Journal
of the
Oriental Institute
M.S. University of Baroda
ISSN: 0030-5324
UGC CARE Group 1

A STUDY ON MACHINE LEARNING APPROACHES FOR REAL-WORLD INDUSTRIAL INTRUSION DETECTION

Mussaveer Tungal
Department of Computer Science and Engineering,
Institute of engineering and technology, Mangalayatan University, Beswan,
Aligarh - 202146
20200969_mussaveer @mangalayatan.edu.in
Dr. Meena Chaudhary.
Assistant professor, Research Guide,
Department of Computer Science and Engineering,
Institute of engineering and technology, Mangalayatan university, Beswan ,
Aligarh - 202146.
meenachaudhary9350@gmail.com

ABSTRACT:

As industrial systems increasingly depend on digital networks and automation, the danger of cyberattacks presents substantial threats to operational integrity and safety. This paper examines diverse machine learning methodologies for practical industrial intrusion detection, emphasizing their efficacy in recognizing and alleviating security concerns. The research seeks to assess the efficacy of these algorithms by utilizing extensive datasets that accurately represent real industrial contexts, emphasizing their capacity to adjust to dynamic and developing attack vectors. The chief objective of this project is to investigate machine learning methodologies for real-world industrial intrusion detection. Methods such as the Maximum Posterior Dichotomous Quadratic Discriminant Analysis (MPDQDJREBC) and the Weibull Distributive Generalized Multidimensional Scaling-Multivariate Censored Phi Extreme Learning Machines for Attack Detection (WDGMS-MCP ELM-AD) in the Internet of Things (IoT) domain exemplify the diverse techniques that have been suggested. The newly proposed technology enhances attack detection accuracy while simultaneously decreasing the time required and the rate of false positives comprehensively. This study has significant consequences, as it improves the comprehension of machine learning's function in protecting industrial systems and offers practical insights for firms aiming to strengthen their cybersecurity frameworks. This research is a preliminary effort to incorporate advanced analytics into industrial cybersecurity initiatives, enhancing the safety and security of operational settings.

Keywords: Machine Learning Approaches; IoT; Real – World; Intrusion Detection; Industrial Systems.

ISSN: 0030-5324
UGC CARE Group 1

INTRODUCTION:

The research on machine learning methodologies for practical industrial intrusion detection arises from the increasing necessity to safeguard industrial settings against advancing cyber threats. Industrial systems, especially those related to essential infrastructure such as manufacturing facilities, power grids, and smart factories, are progressively dependent on interconnected devices and intricate networks, rendering them vulnerable to advanced cyber-attacks (Kus et al., 2022). Conventional security methods, like rule-based systems & signature-based intrusion detection, are insufficient against advanced threats such as zero-day assaults, insider threats, and sophisticated malware. Machine learning provides a viable solution by allowing systems to identify anomalies and adjust to emerging attack patterns instantaneously, thus improving the security framework of industrial networks. Recent advancements in machine learning, encompassing supervised, unsupervised, and reinforcement learning methodologies, enable the analysis of substantial data volumes produced by industrial control systems, facilitating the identification of patterns suggestive of hostile activity (Umer et al., 2022; Pinto et al., 2023). Figure 1 below illustrates an overview of the role of machine learning in the intrusion detection process.

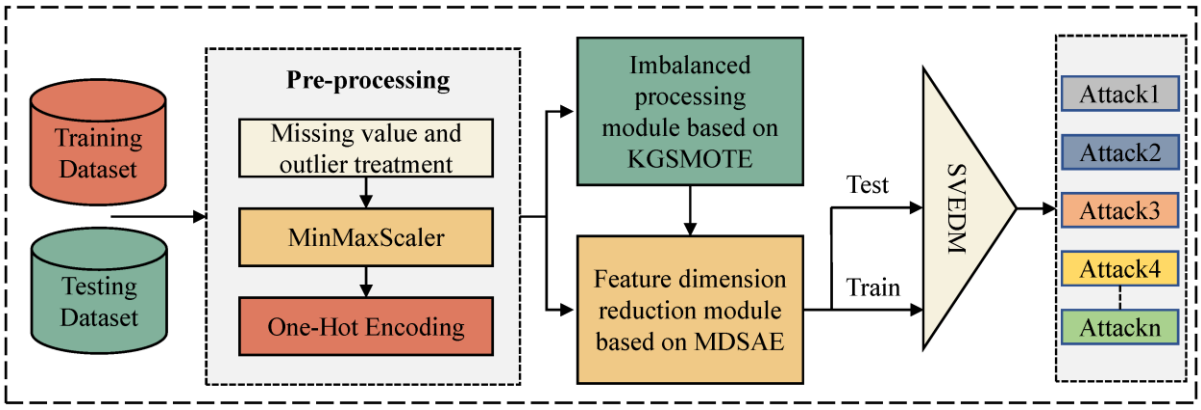

Figure 1: Overview of the role of Machine learning in intrusion detection process (Yang et al., 2023)

This study seeks to examine diverse machine learning techniques for intrusion detection, evaluate their efficacy in industrial settings, and pinpoint the problems and potential for enhancing security in vital industrial infrastructures. The following section elaborates the past literatures related to this study in detail.

LITERATURE REVIEW:

The subsequent table delineates the previous research pertinent to this study on machine learning methodologies for real-world industrial intrusion detection.

Table 1: Studies related to machine learning methodologies in intrusion detection

AUTHORS AND YEARS	METHODOLOGY	FINDINGS
Vinayakumar et al, (2019)	This research used a DNN to create a flexible and effective IDS to detect and categorize unexpected threats.	"The DNN model that did well on KDDCup 99 is benchmarked on NSL-KDD, UNSW-NB15, Kyoto, WSN-DS, and CICIDS 2017''.
MR, (2021)	The article analysed a city-scale water system and identifies obstacles in scaling a data-centric solution for critical infrastructure.	"This study distinguished itself by doing trials on live ICS, unlike previous research that used historical data".
Gaber et al., (2023)	This research introduced a new intrusion detection model using "Particle Swarm Optimization (PSO) and Bat algorithm (BA) for feature selection and the Random Forest (RF)" classifier to classify IIoT-based network traffic harmful behaviours.	"Compared to recent state-of-the-art ML and multiobjective algorithms, the outcomes were better. RF+BA was the best classifier."
Al Lail et al., (2023)	Suggested developing an NIDS system employing ML to identify contemporary attacks with high accuracy.	"Modern network threats are detected by the random forest model 97% of the time, outperforming other models".

Research Gap: The lack of industrial application & validation of machine learning systems for industrial intrusion detection is a research gap. There is little study on machine learning models in real-time, complex industrial networks with different data sources and operational restrictions, despite their success in controlled or simulated contexts. Underexplored issues include high false positives, insufficient labelled data, model flexibility to new threats, and interface with industrial control systems. To create resilient, scalable, and successful industrial intrusion detection systems, these deficiencies must be addressed.

ISSN: 0030-5324
UGC CARE Group 1

METHODOLOGY:

The figure below shows how to detect data attacks by analysing multiple data sets with different properties. The suggested method aims to anticipate attack data from datasets faster and more accurately. Performance investigation of the MPDQDJREBC and WDGMS-MCP-ELM-AD-IoT approaches is done using Python-Tensor flow on a CPU core i7. The UNSW-NB15 dataset is used for experiments. Kaggle provided the IIoT network traffic dataset at https://www.kaggle.com/mrwellsdavid/unsw-nb15. The dataset is varied. Using CSV files. For them, training. For simulation, CSV files are considered. Training is it. CSV files include 1,75,341 records with 45 attributes. The simulation considers 5000–50,000 data points from the dataset.

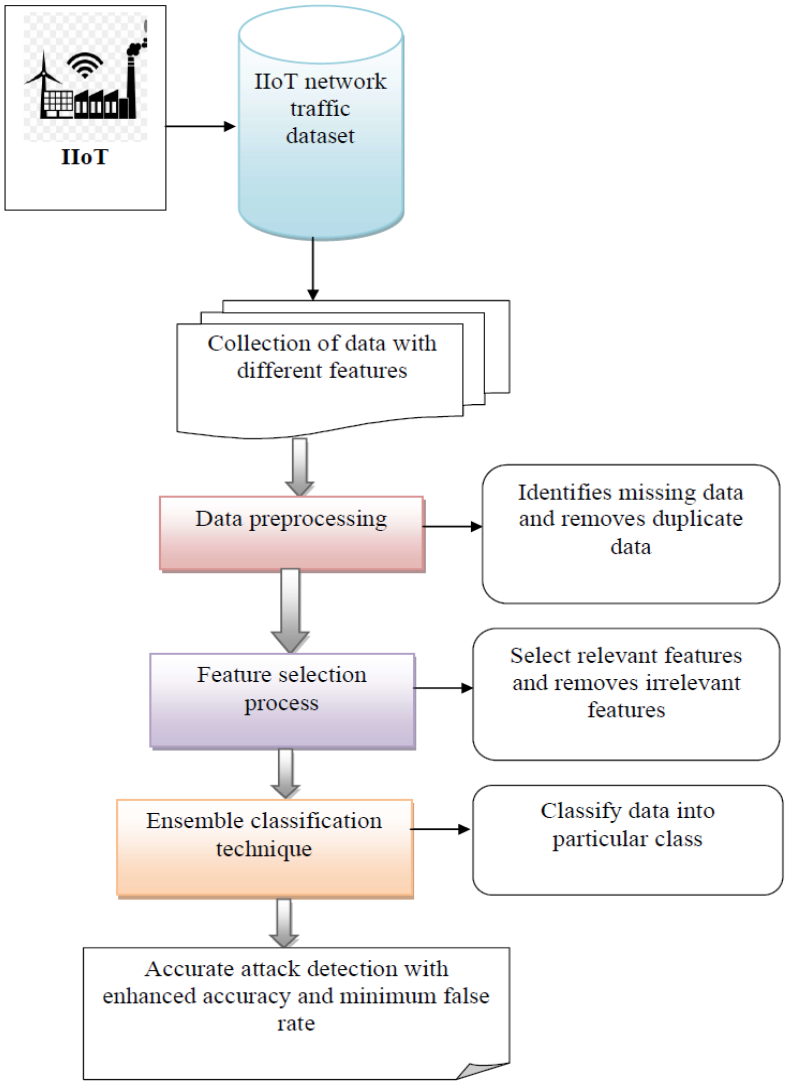

Figure 2: Process of attack detection using IIoT network traffic dataset

RESULTS AND DISCUSSIONS:

The accuracy is defined as the ratio of the number of data that are correctly identified as normal and malicious data based on the classifier process according to the total number of input data. This ratio is applied to the total number of data that is input. On the basis of certain significant characteristics, the classification of the data is carried out. The degree of significance is expressed as a percentage (%).

Table 2: Tabulation for accuracy

Number of data	Accuracy (%)			
	Existing HDRaNN	**Existing DRaNN-AD-IoT**	**Proposed MPDQDJREBC**	**Proposed WDGMS-MCP-ELM-AD-IOT**
5000	88	90	92	93
10000	90	93	94	95
15000	89	92	94	96
20000	91	92	95	95
25000	90	91	94	95
30000	88	90	92	94
35000	90	92	93	96
40000	89	91	94	95
45000	87	89	93	94
50000	90	91	94	95

The experiment's results on the accuracy of attack data identification by classification are presented in the table above, categorized by data type. The data that are evaluated for the purpose of conducting the experiment here are those that fall between the range of 5000 to 50000. Taking into account the values of the table, the accuracy of the classification of data is also becoming more variable across all of the approaches as the number of data increases.

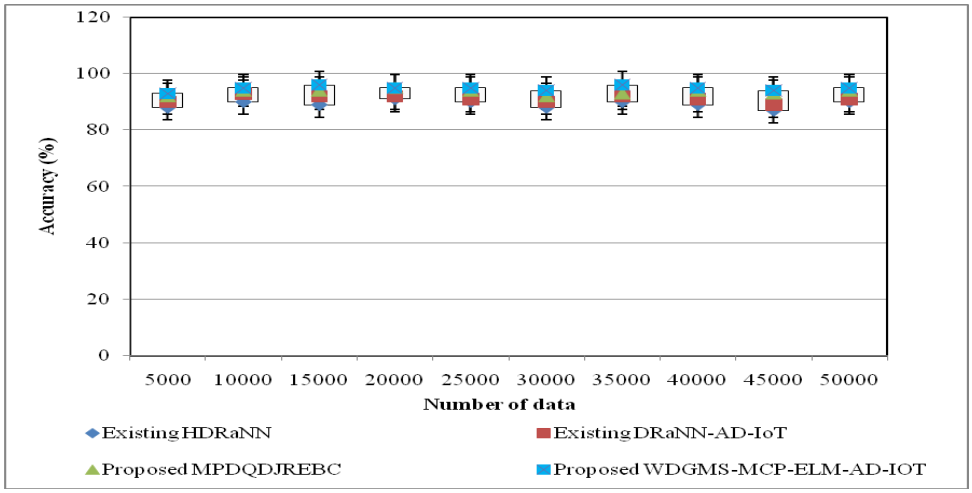

Figure 4.12: Measure of accuracy

The figure above shows accuracy performance analysis on identifying input data from various sources. The experiment uses data from the network dataset between 5000 and 50,000. The figure compares different proposed solutions with the Hybrid Deep Random Neural Network and Deep Random Neural Network mode for attack detection in industrial Internet of Things technologies.

The table that can be found above presents a comparison between the suggested MPDQDJREBC approach and the WDGMS-MCP-ELM-AD-IoT technique and other ways that are already in use. These methods include HDRaNN, which was developed by Zil Huma et al. (2021), and DRaNN-AD-IoT, which was discussed by Shahid Latif et al. respectively. Because of this, the WDGMS-MCP-ELM-AD-IoT strategy provides equivalent results to those of the other methods when it comes to the accuracy of attack detection on data. In order to achieve better performance, the graph that is displayed below was generated based on the parameters mentioned in the table above.

CONCLUSION:

The study concludes that improved machine learning methods are needed to improve intrusion detection in real-world industrial situations where standard methods fail against developing cyber threats. Machine learning approaches may detect sophisticated threats and adapt to dynamic threat landscapes, but high false positive rates, data restrictions, and integration issues make them challenging to apply.

REFERENCES:

Kus, D., Wagner, E., Pennekamp, J., Wolsing, K., Fink, I. B., Dahlmanns, M., ... & Henze, M. (2022, May). A false sense of security? Revisiting the state of machine learning-based

industrial intrusion detection. In *Proceedings of the 8th ACM on Cyber-Physical System Security Workshop* (pp. 73-84).

Umer, M. A., Junejo, K. N., Jilani, M. T., & Mathur, A. P. (2022). Machine learning for intrusion detection in industrial control systems: Applications, challenges, and recommendations. *International Journal of Critical Infrastructure Protection*, *38*, 100516.

Pinto, A., Herrera, L. C., Donoso, Y., & Gutierrez, J. A. (2023). Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure. *Sensors*, *23*(5), 2415.

Yang, Y., Gu, Y., & Yan, Y. (2023). Machine learning-based intrusion detection for rare-class network attacks. *Electronics*, *12*(18), 3911.

MR, G. R., Ahmed, C. M., & Mathur, A. (2021). Machine learning for intrusion detection in industrial control systems: challenges and lessons from experimental evaluation. *Cybersecurity*, *4*(1), 27.

Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al-Nemrat, A., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. *Ieee Access*, *7*, 41525-41550.

Gaber, T., Awotunde, J. B., Folorunso, S. O., Ajagbe, S. A., & Eldesouky, E. (2023). Industrial internet of things intrusion detection method using machine learning and optimization techniques. *Wireless Communications and Mobile Computing*, *2023*(1), 3939895.

Al Lail, M., Garcia, A., & Olivo, S. (2023). Machine learning for network intrusion detection—a comparative study. *Future Internet*, *15*(7), 243.

Huma, Z. E., Latif, S., Ahmad, J., Idrees, Z., Ibrar, A., Zou, Z., ... & Baothman, F. (2021). A hybrid deep random neural network for cyberattack detection in the industrial internet of things. *IEEE access"*, *9*, 55595-55605.

Latif, S., Driss, M., Boulila, W., Huma, Z. E., Jamal, S. S., Idrees, Z., & Ahmad, J. (2021). Deep learning for the industrial internet of things (iiot): A comprehensive survey of techniques, implementation frameworks, potential applications, and future directions. *Sensors*, *21*(22), 7518.