Advancing IoT Cybersecurity: Adaptive Threat Identification with Deep Learning in Cyber-Physical Systems

C. Atheeq
GITAM University, India
atheeq.prof@gmail.com (corresponding author)

Ruhiat Sultana
Lords Institute of Engineering and Technology, India
ruhiatsultana@lords.ac.in

Syeda Asfiya Sabahath
King Khalid University, Saudi Arabia
assyed@kku.edu.sa

Murtuza Ahmed Khan Mohammed
Universiti Teknologi Malaysia
ahmedkhan@graduate.utm.my

Received: 27 January 2024 | Revised: 16 February 2024 | Accepted: 18 February 2024

 DOI: https://doi.org/10.48084/etasr.6969

ABSTRACT

Securing Internet of Things (IoT)-enabled Cyber-Physical Systems (CPSs) can be challenging because security solutions intended for typical IT/OT systems may not be as effective in a CPS setting. The goal of this study is to create a mechanism for identifying and attributing two-level ensemble attacks that are specifically designed for use against Industrial Control Systems (ICSs). An original ensemble deep representation learning model is combined with decision tree algorithm to identify assaults on unbalanced ICS environments at the first level. An attack attribution network, which constitutes a collection of deep neural networks, is formed at the second level. The proposed model is tested using real-world datasets, notably those pertaining to water purification and gas pipelines. The results demonstrate that the proposed strategy outperforms other strategies with comparable computing complexity and that the recommended model outperforms the existing mechanisms.

Keywords-cyber-attacks; deep learning; threat detection; industrial control system; industrial IoT; cyber-physical systems

I. INTRODUCTION

Cyber-Physical Systems (CPSs) are increasingly fusing IoT-connected devices, even in areas of critical infrastructure like dams and power plants [1-7]. The Industrial Control System (ICS), which is in charge of ensuring that the infrastructure is operating efficiently, commonly includes connected devices to the Internet of Things (IoT), sometimes referred to as Industrial IoT or IIoT in certain contexts. ICSs may also refer to Distributed Control Systems (DCSs), Programmable Logic Controllers (PLCs), and Modbus protocols in addition to Supervisory Control And Data Acquisition (SCADA) systems. On the other side, linking IoT or ICS to public networks expands their attack surfaces and the likelihood that hackers will target them. One famous example is the 2010 Stuxnet attack, which reportedly caused serious damage to Iranian centrifuges used for nuclear enrichment. An assault against Iranian centrifuges happened in 2010. Another illustration of this is the Illinois pump attack that occurred in 2011 and led to the shutdown of the state's water treatment facility.

Physical behavior analysis and ensuring that system operations are always available require system-level security solutions. Unlike most IT and OT systems, which normally emphasize availability, integrity, and confidentiality, ICS security objectives are prioritized in the reverse order. ICS attacks have the potential to have detrimental effects. This highlights the value of having exceedingly strong security and safety processes in place to identify and prevent attacks

that target ICSs [8–10]. Attack identification and attribution methods based on anomalies and signatures are widely utilized. Hybrid-based detection and attribution systems have been created to solve the recognized limitations of signature-based and anomaly-based detection and attribution techniques. Hybrid-based approaches are effective at identifying anomalous behavior, but they are unreliable since they cause distinctive Intrusion Detection System (IDS) typologies due to frequent network upgrades. Hybrid-based approaches can identify strange conducts well, but are unreliable. Traditional methods for the detection and attribution of attacks heavily rely on the study of network metadata, including things like transmission ports, traffic volume, packet intervals, and IP addresses. Recently, there has been a revived interest in attack detection and attribution techniques based on Deep Neural Networks (DNNs) or Machine Learning (ML). Numerous cyber-physical systems have been created as a result of the IoT proliferation, allowing industrial equipment and IT processes to transmit and receive data over the internet [11-13]. These systems rely on sensors to detect the condition of equipment and report their findings to a centralized server using an internet connection. However, these sensors can be vulnerable to attacks from malicious actors, leading to false reports being sent to the centralized server and incorrect actions being taken. In response to this threat, researchers have developed algorithms to detect attacks on IoT-enabled CPSs. However, the data are usually unbalanced (e.g. normal records are usually significantly more than attack records). This imbalance can lead to inaccurate predictions and detection failures [14-17]. To address this issue, researchers have proposed a novel technique that engages an autoencoder and DNNs to detect and attribute assaults from the IoT on CPSs [18]. The technique consists of two modules. The first one involves training an autoencoder on an imbalanced dataset and utilizing the extracted features to train a decision tree algorithm to predict attack labels. In the second module, a DNN is trained on known and unknown attacks to identify attack labels or classes [19, 20]. To test the effectiveness of this technique, the Secure Water Treatment (SWaT) dataset was utilized, which contains IoT request and response signatures associated with unique attack labels. The dataset includes various types of cyber-attacks, such as command injection, water level manipulation, and flow rate tampering [21-23].

CPSs are made possible by the IoT. However, protecting IoT-enabled CPS might be challenging since security protocols designed for standard IT/OT systems might not work as well in a CPS setting [24-26]. This study aims to provide a framework for categorizing and assigning responsibility for two-level ensemble attacks created specifically for usage in an ICS and designed for CPS. In an original ensemble deep learning method to detect attacks on unbalanced ICS settings at the first level, a decision tree algorithm and a representation learning model are used together. At the second level, a group of DNNs called attack attribution network is created. Real-world datasets are applied to test the suggested model, notably those about water purification and gas pipelines. The results demonstrate that the proposed model outperforms rivals' tactics at a comparable degree of computer complexity.

According to the mechanism put out in [1], the increasing frequency of cyber-physical system attacks in recent years raises questions about the ICS cyber security. The majority of the current ICS cyber security efforts are based on firewalls, data diodes, and other intrusion prevention techniques. However, it is possible that these strategies will not be enough to guard against the rising number of cyber threats posed by determined attackers. A defense-in-depth-based cyber-attack detection system is currently being developed to raise the level of cyber security provided by ICS. Data from the host system, network traffic, and process parameter measurements will be used by this system. The latter also provides multiple levels of defense in order to buy the defenders some crucial extra time before the physical system is irreparably harmed.

The model developed in [2] is based on the Stealthy Attack against Redundant Controller Architecture. The controller is a crucial component that must be present in an Industrial Cyber-Physical System (ICPS) to guarantee dependability and stability. Several businesses employ the redundant controller architecture technique, including those that use DCS, SCADA, and other common ICPSs. Power production, chemical industry, water treatment, and other crucial industrial processes are under their strict observation and control. Given that some mechanical failures are unpredictable, redundant controller architecture has been created and, to a large part, implemented. This structure, which was first recommended for ensuring dependability and safety, nevertheless, has the potential to increase the surface area vulnerable to cyberattacks. As a result, there is a chance that a hostile entity could use its design as a cover for stealth strikes. The vulnerability caused by the redundant controller design is analyzed, and a combined attack methodology that may be utilized covertly against systems that employ the redundant controller architecture is provided.

Electric grids, water networks, and transport systems are all important metropolitan infrastructure that is frequently the target of cyber-attacks. A method for guaranteeing the cyber security of such critical urban infrastructure was devised in [4]. The network of linked objects that makes up these systems is referred to as the IIoT. An assault on key urban infrastructure IIoT would have a profoundly detrimental effect on society. SCADA systems are frequently used in IIoT control for critical urban infrastructure. Although it is vital to comprehend the cyber threat to the latter, there is currently no data-driven method for assessing the risk that SCADA software poses to IIoT devices. Using cosine similarity tests, this paper involves a comparison of SCADA systems to other types of control systems and it is found that SCADA, as a software subclass, carries particular risk characteristics for the IIoT. The widely held belief that the SCADA subclass of software is not vulnerable to attack is refuted based on the standard vulnerability score system risk criteria of exploitability and impact. In order to identify

SCADA risk metrics, a variety of statistical models were created. These models can be used to calculate the likelihood that a SCADA vulnerability will be exploited. Authors in [5] describe a method for the identification of data abnormalities. A developing trend in conventional industrial systems is the integration of the physical and cyber realms. The goal of this integration is to increase the adaptability and effectiveness of management, control, and supervision. The risk of security breaches has increased due to the deep integration of ICPSs. An essential component of the overall security protection offered is the initial protective barrier that attack detection generates. Contrarily, the majority of conventional systems focuses primarily on digital information and ignores any potential limitations that might result from the characteristics of the physical world. This study develops a zone partition-based strategy for the detection of anomalies in ICPSs. To ensure that important system states can be observed in several zones, an automatic zone division mechanism is created in the initial step of this process. Authors in [7] suggested an intrusion detection system similar to that in [6]. ICS relied on "air-gap" security measures until recently. As a result, every node of the ICS network was physically cut off from the Internet and all other networks. Businesses and professionals that use ICS networks benefit from connecting them to the internet. The protocols employed by ICSs, however, entail extremely few or no security measures because these systems were created for use in an environment with a high level of air-gapped security, which makes them vulnerable to a number of attacks. The described approach for detecting intrusions into network-attached ICSs employs network telemetry. With the aid of simulated PLC units, the newly developed IDS was able to distinguish between the machines used by an engineer and an attacker on the same network with an accuracy rate of 94.3% and between an engineer and an attacker on the Internet with a rate of 99.5%.

The literature review agrees that cyber-physical systems are becoming more risky, emphasizing the need for ICS cyber security. Most research focuses on redundant controller topologies and covert attacks or on SCADA systems' IIoT concerns in important urban infrastructures. Further research demonstrates that ICPSs' integration of cyber and physical realms requires increased anomaly detection and that the transition from air-gapped to network-attached systems demands unique intrusion detection to protect against sophisticated cyber-attacks.

II. THE PROPOSED SYSTEM

The increasing reliance on IoT-enabled CPSs, including operational IT and industrial machinery, has led to a growing concern about cyber-attacks targeting these systems. Attack detection and attribution in these systems can be challenging due to the imbalanced nature of the training data, where one class may contain significantly more samples than the other. Traditional machine learning algorithms often struggle to accurately detect attacks in such imbalanced datasets. This research study presents an original deep learning-based assault detection and attribution approach using a two-stage ensemble methodology that has been specifically created for imbalanced ICS data.

Traditional sampling methods, such as over- and under-sampling, might be ineffective when dealing with imbalanced datasets. In order to enable the DNN to handle unbalanced datasets, this research aims to introduce a novel deep representation learning approach that does not alter, produce, or remove any samples. Although the suggested framework has a complex design, the amount of training samples required (n) is comparable to that of other known DNN-based techniques. To evaluate the proposed framework's effectiveness, we used the SWaT dataset, which contains IoT request and response signatures associated with unique attack labels. The proposed framework demonstrated better recall and f-measure in detecting and attributing attacks than those exhibited by previous works.

The proliferation of data volume has increased as a result of IoT technologies in CPSs, including industrial machinery and operational ITs. These systems use sensors to monitor the condition of equipment, which sends data to a centralized server using internet connections. However, there is a risk of malicious users hacking into these sensors and altering the transmitted data, which could result in false actions being taken. To address this problem, attack detection algorithms have been developed. However, these algorithms frequently are faced with imbalanced datasets that may lead to inaccurate predictions.

The first part of the proposed technique involves implementing an autoencoder deep learning algorithm to extract features from an imbalanced dataset. The autoencoder is trained on the dataset, and then the extracted features are utilized to train a decision tree algorithm to predict whether an attack is labeled as known or unknown. Principal Component Analysis (PCA)'s reduced set of features is applied to train the decision tree. By using an autoencoder, we are able to extract meaningful features from the imbalanced dataset, which can then be employed to train a more accurate decision tree algorithm. The second component of the proposed technique puts into service a DNN algorithm to train on both known and unidentified threats. The DNN will recognize the attack label or class and attribute it if a record includes an attack signature. The ability to precisely identify both known and unidentified attacks using a DNN is crucial for spotting and avoiding cyber-attacks on CPS with IoT support. For the goal of locating and determining who is responsible for cyberattacks in IoT-enabled CPSs, the proposed technique has several advantages over the existing techniques. First, it avoids under- or over-sampling methods, which might produce incorrect predictions. Second, the proposed method enhances the decision tree algorithm's accuracy by employing an autoencoder to extract useful features from unbalanced datasets. Also, the proposed method utilizes a DNN to precisely identify both known and unidentified attacks, which is essential for spotting and avoiding cyberattacks in IoT-enabled CPSs. For imbalanced ICS data, assault detection and attack attribution is conducted adopting a

unique two-stage ensemble deep learning approach. An autoencoder and a DNN are utilized to accurately detect known and unidentified assaults and extract features, respectively. The SWaT dataset was used to test the suggested method, which performed better than other known methods that employed under- or over-sampling strategies. The proposed method could be utilized to identify and stop cyber-attacks in IoT-enabled CPSs due to its advantages over the current techniques.

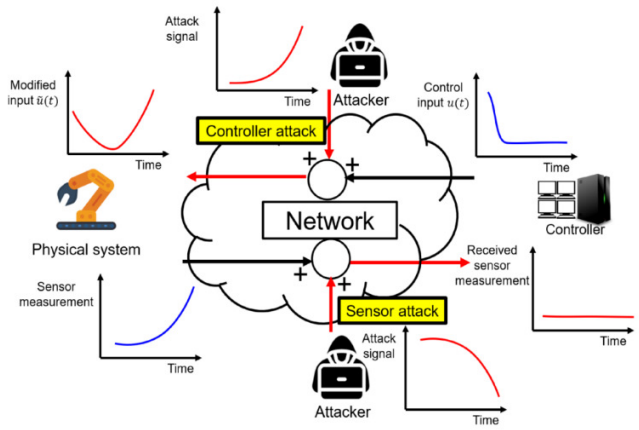

Fig. 1. The proposed architecture.

A. *Algorithm: Two-Stage Ensemble Deep Learning for IoT Cyber-Attack Detection and Attribution (2SDL-CA)*

1) *Step 1: Data Collection*

Collect data from sensors and define the dataset D as a set of records:

$$D = \{(x_i, y_i) | x_i \in R^n, y_i \in \{0,1\}\} \quad (1)$$

where x_i represents a data vector of dimension n and y_i represents labels (0 for normal, 1 for attack).

2) *Step 2: Data Preprocessing*

Split the dataset D into D_{normal} (normal records) and D_{attack} (attack records):

$$D_{normal} = \{(x_i, y_i) | (x_i, y_i) \in D, y_i = 0\} \quad (2)$$

$$D_{attack} = \{(x_i, y_i) | (x_i, y_i) \in D, y_i = 1\} \quad (3)$$

3) *Step 3: Autoencoder Feature Extraction*

Utilize an autoencoder deep learning algorithm to extract features from the imbalanced dataset D. The autoencoder is trained on the dataset, and the extracted features are used as D_{PCA} to train a decision tree: $D_{PCA} = Autoencoder(D)$.

4) *Step 4: Decision Tree Training*

Train a decision tree algorithm (*DT*) employing the reduced features obtained from PCA:

$$DT(D_{PCA}) = argmind\left(\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij} - \mu_j)^2\right) \quad (4)$$

where D_{PCA} contains the reduced features z_i obtained through PCA, n is the number of features, m is the number of data points, x_{ij} is the value of feature i for data point j, and μ_j is the mean value of feature i for all data points.

5) *Step 5: Decision Tree Prediction*

Use the trained *DT* to predict whether an attack is known or unknown for new data points.

6) *Step 6: Deep Neural Network (DNN) Training*

Train a DNN model (*DNN)* on both known and unknown threats:

$$DNN(D) = argmin_\theta\left(\sum_{i=1}^{m} L(y_i, f_\theta(x_i))\right) \quad (5)$$

where D represents the entire dataset, m is the number of data points, θ represents the model parameters, L is the loss function, y_i is the true label, and $f_\theta(x_i)$ is the predicted label for data point i.

7) *Step 7: DNN Prediction*

Use the trained *DNN* to identify the attack label or class and attribute it when a record includes an attack signature.

8) *Step 8: Evaluation and Testing*

Evaluate the proposed method's performance utilizing datasets. Performance metrics include precision (*P*), recall (*R*), F1-score (*F1*), and accuracy (*ACC*), which are computed based on the true labels y_i. F1 is the harmonic mean of *P* and *R*, providing a balance between the two metrics.

$$P = \frac{TP}{TP+FP} \quad (6)$$

$$R = \frac{TP}{TP+FN} \quad (7)$$

$$F1 = \frac{2.P.R}{P+R} \quad (8)$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

where *TP* (True Positives) represents the number of correctly identified attacks, *FP* (False Positives) represents the number of normal instances incorrectly identified as attacks, *FN* (False Negatives) represents the number of actual attacks incorrectly classified as normal, and TN (True Negative) represents the correctly identified normal instances. The DNN Prediction Probability (P_{pred}) is:

$$P_{pred}(y_i = 1) = sigmoid\left(\sum_{j=1}^{n} w_j, x_{ij} + b\right) \quad (10)$$

where n is the number of features, w_j symbolizes the weights associated with each feature, x_{ij} represents the j-th feature of data point I, b is the bias term, and sigmoid(·) is the sigmoid activation function.

DNN Model Loss (L) is defined by:

$$L(y_i, f_\theta(x_i)) = -\left(y_i \log\left(P_{pred}(y_i)\right) + \left((1 - y_i) \log\left(1 - P_{pred}(y_i)\right)\right)\right) \quad (11)$$

The loss function L measures the dissimilarity between true labels y_i and predicted probabilities $P_{pred}(y_i)$ for each data point. It is commonly used in binary classification problems. These complex formulas help assess the performance of the model through various metrics and describe the probability prediction generated by the DNN during testing. Partially connected components, inspired by dropout techniques, introduce random dropout of neurons during training. The fully connected layers integrate the high-level features extracted by the autoencoder. The output y_i of a neuron in the fully connected layer is calculated adopting the parametric rectified linear unit (PReLU) activation function. The mathematical formulation for dropout during training is given by:

$$x_{out} = x_{in} \odot \left(Mask. \frac{1}{1-dropout_{rate}}\right) \quad (12)$$

$$Mask \sim Bernoulli(1 - dropout_{rate})$$

$$y_i = PReLU\left(\sum_j w_{ij}.x_j + b_i\right) \quad (13)$$

$$PReLU(x) = max(0,x) + a_i.min(0,x) \quad (14)$$

$$a_i \sim Uniform(0.01,0.1)$$

where $\odot$ represents element wise multiplication, w_{ij} represents the weight between the i-th neuron and the j-th input, x_j is the j-th input, b_i is the bias term for the i-th neuron, and a_i is a learnable parameter.

B. Attack Attribution Mechanism:

The DNN is trained to recognize and attribute cyber-attacks based on learned features. Let X represent the input features, Y denote the true labels, and Y' signify the predicted labels. The training process involves minimizing a novel attack attribution loss function:

$$L_{attr}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} Y_{ic}.log(Y'_{ic}).(1 - Attribution_{ic}) \quad (15)$$

where $Attribution_{ic}$ is a binary indicator (0 or 1) specifying whether an instance is attributed to a known attack class. The ensemble DNN excels in handling both known and unknown threats. By employing a one-vs-all approach, individual classifiers within the ensemble are specialized for different attack attributes. The final prediction is determined through a weighted combination of these individual predictions, promoting a comprehensive and nuanced attribution process.

$$Final\ Prediction = softmax(\sum_i w_i.Classifier_i(X)) \quad (16)$$

where w_i represents the weight associated with the i-th classifier $Classifier_i$.

C. Model Robustness

Ensuring the robustness of the proposed model is crucial for practical deployment in ICS environments. To enhance model robustness, adversarial training techniques, such as the Fast Gradient Sign Method (FGSM), were incorporated during the training phase. Additionally, to improve model robustness, integrate methods like Layer-wise Relevance Propagation (LRP) were utilized to highlight important features contributing to the final decision. To strengthen the model's defenses against possible attacks, hostile instances are added to the training dataset through the process of adversarial training. Generating adversarial instances is a common application of the Fast Gradient Sign Method (FGSM). The definition of the adversarial loss function (L_{adv}) is:

$$L_{adv}(\theta) = \frac{1}{N}\sum_{I=0}^{N} max\left(0, \| \theta - \epsilon.sign\left(\nabla_\theta J(\theta, x_i, y_i)\right)\|_2 - \|\theta\|_2\right) \quad (17)$$

where θ represents the model parameters, N is the number of training samples, ϵ controls the magnitude of the perturbation, $J(\theta, x_i, y_i)$ is the model's loss function for the input x_i and true label y_i, $\nabla_\theta J(\theta, x_i, y_i)$ is the gradient of the loss with respect to the model parameters.

LRP is a technique used to attribute the model's decision to input features, providing insight into feature importance. The relevance (R_i) for each input feature i is calculated applying the LRP formula. This recursive formula propagates relevance from the output layer to the input layer, highlighting the contribution of each feature to the final decision. To optimize model robustness, an integrated objective ($L_{integrated}$) can be defined as a combination of the adversarial loss and the LRP-based loss:

$$R_i = \sum_j \frac{a_{ij}.R_j}{\sum_k a_{ik}} \quad (18)$$

$$L_{integrated}(\theta) = \alpha.L_{adv}(\theta) - \beta.L_{expl}(\theta) \quad (19)$$

where α and β are hyperparameters controlling the trade-off for adversarial training, $L_{expl}(\theta)$ is the explainability loss derived from LRP, a_{ij} represents the connection weight between the i-th input and the j-th neuron in the preceding layer, and R_j is the relevance of the j-th neuron in the preceding layer.

D. Computational Efficiency Considerations

Efficiency is a critical factor for the practical deployment of the proposed ensemble DNN in large-scale ICS environments. Focus is placed on two aspects: model size reduction through weight pruning and knowledge distillation, and inference time reduction through model quantization. To reduce the model size by eliminating less significant weights, the pruning mask M_{ij} is applied to the weight w_{ij}. To transfer knowledge from the ensemble DNN ($DNN_{esemble}$) to a smaller, distilled DNN ($DNN_{distill}$), the distillation loss is minimized during training:

$$w_{ij}^{pruned} = w_{ij}.M_{ij} \quad (20)$$

$$L_{distill} = \frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} T_c.\log\left(\sigma\left(\frac{Y_{ic}}{T}\right)\right) \quad (21)$$

where M_{ij} is determined based on a predefined threshold, T_c is the softened target distribution from the ensemble model, σ is the softmax function, and T is the temperature parameter.

Quantization reduces the precision of weights and activations, decreasing the memory footprint and speeding

up inference. The quantized weight $w_{ij}^{quantized}$ is given in (22). The predicted inference time t_{pred} can be estimated using a linear regression model:

$$w_{ij}^{quantized} = Round\left(\frac{w_{ij}}{\Delta}\right).\Delta \quad (22)$$

$$t_{pred} = \alpha.Size_{model} + \beta.Size_{input} + \gamma \quad (23)$$

where Δ is the quantization step size, α, β, and γ are the regression coefficients, and $Size_{model}$ and $Size_{input}$ represent the sizes of the model and input data, respectively.

III. RESULTS AND DISCUSSION

To implement the proposed technique, the SWaT dataset, which contains IoT request and response signatures, associated with unique attack labels was utilized. The dataset includes various types of cyber-attack labels, such as replay, denial-of-service, and false data injection attacks. The proposed technique was evaluated based on its ability to use the SWaT dataset to identify and attribute cyber-attacks. The findings demonstrate that the proposed method can reliably identify and attribute cyber-attacks even in the presence of imbalanced data. The proposed technique outperformed existing algorithms that used under- or over-sampling techniques, indicating its effectiveness.

In Table I, Test Case ID is a unique identifier for each test case conducted in this study. Test Case Name provides a brief description of what is being tested each time. Test Case Description provides further details on the test case and its purpose. Test Step field outlines the steps taken to complete each test case. Test Case Status describes whether the test case was successful or not. Test Priority indicates the level of importance assigned to each test case. The first test case, Test Case ID 01, aimed to test whether the Personality Dataset was successfully uploaded into the system. Further operations could not be conducted until the dataset was successfully uploaded. This test case was assigned a high priority due to its importance in the study. The second test case, Test Case ID 02, aimed to test whether the preprocessed dataset was successfully uploaded into the system, which was the expected result. In this case, further operations could not be conducted until the dataset's successful upload was achieved. This test case was also assigned a high priority due to its importance in the study.

TABLE I. ACTIONS ON THE DATASET FOR THE PROPOSED MODEL

Test Case Id	Test Case Name	Test Case Description	Test Steps			Test Case Status	Test Priority
			Step	Negative	Positive		
01	Upload SWaT dataset	Check if the dataset has been uploaded to the system.	Perhaps the dataset was not uploaded.	Any further operations cannot be performed.	Dataset uploaded. Further operations will be done.	High	High
02	Upload preprocessed dataset	Check if the preprocessed dataset has been uploaded to the system.	Perhaps the dataset was not uploaded.	Any further operations cannot be performed.	Dataset uploaded. Further operations will be done.	High	High
03	Run auto encoder algorithm	Test whether auto encoder algorithm was run successfully or not.	Is the encoder sent.	The algorithm cannot be run.	Further operations can be conducted.	High	High
04	Run decision tree with PCA	Test whether decision tree was run successfully or not.	PCA.	Extended PCA cannot be performed.	Further operations can be conducted.	High	High
05	Run DNN algorithm	Verify run rule biased attack detector.	Detection.	DNN cannot be run.	Further operations can be conducted.	High	High
06	Detection and attack attribute type	Verify if the attack detecion graph was successfull or not.	Attribute detection.	Attribute detection cannot be done.	Further operations can be conducted.	High	high
07	Comparison graph	Verify run rule biased attack detector.	Without comparison.	The graph cannot be done.	Further operations can be conducted.	High	High
08	Comparison table	Verify if the attack detecion graph was successfull or not.	Without comparison table.	The table cannot be done.	Further operations can be conducted.	High	high

Test Case ID 03 tested the auto encoder algorithm success. The method was anticipated to run, but if the encoder was not supplied, it may not have achieved this target. Without fixing the algorithm, no more operations could be done. High priority was given to this test case also. All the test cases were completed and the findings were reported. The dataset is shown in Figure 2 as a graph with the name of the attack on the x-axis and the number of times that it was detected in the dataset on the y-axis. As it can be observed, the attacks have fewer records than the Normal class, which results in data imbalance. This issue can be corrected using DNN, Autoencoder, or Decision Tree. The dataset was then preprocessed to fill in any missing values, and then the option "Normalize Values Using MIN-MAX" was chosen to normalize the numbers. The result section of the current research presents the outcomes of the proposed methodology. The dataset was first normalized to convert all values between 0 and 1, and the total number of records in the dataset, along with the train and test split counts, were displayed on the screen. The Autoencoder algorithm was trained on the dataset, resulting in 90% accuracy. To further enhance accuracy, the Decision Tree algorithm was implemented with PCA, and the accuracy and precision values were improved.

The graph in Figure 3 depicts the performance over time for the proposed and the existing method. The proposed method's performance exhibits an overall upward trend with some

fluctuations throughout the observation period, suggesting an increase in performance or output over time. On the other hand, the existing method also shows fluctuations but with a general trend that seems to plateau or increase less steeply compared to the proposed method. This representation implies that while both methods improve over time, the proposed method might offer a more pronounced enhancement in performance.

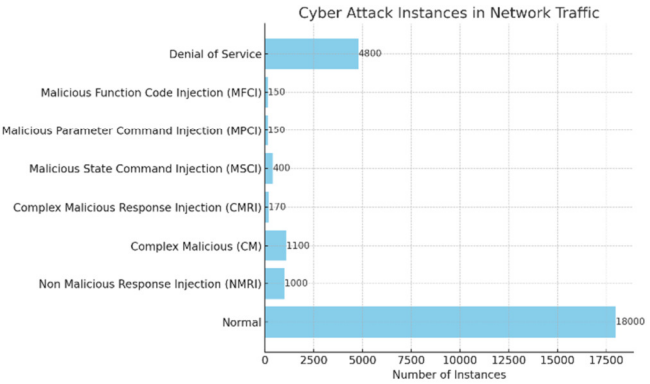

Fig. 2. Attack instance count.

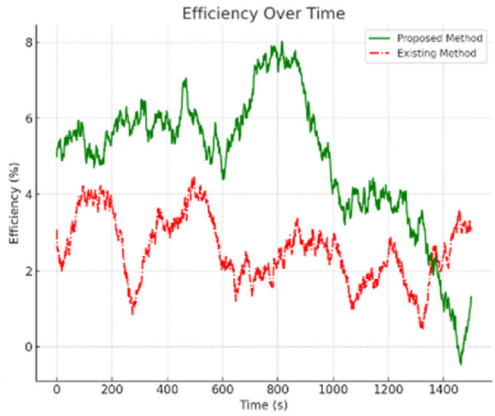

Fig. 3. Effeciency comparison.

In Figure 4, the efficiency of both methods is assessed over the same time span. This graph uses a percentage scale on the y-axis to measure efficiency, which might correspond to the effectiveness or resource utilization of the methods. The proposed method displays a dynamic pattern, with its efficiency percentage rising and falling sharply, indicating significant variability in its efficiency. The existing method, follows a similar trend with less pronounced peaks and troughs, which may suggest a more stable but less efficient performance. The comparison reflects that the proposed method has moments of high efficacy, possibly outperforming the existing method, but also periods where its effectiveness crucially drops. The DNN algorithm was then applied to the dataset, which resulted in 99% accuracy. The detection and attribute of attack types were performed on the uploaded test data. Various types of attacks were spotted, and the results were displayed in a text area. Overall, the proposed methodology proved to be effective in identifying and attributing attack types on the dataset. The normalization of the data, followed by the application of Autoencoder, Decision Tree with PCA, and DNN algorithms, resulted in high accuracy and precision values. The Comparison graph provides a visual representation of the encountered attacks, which could be useful in analyzing and understanding the data. Figure 5 and Table II display the precision, recall, accuracy, and F1 score for each considered algorithm.

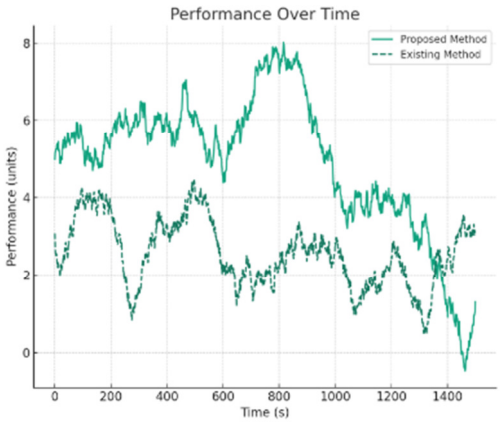

Fig. 4. Performance comparison.

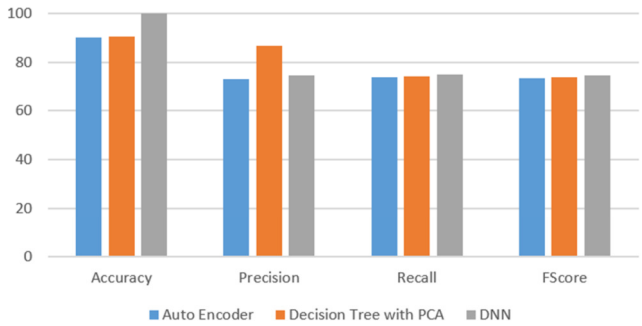

Fig 5. Algorithm performance comparison.

IV. CONCLUSION

Two-stage ensemble deep learning was used in this paper to offer a unique assault detection and attribution methodology for imbalanced ICS data. Deep representation learning converts the data to a higher-dimensional space for attack detection and Decision Tree was utilized to find plausible assault samples. This level can identify new assaults and survive skewed datasets. Multiple one-vs-all classifiers trained on different assault features are merged during attack attribution. The model is a complicated DNN with partially linked and completely connected components that can appropriately attribute cyber-attacks. As with earlier DNN-based methodologies, the proposed framework's training and testing phases are computationally intensive. This applies despite the framework's complex architecture. The proposed framework has greater recall and f-measure values than previous efforts and can swiftly recognize and attribute samples.

TABLE II. PERFORMANCE COMPARISON OF THE PROPOSED MODEL WITH EXISTING METHODS

Algorithm	Accuracy	Precision	Recall	F1 score
Auto encoder	90.091911764	73.168251465	74.120354616	73.598642216
DT withPCA	90.459558826	86.654346153	74.235468468	73.895321646
DNN	99.963512465	74.562458621	75.246864126	74.774652165

REFERENCES

[1] F. Zhang, H. A. D. E. Kodituwakku, J. W. Hines, and J. Coble, "Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network, System, and Process Data," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4362–4369, Jul. 2019, https://doi.org/10.1109/TII.2019.2891261.

[2] R. Ma, P. Cheng, Z. Zhang, W. Liu, Q. Wang, and Q. Wei, "Stealthy Attack Against Redundant Controller Architecture of Industrial Cyber-Physical System," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9783–9793, Sep. 2019, https://doi.org/10.1109/JIOT.2019.2931349.

[3] E. Nakashima, "Foreign hackers targeted U.S. water plant in apparent malicious cyber attack, expert says," *Washington Post*, Jun. 30, 2023. https://www.washingtonpost.com/blogs/checkpoint-washington/post/foreign-hackers-broke-into-illinois-water-plant-control-system-industry-expert-says/2011/11/18/gIQAgmTZYN_blog.html.

[4] G. Falco, C. Caldera, and H. Shrobe, "IIoT Cybersecurity Risk Modeling for SCADA Systems," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4486–4495, Dec. 2018, https://doi.org/10.1109/JIOT.2018.2822842.

[5] J. Yang, C. Zhou, S. Yang, H. Xu, and B. Hu, "Anomaly Detection Based on Zone Partition for Security Protection of Industrial Cyber-Physical Systems," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4257–4267, May 2018, https://doi.org/10.1109/TIE.2017.2772190.

[6] S. A. Alshaya, "IoT Device Identification and Cybersecurity: Advancements, Challenges, and an LSTM-MLP Solution," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 11992–12000, Dec. 2023, https://doi.org/10.48084/etasr.6295.

[7] S. Ponomarev and T. Atkison, "Industrial Control System Network Intrusion Detection by Telemetry Analysis," *IEEE Transactions on Dependable and Secure Computing*, vol. 13, no. 2, pp. 252–260, Mar. 2016, https://doi.org/10.1109/TDSC.2015.2443793.

[8] J. F. Clemente, "No cyber security for critical energy infrastructure," Ph.D. dissertation, Naval Postgraduate School, Monterey, CA, USA, 2018.

[9] C. Bellinger, S. Sharma, and N. Japkowicz, "One-Class versus Binary Classification: Which and When?," in *11th International Conference on Machine Learning and Applications*, Boca Raton, FL, USA, Dec. 2012, vol. 2, pp. 102–106, https://doi.org/10.1109/ICMLA.2012.212.

[10] M. A. Lateef, C. Atheeq, M. A. Rahman, and M. A. Faizan, "Data Aegis Using Chebyshev Chaotic Map-Based Key Authentication Protocol," in *Intelligent Manufacturing and Energy Sustainability*, A. R. Manchuri, D. Marla, and V. V. Rao, Eds. New York, NY, USA: Springer, 2023, pp. 187–195.

[11] M. M. N. Aboelwafa, K. G. Seddik, M. H. Eldefrawy, Y. Gadallah, and M. Gidlund, "A Machine-Learning-Based Technique for False Data Injection Attacks Detection in Industrial IoT," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8462–8471, Sep. 2020, https://doi.org/10.1109/JIOT.2020.2991693.

[12] W. Yan, L. K. Mestha, and M. Abbaszadeh, "Attack Detection for Securing Cyber Physical Systems," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8471–8481, Oct. 2019, https://doi.org/10.1109/JIOT.2019.2919635.

[13] M. A. Alqarni and S. H. Chauhdary, "A Security Scheme for Statistical Anomaly Detection and the Mitigation of Rank Attacks in RPL Networks (IoT Environment)," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12409–12414, Dec. 2023, https://doi.org/10.48084/etasr.6433.

[14] T. K. Das, S. Adepu, and J. Zhou, "Anomaly detection in Industrial Control Systems using Logical Analysis of Data," *Computers & Security*, vol. 96, Sep. 2020, Art. no. 101935, https://doi.org/10.1016/j.cose.2020.101935.

[15] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013, https://doi.org/10.1109/TPAMI.2013.50.

[16] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, "Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6822–6834, Aug. 2019, https://doi.org/10.1109/JIOT.2019.2912022.

[17] I. A. Khan, D. Pi, Z. U. Khan, Y. Hussain, and A. Nawaz, "HML-IDS: A Hybrid-Multilevel Anomaly Prediction Approach for Intrusion Detection in SCADA Systems," *IEEE Access*, vol. 7, pp. 89507–89521, 2019, https://doi.org/10.1109/ACCESS.2019.2925838.

[18] C. Atheeq and M. M. A. Rabbani, "Mutually authenticated key agreement protocol based on chaos theory in integration of internet and MANET," *International Journal of Computer Applications in Technology*, vol. 56, no. 4, pp. 309–318, Jan. 2017, https://doi.org/10.1504/IJCAT.2017.089088.

[19] R. Alsulami, B. Alqarni, R. Alshomrani, F. Mashat, and T. Gazdar, "IoT Protocol-Enabled IDS based on Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12373–12380, Dec. 2023, https://doi.org/10.48084/etasr.6421.

[20] J. J. Q. Yu, Y. Hou, and V. O. K. Li, "Online False Data Injection Attack Detection With Wavelet Transform and Deep Neural Networks," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3271–3280, Jul. 2018, https://doi.org/10.1109/TII.2018.2825243.

[21] A. Cook, A. Nicholson, H. Janicke, L. Maglaras, and R. Smith, "Attribution of Cyber Attacks on Industrial Control Systems," *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 3, no. 7, Apr. 2016, Art. no. e3, https://doi.org/10.4108/eai.21-4-2016.151158.

[22] N. A. Alsharif, S. Mishra, and M. Alshehri, "IDS in IoT using Machine Learning and Blockchain," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11197–11203, Aug. 2023, https://doi.org/10.48084/etasr.5992.

[23] L. Maglaras, M. A. Ferrag, A. Derhab, M. Mukherjee, H. Janicke, and S. Rallis, "Threats, Countermeasures and Attribution of Cyber Attacks on Critical Infrastructures," *EAI Endorsed Transactions on Security and Safety*, vol. 5, no. 16, Oct. 2018, Art. no. e1, https://doi.org/10.4108/eai.15-10-2018.155856.

[24] C. Atheeq and M. M. A. Rabbani, "CACK—A Counter Based Authenticated ACK to Mitigate Misbehaving Nodes from MANETs," *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, vol. 14, no. 3, pp. 837–847, Apr. 2021, https://doi.org/10.2174/2213275912666190809104054.

[25] M. Alaeiyan, A. Dehghantanha, T. Dargahi, M. Conti, and S. Parsa, "A Multilabel Fuzzy Relevance Clustering System for Malware Attack Attribution in the Edge Layer of Cyber-Physical Networks," *ACM Transactions on Cyber-Physical Systems*, vol. 4, no. 3, Nov. 2020, Art. no. 31, https://doi.org/10.1145/3351881.

[26] U. Noor, Z. Anwar, T. Amjad, and K.-K. R. Choo, "A machine learning-based FinTech cyber threat attribution framework using high-level indicators of compromise," *Future Generation Computer Systems*, vol. 96, pp. 227–242, Jul. 2019, https://doi.org/10.1016/j.future.2019.02.013.