EPiC Series in Computing

Volume 88, 2022, Pages 18–27

Proceedings of 31st International Conference
on Software Engineering and Data Engineering

EA-NET: A Hybrid and Ensemble Multi-Level Approach For Robust Anomaly Detection

Jayesh Soni[1], Nagarajan Prabakar[2], Himanshu Upadhyay[3]
[1,2] Knight School Of Computing and Information Sciences
[3] Applied Research Center
Florida International University, Miami, USA
jsoni002@fiu.edu, prabakar@cis.fiu.edu, upadhyay@fiu.edu

Abstract

In the current world, the applications of anomaly detection range from fraud detection to diagnosis in the medical area. Most of the current methodologies are applicable only when a particular dataset pertains to certain assumptions and a distinct domain. Such assumptions require prior knowledge of the dataset. The training development cycle time to find the best single model is time-consuming and challenging. Unsupervised anomaly detection methods do not use the target label for training. However, they result in high false positive rates. In this paper, we address the problem of the ensemble anomaly detection approach that generalizes well across multiple domains. We design a multi-level hybrid approach. At the first level, we train several weak classifiers (weak one class classifiers). Next, we utilize deep learning-based AutoEncoder to reduce the dimension of the dataset. These are the two sets of hybrid features. Next, different one-class classifiers have their strength and limitations. Thus, we propose an adaptive weightage approach that gives the weight to each classifier. Next, this input is passed to the second level. At this level, we have a deep neural network that learns the patterns of the dataset and generates an adaptive dynamic threshold to discriminate the input feature as an anomaly or benign. The major benefit of this approach is the low false-positive rate. The training time is reduced due to the reduction of the input feature dimensions at the first level.

1 Introduction

Anomaly Detection refers to the methodology of finding the data observations that deviate from the expected normal patterns or behavior of data. Developing an efficient anomaly detection solution is always a challenging task, even with the recent surge in the development of learning-based algorithms. Most of the prior work conveys that the usage of supervised-based machine learning algorithms can only recognize the anomalies available in the dataset used for training the model. Nonetheless, any

observation that diverges from the expected behavior has been termed an irregularity. Therefore, such irregularities may not be similar to those already available in the dataset [1]. Secondly, different detection-based techniques rely on diverse and distinct rules in the dataset. Often such algorithms are specific to a particular domain application. Thus, detecting anomalies across multiple domains and in a wide variety of scenarios by a single model is challenging [2]. Simply training multiple one-class classifiers [3, 4] iteratively with different hyper-parameter optimization techniques is a time-consuming task. Furthermore, the anomaly detection approach based on traditional methods often requires features that are processed and engineered in a particular manner. This requires a high amount of computational power and memory. Deep learning-based anomaly detection algorithms [5] have computed higher efficiency to address the abovementioned challenges. Nonetheless, their approach requires the data to be in a particular distribution, and also, the developed methodology lacks generalizability across multiple domains. Thus, in this work, we propose a hybrid multi-level ensemble anomaly detection that learns to combine the predictions from multiple one-class classifiers and trains a deep neural network that gives the final probability of the observation as being normal or anomalous.

2 Literature Review

Based on the availability of the data, the anomaly detection approach is divided into three main categories: supervised, semi-supervised and unsupervised. The supervised-based approach trains the model on binary/multi-class data. It is not used widely for anomaly detection due to the lack of class imbalance problem and training data [6]. The unsupervised approach detects abnormalities based solely on the normal class of data. The conventional approach includes support vector machines [7] and data descriptors [8]. Such algorithms assume data to be normal. The major limitations of traditional approaches are: that the outcome is highly sensitive to the complex hyper-parameters. The trained model cannot be extended to the multi-class dataset. The clustering approach is utilized in [9, 10]. The limitations of these approaches are high computational time, and the results are biased towards the static threshold value. Deep learning-based AutoEncoder is trained, which generates the reconstruction error. This error is used to compute the anomaly score [11]. Compared to traditional approaches, anomaly detection algorithms based on deep learning have shown high results in extracting the complex feature representations of the data [12]. Scalability is one of the advantages of such an approach. Recently, a hybrid approach is being implemented where authors in [13] use autoencoder to learn the latent space of high dimensional complex dataset. This learned latent space is given as input to the one-class classifiers for anomaly detection. It combines the feature extraction capability of the neural network with the discriminative capabilities of the one-class classifiers. The limitation of this approach is to rely solely on the AutoEncoder for feature extraction. To overcome this problem, we enhance the approach that not only uses the AE for feature extraction but also several weak one-class classifiers. This results in low false-positive rates.

3 Contribution of the work

We develop a hybrid and multi-level ensemble anomaly detection framework. At the first level, we reduce the feature dimensionality of the dataset. These features are hybrid since we train multiple one-class classifiers and an AutoEncoder model. Such features have high information gain. Different one-class classifiers have different characteristics. Thus, we apply weightage to each of these weak classifiers. Next, we use these features at the second level to train a deep neural network that outputs the anomaly score. Here, we propose an adaptive threshold approach to decide the boundary. The proposed framework has a low false-positive rate and trains the model to reduce computational time.

The rest of the chapter is structured in the following ways. Section 4 explains the proposed ensemble anomaly detection framework. Section 5 demonstrates experimental results on the open source benchmark dataset. Finally, we conclude in section 6.

4 Proposed Framework

In this section, we explain the proposed Ensemble Anomaly Detection Algorithm. The pictorial view is depicted in Figure 1. It comprises two levels. Hybrid Feature Extraction and Anomaly Detector.

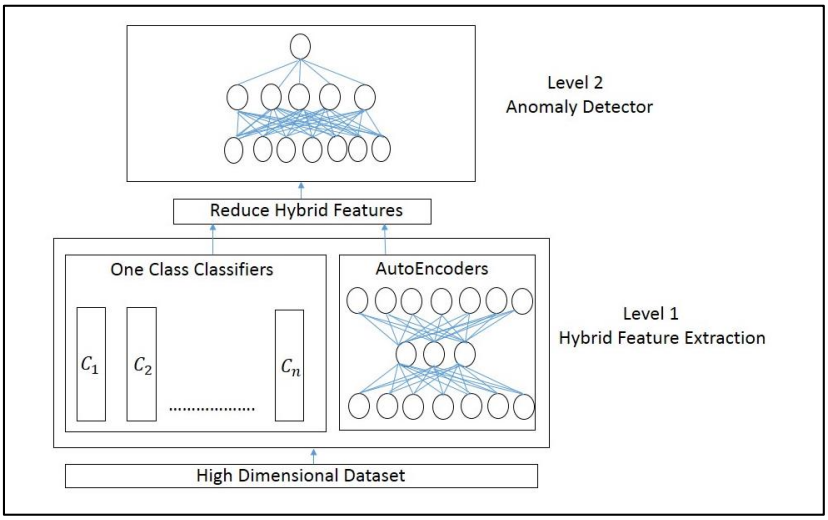

Figure 1: Ensemble Anomaly Framework

4.1 Hybrid Feature Extraction

We extract hybrid features from the high-dimensional dataset in this component. It is the combination of multiple one-class classifiers and variational AutoEncoder. Figure 2 shows the feature extraction mechanism from multiple one-class learner models, namely: One Class Support Vector Machine (OCSVM), Isolation Forest, Mahalanobis Classifier, Local Outlier Factor, and Elliptical Envelope. We fed data of normal class to each learner model ($\mathcal{L}$) to get the anomaly scores. Each one class classifier has its unique characteristics. Thus, we apply an adaptive weightage to each of these algorithms. Next, we apply the K-Fold cross-validation technique, where the value of K is set to 10. We calculate the total number of False Positives produced by the algorithm each time and generate the cumulative error.

$$Avg.\ FP(\mathcal{L}_1) = \frac{\sum_{i=1}^{k} FP(\mathcal{L}_1)_k}{|Val\ Data| * k} \qquad (1)$$

Now, based on the above equation, we calculate the weight of each of the classifiers as follows:

$$Weight_{Classifier} = \ = 1 - \ Avg.\ FP(\mathcal{L}_1) \qquad (2)$$

The output from the multiple one-class classifiers becomes one set of features.

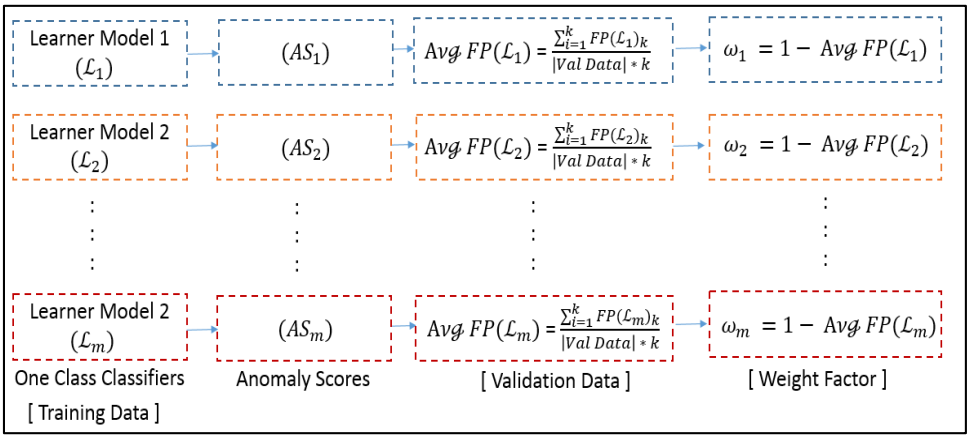

Figure 2: One Class Classifiers

Next, we train deep learning-based variational AutoEncoder to reduce the dimensionality of the dataset to a smaller latent space, as shown in Figure 3. This algorithm takes as input the feature set and will reduce it to a lower dimension.

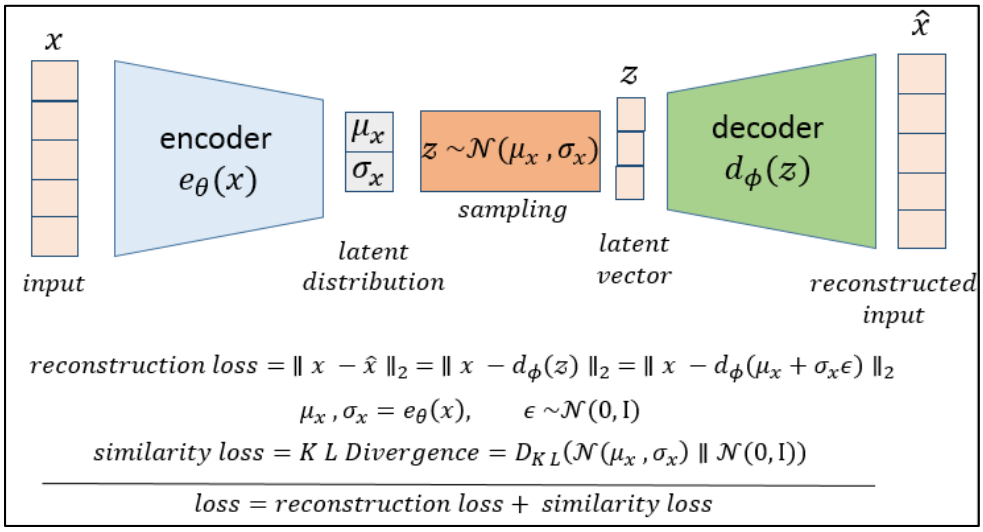

Figure 3: Variational AutoEncoder for Low Dimensional Embedding

Next, it will reconstruct the original feature from the compressed space. The error in reconstruction is the loss. The backpropagation algorithm is applied to update the weight and reduce the loss. We use KL Divergence loss for the backpropagation.
Thus, these hybrid sets of features are then fed to Anomaly Detector. Algorithm 1 depicts the two-step process for anomaly detection.

Algorithm 1 Ensemble Anomaly Algorithm
Input: DataSet
Output: Normal or Anomalous Data Points
1: N = Number of Rows

2: $Classifier_{Output}$ = Train multiple One Class Classifiers and Generate Prediction
3: FP = False Positives on the $Validation_{Data}$
4: $Avg.FP(\mathcal{L}_1) = \frac{\sum_{i=1}^{k} FP(\mathcal{L}_1)_k}{|Val\ Data| * k}$
5: $Weight_{Classifier} = 1 - Avg.FP(\mathcal{L}_1)$
6: $Weighted_{Features}$ = $Classifier_{Output}$ * $Weight_{Classifier}$
7: AE_{Output} = Output from trained *Variational AutoEncoder*
8: $Combined_{Features}$ = $Weighted_{Features}$ U AE_{Output}
9: *DNN* = Trained Neural Net on $Combined_{Features}$
10: **for** *i* in range *0 to N* **do**
11: $Output_{DNN}$ = Prediction using *DNN* for $Data_i$
12: **if** $Output_{DNN} > Adaptive_{Threshold}$ **then**
13: *Data point is anomalous*
14: **else**
15: *Data point is normal*
16: **end if**
17: **end for**

4.2 Anomaly Detector

This is the second level of the proposed framework. It is basically a deep neural network with one hidden layer of 10 units. It takes as input the hybrid features generated from Level 1. Next, it trains the deep neural network and outputs the probability of observation as normal or anomalous. Here, we use K-Fold Cross Validation to generate the value for Dynamic Threshold to decide whether the incoming test data row is normal or anomalous.

5 Experimental Result Analysis

This section discusses the results of applying the proposed algorithm to two intrusion detection datasets: CIC-IDS2017 and UNSW-NB15. Each dataset is unique and has a varying size feature set.

CIC-IDS2017 Dataset
It is one of the intrusion detection datasets released in 2017. There are a total of 2.8 million records with 79 features. This dataset is generated by Canadian Institute for CyberSecurity. It is generated over a period of five days. This dataset contains information on real-world network traffic, which include the normal traces and the malicious traces in the PCAP format.

UNSW-NB15 Dataset
This dataset is developed using the IXIA PerfectStorm tool. It was created in the Australian Center for Cyber Security (ACCS) lab. It has a total of two million records with 44 features. The dataset is a hybrid that captures the real-world scenario of normal activities. On the other hand, it captures the synthetic attack behavior of the network traffic. There are nine different types of attacks recorded in this dataset. The following evaluation metrics are used:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1 - Score = \frac{2*Precision*Recall}{Precision+Recall} \quad (6)$$

Where TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative.
Table 8.1 compares the Proposed Ensemble Anomaly Detection algorithm with the other detection methods for the CIC-IDS2017 dataset.

Technique	**False Positive Rate**
Consolidated J-48 [14]	6.64
LIBSVM [15]	5.13
FURIA [16]	3.16
WiSarD [17]	2.86
DT-Rule [18]	1.14
Proposed Approach	**0.56**

Table 1: Metrics for CIC-IDS2017 Dataset

The authors [14] applied different resampling strategies to train the classification-based machine learning algorithms. Their approach is based on the class distribution of the training dataset. In FURIA [16], authors proposed a novel fuzzy rule-based method for classification purposes. The model learns the fuzzy rules instead of traditional ones, often based on conventional unordered sets. LIBSVM [15] applies quadratic minimization to the traditional SVM algorithm. WiSarD [17] transform the data into patterns of the n-tuple recognizer and further trains the model by passing tuples as input. DT-Rule framework proposed by Ahmed et al. [18] trains an ensemble of JRip, Forest PA, and REP tree. Most of the traditional approaches are based on binary classification. Our proposed ensemble anomaly approach provides the least FPR of 0.56% based on the comparative analysis. Figure 4 and 5 shows the evaluated metrics of our approach on CICIDS2017 compared to other models. Table 2 shows the comparison results of the proposed Ensemble Anomaly Detection algorithm with the other detection methods for the UNSW-NB15 dataset.

Technique	**False Positive Rate**
E-Max [19]	23.79
Two-level Classification [20]	15.64
Stack Ensemble [21]	8.90
GBM [22]	8.60
Proposed Approach	**4.37**

Table 2: Metrics for UNSW-NB15 Dataset

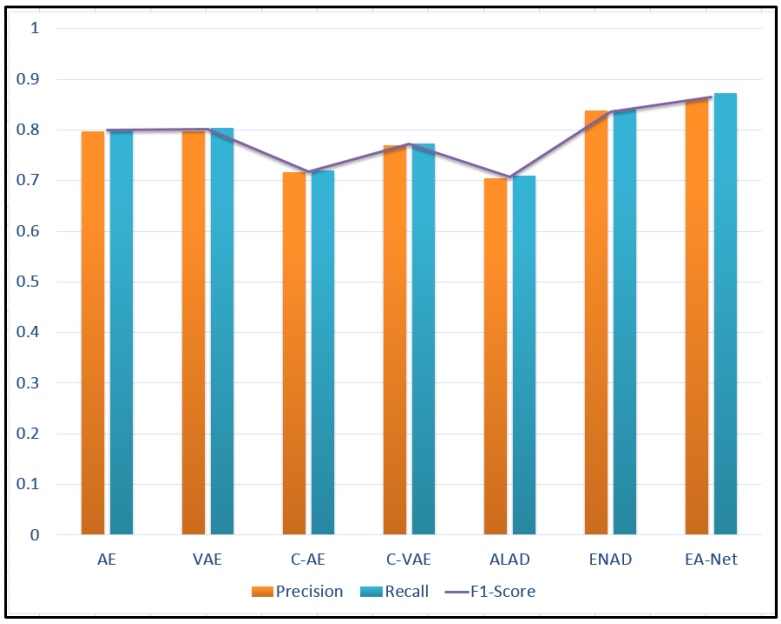

Figure 4: Evaluation Metrics for CICIDS2017 Dataset

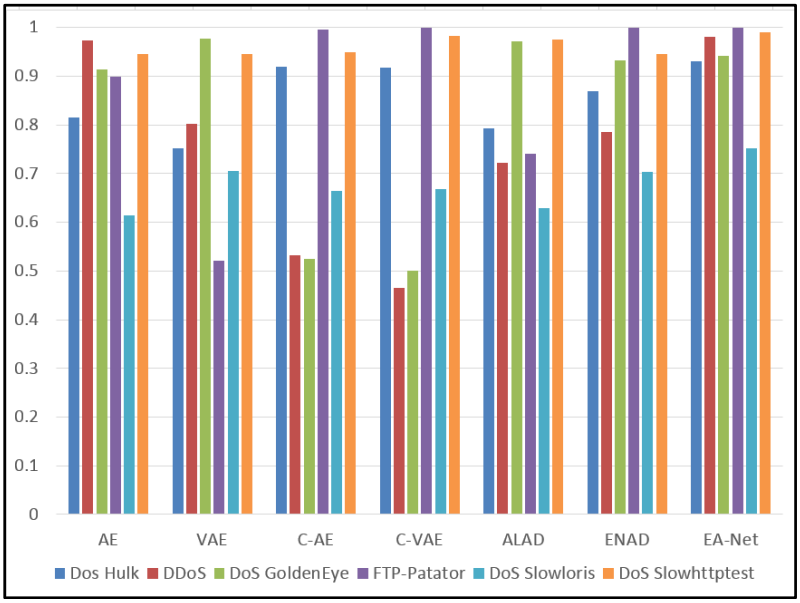

Figure 5: Accuracy for CICIDS2017 Dataset

The performance result of our proposed approach has shown a considerable improvement compared to the existing works. E-Max [19] uses statistical analysis for ranking the attributes and then uses features correlation techniques. They finally trained five different classification algorithms. Two-level classification is employed by Zong et al [20]. They train the model to detect the majority and minority classes of the dataset. The two-level ensemble is proposed in [21], where authors developed a feature selection method and ensemble of two-level classification. Gradient Boosting Classifier is trained by Tama et al. [22] with grid search optimization techniques. The major limitation of this approach is the

training time due to the high complexity of optimizing the hyper-parameters. Our proposed ensemble anomaly approach provides the least FPR of 4.37% based on the comparative analysis. Figure 6 and 7 shows the evaluated metrics of our approach on UNSW-NB15 compared to other models.

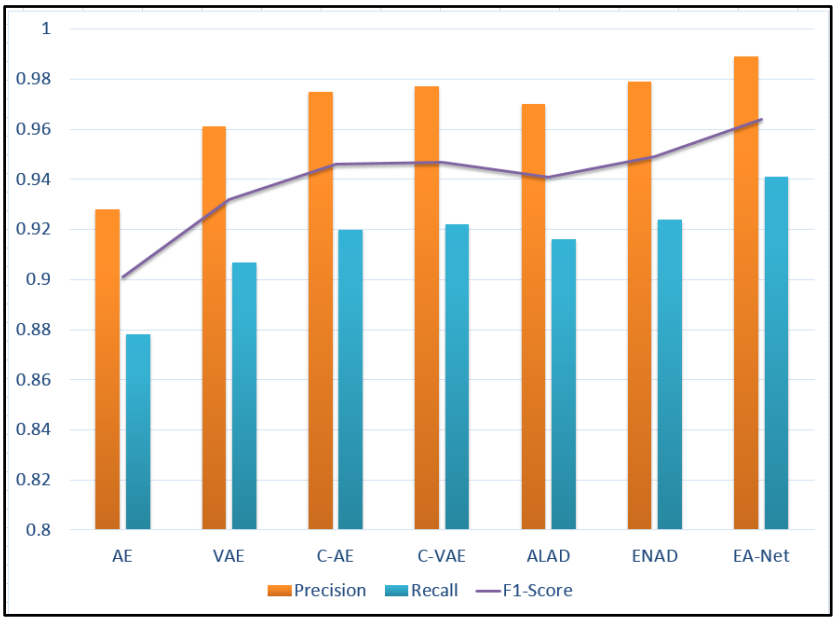

Figure 6: Evaluation Metrics for UNSW-NB15 Dataset

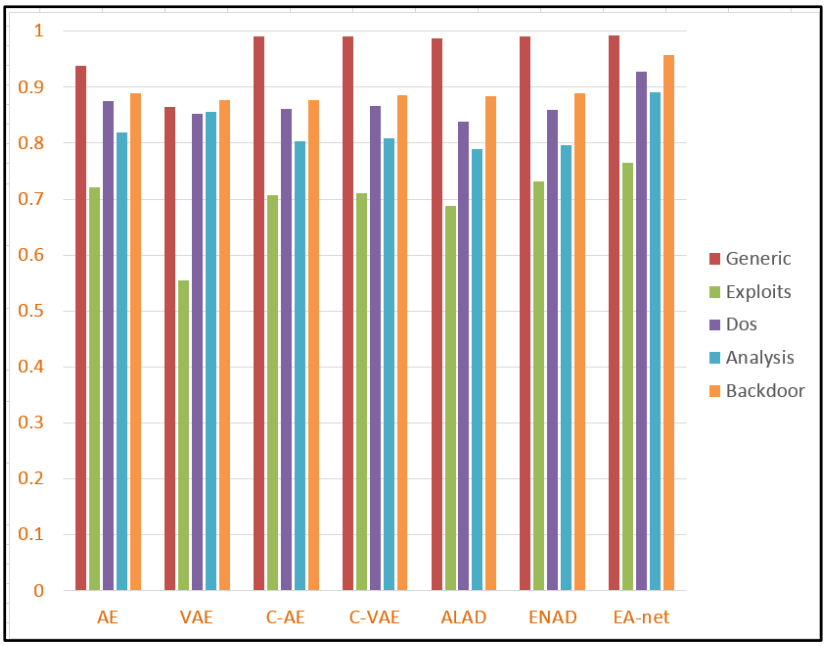

Figure 7: Accuracy for UNSW-NB15 Dataset

6 Conclusion

This study explores anomaly detection for various highly imbalanced classes of the dataset. Binary class and Multiclass are less efficient in detecting the new anomaly since they are trained on the labeled dataset. Currently, various one-class classifiers have been developed, which take the normal class of the dataset as input and learn the normal behavior of the dataset. Anything that deviates from the normal decision boundary is considered an anomaly. Each one class classifier has its characteristics. Thus, training only one algorithm is not efficient for the highly complex real-world dataset with high dimensionality. Therefore, we propose a hybrid two-level anomaly detection framework in this study. We train several one-class classifiers and an AutoEncoder algorithm at the first level. Next, we apply the weight to each one class classifiers algorithm. These reduced feature sets will be passed to the second level. The second level trains a deep neural network that outputs the probability value for the normal and anomalous points. We evaluated our proposed approach on open-source benchmark CIC-IDS2017 and UNSW-NB15 datasets. Our proposed approach results in a low false-positive rate compared to the previous work.

References

[1] Ruff, L., Vandermeulen, R.A., Görnitz, N., Binder, A., Müller, E., Müller, K.R. and Kloft, M., 2019. Deep semi-supervised anomaly detection. arXiv preprint arXiv:1906.02694.

[2] Aggarwal, C.C., 2013. Outlier ensembles: position paper. ACM SIGKDD Explorations Newsletter, 14(2), pp.49-58.

[3] Soni, J., Peddoju, S.K., Prabakar, N. and Upadhyay, H., 2021. Comparative Analysis of LSTM, One-Class SVM, and PCA to Monitor Real-Time Malware Threats Using System Call Sequences and Virtual Machine Introspection. In International Conference on Communication, Computing and Electronics Systems (pp. 113-127). Springer, Singapore.

[4] Soni, J. and Prabakar, N., 2021, December. KeyNet: Enhancing Cybersecurity with Deep Learning-Based LSTM on Keystroke Dynamics for Authentication. In International Conference on Intelligent Human Computer Interaction (pp. 761-771). Springer, Cham.

[5] Pang, Guansong, Chunhua Shen, and Anton van den Hengel. "Deep anomaly detection with deviation networks." In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 353-362. 2019.

[6] Chandola, Varun. "Anomaly detection: A survey varun chandola, arindam banerjee, and vipin kumar." (2007)

[7] Schölkopf, Bernhard, John C. Platt, John Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. "Estimating the support of a high-dimensional distribution." Neural computation 13, no. 7 (2001): 1443-1471.

[8] Tax, D.M. and Duin, R.P., 1999. Support vector domain description. Pattern recognition letters, 20(11-13), pp.1191-1199.

[9] Eskin, E., Arnold, A., Prerau, M., Portnoy, L. and Stolfo, S., 2002. A geometric framework for unsupervised anomaly detection. In Applications of data mining in computer security (pp. 77-101). Springer, Boston, MA.

[10] McInnes, L., Healy, J. and Astels, S., 2017. hdbscan: Hierarchical density based clustering. J. Open Source Softw., 2(11), p.205.

[11] Zhou, Chong, and Randy C. Paffenroth. "Anomaly detection with robust deep autoencoders." In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 665-674. 2017.

[12] Soni, J., Prabakar, N., & Upadhyay, H. (2019, December). Behavioral analysis of system call sequences using LSTM Seq-Seq, cosine similarity and jaccard similarity for real-time anomaly detection. In 2019 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 214-219). IEEE.

[13] Erfani, Sarah M., Sutharshan Rajasegarar, Shanika Karunasekera, and Christopher Leckie. "High-dimensional and large-scale anomaly detection using a linear oneclass SVM with deep learning." Pattern Recognition 58 (2016): 121-134.

[14] Ibarguren, Igor, Jesús M. Pérez, Javier Muguerza, Ibai Gurrutxaga, and Olatz Arbelaitz. "Coverage-based resampling: Building robust consolidated decision trees." Knowledge-Based Systems 79 (2015): 51-67.

[15] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: a library for support vector machines." ACM transactions on intelligent systems and technology (TIST) 2, no. 3 (2011): 1-27.

[16] Hühn, Jens, and Eyke Hüllermeier. "FURIA: an algorithm for unordered fuzzy rule induction." Data Mining and Knowledge Discovery 19, no. 3 (2009): 293-319.

[17] De Gregorio, Massimo, and Maurizio Giordano. "An experimental evaluation of weightless neural networks for multi-class classification." Applied Soft Computing 72 (2018): 338-354.

[18] Ahmim, Ahmed, Leandros Maglaras, Mohamed Amine Ferrag, Makhlouf Derdour, and Helge Janicke. "A novel hierarchical intrusion detection system based on decision tree and rules-based models." In 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), pp. 228-233. IEEE, 2019.

[19] Moustafa, N. and Slay, J., 2016. The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. Information Security Journal: A Global Perspective, *25*(1-3), pp.18-31.

[20] Zong, Wei, Yang-Wai Chow, and Willy Susilo. "A two-stage classifier approach for network intrusion detection." In International Conference on Information Security Practice and Experience, pp. 329-340. Springer, Cham, 2018.

[21] Tama, Bayu Adhi, Marco Comuzzi, and Kyung-Hyune Rhee. "TSE-IDS: A twostage classifier ensemble for intelligent anomaly-based intrusion detection system." IEEE Access 7 (2019): 94497-94507.

[22] Tama, Bayu Adhi, and Kyung-Hyune Rhee. "An in-depth experimental study of anomaly detection using gradient boosted machine." Neural Computing and Applications 31, no. 4 (2019): 955-965.