Enhancing Cybersecurity through Integrity Assessment and Intrusion Detection Using ACGAN-Powered Machine Learning for Unbalanced Data

# B Mukesh Kumar

Department Of Computer Science Engineering St Joseph’s College Of Engineering

Chennai, India

# R Nidhis Kumar

Department Of Computer Science Engineering St Joseph’s College Of Engineering

Chennai, India

# Dr. V Muthulakshmi

Head Of The Department Computer Science Engineering St Joseph’s College Of Engineering

Chennai, India

***Abstract*— This study investigates the use of Auxiliary Classi- fier Generative Adversarial Networks (ACGANs) in addressing imbalanced data within the realm of cybersecurity. ACGANs generate synthetic data mimicking network attacks, contribut- ing to dataset balancing for improved model training. The research focuses on enhancing cybersecurity decision making by refining the accuracy of distinguishing between legitimate and malicious traffic. By leveraging ACGAN-powered machine learning, this project work demonstrates the potential for stronger, more accurate threat detection and integrity assess- ment, ultimately fostering more advanced and resilient intrusion detection systems.**

## INTRODUCTION

The fast-changing field of predictive analytics today, es- pecially in criminology and network security, highlights the growing need for advanced methods to solve tough chal- lenges. In crime prediction, two key focuses are identifying crime risk zones and spotting crime hotspots. Predicting crime risk zones means looking at how environmental and situational factors affect criminal activities. Researchers often use such concepts as routine activity theory in their search for hotspots connecting specific environmental features and the potential for crimes to happen. In this approach, it pools environmental information along with the history of the com- mission of crimes in the area to look for potential hotspots with high risks. The predictive model for crime hotspot puts much focus on specific points that could potentially host crime. This service tends to use a compilation of historical crime data alongside real-time trends that would suppos- edly foreshadow areas most probably bound to experience incidents shortly in the near future. In any case, however, most of these models presume a given past pattern should go unreformed. This sometimes fails, and it becomes too heavy for these models to sustain as criminal activities vary perpetually in nature. This lack of adaptability can be a challenge in scenarios where crime patterns change rapidly due to different factors. Likewise, in network security, it is tough to detect malicious activities because datasets are generally imbalanced. Most network behavior is benign, and only a small fraction of it are malicious, such as hacking or security intrusion. The imbalance can make the traditional detection mechanisms become overfit to benign activity, and therefore less responsive to rare but more dangerous threats. To address this problem, Auxiliary Classification Generative Adversarial Networks offer a promising solution.

Deep learning ACGANs generate synthetic data, and that includes fake scenarios of attacks to enrich Datasets. This provides some degree of data balancing. It also ensures more diversity in attack patterns through machine learning models. The efficiency of predictive models integrating ACGANs is highly significant in regards to the recognition and mitigation of threats by network security systems. While accuracy improves, the decision-making ability also is better which allows the systems to handle risks more effectively. This is a marked method wherein crime prediction methods improve while innovating the network security with novel data gen- eration techniques. An important step in predictive analytics, this marks a landmark development. These modern methods give criminology and cybersecurity better tools to respond to their challenges. As they continue to develop, they’re likely to make predictive models more accurate, flexible, and reliable thus leading to stronger crime prevention as well as cybersecurity solutions.

## METHODOLOGIES

1. *objectives*

There seems to be a major justification towards Auxiliary Classifier Generative Adversarial Networks in network se- curity due to the resolution of issues related to imbalanced datasets being a major hindrance that prevents the detection of threats as cyber with precision. Malicious activity is not largely there in network settings compared to the benign traffic, which further makes the dataset highly unbalanced towards non-malicious data. This imbalance prevents tradi- tional machine learning models from accurately identifying and responding to rare but critical security threats, such as cyber attacks, data breaches, or attempts at unauthorized access. Utilizing ACGANs is a promising solution as they can generate synthetic data that closely resembles real- world attack patterns. ACGANs help in balancing the dataset by including a greater volume of attack data, so machine learning models are exposed to a more representative mix of both benign and malicious network behaviors. Augmentation not only addresses the problem of class imbalance but also makes the learning process more effective, as it makes the models better at differentiating normal network traffic from potential threats. Consequently, the machine learning algorithms that are trained on this improved dataset will be more accurate in the detection of threats because there is

a lesser chance of missing some rare instances of attacks. Furthermore, ACGANs assist in generating more robust and resilient models for security, capable of handling a wide range of cyber threats. These advanced models are more capable of detecting new or unknown patterns of attacks, thus making the network security systems more capable of real- time identification and mitigation of risks. Addressing the basic problem of unbalanced data, ACGAN-based machine learning integration into the field of network security has led to more accurate, adaptive, and proactive defense mech- anisms that can significantly strengthen the security posture of organizations and their networks. This approach not only optimizes the detection of malicious activity but also supports more informed decision making, thus allowing security teams to respond with greater precision and agility toward potential threats. Ultimately, ACGAN-based techniques might end up revolutionizing how network security systems operate, mak- ing them effective in an increasingly complex and dynamic threat landscape.

1. *Problem Statement*

This is a problem study concerning data imbalance in network security; traditional methods are incapable of dis- tinguishing normal network behavior from malicious attacks. Such imbalanced datasets result in a loss of effectiveness for the detection of threats and responses. Further, it im- pacts decision-making as far as security practitioners are concerned. The current research is an attempt to make use of Auxiliary Classifier Generative Adversarial Networks to develop synthetic attack data for making datasets balanced and more effective towards the identification and mitigation of network threats.

## ALGORITHMS USED

*A. Auxiliary Classifier Generative Adversarial Networks (ACGANs)*

The auxiliary classifier is utilized in an extension to the original Generative Adversarial Network architecture; these are known as ACGANs. This combines the discriminator with the assistance of an auxiliary classifier that would en- hance the capabilities of a typical GAN model by compelling it to discriminate on one more class for both the discriminator and generator to converge more strongly towards synthetic and real data respectively. This dual-purpose design enables the discriminator to provide richer feedback as it is evaluating whether the data is authentic, and the categorical accuracy for the data. In training, the generator will try to make the generated data indistinguishable from real samples by the discriminator and correctly classify them using the auxiliary classifier. Improvements in data quality and diversity are outcomes of such a training. Specifically, ACGANs are best suited for the usage where imbalanced datasets need to be handled because synthetic underrepresented class instances could be generated to tackle such shortages and improve machine learning model performance. Integrating classifica- tion tasks into the GAN model contributes a lot towards more realistic and useful generation of data, hence in effect

improving the effectiveness in overall predictive models and also applications based on data-driven operations.

Auxiliary Classifier Generative Adversarial Networks is promising for solving imbalanced data challenges in network security. An imbalanced dataset, such as a dataset containing sparse instances of malicious activities compared to normal traffic, often affects machine learning models, making it difficult to identify and classify network threats. ACGANs, designed to generate synthetic data, can also produce more examples of attack scenarios not sufficiently represented in the dataset. With this augmentation of the training data with the generated samples, the model will be better trained on a more balanced dataset, which leads to better detection capabilities. This improvement in data balance has further made the machine learning models discern the difference between benign and malicious network behavior better. So, it would really make the ACGANs applied with significant strength in making the threat detection system more accurate and reliable so that decision-making and response strategies of networks are highly effective. This approach not only deals with the problem of data imbalance but also adds to the overall robustness of network security measures by providing models that are more effective in recognizing and responding to potential security threats.

## SYSTEM DESIGN

By synthesizing data mimicking almost any network at- tack, ACGANs are effective in offsetting imbalanced nor- mal vs. malicious traffic. Enhanced datasets yield improved model trainings; hence, assured enhanced accuracy towards threat detection. The deployment of generated data from the ACGAN in a machine learning-based setup is aimed to reduce the amount of false positives as well as the false negatives in a quicker adaptation to dynamic attack behavior with enhancing decision-making processes during real time network security procedures. Synthetic augmentation enables the system to train models on a more representative set of data. This implies benign and malicious net flow traffic. Therefore, improving the accuracy of detection made by a model does depend on how the dataset may improve.

1. *Hardware Setup*

The hardware requirements include compatibility with Windows operating systems (Windows 7, 8, and 10) for both 32-bit and 64-bit architectures, along with a minimum of 4GB RAM

1. *Software Setup*

Python or Anaconda Navigator is needed to program. In specific, the development environment should support Python language, and Jupyter Notebook is recommended to use for interactive coding and data analysis tasks. These software components enhance the efficient development and execution of projects and data analysis workflows based on Python language.

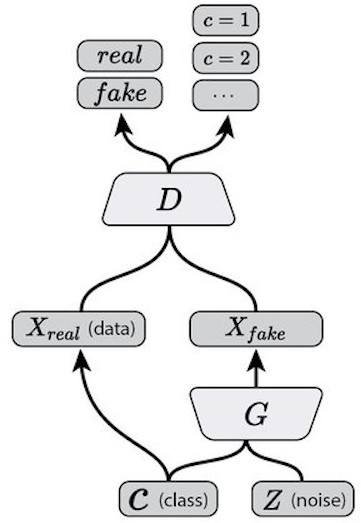


Fig. 1. Architecture of Auxilary Classifier generative Adversarial Networks

1. *Collection of Data*

Systematic collection of numerous data that reflects nor- mal operation of the network and malicious activities. It involves, first of all, using many network monitoring tools, intrusion detection systems, and historical logs from network appliances so as to gain the most extensive spectrum of network traffic, which may include typical communication patterns and security threats. This step ensures that the dataset captures both legitimate and abnormal activities, which would reflect the complex nature of real-world net- work environments. A good model requires a diversified and balanced dataset. It has to contain a wide variety of network behaviors, including various protocols, traffic volumes, and attack vectors. This diversity enables the model to learn and discern specific patterns that characterize normal traffic and those related to network anomalies. In addition, actual malicious activity within the system ensures that the model is trained and exposed to the entire breadth of potential threats, from simple attacks to sophisticated attacks, ensuring it de- tects most possible security risks. Quality and completeness of data are the crucial characteristics that ensure a successful model. A complete and accurate dataset forms a premise for the building up of strong machine learning algorithms and,

consequently, would easily distinguish between good or bad traffic.

1. *Pre-processing of the Data*

After data collection, it is a major pre-processing stage for preparation for further analysis. Some of the main activities during this step include data cleaning, normalization, and transformation. Data cleaning handles missing values, noise, and inconsistencies in the data set, while normalization scales all the features into a uniform range so that machine learning algorithms may work more effectively. Additional transfor- mation of the data can be performed for modifying the raw data into an appropriate format for extracting features. Proper preprocessing is essential to enhance the quality of the data as well as ensure that subsequent analysis and model training are carried out using correct and representative information.

1. *Data Cleaning*

The process of data cleaning should be performed in such a manner that it enhances the quality of the dataset by eliminat- ing missing values, noise, and inconsistencies that can affect the performance of the model. Incomplete and erroneous data may lead to bias or inaccuracy in the predictions of the models. Therefore, these techniques include imputation for missing values, identification, and correction of outliers, filtering of irrelevant information to ensure complete and reliable data. Data cleaning addresses all these problems with the improvement of the quality of the dataset and how well the model is to learn and produce the best results.

1. *Normalization*

Normalization refers to the scaling of the features in the data so that each attribute would be able to contribute equally while the model is training. In a dataset, there are features with different units or magnitudes. Therefore, during learning, some features become dominating and neglect the other ones. Data normalization typically applies min-max scaling or z-score standardization for this purpose. This step has improved the rate of convergence for machine learning algorithms, and it allows a model to perform at optimum levels without any feature biasing the model’s decisions.

1. *Transformation*

Data transformation is the process of transforming raw, unstructured data into a format more amenable to fea- ture extraction and input into models. This can be encod- ing categorical variables, discretizing continuous features, or aggregating data into more informative representations. Complex and raw inputs are made accessible to machine learning algorithms through transformation of the data. Right transformations enable models to pick on the underlying patterns within a more effective way in data thus leading to effective training capabilities and improved predictive capability, which is an important activity in preparing data in meeting the requirements of selected modelling techniques.

1. *Feature Extraction*

This step is also a central phase in the data pipeline, as extraction of features occurs by discovering significant attributes and characteristics in contribution to the model. This phase is vital and yet it processes network traffic data to come up with meaningful features for recognition between the normal and attack traffics. Techniques such as statistical analysis, pattern recognition, and domain-specific algorithms are adopted to extract features that signify a wide range of network behaviors. This process diminishes the dimensionality of the data while emphasizing relevant features that enhance the performance of the model in the discrimination of different types of network activities.

1. *Synthetic Data Generation Using ACGANs*

ACGANs address the imbalanced dataset. This framework creates synthetic samples using the generator that resemble rare attack scenarios with a very high fidelity. Simultane- ously, the discriminator would judge the authenticity of those samples and classify them under pre-defined classes such as attacks or normal traffic. Thus, the quality of the synthetic data generated along with its diversity ensures effective balancing of the dataset. Adding these realistic synthetic instances to the training data makes the system improve the model’s ability to identify rare and evolving threats.

1. *Model Selection and Training*

The machine learning model is trained using a balanced dataset that contains both real and synthetic samples. Several algorithms are tested to find the best one for network traffic classification, and deep learning models often perform better in capturing intricate patterns. Synthetic data generated by ACGANs enhances the ability of the model to differentiate between legitimate and malicious traffic, thereby improving its generalization across unseen network scenarios.

During model selection, multiple algorithms of machine learning are being compared based on how they do in classifying traffic for a network. Models like a decision tree, support vector machines, and neural networks have to compete to become the most accurate and efficient model. Computational complexities, train time, or the fact that it can handle imbalanced data are some of the considerations in choosing the best-performing model. The idea is to choose a model which performs well on historical data and generalizes well to novel network traffic scenarios.

1. *Evaluation and Validation*

This step will be critical in the evaluation of the model selected in terms of performance and its effectiveness in a real-world application. The cross-validation, performance on unseen data, and other metrics such as accuracy, precision, recall, and F1-score are used to measure how well the model can identify and classify threats in the network. This also includes consideration of the model’s response to various attack scenarios and how it can change in response to evolving threats. This evaluation ensures that it is thoroughly tuned and fully capable of providing accurate, timely security insight.

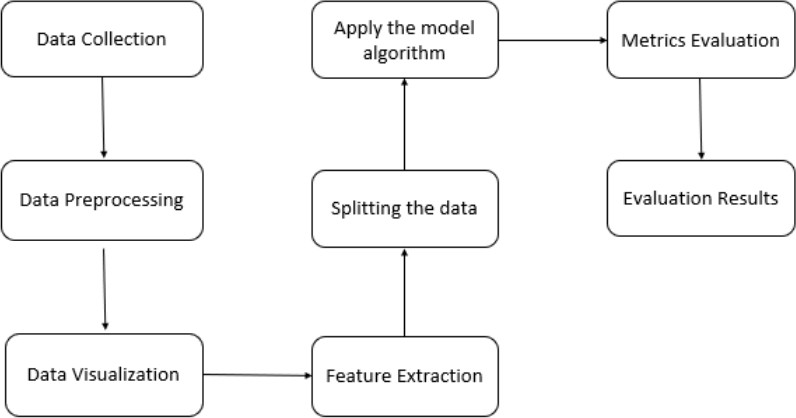


Fig. 2. Training Model Architecture

1. *Deployment and Real-Time Decision Making*

After validation, this model is put into the network en- vironment of the real world; the system monitors traffic persistently. It is in a position to minimize both false positives and false negatives because it has incorporated generated data from ACGAN, and the system enables accurate quick decisions.

The model is made to change dynamically with new and changing threats. This would ensure that the model stays effective over time. The flexibility ensures that the system gives continuous, real-time defense against emerging security challenges.

1. *System Monitoring and Iterative Updates*

After deployment, the system is continuously monitored to measure its performance over time. Information about newly encountered threats is gathered through feedback loops, which are used to retrain and fine-tune the model periodically. This means that the intrusion detection system remains effective and evolves along with the threat landscape.

1. Existing System

The current landscape of network security systems shows tremendous shortfalls, especially in dealing with imbalanced datasets; this is one of the most common issues related to intrusion detection. Most of these datasets contain a significant amount of normal network traffic as opposed to malicious activity instances. Such imbalances prevent conventional machine learning models from distinguishing between benign and malicious behavior effectively. Such unevenness results in distorted training processes in models as they focus heavily on the majority class, normal traffic, while neglecting to focus adequately on the minority class, network attacks. This usually results in the occurrence of more false negatives, meaning actual threats might not be de- tected, or false positives in which legitimate traffic is tagged as malicious, thus damaging the reliability and effectiveness of these systems that jeopardize network integrity and higher critical infrastructures. Thus, the methodologies employed so far exacerbate the problem as they rely on manually-curated datasets or small, static sets of samples from network traffic.

The main datasets thus don’t portray the dynamic and constantly changing nature of cyber threats and render mod- els underprepared to meet actual complexities of the world. Frequency: these types of attacks are not only growing, but also becoming much advanced while adopting techniques that exploit loopholes in the detection system. Traditional machine learning algorithms being unable to adapt quickly suffer from poor generalization properties to unseen attack patterns due to this static nature leads to systems that react too often and rarely proactively change, hence providing a static system with little or no adaptability, thereby very ineffective in keeping up with constantly emerging threats. It has more serious issues because of non-diverse and represen- tative training data the systems have generally compromised decision-making processes.

The practitioner is left working through the models that produce unreliable predictions, which creates inefficiency in operations and delays in response to suspected breaches. This does not allow differentiation of small deviations in network behavior; this makes such a case even more complicated because the sophisticated attacks will mimic legitimate traffic patterns to avoid being detected. Hence, serious security incidents either go unobserved or would necessitate manual intervention which is both time-consuming and susceptible to human error. Overall, the system existing in network security reflects a serious gap in its capacity to address the changing landscape of threats. The use of outdated methodologies, together with problems arising from unbalanced datasets and static models, is what limits the effectiveness of intrusion detection systems. Present deficiencies in these approaches necessitate novel methods that will be able to overcome current barriers, providing solutions robust, adaptive, and scalable enough to face the complexities of modern network environments.

*A. Drawbacks*

* Imbalanced Dataset.
* Limited Model Performance
* Static and Data Model collection.
* Reduced Adaptability.

1. Proposed System:

The proposed system introduces an innovative approach to network security in using Auxiliary Classifier Generative Adversarial Networks. An ACGAN generates synthesized data that looks like multiple network attacks. It works on the imbalance between normal and attack traffic in networks. This synthesized data improves accuracy in model training, bettering the identification of threats. The system is designed to minimize false positives and negatives, increase adapt- ability to evolving attack patterns, and strengthen real-time decision-making in network security by integrating ACGAN- generated data into machine learning models.

The architecture of ACGANs is especially suitable for this task because it includes an auxiliary classifier along with the traditional adversarial framework. This dual functional- ity allows the discriminator to simultaneously evaluate the

authenticity of generated data and classify it into predefined categories, such as specific types of network attacks. As the generator iteratively learns to produce data that is both real- istic and correctly classified, the resulting synthetic samples are highly representative of real-world network threats.

This process not only enriches the data diversity but also ensures the training dataset is representative of modern cyberattacks as complex and dynamic. This enriched data set enables the machine learning models to recognize subtle and nuanced patterns associated with malicious activities and thus, improve their ability to detect and remove po- tential threats. Besides balancing the data, adding ACGAN- generated samples to machine learning workflows signifi- cantly improves the performance of models. The models become more accurate in their predictions as they are trained on a more representative dataset, thus reducing the chances of false positives and false negatives. This is essential in practi- cal deployment where false negatives might allow undetected intrusions and false positives may allocate resources unnec- essarily, causing inefficiency in operations. The strength of the proposed system to reduce these errors will fortify the defense in the network and build confidence for the security practitioner in an automated detection system.

The proposed approach is also adaptive. Cyber threats evolve; attackers now use sophisticated ways of avoiding detection. Since the synthetic samples produced by ACGAN are inherently diverse and dynamic, this system proves effec- tive against emerging attack vectors. Continuously updated models increase adaptability due to incorporation of newly produced data that enhances detection over time. Thus, in a changing threat landscape, this system proposed does not become static but rather changes over time as part of an active defense mechanism predicting and responding to emerging vulnerabilities.

While technical performance is a valid criterion, real- time decision-making is a need required by modern network security. The system allows the enhanced dataset and im- proved machine learning models to make faster and more accurate threat identification possible; this enables security teams to react in time to incidents. This real-time capability is complemented by the scalability of the system, which ensures it will operate effectively within large and complex network environments without impacting performance. The design of the system ensures that high volumes of data can be handled in its deployments, making it suitable for enterprise- grade infrastructures and other high-stakes environments. In summary, the proposed system is a paradigm shift in network security with the use of ACGANs to address persistent challenges with data imbalance, adaptability, and real-time threat detection.

It generates high-fidelity synthetic data and improves ma- chine learning model performance for a robust, scalable, and proactive defense against a wide spectrum of cyber threats. It is the kind of system that reduces errors, keeps track of new emerging attack patterns, and facilitates rapid decision- making in the pursuit of delivering an integrated solution per- fectly matched to the requirements of today’s cybersecurity.

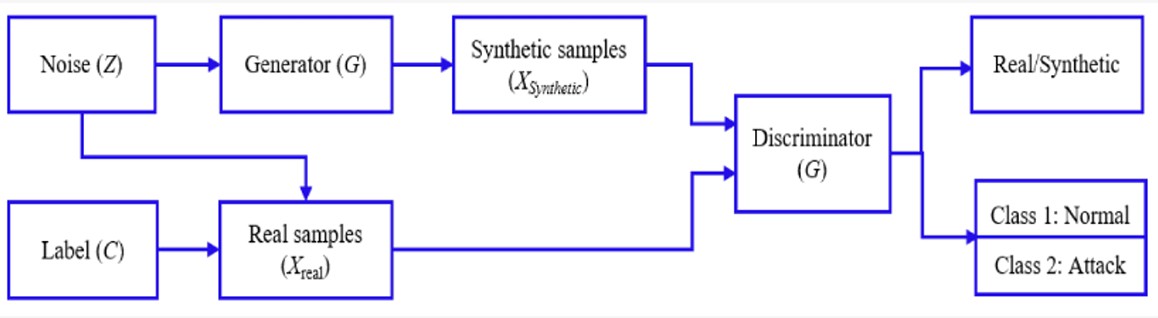


Fig. 3. Working of ACGAN

This is not only strengthening the already-existing defense systems but also paving the way toward next-generation intrusion detection systems capable of safeguarding critical assets within the ever-hostile digital landscape.

*A. Advantages*

* Improved Accuracy
* Addressing Imbalanced Data
* Increased Detection Capability
* Real-time Decision-making

## FUTURE ENHANCEMENT

The fast-evolving nature of cyber threats demands ad- vanced, as well as flexible solutions in protecting network infrastructures. To meet these challenges proposed the system applies Auxiliary Classifier Generative Adversarial Networks and provides a new way for the synthesis of data replicating different network attack scenarios that could be of greater importance for tackling one of the biggest challenges in Network intrusion detection-imbalanced datasets. They make most traditional traffic datasets overrepresent normal traffic compared to the attack traffic, thus developing biased models and decreasing their ability to detect attacks. An ACGAN will alleviate these problems by generating realistic traffic attacks, thus providing more balanced and representative data for training.

The addition of ACGAN-generated synthetic data into ma- chine learning-based intrusion detection models has several benefits. The enhanced dataset for training models improves accuracy as well as robustness. These models are more sensitive to the subtleties of both normal and malicious network behavior, thus making them more sensitive and specific. This reduces both false positives—where benign traffic is incorrectly flagged as malicious—and false nega- tives—where actual threats are overlooked. This is important in modern network environments where the minimization of detection errors is critical to maintaining operational security. In addition, the adaptability of ACGANs to the changing nature of attack patterns ensures that the system remains effective against emerging threats. The continuous generation of synthetic data reflecting new attack strategies helps keep the models relevant and responsive to the dynamic nature of cyber threats. This flexibility enhances the overall resilience of network security, providing a proactive approach to identifying previously unseen attack vectors.

The proposed system would significantly contribute to real-time network security decision-making. When the data generated by the ACGAN is used by the intrusion detection

models, it will analyze live network traffic more accurately and with better speed. This will make the system have more potential in detecting threats and mitigating damage from cyberattacks in real-time. Such integration of advanced techniques will not only strengthen the immediate defense of the network but also lay the groundwork for more adaptive and intelligent security strategies in the future. In a nutshell, the innovative use of ACGANs in this system provides a transformative approach to addressing long-standing chal- lenges in network intrusion detection. This methodology provides for a scalable solution to enhance network security by balancing datasets, increasing detection accuracy, and adapting to emerging threats.

## CONCLUSIONS

In short, ACGANs provide a robust approach to network security’s The imbalanced data challenges are improved by their ability to generate synthetic attack data, thus en- hancing model accuracy, adaptability to evolving threats, and decision-making speed. Promising as this may seem, further research is required for optimization and ethical considerations. Overall, the integration of ACGANs marks a significant stride in fortifying network security against di- verse threats and ensuring more robust defense mechanisms.

References

1. 1.J. P. Anderson, Computer security threat monitoring and surveil- lance, 1980
2. .J. H. Lee and K. H. Park, "AE-CGAN model based high performance network intrusion detection system", Appl. Sci., vol. 9, no. 20, pp. 4221-4229, Oct. 2019.
3. K. Wu, Z. Chen and W. Li, "A novel intrusion detection model for a massive network using convolutional neural networks", IEEE Access, vol. 6, pp. 50850-50859, 2018.
4. W. Zhendong, L. Yaodi and H. Zhongdong, "Optimize intrusion detection of BP neural networks with improved gray wolf algorithms", Small Microcomput. Syst., vol. 42, no. 4, pp. 875-884, 2021.
5. F. Folino, G. Folino, M. Guarascio, F. S. Pisani and L. Pontieri, "On learning effective ensembles of deep neural networks for intrusion detection", Inf. Fusion, vol. 72, pp. 48-69, Aug. 2021.
6. .A. Ahmim, M. Derdour and M. A. Ferrag, "An intrusion detection system based on combining probability predictions of a tree of classifiers", Int. J. Commun. Syst., vol. 31, no. 9, Feb. 2018.
7. W. Daozhou, Design and Implementation of Network Intrusion De- tection System Based on Linux Platform, Wuhan, China:Hust, vol. 5, 2019.