Hindawi
Security and Communication Networks
Volume 2021, Article ID 4259629, 10 pages
https://doi.org/10.1155/2021/4259629

Review Article

A Survey of Few-Shot Learning: An Effective Method for Intrusion Detection

Ruixue Duan,[1,2] **Dan Li**,[1] **Qiang Tong**,[1,3] **Tao Yang**,[1] **Xiaotong Liu**,[1,3] **and Xiulei Liu**[1,3]

[1]*Beijing Advanced Innovation Center for Materials Genome Engineering, Beijing Information Science and Technology University, Beijing 100101, China*
[2]*Beijing Laboratory of National Economic Security Early-Warning Engineering, Beijing 100044, China*
[3]*Laboratory of Data Science and Information Studies, Beijing Information Science and Technology University, Beijing 100101, China*

Correspondence should be addressed to Xiulei Liu; liuxiulei@bistu.edu.cn

Received 12 September 2021; Accepted 5 October 2021; Published 31 October 2021

Academic Editor: Yunchuan Guo

Few-shot learning (FSL) is a core topic in the domain of machine learning (ML), in which the focus is on the use of small datasets to train the model. In recent years, there have been many important data-driven ML applications for intrusion detection. Despite these great achievements, however, gathering a large amount of reliable data remains expensive and time-consuming, or even impossible. In this regard, FSL has been shown to have advantages in terms of processing small, abnormal data samples in the huge application space of intrusion detection. FSL can improve ML for scarce data at three levels: the data, the model, and the algorithm levels. Previous knowledge plays an important role in all three approaches. Many promising methods such as data enrichment, the graph neural network model, and multitask learning have also been developed. In this paper, we present a comprehensive review of the latest research progress in the area of FSL. We first introduce the theoretical background to ML and FSL and then describe the general features, advantages, and main methods of FSL. FSL methods such as embedded learning, multitask learning, and generative models are applied to intrusion detection to improve the detection accuracy effectively. Then, the application of FSL to intrusion detection is reviewed in detail, including enriching the dataset by extracting intermediate features, using graph embedding and meta-learning methods to improve the model. Finally, the difficulties of this approach and its prospects for development in the field of intrusion detection are identified based on the previous discussion.

1. Introduction

Machine learning (ML) is a method that allows machines to perform functions that cannot be achieved by direct programming. Its main purpose is to allow computers to learn from data without human intervention, and to adjust the internal parameters of the algorithm accordingly for regression or classification purposes. ML enables computers to automatically extract useful patterns without explicit programming, and to predict the unseen and take action by learning from existing experience [1]. Here, experience refers to information that is previously obtained by learners through observation, collection, etc. and is typically in the form of data that can be analyzed. However, the quality and quantity of data are critical to the prediction results generated by learners.

ML has made significant progress in various fields such as computer vision [2], natural language processing [3], and robot decision-making [4], and methods based on ML are currently creating breakthroughs in the field of intrusion detection. In recent years, due to the rapid development of information technology, the exploitation of vulnerabilities in networks and information systems for intrusion [5] are likely to pose significant threats to the security of the Internet and will cause serious losses to countries, enterprises, and individuals. Network security has, therefore, become a hot topic, and increasing numbers of defensive technologies are constantly being upgraded and improved. Intrusion

detection [6] is used to detect attacks or attempts to violate the established security strategy in a computer network or system, to help the system deal with network attacks. Intrusion Detection Systems (IDSs) [5] were developed to identify malicious traffic and detect network attacks, and they are an important means of maintaining network security. Some scholars have applied various ML methods in the field of IDS research, such as approaches based on Random Forest (RF) [7], Support Vector Machine (SVM), and neural networks [8]; as long as the number of training samples is sufficient, unknown types of attack can be detected in order to deal with the threat of an intrusion that has never occurred before. Lee and Stolfo [9] proposed a supervised learning-based framework called MADAM ID that used a large amount of audit data to calculate activity patterns and construct features and used ML methods to generate intrusion detection models. Hwang et al. [10] proposed a three-layer IDS that combined a blacklist, a whitelist, and a Smooth SVM (SSVM) classifier. The blacklist was used for the initial screening of known attacks, while the whitelist was used to identify normal traffic to reduce the false positive rate, and the abnormal traffic was classified using an SSVM. Li et al. [11] proposed an intrusion detection model based on ensemble learning, which effectively improved the accuracy of detection by integrating the output results of three classifiers: Bayes, J48, and SVM.

An IDS based on ML is essentially a classification system and requires the construction of a classifier by learning from a training set to distinguish normal behaviors from abnormal ones. Traditional intrusion detection algorithms are based on supervised learning and usually only perform well on tasks that contain a large number of available examples. When the size of the training sample is reduced, the ML model often cannot be reliably generalized [12]. Network attacks represent anomalous samples, and the proportion of attack samples in the entire network traffic is extremely small. For example, mainstream datasets in the field of host intrusion detection (such as the number of abnormal samples published by ADFA and UMN) are far smaller than normal samples, resulting in a serious data imbalance problem. Although many scholars have tried to apply the various ML algorithms that have emerged in recent years to intrusion detection, their results have not been ideal in terms of the detection of attack samples (few-shot samples) [13]. Since traditional ML classification algorithms are difficult to train effectively on small samples, lower detection rates and higher missing report rates may occur in imbalanced datasets, which affects the output results of the classification model. The issue of how to deal with imbalanced datasets and improve the accuracy of detection is a major problem in the field of intrusion detection [14].

In view of the problems described above, some scholars have conducted a great deal of research on methods of undersampling, oversampling, and mixtures of undersampling and oversampling [15]. Undersampling [16] may cause the loss of important features in the process of reducing the number of normal samples. Chawla et al. [17] were the first to propose an oversampling method, using a small number of synthesis algorithms. This method requires the user to decide how many neighbor samples to choose, so there is a certain degree of blindness in operation. In addition, if the negative sample is at the edge of the distribution of the negative sample set, the "artificial" samples will be generated nearby the negative sample and adjacent samples, resulting in the problem of data marginalization and increased difficulty of classification by the classification algorithm eventually. Liu et al. [18] proposed a hybrid intrusion detection method that combined a Convolutional Neural Network (CNN) with SMOTE oversampling and undersampling techniques. PCA dimensionality reduction, k-means, and other technologies were combined to reduce the missing report rate by 16% when detecting abnormal samples. However, this method was not suitable for discretely distributed data, which cannot be used to synthesize minority samples of the same type based on nearest-neighbor relationships.

In recent years, researchers have developed the technique of few-shot learning (FSL) [19], in which the aim is to train the model quickly on small numbers of labelled data. The core idea underlying FSL is similar to the process of learning by humans: it uses previously acquired information to learn new tasks, thereby reducing the amount of data required for training. Research results show that FSL can overcome some of the shortcomings of ML, for example, by significantly reducing the amount of data needed to train ML models and thereby reducing the time required to collect and label large datasets. Areas of application include image classification [20], image retrieval [21], and gesture recognition [22]. Fink [23] proposed a framework that could learn object classifiers from a single example, meaning that the data collection work involved in image classification tasks could be greatly reduced. In addition, FSL makes it possible to train models that are suitable for certain rare situations. A typical example is the huge contribution to the field of drug discovery that have been made based on the latest research progress in FSL [24]. When predicting whether a given molecule is toxic, clinical biological data on the molecule are extremely limited, which often hinders traditional ML methods. Altae-Tran et al. [25] introduced a scheme called Iterative Refinement Long Short-Term Memory (*Iter-RefLSTM*), based on one-shot learning, with the aim of achieving accurate prediction on small datasets. Their experiments proved that a one-shot learning method could transfer information between related but completely different learning tasks and that it had strong generalizability. In many applications, it is extremely important to be able to learn efficiently from small samples.

Anomalous samples are scarce in the field of intrusion detection, and researchers often face the problem of an imbalance in the numbers of normal and abnormal samples. The traditional ML model has a low detection accuracy for abnormal samples with small amounts of data. The development of a technology that can effectively handle limited datasets is, therefore, urgent. FSL is a general term for this type of technology; it can use the few available data to train models and thus improve the efficiency of data utilization and, therefore, has advantages in terms of intrusion detection.

The rest of the paper is organized as follows. In Section 2, an overview of FSL is presented, including a definition, application scenarios, and its advantages. In Section 3, the current work related to FSL is described and classified from three perspectives. In Section 4, examples of the application of FSL in the field of intrusion detection are described in detail. We discuss the challenges faced by the current small sample learning research progress and propose possible directions for future research in Section 5. Finally, conclusions are drawn in Section 6.

2. Overview of Few-Shot Learning

The main goal of ML is to learn by acquiring new knowledge from external sources, and to make this more effective by modifying existing knowledge. In this process, data are refined into "knowledge" and stored for future use [26]. Although ML has achieved great success in many respects, it depends on the availability of massive amounts of data. Standard ML techniques, such as RF and simple deep networks, need to learn meaningful information from thousands of samples [27]. Driven by the academic goal of artificial intelligence approaching the level of human reasoning, and in order to reduce the cost of learning, FSL has attracted widespread attention in recent years. In this section, we will first explain what FSL is, and then we will describe three application scenarios. Finally, we summarize the advantages of FSL.

2.1. What Is Few-Shot Learning? FSL refers to the process of training a model with extremely little training data. In a case where extremely limited training examples are available, this approach can use previously acquired knowledge to improve the performance on new tasks [28]. Due to constraints arising from privacy, ethics, security, computing resources, etc., there may be extremely few examples of an instance, which makes it impossible to obtain sufficient training data for learning in some fields [12]. When insufficient supervision information is available for the target task, there is a decrease in the performance of the model, meaning that supervised learning methods [29] are often unsuccessful when small datasets are used.

Unlike traditional ML methods, FSL uses information that has already been learned as prior knowledge and experience to help in learning new tasks. This makes it possible to learn rare tasks with only a small amount of labelled data [19].

2.2. Advantages of Few-Shot Learning. Humans are able to generalize new knowledge based on only few examples, while artificial intelligence usually requires thousands of examples to achieve similar results. Inspired by the rapid learning ability of humans, researchers aimed to develop a ML model that could learn a new category quickly with only a few sample data after it has learned a large quantity of data for certain categories. This is the problem that FSL aims to solve. Such as the generation of character samples [30], this task only provides a few observable examples.

Research has shown that FSL can combine one or a few examples with previously learned information to learn rare situations where it is difficult to obtain supervision information. For example, FSL can achieve a high level of prediction accuracy on drug discovery tasks with few biological data [25]. This can address the problem of the need for thousands of supervised samples by the ML algorithm to ensure the generalization ability of the model. The use of FSL can also help reduce the high cost of data preparation and shorten the training cycle.

3. Methods of Few-Shot Learning

The existing FSL problem is a supervised learning problem that requires the use of labelled data for training [29]. Since labelled data are usually scarce, FSL must use prior knowledge. Based on the use of prior knowledge, FSL methods can be classified from three perspectives, as data, model, or tuning algorithms.

3.1. Data Algorithms. The first type of FSL method focuses on data. Humans can typically successfully learn new knowledge from one or a few instances, while ML algorithms usually require dozens or hundreds of examples to achieve similar accuracy [30]. These methods enrich the empirical supervision information by expanding the dataset. A model trained directly on the FSL task will encounter the problem of overfitting [31], and the use of more data can avoid the phenomena of underfitting and overfitting [12].

In the following, we describe three FSL methods based on data, and these are shown in Figure 1. This kind of method expands the training sets by transforming other datasets such as similar datasets and weakly supervised or unsupervised datasets.

Extensive use of data from other tasks: Qi et al. [32] proposed an FSL method that used imprint weights for the classification of multilabel small samples. In order to allow the classifier to achieve learning capabilities similar to humans, the method directly sets the final layer weights based on the new training samples added.

Transforming the samples in the dataset related to the task: this type of method adjusts input-output pairs from similar training sets that contain more data to training samples for specific tasks to increase the amount of data. It is usually difficult to perform one-shot learning with Deep Neural Networks (DNNs) [33], since each category learned at a time has only one labelled training example, and it is difficult to meet the data requirements for training. Tsai and Salakhutdinov [34] proposed that the information lost between classes could be compensated for by fusing supplementary information, thereby improving the effect of one-shot learning. A mechanism was introduced to effectively integrate the instances belonging to the "large number of examples" category into the data representation learning, so that they could be treated as quasi-instances of the "one example" category.

Converting samples from weakly supervised or unsupervised datasets: in the field of gesture recognition [22], most of the current methods rely on strongly supervised

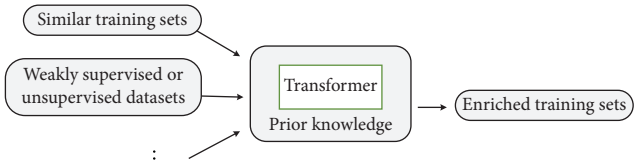

Figure 1: Methods of solving FSL problems from a data perspective.

learning and require large amounts of training data; these consume a lot of resources and cannot be extended to the ever-developing large-scale gesture language. Currently, a gesture classifier can be trained on a weakly supervised example library, where a single example and the weakly supervised library come from different domains.

3.2. Model Algorithms. The second type of FSL method focuses on the model. Prior knowledge is used to constrain the complexity of the hypothesis space, thereby giving a relatively small hypothesis space and reducing the prediction errors that occur in ML problems, so that small datasets can also give accurate training results.

There are three FSL methods that focus on the model, and these are described below.

In *Multitask Learning* (MTL), learned information is shared by learning multiple related tasks at the same time. This means that MTL can give better generalization performance than single-task learning [35]. This is mainly achieved through the use of shared representation when learning tasks in parallel; the content learned for each task can help improve the performance on other tasks [36]. Zhou and Zhao [37] proposed the Clustered Multitask Learning (CMTL) approach, which described an arbitrary task with multiple representative tasks to give an accurate representation. Sun et al. [38] combined MTL with oversampling for intrusion detection. Their method used MTL to learn relevant information from multiple tasks at the same time and then used the learned information for a single task. Experiments showed that the detection rate using this method could reach more than 90%.

In *embedded learning*, each sample is embedded into a lower dimension, and embedding needs to support embedded g to define query embedding f. Similar samples are closer in this low dimension, so that a smaller hypothesis space can be constructed, meaning that fewer training samples are needed. Altae-Tran et al. [25] proposed *Iter-RefLSTM* to allow better embedding functions to be designed and achieved accurate prediction of molecular toxicity with a small dataset through the use of one-shot learning. Zerhoudi et al. [39] supplemented the dataset with semantic descriptions and used the graph embedding method to encode these semantic descriptions. The authors expanded the traditional IDS to realize zero-shot detection.

Generative model [40] methods use prior knowledge to estimate the probability distribution Px of an input value x_i. There is usually a latent variable z whose prior probability Pz is learned from other tasks. The posterior probability Px is obtained by combining the dataset and Pz when learning a new task, meaning that Px is constrained. A hierarchical Bayesian model [41] can learn categories from a single training instance. This model groups datasets based on the similarity between tasks, and all the datasets for each group are then used together for learning, to obtain the prior probability of the class. When an example of a new category is given, the group to which the new category belongs is first inferred, and then the prior probability extracted from this group is used to model it. Lopez-Martin et al. [42] proposed a Variational Generative Model (VGM) that used a specific class label as input to synthesize new data with the same label. Although the generated training samples are different from the real samples, they have a high level of similarity. The use of synthetic data effectively improves the performance of the ML classifier.

3.3. Tuning Algorithms. The third type of FSL method focuses on the algorithm. When the number of training samples is insufficient, the strategy of searching for the parameter θ of the best hypothesis h^* in the hypothesis space H is adapted based on the use of prior knowledge, so that more reliable training results can be obtained for the dataset.

In the following, we introduce three FSL methods that focus on the use of a parameter adjustment algorithm.

The method of *refining existing parameters* takes a parameter θ_0 learned from the related task as the initialization parameter and uses the training set for the new task to optimize θ_0 to make it suitable for the task. Wang and Hebert [43] proposed a method that used unsupervised data sources to group similar samples into a group. Unlabeled data were used for pretraining in order to cluster and separate samples, and CNNs were then applied to adapt the pretraining results to the small sample training set of the new task. This approach could identify new categories from a few examples.

In the method of *refining meta-learned parameters* [44], the initialization parameter θ_0 is meta-learned from a set of tasks, which must have the same distribution PT as the new task. The meta-learner continuously improves θ_0 by using the training set to quickly adapt to new tasks. The model-independent meta-learning scheme proposed by Finn et al. [45] is an iterative process. The meta-learning parameters are adjusted to a specific task using a gradient descent optimization algorithm [46]. This approach can be applied to all models trained with gradient descent and can, therefore, be used to solve small sample learning tasks.

The *optimizer learning* method is different from the method of refining meta-learned parameters. It does not use the gradient descent algorithm to adjust the initialization parameter θ_0; instead, a general optimizer is trained that adapts to multiple tasks and directly provides learners with an updated search strategy. The optimization model [47] trains a meta-learner as an optimizer using a set of tasks obtained from PT and directly outputs the error signal obtained from each iteration to update the parameters of a specific task, to improve the search strategy.

4. Research Progress in Applying Few-Shot Learning to Intrusion Detection

Some researchers have applied ML to IDS and have obtained higher accuracy rates compared with other detection methods. However, ML relies on statistical algorithms, the performance of which improves with training, and large amounts of data are typically required to allow them to learn effectively. In the field of intrusion detection, the scarcity of abnormal samples means that traditional ML cannot further improve the detection results. For example, U2R traffic samples account for only 0.04% of the KDDTrain + dataset, resulting in significantly lower detection rates compared to other samples [48, 49]. When a limited number of attack samples are used for intrusion detection, this is known as few-shot detection. FSL is designed to deal with limited training samples. Intrusion detection methods using FSL can effectively solve the problem of the small number of abnormal samples, thereby improving the detection rate on abnormal samples and the overall accuracy. This section will focus on recent research work on the application of FSL to intrusion detection.

4.1. Few-Shot Deep Learning. Chowdhury et al. [50] combined traditional oversampling and undersampling techniques with FSL to improve the detection rate of abnormal samples. They first trained a deep CNN for the extraction of intermediate features and then input the extracted features as a new dataset into SVM and 1-NN classifiers. The process used in this method is shown in Figure 2. After testing on the KDD and NSL-KDD datasets, compared with the previous literature, this method obtained the best performance. The test accuracy of the SVM classifier reached 94.62% when using the features extracted by 13 layers, and the average classification accuracy was the highest on the nine times oversampling U2R. This showed that the use of FSL could improve the overall performance of intrusion detection while improving the classification accuracy of few-shot samples.

4.2. Few-Shot Detection Based on Meta-Learning. Some scholars have proposed that meta-learning can be used to solve the FSL problem [51], and this approach is also applicable to intrusion detection. Xu et al. [52] developed an end-to-end DNN called FC-Net based on meta-learning, which was used for few-shot network traffic classification tasks. The network consisted of two parts, F-Net and C-Net, which were used to extract and compare features, respectively. The overall architectures of F-Net and C-Net are shown in Figures 3 and 4, respectively. Since FC-Net obtains sufficient prior knowledge for classification from the original dataset, only a small number of label samples are needed to achieve better performance in the task of detecting new types of attacks. Two reconstructed datasets, ISCX2012FS and CICDIS2017FS, were used to evaluate the performance of FC-Net and to compare it with other methods. FC-Net achieved the highest average detection rate of 98.88% by testing on the training set; when testing and training were carried out using different datasets, an even higher average detection rate of 99.62% could be achieved. FC-Net not only improved the classification accuracy of few-shot detection, but also successfully detected new types of attacks that contained only a small number of labelled samples based on the learned prior knowledge.

4.3. Siamese Network Based on One-Shot Learning. Hindy et al. [53] proposed a Siamese network model based on the one-shot learning [19] method, which used a similarity comparison to identify new categories of attack that had not previously appeared. The modelling process used in this method is shown in Figure 5. The performance of the model was evaluated on three datasets: CICIDS2017, KDD Cup '99, and NSL-KDD [54]. The accuracy rate on the CICIDS2017 dataset was more than 80%, while, for KDD Cup '99 and NSL-KDD, it reached more than 72%. This demonstrates that one-shot learning makes it possible to recognize new types of attacks without retraining, and it can achieve acceptable accuracy.

4.4. A Gated Few-Shot Learning Model for Anomaly Detection. Huang et al. [55] proposed a FSL model for anomaly detection, for which the overall architecture is shown in Figure 6. This model used existing data for training, and there is no need for retraining after the introduction of a small amount of abnormal data, and through the gating structure to aggregate the known and unknown abnormal types, so as to solve the problem of dataset imbalance. The architecture of the gated FSL model is shown in Figure 7. The author used the NSL-KDD dataset to conduct experiments to evaluate the performance of the model. Compared with other baseline models, the gated FSL model achieved a higher overall accuracy. When the unknown anomaly class contained five labelled data points, the detection accuracy of this model reached 97.49%. The experimental results showed that the gated FSL model effectively improved the classification performance of the overall and unknown anomaly classes and did not greatly affect the accuracy of the known anomaly classes. By comparing the experimental results with and without a gated structure, it was found that better performance could be obtained by using a gated structure.

4.5. Zero-Shot Recognition with Graph Embedding. When a task has no training samples with supervised information, FSL is transformed into zero-shot learning (ZSL) [56]. There are recent incidents in which insiders have circumvented network security abound and insider threats have therefore received widespread attention. Zerhoudi et al. [39] used graph embedding to integrate ZSL into an existing IDS for insider threat detection, and the overall architecture is shown in Figure 8. The lower part of the figure is the baseline system, and the upper part is an extension of the baseline system based on ZSL. The authors designed two real scenarios to evaluate their method. When budgets are 3500, the use of graph embedding increased the normalized cumulative recall [57] to 0.63 in an application scenario in which

Figure 2: Flow chart of few-shot deep learning approach for intrusion detection.

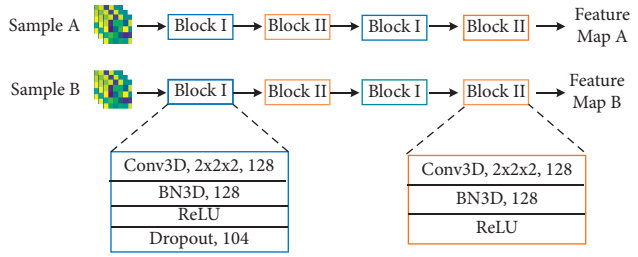

Figure 3: The architecture of F-Net.

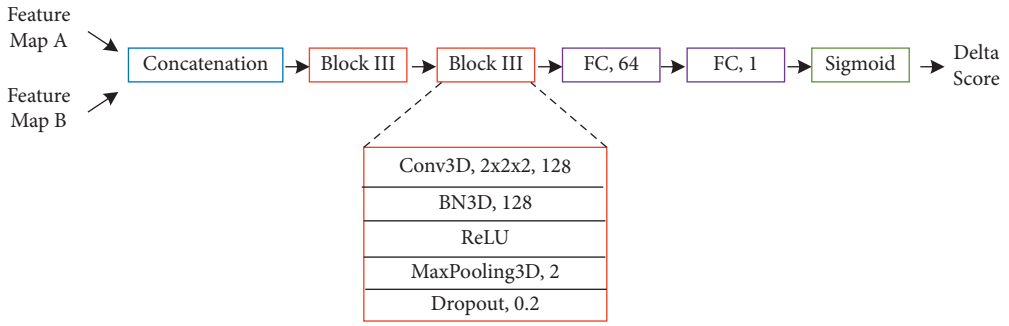

Figure 4: The architecture of C-Net.

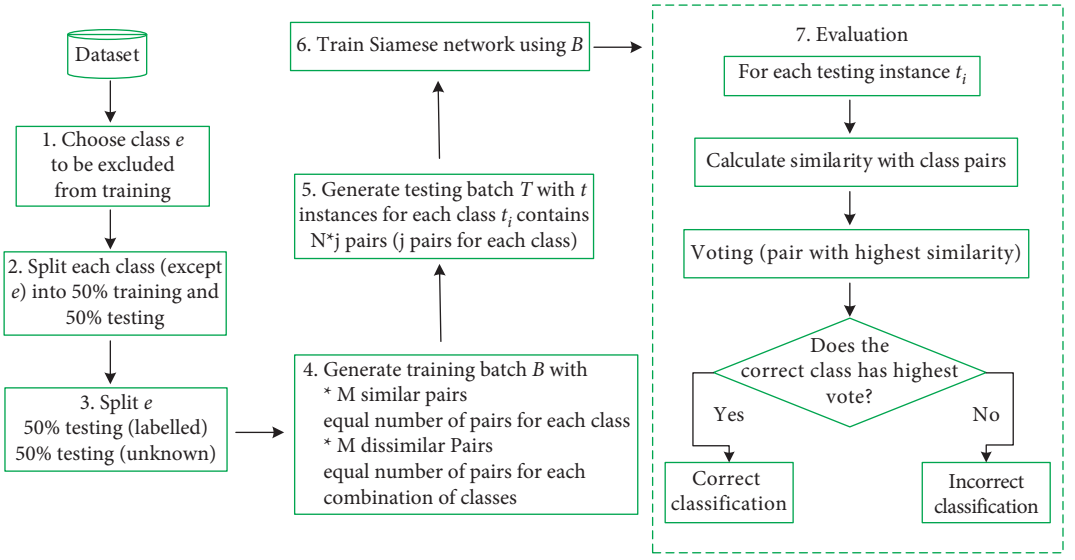

Figure 5: The process of establishing a one-shot Siamese network model for IDS.

new users were added; this was increased further to 0.65 when graph embedding was used in a scenario involving project changes. These results prove that, in the absence of available data, the use of graph embedding is an effective means of identifying abnormal behaviors. Compared with the baseline model that does not use graph embedding, the model that uses graph embedding in the literature can improve the detection performance.

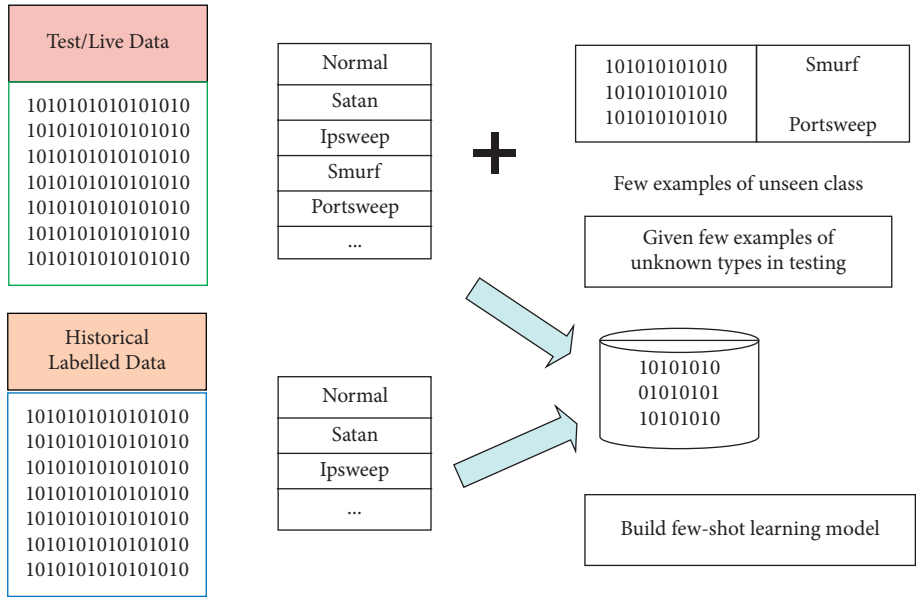

Figure 6: Architecture of FSL for anomaly detection.

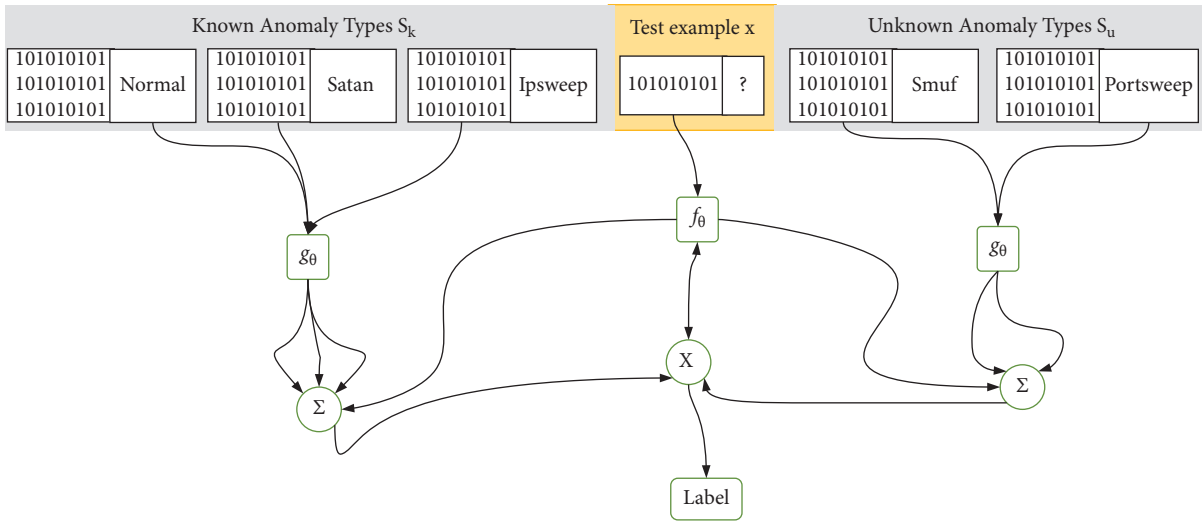

Figure 7: Architecture of the gated FSL model.

5. Open Challenges and Directions for Future Research

Most IDSs rely on statistical characteristics and the knowledge of domain experts and cannot respond well to unknown attacks, meaning that they cannot effectively maintain network security [58]. Current research results have proved that ML has unique advantages in the field of intrusion detection and can help identify new types of attacks; however, the training and performance of ML models are highly dependent on the availability of sufficient amounts of labelled data. The number of abnormal behaviors that can be used as training data in intrusion detection is extremely limited and is not adequate to allow traditional ML methods to yield high performance [59], thus making few-shot detection necessary. The issue of how to improve the accuracy of ML algorithms on small datasets has been the subject of intensive research in recent years, and FSL has played a huge role in this. Several methods have been developed for intrusion detection based on small datasets, and new research results continue to emerge. The issue of how to apply FSL to intrusion detection more comprehensively, from the three perspectives of the data, model, and algorithm, will be a key research direction in future.

Although FSL has now begun to be used in the field of intrusion detection, it still has certain shortcomings. For example, when an FSL model that focuses on the data perspective is used, its ability is extremely limited, although there are many ways to enhance the dataset, which to a certain extent can alleviate the problem of insufficient data. Each enhancement strategy is only designed for specific datasets and cannot be extended to others. To solve this

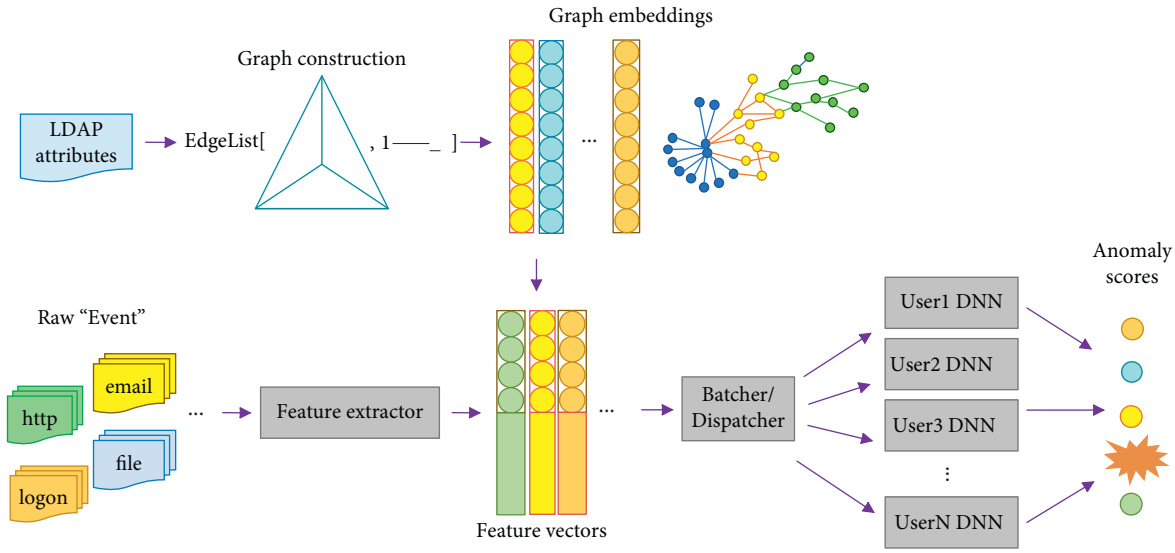

Figure 8: Architecture of IDS based on ZSL graph embedding.

problem, an automatic enhancement approach [60] has recently been proposed that can automatically search for improved data enhancement strategies for deep network training. The use of a model with a small hypothesis space to solve the FSL problem is usually only suitable for certain simple tasks and will cause a large approximation error for complex tasks; this method is therefore unsuitable for attribute prediction problems with many features. In methods in which parameters are optimized through iterative updates, a decrease in accuracy is often seen, since θ_0 is not learned from the current task. Moreover, these methods are often hindered by insufficient amounts of feature data. In addition, the choice of which strategy to use depends on the specific application, and different solutions need to be selected for different types of learning tasks.

As the range of applications of FSL within the field of intrusion detection continues to widen, one future direction for development would be to combine the advantages of multiple solutions and to develop general models that are suitable for different types of tasks, so that FSL can be applied in a more convenient and efficient way to intrusion detection problems.

6. Conclusion

ML refers to the process of using algorithms to guide a computer to construct an appropriate model based on known data, and the use of this model to assess a new situation. However, since the ML model contains numerous parameters, large amounts of labelled data are typically required to train the model, which severely limits its applications. In many scenarios, it is extremely expensive, difficult, or even impossible to collect large amounts of labelled data (for example, medical data, data that are manually annotated by users on mobile phones, and abnormal behavior data for intrusion detection). The question of whether a good model can be obtained by training with only a small amount of labelled data is an important one, and FSL, therefore, forms a key aspect of the development of ML. It can be used to learn rare cases with a small number of features, an approach that can effectively solve the problem of dataset imbalance in intrusion detection. In this paper, we have surveyed and summarized the research efforts that have been made in the field of FSL. We first introduced and defined FSL, in which new tasks are learned by using previously obtained information. We then described applicable scenarios and the advantages of FSL. This approach was shown to be suitable for learning rare tasks, for which only small amounts of training data are available, and for reducing the burden of collecting the massive datasets required for ML. We have discussed the applicability of FSL to the field of intrusion detection and have described in detail some specific examples of applications in this field. Studies have shown that FSL can greatly improve the detection performance on small datasets. Finally, we have identified the challenges faced by FSL and suggested future research directions for this promising approach.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper was supported by the funding of Beijing Natural Science Foundation (Grant No. 4204100), School Foundation of Beijing Information Science and Technology University (Grant No. 1825023), Subject and Graduate Education Program of Beijing Information Science and Technology University (Grant No. 5112111004), and 2021 Promote Connotation Development of Universities-Scientific Research Training Program for College Students of BISTU (Grant No. 5102110805).

References

[1] E.-S. M. El-Alfy and S. A. Mohammed, "A review of machine learning for big data analytics: bibliometric approach," *Technology Analysis & Strategic Management*, vol. 32, no. 8, pp. 984–1005, 2020.

[2] J. Janai, F. Güney, A. Behl, and A. Geiger, "Computer vision for autonomous vehicles: problems, datasets and state of the art," *Foundations and Trends® in Computer Graphics and Vision*, vol. 12, no. 1–3, pp. 1–308, 2020.

[3] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: an introduction," *Journal of the American Medical Informatics Association*, vol. 18, no. 5, pp. 544–551, 2011.

[4] J. Morison and A. Harkens, "Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making," *Legal Studies*, vol. 39, no. 4, pp. 618–635, 2019.

[5] H.-J. Liao, C.-H. Richard Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: a comprehensive review," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 16–24, 2013.

[6] R. A. Kemmerer and G. Vigna, "Intrusion detection: a brief history and overview," *Computer*, vol. 35, no. 4, pp. supl27–supl30, 2002.

[7] P. A. A. Resende and A. C. Drummond, "A survey of random forest based methods for intrusion detection systems," *ACM Computing Surveys*, vol. 51, no. 3, pp. 1–36, 2018.

[8] S. Mukkamala, G. Janoski, and A. Sung, "Intrusion detection using neural networks and support vector machines," in *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, vol. 2, pp. 1702–1707, Honolulu, HI, USA, May 2002.

[9] W. Lee and S. J. Stolfo, "A framework for constructing features and models for intrusion detection systems," *ACM Transactions on Information and System Security*, vol. 3, no. 4, pp. 227–261, 2000.

[10] T. S. Hwang, T. J. Lee, and Y. J. Lee, "A three-tier IDS via data mining approach," in *Proceedings of the 3rd annual ACM workshop on Mining network data*, pp. 1–6, San Diego, CA, USA, June 2007.

[11] Y. Li, J. L. Li, S. J. Yue, and Z. Wang, "Research of intrusion detection based on ensemble learning model," *Applied Mechanics and Materials*, Trans Tech Publications Ltd, vol. 336-338, pp. 2376–2380, 2013.

[12] W. Wang, L. Zhang, M. Zhang, and Z. Wang, "Few shot learning for multi-class classification based on nested ensemble DSVM," *Ad Hoc Networks*, vol. 98, Article ID 102055, 2020.

[13] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: review of methods and applications," *Expert Systems with Applications*, vol. 73, pp. 220–239, 2017.

[14] R. Barandela, J. S. Sánchez, V. Garćia, and E. Rangel, "Strategies for learning in class imbalance problems," *Pattern Recognition*, vol. 36, no. 3, pp. 849–851, 2003.

[15] S. Cateni, V. Colla, and M. Vannucci, "A method for resampling imbalanced datasets in binary classification tasks for real-world problems," *Neurocomputing*, vol. 135, pp. 32–41, 2014.

[16] X. Y. Liu, J. Wu, and Z. H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 2, pp. 539–550, 2008.

[17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[18] K. Liu, Z. Fan, M. Liu, and S. Zhang, "Hybrid intrusion detection method based on k-means and cnn for smart home," in *Proceedings of the 2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, pp. 312–317, IEEE, Tianjin, China, July 2018.

[19] L. Li Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 594–611, 2006.

[20] S. Biswas and S. Barma, "A large-scale optical microscopy image dataset of potato tuber for deep learning based plant cell assessment," *Scientific Data*, vol. 7, no. 1, pp. 371–411, 2020.

[21] A. Tompary, W. Zhou, and L. Davachi, "Schematic memories develop quickly, but are not expressed unless necessary," *Scientific Reports*, vol. 10, no. 1, pp. 16968–17017, 2020.

[22] M. Yasen and S. Jusoh, "A systematic review on hand gesture recognition techniques, challenges and applications," *PeerJ Computer Science*, vol. 5, Article ID e218, 2019.

[23] M. Fink, "Object classification from a single example utilizing class relevance metrics," *Advances in Neural Information Processing Systems*, vol. 17, pp. 449–456, 2005.

[24] J. Jiménez-Luna, F. Grisoni, and G. Schneider, "Drug discovery with explainable artificial intelligence," *Nature Machine Intelligence*, vol. 2, no. 10, pp. 573–584, 2020.

[25] H. Altae-Tran, B. Ramsundar, A. S. Pappu, and V. Pande, "Low data drug discovery with one-shot learning," *ACS Central Science*, vol. 3, no. 4, pp. 283–293, 2017.

[26] D. M. Dutton and G. V. Conroy, "A review of machine learning," *The Knowledge Engineering Review*, vol. 12, no. 4, pp. 341–367, 1997.

[27] G. Subramanian, B. Ramsundar, V. Pande, and R. A. Denny, "Computational modeling of β-s-1) inhibitors using ligand based approaches," *Journal of Chemical Information and Modeling*, vol. 56, no. 10, pp. 1936–1949, 2016.

[28] Y. Xie, H. Wang, B. Yu, and C. Zhang, "Secure collaborative few-shot learning," *Knowledge-Based Systems*, vol. 203, Article ID 106157, 2020.

[29] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," *EURASIP Journal on Applied Signal Processing*, vol. 2016, no. 1, pp. 1–16, 2016.

[30] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "Human-level concept learning through probabilistic program induction," *Science*, vol. 350, no. 6266, pp. 1332–1338, 2015.

[31] H.-J. Ye, X.-R. Sheng, and D.-C. Zhan, "Few-shot learning with adaptively initialized task optimizer: a practical meta-learning approach," *Machine Learning*, vol. 109, no. 3, pp. 643–664, 2020.

[32] H. Qi, M. Brown, and D. G. Lowe, "Low-shot learning with imprinted weights," 2018, https://arxiv.org/abs/1712.07136.

[33] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[34] Y. H. H. Tsai and R. Salakhutdinov, "Improving one-shot learning through fusing side information," 2017, https://arxiv.org/abs/1710.08347.

[35] K.-H. Thung and C.-Y. Wee, "A brief review on multi-task learning," *Multimedia Tools and Applications*, vol. 77, no. 22, pp. 29705–29725, 2018.

[36] R. Caruana, "Multitask learning," *Machine Learning*, vol. 28, no. 1, pp. 41–75, 1997.

[37] Q. Zhou and Q. Zhao, "Flexible clustered multi-task learning by learning representative tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 266–278, 2016.

[38] L. Sun, Y. Zhou, Y. Wang, C. Zhu, and W. Zhang, "The effective methods for intrusion detection with limited network attack data: multi-task learning and oversampling," *IEEE Access*, vol. 8, pp. 185384–185398, 2020.

[39] S. Zerhoudi, M. Granitzer, and M. Garchery, "Improving intrusion detection systems using zero-shot recognition via graph embeddings," in *Proceedings of the 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 790–797, IEEE, Madrid, Spain, July 2020.

[40] S. H. Hong, S. Ryu, J. Lim, and W. Y. Kim, "Molecular generative model based on an adversarially regularized autoencoder," *Journal of Chemical Information and Modeling*, vol. 60, no. 1, pp. 29–36, 2020.

[41] R. Mitra, R. Gill, S. Sikdar, and S. Datta, "Bayesian hierarchical model for protein identifications," *Journal of Applied Statistics*, vol. 46, no. 1, pp. 30–46, 2019.

[42] M. Lopez-Martin, B. Carro, and A. Sanchez-Esguevillas, "Variational data generative model for intrusion detection," *Knowledge and Information Systems*, vol. 60, no. 1, pp. 569–590, 2019.

[43] Y.-X. Wang and M. Hebert, "Learning to learn: model regression networks for easy small sample learning," in *Proceedings of the Computer Vision–ECCV 2016*, pp. 616–634, Amsterdam, The Netherlands, October 2016.

[44] R. Vilalta and Y. Drissi, "A perspective view and survey of meta-learning," *Artificial Intelligence Review*, vol. 18, no. 2, pp. 77–95, 2002.

[45] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," 2017, https://arxiv.org/abs/1703.03400.

[46] D. P. Mandic, "A generalized normalized gradient descent algorithm," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 115–118, 2004.

[47] H. Fei, Q. Li, and D. Sun, "A survey of recent research on optimization models and algorithms for operations management from the process view," *Scientific Programming*, vol. 2017, Article ID 7219656, 19 pages, 2017.

[48] 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.

[49] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," *Ieee Access*, vol. 5, pp. 21954–21961, 2017.

[50] M. M. U. Chowdhury, F. Hammond, G. Konowicz, C. Xin, H. Wu, and J. Li, "A few-shot deep learning approach for improved intrusion detection," in *Proceedings of the 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, pp. 456–462, IEEE, New York, NY, USA, October 2017.

[51] Z. Li, F. Zhou, F. Chen, and H. Li, "Meta-sgd: learning to learn quickly for few-shot learning.," 2017, https://arxiv.org/abs/1707.09835.

[52] C. Xu, J. Shen, and X. Du, "A method of few-shot network intrusion detection based on meta-learning framework," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3540–3552, 2020.

[53] H. Hindy, C. Tachtatzis, R. Atkinson, D. Brosset, and M. Bures, "Leveraging siamese networks for one-shot intrusion detection model," arXiv preprint arXiv:2006.15343, 2020.

[54] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *Proceedings of the 2009 IEEE symposium on computational intelligence for security and defense applications*, pp. 1–6, IEEE, Ottawa, Canada, July 2009.

[55] S. Huang, Y. Liu, C. Fung et al., "A gated few-shot learning model for anomaly detection," in *Proceedings of the 2020 International Conference on Information Networking (ICOIN)*, pp. 505–509, IEEE, Barcelona, Spain, January 2020.

[56] C. H. Lampert, H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 951–958, IEEE, Miami, FL, USA, June 2009.

[57] A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, and S. Robinson, "Deep learning for unsupervised insider threat detection in structured cybersecurity data streams," in *Proceedings of the Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*, San Francisco, CA, USA, February 2017.

[58] L. Yu, J. Dong, L. Chen et al., "PBCNN: packet bytes-based convolutional neural network for network intrusion detection," *Computer Networks*, vol. 194, Article ID 108117, 2021.

[59] M. He, X. Wang, J. Zhou, Y. Xi, L. Jin, and X. Wang, "Deep-feature-based autoencoder network for few-shot malicious traffic detection," *Security and Communication Networks*, vol. 2021, Article ID 6659022, 13 pages, 2021.

[60] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "AutoAugment: learning augmentation policies from data," 2018, https://arxiv.org/abs/1805.09501.