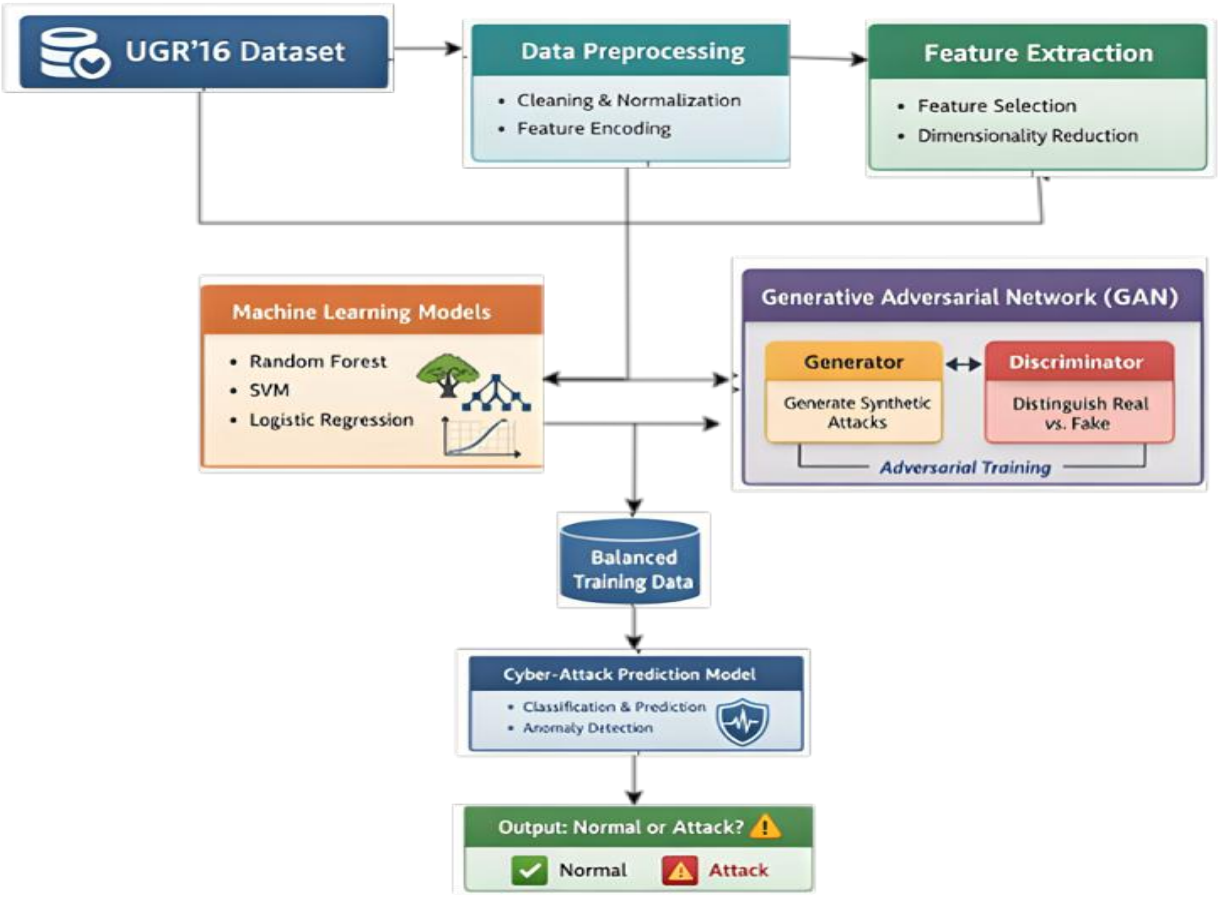

Fig. 1. GAN-based cyber-attack detection architecture

Fig. 1 illustrates; this paper recommends a decision-based hybrid deep learning model to identify a cyber-attack based on UGR16 dataset. Preprocessing data entails cleaning, normalization and coding of the raw data of traffic and then feature extraction is performed to select the important attributes as well as downsize the data. The attack samples are generated synthetically using GAN to address the imbalance between the classes and the outputs generated are utilized together with the classifiers to give a balanced training set. Finally, the prediction model determines network traffic to be either normal or malicious with the assurance that it is sound and precise in identification of the cyber-attacks.

Algorithm 1: GAN-Based Cyber Attack Prediction System
Input: Network traffic dataset D Output: Predicted class (Normal / Attack) Step 1: Initialize the Generator Gand Discriminator Dwith random weights. Step 2: Input real data samples xfrom the dataset. Step 3: Generate random noise vector zfrom a prior distribution. Step 4: Produce synthetic data using the generator: $x_{fake} = G(z)$. Step 5: Train the discriminator using both real data xand fake data x_{fake}. Step 6: Update discriminator weights to maximize its ability to distinguish real and fake data. Step 7: Generate new fake samples using the generator. Step 8: Train the generator to fool the discriminator by minimizing classification error. Step 9: Update generator weights based on feedback from the discriminator. Step 10: Repeat the process iteratively until the model converges.

The GAN is significant in addressing the problem of the imbalance of classes through the production of synthetic samples that are authentic in the minority classes. The generator is conditioned to produce data that aligns with the actual attack patterns and the

discriminator continuously checks itself and attempts to be more suited to be able to identify both real and fake samples. The adversarial training ensures that that artificial data is utilized to equalization the dataset to reduce the biasness of the majority classes.

IV. RESULTS AND DISCUSSIONS

The UGR'16 data set is employed to investigate the proposed system of cyber-attack prediction in terms of its efficiency to identify network traffic that would be considered as malicious. As the results of the experiment indicate, the specified hybrid model is characterized by the overall accuracy of 95; this fact demonstrates the prospects of the proposed model to be effective at detecting the normal and attack traffic. GAN integration can be applied extremely helpful in the identification of minority attack classes and a reduced bias in classification.

A. Experimental Outcome

The experiment assessment will be undertaken on the processed UGR'16 data set, which includes malicious and normal traffic samples. The dataset is made more balanced following the preprocessing and GAN-based data augmentation, which facilitates classification model learning. The findings indicate that the suggested model is fairly effective in terms of various types of attacks.

Table I Performance Metrics of Proposed Model

Metric	Value (%)
Accuracy	95
Precision	94
Recall	93
F1-Score	93.5

In Table I, the GAN-based machine learning model is tested on the UGR, which is 16 datasets. The model has an accuracy of 95 which indicates that it is able to classify well in regard to distinguishing normal and attack traffic. The precision and the recall values indicate that the model is able to effectively identify true positives and minimize the false detections.

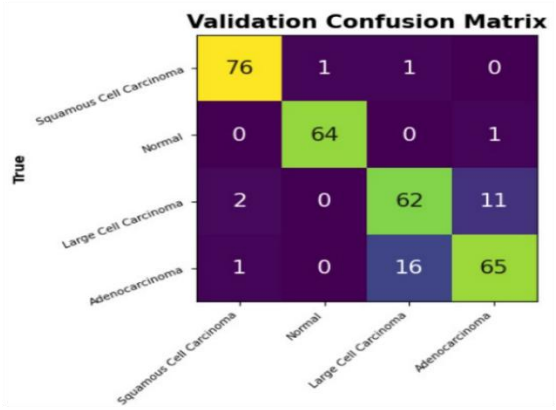

Fig. 2. Validation Confusion Matrix

Figure 2 represented the validation confusion matrix of the proposed classification model in four classes, which are, Squamous Cell Carcinoma, Normal, Large Cell Carcinoma and Adenocarcinoma. On the diagonal, the model has good results with good predictions that are accurate as depicted 76 with Squamous Cell Carcinoma and 65 with Adenocarcinoma. Normal cases have few misclassifications and this indicates that it has a good detection of the healthy samples. It has however, a mixed up of Large Cell Carcinoma and Adenocarcinoma as indicated by 11 and 16 misclassified cases. Overall, the matrix is very accurate in making classification with slight intersection of similar types of cancer.

B. Model Analysis

The model analysis will serve the purpose to test learning behavior and the generalization ability of the proposed system. Monitoring training and testing activities is done to ensure that the model is not over-fitting or under-fitting the data. The results indicate no significant differences whether convergence and performance are between training and testing data sets.

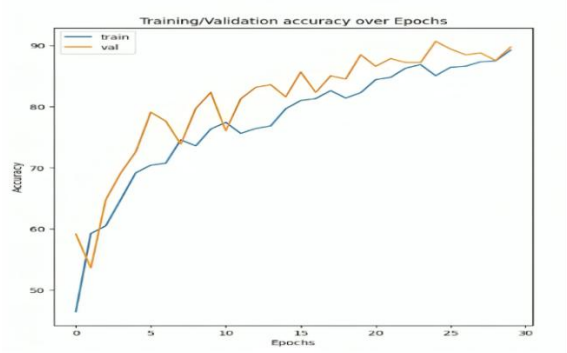

Fig. 3. Training and Validation Accuracy over Epochs

The figure shows how the training and validation accuracy varying with the use of different epochs. It is possible to observe that training and validation accuracy gradually increase with increasing the number of the epochs, which ensures that the model learns successfully. The validation accuracy follows the training curve with slight variations indicating that it has good generalization ability. As Fig. 3 shows, the model has high accuracy of about 90 indicating consistent convergence with no overfitting.

Fig. 4. Training and Validation Loss over Epochs

This is a difference in the training and validation loss during training. The values of the losses also decrease continuously as the epochs are increasing in number and this is an indication that indeed the model is minimizing error. The validation loss is of the same trend as the training loss with slight differences depending on the complexity of the data. Both curves intersect at the point as presented in Fig. 4 which means that the model is well-fitted and there is no serious overfitting or underfitting.

Table III Baseline Comparison of Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	88	86	85	85
Support Vector Machine (SVM)	91	90	89	89
Random Forest	93	92	91	91
Proposed GAN + ML	95	94	93	93.5

According to the results provided in Table III, the suggested GAN-based model is more effective than all the other baseline machine learning models in all the metrics. The performance of LR and SVM is rather poor due to their inabilities to address the cases of imbalanced sets of data. Although the best method is to use the Random Forest because it is an ensemble method, it is inferior to the one proposed. This comparison goes to prove categorically the strength and force of the hybrid structure.

V. DISCUSSION

Using the findings of the experiment, one can conclude that GAN combined with machine learning can promote the quality of cyber-attack detection to a greater extent. Pseudo sample generation helps to balance the sample that enhances the learning capacity of classification models. The system proposed is highly accurate and it possesses high precision and recall ratio. The model is also noted with high capability of generalization; therefore, it can be applied to the real world. This study confirms that the hybrid methodology that incorporates the generation with the conventional techniques is a justifiable way out to intrusion detection issues

VI. CONCLUSION

The current paper is a successful hybrid-type model of cyber-attacks prediction through the combination of machine learning and GANs. The goal of the study is to overcome the limitations of the traditional intrusion detecting systems, particularly the issues of data discrepancies, inefficient generalization and capability to detect new patterns of attacks. The proposed system

will offer realistic simulation of an environment with diversity of network traffic data through use of the UGR16 data set. To remove the issue of unequal representation, the GAN is used to generate the fake samples of the attack, which increases the balance of classes. This kind of integration achieves the effect of augmenting the learning capability of the classification models and reducing bias to the majority classes.

The provided model of the incorporation of GAN, as the results of the experiment have demonstrated, has a high accuracy of 95, which is greater than the current machine learning methods. The model has also been proven to be of high generalization capability by showing that the model has a constant training and validation performance. In a conclusive manner, the presented system provides a scalable and powerful solution to cyber-attack detection. It is an effective combination of generative and predictive to increase the degree of detection accuracy and reliability. This can find a huge potential application in the literal meaning of the cybersecurity particularly in the dynamic nature where new and advanced threats are yet to materialize.

REFERENCES

A. S. George, T. Baskar, and P. B. Srikaanth, "Cyber threats to critical infrastructure: Assessing vulnerabilities across key sectors," *Partners Universal International Innovation Journal*, vol. 2, no. 1, pp. 51–75, 2024.

J. Hudson, "Artificial Intelligence and Cybersecurity Integration: Modern Database Techniques for Securing AI Models," 2024.

R. Padmavathy and N. Hurrah, "Frontiers in Cybersecurity: Battling Zero-Day Attacks and Advanced Persistent Threats," *Exploring Frontiers in Artificial Intelligence and Machine Learning Technologies*, 2024.

S. Tanwar, Q. Bhatia, P. Patel, A. Kumari, P. K. Singh, and W.-C. Hong, "Machine learning adoption in blockchain-based smart applications: The challenges, and a way forward," *IEEE Access*, vol. 8, pp. 474–488, 2020.

M. Sannigrahi and R. Thandeeswaran, "Predictive analysis of network-based attacks by hybrid machine learning algorithms utilizing Bayesian optimization, logistic regression, and random forest algorithm," *IEEE Access*, vol. 12, pp. 142721–142732, 2024.

L. A. Babatunde *et al.*, "Adversarial machine learning in cybersecurity: Vulnerabilities and defense strategies," *Journal of Frontiers in Multidisciplinary Research*, vol. 1, no. 2, pp. 31–45, 2020.

R. SinhaRoy and A. Sen, "A hybrid deep learning framework to predict Alzheimer's disease progression using generative adversarial networks and deep convolutional neural networks," *Arabian Journal for Science and Engineering*, vol. 49, no. 3, pp. 3267–3284, 2024.

V. Dutta, M. Choraś, M. Pawlicki, and R. Kozik, "A deep learning ensemble for network anomaly and cyber-attack detection," *Sensors*, vol. 20, no. 16, Art. no. 4583, 2020.

A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in *Proc. 9th EAI Int. Conf. Bio-inspired Information and Communications Technologies (BIONETICS)*, New York, NY, USA, 2016. doi: 10.4108/eai.3-12-2015.2262516.

M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *Proc. IEEE Symp. Computational Intelligence for Security and Defense Applications (CISDA)*, Ottawa, ON, Canada, Jul. 2009, pp. 1–6. doi: 10.1109/CISDA.2009.5356528.

G. Creech and J. Hu, "A semantic approach to host-based intrusion detection systems using contiguous and discontiguous system call patterns," *IEEE Transactions on Computers*, vol. 63, no. 4, pp. 807–819, Apr. 2014. doi: 10.1109/TC.2013.13.

A. Mora, P. Merino, D. Hernández, P. García-Sánchez, and A. Fernández-Ares, "Applying evolutionary methods for the optimization of an intrusion detection system to detect anomalies in network traffic flows," in *Advances in Computers*, vol. 135. Amsterdam, The Netherlands: Elsevier, 2024, pp. 313–347.

I. Goodfellow *et al.*, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020. doi: 10.1145/3422622.

W. Wang *et al.*, "HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection," *IEEE Access*, vol. 6, pp. 1792–1806, 2018. doi: 10.1109/ACCESS.2017.2780250.

J. Brownlee, *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning*. Melbourne, Australia: Machine Learning Mastery, 2020.

J. Kim, J. Kim, H. L. T. Thu, and H. Kim, "Long short-term memory recurrent neural network classifier for intrusion detection," in *Proc. Int. Conf. Platform Technology and Service (PlatCon)*, Jeju, South Korea, 2016, pp. 1–5.

University of Granada, "UGR'16 Dataset: Network Traffic Data for Intrusion Detection," 2016. [Online]. Available: https://nesg.ugr.es/nesg-ugr16/. [Accessed: Jun. 2026].