Journal of Theoretical and Applied Information Technology
30th April 2023. Vol.101. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ANOMALY DETECTION IN CYBER-PHYSICAL SYSTEMS USING EXPLAINABLE ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

SHERIF KAMEL HUSSEIN [1,2] , MOHAMED A. EL-DOSUKY [1,3]

[1] Dep. of Computer Science, Arab East Colleges, Riyadh, Saudi Arabia

[2] Dep. of Communications &Computer Engineering, October University for Modern Sciences &Arts, Egypt

[3] Dep. of Computer Science, Faculty of Computers & Information, Mansoura University, Egypt
E-mail: skhussein@arabeast.edu.sa, maldosuky@arabeast.edu.sa

ABSTRACT

Cyber-Physical Systems (CPS) embrace integration between digital & physical components of production environments. Data analysis approaches operate on big data, which makes them somewhat limited in industrial applications. Not all of anomaly detection techniques are applicable in ensuring security of CPSs. These techniques face huge volumes of data and require domain-specific knowledge, which necessitates the invention of solutions that integrate advanced AI technologies and models. This paper utilizes Explainable Artificial Intelligence (XAI) & Machine Learning (ML) approaches for detecting the anomalies in CPS. The proposed model improves our understanding of the complex phenomena in CPSs by analyzing the extracted features using feature engineering selection and detecting the outliers of each class labels. Hence, the main motivation of this paper is to scrutinize challenges and emerging trends in Anomaly Detection for CPSs. Furthermore, studying the outlier detection algorithms such as Angle-based Outlier Detection (ABOD) and Clustering Based Local Outlier Factor (CBLOF) to be compared with the proposed approach. Neither of ABOD nor CBLOF succeeds in distinguishing the outlier class. Therefore, the proposed approach attempts to handle the outlier detection by using feature engineering and XAI approaches. Moreover, ML based Random Forest (RF) achieves better results than Support Vector Machine (SVM), Naïve Bayes (NB), and multi-layer perceptron (MLP).

Keywords: *Anomaly Detection, Machine Learning, Cyber-Physical Systems, Explainable Artificial Intelligence, Outlier Detection*

1. INTRODUCTION

The development of Cyber-Physical Systems (CPS) is due to the integration between digital and physical components of production environments. The application of machine learning and deep learning approaches to CPSs has shown a great potential. However, these data analytic approaches operate on big data, making them limited in industrial applications to some extent [1].

Recently, many anomaly detection techniques are proposed. Yet, not all of these techniques are applicable in ensuring security of CPSs [2]. These techniques are confronted with massive data volumes and require domain-specific knowledge. This necessitates the invention of solutions that integrate advanced artificial intelligence techniques and models. Cyber-physical attacks such as Stuxnet and Triton pose many challenges and issues [3], [4]. Thus, data analytic approaches attempt exploiting many features of certain industrial equipment to detect possible failures, especially those created by malicious activities [5].

As we shall see later in this paper, outlier detection algorithms such as Angle-based Outlier Detection (ABOD) and Clustering Based Local Outlier Factor (CBLOF) are tried. Neither of them succeeds in distinguishing outlier class. This is partially because data is very intertwined. Then, machine learning algorithms are tried after feature engineering and explainable Artificial Intelligence. The aim of this paper is to scrutinize challenges and emerging trends [6], [7] in Anomaly Detection for CPSs. Furthermore, understanding the theoretical models such as explainable Artificial Intelligence models and analytics of Anomaly Detection for CPSs to improve our understanding of complex phenomena in CPSs.

Journal of Theoretical and Applied Information Technology
30th April 2023. Vol.101. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The main contributions of this paper are:

- Proposing XAI for extracting the most relevant features of the applied CPS dataset.
- Utilizing ML (RF, SVM, NB, MLP) and deep learning CNN approaches to detect the anomalies resulting from the outlier values of CPS dataset.
- Comparing proposed approach with the ABOD and CBLOF to ensure the reliability and superiority of proposed method.

The rest of this paper is organized as follow. In Section 2, a gentle review of Power Systems as CPSs is provided before presenting the related work. In Section 3, the methodology and tools of this paper are presented. In Section 4, the results and discussions are provided. Finally, the conclusion and future work are in Section 5.

2. RELATED WORK

2.1 Power Systems as CPSs

A power system can be considered as a CPS [8]. Figure 1 shows network diagram of an exemplar power system. The whole system is alimented by two power generators (G1 and G2). These generators are directedly connected to Intelligent Electronic Devices (IEDs; R1 through R4) and breakers (BR1 through BR4). Wireless communication is used to connect those IEDs to the Substation Switch, which in turn is connected with the Control Room and Primary Domain Controller (PDC) [9] .

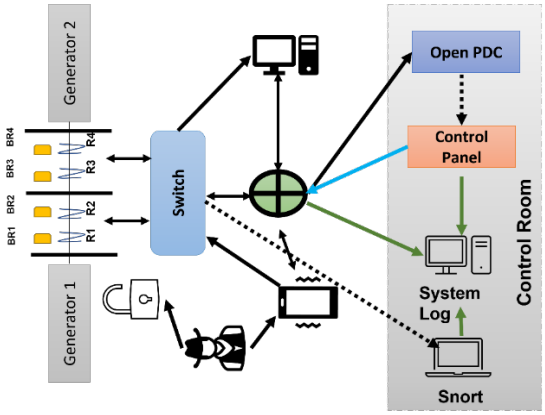

Figure 1:Network diagram of exemplar power system [8].

Regarding Intrusion Prevention Systems (IPSs), Snort is the world's most important open source tool [10]. It applies rules that aid in defining harmful network behavior. These set of rules are then applied in locating packets that match against them and produce warning messages[11].

Snort is employed in line for blocking such packets and as a sniffer for packets, as a packet logger or as a full-fledged network IPS. It may be configured for personal and commercial use [12].

There are 3 datasets downloadable from [13] representing 78,377 events in 15 files: binary dataset (normal or attack), three-class dataset (normal fault, attack, & no-events) and multiclass dataset (distinguishing thirty-seven scenarios). Figure 2 shows three-class dataset distribution.

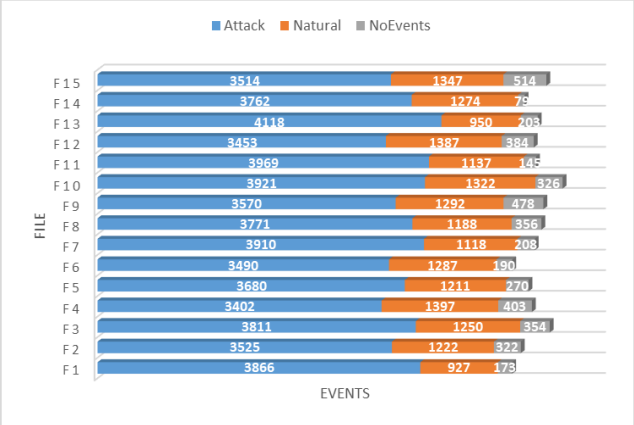

Figure 2: Three-class dataset scenario distribution

Table 1 shows the features and their descriptions. There are 128 features, 116 features of them are given by 4 IEDs. R3-PM5:I, for instance, stands for phase B current phase magnitude at R3. The remaining12 features are for control panel logs, snort logs, and relay logs.

Table 1: The description of the applied features

Feature	Description
PA1: VH through PA3: VH	Phase Angle of Phase A/C Voltage
PM1: V through PM3: V	Magnitude of Phase A/C Voltage
PA4: IH through PA6: IH	Phase Angle of Phase A/C Current
PM4: I through PM6: I	Magnitude of Phase A/C Current
PA7: VH through PA9: VH	Positive - Negative - Zero (PNZ) Voltage Phase Angle
PM7: V through PM9: V	PNZ Voltage Magnitude
PA10: VH through PA12: VH	PNZ Current Phase Angle
PM10: V through PM12: V	PNZ Current Magnitude
F	Relays'Frequency
DF	Relays'Frequency Delta
PA: Z	Impedance at every relay
PA: ZH	Impedance Angle at every relay
S	Relays'Status Flag

2.2 Supervised and Unsupervised Approaches for Anomaly Detection in CPS

Various anomaly detection techniques dependent on the number of labels present in the dataset. A fully labelled training dataset is used in supervised anomaly detection. While the training dataset for semi-supervised anomaly identification is free of anomalies. Then after, anomalies are found by comparing the test data to the normal model and looking for deviations. Finally, unsupervised anomaly detection algorithms simply use the data's inherent information to identify cases that deviate from the rest of the data.

The configuration known as supervised anomaly detection uses fully labelled training and test sets of data. A simple classifier can be employed after being learned. Apart from how frequently classes are wildly unbalanced, this scenario is very comparable to conventional pattern recognition. Because of this, not every classification method is ideal for this purpose. For instance, Artificial Neural Networks (ANN) or Support Vector Machines (SVM) should perform better when dealing with unbalanced data than decision trees like C4.5. The hypothesis that anomalies are recognized and appropriately labelled renders this configuration basically irrelevant. Anomalies may not be anticipated for many applications, or they may appear unexpectedly as innovations during the testing phase.

Unsupervised anomaly detection method grades the data only based on inherent features of the dataset. Distances or densities are frequently used to estimate what is normal and what is an exception. Fig. 3 investigated the most common anomaly detection based on supervised, semi-supervised and unsupervised learning approaches. Here, the supervised learning most used binary classification to classify the anomaly and normal cases in CPS. The semi-supervised might consider the one-class SVM and the semi-supervised SVM which name as S^3VM. The unsupervised learning classified to statistical, and distance-based approaches. The statistical based approaches classified to principal component analysis (PCA), and histogram-based outlier score (HBOS). The distance based might contains the clustering-based K-means, density based local outlier factor. The neighbor based in unsupervised learning of anomaly detection might consider K-nearest neighbor (KNN) and outlier detection using indegree number (ODIN) [14]–[16].

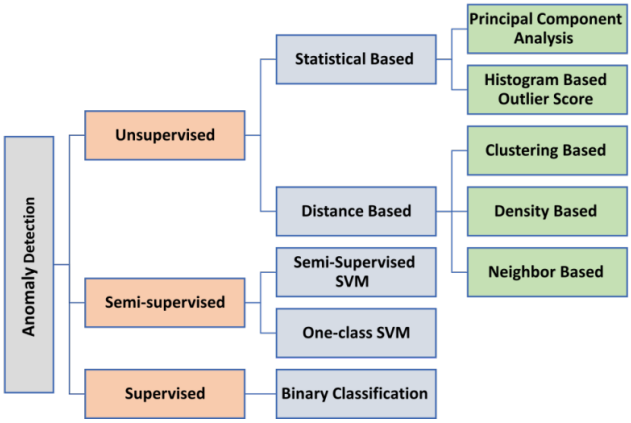

Figure 3: The most common supervised, semi-supervised and unsupervised learning approaches for anomaly detection

To ensure the security of Physical Cyber Systems (CPS), anomaly detection is essential. This is because traditional anomaly detection methods are often unable to properly process the increasing volume of data and the high complexity of CPSs due to multiple attacks. Reference [17] developed Deep learning-based anomaly detection (DLAD). In this reference, they examine contemporary DLAD techniques in CPSs. To grasp the key characteristics of the present approaches, they provide a taxonomy based on the kinds of anomalies, strategies, implementation, and assessment measures. Additionally, they use this taxonomy to point out fresh features and patterns in each CPS domain. Countermeasures are provided for both sensor measurements and innovation sequencing, which can be checked based on the data without relying on a model of the underlying dynamic system. The links between countermeasures and system properties, as well as noise statistics, are then examined to identify conditions that ensure time convergence between countermeasures and thus evaluate the efficiency of the counter entropy measures in attack detection. Denial of service attacks, remodeling, innovation-based spoofing, and data injection attacks are the four types of attacks considered that behave abnormally after a transfer [18], [19].

The efficiency of transfer entropy countermeasures in attack detection is evaluated using theoretical studies, numerical demonstrations, and comparative simulations with typical 2 detectors. Denial-of-service, replay, innovation-based deception, and data injection attacks are the four types of attacks considered. The transfer entropy exhibits anomalous behavior following each of these attacks.

The second recommendation is an anomaly detection module that estimates the posterior

probabilities of normal and abnormal occurrences using a Kalman Filter (KF) and a Gaussian Mixture Model (GMM). The PPAD-CPS architecture is evaluated using two open datasets, the Power System and the UNSW-NB15 dataset. The outcomes of the experiments demonstrate that the framework is superior to four current strategies for achieving high privacy levels. In terms of detection rate, false positive rate, and processing time, the framework outperforms seven competing anomaly detection methods [20].

Existing anomaly detection frameworks frequently do not consider the increased variety of affected systems, even though the ever-increasing interconnectedness of cyber-physical systems increases their attack surface. Existing frameworks either concentrate on a specific fieldbus protocol or necessitate a deeper understanding of the cyber-physical system in question. As a result, we present a standardized approach and framework for using anomaly detection with different fieldbus protocols. They created a feature learning and packet classification approach in one step using stacked denoising autoencoders. Neither specialized knowledge of the application nor specific protocols are required because the method is based on the network traffic's raw bytes stream. They also focus on developing a framework that is effective and can manage the increased communication seen in cyber-physical systems. They have demonstrated using an Ethernet/IP and a Modbus dataset that we can acquire network packets up to 100 times faster than approaches that rely on packet processing. The datasets are from Secure Water Treatment. For longer-lasting attacks, we still manage to reach precision and recall scores of above 99% [21].

2.3 Federated learning for anomaly detection in CPS

Federated learning intrusion detection system (FLIDS) for medical cyber physical systems are proposed. A distributed machine learning approach called federated learning develops a global model by averaging weights from many devices over a number of communication cycles. They alter the initial federated learning algorithm to better detect breaches into medical cyber-physical systems (MCPS). The server serves as the central authority and is in charge of registering the mobile devices, computing the federated model, and storing the model. This is all done through the intrusion detection architecture, which uses the computational resources of the mobile devices to run the detection module. Security flaws and unauthorized access to the private and sensitive medical and health information that MCPS holds can have detrimental impacts on the patient and the hospital, including misuse, liability, loss of privacy, bodily harm, and other harm. The wide range of the systems' participating devices (such as mobile and body sensor nodes) presents large attack surfaces, necessitating the creation of effective security controls for these environments [22].

Due to the quick integration of intelligent networking in traditional industrial infrastructures, the attack surface of Industrial Cyber-Physical Systems (ICPSs) has substantially increased. But protecting such complex large ICPSs from cyber-attacks is very challenging due to the lack of attack cases. To identify online dangers to ICPSs, they presented the DeepFed federated deep learning system. We specifically combine a convolutional neural network and a gated recurrent unit to create a new deep learning-based intrusion detection model for ICPSs. Second, they create a federated learning framework that enables numerous ICPSs to jointly create an extensive intrusion detection model while maintaining privacy. Additionally, a Paillier cryptosystem-based secure communication protocol is designed to maintain the confidentiality and security of model parameters during training. Extensive tests on a genuine ICPSs dataset show the proposed DeepFed scheme's high performance in identifying different kinds of cyber-threats to ICPSs as well as its advantages over cutting-edge techniques [23].

Deep learning is used to address diverse industrial challenges by utilizing ICPSs. Traditional centralized learning (CL) may be inappropriate for various industrial applications involving sensitive data, such as smart medicine, due to privacy regulatory considerations. Federated learning (FL) has recently attracted a lot of attention as a revolutionary cooperation learning strategy that can break down data barriers between various institutions and increase model performance. However, the industrial agents' personal information can be deduced from their shared parameters. They introduced the Privacy-Enhanced Momentum Federated Learning (PEMFL) framework, which combines differential privacy (DP), Momentum FL (MFL), and a chaos-based encryption approach. During training, differential privacy is employed to disrupt the gradient parameters of the industrial agents in order to maintain their privacy information [24].

2.4 Multi-dataset time series for anomaly detection in CPS

For contemporary industrial applications, effective anomaly identification and diagnosis in multivariate time-series data is crucial. Building a system that can quickly and precisely identify abnormal observations is a difficult task, though. This is because modern applications require extremely quick inference times, there are few anomaly labels, and there is substantial data volatility. Only a select few deep learning algorithms for anomaly detection can solve all of these issues, despite recent improvements in the field. They introduced TranAD, an anomaly detection and diagnosis model based on deep transformer networks that leverages attention-based sequence encoders to quickly execute inference while keeping track of the data's larger temporal patterns. Additionally, we can train the model with little data thanks to model-agnostic meta learning (MAML). Extensive empirical experiments on six publicly accessible datasets show that TranAD can perform better in detection and diagnosis than state-of-the-art baseline approaches with data- and time-efficient training. Particularly, TranAD decreases training times by up to 99 percent while increasing F1 scores by up to 17% [25].

There has been a lot of academic and commercial interest in the detection of anomalies in time series. To evaluate time-series anomaly detection techniques, however, there isn't a complete benchmark available. It is typical to employ either I a small number of publicly available datasets, which are frequently biassed to favour particular conclusions, or (ii) proprietary or manufactured data. As a result, we frequently see algorithms that perform remarkably well on one dataset but disappointingly poorly on another, giving the impression of advancement. They carefully reviewed more than 100 papers to find, gather, process, and systematically format datasets proposed in earlier decades to address the aforementioned problems. The efforts presented by [26] in TSB-UAD, demonstrated a new benchmark to evaluation of univariate time-series anomaly detection methods. The total number of time series with labelled anomalies in the TSB-UAD is 13766, and they span a variety of domains with a wide range of anomaly types, ratios, and sizes. 18 previously proposed datasets with 1980 time series are already included in TSB-UAD, and they contribute two dataset collections. They specifically create 958 time series by converting 126 time-series classification datasets into time series with tagged anomalies using a systematic methodology. Additionally, they demonstrate data manipulations that allow us to add fresh anomalies, producing 10828 time series of varying complexity for anomaly detection. Finally, they review 12 example techniques to show that TSB-UAD is a reliable resource for evaluating anomalies detection methods.

2.5 Explainable Artificial Intelligence techniques for anomaly detection in CPS

Industry 4.0, a paradigm that incorporates modern technology and advances, is now a reality. Artificial intelligence (AI) is the primary driver of the industrial transformation because it enables intelligent equipment to do self-monitoring, interpretation, diagnosis, and analysis. Machine learning and deep learning, in particular, assist manufacturers and sectors in forecasting maintenance needs and reducing downtime. Explainable artificial intelligence (XAI) explores and develops methods, algorithms, and tools that produce human-comprehensible information and judgments generated by AI-based systems [27].

XAI is in the process of being integrated into prognostics and health management systems (PHM). The research on PHM-XAI is lacking in terms of uncertainty quantification and explanation evaluation criteria. Authors in [28] present a method of anomaly detection and prognostics for turbines of gas applying Bayesian deep-learning and Shapley additive explanations (SHAP). The method explains the anomaly detection and prognostics, and improves the performance of the prognostics.

In terms of informative index size, informational collection quality, extraction procedures, hyper boundary set, enactment capabilities, and improvement calculations, input data and goal (class) data can currently be prepared with the elite and tested with new information input in traditional deep learning algorithms. These deep learning procedures can give extremely viable results with multiple layers in a profound system allowing it to perceive things at different levels of deliberation. In a system designed to recognize hounds, for example, the lower layers notice fundamental features like schematics or shade; the top layers perceive progressively complicated things like hiding or eyes; and the upper layers describe everything as a canine.

Due to their potential to link computational resources with physical systems, CPSs are crucial components of our modern infrastructure. As a result, the research community continues to pay more attention to issues including the

Journal of Theoretical and Applied Information Technology
30th April 2023. Vol.101. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

dependability, performance, and security of CPSs. Massive amounts of data generated by CPSs present opportunities for the use of predictive Machine Learning (ML) models for performance optimization, preventative maintenance, and threat detection. However, when applied in safety-critical systems like CPSs, the "black-box" character of complicated ML models is a disadvantage. While explainable ML has been a hot topic in recent years, supervised learning has received most of the attention. Relying solely on supervised learning is insufficient for data-driven decision making in CPSs due to the rapid production of enormous amounts of unlabeled data. Consequently, explainable unsupervised ML models are required if we are to make the most of ML in CPSs. In this study, we present a possible use of unsupervised explainable ML in CPSs. We examine the state-of-the-art in unsupervised machine learning, provide the starting requirements of explainable unsupervised ML for CPS, and introduce an explainable clustering methodology based on Self-Organizing Maps that produces both global and local explanations. Authors in [29] evaluated the fidelity of the generated explanations using feature perturbation techniques.

Building a resilient automation system and utilizing advanced ML to develop mitigation and elimination strategies are all necessary to successfully enable Industry 4.0. Authors in [30] present a visual analytics framework and method for situational awareness that includes automatically tracking, analyzing, and forecasting the condition of cyber-physical systems [31]. To identify potential flaws, threats, and malicious assaults, their method relies on visual characterizations of multivariate time series and real-time predictive analytics. They use several aviation datasets available from NASA to confirm utility of their approach [32].

3. PROPOSED FRAMEWORK

This work proposes a framework, as shown in Figure 4. First, the files are merged. Then preprocessing is performed to clean the data from Null entries. Normalization is then performed to ensure that data falls in a range between a minimum and a maximum value. Exploratory Data Analysis (EDA) is performed to spot patterns and anomalies in the data. Explainable AI (XAI) algorithm such as SHAP (SHapley Additive exPlanations) is tried. Outlier Detection algorithms such as Angle-based Outlier Detection (ABOD) and Clustering Based Local Outlier Factor (CBLOF) are tried. The dataset is decomposed into training part and testing part. Training builds a model which is verified in the testing phase. Accuracy of the model is evaluated according to certain metrics.

Then preprocessing is performed to clean the data from Null entries. Normalization is then performed to ensure that data falls in a range, as in Eq1 & Eq2.

$$X_{std} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

$$X_{scaled} = X_{std} \times (\max - \min) + \min \quad (2)$$

where max and min the target maximum and minimum values. The z feature, i.e., Impedance at every relay, is removed because it is not a numeric nor nominal, as noted by Weka tool

Exploratory Data Analysis (EDA) is performed to spot patterns and anomalies in the data. Descriptive statistics visualizations are very useful.

Measures from information theory are helpful such as Gini Impurity which specifies the feature probability to be incorrectly classified when selected at random, as shown in Eq3:

$$Gini = 1 - \sum_{i=1}^{n} (p_i)^2 \quad (3)$$

where p_i is feature probability to be classified for class i, and n is the number of classes.

XAI algorithm such as SHAP (SHapley Additive exPlanations) is tried. It is a game-theoretic method for explaining machine learning model output, by connecting optimal credit allocation with local explanations utilizing classic Shapley values from game theory [33]. Figure 5 shows SHAP in action.

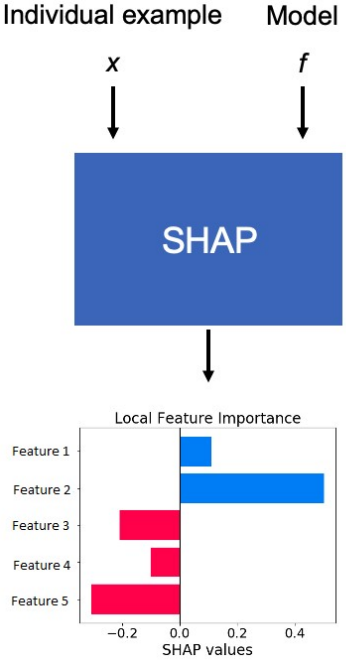

Figure 5: SHAP in action

Journal of Theoretical and Applied Information Technology
30th April 2023. Vol.101. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SHAP takes an example (denoted As X) and a model (denoted as F) and produces local feature importance of the features. It can be considered as a feature engineering step before applying machine learning algorithms.

In data science, an outlier is considered as a noise. It is an observation that is considered different from the remainder observations. Clustering approaches such as DBSCAN[34] and ROCK[35] can handle outliers. However, such algorithms aim at clustering whereas the outlier is considered as a noise. Outlier Detection algorithms such as Angle-based Outlier Detection (ABOD)[36] and Clustering Based Local Outlier Factor (CBLOF) [37] are tried.

ABOD has the algorithm listed in Algorithm 1. ABOD algorithm iterates over all the data points. In each iteration, the angle of every point that pivots from every other data pairs are calculated. These angles are stored. Then the variance of the AngleList is calculated. Variance values less than a threshold are potential anomalies The time ABOD takes is $O(n^3)$ which is very gross.

Step1: Iterate over all data points Ps,
Step 1.1 Calculate the angle a point pivots forms with all other data pairs
Step 1.2 Store angles in AngleList.
Step 2: Calculate variance of AngleList.
Step 3: Variance values less than a threshold are potential anomalies.

Algorithm 1: ABOD algorithm [36]

CBLOF has the algorithm listed in Algorithm 2. Initially, CBLOF applies the Squeezer Algorithm[38] to cluster the dataset. Then, LargeCluster and SmallCluster are calculated. Then, CBLOF iterates over the dataset. For each record in the dataset, if this record belongs to a cluster that belongs to SmallCluster, then the CBLOF= | Ci | * min(distance(r, Cj)), otherwise CBLOF= | Ci | * distance(r, Ci). The time CBLOF takes is O(n) which is very admissible.

Step 1: Cluster the dataset D using Squeezer Algorithm,
Step 2: Calculate LargeCluster and SmallCluster
Step 3: Iterate each record r in the dataset
Step 3.1 If r ∈ Ci and Ci ∈ SmallCluster
Step 3.1.1 CBLOF= | Ci | * min(distance(r, Cj))
Step 3.2 Else
Step 3.2.1 CBLOF= | Ci | * distance(r, Ci)

Algorithm 2: CBLOF algorithm [37]

The dataset is dichotomized into training part and testing part. Number of records in training data and testing data are 57658 and 14415 respectively. Training builds a model which is verified in the testing phase. Accuracy of any model is evaluated according to certain metrics such as accuracy, mean square error and the loss. Applied machine learning algorithms are Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), multi-layer perceptron (MLP), and Convolutional Neural Network (CNN).

RF is an ensemble of n decision trees, as depicted in Figure 6. The parameters of RF in Scikit-learn are: n_estimators =100, and criterion = "gini". RF split the training data into n parts, each of which is fed to n generated decision trees, denoted as DT-1 through DT-n. A voting step is performed on those decision trees by applying the average or the majority.

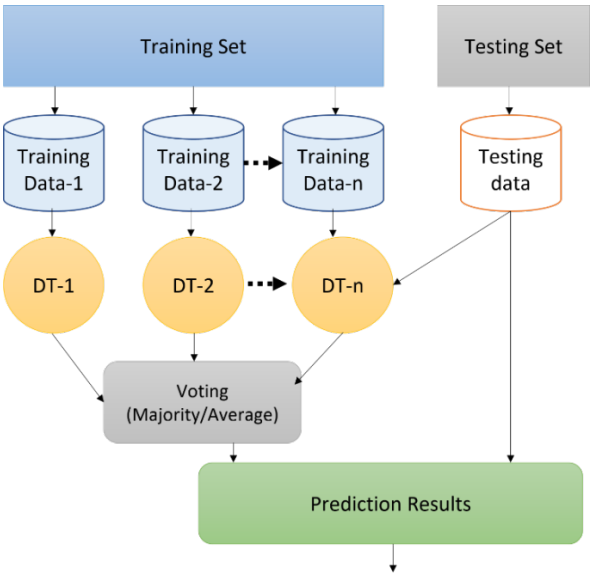

Figure 6 Random Forest [39]

SVM aims at creating a hyper-plane which is the best boundary or line that segregates n dimensional space into classes, as depicted in Figure 7. The hyper-plane that most accurately depicts the distance between the two classes is considered to be the best one.

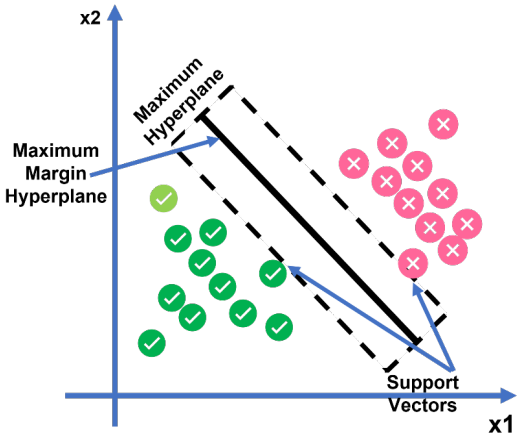

Figure 7: Support Vector Machine [40]

NB classifier is either Gaussian or multinomial. In the Gaussian case, probabilities follow Gaussian distribution, whereas in the multinomial case they are calculated the classical way. Gaussian case is used.

MLP follow the architecture of brains. It is a set of inter-connected neurons, decomposed into an input layer, hidden layers and an output layer. MLP parameters are: hidden layer sizes =20, activation function = "relu", optimizer = "adam", alpha = 10^{-4}, and learning rate = 10^{-3}.

For the CNN, it has the architecture shown in Figure 8. First, the input layer has the dimension of 128 which correspond to the features. Then there are two convolutional layers, followed by one max pooling layer. Then there are two convolutional layers, followed by one max pooling layer. Then there is a flatten layer, followed by a dense layer, followed by a dropout layer. Finally, there is the output dense layer.

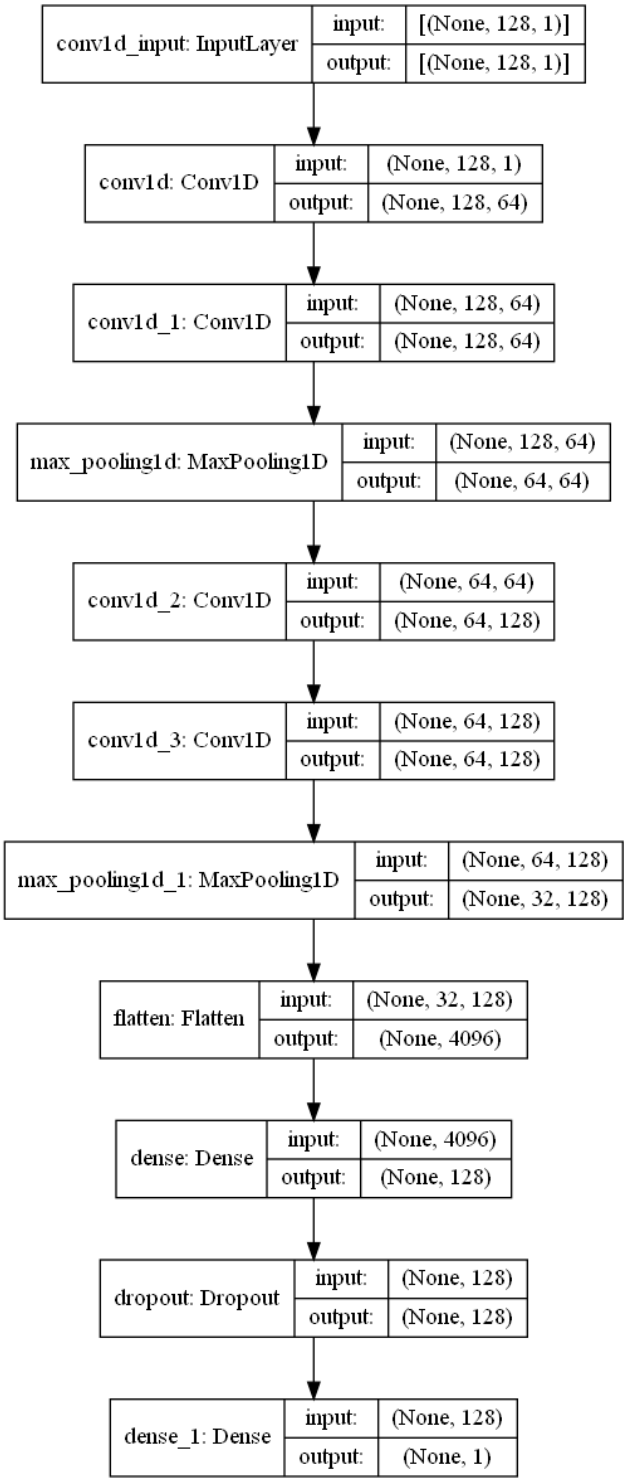

Figure 8: CNN architecture

4. RESULTS

To reproduce results in [8], Weka tool is used. It is an acronym for Waikato Environment for Knowledge Analysis and is built using Java. KazAnova Light, another tool that is built using Java, is used for EDA to gain a better understanding of the data set.

Python code is written for the rest of activities. Scikit-learn is the most used library. However, it does not support deep learning. This necessitates writing code for deep learning approaches using TensorFlow library

Journal of Theoretical and Applied Information Technology
30th April 2023. Vol.101. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 9 shows histogram with Kernel Density Estimation (KDE) for R4-PA10:IH. KDE is a non-parametric tool that performs kernel smoothing for estimating the probability density function (PDF) of random variables by considering kernels as weights.

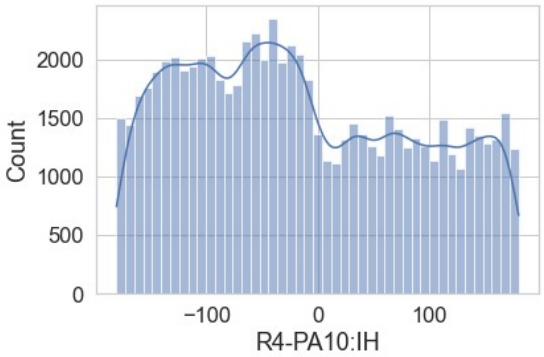

Figure 9: Histogram with KDE for R4-PA10:IH

Figure 10 shows Violin Plot for R4-PA10:IH. Violin plots hybridize box plot and KDE, showing data peaks and depicting summary statistics and the density of the random variable. The white dot denotes the median. The thick gray bar denotes the inter-quartile range. The thin gray line denotes the remaining of the distribution. On each side is a KDE to depict the shape of distribution of data.

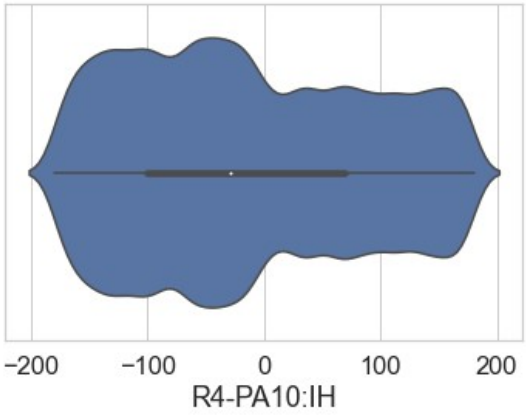

Figure 10: Violin Plot for R4-PA10:IH

The curve known as Receiver Operating Characteristic (ROC) reveals how well the model can differentiate between two classes. The ratio (1-specificity) will rise as the sensitivity does. We may calculate the True Negative Rate using Specificity, and the False Positive Rate using (1-Specificity). The degree to which the probabilities from the positive classes are distinguished from the negative classes is thus shown by the Area Under the Curve (AUC).

Gini coefficient is a tool to fix the Area under the Curve (AUC) to make it more meaningful. It has the range of values [-1, 1]. A perfect model has 1 while reversing model has a negative sign. AUC for R4_PA12_IH is shown in Figure 11, AUC for R4_PA10_IH is shown in Figure 12, and AUC for R4_PA_Z is shown in Figure 13.

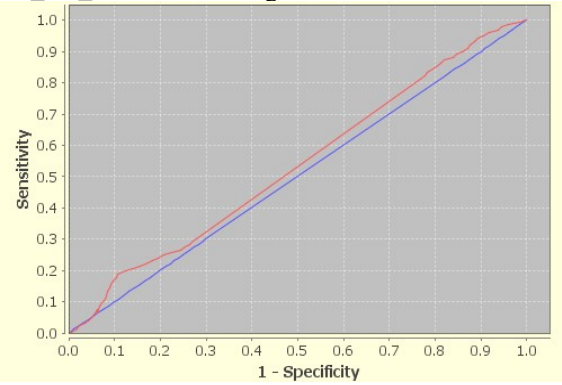

Figure 11: AUC for R4_PA12_IH

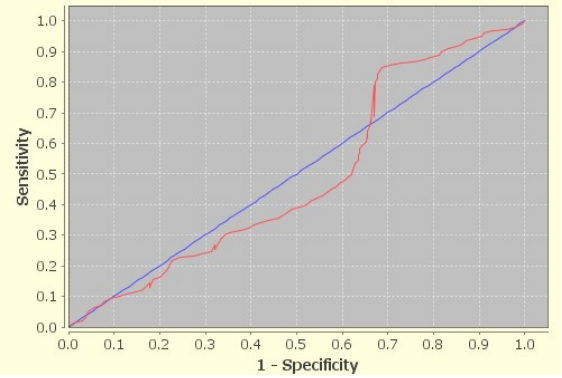

Figure 12: AUC for R4_PA10_IH

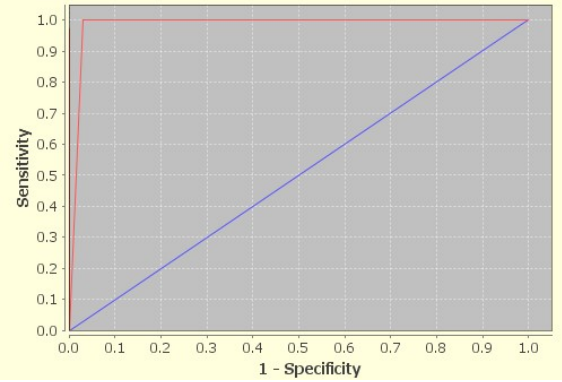

Figure 13: AUC for R4_PA_Z

Not all features contribute very much to the target. AUC for R4_PA12_IH indicates that it does not contribute too much to the target. AUC for R4_PA10_IH indicates that it gets "confused".

Journal of Theoretical and Applied Information Technology
30th April 2023. Vol.101. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

While AUC for R4_PA_Z indicates that it contributes too much to the target.

Figure 14 shows the Summary Plot, while Figure 15 shows the Decision Plot. Summary Plots provide a high-level overview of feature importance and the factors that influence it. It depicts the mean influence on model output.

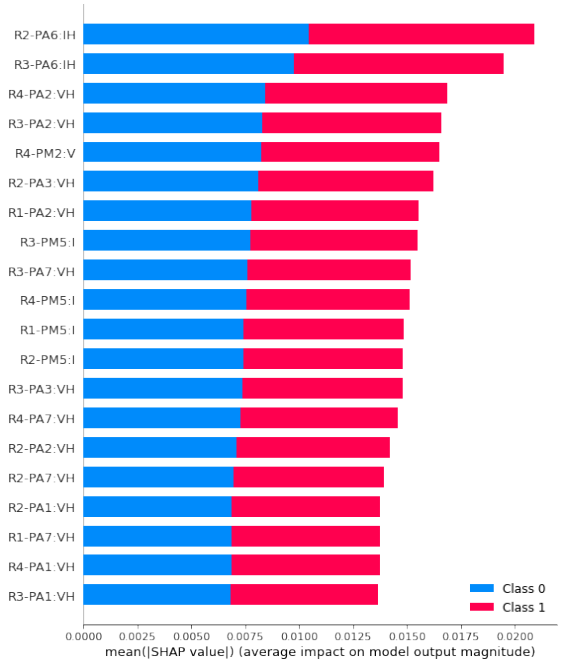

Figure 14:Summary plot of the proposed approach

In decision plot, X-axis denotes the output of the model. The plotting is centered on the x-axis at the expected value. Y-axis denotes the features of the model ordered by descending importance. Decision plots support hierarchical cluster feature ordering. Every prediction of an observation is depicted by a line with certain color on a spectrum. Atop the figure, all lines hit x-axis at their corresponding value of prediction. On the plot and moving bottom-up, SHAP values for all features are added to the base-value of the model.

Force Plot for all the features and for R1-PA-Z are shown in Figure 16 and Figure 17 respectively. Force plot, like feature importance, is an interactive tool for understanding the influences of all features on prediction of the models, based on game theory. Features are considered as players who are capable on constituting coalitions and playing games. Feature importance is the mean contribution to various coalitions. Base-value is the mean predicted probability across all samples. A red arrow depicts a feature that has negative influences on the prediction. Bold values are the actual predictions for such sample. A negative influence does not mean a worse effect but means the proximity or direction towards 1.

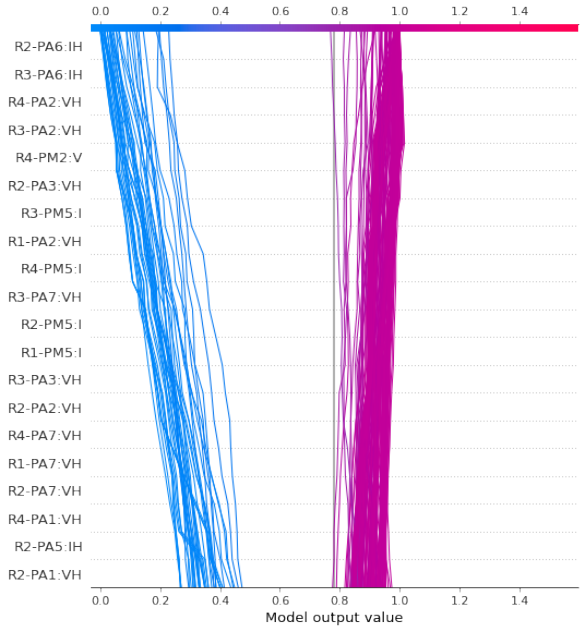

Figure 15: Decision plot of the proposed approach

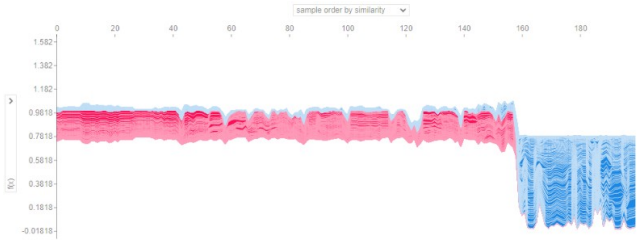

Figure 16 : Force plot of the proposed approach

Figure 17: Force plot for R1-PA-Z of the proposed approach

Both ABOD and CBLOF have mean squared error 0.33. Neither of them succeeds in distinguishing the two classes. This is partially because the data is very intertwined.

An attempt to distinguish the two classes using t-SNE [41] is used. It stands for t-distributed stochastic neighbor embedding. It is a statistical tool to visualize high dimensional data by giving every data-point a location in a 2D or 3D map. The t-SNE projection is shown in Figure 18.

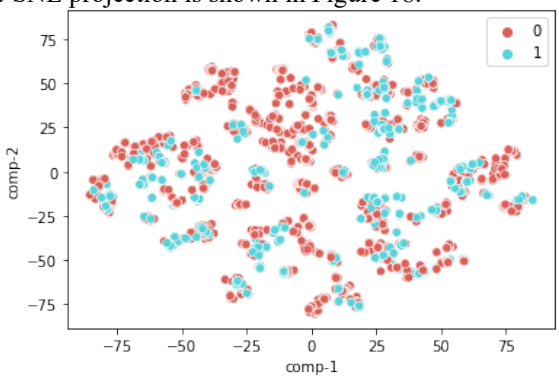

Figure 18: t-SNE projection

Figure 19 shows the accuracy of machine learning approaches. RF outperforms other approaches as it has 98% accuracy. The lowest accuracy is contributed by NB with 33%. Both SVM and CNN have modest accuracy around 73% and 71% respectively.

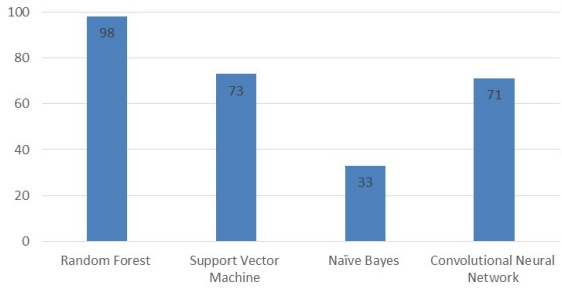

Figure 19: Accuracy of Machine Learning Approaches

5. CONCLUSION

In this work, an effective framework was constructed. For sake of comparing the proposed framework with other related work, we operate on the datasets without modifying them.

The framework merged the dataset files. Then preprocessing is performed to clean the data from Null entries. Normalization is then performed to ensure that data falls in a range between a minimum and a maximum value. EDA is performed to spot patterns and anomalies in the data using KDE and Violin plots. Not all features contribute very much to the target as indicated by the Gini index. SHAP, an XAI algorithm, is tried to generate summary plot, decision plot, and force plot. Outlier Detection algorithms ABOD and CBLOF are tried but neither of them succeeds in distinguishing the two classes because the data is very intertwined. This forced us to apply t-SNE projection. The dataset is decomposed into training part and testing part. Training builds a model which is verified in the testing phase. Applied machine learning algorithms are RF, SVM, NB, MLP, and CNN. Accuracy of the model is evaluated according to certain metrics. RF outperforms other approaches.

The dataset is imbalanced. One possible future work may consider dealing with this issue or using other techniques such as federated learning [42]. Another possible future direction may consider other datasets available at [13] such as Gas Pipeline and Water Storage Tank

REFERENCES:

[1] M. Bertolini, D. Mezzogori, M. Neroni, and F. Zammori, “Machine Learning for industrial applications: A comprehensive literature review,” Expert Systems with Applications, vol. 175, p. 114820, 2021.

[2] Z. Yu, Z. Kaplan, Q. Yan, and N. Zhang, “Security and privacy in the emerging cyber-physical world: A survey,” IEEE Communications Surveys & Tutorials, vol. 23, no. 3, pp. 1879–1919, 2021.

[3] S. Yuan and X. Wu, “Deep learning for insider threat detection: Review, challenges and opportunities,” Computers & Security, vol. 104, p. 102221, 2021.

[4] Q. Zhang, A. Z. Mohammed, Z. Wan, J.-H. Cho, and T. J. Moore, “Diversity-by-design for dependable and secure cyber-physical systems: A survey,” IEEE Transactions on Network and Service Management, vol. 19, no. 1, pp. 706–728, 2021.

[5] T. K. Das, S. Adepu, and J. Zhou, “Anomaly detection in industrial control systems using logical analysis of data,” Computers & Security, vol. 96, p. 101935, 2020.

[6] A. Chevrot, A. Vernotte, and B. Legeard, “CAE: Contextual auto-encoder for multivariate time-series anomaly detection in air transportation,” Computers & Security, vol. 116, p. 102652, 2022.

[7] A. Barbado, Ó. Corcho, and R. Benjamins, “Rule extraction in unsupervised anomaly detection for model explainability: Application to OneClass SVM,” Expert Systems with Applications, vol. 189, p. 116100, 2022.

[8] R. C. Borges Hink, J. M. Beaver, M. A. Buckner, T. Morris, U. Adhikari, and S. Pan, "Machine learning for power system disturbance and cyber-attack discrimination," in 2014 7th International Symposium on Resilient Control Systems (ISRCS), Aug. 2014, pp. 1–8. doi: 10.1109/ISRCS.2014.6900095.

[9] G. Ravikumar and M. Govindarasu, "Anomaly detection and mitigation for wide-area damping control using machine learning," IEEE Transactions on Smart Grid, 2020.

[10] V. Kumar and O. P. Sangwan, "Signature based intrusion detection system using SNORT," International Journal of Computer Applications & Information Technology, vol. 1, no. 3, pp. 35–41, 2012.

[11] F. Rafa, Z. Rahman, M. M. Mishu, M. Hasan, R. Rahman, and D. Nandi, "Detecting Intrusion in Cloud using Snort: An Application towards Cyber-Security," in Proceedings of the 2nd International Conference on Computing Advancements, 2022, pp. 199–206.

[12] H. Asad and I. Gashi, "Dynamical Analysis of Diversity in Rule-Based Open Source Network Intrusion Detection Systems," Empirical Software Engineering, vol. 27, no. 1, p. 4, 2022.

[13] T. Morris, "Industrial Control System (ICS) Cyber Attack Datasets," https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets Last Access September 1, 2022.

[14] G. Costa Silva, R. M. Palhares, and W. M. Caminhas, "A transitional view of immune inspired techniques for anomaly detection," in International Conference on Intelligent Data Engineering and Automated Learning, 2012, pp. 568–577.

[15] M. Ebrahimi, C. Y. Suen, O. Ormandjieva, and A. Krzyzak, "Recognizing predatory chat documents using semi-supervised anomaly detection," Electronic Imaging, vol. 2016, no. 17, pp. 1–9, 2016.

[16] L. Basora, X. Olive, and T. Dubot, "Recent advances in anomaly detection methods applied to aviation," Aerospace, vol. 6, no. 11, p. 117, 2019.

[17] Y. Luo, Y. Xiao, L. Cheng, G. Peng, and D. (Daphne) Yao, "Deep Learning-based Anomaly Detection in Cyber-physical Systems: Progress and Opportunities," ACM Comput. Surv., vol. 54, no. 5, p. 106:1-106:36, May 2021, doi: 10.1145/3453155.

[18] A. Jones, Z. Kong, and C. Belta, "Anomaly detection in cyber-physical systems: A formal methods approach," in 53rd IEEE Conference on Decision and Control, Dec. 2014, pp. 848–853. doi: 10.1109/CDC.2014.7039487.

[19] D. Shi, Z. Guo, K. H. Johansson, and L. Shi, "Causality Countermeasures for Anomaly Detection in Cyber-Physical Systems," IEEE Transactions on Automatic Control, vol. 63, no. 2, pp. 386–401, Feb. 2018, doi: 10.1109/TAC.2017.2714646.

[20] M. Keshk, E. Sitnikova, N. Moustafa, J. Hu, and I. Khalil, "An Integrated Framework for Privacy-Preserving Based Anomaly Detection for Cyber-Physical Systems," IEEE Transactions on Sustainable Computing, vol. 6, no. 1, pp. 66–79, Jan. 2021, doi: 10.1109/TSUSC.2019.2906657.

[21] P. Schneider and K. Böttinger, "High-Performance Unsupervised Anomaly Detection for Cyber-Physical System Networks," in Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy, New York, NY, USA, Jan. 2018, pp. 1–12. doi: 10.1145/3264888.3264890.

[22] W. Schneble and G. Thamilarasu, "Attack detection using federated learning in medical cyber-physical systems," in 28th International conference on computer communications and networks (icccn), 2019, pp. 1–8.

[23] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, "DeepFed: Federated Deep Learning for Intrusion Detection in Industrial Cyber–Physical Systems," IEEE Transactions on Industrial Informatics, vol. 17, no. 8, pp. 5615–5624, Aug. 2021, doi: 10.1109/TII.2020.3023430.

[24] Z. Zhang, L. Zhang, Q. Li, K. Wang, N. He, and T. Gao, "Privacy-enhanced momentum federated learning via differential privacy and chaotic system in industrial Cyber–Physical systems," ISA Transactions, Sep. 2021, doi: 10.1016/j.isatra.2021.09.007.

[25] S. Tuli, G. Casale, and N. R. Jennings, "TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data." arXiv, May 14, 2022. doi: 10.48550/arXiv.2201.07284.

[26] J. Paparrizos, Y. Kang, P. Boniol, R. S. Tsay, T. Palpanas, and M. J. Franklin, "TSB-UAD: an end-to-end benchmark suite for univariate

time-series anomaly detection," Proc. VLDB Endow., vol. 15, no. 8, pp. 1697–1711, Apr. 2022, doi: 10.14778/3529337.3529354.

[27] I. Ahmed, G. Jeon, and F. Piccialli, "From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, How, and Where," IEEE Transactions on Industrial Informatics, vol. 18, no. 8, pp. 5031–5042, Aug. 2022, doi: 10.1109/TII.2022.3146552.

[28] A. K. Nor, S. R. Pedapati, M. Muhammad, and V. Leiva, "Explainable artificial intelligence for anomaly detection and prognostic of gas turbines using uncertainty quantification with sensor-related data explainable artificial intelligence for anomaly detection and prognostic of gas turbines using uncertainty Q," no. September. DOI: https://doi. org/10.20944/preprints202109, vol. 34, p. v2, 2021.

[29] C. S. Wickramasinghe, K. Amarasinghe, D. L. Marino, C. Rieger, and M. Manic, "Explainable Unsupervised Machine Learning for Cyber-Physical Systems," IEEE Access, vol. 9, pp. 131824–131843, 2021, doi: 10.1109/ACCESS.2021.3112397.

[30] D. D. Le, V. Pham, H. N. Nguyen, and T. Dang, "Visualization and explainable machine learning for efficient manufacturing and system operations," 2019.

[31] M. Sayed-Mouchaweh, Explainable AI Within the Digital Transformation and Cyber Physical Systems. Springer, 2021.

[32] J. Souza and C. K. Leung, "Explainable Artificial Intelligence for Predictive Analytics on Customer Turnover: A User-Friendly Interface for Non-expert Users," in Explainable AI Within the Digital Transformation and Cyber Physical Systems, Springer, 2021, pp. 47–67.

[33] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in neural information processing systems. 2017;30.

[34] Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. Inkdd 1996 Aug 2 (Vol. 96, No. 34, pp. 226-231).

[35] Guha S, Rastogi R, Shim K. ROCK: A robust clustering algorithm for categorical attributes. Information systems. 2000 Jul 1;25(5):345-66.

[36] Kriegel HP, Schubert M, Zimek A. Angle-based outlier detection in high-dimensional data. InProceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining 2008 Aug 24 (pp. 444-452).

[37] Z. He, X. Xu, and S. Deng, "Discovering cluster-based local outliers," Pattern recognition letters, vol. 24, no. 9–10, pp. 1641–1650, 2003.

[38] He Z, Xu X, Deng S. Squeezer: an efficient algorithm for clustering categorical data. Journal of Computer Science and Technology. 2002 Sep;17(5):611-24.

[39] _____, "Random Forest Algorithm," https://www.javatpoint.com/machine-learning-random-forest-algorithm Last Access September 1, 2022.

[40] _____, "Support Vector Machine Algorithm https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm Last Access September 1, 2022," https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm Last Access September 1, 2022.

[41] Van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of machine learning research. 2008 Nov 1;9(11).

[42] Zhang C, Liu X, Zheng X, Li R, Liu H. Fenghuolun: A federated learning based edge computing platform for cyber-physical systems. In2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops) 2020 Mar 23 (pp. 1-4). IEEE.

Journal of Theoretical and Applied Information Technology
30th April 2023. Vol.101. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

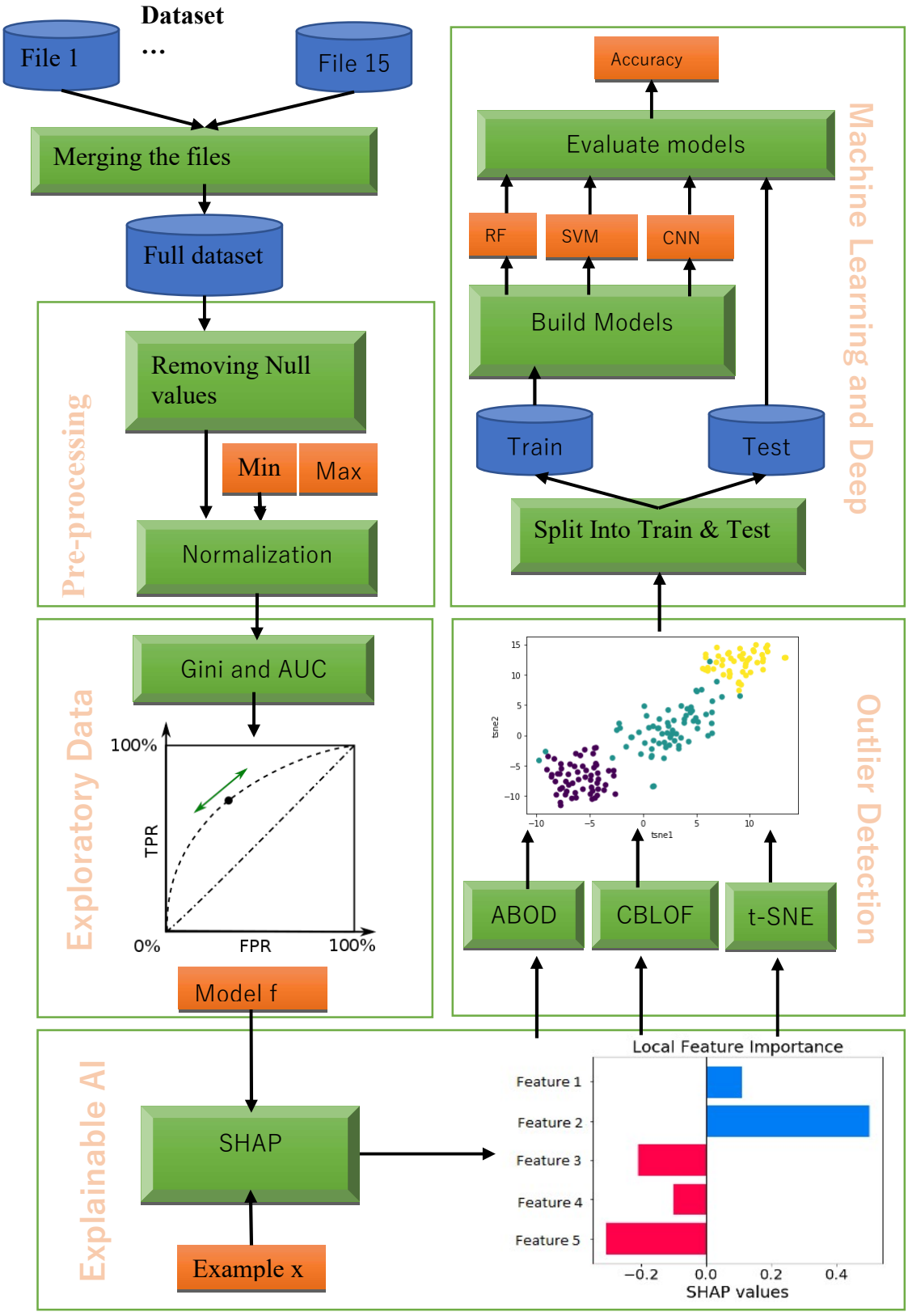

Figure 4: The Proposed Framework of the CPS anomaly detection