Computer Fraud and Security
ISSN (online): 1873-7056

Weakly Supervised Traffic Identification Method based on Fast Detection of Angle Outliers

Xiaolei Liu[1], Xiaohu Wu[1*], Mingyuan Zhang[1], Xiaojian Zhang[1]

[1]Jiangsu Electric Power Information Technology Co. Ltd, Nanjing, China
Corresponding Author's Email: 15895959284@163.com

Abstract: Current information and communication systems are facing network attacks with strong unknown characteristics, and existing network attack detection methods often fail to effectively detect unknown attacks. To address this limitation, this paper develops a novel weakly supervised traffic detection method named CTFABOD, designed to enhance the detection of unknown network attacks in contemporary information and communication systems. CTFABOD method uses generative adversarial network to enhance the performance of fast detection of angle outliers, to achieve precise detection of unknown network attacks. This paper tests the CTFABOD model on the classical NSL-KDD dataset, selects four classic weakly supervised models as comparison models, and uses AUC and precision as the evaluation indicators. The experimental results show that CTFABOD method has achieved the highest AUC and precision scores on the NSL-KDD dataset, and increased the precision score by 10.31%. This study highlights the effectiveness of combining angle-based outlier detection with generative adversarial network in improving the detection of unknown network attacks and suggests promising applications in various high-dimensional data analysis tasks.

Keywords: unknown network attacks; generative adversarial network; intrusion detection; weak supervision

1. Introduction

Cloud computing, Internet of Things (IoT)[1], Blochchain based architectures and Blockchain based Industrial Internet of Things (IIoT)[2] and contemporary information and communication technology (ICT)[3], these are some of the leading technologies the societies and countries are using these days. There is a consensus that contemporary ICT systems, including industrial control systems, medical support systems, virtual environments, and the internet of things, are prime targets for potential attackers [4,5]. Substantial existing evidence regarding the susceptibility of these systems to cyberattacks encompasses the probability of being targeted as well as the substantial costs and impact associated with successful attacks[6]. Outcomes of a successful cyberattack can vary from breaches in confidentiality to decreased availability or the loss of critical data, leading to integrity concerns [7]. It is crucial to note that security threats may extend to safety implications. For instance, an attack aimed at disabling a vehicle's automatic braking system could have severe repercussions on the driver's well-being, infrastructure, and environment. Consequently, systems need to be planned, developed, and executed with a focus on meeting the necessary security standards.

Intrusion Detection Systems (IDS) are widely recognized as an effective means of promptly identifying attacks[8]. When tasked with safeguarding a target system, an IDS continuously monitors various performance metrics, such as memory usage, bus throughput, active sessions, and system calls. The set of values being monitored by an IDS at any given moment is referred to as a data point, and these data points are typically collected and organized into a tabular dataset[9]. Machine learning (ML) algorithms are incorporated within IDS to facilitate binary classification, allowing them to differentiate between data points corresponding to attacks and those representing normal system behaviour. The ML algorithm undergoes a training phase during which it analyses a training dataset, enabling it to develop a model that can later be deployed in a production environment to detect attacks occurring in real-time.

Outlier detection is a fundamental task in data mining, aiming to pinpoint data objects that deviate from the typical data distribution[10]. Its applications encompass diverse fields such as fraud detection, scientific data error rectification, and sports data analysis. Successful outlier detection includes identifying stylistic elements from various sources in written works as an indication of potential plagiarism, or uncovering biases in scientific data as a sign of equipment failure, human error in data processing, or suboptimal experimental setups. The algorithmic approaches to outlier detection are as varied as the scenarios in which they are applied. Prominent and highly effective methods hinge on density estimation using k-nearest neighbour distances, whereby the point with the lowest density estimate is identified as the most compelling outlier candidate[11]. An inherent challenge in outlier detection is the degradation in the quality of density estimation as the dimensionality of the data increases. Numerous studies have examined various facets of the "curse of dimensionality," including phenomena such as the concentration of distances[12].

Outlier detection refers to the process of identifying samples that significantly deviate from the overall

representation within a dataset. This task can be approached through supervised or unsupervised methods, with this paper primarily focusing on the use of unsupervised models [13]. The fundamental concept of outlier detection revolves around identifying data objects that do not align with the general data distribution. This is a critical task within data mining and holds significant applications across various fields, such as uncovering instances of credit card abuse in financial transaction data or pinpointing measurement errors in scientific data. The rationale behind this lies in the fact that data objects are typically generated by specific mechanisms or statistical processes. Observable deviations from the primary distribution are thought to stem from distinct mechanisms. These mechanisms could encompass fraudulent activities, disturbances that compromise sensor integrity, or simply inaccuracies in measuring device readings. However, they could also represent unexpected behaviors that necessitate adjustments to the underlying theory of the relevant experiment.

The computational complexity of certain methods renders them impractical for high-dimensional data. However, all known techniques that are theoretically applicable to high-dimensional data are based on evaluating queries or k-nearest neighbors using local methods. Consequently, these methods are generally considered unsuitable for high-dimensional data due to the "curse of dimensionality." One prominent consequence of this challenge in mining high-dimensional data is the diminishing significance of concepts like proximity, distance, or nearest neighbor as the dataset's dimensionality increases. In general, it has been demonstrated that the relative contrast between the farthest and closest points approaches 0 with increasing dimensionality d:

$$\lim_{d\to\infty} \frac{dist_{max} - dist_{min}}{dist_{min}} \to 0 \quad (1)$$

This indicates that within high-dimensional spaces, the distinction between nearest and farthest neighbors becomes considerably less reliable. These observations hold true for a broad spectrum of data distributions, independent of the relevance of all attributes, and are solely based on the number of dimensions. Additionally, the problem is compounded by the presence of irrelevant attributes, often referred to as "noise," in high-dimensional data. However, global feature reduction methods may not suffice to eliminate noisy attributes, as there is typically no universal noise present; rather, certain attributes are only influential for specific sets of objects. All these effects extend beyond mere complexity issues and have prompted the search for data mining methods that are less reliant on distances between objects. Angle-based outlier detection (ABOD) algorithms still take distance into consideration, but primarily as a secondary measure to normalize the results, placing greater emphasis on the variance of angles between different vectors of the data objects. This approach exhibits much lower sensitivity to the escalation in dataset dimensionality compared to distance-based criteria.

Despite these advancements, existing network attack detection methods still face challenges in effectively detecting unknown and sophisticated attacks[14]. The need for robust and adaptive detection models that can handle high-dimensional data remains a critical research area. This paper proposes a novel weakly supervised traffic detection method named CTFABOD, which combines the strengths of angle-based outlier detection (FastABOD) and generative adversarial network (CTGAN) to enhance the detection of unknown network attacks. The effectiveness of the CTFABOD model is evaluated on the NSL-KDD dataset, and compared to existing weakly supervised models. CTFABOD increased the precision score by 10.31%. Our experiments demonstrate that improving the performance of weakly supervised models through generative adversarial network is an effective way to deal with unknown network attacks.

2. Related Work

Network intrusion detection has long been a critical area of research, given the increasing sophistication and frequency of cyberattacks. Several studies have showcased the effective utilization of supervised classifiers in intrusion detection. For instance, the research outlined involved the comparison of six tree-based supervised classifiers for detecting intrusions in the UNSW-NB15 dataset[15], with results indicating that decision trees with pruning generally outperformed other methods such as random forests[16]. Furthermore, the authors of conducted a comparative analysis of support vector machines (SVMs) and deep convolutional neural networks (CNNs), concluding that feature selection contributes to enhancing the classification accuracy of the NSL-KDD dataset, and deep CNNs outperform SVMs to a significant degree[17]. Recent studies have raised doubts about the effectiveness of deep classifiers when handling tabular data, noting that they frequently exhibit subpar performance compared to supervised classifiers. Notably, the research outlined a comparison between four cutting-edge deep learners and non-neural network classifiers, with findings indicating that the latter demonstrated superior classification performance over the former[18]. Traditionally, most anomaly detectors for tabular data are developed using supervised and deep classifiers[17]. These detectors necessitate training data with labeled instances to construct a model that typically achieves a high level of accuracy, resulting in minimal misclassifications.

Deep models are neural network classifiers that consist of numerous hidden layers, commonly referred to as deep neural networks or deep learners[19]. While deep learners are widely recognized as the standard for

classifying unstructured data such as images, audio, lidar point clouds, and video, they often encounter challenges when tasked with classifying tabular data. For instance, certain studies have revealed that deep neural network classifiers applied to tabular data exhibit inferior classification performance compared to other supervised classifiers. Additionally, there are research efforts aimed at converting tabular data into images to fully harness the potential of deep learners in processing images[20], but these endeavors have not yielded substantial improvements in classification performance.

Detecting outliers in high-dimensional data poses unique challenges. Recent research has explored new approaches, such as treating the problem as "subspace" outlier detection and developing models that exhibit greater stability as data dimensionality increases, such as angle-based outlier models[5,13]. Additionally, ensemble methods for outlier detection are considered particularly beneficial for high-dimensional data[21]. Another area of research focuses on addressing efficiency concerns by employing advanced acceleration techniques tailored for high-dimensional data, such as random projections or locality-sensitive hashing (LSH)[22]. These methods collectively tackle different aspects of the "curse of dimensionality."

3. Method

Network attack detection methods based on weakly supervised learning are highly applicable to unknown attacks. In high-dimensional data, since angles are more stable than distances, outliers typically do not lie in the same direction as many points. Adversarial sample generating is a method to enhance the robustness of a model by adding adversarial examples to the training data. This study develops the CTFABOD model to detect the network attacks on the NSL-KDD dataset. The CTFABOD model utilized a fast angle-based outlier detection method (FastABOD) to detect the network attacks and utilized conditional tabular GANs (CTGAN) to generate augmentation samples as shown in Figure 1. Generated samples are utilized to improve the performance of FastABOD.

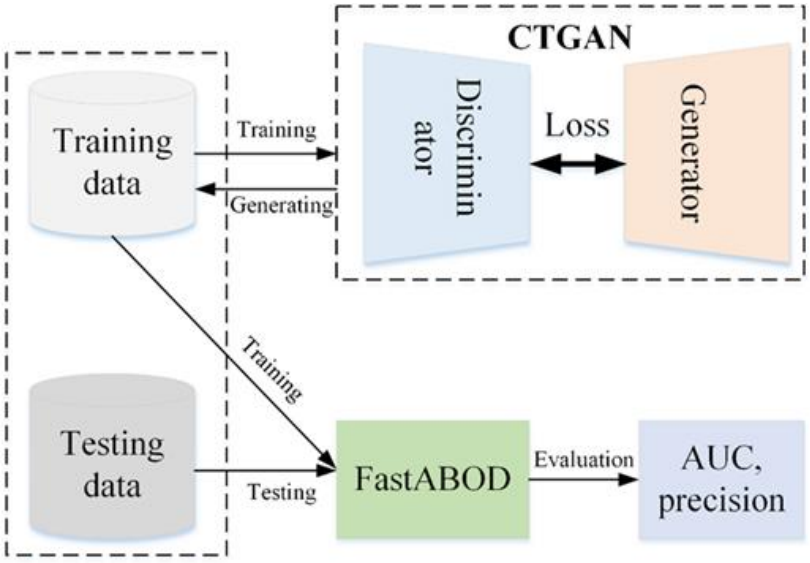

Figure 1. The workflow of CTFABOD.

3.1 CTGAN

A distinctive feature of CTGAN is its utilization of a variational gaussian mixture model to ascertain the number of modes for each continuous column and to fit a Gaussian mixture model accordingly[23]. Each data value is then encoded as a one-hot vector that signifies the mode, accompanied by a scalar that denotes the value within that mode. This methodology adeptly manages the complexities of multimodal and non-Gaussian distributions prevalent in continuous columns. Tabular data often grapple with the issue of category imbalance within categorical columns. Random sampling of training data can lead to the underrepresentation of minority categories, thereby impeding the generator's ability to accurately learn these categories. To counteract this, CTGAN incorporates a conditional generator alongside a training sampling mechanism. The conditional generator produces samples conditioned on specific values of a categorical column. Concurrently, the training sampling mechanism ensures equitable representation of all categories during the training phase by sampling in accordance with the frequency of the category column. This dual approach not only mitigates the category imbalance problem but also preserves the fidelity of the real data distribution.

The architectural backbone of CTGAN is a fully connected network that forms the foundation for both the

generator and the discriminator. The generator is equipped with two hidden layers, employing batch normalization and the ReLU activation function. It outputs a blend of activation functions: tanh for continuous values and Gumbel softmax for modality indicators and discrete values. The discriminator mirrors this structure with two hidden layers, utilizing the leaky ReLU activation function and dropout to enhance robustness. To avert the pitfall of mode collapse, CTGAN leverages the PacGAN framework, generating and evaluating ten samples in each iteration. This strategy bolsters the model's capacity to produce diverse and representative synthetic data, thereby elevating the overall quality and utility of the generated dataset.

3.2 FastABOD

ABOD is an angle-based outlier factor (ABOF) method that mines high-dimensional data to identify outliers[24]. In high-dimensional space, the distance between data points becomes meaningless. This method shows that the distance vector angle between points in the vector space is more suitable for anomaly detection tasks in high-dimensional space. Differentiate between inliers and outliers by comparing their angles with other points. Core of ABOF method is to organize insiders into clusters of data points. The angle between a line formed by connecting an inner line to two other points should have completely different values because there are usually other points around the inner line; this results in a significant change in the angle. The ABOF method provides a way to quantify the divergence of objects in their orientations relative to each other, helping to identify outliers in high-dimensional data based on the angles between observed distance vectors [25]. This method can be expressed as:

Considering the space of data points $D \subseteq \mathbb{R}^m$, and three related points $x_{\vec{A}}$, $x_{\vec{B}}$, $x_{\vec{C}}$, the angle-based anomaly factor ABOF ($x_{\vec{A}}$) is expressed as the variance of the angle between the difference vector of $x_{\vec{A}}$ and all other pairs, and the distance between points is used as the weight:

$$VAR_{x_{\vec{B}}, x_{\vec{C}} \in D}\left(\frac{\left(x_{\vec{A}} - x_{\vec{B}}\right)^{\mathrm{T}}\left(x_{\vec{A}} - x_{\vec{C}}\right)}{\left\|x_{\vec{A}} - x_{\vec{B}}\right\|^2 \cdot \left\|x_{\vec{A}} - x_{\vec{C}}\right\|^2}\right) \tag{2}$$

Considering the construction of triple pairs, the time complexity of the current method is $O(n^3)$. An approximation algorithm called FastABOD can be used to solve this problem. This method approximates the angle-based outlier factor (ABOF) using only a small subsample of the available data points. In FastABOD method, only points with the strongest weights are considered for variance calculation. Usually, the K-nearest neighbor method is used to identify relevant data points. This approximation is more effective, especially in low-dimensional datasets where distances are more meaningful. Using the nearest neighbors provides a better approximation of the ABOF. FastABOD method is more suitable for large datasets, reducing the computation to $O(n^2 + nk^2)$, where k represents the number of nearest neighbors. On the other hand, the performance of the algorithm depends on the neighbors chosen. Using a large number of neighbors can improve the quality of the results, but it also affects the time complexity. Considering the large nature of the existing datasets, FastABOD is used. The ABOD function can be adapted using different distance metrics.

4. Experiment

4.1. NSL-KDD dataset

The dataset NSL-KDD used in the study is the revised version of the KDD99 dataset. Although the KDD99 dataset is widely used, it has problems such as redundancy and duplicate records, which affect the accuracy of experimental results. NSL-KDD improves the quality of the KDD99 dataset by reducing redundancy and balancing data distribution. Each record in the NSL-KDD dataset contains 43 features, of which 41 features refer to the traffic input itself, and the last two are labels (normal or attack) and scores. The dataset includes a class of normal data (Normal) and four attack methods, namely, denial of service (DoS), probe (Probe), user to root (U2R), and remote to local (R2L), as shown in Table 1.

Table 1. Dataset Composition.

Dataset	Quantity Details					
	Sum	Normal	DoS	Probe	U2R	R2L
KDDTrain + 20%	25,192	13,449 (53%)	9234 (37%)	2289 (9.16%)	11 (0.04%)	209 (0.8%)

KDDTrain+	125,973	67,343 (53%)	45,927 (37%)	11,656 (9.11%)	52 (0.04%)	995 (0.85%)	In
KDDTest+	22,544	9711 (43%)	7458 (33%)	2421 (11%)	200 (0.9%)	2654 (21.1%)	addition, to avoid the

interference of irrelevant features, the first four non-numeric features were removed from the above 41 features, and the remaining 37 numeric features were retained.

4.2 Evaluation metrics

In this study, we develop the CTFABOD model for network intrusion detection. Existing four weak supervision models including KNN, K-means, ABOD and FastABOD are used as comparison methods. To assess models' performance on network intrusion detection, AUC (Area Under the Curve) and Precision are used as evaluation metrics. The confusion matrix is a method used to evaluate the performance of classification models, particularly for binary and multi-class classification. For multi-class classification, rows represent the actual classes, and columns represent the predicted classes. Each cell shows the number of samples for the corresponding class. The AUC and precision can be calculated by the confusion matrix, which is shown as following formulas.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Where TP represents that the model correctly predicts the positive class, FP represents that the model incorrectly predicts the positive class.

AUC is the area under receiver operating characteristic curve, which is used to evaluate the performance of a binary classification model. The AUC value provides a comprehensive measure of the model's ability to distinguish between positive and negative samples, which is particularly useful in the case of class imbalance.

5. Results and Discussions

This study proposes the CTFABOD for network intrusion detection on the NSL-KDD dataset. CTFABOD model ensemble FastABOD and CTGAN methods, FastABOD is used to detect network intrusion, and CTGAN is used to generate samples. Network intrusion detection is a binary classification, so we utilized AUC and precision to assess its performance. As we all know, KNN and K-means are weak supervision algorithms and applied to detect network intrusion. The FastABOD method is used as the classifier in this study, which is based on the ABOD algorithm. So we select KNN, K-means, ABOD, and FastABOD models as comparison methods. This study tests all models on the NSL-KDD dataset.

CTFABOD was used for binary classification weak supervision training and tested on the test set. The experimental results are shown in Table 2. The experimental results demonstrate the performance of various models in terms of AUC and precision. The KNN model achieved an AUC of 0.4991 and a precision of 0.4796. FastABOD slightly outperformed ABOD, achieving an AUC of 0.7361 and a precision of 0.6954. Notably, CTFABOD exhibited the best performance among all models, with the highest AUC of 0.7844 and the highest precision of 0.7671. These results indicate that CTFABOD is the most effective model and increases the precision score by 10.31% on the NSL-KDD dataset. CTFABOD achieved the best experimental results mainly because augmentation samples generated by CTGAN model improved the robustness of FastABOD, making FastABOD more accurate in identifying abnormal traffic. AUC of CTFABOD on testing data is 0.7844 and the precision is 0.7671. Clustering images of the real distribution and the predicted distribution are given in Figure 2. Since there are five types of samples, four of which are attack samples and one is a normal sample, the attack is marked as 0 and the normal sample is marked as 1 for distinction. In the real distribution, the normal sample appears as a deviation sample (yellow and red marks). It can be seen in the predicted distribution that although there is some overlap in space, CTFABOD can still distinguish the distribution of the two data types well. CTFABOD highlights the potential of weakly supervised learning and adversarial sample generation in enhancing the robustness and adaptability of intrusion detection systems. This approach could be further explored and refined to address other types of cyber threats and attack vectors beyond those tested in this study.

Table 2. Performance on the testing data of NSL-KDD.

Models	AUC	Precision
KNN	0.4991	0.4796
K-means	0.1316	0.1030
ABOD	0.7343	0.6945
FastABOD	0.7361	0.6954

CTFABOD	0.7844	0.7671

Figure 2. NSL-KDD data clustering sample.

The use of CTGAN for adversarial sample generation provides a promising direction for improving the performance of ML models in the face of evolving threats. By generating realistic augmentation samples, model can be better prepared to handle novel attack patterns. This approach could be extended to other ML tasks, such as image classification, natural language processing, and medical diagnostics, where robustness against adversarial attacks is crucial.

7. Conclusions

Current information and communication systems face network attacks with strong unknown characteristics, and existing network attack detection methods cannot achieve effective detection. This paper develops the CTFABOD model to detect abnormal traffic data. CTFABOD model ensemble FastABOD and CTGAN methods are used to detect network intrusion and generate augmentation samples. Generated augmentation samples are utilized to improve the performance of FastABOD. Compared with the existing four network attack detection methods, the CTFABOD achieved the highest AUC and precision on the NSL-KDD dataset. Thus, CTFABOD increased the precision score by 10.31%. Our experiments demonstrate that improving the performance of weakly supervised models through generation adversarial methods is an effective way to deal with unknown network attacks.

Author Contributions:

Conceptualization, X.L. and X.W.; methodology, X.W. and M.Z.; software, M.Z. and X.Z.; validation, X.L. and M.Z.; formal analysis, X.Z.; investigation, X.L.; resources, X.Z.; data curation, X.Z.; writing—original draft preparation, X.Z.; writing—review and editing, X.L..; visualization, X.W.; supervision, X.W.; project administration, X.L.; funding acquisition, M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key R&D Program of Jiangsu (BE2022081).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors will make the data available upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Laghari, A.A.; Li, H.; Khan, A.A.; Shoulin, Y.; Karim, S.; Khani, M.A.K. Internet of Things (IoT) applications security trends and challenges. *Discover Internet of Things* **2024**, *4*, 1-22.
2. Pandey, S.; Kumar De, A.; Choudhary, S.; Bhushan, B.; Bhatia, S. Leveraging Blockchain Technology in Industry 4.0 and Industrial Internet of Things (IIoT) Scenarios. **2023**.
3. Tindan, T.N.; Tang, G. The Effects of Integrating Information and Communication Technology (ICT) In Teaching the Atomic Structure in Chemistry among Senior High School Students. *International Journal of Latest Technology in Engineering Management & Applied Science* **2025**, *13*, 42-48.
4. Makhdoom, I.; Abolhasan, M.; Lipman, J.; Liu, R.P.; Ni, W. Anatomy of threats to the internet of things. *IEEE communications surveys & tutorials* **2018**, *21*, 1636-1675.
5. Williams, P.; Dutta, I.K.; Daoud, H.; Bayoumi, M. A survey on security in internet of things with a focus on the impact of emerging technologies. *Internet of Things* **2022**, *19*, 100564.

6. Abdelkader, S.; Amissah, J.; Kinga, S.; Mugerwa, G.; Emmanuel, E.; Mansour, D.-E.A.; Bajaj, M.; Blazek, V.; Prokop, L. Securing modern power systems: Implementing comprehensive strategies to enhance resilience and reliability against cyber-attacks. *Results in engineering* **2024**, 102647.
7. Vadisetty, R. The Effects of Cyber Security Attacks on Data Integrity in AI. In Proceedings of the 2024 International Conference on Intelligent Computing and Emerging Communication Technologies (ICEC), 2024; pp. 1-6.
8. Wan, X.; Xue, G.; Zhong, Y.; Wang, Z. Separating Prediction and Explanation: An Approach Based on Explainable Artificial Intelligence for Analyzing Network Intrusion. *Journal of Network and Systems Management* **2025**, *33*, 1-29.
9. Bourou, S.; Saer, A.E.; Velivassaki, T.H.; Voulkidis, A.; Zahariadis, T. A Review of Tabular Data Synthesis Using GANs on an IDS Dataset. *Information* **2021**, *12*, 14.
10. Zhang, Z.; Wang, K.; Dong, J.; Li, S. SDROF: outlier detection algorithm based on relative skewness density ratio outlier factor. *Applied Intelligence* **2025**, *55*.
11. Mahmud, M.Z.; Islam, S.; Alve, S.R.; Pial, A.J. Optimized IoT Intrusion Detection using Machine Learning Technique. **2024**.
12. Arumugam, S.R.; Paul, P.M.; Issac, B.J.J.; Ananth, J.P. Hybrid deep architecture for intrusion detection in cyber-physical system: An optimization-based approach. *International Journal of Adaptive Control & Signal Processing* **2024**, *38*.
13. Zoppi, T.; Gharib, M.; Atif, M.; Bondavalli, A. Meta-learning to improve unsupervised intrusion detection in cyber-physical systems. *ACM Transactions on Cyber-Physical Systems (TCPS)* **2021**, *5*, 1-27.
14. Zoppi, T.; Ceccarelli, A.; Puccetti, T.; Bondavalli, A. Which algorithm can detect unknown attacks? Comparison of supervised, unsupervised and meta-learning algorithms for intrusion detection. *Computers & Security* **2023**, *127*, 103107.
15. Moustafa, N.; Slay, J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In Proceedings of the 2015 military communications and information systems conference (MilCIS), 2015; pp. 1-6.
16. Chkirbene, Z.; Eltanbouly, S.; Bashendy, M.; AlNaimi, N.; Erbad, A. Hybrid machine learning for network anomaly intrusion detection. In Proceedings of the 2020 IEEE international conference on informatics, IoT, and enabling technologies (ICIoT), 2020; pp. 163-170.
17. Taher, K.A.; Jisan, B.M.Y.; Rahman, M.M. Network intrusion detection using supervised machine learning technique with feature selection. In Proceedings of the 2019 International conference on robotics, electrical and signal processing techniques (ICREST), 2019; pp. 643-646.
18. Shwartz-Ziv, R.; Armon, A. Tabular data: Deep learning is not all you need. *Information Fusion* **2022**, *81*, 84-90.
19. Zhang, C.; Jia, D.; Wang, L.; Wang, W.; Liu, F.; Yang, A. Comparative research on network intrusion detection methods based on machine learning. *Computers & Security* **2022**, *121*, 102861.
20. Zhu, Y.; Brettin, T.; Xia, F.; Partin, A.; Shukla, M.; Yoo, H.; Evrard, Y.A.; Doroshow, J.H.; Stevens, R.L. Converting tabular data into images for deep learning with convolutional neural networks. *Scientific reports* **2021**, *11*, 11325.
21. Zimek, A.; Campello, R.J.; Sander, J. Ensembles for unsupervised outlier detection: challenges and research questions a position paper. *Acm Sigkdd Explorations Newsletter* **2014**, *15*, 11-22.
22. Liu, R.; Zhao, J.; Chu, W.H., Jing. Can LSH (locality-sensitive hashing) be replaced by neural network? *Soft computing: A fusion of foundations, methodologies and applications* **2024**, *28*, 887-902.

23. Habibi, O.; Chemmakha, M.; Lazaar, M. Imbalanced tabular data modelization using CTGAN and machine learning to improve IoT Botnet attacks detection. *Engineering Applications of Artificial Intelligence* **2023**, *118*, 105669.
24. Utilizing angle-based outlier detection method with sliding window mechanism to identify real-time crash risk. *Journal of Transportation Safety And Security* **2024**, *16*, 157-174.
25. Ilie-Ablachim, D.C.; Dumitrescu, B. Angle-Based Dictionary Learning for Outlier Detection. In Proceedings of the 2023 IEEE Third International Conference on Signal, Control and Communication (SCC), 2023; pp. 01-06.