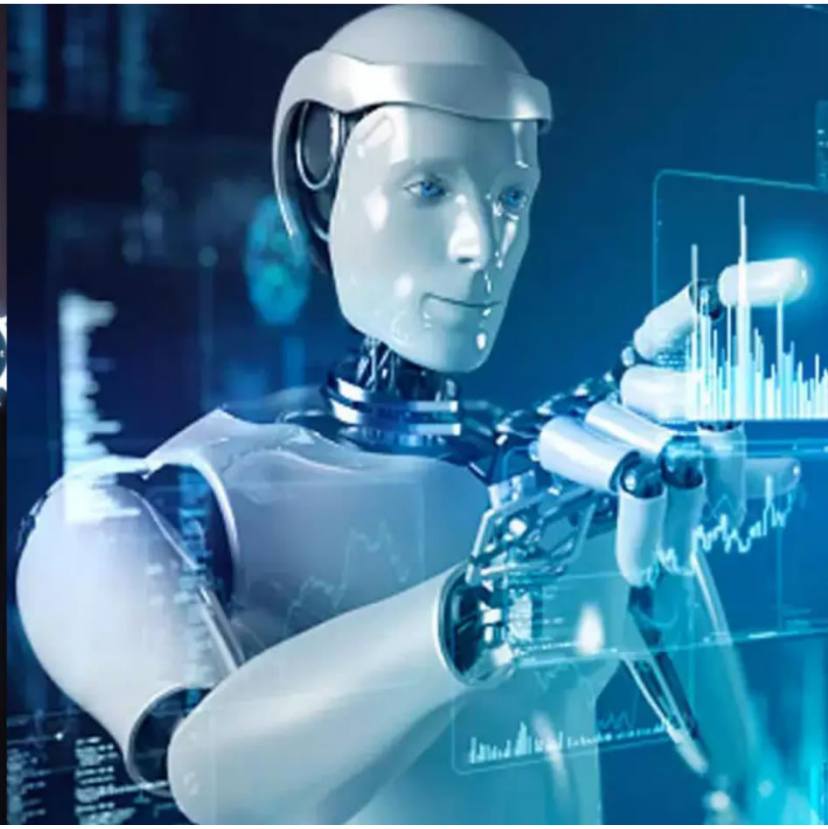
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*

**Impact Factor: 8.699**

**Volume 14, Issue 4, April 2025**

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404709|

# AI-Based Network Intrusion Detection System using Generative Adversarial Networks

**Chaudhari Komal[1], Prof. Joshi S. G.[2]**

P.G. Student, Dept. of Computer Engineering, Vishwabharati Academy's COE, Ahmednagar, Maharashtra India[1]

Assistant Prof., Dept. of Computer Engineering, Vishwabharati Academy's COE, Ahmednagar, Maharashtra, India[2]

**ABSTRACT:** As communication technology advances, various and heterogeneous data are communicated in distributed environments through network systems. Meanwhile, along with the development of communication technology, the attack surface has expanded, and concerns regarding network security have increased. Accordingly, to deal with potential threats, research on network intrusion detection systems (NIDSs) has been actively conducted. Among the various NIDS technologies, recent interest is focused on artificial intelligence (AI)-based anomaly detection systems, and various models have been proposed to improve the performance of NIDS. However, there still exists the problem of data imbalance, in which AI models cannot sufficiently learn malicious behavior and thus fail to detect network threats accu rately. In this study, we propose a novel AI-based NIDS that can efficiently resolve the data imbalance problem and improve the performance of the previous systems. To address the mentioned problem, we leveraged a state-of-the-art generative model that could generate plausible synthetic data for minor attack traffic. In particular, we focused on the reconstruction error and Wasserstein distance-based generative adversarial networks, and autoencoder-driven deep learning models. To demonstrate the effectiveness of our system, we performed comprehensive evaluations over various data sets and demonstrated that the proposed systems significantly outperformed the previous AI-based NIDS.

**KEYWORDS**: Anomaly detection, generative adversarial network (GAN), network intrusion detection system (NIDS), network security.

## I. INTRODUCTION

Mobile One of the fundamental challenges in cybersecurity is the detection of network threats, and various results have been reported in the field of network intrusion detection systems (NIDSs). In particular, the most recent studies have been focused on applying the artificial intelligence (AI) technology to NIDS, and AI-based intrusion detection systems have achieved remarkable performance. Initially, the research primarily focused on applying traditional machine learning models, such as decision trees [1] (DTs) and support vector machines [2] (SVMs) to existing intrusion detection systems, and it has now been extended to deep learning approaches [3], such as convolutional neural networks (CNNs), long shortterm memory (LSTM), and autoencoders. Although these results have achieved remarkable performance in detecting anomalies, there still exist limitations in deploying them in real systems. In general, most of the network flow data is normal traffic, and malicious behavior that can cause service failure occurs rarely. Moreover, within the category of malicious behavior, most of the data are well-known attacks, and specific types of attacks are extremely rare. Due to this data imbalance problem, AI models deployed in NIDS cannot sufficiently learn the char acteristics of specific network threats, and this may leave the network systems vulnerable to the attacks owing to the poor detection performance. In this study, to address this inherent problem, we propose a novel AI-based NIDS that can resolve the data imbal ance problem and improve the performance of the previous systems. To address the aforementioned problem, we leveraged a state-of-the-art deep learning architecture, generative adversarial networks [4] (GANs), to generate synthetic network traffic data. In particular, we focused on the reconstruction error and Wasserstein distance-based GAN architecture [5], which can generate plausible synthetic data for minor attack traffic. By combining the generative model with anomaly detection models, we demonstrated that the proposed systems outperformed previous results in terms of the classification performance.

The main contributions of the proposed approach can be summarized as follows. 1) By combining the state-of-the-art GAN model that can generate plausible synthetic data and measure the convergence of training, we show that the proposed system outperforms existing AI-based NIDS in terms of detection rate. 2) Through comparative experiments

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404709|

with various deep learning models, we present that the detection performance for rare attacks can be improved by apply ing our methodology it as a base module. 3) By experimenting with data sets collected from vari ous scenarios, we show that the proposed system can be effectively applied to real-world environments. The remainder of this article is organized as follows. Section II briefly reviews related research from the perspective of NIDS based on machine learning and deep learning approaches, and Section III provides a background with a focus on autoencoders and GANs. In Section IV, we describe our methodology and the proposed framework as well as the four main stages in detail. In Section V, we evaluate the proposed system in various environments and present exper imental results with detailed analysis. Finally, we present concluding remarks and future work directions of this study in Section VI.

## II. RELATED WORK

In the field of AI-based NIDSs, many studies have been conducted to apply machine learning and deep learning tech nologies as anomaly detection. Ingre and Yadav [10] proposed multilayer perceptron-based intrusion detection system and showed that the proposed approach achieve 81% and 79.9% accuracy in experiments on the NSL-KDD data set for binary and multiclassification, respectively. Gao et al. [11] proposed a semi-supervised learning approach for NIDSs based on fuzzy and ensemble learning and reported that the proposed system achieved 84.54% accuracy on the NSL-KDD data set. By applying the deep belief network (DBN) model, Alrawashdeh and Purdy, [12] developed an anomaly intrusion detection system and showed that the proposed DBN-based IDS exhibited a superior classification performance in subsampled testing sets (sampled subsets from the original data set). By considering the software defined networking environment, Tang et al. [13] proposed a DNN-based anomaly detection system and reported that the DNN-based approach outperformed traditional machine learn ing model approaches (e.g., Naïve Bayes, SVM, and DT). Imamverdiyev and Abdullayeva [14] proposed a restricted Boltzmann machine (RBM)-based intrusion detection system and showed that the Gaussian–Bernoulli RBM model outperformed other RMB-based models (such as Bernoulli-Bernoulli RBM and DBN). From the perspective of utilizing both behavioural (network traffic characteristics) and content features (payload information), Zhong et al. [15] introduced a big data and tree architecture-driven deep learning system into the intrusion detection system, where the authors combined shallow learning and deep learning strategies and showed that the system is particularly effective at detecting subtle patterns for intrusion attacks. With the ensemble model-like approach, Haghighat et al. [16] proposed an intrusion detection system based on deep learning and voting mechanisms. Haghighat and Li [16] aggregated the best model results and showed that the system can provide more accurate detections. Moreover, they showed that the false alarms can be reduced up to 75% compared to the conventional deep learning approaches. Considering data streams in industrial IoT environments, Yang et al. [17] proposed a tree structure-based anomaly detection system, where the authors incorporate the window sliding, detection strategy changing, and model updating mechanisms into the locality-sensitive hashing-based iForest model [18], [19] to handle the infiniteness of data streams in real-time scenario. Similarly, Qi et al. [20] proposed an intrusion detection system for multiaspect data streams by combining locality-sensitive hashing, isolation forest, and principal component analysis (PCA) techniques. Qi et al. [20] showed that the proposed system can effectively detect group anomalies while dealing with multiaspect data and process each data row faster than the previous approaches.

## III. BACKGROUND

In this section, briefly illustrate the concepts of autoencoders and GAN, which are key components of our anomaly detection system.

A. Autoencoder The autoencoder is one of the fundamental deep learning models and is trained with an unsupervised learning process. The objective of autoencoders is to return the output as close to the original input as possible. Therefore, the parameters are updated progressively during the training process to minimize the reconstruction error. In general, the architecture of an autoencoder consists of two components: 1) an encoder and 2) a decoder (see Fig. 1).

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404709|

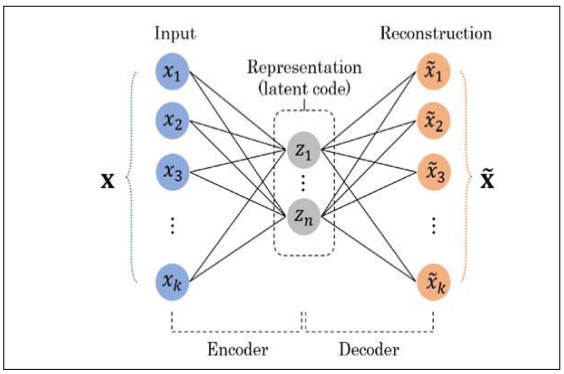



Fig 1: Basic Architecture Autoencoder

One of the fundamental characteristics of the autoencoder is to represent high-dimensional input data as lower dimensional information (summarized but meaningful information). Herein, we utilized autoencoders with the aim of feature extraction Fig. 3. Basic architecture of generative adversarial networks. (dimension reduction) on the input data. Although PCA has traditionally been utilized to project high-dimensional data into a lower dimensional space, we leveraged the autoencoders for nonlinear transformations on complex data sets. Although we only present the basic architecture of autoencoders, models can be built in multiple layers and an asymmetric manner.

B. Generative Adversarial Networks Generative models are designed to approximate the probability distribution of a training dataset and aim to generate synthetic data that is close to the real data (training data). Recently, among these generative models, research on GAN [4] has been of significant interest. Accordingly, various GAN models have been proposed to improve the performance and advance functionality. A GAN model consists of two neural network-based models: 1) a generator G and 2) a discriminator D (see Fig. 2).

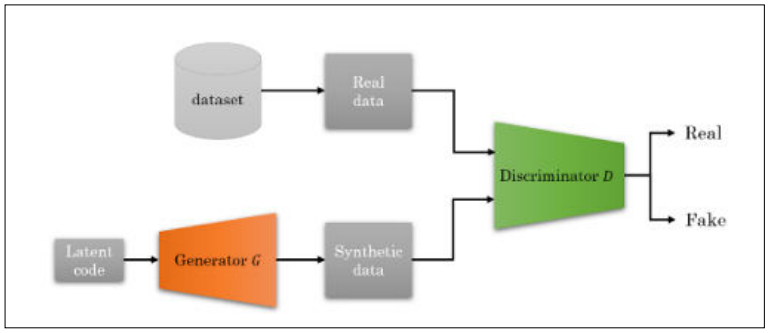



Fig 2: Basic Architecture of GAN

The generator G aims to generate synthetic data (fake data) that is close to the real data, while the discriminator D aims to discriminate between the real and fake data. In other words, these two components have opposing objectives during the training process. More formally, let pz and pdata be the probability distributions of the latent code and the real data, respectively.

Therefore, the discriminator is trained to output a higher confidence value in real data, and the generator is trained to generate synthetic data that can maximize the confidence score in the discriminator. After a sufficient number of iterations of this training process, both the discriminator and generator will settle to a point, where there is no scope for further improvement (i.e., a Nash equilibrium is achieved). Since the basic concept of the GAN model was introduced, numerous variants have been proposed to develop the original model by adjusting the objective function or by modifying the model architecture.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404709|

## IV. PROPOSED METHODOLOGY

As shown in Fig. 1, the entire architecture of the proposed AI-based NIDS consists of four main streams: 1) preprocess ing; 2) generative model training; 3) autoencoder training; and 4) predictive model training. In this section, we describe the proposed methodology and each module (process) in detail.

### C. Preprocessing

Before building and training AI models, the system refines a given raw data set via the preprocessing module that consists of three subprocesses: 1) outlier analysis; 2) one-hot encoding; and 3) feature scaling.

In the outlier analysis phase, the system eliminates outliers, which can negatively affect the model training. Typically, outliers are detected by quantifying the statistical distribution of the data sets via robust measures of scale. There are several standard robust measures of scale for detecting outliers, such as interquartile range (IQR) and median absolute deviation (MAD).

After filtering out the outliers, the system transforms nomi nal attributes into one-hot vectors. Each nominal (categorical) attribute is represented as a binary vector with the size of the number of attribute values, where 1 is assigned only to a point corresponding to the expressed value and 0 to all others. For example, in the case of the "protocol" attribute (commonly included in network traffic data) with the values tcp, udp, and icmp, the attribute is transformed into a binary vector of length 3, and the attribute values are converted into [1, 0, 0], [0, 1, 0], and [0, 0, 1], respectively.

In general, existing deep-learning-based approaches consider feature extraction (e.g., PCA, Pearson correlation coefficient, etc.) at this step to feed the model as many informative features as possible, and, consequently, feature extraction can significantly impact the performance of models in anomaly detection. However, we do not consider the computational feature extraction process, as our framework embeds an autoencoder model that can replace functionalities of feature extraction.

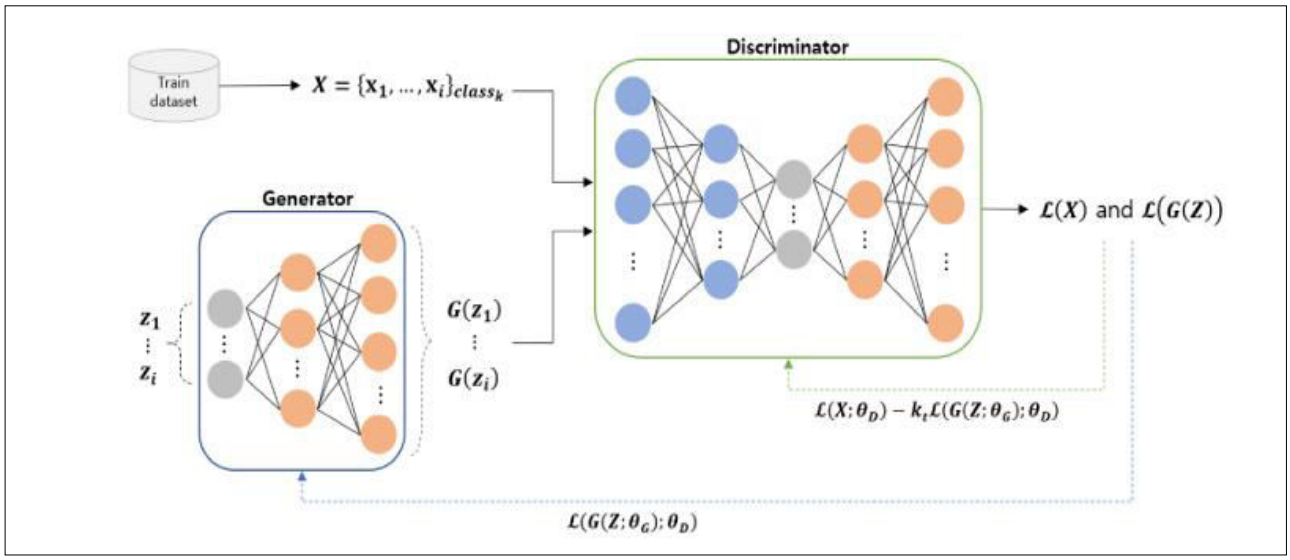



Fig 3: Architecture Of Generative Model in System

### D. Synthetic Data Generation With Generative Model

The synthetic data generation module builds and trains generative models using the data set refined in the data pre processing module. One of the important factors that must be considered when applying GAN models to NIDS is the determination of the termination criteria of training, which has a significant impact on the performance of anomaly

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404709|

detection, as it is directly related to the quality of the synthetic data to be trained on the detection model. The determination of the termination criteria stems from the tracking of the training convergence, and this is a difficult problem, as the objective function of GAN models is defined to have the properties of a zero sum game. In general, monitoring the training progress has been conducted indirectly through visual inspection of syn thetic (generated) data. However, even this approach is not feasible in NIDS environments because the data being han dled is not in the form of an image.

After training the generative model, the system generates synthetic data according to the classes using the trained gen erator and integrates the generated data set into the original training data set. This expanded data set is used to train the autoencoder and detection model in the next stage. Note that although we designed the synthetic data generation module to build multiple generative models according to the number of classes, it can be built as a single model by integrating the concept of the conditional GAN architecture [41], where class attributes are embedded in the input space.

E. Learning the Autoencoder and Detection Model

To build the intrusion detection model, the system first trains an autoencoder model that can provide feature extraction and dimensionality reduction functionalities. In our framework, we designed the autoencoder to possess the same architecture as the discriminator of the generative model. For detection models, we utilized the basic DNN, CNN, and LSTM as classifiers. We designed the DNN model to possess two hidden layers, and it could naturally process the refined network traffic data in terms of the model training and classification task. In the case of the CNN model, because the model was originally designed to be more suitable for analyzing image data, it required additional transformation processes in the input data space or the layers of the model depending on the approach followed. In our system, we built the CNN model with one-dimensional (1-D) convolutional layers to process the network traffic data, rather than converting the input data (i.e., network traffic data) into a 2-D space. Additionally, we subdivide the whole system into subsystems for a comprehensive comparison. In particular, we consider the DNN, CNN, and LSTM models as naïve deep learning models and DNNAE and CNNAE, which are models combined with the autoencoder, as advanced deep learning models. In the experiment, we conducted a comparative analysis of G-LSTM, G-DNNAE, and G-DNNAE with the subsystems.

## V. PROPOSED ALGORITHM

**Algorithm 1** Autoencoder Training With Generators

**Input**: training dataset $\mathcal{D}_{train}$, a set of generators $\mathbf{G}$

1: **Initialize** Autoencoder parameters $\theta_{AE}^{0}$
2: **for** $G_i \in \mathbf{G}$, where $1 \leq i \leq k$ **do**
3: sample $\mathbf{z} = \{z_j\}_{j=1,\ldots,m_i}$ from the latent space
4: $\hat{\mathcal{D}}_i = G_i(\mathbf{z})$
5: **end for**
6: $\tilde{\mathcal{D}} = \mathcal{D}_{train} \cup \hat{\mathcal{D}}_1 \cup \cdots \cup \hat{\mathcal{D}}_k$
7: $\theta_{AE} =$ **Train_Autoencoder**$(\theta_{AE}^{0}, \tilde{\mathcal{D}})$
8: $\theta_{enc} =$ **Extract_Encoder**$(\theta_{AE})$

**Output**: trained encoder $\theta_{enc}$

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404709|

**Algorithm 2** Classifier Training With Generators

**Input**: training dataset $\mathcal{D}_{train}$, a set of generators **G**, trained encoder $\theta_{enc}$

1: **Initialize** classifier parameters $\mathcal{W}^0$
2: **for** $G_i \in \mathbf{G}$, where $1 \leq i \leq k$ **do**
3: sample $\mathbf{z} = \{z_j\}_{j=1,\ldots,m_i}$ from the latent space
4: $\hat{\mathcal{D}}_i = G_i(\mathbf{z})$
5: **end for**
6: $\tilde{\mathcal{D}} = \mathcal{D}_{train} \cup \hat{\mathcal{D}}_1 \cup \cdots \cup \hat{\mathcal{D}}_k$
7: **Set** Trainable_State on $\theta_{enc}$ = **False**
8: **Build** $\mathcal{W}^0_{\theta_{enc}}$ = **Concatenate_Models**$(\theta_{AE}, \mathcal{W}^0)$
9: $\mathcal{W}_{\theta_{enc}}$ = **Train_Classifier**$(\mathcal{W}^0_{\theta_{enc}}, \tilde{\mathcal{D}})$

**Output**: trained classifier $\mathcal{W}_{\theta_{enc}}$

## VI. CONCLUSION AND FUTURE WORK

In this study, presentation of a novel AI-based NIDS that can efficiently resolve the data imbalance problem and improve the classification performance of the previous systems. To address the data imbalance problem, we leveraged a state-of the-art generative model that could generate plausible synthetic data and measure the convergence of training. Moreover, we implemented autoencoder-driven detection models based on DNN and CNN and demonstrated that the proposed models outperform previous machine learning and deep learning approaches. The proposed system was analyzed on various data sets, including two benchmark data sets, an IoT data set, and a real data set. In particular, the proposed models achieved accuracies of up to 93.2% and 87% on the NSL-KDD data set and the UNSW-NB15 data set, respectively, and showed remarkable performance improvement in the minor classes. In addition, through experiments on an IoT data set, we demonstrated that the proposed system can efficiently detect network threats in a distributed environment. Moreover, in order to investigate the feasibility in real-world environments, we collected real data from a large enterprise system and evaluated the proposed model on the collected data set. Through this experiment, we demonstrated that the proposed model can significantly improve the detection rate of network threats by resolving the data imbalance problem in the real environment. In the future, by considering practical distributed environments, we will focus on applying our framework to federated learning systems and ensemble AI systems to enhance network threat detection. In addition, we will study adversarial attacks that can bypass AI-based NIDS through vulnerabilities in AI models and conduct research on enhanced NIDS that can resist these attacks in real-world environments.

## REFERENCES

1. J. R. Quinlan, C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning). San Mateo, CA, USA: Morgan Kaufmann, 1993.
2. N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge, U.K.: Cambridge Univ. Press, 2000.
3. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
4. J. Goodfellow et al., “Generative adversarial nets,” in Proc. 27th Int. Conf. Neural Inf. Process. Syst. (NIPS), 2014, pp. 2672–2680.
5. D. Berthelot, T. Schumm, and L. Metz, “BEGAN: Boundary equilibrium generative adversarial networks,”2017, arXiv:1703.10717.
6. S. Hettich and S. D. Bay. “KDD cup 1999 data.” 1999. [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
7. M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed anal ysis of the KDD CUP 99 data set,” in Proc. IEEE Symp. Comput. Intell. Secur. Defense Appl., Jul. 2009, pp. 1–6.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404709|

8. N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set),"in Proc. Military Commun. Inf. Syst. Conf. (MilCIS), 2015, pp. 1–6.
9. Parmisano, S. Garcia, and M. J. Erquiaga, "A labeled dataset with malicious and benign IoT network traffic." 2020. [Online]. Available: https://www.stratosphereips.org/datasets-iot23
10. Ingre and A. Yadav, "Performance analysis of NSL-KDD dataset using ANN," in Proc. Int. Conf. Signal Process. Commun. Eng. Syst., Andhra Pradesh, India, Jan. 2015, pp. 92–96.
11. Y. Gao, Y. Liu, Y. Jin, J. Chen, and H. Wu, "A novel semi-supervised learning approach for network intrusion detection on cloud-based robotic system," IEEE Access, vol. 6, pp. 50927–50938, 2018.
12. K. Alrawashdeh and C. Purdy, "Toward an online anomaly intrusion detection system based on deep learning," in Proc. IEEE 15th Int. Conf. Mach. Learn. Appl. (ICMLA), Anaheim, CA, USA, 2016, pp. 195–200.
13. T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho, "Deep learning approach for network intrusion detection in software defined networking," in Proc. Int. Conf. Wireless Netw. Mobile Commun. (WINCOM), 2016, pp. 258–263.
14. Y. Imamverdiyev and F. Abdullayeva, "Deep learning method for denial of service attack detection based on restricted Boltzmann machine," Big Data, vol. 6, no. 2, pp. 159–169, Jun. 2018.
15. W. Zhong, N. Yu, and C. Ai, "Applying big data based deep learn ing system to intrusion detection," Big Data Min. Anal., vol. 3, no. 3, pp. 181–195, Sep. 2020.
16. M. H. Haghighat and J. Li, "Intrusion detection system using vot ingbased neural network," Tsinghua Sci. Technol., vol. 26, no. 4, pp. 484–495, Aug. 2021.
17. Y. Yang et al., "ASTREAM: Data-stream-driven scalable anomaly detec tion with accuracy guarantee in IIoT environment," IEEE Trans. Netw. Sci. Eng., early access, Mar. 8, 2022, doi: 10.1109/TNSE.2022.3157730.
18. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," ACM Trans. Knowl. Discov. Data, vol. 6, no. 1, pp. 1–39, Mar. 2012.
19. X. Zhang et al., "LSHiForest: A generic framework for fast tree isolation based ensemble anomaly analysis," in Proc. IEEE 33rd Int. Conf. Data Eng. (ICDE), Apr. 2017, pp. 983–994.
20. L. Qi, Y. Yang, X. Zhou, W. Rafique, and J. Ma, "Fast anomaly identification based on multi-aspect data streams for intelligent intru sion detection toward secure industry 4.0," IEEE Trans. Ind. Informat., vol. 18, no.9, pp. 6503–6511, Sep. 2022.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details