MLE-NET: A Multi-Layered Ensemble Approach for an Enhanced Anomaly Detection

Jayesh Soni*, Nagarajan Prabakar†, Himanshu Upadhyay‡, and Leonel Lagos
Florida International University, Miami, FL 33174, USA.

Abstract

Anomaly detection is an important task in many areas, including cybersecurity, healthcare, and finance, where it is crucial to identify abnormal behaviors or patterns. However, traditional anomaly detection methods can be sensitive to outliers and lack robustness to distributional changes in the data. In order to overcome these limitations, a hybrid and ensemble multi-layered approach for robust anomaly detection has been proposed in this work. The approach consists of a combination of multiple one class classifiers, each trained on a different subset of the data, and a Variational Autoencoder (VAE). The one class classifiers are used to identify local anomalies, while the VAE is used to model the underlying distribution of the data and detect global anomalies. These are the two sets of hybrid features. Next, different one-class classifiers have their strength and limitations. The final decision on whether an instance is anomalous is made by combining the outputs of the one class classifiers and the VAE through an ensemble learning mechanism. Thus, we propose an adaptive weightage approach that gives the weight to each classifier. Next, these reduced hybrid features are passed as input to the second phase. In this phase, 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 results showed that the hybrid and ensemble multi-layered approach outperforms state-of-the-art anomaly detection methods in terms of robustness and accuracy. Furthermore, the combination of the one class classifiers and the VAE provides a complementary approach that captures both local and global anomalies, making the approach more comprehensive than traditional methods. In conclusion, this work presents a novel hybrid and ensemble multi-layered approach for robust anomaly detection that can effectively address the limitations of traditional methods. The approach has the potential to be applied in a wide range of applications.

Key words: Hybrid multi-layered ensemble, anomaly detection, one class classifiers, variational auto encoders (VAEs), adaptive weightage.

* Applied Research Center.
† Knight Foundation School of Computing and Information Sciences.
‡ Electrical and Computer Engineering.

1 Introduction

Anomaly detection is a crucial task in many domains, including cybersecurity, healthcare, and finance, where the ability to identify abnormal behavior patterns is essential. Traditional anomaly detection methods, such as statistical methods and distance-based methods, can be sensitive to outliers and lack robustness to distributional changes in the data. These limitations can lead to false positive or false negative detections, which can have significant consequences in applications such as fraud detection or network security.

Most previous studies suggest that supervised machine learning algorithms can only identify anomalies present in the training dataset. Nonetheless, deviations from normal behavior are referred to as irregularities. As a result, these irregularities may not resemble those already present in the dataset [15]. Additionally, various anomaly detection techniques rely on different and specific rules in the dataset. These algorithms are often specific to a particular domain and detecting anomalies across multiple domains and scenarios with a single model is challenging [1]. The process of training multiple one-class classifiers [17-18] repeatedly with different hyper-parameter optimization techniques is time-consuming. The traditional anomaly detection approach also requires features that are processed in a specific way, which consumes a significant amount of computational resources. While deep learning-based anomaly detection algorithms [14] have shown improved efficiency, they require the data to be in a specific distribution and the developed methods are not easily transferable across domains. To address these limitations, recent research has focused on developing more robust anomaly detection methods that can effectively handle distributional changes and outliers. One promising direction is the use of deep learning-based methods, such as Variational Autoencoders (VAEs), which have shown great promise in modeling the underlying distributions of complex data. VAEs can be used to detect anomalies by identifying instances that deviate significantly from the modeled distribution.

However, VAEs are known to have limitations when it comes to detecting local anomalies, which are anomalies that are specific to a certain region of the data. To address this issue, multiple one class classifiers can be used to identify local anomalies by training each classifier on a different subset of the data. The outputs of the one class classifiers can then be combined to make the final decision on whether an instance is anomalous.

2 Literature Review

Anomaly detection approaches are classified into three categories based on the availability of data: supervised [8, 11, 24], semi-supervised, and unsupervised. The supervised approach trains the model using binary or multi-class data, but it is not commonly used for anomaly detection due to the class imbalance issue and limited training data [3]. The unsupervised approach detects anomalies solely based on the normal class of data, using methods such as support vector machines [16] and data descriptors [23]. These algorithms have the drawback of being highly sensitive to complex hyper-parameters and not being applicable to multi-class datasets. Clustering techniques [7, 12] have also been used, but these approaches consume a significant amount of computational time and are biased towards a static threshold value. An anomaly detection approach based on deep learning involves training an AutoEncoder and computing the anomaly score based on the reconstruction error [25]. Compared to traditional methods, deep learning-based anomaly detection algorithms have demonstrated better results in capturing complex features of the data [19]. They also offer scalability as an advantage. A recent hybrid approach, as implemented in [6], uses an autoencoder to learn the latent space of high dimensional complex data and provides this learned latent space as input to one-class classifiers for anomaly detection, combining the feature extraction ability of the neural network with the discriminative capabilities of the one-class classifiers. However, this approach relies solely on the autoencoder for feature extraction. To address this issue, we propose an enhanced approach based on EA-Net [20] that not only uses the autoencoder for feature extraction but also integrates several weak one-class classifiers with repeated level of feature detector, resulting in low false-positive rates.

3 Contribution of the work

In this work, we propose a hybrid and ensemble multi-layered approach for robust anomaly detection that combines the strengths of VAEs and one class classifiers. The approach consists of multiple one class classifiers, each trained on a different subset of the data, and a VAE that models the underlying distribution of the data. The final decision on whether an instance is anomalous is made by combining the outputs of the one class classifiers using weightage approach and the VAE through an ensemble learning mechanism. The combination of the one class classifiers and the VAE provides a complementary approach that captures both local and global anomalies, making the approach more comprehensive than traditional methods. We perform multiple experiments where the layer of multiple one class classifier and VAE is repeated to reduce the feature dimension. At the end, we train a deep neural network to provide the final probability of an observation being normal or anomalous.

The rest of this paper is organized as follows: Section 4 presents the proposed hybrid and ensemble multi-layered approach for robust anomaly detection. Section 5 describes the experimental results conducted to evaluate the performance of the proposed method and a comparison with state-of-the-art methods. Finally, Section 6 concludes the paper and discusses potential future work.

4 Proposed Framework

In this section, we describe the proposed MLE Framework, which is illustrated in Figure 1. Figure 1a represents MLE Framework with One Layer Feature Reduction, Figure 1b represents MLE Framework with Two Layer Feature Reduction and Figure 1c represents MLE Framework with Three Layer Feature Reduction. The framework consists of two phases: Hybrid Feature Extraction with different layer and Anomaly Detection.

4.1 Hybrid Feature Extraction

In the Hybrid Feature Extraction component, we derive hybrid features from the high-dimensional data. This is achieved through a combination of multiple one-class classifiers and a variational AutoEncoder. The feature extraction

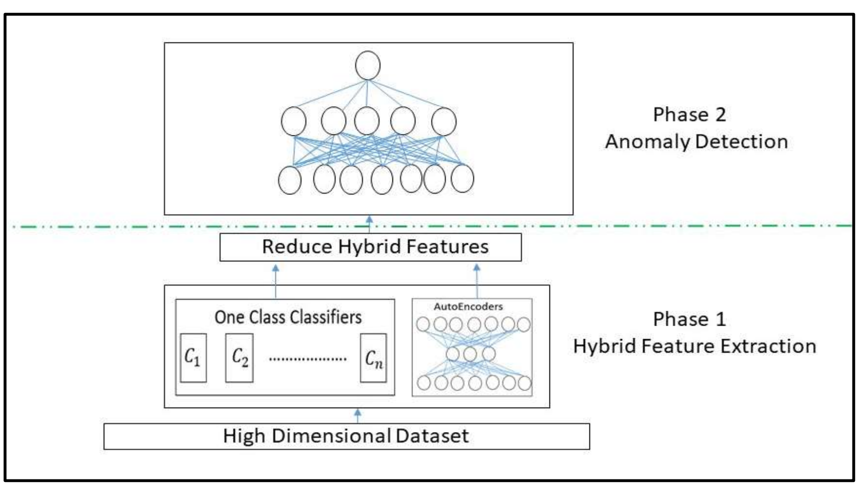

Figure 1a: MLE framework with one layer feature reduction

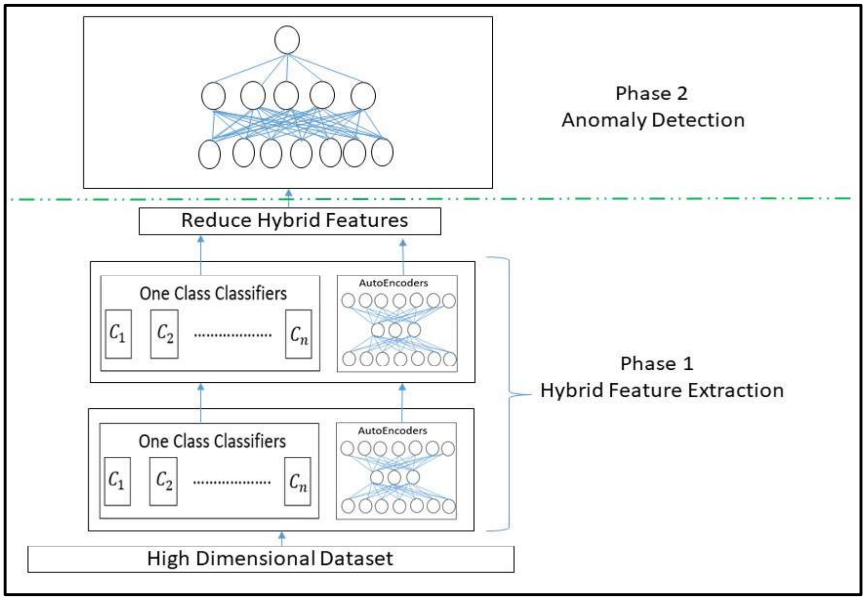

Figure 1b: MLE framework with two-layer feature reduction

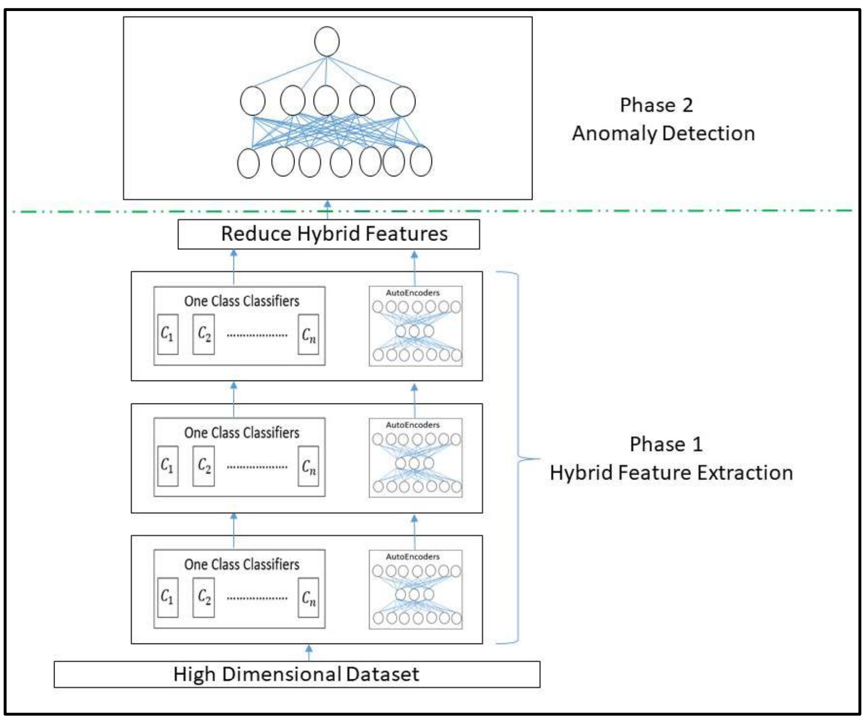

Figure 1c: MLE framework with three-layer feature reduction

mechanism is demonstrated in Figure 2 and involves the following one-class learner models: One Class Support Vector Machine (OCSVM), Isolation Forest, Mahalanobis Classifier, Local Outlier Factor, and Elliptical Envelope. The normal class data is input into each learner model ($\mathcal{L}$) to obtain anomaly scores. Each one-class classifier has distinct characteristics, and thus, we apply an adaptive weighting to each of these algorithms. After that, we implement the K-Fold cross-validation technique with a value of K set to 10. The cumulative error is calculated by determining the total number of False

Positives produced by the algorithm at each iteration.

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

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.

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.

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.

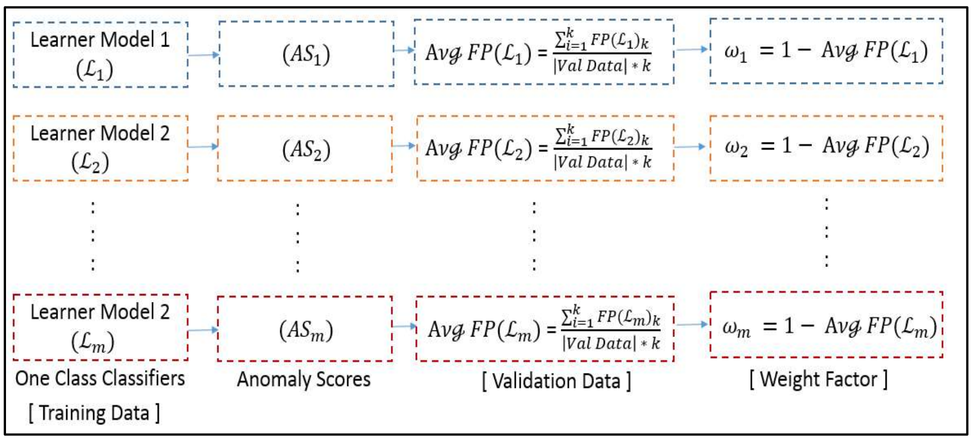

Figure 2: One class classifier

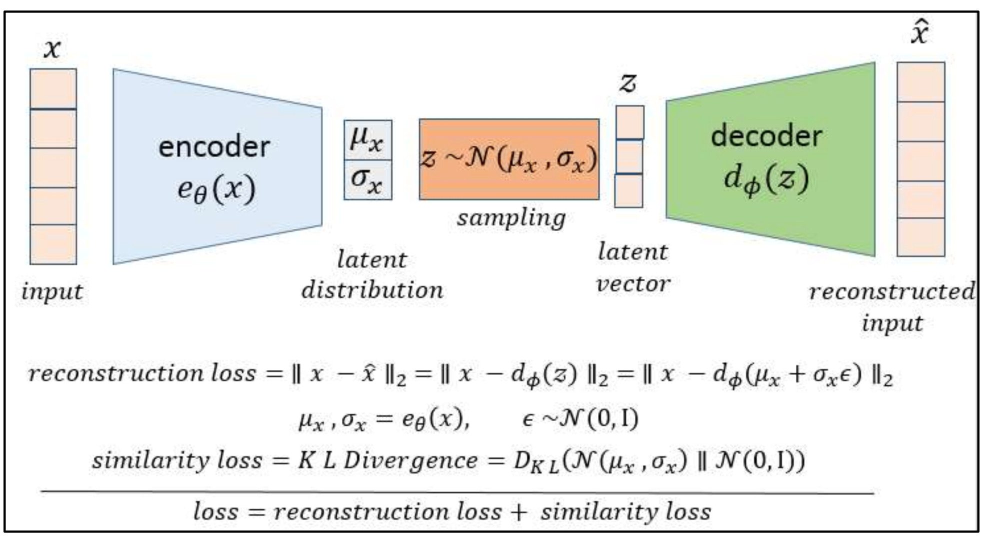

Figure 3: Variational autoencoder for low dimensional embedding

Algorithm 1 Multi-layered Ensemble Anomaly Algorithm
Input: DataSet
Output: Normal or Anomalous Data Points

N = Number of Rows
L = Number of Feature Reduction Layers
for *k* in range *0 to L* **do**
  $Classifier_{Output}$ = Train multiple One Class Classifiers on subset of N and Generate Prediction
  FP = False Positives on the $Validation_{Data}$
  $Avg\ FP(\mathcal{L}_1) = \frac{\sum_{i=1}^{k} FP(\mathcal{L}_1)_k}{|Val\ Data| * k}$
  $Weight_{Classifier} = 1 - Avg\ FP(\mathcal{L}_1)$
  $Weighted_{Features} = Classifier_{Output} * Weight_{Classifier}$
  AE_{Output} = Output from trained *Variational AutoEncoder*
end for
$Combined_{Features} = Weighted_{Features} \cup AE_{Output}$
DNN = Trained Neural Net on $Combined_{Features}$
for *i* in range *0 to N* **do**
  $Output_{DNN}$ = Prediction using *DNN* for $Data_i$
  if $Output_{DNN} > Adaptive_{Threshold}$ **then**
    Data point is anomalous
  else
    Data point is normal
  end if
end for

4.2 Anomaly Detector

The proposed framework has a second level which comprises of a one-hidden-layer deep neural network containing 10 units. The input for this level is the hybrid features generated from the first level. The deep neural network is trained to output the probability of an observation being normal or anomalous. To determine if the incoming test data row is normal or anomalous, K-Fold Cross Validation is used to calculate the value for the dynamic threshold.

5 Experimental Result Analysis

The performance of the proposed algorithm is evaluated on two intrusion detection datasets, CIC-ID2017 and UNSW-NB15. Both datasets have unique characteristics and feature sets of varying sizes.

CIC-ID2017 dataset is a collection of 2.8 million records with 79 features released by the Canadian Institute for CyberSecurity in 2017. The dataset was generated over a period of five days and contains information on real-world network traffic, including normal and malicious traces in PCAP format.

UNSW-NB15 is a dataset created in the Australian Center for Cyber Security (ACCS) lab using the IXIA PerfectStorm tool. It consists of two million records with 44 features and provides a realistic representation of normal network activities and synthetic attack behaviors. The dataset includes nine different types of recorded attacks.

The following evaluation metrics are used:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

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

$$F1-Score = \frac{2*Precision*Recall}{Precision+Recall} \tag{6}$$

Where TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative.

Table 1 compares the Proposed Ensemble Anomaly Detection algorithm with the other detection methods for the CIC-IDS2017 dataset.

Table 1: Metrics for CIC-IDS2017 dataset

Technique	False Positive Rate
Consolidated J-48 [10]	6.64
LIBSVM [4]	5.13
FURIA [9]	3.16
WiSarD [5]	2.86
DT-Rule [2]	1.14
MLE-Net with one layer	**0.56**
MLE-Net with two layers	**0.39**
MLE-Net with three layers	**0.45**

In reference [10], the authors utilized different resampling methods to train classification-based machine learning models that are based on the class distribution of the training data. In FURIA [9], the authors introduced a unique fuzzy rule-based classification method that learns from fuzzy rules rather than traditional ones based on unordered sets. LIBSVM [4] enhances the traditional SVM algorithm using quadratic minimization. WiSarD [5] transforms data into n-tuple patterns for training the model using tuples as inputs. The DT-Rule framework by Ahmed et al. [2] trains an ensemble of JRip, Forest PA, and REP tree models. Traditional methods mostly focus on binary classification; however, our proposed ensemble anomaly approach outperforms others with a minimum FPR of 0.56% with one layer of Feature Detector, 0.39% with one layer of Feature Detector and 0.45% with one layer of Feature Detector. Figures 4 and 5 show 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.

Table 2: Metrics for UNSW-NB15 dataset

Technique	False Positive Rate
E-Max [13]	23.79
Two-level Classification [6]	15.64
Stack Ensemble [21]	8.90
GBM [22]	8.60
Proposed Approach with one layer	**4.37**
Proposed Approach with two layers	**2.93**
Proposed Approach with three layers	**3.57**

Our proposed approach has demonstrated a substantial improvement in performance compared to previous works. For example, E-Max [13] uses statistical analysis to rank attributes and employs correlation techniques to determine features, which are then used to train five different classification algorithms. Zong et al [26] use a two-level classification approach, training the model to detect majority and minority classes in the dataset. The authors of [21] propose a two-level ensemble with a feature selection method and two-level classification. Tama et al [22] use the Gradient Boosting Classifier, trained with grid search optimization techniques, but the major drawback of this approach is the lengthy training time due to the complexity of hyper-parameter optimization. Our proposed ensemble anomaly approach was found to have the lowest false positive rate (FPR) of 4.37% with one layer of Feature Detector, 2.93% with one layer of Feature Detector and 3.57% with one layer of Feature Detector. Figures 6 and 7 show the evaluated metrics of our approach on UNSW-NB15 compared to other models.

6 Conclusion

In this study, we investigate anomaly detection for datasets with highly imbalanced classes. Traditional binary and multiclass classifiers are less effective at detecting anomalies as they are only trained on labeled data. To address this, various one-class classifiers have been developed, which learn the normal behavior on the subset of the dataset by using the normal class as input. Any deviation from the normal decision boundary is considered an anomaly. However, relying on only one classifier is not sufficient for highly complex, high-dimensional real-world datasets. To tackle this, we propose a hybrid two-phase anomaly detection framework. We first train multiple one-class classifiers and an AutoEncoder algorithm at

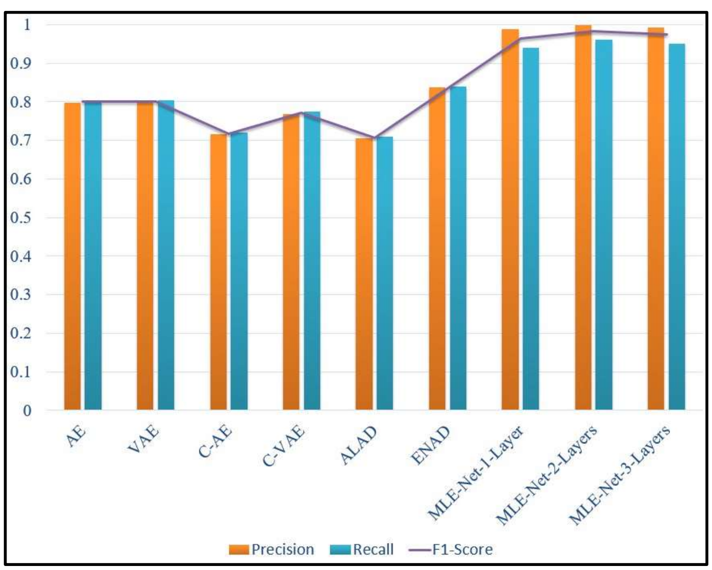

Figure 4: Evaluation metrics for CICIDS2017 dataset

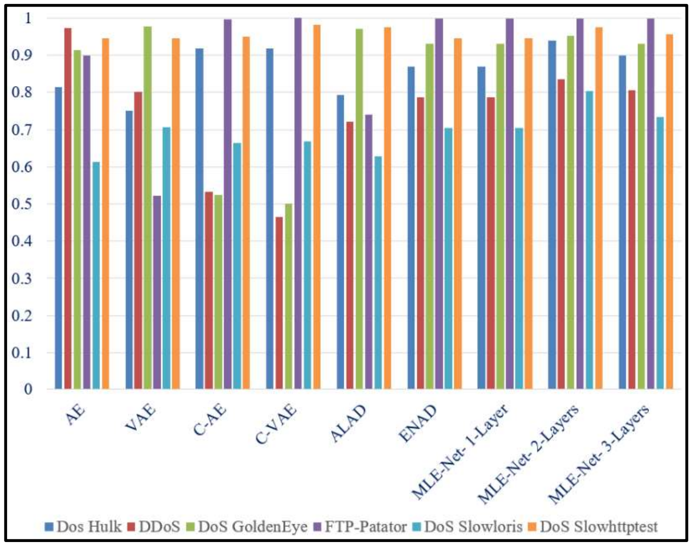

Figure 5: Accuracy for CICIDS2017 dataset

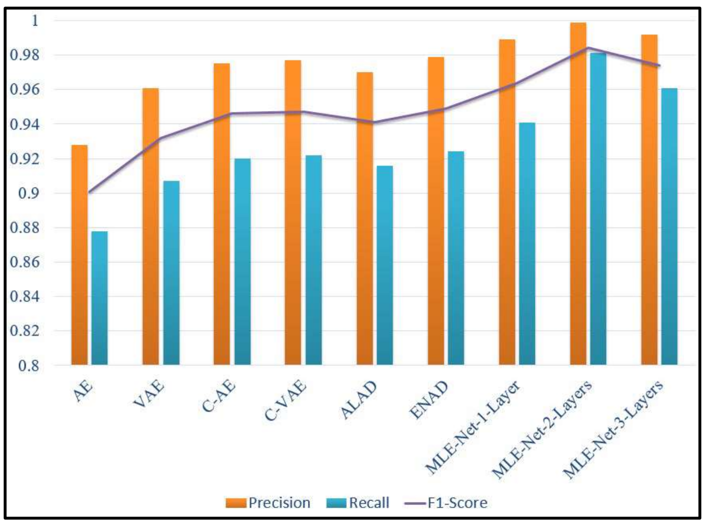

Figure 6: Evaluation metrics for UNSW-NB15 dataset

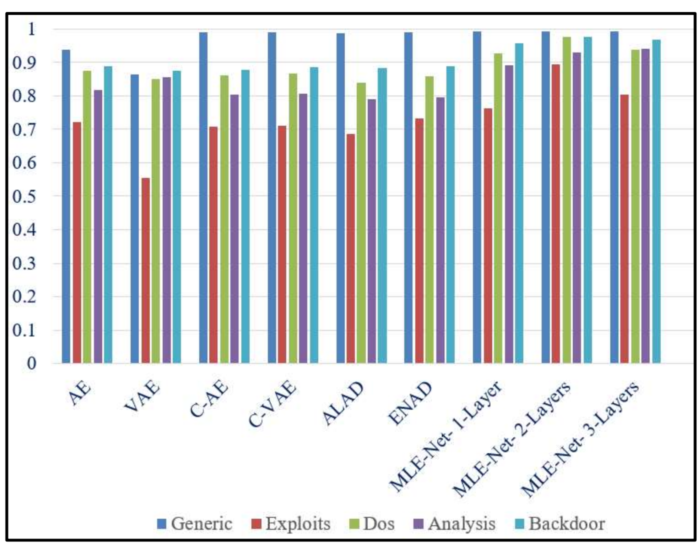

Figure 7: Accuracy for UNSW-NB15 dataset

the first phase, then apply weights to the results from each classifier. We introduced layers in the first phase and experimented three layered versions. The reduced feature sets are then passed to the second phase, which trains a deep neural network to output the probability of normal and anomalous points. Our approach was evaluated on the open-source benchmark datasets CIC-ID2017 and UNSW-NB15 and was found to have a low false-positive rate with two layers in the first phase.

References

[1] C. C. Aggarwal, "Outlier Ensembles: Position Paper," *ACM SIGKDD Explorations Newsletter*, 14(2):49-58, 2013.

[2] Ahmed Ahmim, , Leandros Maglaras, Mohamed Amine Ferrag, Makhlouf Derdour, and Helge Janicke, "A Novel Hierarchical Intrusion Detection System based on Decision Tree and Rules-Based Models," 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), IEEE, pp. 228-233, 2019.

[3] Varun Chandola, "Anomaly Detection: A Survey Varun Chandola, Arindam Banerjee, and Vipin Kumar," 2007.

[4] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology (TIST)* 2(3):1-27, 2011.

[5] Massimo De Gregorio and Maurizio Giordano, "An Experimental Evaluation of Weightless Neural Networks for Multi-Class Classification," *Applied Soft Computing* 72:338-354, 2018.

[6] Sarah M. Erfani, Sutharshan Rajasegarar, Shanika Karunasekera, and Christopher Leckie, "High-Dimensional and Large-Scale Anomaly Detection using a Linear Oneclass SVM with Deep Learning," *Pattern Recognition* 58:121-134, 2016.

[7] E. Eskin, A. Arnold, M. Prerau, L. Portnoy, and S. Stolfo, "A Geometric Framework for Unsupervised Anomaly Detection," Applications of Data Mining in Computer Security, Springer, Boston, MA, pp. 77-101, 2002.

[8] P. Gangwani, J. Soni, H. Upadhyay, and S. Joshi, "A Deep Learning Approach for Modeling of Geothermal Energy Prediction," *Int. J. Comput. Sci. Inf. Secur.*, 18(1):62-65, 2020.

[9] Jens Hühn and Eyke Hüllermeier, "FURIA: An Algorithm for Unordered Fuzzy rule Induction," Data Mining and Knowledge Discovery, 19(3):293-319, 2009.

[10] Igor Ibarguren, Jesús M. Pérez, Javier Muguerza, Ibai Gurrutxaga, and Olatz Arbelaitz, "Coverage-Based Resampling: Building Robust Consolidated Decision Trees," *Knowledge-Based Systems*, 79:51-67, 2015.

[11] S. Joshi, H. Upadhyay, L. Lagos, N. S. Akkipeddi, and V. Guerra, "Machine Learning Approach for Malware Detection Using Random Forest Classifier on Process List Data Structure," Proceedings of the 2nd International Conference on Information System and Data Mining - ICISDM '18, pp. 98-102, 2018, https://dl.acm.org/doi/10.1145/3206098.3206113.

[12] L. McInnes, J. Healy, and S. Astels, "HDBscan: Hierarchical Density Based Clustering, *J. Open-Source Software*., 2(11):205, 2017.

[13] N. Moustafa and J. Slay, "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):18-31, 2016.

[14] Guansong Pang, Chunhua Shen, and Anton van den Hengel, "Deep Anomaly Detection with Deviation Networks." Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 353-362, 2019.

[15] L. Ruff, R. A. Vandermeulen, N. Görnitz, A. Binder, E. Müller, K. R. Müller, and M. Kloft, "Deep Semi-Supervised Anomaly Detection," arXiv preprint arXiv:1906.02694, 2019.

[16] Bernhard Schölkopf, John C. Platt, John Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. "Estimating the Support of a High-Dimensional Distribution," *Neural Computation* 13(7):1443-1471, 2001.

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

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

[19] J. Soni, N. Prabakar, and H. Upadhyay, "Behavioral Analysis of System Call Sequences using LSTM Seq-Seq, Cosine Similarity and Jaccard Similarity for Real-Time Anomaly Detection," 2019 International Conference on Computational Science and Computational Intelligence (CSCI), IEEE, pp. 214-219, December 2019.

[20] J. Soni, N. Prabakar, and H. Upadhyay, "EA-NET: A Hybrid and Ensemble Multi-Level Approach for Robust Anomaly Detection," Proceedings of 31st International Conference, 88:18-27, November 2022.

[21] Bayu Adhi Tama, 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] Bayu Adhi Tama and Kyung-Hyune Rhee, "An In-Depth experimental Study of Anomaly Detection using Gradient Boosted Machine," *Neural Computing and Applications* 31(4):955-965, 2019.

[23] D. M. Tax and R. P. Duin, "Support Vector Domain Description," *Pattern Recognition Letters*, 20(11-13):1191-1199, 1999

[24] H. Upadhyay, L. Lagos, S. Joshi, and A. Abrahao, "Big Data Framework with Machine Learning for D and D Applications – 19108,". United States, 2019.
[25] Chong Zhou and Randy C. Paffenroth, "Anomaly Detection with Robust Deep Autoencoders," Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 665-674, 2017.
[26] Wei Zong, Yang-Wai Chow, and Willy Susilo, "A Two-Stage Classifier Approach for Network Intrusion Detection," International Conference on Information Security Practice and Experience, Springer, Cham, pp. 329-340, 2018.

Jayesh Soni is a Postdoctoral Associate at the Applied Research Center (ARC) of Florida International University. He received his Ph.D. degree in Computer Science from Florida International University in 2022. His M.Tech. and B.E. degrees are in Computer Science and Computer Engineering respectively. His research is in the area of Applied AI modeling and development in interdisciplinary domain. His passion for this field stems from his belief that AI has the potential to revolutionize various industries and improve the quality of life for people all over the world. His PhD thesis focused on developing novel deep learning methods for anomaly detection in sequential data. This work resulted in several publications in leading conferences and journals. In addition to developing AI models for cybersecurity, he also has a keen interest in exploring other interdisciplinary domains where AI can be applied to solve complex problems. This includes areas such as natural language processing, image and video analysis, and data privacy. By working at the intersection of AI and other fields, he aims to make meaningful contributions to both the AI and interdisciplinary communities.

Nagarajan Prabakar received the M.Eng. degree in automation from the Indian Institute of Science, Bangalore, and the PhD degree in Computer Science from the University of Queensland, Brisbane, Australia. He is currently an associate professor in the School of Computing and Information Sciences at Florida International University, Miami, USA. His research interests include machine learning-based object detection, anomaly detection for system security, and distributed optimization for real-world problems.

Himanshu Upadhyay is an Associate Professor in the Electrical and Computer Engineering at Florida International University, Miami, USA. His current research focuses on artificial intelligence/machine learning, cyber forensics, malware analysis, cyber analytics/visualization, and big data. He architected a range of tiered and distributed application system using Microsoft.Net technology to address strategic business needs, managing a team of researchers and scientists building secured enterprise information systems and mentoring undergraduate and graduate students. He has 30 years of experience in cybersecurity, artificial intelligence/machine learning, information technology, data science, management and engineering role, serving as co-principal Investigator for multimillion - dollar cybersecurity research project for the Department of Defense and Department of Energy's Office of Environmental Management.

Leonel Lagos is currently the Associate Professor - Moss Department of Construction Management, Director of Research - Applied Research Center at Florida International University (FIU), USA. He is also the Principal Investigator of the Department of Energy-Florida International University Cooperative Agreement. Further, he is serving as the Program Director of Department of Energy-FIU Science and Technology Workforce Development Program. Dr. Lagos received a B.S. in Aerospace Engineering from the University of Florida in 1991, a Masters in Mechanical Engineering, and Ph.D. in Environmental Engineering from Florida International University (FIU) in 1996 and 2007, respectively. Dr. Lagos is an active member of DOE's Energy Facilities Contractors Group (an advisory group supporting DOE's environmental restoration mission) and serves as a group member in EFCOG's D&D and Facility Engineering Working Group. Dr. Lagos also serves as an active Program Advisory Committee (PAC) member for the Waste Management Symposia organization. Dr. Lagos also serves in the Executive Committee for the American Nuclear Society's Robotics and Remote Systems Division.