An Integrated Imbalanced Learning and Deep Neural Network Model for Insider Threat Detection

Mohammed Nasser Al-Mhiqani[1],
Rabiah Ahmed[2]*, Z Zainal Abidin[3]
Center for Advanced Computing Technology
Faculty of Information Communication Technology
Universiti Teknikal Malaysia Melaka
Melaka, Malaysia

Isnin, S.N[4]
Faculty of Technology Management and Technopreneurship
Center for Technopreneurship Development CTeD
Universiti Teknikal Malaysia Melaka
Melaka, Malaysia

***Abstract*—The insider threat is a vital security problem concern in both the private and public sectors. A lot of approaches available for detecting and mitigating insider threats. However, the implementation of an effective system for insider threats detection is still a challenging task. In previous work, the Machine Learning (ML) technique was proposed in the insider threats detection domain since it has a promising solution for a better detection mechanism. Nonetheless, the (ML) techniques could be biased and less accurate when the dataset used is hugely imbalanced. Therefore, in this article, an integrated insider threat detection is named (AD-DNN), which is an integration of adaptive synthetic technique (ADASYN) sampling approach and deep neural network technique (DNN). In the proposed model (AD-DNN), the adaptive synthetic (ADASYN) is used to solve the imbalanced data issue and the deep neural network (DNN) for insider threat detection. The proposed model uses the CERT dataset for the evaluation process. The experimental results show that the proposed integrated model improves the overall detection performance of insider threats. A significant impact on the accuracy performance brings a better solution in the proposed model compared with the current insider threats detection system.**

***Keywords*—Security; insider threat; insider threats detection; machine learning; deep learning; imbalanced data**

I. Introduction

Information systems are facing a security challenge, which comes from outside or inside of an organization. The outside security challenge involves malware and cyber-attack penetrating the network from remote sites. The inside security issue comes from the “trusted” employee within the organization. In which this issue involves both a behavioral and a technical nature [1][2]. Insider threat is commonly known as a problem of utmost importance for information system security management [3].

The malicious insider threat has been defined in the technical report [4] by Cappelli mentioned “a current or former employee, contractor or business partner who has or had authorized access to an organization’s network, system, or data and intentionally exceeded or misused that access in a manner that negatively affected the confidentiality, integrity, or availability of the organization’s information or information systems”. The insider threat activity was conducted by the intentional insiders; such as sabotage of information system, classified information disclosure and theft of intellectual property, or by an unintentional insider, such as losing external devices that contain sensitive information about the organization. Unlike the tasks of the traditional intrusion detection, several insider threat detection challenges come from the nature of the insider where the insider has the authorization to access the computer systems of the organization and has more knowledge about the security levels of the organization [5][6]. Cybersecurity reports show that 63% think insider attacks have become more frequent in the past 12 months. In a recent survey, 53% of the responders believe that detecting insider attacks has become significantly to somewhat harder [7].

The detection of the insider threat is very difficult task; this is because of many challenges. Firstly, as security mechanisms of an organization are not mainly designed for the people who are already inside the organization’s network, this brings a chance for the motivated malicious insider with authorized access to carry out the malicious actions without triggering alerts. Secondly, majority of the attacks initiated by insider are carried out in several phases over a long time. For this reason, effective detection systems for insider threat have to be designed with consideration of long-term monitoring and wide audit data sources range [8][9].

Despite the good performance demonstrated by the current insider threat detection approaches, the traditional machine learning techniques are not able to utilize all the data of user behavior because of the complexity, high-dimensionality, sparsity, and heterogeneity of the data. ML algorithms normally assuming that the used data are balanced in their nature. However, imbalanced data usually produce high accuracy in detecting the majority class, while the accuracy of the minority class is very low. This type of result is not suitable in the situation of insider threats, where the minority class is the important in detection [10][11].

Hence, to deal with the abovementioned challenge, this article proposes an integrated insider threat detection model, called (AD-DNN), which is based on adaptive synthetic sampling approach (ADASYN) and deep neural network (DNN). The proposed of AD-DNN model contains two main parts. Firstly, the ADASYN oversamples the low-frequency samples of insider threats adaptively for increasing these samples, which will lead in helping the machine learning classifiers to learn the low-frequency insider threats attack samples characteristics. Secondly, The DNN is used to classify the samples to normal or malicious insider based on the generated new dataset from the first stag. To evaluate the AD-DNN performance, an experiment is conducted on the CERT 4.2 insider threats dataset [12].

The rest of this paper is organized as follows. The related works is discussed in Section 2. Section 3 presents the methodology. Section 4 discuss the Implementation and Results. Finally, Section 5 concludes the work.

II. Related Work

The importance of machine learning in the domain of insider threats is growing [13]. In several earlier researches, the use of machine learning algorithms has been used to build a classifier that can identify threats from insiders [14][15].

A significant work have been done for the propose of insider threats detection. The Hidden Markov Model (HMM) is used by Wang et al. in [16] to develop an insider threat detection approach. The HMM modeled the normal users' behavior to identify any abnormal behaviors which may differ from the normal behaviors. By utilizing the HMM in modeling the insider threats, the states number of HMM have an high impact on the effectiveness of the method. When the number of states increases the HMM computational cost increases.

ML algorithms have a high powerful ability in improving the insider threats detection performance and self-adaptive capabilities in handling the environment changes of insider threat. Nevertheless, these techniques of ML are still influenced from the effect of imbalanced data in the insider threats domain as well as the lack of in depth knowledge of the insider's behavior patterns [17].

Parveen et al. in [18] utilized the use of one-class support vector machine (OCSVM) technique to model the time series of the daily log, that conceptualizes the insider threat detection issue as a stream mining problem.

Lin et al. [19] proposed a hybrid insider threat detection model using the CERT dataset. The Deep Belief Network (DBN) and OCSVM have been used to build the insider threats detection model. Firstly, the unsupervised DBN is applied to extract the raw data hidden features. And then, the OCSVM is applied for the training of the model utilizing the extracted features.

In recent years, DNN and RNN techniques are widely used in the development of the detection systems of insider threat, Tuor et al. [20] proposed an online unsupervised deep learning approach based on DNN and RNNs to detect anomalous insider activities in real-time from the system logs. Their approach is containing three main parts, firstly the feature extractor, secondly the batcher/dispatcher, and finally the number of Recurrent Neural Networks (RNNs) or DNNs. Long short-term memory (LSTM) techniques have been used to model the user behaviors either alone or in combination with other techniques, Yuan et al. [14] applied the LSTM and Convolutional Neural Network (CNN) based model on user behavior to model the normal users behavior and detect anomalous user behavior. They have dealt with user activities like the natural language modeling. Similar with the previous work, Zhang et al. in [17] employed the LSTM for modeling the log activity of the insider and treat these activities same like the natural language sequences, the proposed solution is worked by extracting the features and detecting the malicious activities when the patterns of the log differ from the training samples. The proposed model evaluation was carried out on a small group of users, only eight users were selected randomly from the CERT experimental dataset. Another work by Sharma et al. [21] also utilized LSTM based Autoencoder using the similar concept to the previous work which models the user behavior using session activities and therefore detect the abnormal data points.

A great efforts have been made by the researchers in the previous literature, however, we believe that there are still way to improve the insider threats detection performance by considering the issue of imbalanced data, and deal with the issue before proceeding the classification task.

III. Methodology

In this part, the basic concepts and methodology components of the proposed AD-DNN model is discussed as shown in Fig. 1.

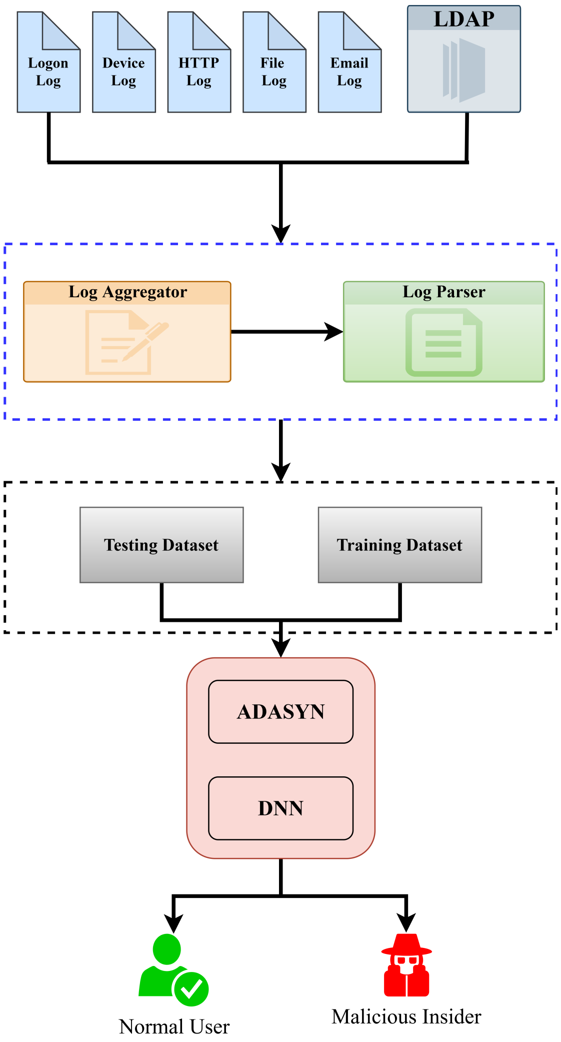

Fig. 1. Proposed Model.

A. Dataset

In this article the CERT r4.2 dataset is used to evaluate the proposed model this is due to the fact that this dataset contains several types of users' event activities, including logon/logoff, device, email, HTTP and files which capture the activities of 1000 employees in an organization for the period of 17 months. Additionally, CERT r4.2 have more instances of malicious insider compared to the other CERT datasets version. The dataset contains 32,770,222 event records generated by the1000 normal and anomalous users. 7323 of the generated activities are malicious insider instances that were manually injected by experts, representing 3 different scenarios of insider threat. The dataset is divided into two sets: the first subsets is used for training and second subsets for testing. 80% of the datasets is used to train the proposed model, the remaining 20% is utilized for the evaluation of the model performance.

B. Log Aggregator and Parser

Firstly, the process of log aggregation starts with the collection of all insider data activity from multiple applications to the main-storage in order to prepare it for the processing task. After the combination of this data done by log parser, it can be saved as a new master dataset. Secondly, to make the data is compatible with machine learning algorithms the log parsing or the parsing engine is created. As the CERT data that has been aggregated in the first stage is mostly in text strings format, which is not readable by the DNN algorithms that we are applying here, the aggregated data need to be transformed to the applicable formats. To transform the data for our model the MaxAbsScaler is used to scales the data between the [-1,1] range automatically based on the absolute maximum.

C. AD-DNN

The idea of sampling methods is either increasing or decreasing the number of samples in the evaluation dataset. The oversampling approach increases the records' frequency, which is a lower sample while under-sampling decreases the records' frequency, which is in a higher sample.

In this article, the oversampling method is used, since the focus on the insider threats, where the minority class is the important in detection, the method used called ADASYN. ADASYN approach is an algorithm that generates synthetic data, the ADASYN main idea is to use a weighted distribution for different examples of minority class according to their difficulty level in learning, the more synthetic data is mainly generated for the examples of minority class which is difficult to learn when it is compared to the other examples of minority classes that are easy to learn [22].

The ADASYN firstly calculates the minority class' K-nearest neighbors of every record in the sample class. Moreover, it draws a line between the neighbors and newly generated random points on that line. Then, it adds some small values randomly on the new point, which makes them similar to the real point. Therefore, these added sample points have more variance than the samples that are taken from their parent samples.

Deep learning (DL): is another machine learning techniques that is based on the learning concept of multi-level representations. The DL creates a hierarchy of features where the lower the level is defining the higher levels and the features of the lower the level helps features are defined at a higher level. The structure of DL is extending the traditional neural networks where more hidden layers are added to the network architecture between the two layers of input and output for modeling the nonlinear and complex relationships. In recent years, this area of research has gained the concern of the researchers due to its great performance for becoming one of the best solutions in many problems. Many DL architectures are existing nowadays, currently, one of the common DL architectures is the convolutional neural networks (CNN), which can carry out complex tasks by using convolution filters. A CNN architecture is a feed-forward layers sequence where the convolutional filters and pooling layers are implemented. CNN adopts many fully-connected layers after the final pooling layer, which work on converting the previous layers 2D feature maps to 1D vector for the classification process. Despite the advantages of the CNN architecture where the feature extraction process is not required before the CNN being applied but the process of CNN training from scratch difficult and time-consuming because it requires large labeled dataset samples to build and train the model before it is prepared for classification. DNN is another type of DL architecture, which is widely utilized and succeeded in both regression and classification in various areas. DNN is a typical feed-forward network where the input flows to the output layer from the input layer using two or more hidden layers. Fig. 2 present the architecture of DNN.

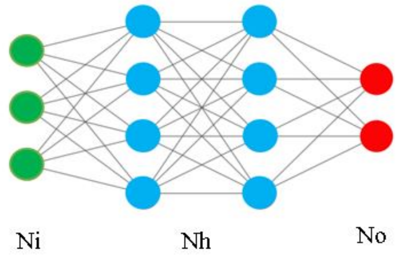

Fig. 2. DNN Architecture.

Fig. 2 shows the typical DNNs architecture where Ni is representing the input layer containing neurons for the input features, the Nh illustrates the hidden layers, and the No is the output layer classes.

IV. Implementation and Results

We have implemented the proposed system on Python with Tensorflow as the backend. The experiment environment is a Ubuntu 18.04.5 LTS operating system runs on a machine with an NVIDIA 1660Ti GPU on a 3.7GHz Intel Core i7-8700HQ, 16GB RAM.

***Model Parameters**:* The proposed model parameters in this article includes the following: (a) The hidden layer of the DNN network, learning rate, number of epochs, and batch size. (b) The ADASYN algorithm oversampling rate and the number of nearest neighbors. We tuned our model with 20 hidden layers, 1e-3 learning rate, 50 epochs, 1024 * 16 batch size, and the Adam optimizer is used.

Metrics: To evaluate the proposed model performance, the parameters used are the average accuracy, average, average false positive rate, average F-Score, average true-negative rate and average false-negative rate. The performance of the proposed model was compared with other classifiers using the same parameter measurements.

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{1}$$

$$F - score = 2.\frac{Precision.Recall}{Precision + Recall} \tag{2}$$

$$FPR = \frac{FP}{FP + TN} \tag{3}$$

$$TNR = \frac{TN}{TN + FP} \tag{4}$$

$$FNR = \frac{FN}{FN + TP} \tag{5}$$

where TP (True Positive), TN (True Negative), FN (False Negative), and FP (False Positive). Additionally, to consider the problem of class imbalance where the insider attacks often carried out by the malicious insiders during the normal work time, which scatters the abnormal insider behavior in large amount of normal employees' behavior, we use the Area Under-Curve (AUC) measurement for evaluating the proposed model. The AD-DN produces a better result compared to the other single classifiers, as shown in Fig. 3 the best result that the AD-DNN gets is AUC = 95%.

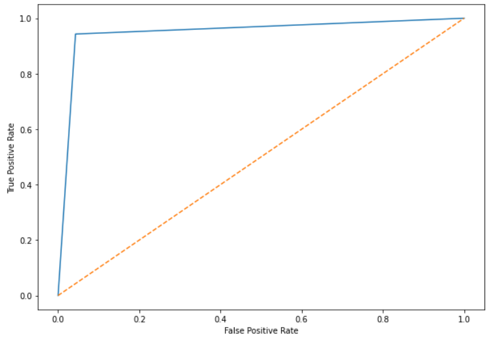

Fig. 3. AD-DNN AUC

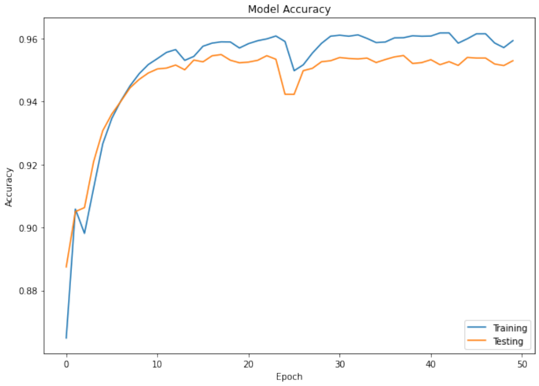

Fig. 4. AD-DNN Accuracy

Fig. 4 presents the average accuracy of the proposed AD-DNN, which shows the accuracy versus the number of epochs. It plots the training and testing performances. As shown in the figure, the proposed AD-DNN obtain good accuracy with average of 96% and there is no major problems indicated with the model since the training and testing curves are very similar to each other and there is no possibility of overfitting.

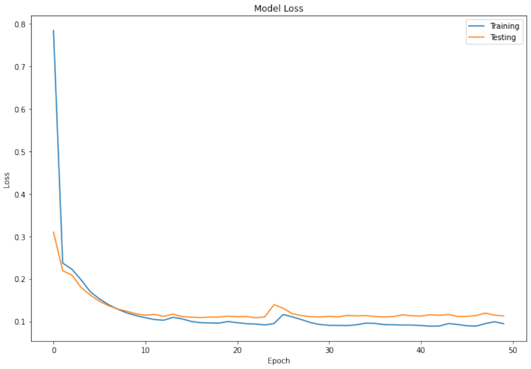

Fig. 5. AD-DNN Loss

Fig. 5 presents the loss of the training and testing in every epoch. In this experiment, the model was stopped after 50 epochs when there is no high testing loss was seen between successive iterations.

Finally, in this article, the designed AD-DNN model is compared with three common methods machine learning techniques (SVM, DNN and LSTM), which have been used in the field of insider threats. The Scikit-Learn library has been implemented to execute three techniques. Additionally, for evaluating the effectiveness of the proposed model using all evaluation matrices, the AD-DNN is compared with some of the recent works as shown in Table I.

TABLE I. COMPARISON SUMMARY OF PROPOSED MODEL

Model	Accuracy	F-Score	AUC	FPR	FNR	TNR
SVM	70%	60%	44%	23.8%	89.10%	76%
LSTM	75%	30%	68%	23.6%	40%	76%
DNN	86%	48%	80%	12.9%	27%	87%
OCSVM based on DBN[19]	87.79%			12.18%		
LSTM Autoencoder[21]	90.17%		95%	9.84%		91%
AD-DNN	96%	95%	95%	4%	5%	96%

On comparative analysis of the well-known classifiers and some of the recent works on detection of the insider threat using the CERT v4.2 dataset, AD-DNN produces a good and promising results. Table I shows that AD-DNN gives the highest accuracy with 96% and the highest F-score, AUC and TNR with 95%, 95% and 96% respectively. Additionally, the AD-DNN achieves the least false rate with 4% FPR and 5% FNR only. It can be seen that AD-DNN is superior to other methods in almost all the evaluation metric, for example the DNN without ADASYN that gives 86% accuracy, 48% F-score, 80% AUC,87% FNR, 12.9% FPR and 27% FNR. This is because AD-DNN consider and solve the imbalance data problem before start training the classifier, and our method can effectively improve the performance of detection.

V. Conclusion

In this article, an integrated insider threat detection model is introduced called as AD-DNN for solving the current challenges in the insider threat detection constructed by employing the theory of machine learning. Firstly, the ADASYN algorithm is used to solve the imbalanced data problem in the situation of insider threats, where the minority class is important in detection. Then, the DNN classifier is designed as the anomaly insider threat detection. The results of the experimental on the CERT dataset shows that the ADASYN algorithm solves the machine-learning algorithms imbalanced the fitting trend of the low-frequency and high-frequency insider data and improves the detection accuracy of the low-frequency insider attack by generating fewer new samples. Furthermore, compared with other recent research works and machine learning techniques used for insider threats detection, the proposed AD-DNN makes the insider threats detection obtains superior and satisfactory results in all the evaluation metrics.

Acknowledgment

Thank you to Research Group of Information Security Forensics and Computer Networking, Center for Advanced Computing Technology (C-ACT), Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka (UTeM). This project is funded by the Ministry of Higher Education Malaysia under the Transdisciplinary Research Grant Scheme (TRGS) with project Number TRGS/1/2016/UTEM/01/3. Its reference is TRGS/1/2016/FTMK-CACT/01/D00006.

References

[1] M. Kandias, A. Mylonas, N. Virvilis, M. Theoharidou, and D. Gritzalis, "An insider threat prediction model," in *International Conference on Trust, Privacy and Security in Digital Business*, S. Bilbao, Ed., vol. 6264 LNCS. Information Security and Critical Infrastructure Protection Research Group, Dept. of Informatics, Athens University of Economics and Business, 76 Patission Ave., GR-10434, Athens, Greece: Springer, 2010, pp. 26–37. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-78049354490{\&}doi=10.1007{\%}2F978-3-642-15152-1{_}3{\&}partnerID=40{\&}md5=9bdad27630b10aae5b5e10b2f4c87ea2

[2] M. Al-Mhiqani, R. Ahmad, K. Abdulkareem, and N. Ali, "Investigation study of Cyber-Physical Systems: Characteristics, application domains, and security challenges," *ARPN Journal of Engineering and Applied Sciences*, vol. 12, no. 22, 2017.

[3] M. Theoharidou, S. Kokolakis, M. Karyda, and E. Kiountouzis, "The insider threat to information systems and the effectiveness of ISO17799," *Computers and Security*, vol. 24, no. 6, pp. 472–484, 2005.

[4] D. M. Cappelli, A. P. Moore, and R. F. Trzeciak, *The CERT guide to insider threats: how to prevent, detect, and respond to information technology crimes (Theft, Sabotage, Fraud)*, 2nd ed. Boston, MA, USA: Addison-Wesley, 2012.

[5] D. C. Le, N. Zincir-Heywood, and M. I. Heywood, "Analyzing Data Granularity Levels for Insider Threat Detection Using Machine Learning," *IEEE Transactions on Network and Service Management*, vol. 17, no. 1, pp. 30–44, 2020.

[6] N. Ameera, N. Mohammad, W. M. Yassin, R. Ahmad, A. Hassan, and M. N. Al-mhiqani, "An Insider Threat Categorization Framework for Automated Manufacturing Execution System," *International Journal of Innovation in Enterprise System*, vol. 3, no. 01, pp. 31–41, 2019.

[7] H. Schulze, "Insider Threat: 2020 Report," Cybersecurity Insiders, Tech. Rep., 2020. [Online]. Available: https://www.cybersecurity-insiders.com/wp-content/uploads/2019/11/2020-Insider-Threat-Report-Gurucul.pdf

[8] L. Liu, C. Chen, J. Zhang, O. De Vel, and Y. Xiang, "Insider Threat Identification Using the Simultaneous Neural Learning of Multi-Source Logs," *IEEE Access*, vol. 7, pp. 183 162–183 176, 2019.

[9] M. N. Al-Mhiqani, R. Ahmad, Z. Z. Abidin, W. Yassin, A. Hassan, K. H. Abdulkareem, N. S. Ali, and Z. Yunos, "A review of insider threat detection: Classification, machine learning techniques, datasets, open challenges, and recommendations," *Applied Sciences (Switzerland)*, vol. 10, no. 15, 2020.

[10] A. Azaria, A. Richardson, S. Kraus, and V. S. Subrahmanian, "Behavioral analysis of insider threat: A survey and bootstrapped prediction in imbalanced data," *IEEE Transactions on Computational Social Systems*, vol. 1, no. 2, pp. 135–155, 2014.

[11] S. Yuan and X. Wu, "Deep Learning for Insider Threat Detection: Review, Challenges and Opportunities," *arXiv*, 2020. [Online]. Available: http://arxiv.org/abs/2005.12433

[12] The CERT Division, "Insider Threat Test Dataset," https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=508099, November 2016, (Accessed on 01/16/2021).

[13] S. Walker-roberts, M. Hammoudeh, A. L. I. Dehghantanha, and S. Member, "A Systematic Review of the Availability and Efficacy of Countermeasures to Internal Threats in Healthcare Critical Infrastructure," *IEEE Access*, vol. 6, pp. 25 167–25 177, 2018.

[14] F. Yuan, Y. Cao, and Y. Shang, "Insider Threat Detection with Deep Neural Network," in *International Conference on Computational Science*, vol. 10860. Springer International Publishing, 2018, pp. 43–54. [Online]. Available: http://link.springer.com/10.1007/978-3-319-93698-7

[15] M. N. Al-Mhiqani, R. Ahmad, Z. Z. Abidin, W. Yassin, A. Hassan, and A. N. Mohammad, "New insider threat detection method based on recurrent neural networks," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 3, pp. 1474–1479, 2019.

[16] C. Wang, G. Zhang, and L. Liu, "A detection method for the resource misuses in information systems," in *Advances in Intelligent and Soft Computing*, vol. 137 AISC. Springer, 2012, pp. 545–552.

[17] D. Zhang, Y. Zheng, Y. Wen, Y. Xu, J. Wang, Y. Yu, and D. Meng, "Role-based Log Analysis Applying Deep Learning for Insider Threat Detection," in *Proceedings of the 1st Workshop on Security-Oriented Designs of Computer Architectures and Processors*, 2018, pp. 18–20. [Online]. Available: http://doi.acm.org/10.1145/3267494.3267495

[18] P. Parveen, Z. R. Weger, B. Thuraisingham, K. Hamlen, and L. Khan, "Supervised learning for insider threat detection using stream mining," in *23rd IEEE International Conference on Tools with Artificial Intelligence Supervised*, 2011, pp. 1032–1039. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84855784162{\&}doi=10.1109{\%}2FICTAI.2011.176{\&}partnerID=40{\&}md5=c556f156272fbd56c9375b1c021e6380

[19] L. Lin, S. Zhong, C. Jia, and K. Chen, "Insider Threat Detection Based on Deep Belief Network Feature Representation," in *2017 International Conference on Green Informatics (ICGI)*, 2017, pp. 54–59.

[20] A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, and S. Robinson, "Deep learning for unsupervised insider threat detection in structured cybersecurity data streams," *AAAI Workshop - Technical Report*, vol. WS-17-01 -, no. 2012, pp. 224–234, 2017.

[21] B. Sharma, P. Pokharel, and B. Joshi, "User Behavior Analytics for Anomaly Detection Using LSTM Autoencoder-Insider Threat Detection," in *ACM International Conference Proceeding Series*, 2020, pp. 1–9.

[22] S. He, H., Bai, Y., Garcia, E., & Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In IEEE International Joint Conference on Neural Networks, 2008," *IJCNN 2008.(IEEE World Congress on Computational Intelligence) (pp. 1322– 1328)*, no. 3, pp. 1322– 1328, 2008.