AHMAD, Z. and PETROVSKI, A. 2024. Securing cyber-physical systems with two-level anomaly detection strategy. In *Proceedings of the 7th IEEE (Institute of Electrical and Electronics Engineers) Industrial cyber-physical systems international conference 2024 (ICPS 2024)*, 12-15 May 2024, St. Louis, USA. Piscataway: IEEE [online], article number 10639983. Available from: https://doi.org/10.1109/ICPS59941.2024.10639983

Securing cyber-physical systems with two-level anomaly detection strategy.

AHMAD, Z. and PETROVSKI, A.

2024

© 2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

This document was downloaded from
https://openair.rgu.ac.uk

Securing Cyber-Physical Systems with Two-level Anomaly Detection Strategy

Zeeshan Ahmad
National Subsea Centre,
Robert Gordon University,
Aberdeen, United Kingdom
z.ahmad1@rgu.ac.uk

Andrei Petrovski
National Subsea Centre,
Robert Gordon University,
Aberdeen, United Kingdom
a.petrovski@rgu.ac.uk

***Abstract* — Cyber-physical system (CPS) represents the integration of digital technologies with physical processes to revolutionize Industry 4.0 by optimizing the industrial processes. However, due to the integration of interconnected devices, the internet, and physical processes, CPS is more susceptible to cyber and physical anomalies. Anomaly detection systems can be implemented to enhance CPS security by actively identifying both physical and cyber irregularities through continuous data monitoring. To this end, this study proposes a two-level detection strategy to secure CPS from all types of anomalies. The first level uses a hybrid Convolutional Neural Network and Long Short-Term Memory to perform the binary classification. Whereas the second level uses a Gradient Boosting Machine to detect the exact type of anomaly. The proposed methodology is evaluated on the physical and network hardware-in-the-loop dataset obtained from a Water Distribution Testbed. The evaluation results demonstrated a high F1-score of 100% and 97.3% on network and physical data respectively, exhibiting its efficiency in accurately predicting anomalies while capturing the most relevant instances to achieve high accuracy.**

Keywords — Anomaly Detection System, Convolutional Neural Network, Cyber-Physical Systems, Gradient Boosting Machine, Long Short-Term Memory

I. Introduction

Cyber-physical systems (CPS) form an important component of the Industrial Internet-of-Things (IIoT) that are believed to play a key role in Industry 4.0 [1]. In CPS, shown in Fig. 1, the cyber and physical parts work together. The cyber side handles computing, networking, and control structures, making sure industrial systems can operate, connect, and work smartly. The physical part includes the manufacturing and automation systems by using industrial devices to do specific production and automation jobs [2]. Due to the technological growth over the last decade, CPSs have become hugely popular and are now embraced by major industries like smart grids [3], oil and natural gas pipelines [4], and wastewater treatment plants [5], among others.

The complex nature of CPS and the widespread connectivity of the interconnected devices to cyberspace make it more susceptible to threats. These threats can be physical in the form of faults such as broken values or pumps [6], or can be cyber security threats such as Man-in-the-Middle attacks (MitM), scanning attacks, Denial of Service (DoS) attacks, and more. This could cause problems like service disruptions, damage to equipment, and environmental pollution, among other consequences [7]. To protect the CPS against all types of anomalies, recently researchers have made efforts to propose reliable intrusion detection systems (IDS) for the CPS. An IDS can be deployed as an extra security shield to protect the CPS from different types of anomalies by constant monitoring of the data for any suspicious or abnormal behavior to detect anomalies. IDS can be categorized based on how they are deployed as host-based IDS or network-based IDS. They can also be categorized based on their detection strategy as signature-based, anomaly detection-based, specification-based, or hybrid detection-based [8].

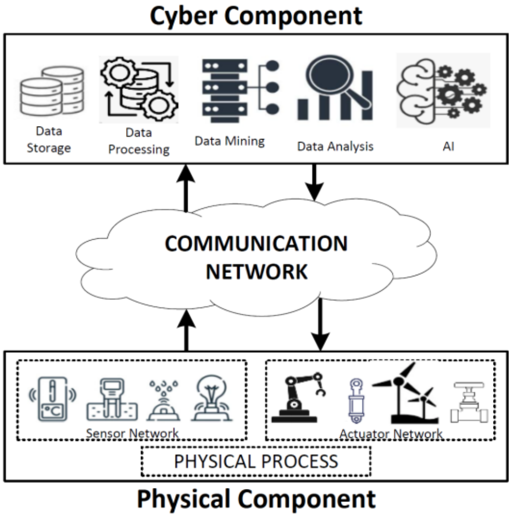


Fig. 1. Cyber Physical System

Over the past decade, anomaly detection-based IDS (AIDS) using AI techniques become hugely popular due to the ability of machine learning (ML) and deep learning (DL) methods to efficiently process data to learn important patterns for the correct prediction. In traditional ML, valuable data features are extracted through feature engineering. In contrast, the complex deep architecture of DL enables automatic learning of essential features, eliminating the need for human input or explicit feature engineering. This makes both ML and DL an ideal tool that can be integrated within AIDS to enhance anomaly detection in a CPS environment. To this end, this study considers both DL and ML-based approaches employed in two levels to propose an effective AIDS strategy for CPS.

The main contributions of this research are 3-fold. (1) To extensively discuss the state-of-the-art AI-based AIDS methodologies proposed for the CPS. (2) To propose an effective two-level anomaly detection strategy for the CPS employing a hybrid model of Convolution Neural Network (CNN) and Long Short-Term Memory (LSTM) in the first level and the Gradient Boosting Machine (GBM) in the second level. (3) To evaluate the performance of our proposed methodology on the Physical and Network hardware-in-the-loop dataset obtained from a Water Distribution Testbed (WDT) [7] and compare the performance of our proposed solution against different AI-based AIDS methodologies.

The rest of the paper is organized as; Section II provides the state-of-the-art relevant work on AIDS methodologies for CPS. Section III details the preliminary concepts and the proposed methodology. Section IV presents the dataset, experimental results, and the discussion. Finally, section V concludes this article.

II. Related Work

The researchers have widely explored ML and DL techniques to propose effective anomaly and fault detection methodologies for the CPS over the last decade. This section discusses some of the notable methodologies proposed in the literature.

Farmondi et al. [9] provided a comparative study of the performance of the ML-based detection methodologies on three different publicly available CPS datasets. They tested ML methodologies to find the effectiveness of these methodologies in efficiently detecting cyber and physical anomalies, finding the specific type of anomaly, and detecting unseen threats.

Sayegh et al. [10] proposed a specialized Intrusion Detection System (IDS) tailored for SCADA systems. This IDS operates by detecting SCADA attacks by analyzing network traffic behavior, specifically focusing on the temporal behavior of prevalent patterns within SCADA protocols. When it detects abnormal behavior, the IDS triggers alarms. The study results indicated that the IDS exhibited a high detection rate for attacks while maintaining a low false alarm rate.

Feng et al. [11] introduced an anomaly detection framework designed for Industrial Control Systems (ICS). This framework involved Bloom filter package-level anomaly detectors alongside an LSTM network-based softmax classifier. The goal was to effectively learn the normal behavioral patterns and subsequently detect anomalies within the system.

Zho et al. [12] proposed a methodology to exploit correlations between sensors to help detect anomalies for predictive maintenance in CPS. They demonstrated the effectiveness of their proposed methodology on electric generators by predicting the failures earlier to reduce maintenance and downtime costs.

Ding et al. [13] proposed an efficient DL-based online error detection and its mitigation using LSTM and LSTM autoencoder. They used the LSTM model for single-step prediction to demonstrate the model efficiently in detecting sensor data spikes, offsets, computing errors, and packet loss. to propose an efficient online error detection and mitigation model. They used the LSTM autoencoder for the multistep prediction to show the model's effectiveness in detecting the long-duration sensor errors faults caused by network delays etc. They also proposed an online error mitigation to replace the faulty values with predicted values to prevent system failures.

Paredes et al. [14] proposed a detection model using the one-dimensional (1D) CNN to detect the DoS and integrity cyber-attacks. Their proposed solution exhibited a high true positive rate to demonstrate the ability to detect and isolate cyber-attacks efficiently.

Du et al. [15] also used the LSTM autoencoder network and the Generative Adversarial Network (GAN) to detect anomalies using the cyber-physical fusion features. Their proposed methodology improved the recall of anomaly detection to overcome the challenges of insufficient labeled samples and unbalanced datasets.

This study takes a slightly different approach by proposing a hybrid approach of detecting anomalies in two levels using DL and ML to prevent CPS.

III. Proposed Solution

This section discusses the preliminary concepts followed by the details of the proposed two-level AIDS for CPS.

A. Hybrid CNN-LSTM Model

In this study, we adopted the approach of integrating the LSTM layers in the CNN (CL model) for binary classification. CNN is a popular supervised DL method that has shown effectiveness in handling data stored in matrices or arrays. In this study, the CNN model comprises an input layer, a series of convolutional layers (CoL) with activation functions and pooling layers (PoL), followed by an LSTM layer, and then an output classification layer. The combination of CoL and PoL layers works to extract important features from the sequences, the LSTM layer captures temporal relationships, and lastly, a dense layer is used for making predictions. This study uses the 1D CNN approach, which is typically used for sequential data like time series or text.

The CoL is the heart of the CNN. It takes sequential data represented in the form of a 1D array e.g., the time-series data, and applies convolution kernels (filters) to learn features, creating a feature map. This map goes through a Rectified Linear Unit (ReLU) activation to produce the layer's output L_i as [8],

$$L_i = max\left(0, b + \sum_j \left(x_{i+j} * k_j\right)\right) \tag{1}$$

where b is the bias term to the input sequence, x is the input. K_j is the convolutional kernel. The symbol * represents the convolution operation.

Passing the output feature map of the CoL layer through the PoL reduces its size by selecting the maximum value within nonoverlapping subsets. This process aims to enhance memory efficiency and prevent overfitting. The output of PoL is then processed through an LSTM layer to learn long-term dependencies from the learned local features. The LSTM utilizes four components: an input gate i_t, an output gate o_t, a forget-gate f_t and a cell gate g_t with a self-recurrent connection [16] as given mathematically for time step t as,

$$i_t = \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{(t-1)+}b_{hi}) \tag{2}$$

$$f_t = \sigma(W_{if}x_t + b_{if} + W_{hf}h_{(t-1)+}b_{hf}) \tag{3}$$

$$g_t = \tanh(W_{ig}x_t + b_{ig} + W_{hg}h_{(t-1)+}b_{hg}) \tag{4}$$

$$o_t = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{(t-1)+}b_{ho}) \tag{5}$$

$$c_t = f_t c_{t-1} + i_t g_t \tag{6}$$

$$h_t = o_t \tanh\,(c_t) \tag{7}$$

Where c_t and h_t represent the cell and hidden state, while W and b weights and biases for different gates and operations. Also, σ represents the sigmoid activation function, and tanh denotes the hyperbolic tangent function.

The output of this layer is then passed through the output classification layer which employs a sigmoid activation layer for binary classification. Mathematically, the output layer activation functions for any input x are given,

$$sigmoid\ (x) = \frac{1}{1+e^{-x}} \tag{8}$$

The architecture of the CL model adopted in this study is sequential. The CoL is configured with 64 filters and a kernel size of 3, utilizing the ReLU activation function and padding to maintain the input shape of (45, 1). Then the PoL reduces the spatial dimensionality by a factor of 2. The LSTM layer, comprising 70 units, processes the sequential information obtained from the CNN layers. A Dropout regularization with a rate of 0.1 is applied to mitigate overfitting. Finally, the classification layer is the Dense layer with a sigmoid activation function, to predict normal and anomaly instances.

B. *Gradient Boosting Machine*

Gradient Boosting Machine (GBM) is a powerful ML technique that obtains predictions by boosting through the sequential approach in building trees [17]. In GBM, the decision tree (DT) predicts the error of the previous DT to boost the gradient (error). The iterative process involves minimizing a predefined loss function by sequentially adding weak learners to an ensemble. The updated prediction $P_{i+1}(x)$ at the stage i is mathematically given as [18],

$$P_{i+1}(x) = P_i(x) + rw_i(x) \tag{9}$$

Where, P_i represents the current prediction, r is the learning rate while w_i is the contribution of the new weak learner. The contribution of the new learner is found by fitting it to the negative gradient of the loss function F for P_i as,

$$w_i(x) = \text{argmin}_w \sum_{j=1}^{N} F\left(y_j, P_i(x_j)\right) \tag{10}$$

Where y_j is the actual target for the jth sample in the dataset while x_j is the jth feature of the sample. The process iterates by sequentially adding weak learners while adjusting the predictions toward minimizing the loss function. This stepwise optimization process enables GBMs to gradually improve the model's predictive performance by focusing on the residuals or errors of the previous models.

For this study, we utilized the GBM classifier with parameters configured as follows: 100 estimators, a learning rate of 0.05, and a maximum depth of 3 for each individual tree.

C. *Methodology*

In this study, we proposed an AIDS model based on CL-GBM for the CPS, as depicted in Fig. 2. The proposed model consists of two main phases, (1) Data collection and Preparation phase and (2) Anomaly detection phase.

(1) Data collection and Preparation phase

This is the first phase of CL-GBM which performs the important tasks of the data collection followed by its preparation to make it suitable to be used for the ML process. The different steps performed in this phase are,

Step-1: The raw data from different sensors will be collected and stored in the CPS dataset.

Step-2: The collected data undergoes a cleaning process by initially eliminating redundant instances containing infinite or empty fields. Afterward, categorical features are encoded through one-hot encoding. Subsequently, each feature is normalized based on its values, scaling them between 0 and 1 using Min-Max scaling.

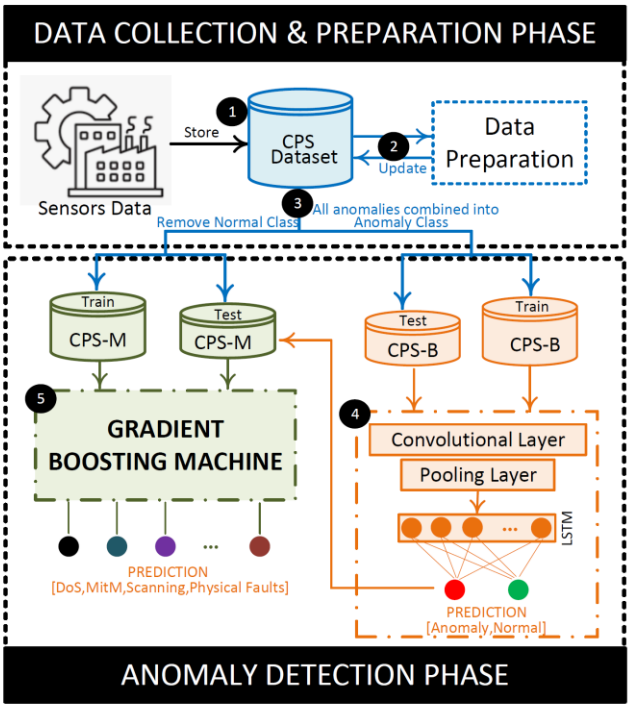


Fig. 2. Proposed CL-GBM

Step-3: The pre-processed and normalized CPS dataset is split into two datasets: CPS-B and CPS-M. In the CPS-B dataset, we combined various anomaly types into a single category named "Anomaly". Additionally, we removed the feature that originally labeled the specific types of anomalies. In CPS-M, we excluded all instances labeled as "Normal" and retained only the instances representing specific anomalies. Subsequently, both CPS-B and CPS-M datasets were divided into corresponding 75% training data and 25% testing data for model development and evaluation.

(2) Anomaly Detection Phase

This phase is the main anomaly detection phase for the CPS, employing a two-level approach. Level-1 prediction involves using the DL-based CL model. Level-2 prediction is executed using the ML-based GBM model. This dual-level approach helps enhance anomaly detection within the CPS by leveraging the strengths of both deep learning and gradient-boosting techniques. The different steps performed in this phase are,

Step-4: The CL model is trained using the CPS-B Train dataset, while the GBM model is trained with the CPS-M Train dataset. After training, the best-performing models for each approach are saved as trained models for subsequent use in the anomaly detection process.

Step-5: First the CL model is tested using the CPS-B Test dataset. If "Normal" is predicted, no further action is taken. However, upon detecting an anomaly, an alarm signal is triggered to notify the administrator for necessary actions. Simultaneously, the identified anomaly is sent to Level-2 for precise anomaly-type detection. All classified anomalies are then forwarded to the trained GBM model, which further categorizes the anomalies into specific types. This multi-level process allows for a more granular identification of anomalies within the CPS.

IV. Experimental Result and Analysis

This section details the dataset, evaluation metrics, experimental configurations, and a comprehensive analysis of the obtained results.

A. *Dataset Description*

To evaluate the performance of the CL-GBM, we used the publicly available physical and network data of the WDT dataset [7]. The dataset is obtained from the water distribution testbed that emulates water flowing between eight tanks through 22 solenoid valves, 6 pumps, 8 pressure, and 4 flow sensors. The dataset contains both cyber-attacks (DoS, MitM, Scanning) and physical faults (Tank water leak and Sensors and pump breakdown) [9]. The physical dataset is provided in CSV format with 41 features while the network dataset is provided in both PCAP and CSV formats with 14 features. For this study, we used the instances of each category extracted after 0.1 seconds from the main dataset. The physical and network data instances for different categories considered in this study are detailed in Table I. Also, for this study, the level-1 of the CL-GBM is trained in a binary fashion. So, all the instances of DoS, MitM, Scanning attacks, and Physical faults are combined into one anomaly class.

TABLE I. Dataset Distribution

Class	*Network dataset*	*Physical dataset*
Normal	44461	7498
DoS	1005	157
MitM	3554	743
Scanning	-	7
Physical Fault	4168	552

B. *Evaluation Metrics*

In this study, the performance evaluation metrics include F1-score, Precision, Recall, and Accuracy, calculated from various elements of the confusion matrix [8] and are given as;

$$\text{Precision} = \frac{TP}{TP+FP} \tag{11}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{12}$$

$$\text{F1 score} = \frac{2(\text{Precision})(\text{Recall})}{\text{Precision}+\text{Recall}} \tag{13}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{14}$$

Where, TP and TN represent the True Positive and True Negative and signify correctly predicted Anomaly and Normal samples, respectively. Whereas FN and FP represent the False Negative and False Positive respectively and denote misclassifications as Normal and Anomaly samples.

C. *Experimental Setup*

All performance evaluation experiments were conducted on an HP laptop featuring an Intel Core i9-10885H processor, 32GB RAM, and a 64-bit Windows 10 operating system. Python (version 3.10.12) served as the primary programming language to implement and evaluate all IDS methodologies within the Google Colab environment using the Keras library.

D. *Results and Discussion*

In this research, the performance of the CL-GBM is compared with five supervised ML approaches such as DT, Naïve Bayes (NB), Support vector machines (SVM), GBM, and Random Forest (RF). Also, CL-GBM performance is compared with three DL approaches CNN, LSTM, and CL. For DL, the adjusted hyperparameters included a batch size of 32, a learning rate of 0.01 with Adam optimizer, binary cross-entropy for binary classification, categorical cross-entropy for multiclass classification, ReLU for hidden layers, and sigmoid/softmax for output layers in binary and multiclass scenarios, respectively. The softmax activation function for any input x is mathematically given as,

$$softmax(x)_k = \frac{e^{x_k}}{\sum_{i=1}^{N} e^{x_i}} \text{ for } k = 1, \cdots, N \tag{15}$$

Table II presents the comprehensive performance evaluation scores, expressed as percentages, achieved by CL-GBM alongside all the ML/DL models considered for both physical and network datasets. Notably, our proposed model outperforms alternative methodologies across all considered

TABLE II. Performance Evaluation Metric Score [%]

	Network dataset				**Physical dataset**			
Model	***F1-score***	***Recall***	***Precision***	***Accuracy***	***F1-score***	***Recall***	***Precision***	***Accuracy***
DT	71.91	75.00	94.50	98.11	55.60	57.79	93.66	97.46
NB	98.55	97.81	99.40	99.83	51.16	66.47	60.83	91.43
SVM	71.82	74.83	94.45	98.07	72.85	72.88	92.84	98.21
GBM	99.97	99.94	99.90	99.99	86.75	81.60	96.59	98.84
RF	95.22	91.91	99.28	98.35	77.33	75.38	**99.88**	99.50
CNN	96.15	98.82	94.21	97.68	72.16	69.34	77.44	98.30
LSTM	89.33	89.18	92.07	97.20	65.06	67.42	65.18	96.96
CL	90.79	90.20	92.17	97.33	65.55	64.13	73.83	98.13
CL-GBM	**100.00**	**100.00**	**100.00**	**100.00**	**97.35**	**93.75**	99.86	**99.72**

evaluation metrics. Specifically, on the network dataset, both GBM and CL-GBM demonstrate exceptional predictive accuracy and capture pertinent instances, yielding high accuracy. On the physical dataset, our model maintains superior performance by obtaining the F1 score of 97.3%, while GBM exhibits a slight decrease in performance. We also observe that in terms of precision, RF performed 0.02% better than CL-GBM. As a whole, the results underscore that employing CL and GBM in a two-level structure notably enhances overall evaluation metric scores for both dataset types.

The percentage improvement in the performance of the CL-GBM over the other models on the network and the physical dataset is shown in Fig. 3 and Fig. 4, respectively. It is obvious that in dealing with the imbalance dataset, the CL-GBM showcased F1-score enhancements ranging from 0.03% to 28.09% on the Network dataset and 10.61% to 46.20% on the Physical dataset comparing other AI-based AIDS methodologies for CPS.

Fig. 5 and Fig. 6 depict the confusion matrix for the two-level CL-GBM considering the Network and Physical datasets. The first level performed the binary classification task using the CL model, followed by the GBM model performing the multiclass classification on the predicted anomalies of the first level to find the exact anomaly type. It is obvious from the level-2 of Fig. 6, that the CL-GBM performed very well even in predicting the scanning anomaly which had very few instances in the dataset. CL-GBM correctly predicted 3 out of 4 scanning anomaly instances.

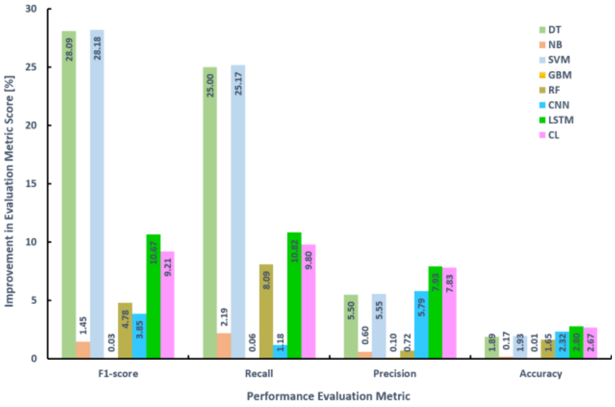


Fig. 3. Performance improvement of CL-GBM over other ML/DL models (Network dataset)

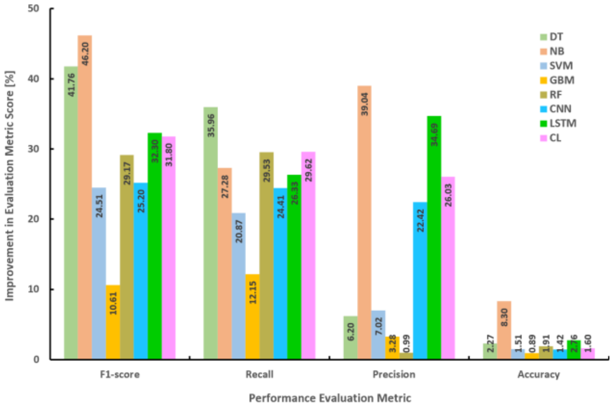
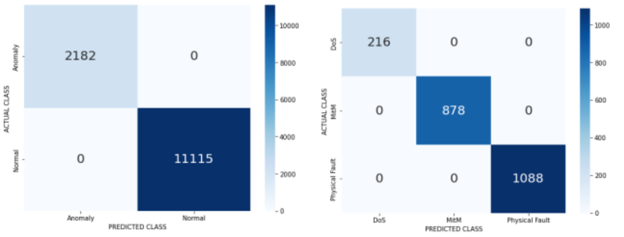
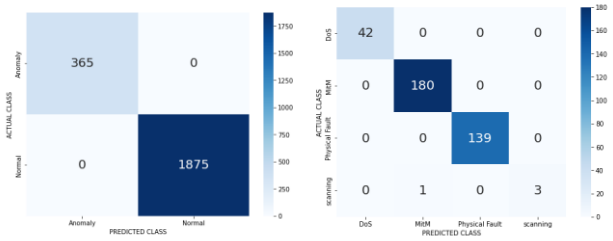


Fig. 4. Performance improvement of CL-GBM over other ML/DL models (Physical dataset)



Level – 1 Level – 2

Fig. 5. Confusion Matrix of CL-GBM (Network dataset)



Level – 1 Level – 2

Fig. 6. Confusion Matrix of CL-GBM (Physical dataset)

Table. III details the comparison of results in this study directly with the results obtained on the Physical and Network dataset, detailed in [7], [9] considering the F1-score as the evaluation metric. We noticed that the ML models used in those studies exhibited better performance on the Physical dataset compared to the Network dataset. However, it remains unclear how many instances of network flows were considered in [7] to evaluate the ML models on the Network dataset. In contrast, our two-level CL-GBM maintained consistent performance across both datasets. We also observed that both RF and CL-GBM exhibited similar performance in terms of F1 score on the Physical dataset.

TABLE III. COMPARISON OF F1-SCORE [%]

Model	***F1-score*** [7] (Network dataset)	***F1-score*** [7], [9] (Physical dataset)
RF	54	97
SVM	20	75
NB	27	77
CL-GBM (This Study)	**100**	**97.35**

V. CONCLUSIONS

This paper introduces a two-level detection method employing a hybrid approach involving CNN, LSTM, and GBM to efficiently detect cyber anomalies like DoS, MitM, and Scanning, as well as physical anomalies such as water leaks in tanks and breakdowns in sensors or pumps. The proposed solution utilizes a hybrid CL model for the effective classification of data into Normal and Anomaly categories in the first level. In the subsequent level, GBM predicts the exact types of classified Anomaly instances. The solution is evaluated using the network and physical data from the WDT dataset, showcasing its effectiveness through high evaluation metric scores achieved on both dataset types.

For future research, we aim to extend this work by evaluating its performance in real-time CPS scenarios. Additionally, we plan to explore the concept of a two-level anomaly detection approach employing unsupervised ML/DL methodologies.

REFERENCES

[1] J.-P. A. Yaacoub, O. Salman, H. N. Noura, N. Kaaniche, A. Chehab, and M. Malli, “Cyber-physical systems security: Limitations, issues and future trends,” *Microprocess Microsyst*, vol. 77, p. 103201, Sep. 2020, doi: 10.1016/j.micpro.2020.103201.

[2] H. Xu, W. Yu, D. Griffith, and N. Golmie, “A Survey on Industrial Internet of Things: A Cyber-Physical Systems Perspective,” *IEEE Access*, vol. 6, pp. 78238–78259, 2018, doi: 10.1109/ACCESS.2018.2884906.

[3] S. Y. Diaba, M. Shafie-khah, and M. Elmusrati, “On the performance metrics for cyber-physical attack detection in smart grid,” *Soft comput*, vol. 26, no. 23, pp. 13109–13118, Dec. 2022, doi: 10.1007/s00500-022-06761-1.

[4] A. S. Mohammed, P. Reinecke, P. Burnap, O. Rana, and E. Anthi, “Cybersecurity Challenges in the Offshore Oil and Gas Industry: An Industrial Cyber-Physical Systems (ICPS) Perspective,” *ACM Transactions on Cyber-Physical Systems*, vol. 6, no. 3, pp. 1–27, Jul. 2022, doi: 10.1145/3548691.

[5] C. Alexandra, K. A. Daniell, J. Guillaume, C. Saraswat, and H. R. Feldman, “Cyber-physical systems in water management and governance,” *Curr Opin Environ Sustain*, vol. 62, p. 101290, Jun. 2023, doi: 10.1016/j.cosust.2023.101290.

[6] Z. Ahmad *et al.*, “S-ADS: Spectrogram Image-based Anomaly Detection System for IoT networks,” in *2022 Applied Informatics International Conference (AiIC)*, IEEE, May 2022, pp. 105–110. doi: 10.1109/AiIC54368.2022.9914599.

[7] L. Faramondi, F. Flammini, S. Guarino, and R. Setola, “A Hardware-in-the-Loop Water Distribution Testbed Dataset for Cyber-Physical Security Testing,” *IEEE Access*, vol. 9, pp. 122385–122396, 2021, doi: 10.1109/ACCESS.2021.3109465.

[8] Z. Ahmad, A. S. Khan, K. Zen, and F. Ahmad, “MS-ADS: Multistage Spectrogram image-based Anomaly Detection System for IoT security,” *Transactions on Emerging Telecommunications Technologies*, vol. 34, no. 8, Aug. 2023, doi: 10.1002/ett.4810.

[9] L. Faramondi, F. Flammini, S. Guarino, and R. Setola, “Evaluating Machine Learning Approaches for Cyber and Physical Anomalies in SCADA Systems,” in *2023 IEEE International Conference on Cyber Security and Resilience (CSR)*, IEEE, Jul. 2023, pp. 412–417. doi: 10.1109/CSR57506.2023.10224915.

[10] N. Sayegh, I. H. Elhajj, A. Kayssi, and A. Chehab, “SCADA Intrusion Detection System based on temporal behavior of frequent patterns,” in *MELECON 2014 - 2014 17th IEEE Mediterranean Electrotechnical Conference*, IEEE, Apr. 2014, pp. 432–438. doi: 10.1109/MELCON.2014.6820573.

[11] C. Feng, T. Li, and D. Chana, “Multi-level Anomaly Detection in Industrial Control Systems via Package Signatures and LSTM Networks,” in *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, IEEE, Jun. 2017, pp. 261–272. doi: 10.1109/DSN.2017.34.

[12] P. Zhao, M. Kurihara, J. Tanaka, T. Noda, S. Chikuma, and T. Suzuki, “Advanced correlation-based anomaly detection method for predictive maintenance,” in *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*, IEEE, Jun. 2017, pp. 78–83. doi: 10.1109/ICPHM.2017.7998309.

[13] K. Ding, S. Ding, A. Morozov, T. Fabarisov, and K. Janschek, “On-Line Error Detection and Mitigation for Time-Series Data of Cyber-Physical Systems using Deep Learning Based Methods,” in *2019 15th European Dependable Computing Conference (EDCC)*, IEEE, Sep. 2019, pp. 7–14. doi: 10.1109/EDCC.2019.00015.

[14] C. M. Paredes, D. Martínez-Castro, V. Ibarra-Junquera, and A. González-Potes, “Detection and Isolation of DoS and Integrity Cyber Attacks in Cyber-Physical Systems with a Neural Network-Based Architecture,” *Electronics (Basel)*, vol. 10, no. 18, p. 2238, Sep. 2021, doi: 10.3390/electronics10182238.

[15] Y. Du, Y. Huang, G. Wan, and P. He, “Deep Learning-Based Cyber–Physical Feature Fusion for Anomaly Detection in Industrial Control Systems,” *Mathematics*, vol. 10, no. 22, p. 4373, Nov. 2022, doi: 10.3390/math10224373.

[16] J. Zhao, X. Mao, and L. Chen, “Speech emotion recognition using deep 1D & 2D CNN LSTM networks,” *Biomed Signal Process Control*, vol. 47, pp. 312–323, Jan. 2019, doi: 10.1016/j.bspc.2018.08.035.

[17] V. K. Ayyadevara, “Gradient Boosting Machine,” in *Pro Machine Learning Algorithms*, Berkeley, CA: Apress, 2018, pp. 117–134. doi: 10.1007/978-1-4842-3564-5_6.

[18] M. H. L. Louk and B. A. Tama, “Dual-IDS: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system,” *Expert Syst Appl*, vol. 213, p. 119030, Mar. 2023, doi: 10.1016/j.eswa.2022.119030.