Received: 29 January 2021 | Revised: 17 May 2021 | Accepted: 24 May 2021

DOI: 10.1049/cps2.12019

IET Cyber-Physical Systems: Theory & Applications

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

A comparative analysis of CGAN-based oversampling for anomaly detection

Rahbar Ahsan[1] | Wei Shi[2] | Xiangyu Ma[2] | William Lee Croft[1]

[1]School of Computer Science, Carleton University, Ottawa, Canada

[2]School of Information Technology, Carleton University, Ottawa, Canada

Correspondence
Xiangyu Ma, School of Information Technology, Carleton University, 1125 Colonel By Dr, Ottawa, Canada.
Email: johnnyma@cmail.carleton.ca

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: RGPIN-2020-06482

Abstract

In this work, the problem of anomaly detection in imbalanced datasets, framed in the context of network intrusion detection is studied. A novel anomaly detection solution that takes both data-level and algorithm-level approaches into account to cope with the class-imbalance problem is proposed. This solution integrates the auto-learning ability of Reinforcement Learning with the oversampling ability of a Conditional Generative Adversarial Network (CGAN). To further investigate the potential of a CGAN, in imbalanced classification tasks, the effect of CGAN-based oversampling on the following classifiers is examined: Naïve Bayes, Multilayer Perceptron, Random Forest and Logistic Regression. Through the experimental results, the authors demonstrate improved performance from the proposed approach, and from CGAN-based oversampling in general, over other oversampling techniques such as Synthetic Minority Oversampling Technique and Adaptive Synthetic.

1 | INTRODUCTION

In recent years, an unprecedented rise in the number of computing apps and network sizes has increased the potential threat of cyber-attacks drastically [1]. Network security has become an integral concern more than ever. It has also become increasingly difficult to capture anomaly signals due to their constantly changing nature [2]. It is therefore crucial to employ automated systems like Intrusion Detection Systems (IDSs), which can accurately detect cyber-attacks [3]. The key roles of an IDS are tracking hosts and networks, evaluating computer system activities, generating warnings, and reacting to unusual behaviours. IDSs that utilize machine learning algorithms can identify intrusion effectively when there is sufficient training data, and the algorithms are flexible enough to identify attack variations and novel threats. Furthermore, machine-learning-based IDSs are simple to develop and construct without having deep domain-specific knowledge [4]. Supervised Learning, Unsupervised Learning and Reinforcement Learning (RL) are the three main machine learning techniques. RL, which lies outside of supervised and unsupervised learning due to its own 'signal sensing' ability [5], overcomes the problem of label scarcity in supervised learning and the problem of poor performance in unsupervised learning.

In data mining, datasets with imbalanced classes are a ubiquitous natural phenomenon. Common IDS datasets suffer from imbalanced representation as the normal traffic behaviour always constitutes the majority of the dataset, whereas intrusion traffic behaviour typically constitutes a relatively small proportion of the dataset. When making a binary classification in detecting fraudulent activities, the class-imbalance issue significantly reduces the effectiveness of binary classifiers, undesirably biasing the results towards the prevailing class, while we are interested in the minority class. Current research on the classification of imbalanced data is mainly summarized into two classes of approaches. The first targets the algorithm level through the use of classifiers designed with imbalanced data in mind. The objective is to enable the classifier to adapt or strengthen its learning process for the minority classes. The second approach involves data-level modifications aimed at balancing the distribution of classes in the training data. Resampling techniques are employed for this purpose and they are divided into oversampling [6], undersampling [7] and hybrid sampling methods [8]. Generative adversarial networks (GANs) [9] are machine learning models, which are used to produce novel instances of samples from a targeted data distribution. A Conditional GAN (CGAN) [10] variant of the model can be used to provide further control over the generated data such as forcing the model to learn a distribution of a dataset conditioned on its class labels.

1.1 | Major contributions

In this study, a novel algorithm, AEGAN, has been proposed. It is a hybrid model that consists of adversarial environment reinforcement learning (AE-RL) and CGAN. The CGAN model is trained on a network intrusion detection dataset, and is used to generate synthetic samples to handle the class-imbalance problem. This data-level approach has been combined with an algorithm-level approach, AE-RL. The combination of these two frameworks is able to provide an improved performance in the network intrusion detection system. Furthermore, a comparative study has been shown between CGAN and other oversampling techniques and a detailed analysis has been conducted over which oversampling techniques are more appropriate to combine with AE-RL. We conduct our experiments on the AWID dataset, comparing various oversampling techniques in their performance for classification over an imbalanced distribution of classes. Our contributions are summarized as follows:

1. We present a novel IDS solution that combines an algorithm-level approach, AE-RL, with CGAN-oversampling concept to achieve a better classification on imbalanced data.
2. We compare and analyse the performance of alternate classifiers (Naïve Bayes [NB], Multilayer Perceptron [MLP], Random Forest [RF] and Logistic Regression [LR]) when combined with CGAN.
3. We analyse the performance of different combinations of oversampling techniques and base classifiers and demonstrate improved classification performance from our proposed solution.

2 | RELATED WORK

In the cybersecurity field, detecting cyber-attacks has become a crucial task. Cyber-attacks are rapidly evolving and becoming increasingly difficult to detect [2]. Traditional machine learning approaches are ill suited to the task of detecting evolving cyber-attacks due to the class-imbalance property of common IDS datasets. We provide a review of both algorithm-level and data-level approaches to address the class-imbalance problem.

2.1 | A review of algorithm-level approaches

With algorithm-level approaches, the objective is to strengthen a classifier in terms of algorithm architecture to cope with the class-imbalance problem. AE-RL is an example in this context [11]. AE-RL learns based on a reward function, which has been obtained through the interaction between two agents. However, although AE-RL offers improved performance in the imbalanced domain, it also suffers from the lack of positive samples in cyber-attack datasets [12]. Due to insufficient variational data, the classifier struggles to predict unknown classes. In [13], the authors propose a solution for IDS using boosting-based feature selection to evaluate the relative importance of individual features. This work accounts for the imbalanced classes by assigning different costs for positive and negative samples and applies feature selection on the modified dataset. In [14], the authors address the class-imbalance problem by consolidating stratified sampling, the use of a cost function, and a weighted support vector machine (WSVM). The authors assign the records of minority classes that have greater weights than those of the majority classes. In [15], the authors propose a double-layer detection and classification approach that consists of a neural network model with two layers and multiple ensembled techniques to better categorize subtypes in the attacks. Their experiments employ various classifiers including support vector machine (SVM), RF, Adaboost and Gradient-Boosted Decision Trees (GBDT). However, only precision and recall scores are compared. AE-RL showed the best performance on addressing the class-imbalance problem from the algorithm-level perspective. Moreover, the auto-learning nature of reinforcement learning gives AE-RL a unique advantage that makes AE-RL not only capable of performing prediction without human supervision, but also has great potential on performing prediction on unknown classes.

2.2 | A review of data-level approaches

The objective of data-level approaches is to alter on the training dataset such that it becomes sufficiently balanced, enabling more effective learning for classification algorithms. A typical strategy is to generate synthetic data samples. In [16], the authors propose the Two-Layer Multi-class Detection (TLMD), which consists of a combination of a C5 Decision Tree and NB to perform adaptive network intrusion detection. They handle the imbalanced dataset problem by extracting subsets of data from the training dataset. The work in [17] proposes Synthetic Minority Oversampling Technique (SMOTE), which can produce synthetic samples via interpolation using distance measures within the K-nearest neighbouring samples. In [18], it is stated that traditional oversampling approaches such as SMOTE are restricted to only generating samples based on local information. Hence, when applying oversampling techniques, data generated by SMOTE has a disadvantage due to the limited local information. In [19], The authors optimize the SMOTE ratios for the minority classes on the KDDCUP1999 dataset by adding a support vector regression to help in creating the model. By conducting the experiments using the best ratios, the results are significantly better than the original SMOTE. In [20], the authors propose a novel class-imblance processing technology that combines with SMOTE and under-sampling for clustering based on Gaussian mixture model (GMM). They investigate the impact of different numbers of convolution kernels and different learning rates on model performance. Through experiments on the UNSW-NB15 and CICIDS2017 datasets, results show that their proposed method outperforms the state-of-the-art intrusion deteciton methods. In [21], the authors present Adaptive Synthetic (ADASYN) sampling approach for imbalanced learning , which improves learning with respect to the data distribution by reducing the bias due to the class-imbalance and adaptively shifting the classification decision boundary towards the difficult examples. Similarly, to SMOTE, it has a disadvantage in limited local information based on nearest neighbour distance measures.

2.2.1 | GAN oversampling techniques

The GAN framework is a popular machine learning approach for creating synthetic data. Oftentimes, research works apply the GAN framework in the context of producing synthetic samples for cyberattack datasets [22]. GAN has a disadvantage owing to mode collapse, unstable training and lacks the consideration of the majority class samples that affect the classification boundary [18]. In [23], the authors propose an Outlier Detectable-GAN (OD-GAN), which uses a discriminator as an outlier detector to quantify the difference between the distributions of the majority and minority classes. In [24], the authors propose a novel Single-Objective Generative Adversarial Active Learning (SO-GAAL) method to generate potential outliers for data in high-dimensional space as these outliers may provide information to assist the classifier in describing a boundary that can separate outliers from normal data effectively. In [25], the authors propose a conditional variational autoencoder (VAE) that is able to learn class-dependent distributions. In their results, they discover that deep generative models outperform traditional oversampling methods in many circumstances, especially in cases of severe imbalance. In [26], the authors claim that CGAN is better at approximating the true data distribution and generate data for the minority class of various imbalanced datasets. They apply CGAN on binary-class-imbalanced datasets, where the CGAN conditional information is the data sample class label. In their experiments on various datasets with different imbalance ratios, the performance of CGAN is compared against several oversampling methods including Random Oversampling, SMOTE, Borderline SMOTE, ADASYN and Cluster-SMOTE and decent improvement have been found. In [27], the authors propose to add constraints to the networks of the CGAN to limit the degree of convergence freedom, which mitigates the phenomenon of slow convergence or failure to converge due to the high degree of freedom of traditional GAN. In [28], the authors train the CGAN model to generate synthetic samples from minority classes using the KL-divergence to guide the model towards learning the true minority class distribution. Before training the model, they have also reduced the size majority class by using undersampling techniques. Their experiments are performed on the NSL-KDD and UNSW-NB15 datasets and decent performance has been achieved. In [29], the authors propose a solution for classification on credit card fraud datasets, which are strongly imbalanced. It adopts the GAN framework to output synthetic minority class samples and then merges them with the original training set to form an augmented set. The experiments show a significant improvement on the performance of a classifier trained on the augmented set over that of the same classifier trained on the original data. The authors in [30] propose an Imbalanced Generative Adversarial Network-based Intrusion Detection System (IGAN-IDS), which introduces an imbalanced data filter and convolutional layers to the typical architecture. The framework utilizes a data filter, which restricts the GAN training data to only the minority classes. The work in [31] introduces Autoencoder-Conditional GAN (AE-CGAN), a novel framework that processes data characteristics to a lower level by using an autoencoder prior to GAN oversampling. The use of a Conditional Wasserstein GAN with Gradient Penalty (CWGAN-GP) for class balancing is proposed in [18]. The Wasserstein objective function provides greater stability during training, mitigating issues such as mode collapse that are common in standard GAN models. In [32], the authors propose Supervised Adversarial Variational Auto-encoder with Regularization and Deep Neural Network (SAVAER-DNN) that can detect both minority class samples, and unknown attacks. The framework consists of a combination of a supervised variational autoencoder with regularisation and a Wasserstein GAN with Gradient Penalty (WGAN-GP). It uses the encoder-decoder process to synthesise minority class samples and unknown attacks for the purpose of balancing training data.

3 | CGAN AS AN OVERSAMPLING TECHNIQUE

In this section, we describe the general structure and training details of the GAN framework, the CGAN variant and how they are being used as an oversampling technique.

3.1 | Generative adversarial networks

When training a GAN, the goal is to learn a mapping of noise drawn from a random distribution to an approximation of the desired data distribution. When successfully trained, the model is able to produce highly realistic samples. This ability to produce novel, high quality data is of particular use for the task of data augmentation and has been investigated as a means to balance class distributions in unbalanced datasets by oversampling minority classes [26, 29].

A GAN typically consists of two multi-layered, feed-forward neural networks: a generator and a discriminator. The task of the generator is to learn the mapping from a multi-dimensional latent space to the data distribution. The latent space is often sampled using a normal or uniform distribution. The discriminator is used to scrutinize the quality of the generated data by learning to classify the samples as being either real or fake (i.e. generated). The output of the discriminator is thus a scalar value representing the probability that the input sample is real. A diagram depicting the overall structure of the model is given in Figure 1. Dashed arrows denote the auxiliary information present only in the CGAN variant. In

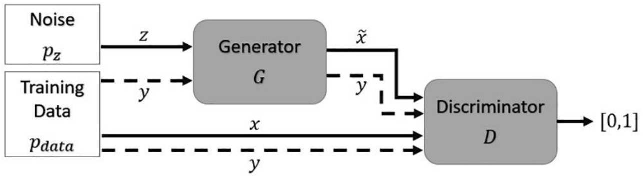

FIGURE 1 Flow of data through the generative adversarial networks model

both variants, the generator G accepts as input random noise z sampled from a distribution p_z and produces as output fake data $\tilde{x}$ from an approximation of the training data distribution p_{data}. The discriminator D accepts input fake data $\tilde{x}$ or real data x sampled from p_{data} and outputs a value in the range [0, 1]. In the CGAN variant, both networks additionally accept input auxiliary information y sampled from p_{data}.

The two networks compete in a minimax game during training. The discriminator aims to correctly distinguish between real training samples and fake samples while the generator aims to produce high quality samples capable of fooling the discriminator. The adversarial nature of the training forces both networks to continually improve in order to thwart the other. The networks alternate in updating their parameters during training in order to ensure that neither improves too rapidly compared with the other, as this would lead to training instability. Both networks share the same objective function:

$$\min_G \max_D \mathbb{E}_{x\sim p_{data}(x)}[\log(D(x))] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))],$$

where G and D are the generator and discriminator networks, respectively, p_{data} is the training data distribution and p_z is the distribution of random noise used as input for G. The generator aims to minimise this value while the discriminator aims to maximize it.

A CGAN [10] variant of the model can also be employed when the goal is to generate samples from a conditional data distribution. This can be used to provide more control over the generated data. For instance, by training the model to learn a distribution conditioned on class labels, it becomes possible to specify the class of samples to be generated by the trained model. In the context of oversampling for the purpose of data balancing, this allows for labels of minority classes to be specified during data generation in order to selectively augment the classes in need of additional samples. At the implementation level, the generator and discriminator networks are modified to accept an additional input that captures the auxiliary information (e.g., a class label). The flow of the auxiliary information is shown by the dashed arrows in Figure 1. The updated objective function is given as follows:

$$\min_G \max_D \mathbb{E}_{x,y\sim p_{data}(x,y)}[\log(D(x,y))] \\ + \mathbb{E}_{z\sim p_z(z),y\sim p_y(y)}[\log(1 - D(G(z,y),y))],$$

where p_{data} is now a conditional distribution and p_y is the distribution of the auxiliary information.

4 | AEGAN

In this section, we present details on applying CGAN oversampling to an algorithm approach AE-RL. Q-learning is used to perform reinforcement learning. With the defined Q-values, we further explain how to setup AE-RL to run on the AWID dataset. Finally, we present details on how to combine the data-level approach CGAN to the chosen algorithm-approach.

4.1 | *Q*-learning

Q-learning [33] is a model-free RL algorithm that we use in our proposed approaches. The objective of Q-learning is to find a policy that is optimal in the sense that the expected return over all successive time steps is the maximum achievable. The algorithm consists of a value iteration process which iteratively updates the Q-values for each state action pair using the Bellman equation until the Q function converges to the optimal Q function, Q*. Among all possible functions, there exists an optimal value function which has the highest value, denoted as $V^*(s) = \max_\pi V^\pi(s)$ and $\pi^* = \arg\max_\pi V^\pi(s)$ is the optimal policy which maximizes the action value achievable for state s. The relationship is defined as follows:

$$Q^*(s,a) = R(s,a) + \gamma E_{s'}[V^*(s')] = R(s,a) \\ + \gamma \sum_{s'\in S} P(s'|s,a)V^*(s') \text{ [33]}$$

where $R(s, a)$ is the immediate expected reward after performing action a at state s and $\gamma E_{s'}[V^*(s')]$ is the expected, discounted, accumulated, future reward after the transition to the next state s'. The learning function is defined as follows:

$$Q^{new}(s_t, a_t) \leftarrow (1 - \alpha)Q(s_t, a_t) + \alpha(r_t + \gamma \max_a Q(s_{t+1}, a_{t+1}))$$ [33]

where Q^{new} is being updated at a discount rate γ based on the learning rate α.

4.2 | AE-RL setup

AE-RL is set up with two learning agents, with Q-values defined as $Q_c(s_i, a_{ci})$ and $Q_e(s_i, a_{ei})$. They are arbitrarily initialized at the beginning of the training process and tuned as RL proceeds. For every training episode, a random sample is selected as the initial state before the RL process. Then, it calculates the states, actions, and rewards of both agents. Following their given rewards, the classifier makes a prediction on the current class type, a_{ci}, based on its policy, $Q_c(s, a)$, while the environment provides the next training sample a_{ei}. a_{ei} is calculated based on the previous classifier's performance, $Q_e(s, a)$. Similarly, r_{ci} is either a positive integer or zero, as the reward corresponds to classifier's correct or incorrect classification, respectively. Meanwhile, r_{ei} takes opposite rewards as the classifier agent. The next state, s_{i+1}, is derived from randomly picking a sample from $a_{e(i+1)}$, the resulting class decided by the environment agent.

4.3 | AEGAN details

In general, AE-RL works well on imbalanced multi-class classification problems because of its ability to select samples in a balanced way. However, when there are very few samples in a minority class, even AE-RL suffers from a lack of variation in the data. We aim to solve this problem by creating relevant synthetic samples of the minority classes using CGAN. The CGAN model trains on the original dataset using the labels associated with each class as the conditional information. After this training, we select minority classes that are in need of oversampling and then produce new records by passing random noise to the CGAN combined with the class label to indicate what class to generate. CGAN is excellent at mimicking the original data distribution. Figure 2 illustrates the algorithm architecture of the proposed IDS. At first, various basic classifiers are trained on the original training dataset in order to gain general exploration knowledge. By observing each performance metric, we identify the minority classes that are in most need of augmentation. Then, we resample the chosen minority classes using oversampling techniques. This phase is considered as the data-level approach portion of the solution to the class-imbalance problem. The oversampling process is used to raise the number of instances in the minority classes to levels comparable with the majority class. Both original and synthetic records are concatenated into a new training dataset. This modified dataset is then passed in as an input for classifier training. In this study, various classifiers are compared. The expected use of AE-RL is to tackle the class-imbalance problem as an algorithm-level approach whereas other classifiers are used to examine the ability of CGAN as a data-level approach. This also enables us to examine whether the algorithm-level approach has a counter effect on the data-level approach. From the probabilistic family, we have selected NB. Then, RF is used as an ensemble technique. We have selected an MLP as an implementation of a neural network. Finally, Logistic Regression from the linear model family is selected.

5 | EXPERIMENTAL SETUP

In this section, we first provide information on the training dataset. We then describe the settings for training our CGAN framework and the conceptual details of all oversampling techniques and learning classifiers that we have used.

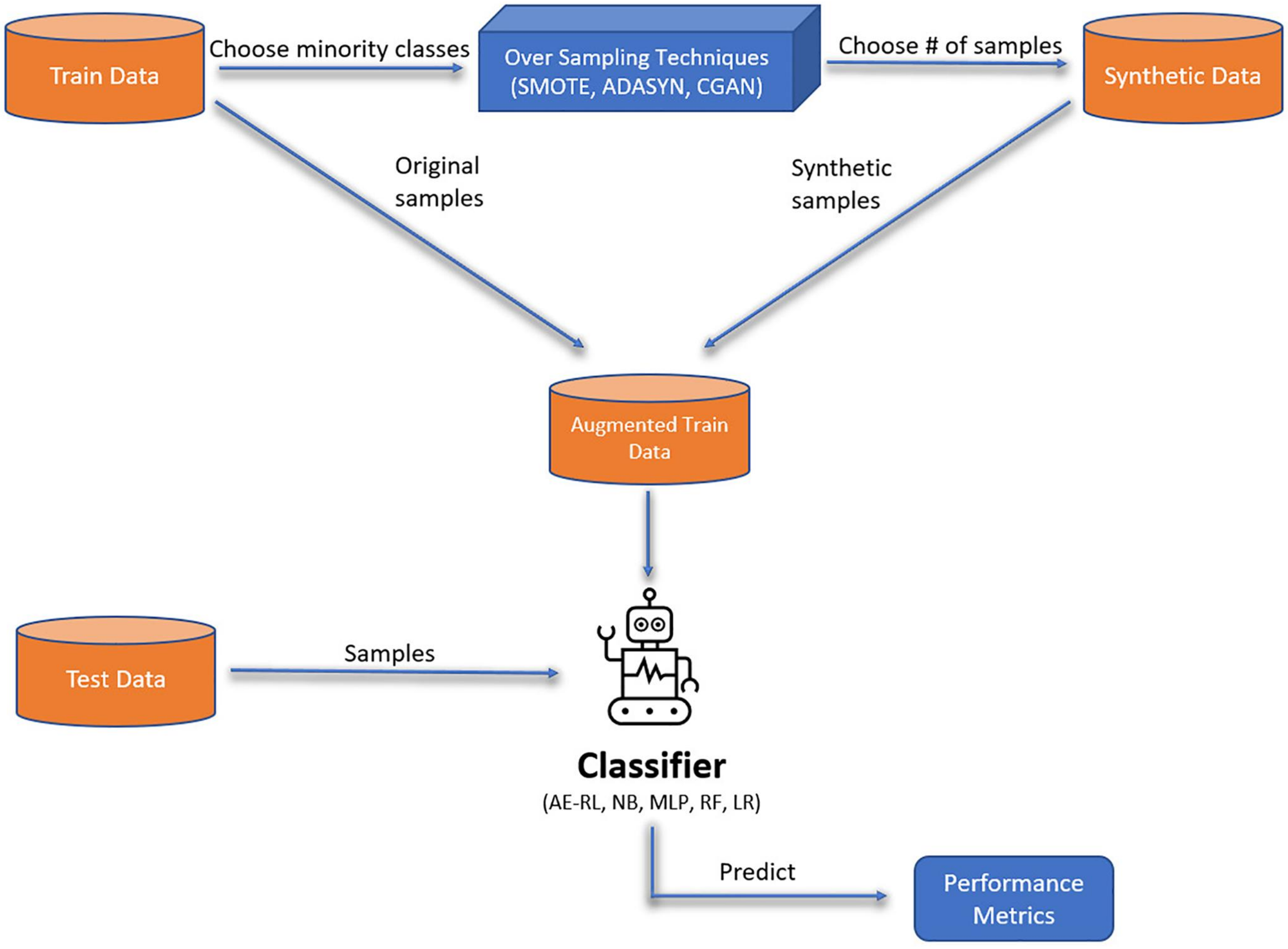

FIGURE 2 Algorithm architecture of the detection system

5.1 | AWID

The Aegean WIFI Intrusion Dataset (AWID) is a well-known publicly available dataset [34].This dataset has been created based on real-world network traffic with a fine mixture of normal activity and anomalous activities. Among the available variants of the dataset, we have selected AWID-CLS-R. Here, the attack types are classified into three major categories (Flooding, Injection and Impersonation), with normal activity as a fourth category. A detailed description of the attack types can be found in [34]. The dataset contains 1,795,574 training samples and 575,642 testing samples. It has in total 154 features (continuous and categorical ones). This dataset is extremely imbalanced, having 1,633,189 normal instances with only 162,385 instances across all attack classes. In Figure 3, the class distribution of the AWID dataset is shown. The minority classes are similarly sized with each other in regard to their number of samples in the training dataset. For the preprocessing, we have followed the steps detailed in [11].

5.2 | Oversampling techniques

In this study, three oversampling techniques are studied: two traditional oversampling techniques SMOTE and ADASYN, as well as CGAN. The default parameters are used for both SMOTE and ADASYN. In CGAN, there are two main modules: a generator and a discriminator, each of which is embodied by a neural network. The generator has four hidden layers; the numbers of output units for each layer are 32, 64, 128 and 256, respectively. Similarly, the discriminator has four hidden layers with 256, 128, 64 and 32 output units. The hidden layers of both modules are configured based on the description of [18]. The SGD optimizer is used along with 30,000 steps. During the training process, the generator and discriminator execute training steps alternately to update their parameters. A single training step is configured to be executed in each network. A learning rate of 0.0001 is set to help CGAN learn smoothly. For the CGAN model, both the generator and discriminator accept class labels as conditional information along with their standard input features. After the training stage, the generator can be employed to create synthetic samples of a specific class. After initial experimentation, we have chosen to use a 100-dimensional vector of Gaussian noise as the input for the generator network.

From the AWID dataset, we observe that the injection class has a very unique distribution, which enables it to be classified perfectly in the testing phase for all studied classifiers. Since injection has nearly 100% detection rate, we have decided to create synthetic samples only for the flooding and impersonation classes. It is crucial to decide the number of synthetic samples to create. We have created 1 million synthetic samples for each of the flooding and impersonation classes as this not only balances the distribution, but also provides enough variation to these minority classes to assist classifiers such as AE-RL to be more effective.

5.3 | Machine learning classifiers

In this study, the main machine learning classifier used is AE-RL. It is a complex classifier to manipulate. However, it offers great potential for handling an imbalanced dataset. The parameters of AE-RL are set up as described in [11]. Other standard machine learning algorithms are also included in our experiments in order to compare their results with those of AE-RL. The NB classifier sets a rigid independence assumption on the feature variables and takes the approach of calculating a conditional probability for all possible prediction outcomes [35]. MLP is similar to a regular feedforward neural network, and it consists of at least three layers of nodes, which include input, hidden and output layers. It uses backpropagation to train on labelled datasets [36]. RF is an ensemble learning method for supervised classification that combines a large number of decision trees and makes prediction on the class with the most votes [37]. LR is a statistical approach that makes estimates by using a logistic function to model a binary dependent variable [38] and it can be used for multi-class scenarios by applying it repeatedly as one-against-rest classification. The results of these standard classifiers are recorded both with and without the use of oversampling techniques. A detailed comparison is shown in the Results

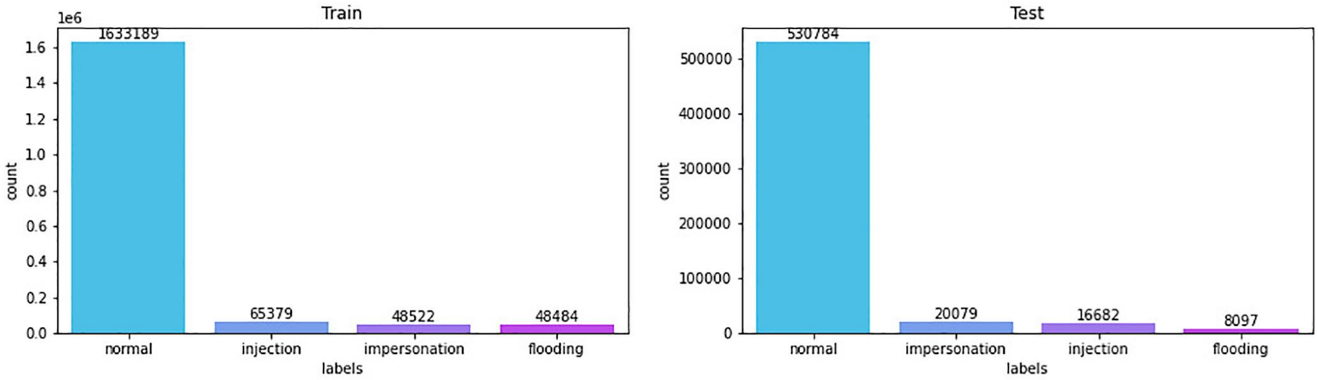

FIGURE 3 Class distribution of the Aegean WIFI Intrusion Dataset dataset

and Analysis section. Except AE-RL, all other classifiers are implemented using the scikit learn library [39].[1] For these classifiers, default parameters are used and all experiments are performed on nodes with the same level of computational power. For each combination of a classifier and an oversampling technique, we perform the experiment 20 times and report the average value of our selected evaluation metrics as the result.

5.4 | Evaluation metrics

For the evaluation of these the above-mentioned models, we have focussed on F1-score because it reflects the trade-off between recall and precision. We used the weighted form of F1-score due to the imbalanced nature of the dataset. The weighted F1-score gives priority based on the number of the samples of each class presented in the testing data. A good F1-score requires a reasonable detection rate and low false alarm rates [10]. The formula of the F1-score is as follows:

$$F1 - score = \frac{2 * (Precision * Recall)}{Precision + Recall} \tag{1}$$

We additionally report the accuracy, precision and recall for all the models.

6 | RESULTS AND ANALYSIS

In this section, we present the results of all classifiers when applied with different oversampling techniques. We report all four evaluation metrics and provide analysis on these results.

Table 1 and Figure 4 illustrate the performance of AE-RL combined with various oversampling techniques. We observe that AE-RL + CGAN achieves the best F1-score of 0.9438, while AE-RL combined with SMOTE also produces results that outperform the original AE-RL. AE-RL combined with ADASYN achieves a similar improved performance. Since AE-RL takes a unique algorithm-level approach to resolve class-imbalance, we were expecting minimal improvements when it was combined with data-level approaches. From the results, we indeed do see there is a smaller improvement in performance while combining data-level approaches with AE-RL compared with other classifiers (as presented later in this subsection). It is beneficial to further add a data-level resampling phase into the AE-RL classifier.

Figures 5 and 6 show a comparison between the original AE-RL and the case when CGAN is combined with AE-RL. A Normalized Confusion Matrix (NCM) is used to illustrate the results. First, we observe an improvement on the prediction of the minority classes: the correct labelling of the Flooding attack class is increased from 62% to 63% and the correct labelling of the Impersonation class is increased from 36% to 72%. Meanwhile, the false-positive rate on the Normal class labelled as the Impersonation class drastically decreased from 64% to 27%. As an effect of these shifted percentages, the AE-RL classifier gains an overall performance improvement from 93.34% to 94.38% in terms of the F1-score.

Table 2 and Figure 7 illustrate the performance of NB with various oversampling techniques. NB combined with CGAN achieves the best F1-score of 0.9428, while NB combined

TABLE 1 Results of AE-RL on AWID

Method	F1-score	Accuracy	Precision	Recall
AE-RL	0.9334	0.9289	0.9415	0.9289
AE-RL + SMOTE	0.9337	0.9388	0.9398	0.9388
AE-RL + ADASYN	0.9320	0.9402	0.9155	0.9578
AE-RL + CGAN	0.9438	0.9371	0.9592	0.9371

Abbreviations: ADASYN, (ADASYN); AE-RL, adversarial environment reinforcement learning; AWID, Aegean WIFI Intrusion Dataset; CGAN, Conditional Generative Adversarial Network; SMOTE, Synthetic Minority Oversampling Technique.

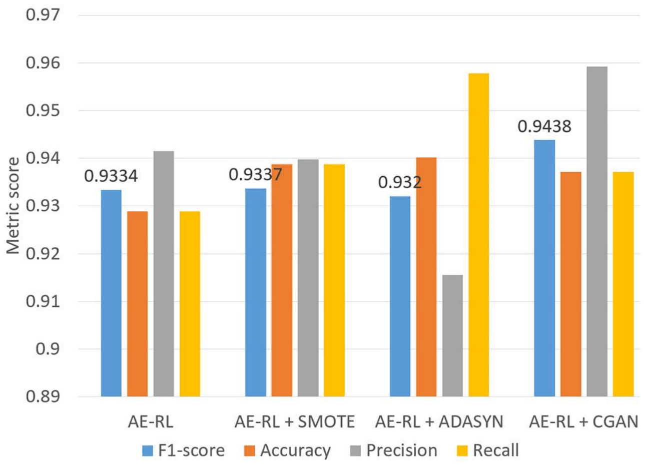

FIGURE 4 Various oversample techniques applied on AE-RL

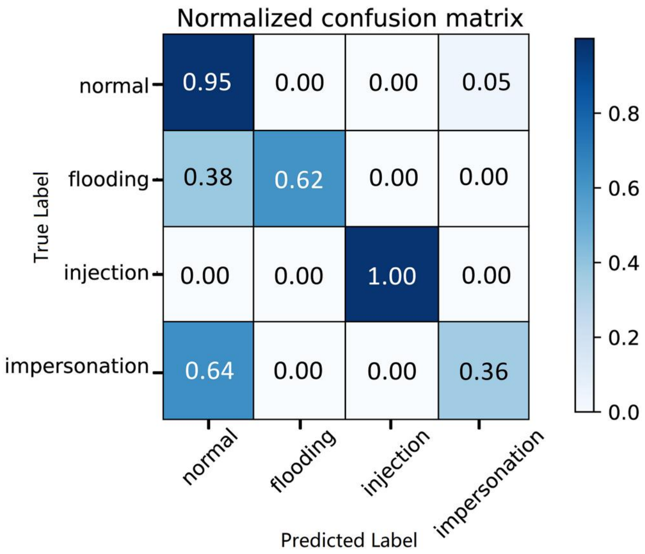

FIGURE 5 Confusion matrix of AE-RL

[1] https://scikit-learn.org/stable/

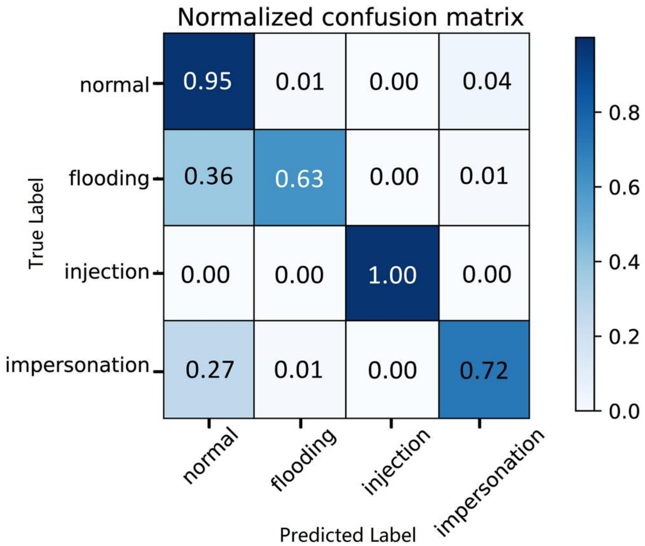

FIGURE 6 Confusion matrix of AEGAN

TABLE 2 Results of NB on AWID

Method	F1-score	Accuracy	Precision	Recall
NB	0.8925	0.8734	0.9154	0.8726
NB + SMOTE	0.8912	0.8731	0.9125	0.8714
NB + ADASYN	0.8757	0.8235	0.9648	0.8232
NB + CGAN	0.9428	0.9504	0.9354	0.9504

Abbreviations: ADASYN, (ADASYN); AWID, Aegean WIFI Intrusion Dataset; CGAN, Conditional Generative Adversarial Network; NB, Naïve Bayes; SMOTE, Synthetic Minority Oversampling Technique.

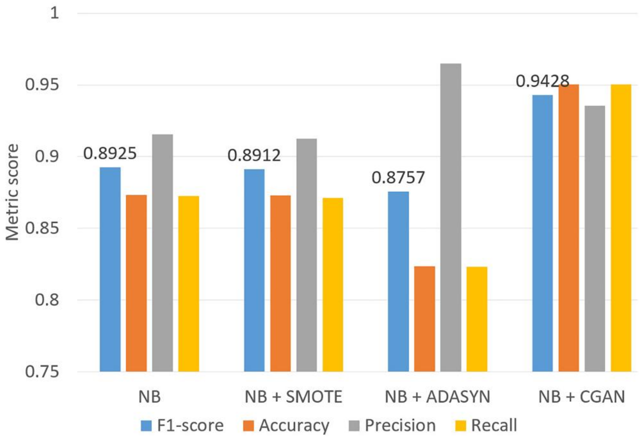

FIGURE 7 Various oversample techniques applied to Naïve Bayes

with SMOTE achieves similar performance to the original NB. Finally, NB combined with ADASYN receives the worst performance. Because NB is a probabilistic model, the distribution of the dataset has a great impact on the classification result. GAN is known for being great at mimicking the distribution of the original dataset whereas SMOTE and ADASYN generate synthetic samples based on distance measures. Therefore, CGAN + NB performs the best compared to applying SMOTE or ADASYN on NB. When the variable independence assumption is enforced, NB has an advantage of having a fast execution time by simply formulating predictions based on the conditional probability model. However, since there is no control put on the augmented portion of the training data, the newly generated samples might change the feature independence. Without this variable independence assumption, the overall performance is not guaranteed.

TABLE 3 Results of MLP on AWID

Method	F1-score	Accuracy	Precision	Recall
MLP	0.9245	0.9324	0.9215	0.9367
MLP + SMOTE	0.9354	0.9314	0.9347	0.9385
MLP + ADASYN	0.9158	0.9226	0.9274	0.9287
MLP + CGAN	0.9523	0.9516	0.9574	0.9516

Abbreviations: ADASYN, (ADASYN); AWID, Aegean WIFI Intrusion Dataset; CGAN, Conditional Generative Adversarial Network; MLP, Multilayer Perceptron; SMOTE, Synthetic Minority Oversampling Technique.

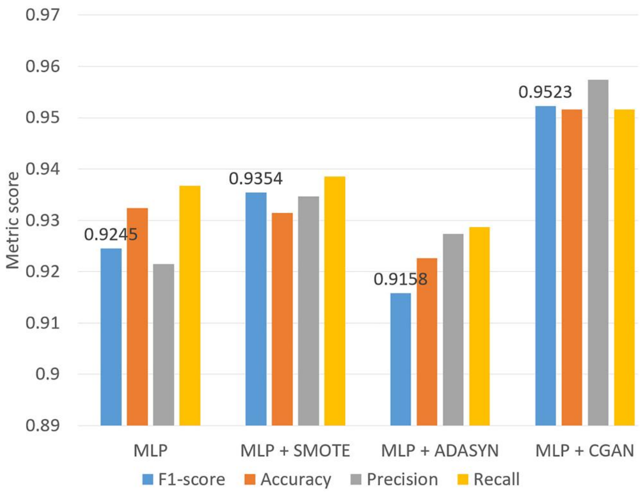

FIGURE 8 Various resample techniques applied to Multilayer Perceptron

Table 3 and Figure 8 illustrate the performance of MLP with various oversampling techniques. The best performance is shown when MLP is combined with CGAN. This combination achieves an F1-score of 0.9523. When MLP is combined with SMOTE, its F1-score ends up reaching 0.9354. When MLP and ADASYN are combined, a slightly worse performance is shown with an F1-score as 0.9158. Although the base AE-RL has a better F1-score than the base MLP, this reverses when applying oversampling, resulting in better performance with MLP combined with CGAN. This demonstrates that applying a data-level approach on a normal classifier can achieve a better result than applying it on an algorithm-level approach.

Table 4 and Figure 9 illustrate the performance of RF when various oversampling techniques are applied. The highest F1-score is achieved by combining RF with CGAN. Its F1-score is 0.9282. When RF is combined with ADASYN, it achieves a better performance than the original RF with F1-score reaching

TABLE 4 Results of RF on AWID

Method	F1-score	Accuracy	Precision	Recall
RF	0.9144	0.9368	0.9347	0.9359
RF + SMOTE	0.8921	0.9238	0.9245	0.9234
RF + ADASYN	0.9242	0.9548	0.9551	0.9548
RF + CGAN	0.9282	0.9473	0.9573	0.9487

Abbreviations: ADASYN, (ADASYN); AWID, Aegean WIFI Intrusion Dataset; CGAN, Conditional Generative Adversarial Network; RF, Random Forest; SMOTE, Synthetic Minority Oversampling Technique.

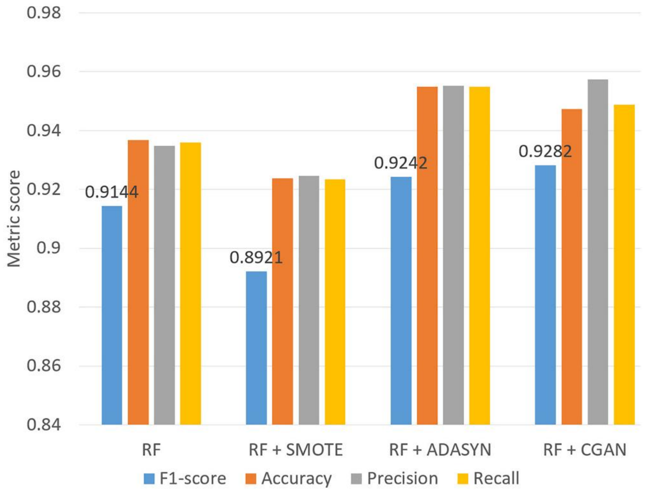

FIGURE 9 Various oversampling techniques applied to Random Forest

0.9242. Since RF is an ensemble of decision trees, it takes the majority decision from the subtrees. When combined with ADASYN, it focusses on the hard-to-learn minority samples, which slightly improves the result. We observe similar results when using the augmented dataset produced using CGAN because it captures the distribution of the minority classes well, which helps to improve the detection rate of the minority classes. On the other hand, the SMOTE-augmented dataset led to worse performance than that of the original dataset. Our intuition is that the augmented dataset creates a bias towards the minority classes that ends up misclassifying some normal class instances. However, further experiments are required to quantify this bias. We will address this in future work.

Table 5 and Figure 10 illustrate the performance of LR when various oversampling techniques are applied. An outstanding performance is achieved when LR is combined with CGAN, reaching an F1-score of 0.985, while LR combined with SMOTE achieves similar performance as the original LR with F1-scores of 0.9568 and 0.9569, respectively. When LR is combined with ADASYN, its performance is comparatively poor reaching an F1-score of 0.9521. We can observe that the overall performance of LR-based classifiers outperforms all other classifiers.

We illustrate a comparison between LR and LR + CGAN using NCMs in Figures 11 and 12 . We observe a significant improvement on one of the minority classes; Flooding: the rate

TABLE 5 Results of LR on AWID

Method	F1-score	Accuracy	Precision	Recall
LR	0.9569	0.9586	0.9663	0.9586
LR + SMOTE	0.9568	0.9503	0.9712	0.9503
LR + ADASYN	0.9521	0.9522	0.9557	0.9522
LR + CGAN	0.9850	0.9849	0.9851	0.9849

Abbreviations: ADASYN, (ADASYN); AWID, Aegean WIFI Intrusion Dataset; CGAN, Conditional Generative Adversarial Network; LR, Logistic Regression; SMOTE, Synthetic Minority Oversampling Technique.

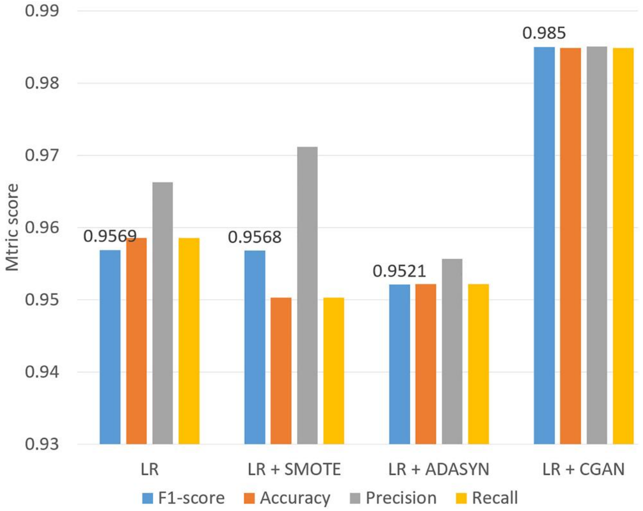

FIGURE 10 Various oversampling techniques applied to Logistic Regression

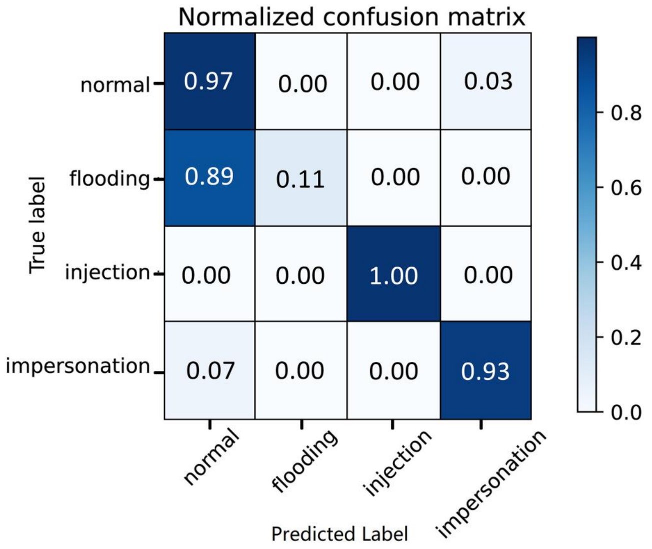

FIGURE 11 Confusion matrix of Logistic Regression on Aegean WIFI Intrusion Dataset

of successful identification of this type of attack is increased from 11% to 62% while eliminating 51% of the false-positive rate on predicting as Normal. Even after the improvement on the detection rate of the Flooding classes, we observe that it is

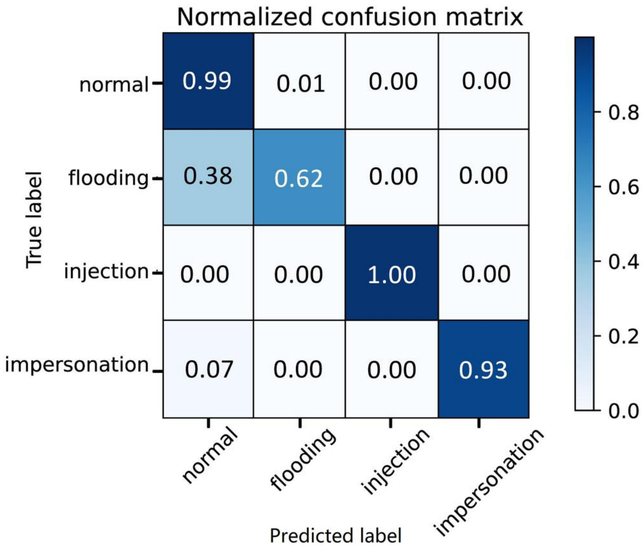

FIGURE 12 Confusion matrix of Logistic Regression + Conditional Generative Adversarial Network on Aegean WIFI Intrusion Dataset

difficult to improve after reaching 62%. Flooding attacks aim to overload the network traffic in order to saturate the victim's system [40]. Since vast numbers of packets following the same network protocols as normal traffic are sent to perform flooding attacks in order to cause a Denial of Service, it is difficult to distinguish such an attack from normal traffic. Meanwhile, the majority class Normal gains further improvement in performance from 97% to 99% while eliminating 3% of false-positive rate on the Impersonation attack class.

In summary, all studied oversampling techniques, namely, SMOTE, ADASYN and CGAN have a positive impact in most of the cases over AE-RL, MLP, RF and LR. A crucial observation is that CGAN always performs better than other oversampling techniques. Furthermore, all classifiers combined with CGAN consistently outperform their original models. Among these results, LR demonstrates the most significant improvement when combined with CGAN.

Lastly, we calculate the population variance for each classifier to inspect the variation in performance induced by oversampling. These measures are calculated using the F1-scores from the base classifiers and each of their oversampling methods. The formula of population variance is as follows:

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$$

where x_i is the sample from each classifier and μ is the mean of the corresponding population. From the results shown in Table 6, we observe that AE-RL has the lowest variance value of 0.000022146 when applying the oversampling techniques. Lower variance values are indicative of less significant changes in the performance of a classifier from the use of oversampling. This suggests that data-level approaches have a less significant impact on algorithm-level approaches compared with other classifiers. Since AE-RL addresses the imbalance problem of the dataset, it does not benefit like other classifiers when data-level approaches are applied. It can further be investigated by testing more algorithm-level approaches. On the other hand, NB has the highest variance value of 0.000638703 after applying oversampling techniques. This is due to NB's rigid assumption of independence, which cannot be avoided in the augmented portion of the training data.

TABLE 6 Population variance on the F1-scores for each classifier

Method	σ^2
AE-RL	0.000022146
NB	0.000638703
MLP	0.000185585
RF	0.000195837
LR	0.000169525

Abbreviations: AE-RL, adversarial environment reinforcement learning; LR, Logistic Regression; MLP, Multilayer Perceptron; NB, Naïve Bayes; RF, Random Forest.

7 | CONCLUSION

IDSs are a critical service that monitor networks for malicious activities through analysing abnormal network traffic. In this study, we propose an IDS architecture that combines a CGAN-based oversampling technique with the existing classifiers. Based on the experimental results, we provide a thorough performance analysis to display the benefit of our models. Overall, most data-level approaches have a positive effect on the performance of classifiers. The results demonstrate that LR combined with CGAN achieves the best F1-score of 0.985 compared with all other classifier-oversampling combinations. The algorithm-level approach AE-RL shows the least amount of variation when combined with oversampling. This suggests there may be limited additional benefit in the combination of data-level approaches with algorithm-level approaches. Expansion of the experiments to include additional algorithm-level approaches is required to further investigate this topic.

As future work, we intend to explore the capabilities of the GAN framework in the context of improving semi-supervised learning. We believe that this can be leveraged to not only create synthetic samples but also to predict class labels. Furthermore, using dimensionality reduction techniques, we hope to speed up the learning process. Moreover, we plan to test our proposed model on other similar datasets to further validate its performance. Finally, we plan to investigate creating a ranking system for majority class samples using a trained CGAN discriminator in order to apply undersampling techniques to the ranked majority class samples in a dataset.

ACKNOWLEDGEMENT

We gratefully acknowledge the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant No. RGPIN-2020-06482.

ORCID

Rahbar Ahsan https://orcid.org/0000-0001-6624-1462
Wei Shi https://orcid.org/0000-0002-3071-8350
Xiangyu Ma https://orcid.org/0000-0002-6956-3851

REFERENCES

1. Bhuyan, M.H., Bhattacharyya, D.K., Kalita, J.K.: Towards generating real-life datasets for network intrusion detection. IJ Netw. Secur. 17(6), 683–701 (2015)
2. Jang-Jaccard, J., Nepal, S.: A survey of emerging threats in cybersecurity. J. Comput. Syst. Sci. 80(5), 973–993 (2014)
3. Bhuyan, M.H., Bhattacharyya, D.K., Kalita, J.K.: Network anomaly detection: methods, systems and tools. IEEE Commun. Surv. Tutor. 16(1), 303–336 (2013)
4. Liu, H., Lang, B.: Machine learning and deep learning methods for intrusion detection systems: a survey. Appl. Sci. 9(20), 4396 (2019)
5. Simeone, O.: A very brief introduction to machine learning with applications to communication systems. IEEE Trans. Cogn. Commun. Netw. 4(4), 648–664 (2018)
6. Chawla, N.V., et al.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357 (2002)
7. Liu, X.Y., Wu, J., Zhou, Z.H.: Exploratory undersampling for class-imbalance learning. IEEE Trans. Syst. Man Cyberne. Part B. 39(2), 539–550 (2008)
8. Batista, G.E., Prati, R.C., Monard, M.C.: A study of the behaviour of several methods for balancing machine learning training data. ACM SIGKDD Explor. Newslett. 6(1), 20–29 (2004)
9. Goodfellow, I., et al.: Generative adversarial nets. Adv. Neural Inf. Process Syst. 27, 2672–2680 (2014)
10. Mirza, M., Osindero, S.: Conditional generative adversarial nets (2014). arXiv preprint arXiv:1411.1784
11. Caminero, G., Lopez-Martin, M., Carro, B.: Adversarial environment reinforcement learning algorithm for intrusion detection. Comput. Netw. 159, 96–109 (2019)
12. Ma, X., Shi, W.: AESMOTE: adversarial reinforcement learning with SMOTE for anomaly detection. IEEE Trans. Netw. Sci. Eng. (2020)
13. Alshawabkeh, M., et al.: Effective virtual machine monitor intrusion detection using feature selection on highly imbalanced data. 2010 Ninth International Conference on Machine Learning and Applications, Washington (2010)
14. Alabdallah, A., Awad, M.: Using weighted support vector machine to address the imbalanced classes problem of intrusion detection system. KSII Trans. Internet Infor. Syst. 12(10) (2018)
15. Sun, C., et al.: A double-layer detection and classification approach for network attacks. International Conference on Computer Communications and Networks (ICCCN), Hangzhou, China (2018)
16. Yuan, Y., Hua, L., Hogrefe, D.: Two layers multi-class detection method for network intrusion detection system. IEEE Symposium on Computers and Communications (ISCC). Heraklion, Greece (2017)
17. Chawla, N., et al.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357 (2002)
18. Zheng, M., et al.: Conditional Wasserstein generative adversarial network-gradient penalty-based approach to alleviating imbalanced data classification. Inf. Sci. 512, 1009–1023 (2020)
19. Seo, J.H., Kim, Y.H.: Machine-learning approach to optimise SMOTE ratio in class imbalance dataset for intrusion detection. Comput. Intell. Neurosci. 2018 (2018)
20. Zhang, H., et al.: An effective convolutional neural network based on SMOTE and Gaussian mixture model for intrusion detection in imbalanced dataset. Comput. Netw. 177, 107315 (2020)
21. He, H., et al.: ADASYN: adaptive synthetic sampling approach for imbalanced learning. 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong (2008)
22. Vu, L., Bui, C.T., Nguyen, Q.U.: A deep learning based method for handling imbalanced problem in network traffic classification. In: Proceedings of the Eighth International Symposium on Information and Communication Technology, pp. 333–339. (2017)
23. Oh, J.H., Hong, J.Y., Baek, J.G.: Oversampling method using outlier detectable generative adversarial network. Expert Syst. Appl. 133, 1–8 (2019)
24. Liu, Y., et al.: Generative adversarial active learning for unsupervised outlier detection. IEEE Trans. Knowl. Data Eng. (2019)
25. Fajardo, V.A., et al.: On oversampling imbalanced data with deep conditional generative models. Expert Syst. Appl. 169, 114463 (2021)
26. Douzas, G., Bacao, F.: Effective data generation for imbalanced learning using conditional generative adversarial networks. Expert Syst. Appl. 91, 464–471 (2018)
27. Ye, J., Fang, Y., Ma, J.: Intrusion detection model based on conditional generative adversarial networks. In: ACAI 2019: Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence, Sanya, China, pp. 350–356. (2019)
28. Dlamini, G., Fahim, M.D.G.M.: A data generative model to improve minority class presence in anomaly detection domain. Neural Comput. Appl. (2021)
29. Fiore, U., et al., Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Infor. Sci. 479, 448–455 (2019)
30. Huang, S., Lei, K.: IGAN-IDS: an imbalanced generative adversarial network towards intrusion detection system. Ad-hoc Netw. 102177 (2020)
31. Lee, J., Park, K.: AE-CGAN model based high performance network intrusion detection system. Appl. Sci. 9(20), 4221 (2019)
32. Yang, Y., et al.: Network intrusion detection based on supervised adversarial variational auto-encoder with regularisation. IEEE Access. 8, 42169–42184 (2020)
33. Watkins, C.J., Dayan, P.: Q-learning. Mach. Learn. 8(3-4), 279–292 (1992)
34. Kolias, C., et al.: Intrusion detection in 802.11 networks: empirical evaluation of threats and a public dataset. IEEECommun. Surv. Tutor. 18(1), 184–208 (2015)
35. Granik, M., Mesyura, V.: Fake news detection using naïve Bayes classifier. 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, Ukraine (2017)
36. Raman, G., Somu, N., Mathur, A.P.: A multilayer perceptron model for anomaly detection in water treatment plants. Int. J. Crit. Infrastruct. Prot. 31, 100393 (2020)
37. Primartha, R., Tama, B.A.: Anomaly detection using random forest: a performance revisited. International Conference on data and Software Engineering (ICODSE). Palembang, Indonesia (2017)
38. Noureen, S.S., et al.: Anomaly Detection in Cyber-Physical System using Logistic Regression Analysis. IEEE Texas Power and Energy Conference (TPEC), College Station, TX, USA, USA (2019)
39. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830 (2011)
40. Vidal, J.M., Orozco, A.L.S., Villalba, L.J.G.: Adaptive artificial immune networks for mitigating DoS flooding attacks. Swarm and Evolutionary Computation. 38, 94–108 (2018)

How to cite this article: Ahsan, R., et al.: A comparative analysis of CGAN-based oversampling for anomaly detection. IET Cyber-Phys. Syst., Theory Appl. 1–11 (2021). https://doi.org/10.1049/cps2.12019