

Next-gen security in IIoT: integrating intrusion detection systems with machine learning for industry 4.0 resilience

Lahcen Idouglid¹, Said Tkatek¹, Khalid Elfayq¹, Azidine Guezzaz²

¹Computer Sciences Research Laboratory, Ibn Tofail University, Kenitra, Morocco

²Computer Science and Mathematics Department, Cadi Ayyad University, Marrakech, Morocco

Article Info

Article history:

Received Jan 19, 2024

Revised Mar 3, 2024

Accepted Mar 5, 2024

Keywords:

Deep learning

Industrial internet of things

Intrusion detection systems

Machine learning

Smart factory

XGBoost

ABSTRACT

In the dynamic landscape of Industry 4.0, characterized by the integration of smart technologies and the industrial internet of things (IIoT), ensuring robust security measures is imperative. This paper explores advanced security solutions tailored for the IIoT, focusing on the integration of intrusion detection systems (IDS) with advanced machine learning (ML) and deep learning (DL) techniques. In this paper, we present a novel intrusion detection model to fortify Industry 4.0 systems against evolving cyber threats by leveraging ML and DL algorithms for dynamic adaptation. To evaluate the performances and effectiveness of our proposed model, we use the improved Coburg intrusion detection data sets (CIDDS) and BoT-IoT datasets, showcasing notable performance attributes with an exceptional 99.99% accuracy, high recall, and precision scores. The model demonstrates computational efficiency, with rapid learning and detection phases. This research contributes to advancing next-gen security solutions for Industry 4.0, offering a promising approach to tackle contemporary cyber.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Lahcen Idouglid

Computer Sciences Research Laboratory, Ibn Tofail University

Kenitra, Morocco

Email: lahcen.idouglid@uit.ac.ma

1. INTRODUCTION

The integration of the industrial internet of things (IIoT) into industrial landscapes holds tremendous promise for optimizing efficiency through real-time monitoring and control of diverse machine activities. However, this transformative potential is met with a critical concern cybersecurity [1]. Traditional IIoT architectures, often centralized, present vulnerabilities susceptible to cyber threats including individual points of failure and bottlenecks.

The ongoing Industry 4.0, often termed the digital transformation, is marked by the integration of modern digital technologies like internet of things (IoT), artificial intelligent (AI), cloud computing, and robotics, creating connected cyber-physical systems [2]. The concept of the IIoT pertains to the interconnectedness of industrial devices, encompassing sensors, controllers, machines, and robots, linking them both internally and to the broader Internet. The IIoT holds the potential to enhance the effectiveness, output, and adaptability of industrial operations while facilitating the emergence of innovative services and business model [3]. Nevertheless, the adoption of IIoT also brings about significant security challenges [4]. These challenges predominantly stem from the prolonged lifespan of components, the extensive scope of networks, and the stringent safety and reliability prerequisites inherent in industrial systems [5].

Addressing the intricate security challenges encountered by the IIoT necessitates inventive methods that go beyond conventional measures. One promising avenue involves the integration of advanced security technologies, specifically intrusion detection systems (IDS) and machine learning (ML) [6], [7]. IDS can play

a pivotal role in fortifying IIoT against vulnerabilities such as weak password protection and unauthorized access. By monitoring network and device activities, IDS can swiftly detect suspicious patterns or anomalies that may indicate security threats [8].

The integration of machine learning adds a layer of intelligence to the security framework. ML algorithms can learn and adapt to evolving attack scenarios, offering a proactive defense mechanism against sophisticated cyber threats. For instance, ML can analyze patterns in device behavior, identifying deviations that may signify potential security breaches [2]. This dynamic approach is particularly crucial in mitigating challenges associated with the lack of regular patches and updates, as ML can adapt to new threats without relying solely on predefined signatures [4].

In the cybersecurity domain, IDS function as defensive mechanisms crafted to detect various forms of cyberattacks. The framework is structured to defend against targeted hacking attempts by incorporating conventional security strategies, including prevention, detection, and mitigation [9], [10]. Although privacy and verification serve as preventive measures against intrusions, they may prove inadequate when dealing with malicious entities. The integration of intrusion detection introduces an additional layer of defense, promptly recognizing and mitigating potential threats. IDSs are categorized based on intrusion type, attacker profile, and detection mechanism [10]. The three primary classifications are misuse-based, hybrid-based, and anomaly-based IDSs [11]. Anomaly-based IDSs, for instance, continuously monitor network activity, identifying irregularities by contrasting them with typical behavior. On the other hand, misuse-based IDSs rely on a predefined catalog of known attacks, presenting challenges when confronted with novel and previously unseen threats [10].

The IIoT constitutes a network of intelligent devices responsible for collecting and processing data across diverse industrial sectors. Comprising three fundamental layers: perception, network, and processing; each layer confronts distinct security challenges necessitating proficient IDS for safeguarding data and devices against malicious attacks [12], [13]. Figure 1 provides a visual exploration of the IIoT. Integrating ML techniques into IDS proves instrumental in enhancing detection accuracy, adaptability, and scalability. ML empowers IDS to learn from data, discern pertinent features, and construct models capable of detecting known and unknown attacks. Additionally, ML facilitates the adaptation of IDS to the complexities of Fog/Edge computing, an emerging paradigm that involves bringing computation and storage closer to the data sources within the IIoT framework [14], [15].

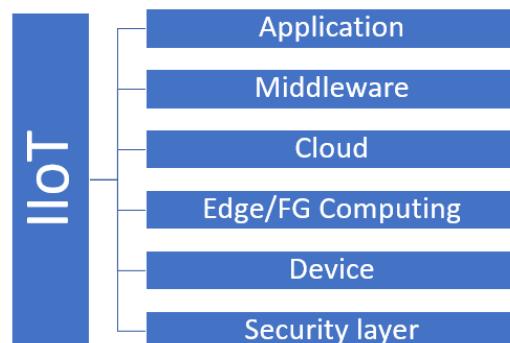


Figure 1. A Visual exploration of the industrial internet of things (IIoT)

This paper suggests a holistic strategy that involves the incorporation of sophisticated machine learning methods into intrusion detection systems. The primary objective is to enhance the security framework of IIoT systems, providing a robust defense against the dynamic spectrum of cyber threats prevalent in industrial environments [16]. Through the implementation of this integrated methodology, the paper actively contributes to the ongoing discussions surrounding the safeguarding of the future of IIoT in industrial applications. It addresses the imperative of risk mitigation and aims to ensure the sustained dependability of interconnected systems in the face of emerging security challenges.

The paper's organization follows a systematic progression. Firstly, the initial section thoroughly delves into intrusion detection studies within the IIoT, placing a particular emphasis on the utilization of machine learning. In the subsequent section, there is an intricate exploration of the proposed innovative architecture. Moving to the next phase, a meticulous presentation and comparison of results are conducted, ultimately leading to insightful conclusions. Lastly, in the concluding segment, the final section provides valuable recommendations for forthcoming studies within the fluid environment of IIoT security.

2. LITERATURE REVIEW

In this section, we thoroughly analyze and consolidate the results, methodologies, and insights gleaned from previous research relevant to our investigation. Specifically, the literature review presented in this paper critically evaluates existing studies concerning IDS within the context of the IIoT. Emphasizing the design considerations for securing IIoT, our review delves into recent advancements and challenges within the rapidly evolving domain of edge-based IIoT.

The first study cited in this section published on 2022, Idrissi *et al.* [17] analyze the feasibility of implementing a host-intrusion detection system (HIDS) based on deep learning-HIDS (DL-HIDS) across diverse commercial IoT devices. Our findings demonstrate significant promise with up to 99.74% accuracy and minimal inference time. The study concludes by emphasizing the necessity of customizing IDS for individual device classes due to their diverse architectures, debunking the practicality of a generalized DL-HIDS for all IoT devices. The proposed methodology provides valuable insights for securing IoT environments while highlighting the importance of complementing device-specific IDS integrating a central intrusion detection system within the layers of fog or cloud for comprehensive security coverage.

In the same year Kumar and Karne [18] tackles modern network and IIoT security challenges by designing an advanced intrusion detection system. Leveraging deep learning technologies, the proposed methodology optimizes network configurations, resulting in a robust IIoT anti-intrusion detection system. Demonstrating superior performance, the IDS showcases heightened detection rates, minimal false positives, and robust data correctness, aligning with privacy laws. The study advocates further validation in complex network scenarios to ensure the applicability of the proposed system in diverse industrial processes.

Alasmary [19] presents an access key agreement (AKA) scheme designed to boost security within the internet of things, with a specific focus on industrial environments. Reliable device-access framework for the industrial IoT (RDAF-IIoT) guarantees user authenticity verification, facilitating secure real-time access to information from industrial IoT devices. The efficacy of the proposed scheme is verified utilizing a random oracle framework and evaluated for its resilience against security attacks through the Scyther tool. Performance assessments demonstrate RDAF-IIoT's superiority, showcasing reduced computational and communication expenses in contrast to analogous security frameworks, coupled with upgraded security attributes.

Yao *et al.* [12] highlights the pivotal role of an intrusion detection system in safeguarding the IIoT, especially in the face of challenges presented by the expansive growth of edge-based IIoT and its decentralized data interactions. The survey delves into contemporary studies on IDS detection methods and system structures, introducing a hybrid intrusion detection system architecture that incorporates a detection method assisted by machine learning. The proposed innovations showcase heightened detection accuracy, decreased training time, and enhanced network security, positioning them as well-suited for the edge scenario of IIoT, capitalizing on diverse devices' capabilities to meet accuracy and training time requirements.

In response to security challenges in the industrial internet of things (IIoT), Guezzaz *et al.* [20] introducing PK-IDS, a cutting-edge hybrid intrusion detection systems for edge-based industrial internet of things. By seamlessly integrating k-nearest neighbors (kNN) and principal component analysis (PCA), it achieves remarkable results: 99.10% accuracy, 98.4% detection rate, and a false alarm rate of 2.7% (NSL-KDD); 98.2% accuracy, 97.6% detection rate, and 2.9% false alarm rate (Bot-IoT). A groundbreaking solution addressing intricate security challenges in edge-based IIoT environments.

Hazman *et al.* [21] introducing IDS-SIoEL, a pioneering intrusion detection framework designed for smart environments based on the IIoT. Utilizing ensemble learning, the model integrates AdaBoost and feature selection techniques, achieving remarkable performance on datasets including Edge-IIoT, IoT-23 and BoT-IoT. With an impressive 99.9% accuracy, recall, and precision, along with swift learning and detection times, the model's anomaly detection approach, boosted by advanced techniques and graphics processing unit (GPU) utilization, provides a robust solution for bolstering IoT security in diverse smart city applications.

Al-Ambusaidi *et al.* [8] address the escalating security concerns in the IoT environment by proposing a ML-enabled IDS. Focusing on the modified random forest (RF) algorithm, our proposed IDS is compared with nine established ML algorithms utilizing UNSW-NB15 and TON_IoT datasets. Evaluation metrics such as accuracy and sensitivity demonstrate the effectiveness of our ML-enabled IDS in safeguarding diverse IoT networks and applications, emphasizing its significance in addressing current security challenges.

Alshathri *et al.* [22] addresses cybersecurity threats in the IoT, particularly in industrial IoT (IIoT) applications, by proposing an effective IDS. Employing machine learning algorithms like k-nearest neighbors, random forest, and logistic regression, the study evaluates on the dataset TON_IoT, showcasing the efficiency of the classification and regression tree (CART) algorithm in mitigating IoT/IIoT intrusion risks. Additionally, it sets the stage for future exploration of deep learning algorithms and multiclass classification for industrial IoT security.

Alkadi *et al.* [23] investigates the impact of data and model quality on the performance of intrusion detection systems in the context of multi-class classification for IoT networks. Experiments using six ML models and four benchmark datasets show that employing quality data preprocessing and model configurations significantly enhances ML-based IDS detection accuracy. multi-layer perceptron (MLP) and clustering-based algorithms outperform others, achieving up to 99.97% accuracy.

Table 1 offers a thorough overview of ML-based security solutions designed for Industry 4.0 and IoT settings, highlighting methodologies, datasets, and achieved accuracies from diverse research endeavors. This compilation serves as an invaluable resource for researchers and practitioners aiming to grasp cutting-edge approaches in bolstering IoT security. By examining this table, readers can obtain a comprehensive understanding of how machine learning techniques are employed to fortify intrusion detection and safeguard IoT applications within Industry 4.0 environments.

Table 1. ML-based security solutions for Industry 4.0 and IoT: research summary

Contribution	Year	ML methods	Dataset	Accuracy
[17]	2022	Convolutional neural network (CNN)	MQTTIOT-IDS2020	99.74%
[18]	2022	Artificial neural network (ANN), CNN	N/A	98.25%
[12]	2019	LightGBM, CNN, and LR	N/A	99.9%
[20]	2022	k-NN and PCA	NSL-KDD, Bot-IoT	99.10% (NSL-KDD), 98.2% (Bot-IoT)
[21]	2023	Ensemble learning, AdaBoost	BoT-IoT, Edge-IIoT, IoT-23	99.9%
[8]	2023	GB, ET, kNN, MLP, RF, AB	UNSW-NB15, TON-IoT	100% (TON-IoT), 92.12% (UNSW-NB15)
[22]	2023	LR, LDA, kNN, NB, CART, RF, AB	TON_IoT	100%
[23]	2023	MLP	Edge-IIoT, ToN-IoT, UNSW-NB15, BOT-IoT	99.97%

Note: LR: logistic regression, GB: gradient boosting, RF: random forest, AB: AdaBoost, LDA: linear discriminant analysis, NB: naive Bayes, CART: classification and regression trees

3. METHOD

The methodology section outlines the structure of our IDS, consisting of three primary components. It provides an extensive summary of the approach used to design, implement, and evaluate our IDS. Justifying the selection of our methodology is vital for guaranteeing the trustworthiness and dependability of our research findings. This justification may involve analyzing existing methodologies in intrusion detection, providing theoretical justification based on established principles, or a combination of both. By detailing the specific algorithms, techniques, and processes employed, our aim is to provide transparency and clarity, enabling other researchers in the field to confirm and replicate our findings. Before delving into the procedural steps, let's familiarize ourselves with the components of the IDS, as depicted in Figure 2.



Figure 2. The intrusion detection system components

Step 1: Dataset pre-processing

The first task involves selecting a dataset for analysis. Subsequently, comprehensive cleaning, transformation, and normalization processes are applied. These steps ensure the dataset is well-organized, in an appropriate format, and maintains consistent scale. The feature selection phase follows, identifying pertinent features crucial for anomaly detection. This meticulous pre-processing phase sets the stage for subsequent analytical procedures, laying a solid foundation for the effectiveness of our IDS components and their performance evaluation.

Step 2: Data split

The dataset undergoes a crucial phase—data split. This involves segregating the dataset into two subsets: a training set, which accounts for 60% of the data, and a testing set, which comprises the remaining 40%. This partitioning facilitates the training of our model on one subset and the subsequent evaluation of its performance on the other, ensuring a robust and reliable assessment.

Step 3: The IDS development

Utilizing different machine learning algorithms, the intrusion detection system incorporates extreme gradient boosting (XGBoost), support vector machine (SVM), multi-layer perceptron (MLP), and k-nearest neighbors (K-NN). Through parameters tuning, these algorithms are optimized for precise intrusion detection. The IDS integrates the selected algorithms and undergoes rigorous evaluation to confirm its ability to accurately detect and classify anomalies within the dataset.

3.1. Machine learning algorithms**3.1.1. XGBoost**

XGBoost [24] stands out as an efficient and scalable implementation of gradient boosting algorithms, emphasizing speed and performance in its design. By utilizing decision trees as base learners and combining them in a boosting ensemble, XGBoost significantly enhances predictive accuracy. This approach enables XGBoost to excel in various machine learning tasks, particularly in scenarios where large datasets and high-dimensional feature spaces are prevalent.

3.1.2. Support vector machine

Support vector machine (SVM) [25] is a powerful method of supervised learning extensively utilized in regression and classification assignments, distinguished by its pursuit of the ideal hyperplane for effectively segregating various classes within the feature space. By meticulously optimizing the margin between classes, SVM enhances the algorithm's discriminatory capacity, enabling accurate classification even in complex datasets. Its versatility and effectiveness make SVM a popular choice across diverse domains, from finance to bioinformatics.

3.1.3. K-nearest neighbors

K-nearest neighbors (k-NN) [26] emerges as a straightforward yet potent supervised learning method suitable for both classification and regression tasks. By identifying the k nearest data points to a point of inquiry within the feature space, k-NN employs a majority voting scheme for classification or averaging for regression to determine the output. Despite its simplicity, k-NN's effectiveness lies in its ability to capture local patterns in the data, making it a valuable tool in various domains, from healthcare to image recognition.

3.1.4. Multi-layer perceptron

Multi-layer perceptron (MLP) [27] represents a sophisticated form of artificial neural network characterized by its multiple layers of interconnected nodes, or neurons. This architecture enables MLP to effectively tackle supervised learning tasks, leveraging its capability to learn intricate nonlinear relationships present in the data. Due to its versatility and ability to handle complex data structures, MLP finds extensive application across various domains, including image recognition, natural language processing, and financial forecasting.

4. RESULTS AND DISCUSSION

This section dissects the performance metrics across the Coburg intrusion detection data sets (CIDDS) and BOT-IoT datasets, unveiling the strengths and nuances of each machine learning algorithm employed. We present the outcomes and insights obtained from the two main benchmarking datasets, utilizing four algorithms: MLP, kNN, XGBoost and SVM. Various performance measures such as precision, F1-measure, recall and accuracy, have been employed in research literature to evaluate the efficacy of the models.

4.1. Datasets

The research community commonly employs various datasets to assess the effectiveness of IDSs with ML models. In this study, two publicly accessible IoT-based IDS datasets were selected, specifically designed for IoT applications. These recently released datasets provide substantial traffic data, addressing the need for a more comprehensive evaluation of IDS models in the context of evolving attack scenarios. Table 2 provides an overview of the Bot-IoT and CIDDS datasets.

Table 2. Overview of the Bot-IoT and CIDDS datasets

Dataset Name	Purpose	Content	Usage
Bot-IoT	Intrusion detection in IoT environments	IoT network traffic	Development and research of IoT IDSs
CIDDS	Intrusion detection within cyber-physical systems	Network within cyber-physical systems	Development and research of IDSs for cyber-physical systems

4.2. Bot-IoT datasets

The Bot-IoT [28] dataset, curated by researchers at UNSW Canberra Cyber, is a valuable asset for research in network forensics and security. It includes approximately 73 million botnet attacks in IoT networks, featuring diverse attack types. Alongside normal and simulated IoT traffic, this dataset plays a crucial role for researchers and practitioners in evaluating and strengthening IoT network security through the development and testing of intrusion detection techniques.

4.3. CIDDS datasets

The CIDDS [29] CIDDS dataset is a collection of network traffic data sets for evaluating intrusion detection systems based on anomalies. It was created by researchers from Hochschule Coburg in Germany, and it simulates a small business environment with normal and malicious activities. The data sets include different types of attacks, including port scan, brute force, and denial of service, as well as normal traffic from web browsing, email, and file synchronization. The CIDDS dataset aids in the assessment of the robustness and effectiveness of intrusion detection methodologies across various network scenarios.

4.4. Performance measures

In research literature, various performance metrics are utilized for model evaluation, with four common metrics being precision, recall, accuracy, and F1-measure, which are indirectly derived from a confusion matrix. Among these metrics, accuracy measures the correctness of the model's predictions, providing a fundamental assessment of its overall performance. Precision, also known as positive predictive value, focuses on evaluating the proportion of predicted positive instances that are correctly identified, thus offering insights into the model's ability to avoid false positives.

Recall, also referred to as sensitivity or true positive rate, quantifies the model's capability to identify all relevant cases by measuring the proportion of true positives correctly identified from all actual positive instances. F1-Score serves as a balanced metric between recall and precision, providing a harmonic mean that weighs both aspects equally. These metrics collectively offer a comprehensive evaluation of a model's performance in classification tasks, guiding the understanding of its strengths and weaknesses. This is the equation for those four-performance metrics:

$$\text{Accuracy} = \frac{TN+TP}{TN + FP + FN + TP} \quad (1)$$

$$\text{Precision} = \frac{TP}{FP+TP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1 - Score} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4)$$

Our meticulously curated testing environment featured a cutting-edge Kaggle platform equipped with 15 gigabytes of GPU memory and 13 gigabytes of random-access memory (RAM), all running seamlessly on the latest Python version (3.11.2) within a 64-bit operating system. This robust setup, accessible through Jupyter Lab, facilitated the precise measurement of performance metrics, ensuring comprehensive evaluation and validation of machine learning models with unparalleled efficiency and accuracy.

4.5. Experiments results

Table 3 provides a comprehensive evaluation of machine learning algorithm performance on the CIDDS dataset for intrusion detection systems. XGBoost shows exceptional accuracy (99.93%), precision (99.90%), F1-score (99.82%), and recall (99.82%), highlighting its robust performance. SVM exhibits solid accuracy (93.77%) but comparatively lower precision (87.01%), while K-NN and MLP demonstrate competitive accuracy and precision. These metrics offer valuable insights for informed decision-making in deploying intrusion detection systems in CIDDS scenarios.

Table 3. The results of performance metrics on the CIDDS dataset

ML algorithm	ACC%	Precision%	F1-score%	Recall%
XGBoost	99.93	99.90	99.82	99.82
SVM	93.77	87.01	87.41	87.83
K-NN	96.66	92.63	93.42	94.30
MLP	99.26	98.74	98.80	98.86

Figure 3 displays the performance metrics of various machine learning algorithms on both the CIDDS and BOT-IoT datasets. The metrics provide insights into the effectiveness of these algorithms in intrusion detection for IoT environments. Figure 1(a) displays metrics on the CIDDS dataset, showing exceptional results for XGBoost and MLP, with XGBoost achieving the highest accuracy (99.93%). SVM and K-NN also perform well, with K-NN excelling in recall. Figure 3(b) illustrates results from the BOT-IoT dataset, indicating high performance for all algorithms (XGBoost, SVM, K-NN, and MLP), with XGBoost leading in accuracy. These findings suggest the proposed intrusion detection system, utilizing these algorithms, effectively identifies and responds to security threats in IoT environments.

Table 4 displays performance metrics of machine learning algorithms on the BOT-IoT dataset. XGBoost achieves remarkable results with 99.99% recall, precision, accuracy, and F1-Score, showcasing exceptional classification capability. SVM, K-NN, and MLP also demonstrate high performance, underscoring their effectiveness in intrusion detection. These findings highlight the reliability and efficiency of these algorithms in enhancing security in IoT environment.

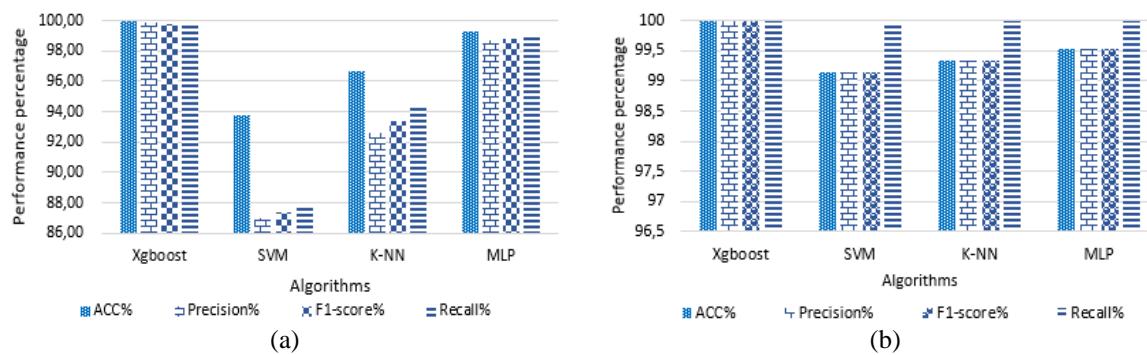


Figure 3. Visualizing performance metrics on: (a) CIDDS and (b) BOT-IoT datasets

Table 4. The BOT-IoT dataset's performance metrics results

ML algorithm	ACC %	Precision %	F1-score %	Recall %
XGBoost	99,99	99,99	99,99	100
SVM	99,15	99,15	99,15	99,95
K-NN	99,33	99,33	99,33	99,99
MLP	99,52	99,52	99,52	99,99

Figure 4 illustrates a comparative performance analysis of ML algorithms on both the CIDDS and BOT-IoT datasets. Figure 4(a) through (d) provide detailed insights into different metrics: accuracy, recall, precision, and F1-Score. These visual representations offer a comprehensive understanding of how various ML algorithms perform across the two datasets, aiding in the evaluation and comparison of their effectiveness in intrusion detection.

The thorough examination of machine learning algorithms' performance on the CIDDS and BOT-IoT datasets offers nuanced insights into their effectiveness within intrusion detection systems (IDS). Across the CIDDS dataset, XGBoost emerges as a standout performer, showcasing exceptional accuracy (99.93%), precision (99.90%), F1-score (99.82%), and recall (99.82%). This robust performance underscores XGBoost's capability in accurately identifying and classifying both benign and malicious instances. Additionally, MLP demonstrates strong performance with a high accuracy of 99.26% and competitive scores in precision, F1-Score, and recall. However, SVM and K-NN exhibit slightly lower performance metrics compared to XGBoost and MLP, indicating potential areas for optimization to enhance their efficacy further.

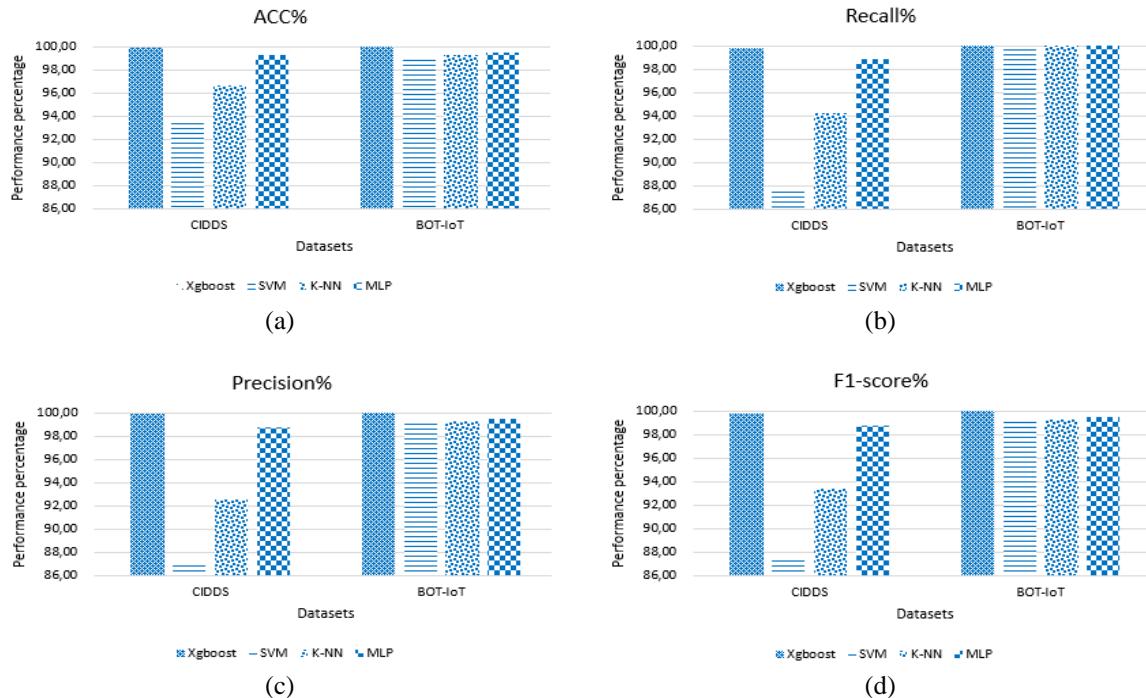


Figure 4. Comparative performance analysis of ML algorithms on CIDDS and BOT-IoT datasets:
 (a) accuracy, (b) recall, (c) precision, and (d) F1-score

On the BOT-IoT dataset, all machine learning algorithms deliver outstanding results, reaffirming their suitability for intrusion detection tasks in IoT environments. Notably, XGBoost shines with near-perfect scores across all metrics, boasting an accuracy of 99.99% and demonstrating exceptional capability in accurately detecting and classifying intrusions. SVM, K-NN, and MLP also exhibit impressive performance, highlighting their reliability and versatility in handling diverse intrusion scenarios. These findings underscore the critical role of machine learning algorithms, particularly XGBoost, in fortifying cybersecurity measures in IoT environments. As IoT ecosystems continue to expand in complexity and scale, the demonstrated efficacy of machine learning algorithms serves as a foundation for developing proactive and adaptive security solutions tailored to mitigate evolving threats effectively, ensuring the integrity and security of interconnected systems and devices.

This study sheds light on the efficacy of machine learning algorithms in IIoT intrusion detection using CIDDS and BOT-IoT datasets. However, limitations exist. Evaluation on specific datasets may not fully reflect the complexities of real-world IIoT deployments. Thus, generalizing findings to broader IoT scenarios may be constrained, warranting further research. Moreover, the focus on technical metrics overlooks practical challenges in real-world implementations.

Future research in securing IIoT environments, including agriculture, smart cities, and smart universities, will integrate intrusion detection systems (IDS) and machine learning (ML) for real-time anomaly detection against threats and attacks. Leveraging ML algorithms like XGBoost and MLP, the study aims to enhance cybersecurity by detecting and mitigating evolving cyber threats in real-time. In agriculture, real-time anomaly detection will safeguard equipment and automated systems against cyberattacks, while in smart cities, it will protect critical infrastructure and citizen data from malicious activities [30], [31]. Similarly, in smart universities, real-time anomaly detection will secure student and administrative data and campus facilities against unauthorized access and cyber breaches [32]. By exploring scalability and resilience, future research will contribute to robust cybersecurity frameworks tailored to IIoT environments, effectively addressing emerging challenges posed by cyber threats and attacks [33], [34].

5. CONCLUSION

In conclusion, this study adds important knowledge on how to combine innovative IDS integrated with ML techniques to strengthen security in the industrial internet of things integrated within the Industry 4.0 framework. The proposed model, evaluated using the enhanced BoT-IoT dataset, demonstrates exceptional performance, achieving a remarkable 99.99% accuracy along with high recall and precision

scores. The model's computational efficiency, with rapid learning and detection phases, positions it as a promising solution for addressing evolving cyber threats in Industry 4.0. For future works, the research can explore scalability to larger datasets, resilience to adversarial attacks, and adaptability to emerging cyber threats. Enhancing model interpretability and continuous refinement are essential for addressing evolving challenges in Industry 4.0 cybersecurity.

REFERENCES

- [1] M. L. M. E, and M. A, "Cybersecurity management for (Industrial) internet of things: challenges and opportunities," *Journal of Information Technology & Software Engineering*, vol. 08, no. 05, 2018, doi: 10.4172/2165-7866.1000250.
- [2] M. Serror, S. Hack, M. Henze, M. Schuba, and K. Wehrle, "Challenges and opportunities in securing the industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 2985–2996, May 2021, doi: 10.1109/TII.2020.3023507.
- [3] T. S. Fun and A. Samsudin, "Recent technologies, security countermeasure and ongoing challenges of industrial internet of things (IIoT): A survey," *Sensors*, vol. 21, no. 19, Oct. 2021, doi: 10.3390/s21196647.
- [4] J. Zhang, H. Chen, L. Gong, J. Cao, and Z. Gu, "The current research of IoT security," in *Proceedings - 2019 IEEE 4th International Conference on Data Science in Cyberspace, DSC 2019*, Jun. 2019, pp. 346–353, doi: 10.1109/DSC.2019.00059.
- [5] N. A. Khan, A. Awang, and S. A. A. Karim, "Security in internet of things: a review," *IEEE Access*, vol. 10, pp. 104649–104670, 2022, doi: 10.1109/ACCESS.2022.3209355.
- [6] Bhupal Naik D. S., V. Dondeti, and S. Balakrishna, "Comparative analysis of machine learning-based algorithms for detection of anomalies in IIoT," *International Journal of Information Retrieval Research*, vol. 12, no. 1, pp. 1–55, May 2022, doi: 10.4018/ijirr.298647.
- [7] K. El Fayq, S. Tkatek, L. Idouglid, and J. Abouchabaka, "Detection and extraction of faces and text lower third techniques for an audiovisual archive system using machine learning," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 9, pp. 625–632, 2022, doi: 10.14569/IJACSA.2022.0130974.
- [8] M. Al-Ambusaidi, Z. Yinjun, Y. Muhammad, and A. Yahya, "ML-IDS: an efficient ML-enabled intrusion detection system for securing IoT networks and applications," *Soft Computing*, vol. 28, no. 2, pp. 1765–1784, Dec. 2024, doi: 10.1007/s00500-023-09452-7.
- [9] W. Gou, H. Zhang, and R. Zhang, "Multi-classification and tree-based ensemble network for the intrusion detection system in the internet of vehicles," *Sensors (Basel, Switzerland)*, vol. 23, no. 21, p. 8788, Oct. 2023, doi: 10.3390/s23218788.
- [10] G. Kim and B. C. Kim, "Classification of functional types of lines in P&IDs using a graph neural network," *IEEE Access*, vol. 11, pp. 73680–73687, 2023, doi: 10.1109/ACCESS.2023.3296223.
- [11] S. Neupane *et al.*, "Explainable intrusion detection systems (X-IDS): a survey of current methods, challenges, and opportunities," *IEEE Access*, vol. 10, pp. 112392–112415, 2022, doi: 10.1109/ACCESS.2022.3216617.
- [12] H. Yao, P. Gao, P. Zhang, J. Wang, C. Jiang, and L. Lu, "Hybrid intrusion detection system for edge-based IIoT relying on machine-learning-aided detection," *IEEE Network*, vol. 33, no. 5, pp. 75–81, Sep. 2019, doi: 10.1109/MNET.001.1800479.
- [13] S. Haque, F. El-Moussa, N. Komminos, and R. Muttukrishnan, "A systematic review of data-driven attack detection trends in IoT," *Sensors*, vol. 23, no. 16, Aug. 2023, doi: 10.3390/s23167191.
- [14] R. Alghamdi and M. Bellaiche, "An ensemble deep learning based IDS for IoT using Lambda architecture," *Cybersecurity*, vol. 6, no. 1, Mar. 2023, doi: 10.1186/s42400-022-00133-w.
- [15] S. Latif *et al.*, "Deep learning for the industrial internet of things (IIoT): A comprehensive survey of techniques, implementation frameworks, potential applications, and future directions," *Sensors*, vol. 21, no. 22, p. 7518, Nov. 2021, doi: 10.3390/s21227518.
- [16] S. M. Umran, S. Lu, Z. A. Abduljabbar, J. Zhu, and J. Wu, "Secure data of industrial internet of things in a cement factory based on a blockchain technology," *Applied Sciences (Switzerland)*, vol. 11, no. 14, p. 6376, Jul. 2021, doi: 10.3390/app11146376.
- [17] I. Idrissi, M. Azizi, and O. Moussaoui, "A lightweight optimized deep learning-based host-intrusion detection system deployed on the edge for IoT," *International Journal of Computing and Digital Systems*, vol. 11, no. 1, pp. 209–216, Jan. 2022, doi: 10.12785/ijcds/110117.
- [18] A. Arun Kumar and R. Krishna Karne, "IIoT-IDS network using inception CNN model," *Journal of Trends in Computer Science and Smart Technology*, vol. 4, no. 3, pp. 126–138, Aug. 2022, doi: 10.36548/jtcsst.2022.3.002.
- [19] H. Alasmary, "RDAF-IIoT: reliable device-access framework for the industrial internet of things," *Mathematics*, vol. 11, no. 12, p. 2710, Jun. 2023, doi: 10.3390/math11122710.
- [20] A. Guezzaz, M. Azrour, S. Benkirane, M. Mohy-Eddine, H. Attou, and M. Douiba, "A lightweight hybrid intrusion detection framework using machine learning for edge-based IIoT security," *International Arab Journal of Information Technology*, vol. 19, no. 5, pp. 822–830, 2022, doi: 10.34028/iajит/19/5/14.
- [21] C. Hazman, A. Guezzaz, S. Benkirane, and M. Azrour, "IIDS-SIoEL: intrusion detection framework for IoT-based smart environments security using ensemble learning," *Cluster Computing*, vol. 26, no. 6, pp. 4069–4083, Nov. 2023, doi: 10.1007/s10586-022-03810-0.
- [22] S. Alshathri, A. El-Sayed, W. El-Shafai, and E. El-Din Hemdan, "An efficient intrusion detection framework for industrial internet of things security," *Computer Systems Science and Engineering*, vol. 46, no. 1, pp. 819–834, 2023, doi: 10.32604/csse.2023.034095.
- [23] S. Alkadi, S. Al-Ahmadi, and M. M. Ben Ismail, "Toward improved machine learning-based intrusion detection for internet of things traffic," *Computers*, vol. 12, no. 8, Jul. 2023, doi: 10.3390/computers12080148.
- [24] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, vol. 13-17-Augu, pp. 785–794, doi: 10.1145/2939672.2939785.
- [25] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.
- [26] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967, doi: 10.1109/TIT.1967.1053964.
- [27] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, doi: 10.1038/323533a0.
- [28] X. Zhang, O. Upton, N. L. Beebe, and K. K. R. Choo, "IoT Botnet forensics: a comprehensive digital forensic case study on Mirai botnet servers," *Forensic Science International: Digital Investigation*, vol. 32, Art. no. 300926, Apr. 2020, doi: 10.1016/j.fsidi.2020.300926.

- [29] J. Carneiro, N. Oliveira, N. Sousa, E. Maia, and I. Praça, "Machine learning for network-based intrusion detection systems: an analysis of the CIDDS-001 dataset," in *Lecture Notes in Networks and Systems*, vol. 327 LNNS, Springer International Publishing, 2022, pp. 148–158.
- [30] L. Elhaloui, S. El Filali, E. H. Benlahmer, M. Tabaa, Y. Tace, and N. Rida, "Machine learning for internet of things classification using network traffic parameters," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 3, pp. 3449–3463, Jun. 2023, doi: 10.11591/ijece.v13i3.pp3449-3463.
- [31] A. Hafian, M. Benbrahim, and M. N. Kabbaj, "IoT-based smart irrigation management system using real-time data," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 6, pp. 7078–7088, Dec. 2023, doi: 10.11591/ijece.v13i6.pp7078-7088.
- [32] M. M. Rashid *et al.*, "Adversarial training for deep learning-based cyberattack detection in IoT-based smart city applications," *Computers and Security*, vol. 120, Art. no. 102783, Sep. 2022, doi: 10.1016/j.cose.2022.102783.
- [33] J. Karande and S. Joshi, "DEDA: An algorithm for early detection of topology attacks in the internet of things," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 2, pp. 1761–1770, Apr. 2021, doi: 10.11591/ijece.v11i2.pp1761-1770.
- [34] H. A. Al-Fatlawi and H. J. Motlak, "Smart ports: towards a high performance, increased productivity, and a better environment," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 2, pp. 1472–1482, Apr. 2023, doi: 10.11591/ijece.v13i2.pp1472-1482.

BIOGRAPHIES OF AUTHORS



Lahcen Idouglid a PhD researcher at Ibn Tofail University's Faculty of Sciences in Kenitra, works within the Computer Science Research Laboratory (LaRIT). His research encompasses computer networks, software engineering, artificial intelligence, and security. He can be contacted at email: lahcen.idouglid@uit.ac.ma.



Said Tkatek a professor of computer science at Ibn Tofail University in Kenitra, he is a member of the Computer Science Research Laboratory (LaRI). His primary research across various fields includes big data, artificial intelligence (AI), and their applications. He can be contacted at email: said.tkatek@uit.ac.ma.



Khalid Elfayq a PhD researcher at Ibn Tofail University's Faculty of Sciences in Kenitra, works within the Computer Science Research Laboratory (LaRIT). His research encompasses software engineering, computer networks, artificial intelligence, and audiovisual technologies. For further communication, he can be contacted at email: khalid.elfayq@uit.ac.ma.



Azidine Guezzaz serves as a professor of computer science and mathematics at Cadi Ayad University in Marrakech. His primary research interests encompass artificial intelligence, security, cryptography, and intrusion detection. He acts as a guest editor for special issues in several journals and also contributes as a reviewer for various scientific publications. He can be contacted at email: a.guzzaz@gmail.com.