Advances in Artificial Intelligence and Machine Learning; Research 4 (3) 2408-2415 Received 15-05-2024; Accepted 12-07-2024; Published 19-07-2024

A Model Proposal of Cybersecurity for the IIoT: Enhancing IIoT Cybersecurity through Machine Learning and Deep Learning Techniques

Atdhe Buja ab29762@seeu.edu.mk
University of South East European (SEEU)
Tetovo, N. Macedonia.

Marika Apostolova m.apostolova@seeu.edu.mk
University of South East European (SEEU)
Tetovo, N. Macedonia.

Artan Luma a.luma@seeu.edu.mk
University of South East European (SEEU)
Tetovo, N. Macedonia.

Corresponding Author: Atdhe Buja

Abstract

The Industrial Internet of Things (IIoT) acceleration has caused automation and data exchange enhancements within industrial surroundings. Yet, this evolution has presented security issues due to the raised exposure of critical infrastructure to cyber threats. This study focuses on designing a thorough model for identifying and mitigating vulnerabilities within IIoT networks utilizing Machine Learning (ML) and Deep Learning (DL) techniques. A replicated IIoT network infrastructure was set to communicate and exchange data for simulation. Use of Python script to execute network scanning and data collection, distinct possible vulnerabilities. Then, ML-DL analysis is handled by employing techniques of gradient boosting, logistic regression, decision trees, random forest, multilayer perceptron, and convolutional neural network. Throughout, gradient boosting has proven to higher performance accuracy rate in recognizing the most impactful vulnerabilities. As well as a model integral part of a Cost-Benefit Analysis (CBA) provides security recommendations to mitigate identified vulnerabilities. According to the CBA model vulnerabilities are prioritized based on the severity, related costs, and potential benefits of mitigation. The proposed Cybersecurity model in addition to high accuracy in vulnerability detection also provides a standardized approach for categorizing Cybersecurity countermeasures according to cost-effectiveness. This study emphasizes the need for a consolidated Cybersecurity model for the IIoT and shows the capability of ML techniques to advance Cybersecurity posture. Future work considered testing the model in a real operation environment of IIoT, refining the model, and enrichment with more knowledge base actionable mitigations.

Keywords: IIoT, Cybersecurity, Machine learning, Deep learning, Vulnerability assessment, Cost-benefit analysis, GNS3, Network scanning.

Citation: Atdhe Buja, et al. A Model Proposal of Cybersecurity for the IIoT: Enhancing IIoT Cybersecurity through Machine Learning and Deep Learning Techniques. Advances in Artificial Intelligence and Machine Learning. 2024;4(3):140.

1. INTRODUCTION

The Industrial Internet of Things (IIoT) defines a fusion of industrial operations with emerging technologies, facilitating automation, data transfer, and process rationalization. This development guided enhancements in many ways but also introduced major Cybersecurity challenges [1]. The IIoT is composed of sensors, actuators, and controllers that collect and transfer data constantly, facilitating smart industrial settings [2]. Nevertheless, the rise and utilization of these devices have extended the attack surface, putting IIoT networks target of various cyber-attacks [3]. The main motivation of this research is to treat the increase in Cybersecurity matters in IIoT by developing a thorough model that identifies vulnerabilities and generates actionable recommendations for further mitigation. The proposed Cybersecurity model uses Machine Learning (ML) and Deep Learning (DL) techniques to advance the precision of vulnerability detection and generate recommendations according to the Cost-Benefit Analysis (CBA) model. In our previous work [4, 5], we explored actual research and highlighted the need for a proactive solution for the protection of IIoT against threats. [6, 7] examined ML-DL capability for detecting threats in IIoT, underlining the potential to advance security countermeasures. The report of McKinsey [8], securing IIoT systems has turned into a vital priority for organizations. In the following sections, we present a comprehensive methodology for experimental setup, data analysis, and model selections [9, 10]. Thereafter, we drill down into the ML-DL analysis and explain the results and insights gained from our research. Yet, we discuss in detail the contribution and limitations of our proposed Cybersecurity model and provide advice on enhancing the Cybersecurity posture of IIoT. In response to the lack [4], of a proactive approach to identify and mitigate before they occur, this paper focuses on building a Cybersecurity model able to identify and mitigate vulnerabilities within the IIoT environment.

2. METHOD

This section introduces the methodology used in applied research [11], focusing on building our proposed Cybersecurity model for IIoT. The methodology outlines the organized approach conducted to handle the research objectives. The key objective of our research is to build a Cybersecurity model that identifies vulnerabilities and provides actionable recommendations for mitigation according to the Cost-Benefit Analysis (CBA) model. This model sets to address current Cybersecurity challenges and threats in IIoT environments. This consists of assessing the security posture, categorizing threats and vulnerabilities into severity levels, obtaining insights existed vulnerabilities, and generating recommendations from a cost-benefit outlook. Our previous research [4, 5], lays the foundations for building the model, based on the past results of existing research work associated with Cybersecurity in IIoT.

2.1 Experimental Setup

The first phase of the methodology included the setup of a thorough virtual network utilizing Graphical Network Simulator-3 (GNS3) to simulate an industrial infrastructure. This virtual network contained a variety of components router, switch, IIoT gateway aggregator, types of IIoT sensors, a server, and a machine used to run Python scripts. This environment was run on the Z Tower Workstation type, 32 GB RAM, intel Xeon CPU. The utilization of tools and software has guided our simulation experiment in obtaining data and organizing them in datasets CSV. Python is seen as the main programming language for implementing our methodology. The setup process intricate activities of installation and configuration of host GNS3 and virtual environment together with components; configuration of the network check all devices are connected and able to exchange data; functioning sensors, IIoT gateway to simulate real industrial operations. Once the setup was completed, custom Python scripts were built to assess real-time network scanning by identifying vulnerabilities and collecting data. Additionally, connection with external sources (CVE, NVD, Exploit-DB, CISA, and NIST [12–16]) with the Python script for a thorough vulnerability analysis. FIGURE 1, shows the comprehensive infrastructure that was built to conduct experimental simulations.

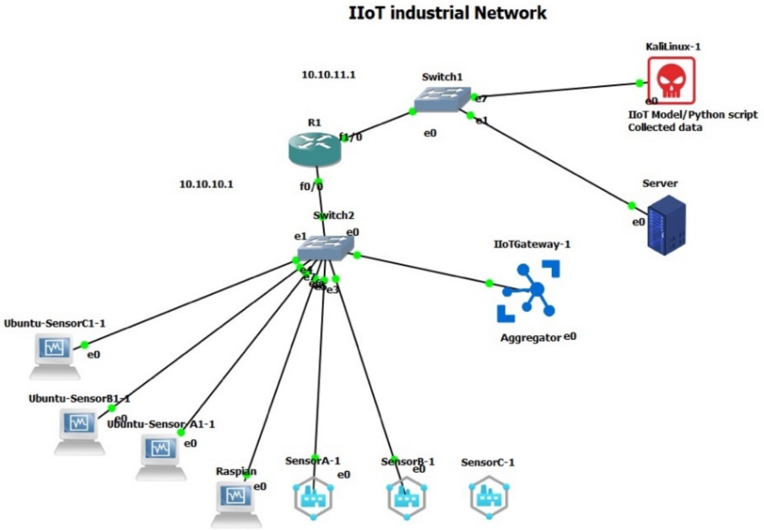

Figure 1: Overview of the infrastructure of the experimental setup.

2.2 Data Analysis

The datasets used were obtained from the network scanning, and connection with external sources that occurred during the simulation, a repository of IIoT vulnerability data. Each record in the datasets carries information on vulnerabilities with all details related to IIoT attacks (host, port, state, service, cve, description, cvss v3 base score, cvss v3 vector string, severity, vendor project, product, required action, known ransomware, edb-id, exploit raw), model references fields (ISO standard and control, NIST). At one point data is acquired, and it went through analysis utilizing Machine Learning (ML) and Deep Learning (DL) techniques. The analysis accompanies proper to the Data science lifecycle (data preprocessing, feature engineering, model selection and training, and evaluation). By engaging in data preprocessing, we were able to clean and organize the data and consolidate the data from multiple sources in one dataset (based on cve id), so we made it suitable for analysis. Drawing out appropriate features from the dataset is key to the result of the analysis. Considering the complexity of IIoT vulnerability data, it was essential to identify key features that could identify vulnerabilities and make accurate predictions. From the capability of statistical analysis, and domain expertise, we selected a category of features that were important to carry out the research objectives.

2.3 ML-DL Analysis and Model

The selection of the ML model was by testing various techniques including Gradient Boosting (GB), Logistic Regression (LR), Decision Trees (DT), and Random Forest (RF) to organize the severity of Common Vulnerabilities and Exposures (CVEs) and generate actionable information. On the other hand, the DL model selected techniques including Multi-Layer Perceptron (MLP) and Convolutional Neural Network (CNN) considering feature importance calculation using RF model, guides in main features related to vulnerability classification. Additionally, the model integrates the generating recommendations based on the CBA model to sort and generate Cybersecurity recommendations grounded on the identified vulnerabilities. The CBA model was defined by costs and benefits analysis for each recommendation, analysis of the severity, and ranking recommendations to deliver actionable steps for mitigating vulnerabilities. The first step was defining the main elements of the CBA model seeing the severity level of CVE, cost estimates, and cost in total. The CBA model distinguishes CVE severity level (critical, high, medium, low, informational) according to score (CVSS) surrounding factors. Then, assess the total costs for remediation (software patches, infrastructure changes, and training). The second step followed by assigning costs and benefits to each action grounded on risk reduction and security posture advance, recommendations are sourced from model references. Lastly, analysis of the severity of vulnerability impact by preprocessing data to pull severity scores and assess scores (CVSS) and potential outcomes to sort vulnerabilities for remediation. Here is a formula to describe the CBA model for sorting Cybersecurity recommendations:

$$\text{CBA score} = \frac{\textit{Benefit of Mitigation}}{\textit{Cost of Mitigation}}$$

3. RESULTS AND DISCUSSION

In this section, we introduce the results and discussion from the ML-DL analysis conducted by using various techniques andi achieving engaging results on the behavior of the model. Our objective is to build a Cybersecurity model that identifies vulnerabilities and generates actionable recommendations for mitigation based on the Cost-Benefit Analysis (CBA) model. The simulation results of our research emphasize the efficiency of the proposed Cybersecurity model for the IIoT. The results are grouped into model performance, vulnerability detection, and recommendations.

3.1 Model Performance

Performing ML-DL analysis techniques of Gradient Boosting (GB), Logistic Regression (LR), Decision Trees (DT), Random Forest (RF), Multi-Layer Perceptron (MLP), and Convolutional Neural Networks (CNN); helps to conclude the evaluated performance, on the dataset obtained from the simulated IIoT infrastructure. GB was engaged in identifying the highly impressive vulnerabilities in IIoT networks. It shows effectiveness in our analysis, gaining excellent metrics (accuracy, precision, recall, and F1-score 1.00 for both classes 1 ‘critical vulnerabilities’ and 3 ‘high vulnerabilities’), which means this model proved the higher performance in classifying the severity of vulnerabilities. LR is used as an inception model and has shown an accuracy of 67% followed by other metrics (precision was 1.00 for classes 1 ‘critical vulnerabilities’ and 3 ‘high vulnerabilities’, but the recall 0.50 for class 1 ‘critical vulnerabilities’ and 1.00 for class 3 ‘high vulnerabilities’), which means this model proved lower performance in identifying some high-severity vulnerabilities. DT was preferred for the facilitated interpretation and capacity to model decision-making processes in identifying vulnerabilities. The results show a higher metric (accuracy, precision, recall, and F1-score of 1.00 for both classes 1 ‘critical vulnerabilities’ and 3 ‘high vulnerabilities’) like GB, meaning a good performance in classifying vulnerability severity. RF is used to improve accuracy and maintain overfitting, achieve higher metrics (accuracy, precision, recall, and F1-score of 1.00 for both classes 1 ‘critical vulnerabilities’ and 3 ‘high vulnerabilities’), confirming its solidity in vulnerability classification. MLP and CNN were chosen for more depth in the analysis, and their feature selection was done after conducting a feature importance analysis (using an RF model). Both models show results of metrics 66.67% (high precision, recall, and F1-score for class 1 ‘critical vulnerabilities’), meaning their effectiveness and verifying the DL techniques in identifying critical vulnerabilities. The merging of these models into our research other than verifying the effectiveness of ML-DL techniques in advancing Cybersecurity same time contributes to a standardized way for classifying vulnerabilities according to the CBA model.

3.2 Vulnerability Detection

The analysis phase exhibited that the ML models (gradient boosting, decision trees, and random forest) effectually identified critical and high-severity vulnerabilities. The models utilized precisely classified vulnerabilities (‘cve’) that need immediate mitigation actions. ML analysis showed a precise classification of CVE severity by applying multiple ML models. DL analysis performance supports the potential of DL techniques in advancing vulnerability detection capabilities.

3.3 Recommendations

Grounded on the ML-DL analysis, three rational Cybersecurity recommendations were generated based on the CBA model. The generated recommendations intent on mitigation of the identified vulnerabilities with the evaluation of cost and benefits. The following TABLE 1 shows the detailed recommendations for each vulnerability (‘cve’) identified by the model analysis (ML-DL), including cost-benefit.

The above-mentioned recommendations are actionable steps for remediating the identified vulnerabilities and highlighting cost-benefit measures. The CBA model incorporation with ML analysis assures that the recommendations also guide in a standardized way for advancing Cybersecurity countermeasures in IIoT environments. Additionally, the recommendations represent practical detailed insights for mitigation strategies valuable for various stakeholders of security operation teams, network security analysts, management, and leaders of organizations.

The research engaged a comprehensive approach to advancing the Cybersecurity of IIoT utilizing the application of Machine Learning (ML) and Deep Learning (DL) techniques. The methodology,

Table 1: Three security recommendations according to the CBA model

	Recommendation	Cost Estimate	Potential Benefits
CVE-2018-0171 (Critical, Class 1) ML Identified Class: 1 DL Identified Class: 1	Apply updates per vendor instructions to prevent data theft or system compromise …………….	$5,500	Significant risk reduction; prevention of potential data theft or system compromise.
CVE-2023-7209 (High, Class 3) ML Identified Class: 3	Update Uniway Router firmware to prevent denial of service …………….	$5,500	Important risk reduction; prevention of potential data breaches or system compromise.
CVE-2023-7211 (High, Class 3) ML Identified Class: 3	Implement access controls and network segmentation to protect against unauthorized access....	$3,000	Important risk reduction; prevention of potential data breaches or system compromise.

simulation, and data analysis were carefully built to ensure precise identification and mitigation of vulnerabilities within IIoT networks. The paper aims to build a well Cybersecurity model that not only identifies vulnerabilities but generates actionable recommendations for further mitigation processes. While the number of cyber-attacks is increasing targeting industrial organizations shows a lack [4], of a proactive approach to identify and mitigate before they occur. The analysis includes ML and DL models to identify which achieves accuracy in vulnerability identification.

The gradient boosting (GB) technique exhibits higher performance metrics through overall evaluation parameters. GB's capability to treat intricate datasets and enhance model accuracy put it at the highest performance in this research. Then logistic regression (LR) offers initial insights into the dataset, although its low-performance LR provided valuable metrics for comparison. The decision trees (DT) provided clear illumination and high accuracy metrics, effective for decision-making processes associated with vulnerability identification. Random forest (RF) advanced accuracy and agility of the predictions, showed to be very effective in classifying vulnerability severity. Furthermore, these DL models MLP and CNN following feature selection using RF, exhibit the ability to identify critical vulnerabilities with high metrics (precision and recall); which confirms the validity of DL techniques in Cybersecurity applications. FIGURE 2 introduces the summarized results of the metric (accuracy) of each model utilized during the analysis using the dataset obtained from the simulation.

The insights gathered from these analyses have essential reasoning for enhancing IIoT Cybersecurity posture and guiding into proactive protection. Security operation teams, network security analysts, management, and leaders of organizations can leverage this model, facilitating proactive measures against potential threats. In addition, the ML-DL analysis of the proposed Cybersecurity model generates actionable recommendations founded in the cost-benefit analysis (CBA) model. The incorporation of the CBA model gives a chance for a structured evaluation of cost-benefits related to each mitigation strategy, giving a standardized way for decision-making in Cybersecurity.

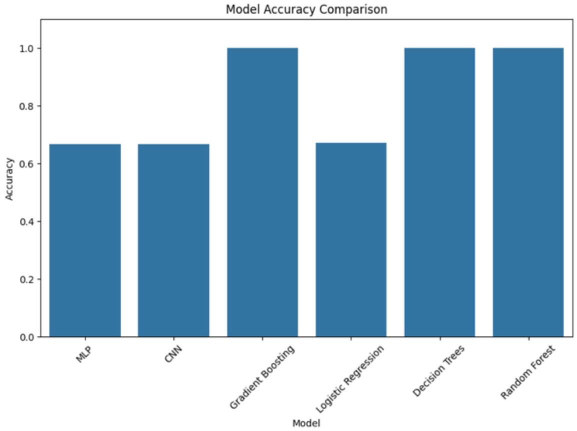

Figure 2: Summary results of accuracy metric comparison among used ML-DL models.

At last, the research emphasizes the potential of ML-DL techniques in enhancing Cybersecurity countermeasures for IIoT. Mostly, models GB and RF with high accuracy reinforce their implementation in real-world cases. Moreover, the integration of the CBA model makes sure that recommendations are useful and economically feasible. Limitations of the Cybersecurity model now may be the lack of validation with real operational data to understand the effectiveness. Since it can be introduced as a challenge, which has a solution through refining the model and improving it with an extensive knowledge base of actionable mitigations will advance its implementation.

4. CONCLUSION

This research focused on addressing Cybersecurity challenges by building a thorough Cybersecurity model that benefits Machine Learning (ML) and Deep Learning (DL) techniques able to identify and mitigate vulnerabilities of IIoT. By utilizing advanced data science techniques and ML-DL algorithm analysis, we propose an enhanced Cybersecurity model that shows abilities to detect vulnerabilities, threats, and potential cyber-attacks. By employing algorithms like gradient boosting, logistic regression, decision trees, random forest, multi-layer perceptron, and convolutional neural networks, we have presented the efficiency of achieving excellent scores in metrics for critical and high-severity vulnerabilities. Furthermore, the incorporation of a CBA model into the Cybersecurity model offers recommendations for remediating identified vulnerabilities. This way makes sure that actionable recommendations are assessed by economic feasibility, guiding informed decision-makers on Cybersecurity investments. Remarked, the limitation of the research is to validate the model with real operational data to quite assess the model's persuasiveness in practical cases. Future work will focus on model refining Machine Learning (ML) and Artificial Intelligence (AI) and integrating an expanding knowledge base of actionable mitigations. Beyond exploration will include enhancements of a more agile Cybersecurity model for its applicability and persistence upon cyber threats. In summary, insights obtained from our research emphasize the potential of ML-DL techniques in advancing the Cybersecurity posture of IIoT. By leveraging from proposed model, we can standardize the process to vulnerability category and mitigation and provide precious input for security operation teams, network analysts, and organizational executives.

References

[1] Zhang ZK, Cho MC, Wang CW, Hsu CW, Chen CK, et al. Iot Security: Ongoing Challenges and Research Opportunities. 7th International Conference on Service-Oriented Computing and Applications, Nov. 2014. IEEE PUBLICATIONS.2014:230-234.

[2] Xu H, Yu W, Griffith D, Golmie N. A Survey on Industrial Internet of Things: A Cyber-Physical Systems Perspective. IEEE Access. 2018;6:78238-78259.

[3] Lee I, Lee K. The Internet of Things (Iot): Applications, Investments, and Challenges for Enterprises. Bus Horiz. Jul. 2015;58:431-440.

[4] Buja A, Apostolova M, Luma A, Januzaj Y. Cyber Security Standards for the Industrial Internet of Things (IIoT)– A Systematic Review. International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). 2022.

[5] Buja A, Apostolova M, Luma A. Enhancing Cyber Security in Industrial Internet of Things Systems: An Experimental Assessment. 12th Mediterranean Conference on Embedded Computing (MECO). 2023;2023:1-5.

[6] Al-Garadi MA, Mohamed A, Al-Ali AK, Du X, Ali I, et al. A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security. IEEE Commun Surv Tutorials. 2020;22:1646-1685.

[7] Yang J. The Application of Deep Learning for Network Traffic Classification. Highlights in Science, Engineering and Technology. 2023;39:979-84.

[8] https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/iot%20value%20set%20to%20accelerate%20through%202030%20where%20and%20how%20to%20capture%20it/the-internet-of-things-catching-up-to-an-accelerating-opportunity-final.pdf.

[9] Li J, Othman MS, Chen H, Yusuf LM. Optimizing IoT Intrusion Detection System: Feature Selection Versus Feature Extraction in Machine Learning. J Big Data. 2024;11:36.

[10] Singh R, Ujjwal RL. Feature Selection Methods for Iot Intrusion Detection System: Comparative Study. Comp Intell. 2023:227-36.

[11] Nielsen C, Lund M, Montemari M, Paolone F, Massaro M, et al. Business Models. Routledge. 2018.

[12] MITRE Corporation. CVE Program. https://cve.mitre.org.

[13] https://nvd.nist.gov/vuln/search

[14] https://www.exploit-db.com

[15] https://www.cisa.gov/known-exploited-vulnerabilities-catalog

[16] https://csrc.nist.gov/publications.