Journal of Intelligent System and Applied Data Science (JISADS)

Journal homepage : https://www.jisads.com

ISSN (2974-9840) Online

ENSEMBLE-BASED MODEL FOR MITIGATING FEATURE DISCREPANCIES FOR ENHANCED THREAT DETECTION USING DOMAIN ADAPTATION

Joshua J. Tom[1], Pius U. Ejodamen[2], Taiwo Fele[3]*

[1]*Department of Mathematics and Computer Science, Elizade University, Ilara Mokin, Nigeria,*
[2]*Department of Computing Sciences, Admiralty University of Nigeria, Ibusa, Nigeria,*
[3]*Computer Science Department, School of Science and Computer Studies, The Federal Polytechnic, Ado-Ekiti, Nigeria,*

drtomjoshua@gmail.com, piusejodamen@adun.edu.ng, fele_ta@fedpolyado.edu.ng

ABSTRACT

In today's highly interconnected digital world, there are varieties of threat actors and threat types which necessitate a deep and robust threat detection system. Algorithms for detecting threats rely on various features of security data to identify potential threats. However, some threats are feature-dependent making it nontrivial to detect all types of threats using the same set of features in the dataset. Discrepancy in security telemetry datasets can be a potential cause of threat misclassification and consequently low threat detection system performance. In this paper, we propose an ensemble technique (Ensemble-DAFE) that integrates two techniques for mitigating feature discrepancy in security data viz domain adaptation (DA) and feature engineering (FE) techniques leveraging the strengths of the two to improve threat detection accuracy. We conducted experiments to determine the impact of feature discrepancies on threat detection performance. We obtained a threat detection performance accuracy of 99.96%. when the combined DA and FE was implemented compared to performance accuracy 96.38% without DA. Our result for Ensemble-DAFE with DA combined with FE outperforms state-of-the-art methods without DA compared to ours in terms of detection accuracy. We evaluate the effectiveness of our Ensemble-DAFE threat detection model using a synthetic dataset of network traffic with real-world security features. Based on the result, we noticed a 3.58% improvement in detection performance due to the integration of DA in the threat detection process and demonstrate its ability to reduce false negatives and false positives compared to individual feature-based detection methods.

Keywords: Feature Discrepancy, Ensemble Model, Domain Adaptation, Threat Detection,

1. INTRODUCTION

The cybersecurity arena has witnessed organizations generating huge amount of security telemetry data including data collected from firewalls, routers and switches, endpoint logs (system logs, application logs, antivirus logs), data from user activity logs, feeds from threat intelligence, cloud infrastructure and services logs and data gathered from security information and events management (SIEM), intrusion detection system (IDS) and intrusion prevention system IPS). These data are collected, analyzed, and monitored for security operations, swift response to cybersecurity incidence, and good organizations' posture. Security data is fundamental in shaping an organization's security

strategy in detecting and mitigating threat. Several threat detection algorithms have utilized different features extracted from these data to detect existential threats. However, there may be some discrepancies in the set of extracted data in the process of threat detection as different features do not have the same effectiveness given all threat types. This can have far-reaching negative consequences on a threat detection model such as reduction in model accuracy due to threat misclassification (false positives), fragmented view of security posture, integration problems arising from the need for unified analysis of data from multiple sources, poor and inefficient incident response. These have been drawbacks of many existing threat detection models. Due to the huge amount of security data generated from different sources, the constantly changing and complex nature of threats, and the dynamic threat landscape, there is the need to be able to adapt to new patterns and the continuously learn from the data for improve real-time detection performance [1]. Machine learning algorithm (MLA) is suitable in this scenario as it can handle vast amount of data in real-time detection prevent breaches and prevent damage. Because of this, most detection systems are based on machine learning, AI, deep learning, etc. for threat detection [2]. The use of ML in threat detection can help improve threat detection compared to traditional approaches such as signature-based detection [3]. These ML algorithms can be integrated with specialized techniques that mitigate the performance degrading effect of discrepancies in dataset [4]. There exist many techniques for handling data feature discrepancy including domain adaptation, feature engineering, feature transformation, multi-view learning, etc. [5].

As our contribution in mitigating this problem, we propose an ensemble-based model comprising extra tree classifier, gradient boosting classifier, and random forest models that mitigates discrepancy in security data features by combining multiple feature-based detection algorithms to achieve efficient threat detection through improved accuracy. We integrate domain adaptation with innovative feature engineering techniques to significantly improve our threat detection model's performance in identifying evolving threats. To the best of our knowledge, our work is the first attempt to combine two mitigating approaches against data feature discrepancy. The first approach adopts the domain adaption framework proposed in [6]. The second approach employs feature engineering technique to determine signatures of traffic data and enable our model to distinguish between normal traffic and threats. In this paper, we have come up with a framework for developing a hybrid model capable of addressing data feature discrepancies in threat detection.

2. THEORETICAL FRAMEWORK

The fast-evolving cyber threats landscape has made it mandatory for organizations to pay specific attention to the development of efficient threat detection systems to identify and mitigate risks in real-time. This is because of increased sophistication in cyber threats dynamics and advancement in development tools [7]. The focus on ML techniques in threat detection has been intensified by the inadequacy of traditional rule-based detection systems. The adoption of Machine learning models in different fields have shown great results in terms of threat detection automation by learning patterns from vast amount security datasets. However, a significant problem to grapple with is data feature discrepancy during training [8]. Some key aspects of feature discrepancy in security telemetry data collected from various security events or logs include inconsistency in naming, differences in data types, missing features, varying feature availability, data granularity, temporal discrepancies, feature redundancy, etc. The presence of these feature discrepancies in datasets arises due to differences in the way in which features are distributed between source and target domains. As such, machine learning models must learn to adapt to a target domain different from a source domain in which it was trained [5].

Domain Adaptation (DA) is a desirable and an important component of a machine learning model whose major purpose is to improve the performance of models trained in one environment (source domain) and tested in a new environment (target domain) [9]. With regards to threat detection, domain adaptation is vital because of the rapidly evolving cyber threats, which changes rapidly and are characterized by varied behaviors. Different organizations require different approach to detecting threats in their digital life but the narrative can change when threat detection models are design to be able to adapt to changing environments such that a model trained on benign network traffic exhibits accurate threat detection capability and not struggle in a strange environment with different features like attack vectors, user behavior patterns, and network architecture. Therefore, to handle any disparity in security data features, domain adaptation techniques are used to bridge the gap between the source and target domains making threat detection models to perform optimally in the face of new and unfamiliar threats.

Domain adaptation mainly consists of a range of methodologies including instance re-weighting, feature transformation-based Maximum Mean Discrepancy, adversarial training-based Generative Adversarial Networks (GAN), and domain-invariant feature learning [10]. These techniques are often engaged to reduce the divergence between the source and target domain distributions. Furthermore, Generalized Adaptive Models (GMAs) have been recently been used in designing various machine learning based models to bring in flexibility to cater for variations in data features [11]. This approach is principally to adapt the feature engineering architecture and learning mechanisms to specific feature distribution characteristics. Figure 1 shows a general conceptual framework of discriminative cross domain adaption.

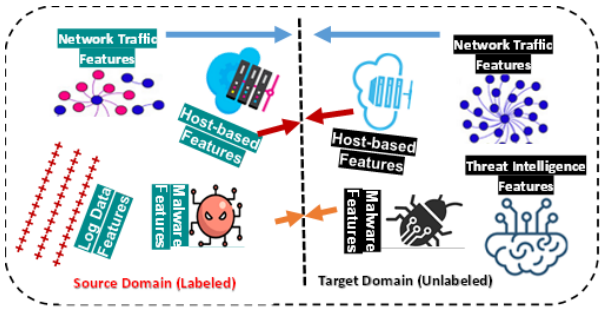

Figure 1: Illustrating Cross Domain Adaptation to Learn Discriminative Information by Mitigating Domain Discrepancy.

Feature engineering forms the foundation of machine learning models for effective threat detection [12]. In this approach, relevant features are selected or modified from the raw data that would enhance the model's understanding of the patterns that may indicate potential threats. Some important aspects of feature engineering in detecting threats are: eliciting specific domain for hunting threats e.g. network traffic or transaction data; the type of features e.g. raw features such as user IDs, timestamps, IP addresses, etc., behavioral features and contextual features; temporal features; statistical features; anomaly detection features, amongst others. For threat detection models to perform optimally, security data features must be carefully and correctly selected to ensure better performance. With the dynamic nature of threats, feature refinement and adaptation are key to combat emerging threats and contend the unpredictability in attacker behaviors [13]. Many research works have been conducted deploying other techniques for handling discrepancies in data features in dataset meant for machine learning analytics including feature alignment, data augmentation, bias correction, ensemble and multimodal approaches, etc.

Furthermore, the potentials of machine learning models in threat detection and domain adaptation cannot be overemphasized. Machine learning algorithms (neural network architectures or ensemble-based learning models) intrinsically providing mechanisms to take care of possible variances in domains can be a better option in handling feature discrepancies [14]. Studies have shown that advancements techniques like transfer learning and adversarial training are good candidates for improving the robustness of machine learning models against domain shifts [15]. Individual machine learning algorithm perform best given different aspects and scenarios. To leverage the diverse strengths of these model we need to bring them into a single framework, ensemble model. In the context of threat detection, the diversified characteristic of ensemble models can capture varying patterns and nuances in datasets leading to higher performance in threat detection [16]. Mostly use ensemble framework include those that use Decision Trees, Support Vector Machines, Neural Networks, Logistic Regression, K-Nearest Neighbor, Gradient Boosting Machines, etc. as the base learners for classification and regression tasks. From the foregoing, the deployment of domain adaptation, feature engineering, and machine learning models is expected to bring some level of innovation and novelty in the context of cybersecurity and can offer the much-envisaged real-time response and adaptability in tackling the rampaging and evolving threats. As organizations continuously grapple with different and sophisticated attack vectors, it is important to consider the design of robust and efficient real-time threat detection systems that can adapt to new environments and changing threat characteristics. This paper aims to model a threat detection system by intersecting domain adaptation, a subset of transfer learning, feature engineering technique and ensembled machine learning models for efficient threat detection. We carry out a systematic review of existing works, methodologies and frameworks, analyze the impact of feature discrepancy on threat detection model, and propose an ensemble-based model [17] integrated with innovative techniques described above to leverage the techniques' individual strengths.

3. LITERATURE REVIEW

3.1 Related Work

There have been several challenges posed by data feature discrepancies for machine learning models, prominent among them is low model performance. In recent times, extensive studies have been carried out by many researchers to solve the problem of feature

discrepancies in machine learning where numerous techniques including ensemble methods, domain adaptation, feature engineering, etc. have proposed multiple techniques, including to mitigate the adverse effects of such discrepancies. In this section, we review a handful of related works in this direction. One of the renowned techniques for improving the accuracy of machine learning models in the face of feature discrepancies in Ensemble Models. Here, techniques such as bagging and boosting [18] have been applied in effectively reducing feature variance and domain shift in datasets. Recently, some researchers have built models on different feature subsets using some sort machine learning model aggregation strategies to achieve robustness against feature discrepancies [19]. [20] proposed a hybrid feature selection with an ensemble classifier to select relevant features and provides consistent attack classification. To achieve effective threat detection, they used CfsSubsetEval, genetic search, and a rule-based engine to effectively select subsets of features with high correlation. They claimed that their model drastically reduces False Alarm Rate (FAR). However, there are specific limitations occasioned by individual techniques deployed by the authors which might degrade the model's performance. The CfsSubsetEval method is sensitive to data distribution, assumes linear relationships between features and ignores temporal variations in data. The genetic search component can introduce the risks of overfitting and parameter sensitivity following genetic algorithm's reliance on parameter tuning [21]. Rule-based feature selection is existentially based on heuristics to select features. This approach has notable drawbacks including lack of adaptability, domain knowledge dependent, limitation in caring for interactions, and there is the potential for bias [22]. [23] introduced a solution to the problem of botnet detection where they used a hybridized feature selection approach (consisting of Categorical Analysis, Mutual Information, and Principal Component Analysis) to enhance the detection capabilities of the ensemble learners. For the ensemble technique, Extra trees was used to help in adapting the detection model to new botnet threats. While the used of the comprehensive feature selection method offers some gains, there are individual feature selection disadvantages which might adversely affect the model's performance. For instance, Categorical Analysis may not be appropriate for continuous variables except the variables are discretized while for Principal Component Analysis method there is the problem of loss of interpretability, assumption of linear relationship among features, and is sensitive to scaling [24]. [25] presented a review on feature selection and ensemble techniques used in anomaly-based IDS research. The paper categorized feature selection techniques to determine individual technique's effectiveness on machine learning-based threat detection models during training and detection phases and concluded that selection of most relevant features in a dataset increases the efficiency of detection in terms of accuracy of the model. They also focused on ensemble techniques employed in anomaly-based IDS models and illustrated how this technique improves the performance of the anomaly-based IDS models. To offer significant improvements in existing intrusion detection systems (IDS) in the Internet of Things domain, [26] proposed an ensemble-based intrusion detection model using logistic regression, naive Bayes, and decision tree as the machine learning algorithms deployed with a voting classifier. The evaluated their model's performance with some existing state-of-the-art techniques using the CICIDS2017 dataset and the result showed significant improvement in terms of accuracy as compared to existing models in terms of both binary and multi-class classification scenarios. However, a stacked ensemble used in the model can make the training of the model very complex and slow since the individual models requires separate training. An existential ensemble model such as Extra Trees (Extremely Randomized Trees) could give a good balance in the bias-variance trade-off by introducing randomness offering robust performance across various datasets. In most cases, the selection of the base models to form stacked ensemble machine learning model in order to gain accuracy with neural models trained in detecting novel threats is a nontrivial problem. [27] presented a novel method named PANACEA to detect cyber-threat, by integrating ensemble learning with adversarial training. The main objective of their work was to enhance the accuracy of neural models addressing threat challenges by creating an ensemble consisting of different base models. The study focused significantly on the selection and pruning of these base models using eXplainable AI (XAI) to improve diversity and improve the accuracy of the ensemble model. They evaluated how different models respond to various input feature subspaces, and used the result of the evaluation to refine the training process, and targeted the models' performances in identifying diverse attack patterns. They conducted empirical validation on several benchmark datasets and results show that the combination of adversarial training, ensemble learning, and XAI techniques was effective in improving multiclass classification accuracy in the datasets. Since the proposed model was based on deep learning, there is

going to be issues of scalability with regard to computational resources for training the model. Wireless Sensor Networks (WSNs) is no doubt backbone of Internet of Things (IoT) and require adequate threat detection strategy. [28] presented a first of its kind method code named Weighted Score Selector (WSS) using ensemble-based machine learning (ML) techniques aimed at improving the detection of threats in WSNs. Their choice of a machine learning approach to the solution is a wise choice because of the recent adoption of this approach in threat detection especially for real-time threat monitoring. The WSS model uses a combination of machine learning classifiers utilizing the strengths of the prominent ones during threat detection in improving the overall performance of the model. Compared with traditional ensemble techniques such as Boosting, Bagging, and Stacking, WSS substantiates the position of the authors in terms of the model's effectiveness in threat detection. Recently, researches have been conducted which integrate ensemble learning techniques for the design of threat detection solutions to handle feature discrepancies. [29] propose a binary classifier approach developed from a machine learning ensemble method to filter and dump malicious traffic to prevent malicious actors from accessing the IoT network and its peripherals. They employed the gradient boosting machine (GBM) ensemble approach to train the binary classifier using pre-processed recorded data packets to detect the anomaly and prevent the IoT networks from zero-day attacks. [30] proposed a domain adaptive ensemble learning (DAEL) framework to address both unified domain adaptation (UDA) and domain generalization (DG) problems. The proposed framework consisted of one CNN feature extractor shared across domains and multiple classifiers with each classifier trained to specialize in a particular source domain thereby acting as an expert in its own domain. Overall, these experts in the DAEL framework collaboratively forms an ensemble and learns complementary information from each other to be efficient in an unfamiliar domain. Under this arrangement, one classifier's source domain becomes another classifier's target domain, which can actively check feature discrepancy across domains. The authors experimented their model on three multi-source UDA datasets and two DG datasets and the results show significant improvement in the state of the art on both problems. This approach leverages the diversity of predictions from multiple models while simultaneously addressing feature discrepancies during training. In [31], a novel approach to solve the problem of multisource domain adaptation (MDA) was proposed using Dual-Level Alignment Network with Ensemble Learning (DANE). In particular, the issue of intradomain and interdomain shifts that hinder knowledge transfer from multiple labeled source domains to an unlabeled target domain was addressed. The objective of the paper was to enhance the performance of the classifier(s) by effectively using knowledge present in the source domains. [32] presented a timely and important contribution to the field of malware detection by proposing an unsupervised domain optimization (UDA)-based malware detection method. It addressed the challenge of detecting of known and unknown malware, hence earning to reduce source (labeled) and target domain (unlabeled) using the distribution divergence between the source and target domain which is minimized with the help of symmetric adversarial learning. The traditional approach often falters in the face of rapidly changing cyber threats. The use of two public datasets in the evaluation enhances the reliability of the proposed method. They found that an accuracy rate of 95.63% was impressive in detecting unknown malicious code and indicates that UDA-based approaches are effective in a variety of situations. Although the proposed method shows promising results, the complexity of implementing symmetric adversarial learning poses challenges for practitioners.

3.2 Research Gap/Problem Statement

Despite the fact that existing threat detection systems are dependent on various algorithms that utilize specific features of security data to identify latent threats, there is significant unattended challenges posed by feature discrepancies across datasets. State-of-the-arts approaches fail to consider the non-homogeneity in the number and types of features present in different security datasets, which is capable of misclassifying threats and diminishing threat detection performance. Furthermore, a number existing methodologies do not adequately leverage the strength in combining domain adaptation (DA) and feature engineering (FE) techniques but often focus on either of these techniques in isolation. Hence, the complementary benefits of these approaches are never harnessed. While earlier results from studies in this direction suggest that ensemble techniques have the tendency to improve detection accuracy, an all-inclusive study that systematically assess the impact of feature discrepancies on the overall effectiveness of threat detection systems is lacking. Therefore, the purpose of this study is to fill the identified gap by proposing the Ensemble-DAFE method, which incorporates the DA and FE techniques to establish a more robust framework for threat detection in security.

3.3 Methodology

This section outlines the hybrid methodology employed to address data feature discrepancies and the ensemble machine learning approach to ensure efficient real-time threat detection. Our approach integrates ensemble methods, domain adaptation techniques, and a systematic process for feature engineering and selection.

We structure our methodology in five phases. Phase 1 handles data collection and preprocessing. In phase 2, we conduct feature engineering and selection. The other phases include domain adaptation phase, model development phase, and lastly the evaluation phase. We first all give the architecture of the proposed model as shown in figure 2.

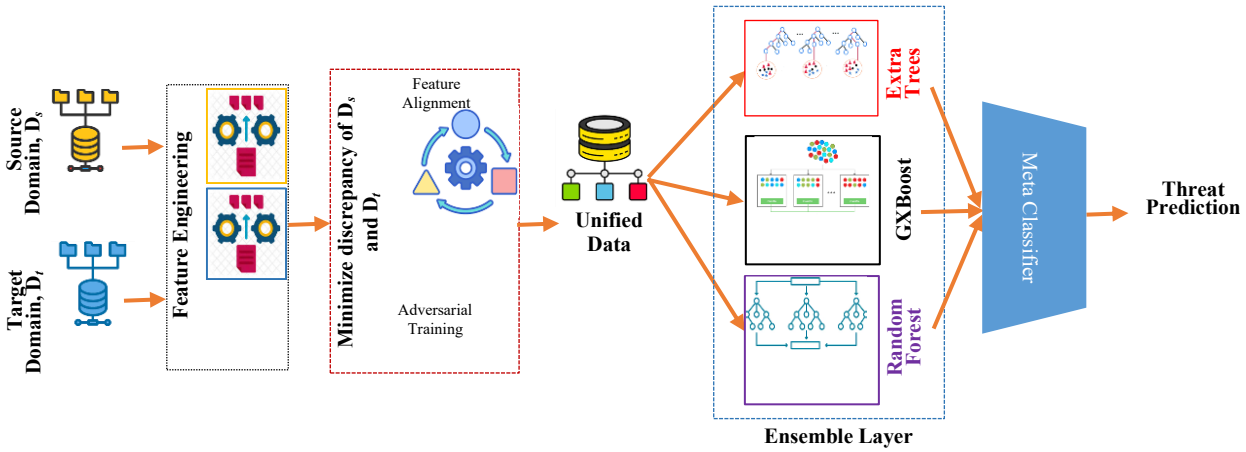

Figure 2: Architecture of Proposed Ensemble-DAFE model for with Feature Discrepancy Mitigation

i. Phase 1: Data Collection and Preprocessing

Dataset

For this experiment, we generated synthetic security dataset code named SYNCyberNet Dataset in Python environment using faker, pandas and NumPy libraries. A base structure of generated synthetic (SYNCyberNet) dataset is defined and comprised of 200,000 rows with 25 security features including Threat_ID, Timestamp, Destination_IP, Source_IP, Byte_Count, Protocol, Port, Attack_Vector, User_Agent, Session_Duration, and Malware_Detected, etc. For simplicity, we labelled 20% of the records as attack signifying that a threat is detected and 80% labeled as normal signifying no threat.

Though we understand that synthetic data cannot replace real world dataset and despite the limitations of synthetic datasets including possible domain gap, overfitting to synthetic data, lack of realism, etc., there are several reasons to justify our use of a synthetic dataset to validate the model proposed in this work. Some of the justifications include:

(i) Using synthetic data in validating our model offers us precise control over the characteristics of the dataset. With this, we are able to control how features are distributed, noise level, etc.

(ii) We also opted to use synthetic data to create data with subtle anomalies or with highly correlated feature combinations. This allows us to perform targeted testing to determine model robustness.

(iii) Using synthetic data ensures repeatable validation experiments due to the deterministic nature of the data generation. This allows varied model versions performance to be consistently compared.

(iv) Real world datasets do not address latest attacks and emerging threats. Synthetic data allows simulating threats not yet observed in real world data, enabling the model to be ready for evolving threats.

Defining Source and Target Domains

To create the source domain, we specified the SYNCyberNet Dataset as the source domain and generate the target domain from the source domain data by introducing domain specific variations in the datasets and simulate discrepancies in the target domain data using 6 of the features including Byte_count, Is_Vulnerable, Threat_Level, Protocol, User_role, and Attack_Vector to generate the target domain data with no labels. In this paper, we model the source domain data and target domain data using equation 1 and equation 2 respectively:

$$D_s = \{(\mathrm{x}_i^s, y_i^s)\}_{i=1}^{n_s} \quad (1)$$

$$D_t = \{(\mathrm{x}_j^t)\}_{j=1}^{n_t} \quad (2)$$

where D_s and D_t represent source domain with labels and target domain without labels and the two have different distributions, n_s and n_t denote the number of threats in D_s and D_t respectively, and y_i^s represents the label of sample x_i^s in D_s.

Feature Discrepancy Analysis

We specified a Python code with libraries such as seaborn, matplotlib with Numpy, and pandas to carry out correlation analysis on the source and target domain datasets using heatmaps to visualized the resulting correlation matrices to understand and access the complex relationship across the two domains. This step is playing a very prominent part in preemptive feature engineering and selection. Correlation matrices for source domain data and target domain data are shown in figures 3.

ii. Phase 2: Feature Engineering and Selection

We conducted feature engineering to create new features by combining the source and target domains based on domain knowledge resulting a unified domain saved in a csv files (unified_dataset.csv). to achieve this, we specified a Python code that import category_encoders to implement Target Categorical Encoding. The source and target domain files were loaded into pandas DataFrame and used to create combined features from selected features such as Byte_Count', 'Attack_Vector', 'Is_Vulnerable', 'Event_Type', 'Threat_Level', and 'User_Role'. We setup sets up a Target Encoder to convert categorical variables into numerical values. Finally, the DataFrame with the new features and encoded values were saved as unified_dataset.csv. Figure 4 captures the output of our Python code for feature engineering showing 5 rows of the combined features.

```
Python 3.12.4 | packaged by Anaconda, Inc. | (main, Jun 18 2024, 15:03:56) [MSC
v.1929 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 8.25.0 -- An enhanced Interactive Python.

Restarting kernel...

In [1]: runfile('C:/Users/Dr. Joshua Tom/combine_features.py', wdir='C:/Users/
Dr. Joshua Tom')
   Byte_Count      Attack_Vector  ...  High_Byte_Vulnerability User_Role_Threat
0        6083            Malware  ...                        1        Admin_Low
1        8286  Denial of Service  ...                        1     Admin_Medium
2        8252           Phishing  ...                        0      User_Medium
3        1529  Network Intrusion  ...                        0     Guest_Medium
4        8090  Network Intrusion  ...                        1     Guest_Medium

[5 rows x 9 columns]
Feature engineering and encoding completed successfully.

In [2]:
```

Fig. 4: Feature Engineering using Target Categorical Encoding using Category Encoders for Verification

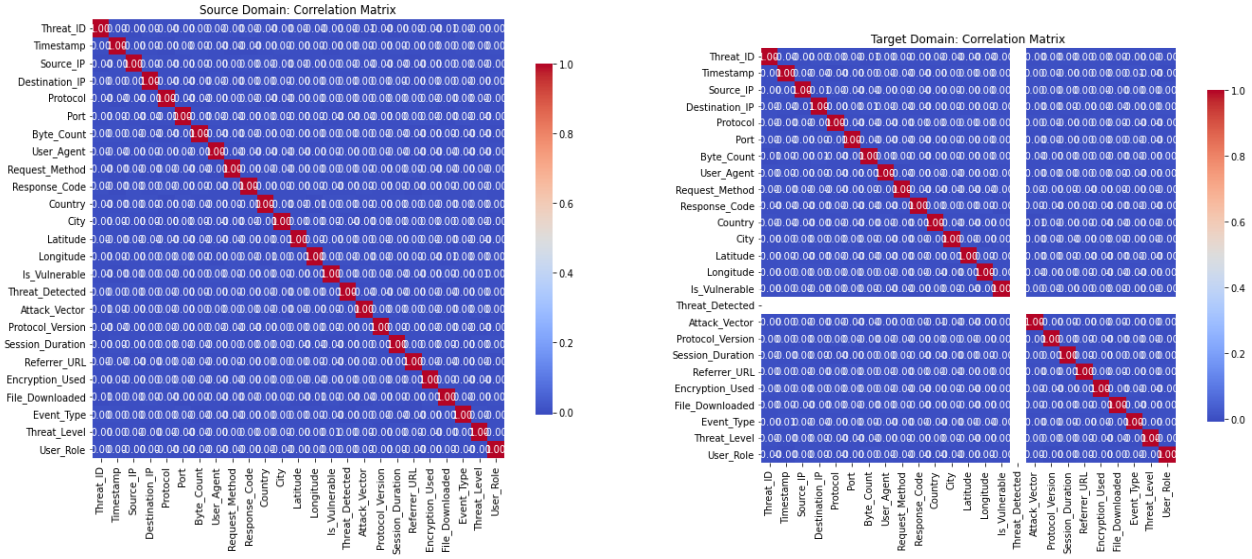

Figure 3: Correlation Matrices for Source and Target Domain

Fig. 3a: Correlation Matrice for Source Domain *Fig. 3b: Correlation Matrice for Target Domain*

We use two (2) methods for feature selection: filter and wrapper to select the associated features, and remove irrelevant or redundant features. We selected features using the filter method (ANOVA F-value as score function). Recursive Feature Elimination (RFE) was the Wrapper Method used.

3.4 Propose Ensemble Model

The ensemble technique creates a unified model by combining multiple feature-based detection methods that share the characteristics of voting. Every feature-based detection method is trained on different dataset and detects a specific threat class.

iii. Phase 3: Cross Domain Adaptation

There are four techniques for adapting a model to target domain. These include pretrained model fine-tuning, adversarial training, self-training, and feature alignment. To address discrepancies and improve model transferability across the source and target domains, we used Feature Alignment technique for domain adaptation known as Maximum Mean Discrepancy, which is based on feature transformation. Maximum mean discrepancy (MMD) is generally a class of nonparametric two-sample tests

aimed at maximizing the mean difference between samples from the source domain, X to the target domain, Y over all choices of data transformations living in some function space [33].

Detailed Implementation of MMD in a Domain Adaptation Setting

We implemented domain adaptation using MMD as follows:

We passed both source and target data through a CNN feature extractor to obtain feature representations for both domains. We choose the Gaussian (RBF) kernel function, $k(x, x')$ for MMD to map the data into a higher dimensional space as given below:

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad (3)$$

where x and $x'$are feature vectors and σ represents the kernel bandwidth.

The MMD loss is calculated as the square of the distance between the mean embeddings of the source and target distributions in the reproducing kernel Hilbert space (RKHS) and is given as:

$$\mathrm{MMD}(\mathrm{P}, \mathrm{Q}) = \left\| \frac{1}{n_s}\sum_{i=1}^{n_s} k(x_i, .) - \frac{1}{n_t}\sum_{i=1}^{n_s} k(y_j, .) \right\|_H^2 \quad (4)$$

where P and Q are the source and target distributions, x_i and x_i are samples from the source domain and target domain respectively, n_s and n_t are the number of source and target samples respectively and H is the Hilbert space. Equation 4 can be evaluated as:

$$\begin{aligned}\mathrm{MMD}(\mathrm{P}, \mathrm{Q}) = &\frac{1}{n_s^2}\sum_{i=1}^{n_s}\sum_{j=1}^{n_s} k(\mathrm{x}_i, \mathrm{x}_j) \\ &+ \frac{1}{n_t^2}\sum_{i=1}^{n_t}\sum_{j=1}^{n_t} k(\mathrm{y}_i, \mathrm{y}_j) \\ &- \frac{2}{n_s n_t}\sum_{i=1}^{n_s}\sum_{j=1}^{n_t} k(x_i, \mathrm{y}_j)\end{aligned} \quad (5)$$

where x_i's are the source domain data points and x_j's are the target domain data points. k is the *pairwise* radial basis function-based kernel provided in the scikit-learn library to calculate the radial basis function (RBF) kernel between pairs of samples in the source and target domain data. This allowed us to determine whether there is a serious feature discrepancy across the source and target data. In this experiment, the threshold for acceptable MMD value was set at 0.15. Firstly, our MMD value was 0.1811 suggesting that we have detected high discrepancy. Therefore, based on this value, domain adaptation using CCA is implemented which tries to overcome or minimize the computed discrepancy. The reduced MMD value of 0.0172, following the CCA minimization process confirms that there is considerable success in reducing initial feature discrepancy within data structure. Figure 5 presents a visualization of source and target domain (left) with transformed sources/targets after CCA minimization applied (right). Unsurprisingly, this decreased difference allowed the threat detection model to operate better. Finally, we minimize the MMD loss by adding the MMD loss as a regularization term to the overall loss function during training, which makes our proposed model to learn to minimize both the classification/regression loss on the labeled source data and the discrepancy between the source and target feature distributions.

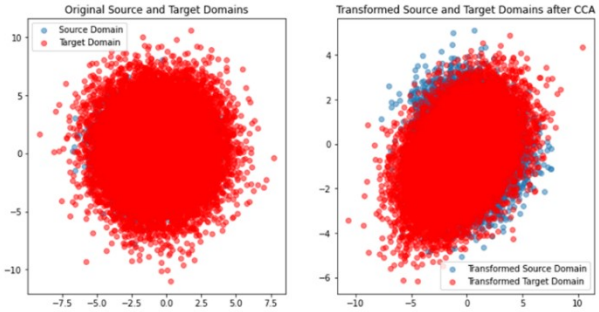

Figure 5: Visualization of the Original Source and Target Domain and the Transformed Source

iv. Phase 4: Model Development

Ensemble Learning Approach

An ensemble of multiple machine learning models made up of RandomForestClassifier, ExtraTreesClassifier, and XGBClassifier was constructed to leverage the strengths of different algorithms with Logistic Regression Classifier as the meta classifier. The base models were selected to form the ensemble model based on the respective strengths and capabilities. The Extra Trees were chosen because of its power of interpretability and effectiveness in capturing non-linear relationships. Random Forests was included to enhance robustness and prevent overfitting while offering improved performance across varying datasets. XGboost classifier was considered to provide for optimization of loss functions using its sequential learning ability. To build the ensemble model, we used StackingClassifier ensemble learning method from the scikit-learn library. We specified the StackingClassifier's initial estimators as the base models and the final estimator as the meta classifier.

Model Training and Tuning

For the training of an already developed model (Ensemble-DAFE), both the grid search and cross validation were used during hyperparameter

optimization so as to obtain the configurations of models to be incorporated in the final ensemble. Two of the source domains were set up for training and validation. The validation data set was 20% of the total source domain while training data set was 80%. In order to build a prediction model for the source domain data, each of the base models was trained and tested separately. All base models' predictions were implemented into the Logistic regression based stacked ensemble model, which was trained and tested as well. The final prognosis was based on the outputs of each model through a process known as majority voting to improve confidence and precision. Their performances in terms of model accuracy are given in table 1.

Table 1: Performances of the Base Models and the Stacked Ensemble Model before minimizing feature Discrepancy

Model Name	Model Accuracy	Precision	Recall	F1-Score
Random Forest Classifier	93.18%	0.9441	0.9193	0.9316
Extra Trees Classifier	94.16%	0.9490	0.9245	0.9366
XGBoost Classifier	94.24%	0.9507	0.9340	0.9423
Proposed Ensemble Model	96.38%	0.9512	0.9368	0.9940

v. Phase 5: Evaluation

Evaluation Metrics

A detailed evaluation of model performance was carried out which included parameters like Accuracy, Precision, Recall, and F1 Score. The values obtained for the different metrics are as given in table 1. In this paper, we provided for model evaluation to be computed through the Accuracy, Precision, Recall and F1 Score as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

where Accuracy is a measure of the ratio of correctly classified samples to the total number of samples. TP is true positives, TN true negatives, FP false positives and FN represents false negatives.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

where Precision is the ratio of positive samples to forecasted positive samples.

$$\text{Recall} = \frac{TP}{TP + FN} \quad 6)$$

where Recall refers to the ratio of predicted positive samples against actual positive samples.

$$\text{F1–score} = \frac{Precision \text{ x Recall}}{Precision + \text{Recall}} \quad (7)$$

where F1-score represents a weighted average of Precision and Recall to balance the effects of accuracy and recall, and to evaluate a classifier comprehensively. Targets were achieved after resolving the feature discrepancies through application of CCA as explained in the Domain Adaptation section above, we obtained the transformed data (figure 5b), and re-trained and re-evaluated the base and stacked ensemble models on the transformed dataset. The outcome of the re-training and re-evaluation is presented in table 2.

Table 2: Performances of the Base Models and the Stacked Ensemble Model on the feature Discrepancy Minimized Dataset

Model Name	Model Accuracy	Propose Ensemble model accuracy
Random Forest Classifier	96.87%	99.96%
Extra Trees Classifier	98.27%	
XGBoost Classifier	98.89%	

4. COMPARATIVE ANALYSIS

The performance of the proposed hybrid model was compared against baseline models trained solely on the target domain and models without domain adaptation. The study in [26] designed an ensemble-based intrusion detection model using logistic regression, naive Bayes, and decision tree as the machine learning algorithms and obtained an average accuracy of 88.94% without implementing domain adaptation for feature discrepancy handling. The proposed model integrated with domain adaptation for feature discrepancy handling outperforms theirs with 99.96 performance accuracy. [29] proposed an anomaly-based intrusion detection system for security against DDoS using gradient boosting machine ensemble classifier with no domain adaptation capability, which performed well for binary classification yielding accuracy of 98.27%. Compared to this, the proposed approach outperformed the one proposed in [29] even with high computational requirement of our model. The work in [32] used symmetric adversarial learning to minimize the distribution divergence between the source and target domain in their proposed unsupervised domain adaptation (UDA)-based malware detection method and obtained a performance accuracy of 95.63%,

which clearly shows that our method outperforms theirs in threat detection accuracy

5. DISCUSSION ON THE RESULTS

The results of the experiments conducted show that the ensemble model developed in this paper, Ensemble-DAFE outperforms individual base detection methods in terms of accuracy and reduced false negatives and false positives. We specifically noticed that our ensemble technique achieves an F1-score of 0.9940 against the highest F1-score of 0.9423 exhibited by the XGBoost classifier as the best individual model. Our ensemble model achieves a performance improvement of 2.51% in detection accuracy due to feature discrepancy minimization. As a result of this discrepancy handling, the developed ensemble technique reduced false negatives by 6% and false positives by 10% compared to individual models used in the ensemble.

6. CONCLUSION

Ensemble methods have been shown to effectively resolve the issue of feature disparity in threat detection algorithms. Improving more than one feature-based detection strategy concurrently allows us to increase the efficiency of detecting a threat while minimizing both false positives and negatives. Our method can be used for different ranges of threat detection systems such as network behavior analysis, system log analysis, or user actions analysis. The method described in this paper for improving threat detection performance by addressing feature differences, between source and target domains in security datasets involves creating a model that can manage data feature variations in threat detection effectively. Stacked ensemble, with domain adaptation and thorough feature engineering and selection processes can improve both detection accuracy and overall performance in a domain setting. The literature review conducted in this paper highlights the increasing importance of dealing with data feature differences to maintain the efficiency of ensemble models. Researchers have been working on improving the performance and reliability of machine learning systems by using domain adaptation techniques and experimenting with methods, across different applications and datasets successfully over the years. In a nutshell although there has been advancement in handling variations in data features using these methods, the intricate nature of real-world situations calls for ongoing exploration of combined and flexible approaches to further boost both model reliability and effectiveness. In the future, we intend to evaluate the performance of our stacked ensemble model on larger and real-world datasets. This paper suggests further investigation into the implications of increasing the number of base models in a given ensemble architecture on the detection performance of a threat detection model. It is also important to explore and determine the appropriateness of other ensemble methods such as bagging and boosting in future threat detection solutions.

REFERENCES

[1] J. Lan X. Liu B. Li et al. A Novel Hierarchical Attention-based Triplet Network with Unsupervised Domain Adaptation for Network Intrusion Detection. Appl Intell 53, 11705–11726, 2023.

[2] S. Sharma and N. S. Yadav. Ensemble-based Machine Learning Techniques for Attack Detection, 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, pp. 1-6, 2021.

[3] S. Dandyala and S. Banik Traditional Methods of Threat Detection. International Journal of Advanced Engineering Technologies and Innovations, 1(2), 161-177, 2021.

[4] L. L. Scientific Ensemble Machine Learning Algorithm Methods for Detecting the Attacks Using Intrusion Detection System. Journal of Theoretical and Applied Information Technology, 102(5), 2024.

[5] F. Khani and P. Liang. Feature Noise Induces Loss Discrepancy Across Groups. In International Conference on Machine Learning (pp. 5209-5219). PMLR, 2020.

[6] J. Yan R. Sun T. Liu S. Duan. Domain-adaptation-based active ensemble learning for improving chemical sensor array performance. Sensors and Actuators A: Physical, 357, 114411, 2023.

[7] T. Zaid and S. Garai. Emerging Trends in Cybersecurity: A Holistic View on Current Threats, Assessing Solutions, and Pioneering New Frontiers. Blockchain in Healthcare Today, 7, 2024.

[8] Dou Y, Yang H, Yang M, Xu Y, Ke D. Dynamically mitigating data discrepancy with balanced focal loss for replay attack detection. In 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 4115-4122),

2021. IEEE.

[9] S. Zhao G. Wang S. Zhang Y. Gu Y. Li Z. Song K. Keutzer. Multi-Source Distilling Domain Adaptation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 12975-12983), 2020.

[10] P. Singhal R. Walambe S. Ramanna K. Kotecha. Domain Adaptation: Challenges, Methods, Datasets, and Applications. IEEE Access, 11, 6973-7020, 2023.

[11] I. Karna A. Madam C. Deokule R. Adhao V. Pachghare. Ensemble-based Filter Feature Selection Technique for Building Flow-based IDS. In 2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS) (pp. 324-328), 2021. IEEE.

[12] P. R. Kothamali S. Banik S. V. Nadimpalli. Feature Engineering for Effective Threat Detection. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence, 12(1), 341-358, 2021.

[13] S. Zhou D. Chen J. Pan J. Shi J. Yang. Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2952-2963), 2024.

[14] H. Do Hoang T. B. Xuan T. N. N. Minh P. T. Duy V. H. Pham. DA-GAN: Domain Adaptation for Generative Adversarial Networks-assisted Cyber Threat Detection. In 2022 RIVF International Conference on Computing and Communication Technologies (RIVF) (pp. 29-34), 2022. IEEE.

[15] A. S. Li, A. Iyengar A. Kundu E. Bertino. Transfer Learning for Security: Challenges and Future Directions. arXiv preprint arXiv:2403.00935, 2024.

[16] A. Odeh and A. Abu Taleb. Ensemble-Based Deep Learning Models for Enhancing IoT Intrusion Detection. Applied Sciences, 13(21), 11985, 2023.

[17] N.I. Haque, M.A Rahman, H. Shahriar. Ensemble-based efficient anomaly detection for smart building control systems. In 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC) (pp. 504-513), 2021. IEEE.

[18] A. Mosavi F. H. Sajedi, B. Choubin M. Goodarzi A. A. Dineva E. Rafiei Sardooi. Ensemble boosting and bagging based machine learning models for groundwater potential prediction. Water Resources Management, 35, 23-37, 2021.

[19] M. Moshawrab M. Adda A. Bouzouane H. Ibrahim A. Raad. Reviewing federated learning aggregation algorithms; strategies, contributions, limitations and future perspectives. Electronics, 12(10), 2287, 2023.

[20] E. Jaw and X. Wang. Feature selection and ensemble-based intrusion detection system: an efficient and comprehensive approach. Symmetry, 13(10), 1764, 2021.

[21] M. Mosayebi and M. Sodhi. Tuning genetic algorithm parameters using design of experiments. In Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion (pp. 1937-1944), 2020.

[22] F. Chiroma M. Cocea H. Liu. Evaluation of rule-based learning and feature selection approaches for classification. In 7th Imperial College Computing Student Workshop (pp. 1-6). Schloss Dagstuhl-Leibniz-Zentrum fur Informatik GmbH, Dagstuhl Publishing, 2019.

[23] M.A. Hossain, M.S. Islam. A novel hybrid feature selection and ensemble-based machine learning approach for botnet detection. Sci Rep 13, 21207. https://doi.org/10.1038/s41598-023-48230-1, 2023.

[24] C. Peng Y. Chen Z. Kang C. Chen Q. Cheng. Robust principal component analysis: A factorization-based approach with linear complexity. Information Sciences, 513, 581-599, 2020.

[25] M. Torabi N.I. Udzir M.T. Abdullah R. Yaakob. A Review on Feature Selection and Ensemble Techniques for Intrusion Detection System. International Journal of Advanced Computer Science and Applications, 12, 2021.

[26] A. Abbas M.A. Khan S. Latif et al. A New Ensemble-Based Intrusion Detection System for Internet of Things. Arab J Sci Eng 47, 1805–1819. https://doi.org/10.1007/s13369-021-06086-5, 2022.

[27] M. AL-Essa G. Andresini A. Appice et al. PANACEA: a neural model ensemble for cyber-threat detection. Mach Learn 113,

5379–5422. https://doi.org/10.1007/s10994-023-06470-2, 2024.

[28] S. Ismail Z. El Mrabet H. Reza. An Ensemble-Based Machine Learning Approach for Cyber-Attacks Detection in Wireless Sensor Networks. Applied Sciences. 13(1):30. 2023.

[29] P. Verma A. Dumka R. Singh A. Ashok A. Gehlot PK Malik M. Hedabou. A novel intrusion detection approach using machine learning ensemble for IoT environments. Applied Sciences, 11(21), 10268, 2021.

[30] K. Zhou Y. Yang Y. Qiao T. Xiang. Domain adaptive ensemble learning. IEEE Transactions on Image Processing, 30, 8008-8018, 2021.

[31] Y. Yang L. Wen P. Zeng B. Yan Y. Wang. DANE: A Dual-level Alignment Network with Ensemble Learning for Multi-Source Domain Adaptation. IEEE Transactions on Instrumentation and Measurement, 2024.

[32] F. Wang G. Chai Q. Li C. Wang. An efficient deep unsupervised domain adaptation for unknown malware detection. Symmetry, 14(2), 296, 2022.

[33] S. Paik, M. Celentano, A. Green, R. J. Tibshirani. Maximum mean discrepancy meets neural networks: The radon-kolmogorov-smirnov test. arXiv preprint arXiv:2309.02422, 2023.