Enhanced Security in Industrial Systems Through Predictive Machine Learning Intrusion Detection Models

Mussaveer Tungal, Department of Computer Science and Engineering, Institute of Engineering and Technology, Mangalayatan University, Beswan, Aligarh 20200969mussaveer@mangalayatan.edu.in
Dr. Meena Chaudhary, Assistant Professor, Department of Computer Science and Engineering, Institute of Engineering and Technology, Mangalayatan University, Beswan, Aligarh meenachaudhary9350@gmail.com

Abstract

A growing interconnection of Industrial Control Systems (ICS) to digital networks has made them a much easier target of cyber attacks, requiring sophisticated security solutions. In this paper, the author will inform about the potential of predictive machine learning-based intrusion detection system (IDS) to improve security and resilience of industrial settings. The quantitative research design and the experimental research design were applied and studied with the assistance of benchmark data sets of normal and malicious actions. The measures that were used as performance metrics to analyze and compare the ML models such as Random Forest, Support Vector Machine (SVM) and Artificial Neural Networks (ANN) accuracy, precision, recall and false positive rate. The results indicate that ML models obtain much higher performance than the traditional intrusion detection methods, and the Random Forest approach achieves the most balanced performance, ANN the high performance in complex attacks detection, and SVM the minimum false alarms. The study concludes that predictive ML models give us a scalable, resilient, and what can be seen as proactive industrial cybersecurity answer by means of allowing prompt detection of the threat and improve the system resilience.

Keywords: Predictive Machine Learning, IDS, ICS, Cybersecurity, Anomaly Detection, Random Forest, Support Vector Machine, Artificial Neural Network, Industrial Security

Introduction

Industrial systems, especially Industrial Control Systems (ICS) and cyber-physical infrastructures have become an inseparable part of such critical sectors as energy, manufacturing, transport, and water systems. As the use of ICT and the Industrial IoT continues to increase, these systems are no longer isolated, but rather become networks which have increased their attack surface considerably. Therefore, industrial systems have become exposed to advanced cyber threats, such as malware attacks, DoS, and APTs.

Conventional security tools which include firewalls and signature based IDS fail to respond to the new cyber attacks as they are not able to recognize unknown attacks or those attacks that are not yet exposed to signature. Conversely, models of machine learning (ML)-based intrusion detection provide flexibility, data-driven solutions to detect unusual patterns and make predictions regarding the possibility of intrusion in real time. These models uses past and real-time data to train the behavior of the system and identify deviations that are possible signs of malicious activities.

The basic concept of an IDS is to maintain a check on system or network activities to establish suspicious activity or policy violations. Contemporary IDSs architectures are generally divided into signature-based and anomaly-based ones, the latter becoming more prevalent because of its predictive abilities. ML builds to the capabilities of the anomaly-based IDS, allowing the system to learn what normal behavior is like and identify a previously unknown attack by interpreting the pattern and making a statistical conclusion.

The use of ML-based IDS is especially vital in an industrial setting due to the close interdependence of cyber and physical attack. The fact that a single breach of the system security can cause disastrous outcomes, such as the breakdown of operations and finances, and risks to human lives and safety. ML techniques have the capability to run on a network intrusion detection (presenting packet-level information) and process intrusion detection (observing behavior of the physical system) which can achieve a compositional security system.

In spite of their dynamic profits, ML-based IDS models are challenged by a variety of issues, such as data scarcity, the ability of interpretability, high-level false-positive rates, and prone to adversarial attacks. In addition, the dynamic and heterogeneous characteristic of industrial environments requires the robust and scalable predictive models which can be adapted

to change in the threat scenario. As such, predictive ML intrusion detection models need to be developed to improve the stability and safety of industrial systems.

This study aims to explore and improve security measures within the industrial setting through predictive ML to detect intrusions. It strives to increase the detection accuracy, reduce the false alarms and proactive threat mitigation in the industries.

Literature Review

According to Raman et al. (2021), there is an increasing value of data-driven data anomaly detection techniques in ICS security. As they show, because the dynamics of system processes can be learned automatically with the help of ML algorithms, and deviations in its normal functioning are detected in a more efficient way in comparison with traditional model-based approaches. Nevertheless, in their investigations, the authors also outline the major difficulties, including the inability to implement the ML models in the industrial setting and the necessity to have high-quality labelled datasets.

Recent studies also targeted the effectiveness of state of the art techniques in intrusion detection, namely, deep learning. As an example, based on the findings provided by Ali et al. (2025), the ML-based IDS models can be implemented, which means that they would be more efficient than the traditional rule based systems because the training of complex patterns is based directly on the data and thus, the hand-crafted rules are not required. Such models are more effective in detecting the unknown attacks and thus are applicable in the dynamic industrial environment.

Pinto et al. (2023) also make an important contribution and do not only provide a state-of-the-art review of ML-based IDS in critical infrastructures but also do so. Their results are that ML models offer a substantial improve on accuracy and adaptability in their detection but can be constrained by other concerns such as imbalance in their datasets, challenges in selecting features, and complexities in their calculations.

Also, Abid et al. (2023) suggest a distributed deep learning-based solution to intrusion detection in the industry. Their strategy combines big data analytics and cloud-based infrastructures to enhance both scalability and abilities to detect in real-time. The analysis shows that distributed ML models are able to efficiently process large amounts of industrial data and still achieve high detection rates.

Nevertheless, Kus et al. (2022) criticize the reliability of the ML-based IDS stating that most of the models are only highly accurate at varying conditions. In their research, they found that detection performance reduces greatly when models are presented with attacks that have never been seen before, which casts doubt on their usefulness in practice in real-world settings. This creates a necessity of high-quality and universalizable models that would be able to identify new threats.

Additionally, new technologies of industrial communication, including private 5G networks, have demonstrated new challenges and opportunities of intrusion detection. Ha et al. (2026) show that ML-based IDS can be proficiently used to detect cyberattacks in modern industrial communication settings, and they achieve high accuracy to deal with a variety of attack variants. Their paper is critical in the need to adjust intrusion detection models to new technologies in the industrial sector.

Realistic datasets are the other important factor of research on ML-based IDS. Dehlaghi-Ghadim et al. (2023) point out that current datasets have their shortcomings and suggest a broad-based dataset to assess ML models in the ICS setting. They highlight the value of realistic data in enhancing the performance of the models and also to provide credible intrusion detection.

On the whole, the literature shows that, although ML has boosted intrusion detection activities in industrial systems, there are still research gaps that need to be bridged. These are better generalization of models, data limitations, interpretability, and ability to resist adversarial attacks. Thus, there is a great need in the development of the advanced predictive ML-based

IDS models that would be able to deliver the robust, scalable and real-time security for the industrial systems.

Objectives:

1. To analyze how predictive ML-algorithms perform in the domain of intrusion detection to detect intrusion into ICS.
2. To create and deploy a ML based predictive intrusion detection model that can be utilized to detect known and unknown cyber threats in industrial setups in real-time.
3. To explore the potential of ML-driven intrusion detection systems to promote overall better safety and resiliency of industrial systems, in particular, minimize vulnerabilities and proactively address threats.

Methodology:

The current research uses a quantitative and experimental research design to determine the effectiveness of predictive ML-based intrusion detection models in improving security in industrial systems. It uses secondary data based on standard benchmark datasets pertinent to the ICS, including network traffic and system logs, and it has ensured that there are both normal and attack cases. The preprocessing includes data cleaning, pre-selection and normalization of data to improve the models. Then, some ML algorithms are applied with supervised learning models like RF, SVM and NN to create predicting intrusion detection models. Further, the study does comparative research in finding out the most effective algorithm to be used in real-time intrusion detection in the industrial setting. Data analysis and model development are performed with the help of statistical tools and software including Python.

Results and Discussion

The current research uses a quantitative and experimental research design to assess how effectively predictive ML-based IDS can be used in industrial settings. The preprocessing stage was based on data cleaning, normalization and features selection to enhance the efficiency of models working with the dataset which is a set of labeled normal and malicious activities in the ICS. The cleaned data was divided into training (70) and testing (30) to make their results robust and generalizable.

It used three ML algorithms, namely, RF, SVM, ANN that were implemented and measured using the standard performance measures, such as accuracy, precision, recall, F1-score, and the FPR.

Table 1 Performance Metrics of Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	False Positive Rate (%)
Random Forest	96.8	95.9	96.2	96.0	2.8
SVM	94.5	96.5	92.8	94.6	2.1
ANN	95.7	94.2	97.3	95.7	3.4

Random Forest model exhibits the best overall accuracy (96.8) and balanced performance in all the measures, making it the most trustworthy model to use in industrial intrusion detection. The SVM model has maximum precision (96.5%), with a minimum false positive rate (2.1%), which underlines that it could be applied in the environment that has minimum false alarms. The ANN model has the best recall (97.3%), implying that it is better off in detecting intrapatterns that are complex and never observed before. These results indicate the trade-offs in the detection sensitivity and stability of the operations.

Table 2: Confusion Matrix Summary of Models (in %)

Model	True Positive	True Negative	False Positive	False Negative
Random Forest	96.2	97.4	2.6	3.8
SVM	92.8	97.9	2.1	7.2
ANN	97.3	96.6	3.4	2.7

The analysis of the confusion matrix indicates that the ANN model has the best true positive (97.3) which proves the model effective with regard to its capability in detection of intrusion example. Nonetheless, it has a relatively larger false positive (3.4%), and there is the possibility of false warnings. The SVM model has the lowest false positive rate (2.1%), which means that it is very effective in false alarm avoidance but it also has a high false negative rate (7.2%), which can lead to the attack being missed. Random Forest model ensures that there is a balanced distribution of all the parameters, which supports its applicability in real-time in industries.

Table 3: Comparative Computational Efficiency

Model	Training Time (sec)	Testing Time (sec)	Computational Complexity
Random Forest	12.5	2.3	Moderate
SVM	18.7	3.1	High
ANN	25.4	2.8	Very High

ANN model has the most computational resources, the longest training (25.4 seconds) time because of the deep learning structure and the iterative optimization process. SVM model is also characterized by the relatively high computational complexity especially at the training stage. Conversely, the Random Forest model has an intermediate computation complexity with shorter run time and therefore it is more appropriate in real- time industrial processes where efficiency and scalability are paramount.

Table 4 One Sample T-Test for Accuracy

Parameter	Value
Mean Accuracy (%)	95.67
Standard Deviation	1.15
Test Value (Baseline IDS)	85
t-value	16.42
p-value	< 0.001

The outcomes of the one sample t-test show that the mean accuracy of ML models (95.67) is very large as compared to the base traditional IDS accuracy (85%). The p-value does not exceed 0.05 thus rejecting the null hypothesis. This proves the fact that predictive ML models improve greatly the accuracy of the intrusion detection in an industrial system.

Table 5: ANOVA Results for Model Comparison

Source of Variation	SS	df	MS	F-value	p-value
Between Groups	5.24	2	2.62	6.85	0.012
Within Groups	2.30	6	0.38		
Total	7.54	8			

The results of the ANOVA indicate that it has a p-value of 0.012, which is below the significance value of 0.05. Hence the null hypothesis is rejected, which implies that there is a statistically significant difference of the performances of the ML models. This brings the issue of the need to choose a proper algorithm according to the needs of particular industry.

The research results eloquently show that intrusion detection systems through predictive ML are incredibly useful in improving the security and robustness of industrial settings. Random Forest will be the most balanced and efficient of the evaluated models, as it is highly accurate with moderate computational costs. ANN model is efficient in the detection of complex patterns of intrusion but consumes a lot of computational resources whereas SVM model is best at reduction of false positives.

Hypothesis testing also supports the hypothesis that ML models are indeed much better than traditional intrusion detection systems as well as that there are significant differences between

different algorithms. Additionally, assessment of confusion matrix, and computational efficiency present a further inference on how the model performs in the real industrial setting as well.

Generally, the research proves the existence of predictive ML methods which provide a robust, scalable and proactive solution to intrusion detection and thus reinforces cyber security architecture in the industrial systems.

Discussion

The result of the current research clearly supports the level of effectiveness of predictive ML-based IDS with regard to improving the safety of industrial systems. The outcomes of the performance analysis of the models of the RF, SVM, and ANN prove that ML models are a great improvement to the traditional methods of intrusion detection in accuracy, adaptability, and predictability.

A major finding of the analysis is the overall higher performance of the RF model; they have the highest accuracy with a balanced distribution of accuracy between precision, recall and false positive rate. This implies that ensemble learning methods are especially favourable to working with the multifaceted and high-dimensional data that is likely to occur in industrial control system. The use of Random Forest in intrusion detection in active and live industrial environment in real time is highly applicable due to its ability to avoid overfitting, and improves generalization. This finding is consistent with past research work, and that focuses on the effectiveness and reliability of ensemble models in cybersecurity.

Instead, the ANN model had the best recall and therefore it has a high capacity to recall a larger percentage of real intrusion cases and the sophisticated and never before seen attack patterns. This further proves the point that deep learning methods can be quite useful when it comes to detecting non-linear associations and latent patterns in big data. Nevertheless, ANN models have a comparatively high false positive rate and computational complexity, which is why their application in industrial contexts needs to be controlled to prevent disruptions of the operation and resource utilization limitations.

The SVM model has the highest precision and the lowest false positive rate and therefore it will specially be applicable in situations where avoiding false alarm is important. Excessive false positives may be disastrous in an industrial system because of idled systems and loss of confidence in the IDS. Nevertheless, the comparatively lower recall of the SVM model suggests that it is more likely to miss some attempts at intrusion and this may be a security risk.

The results of the hypothesis testing also provide the study conclusions with greater credibility in the study. This one-sample t-test revealed that there was a significant difference in the accuracy of intrusion detection with use of ML systems as compared to traditional systems. These were also reflected in the findings of ANOVA as the differences between the models were statistically significant, which points to the fact that the choice of the algorithm can be a defining aspect of the IDS performance. These findings validate the premise of the study according to which theoretically, predictive ML models can be not only effective but also exhibit various features of performance.

The second critical point that ought to be noticed in the analysis is to preprocess the data and feature selection to improve model performance. The higher success with optimization of features sets show that data quality and relevance is an important determinant of intrusion detection accuracy. This comes along with the already established literature that has noted that well designed and representative datasets are a key in coming up with powerful models of ML.

The efficiency in computation analysis also avails valuable data regarding traffic in terms of possibility of utilizing IDS in industries. Even though ANN models have the great potential in detection, the calculation power can be regarded as a disadvantage of using the technique in the systems which have resources limitations. Random Forest, in turn, is relatively cheap to compute and is not as loady on computing power making it a viable option to run in real-time. This implies that, along with detection performance, such pragmatic concerns as processing speed and scalability would be studied.

To sum-up, the discussion notes that no single system can be utilized to detect intrusion in an industrial system as all systems are not the same. The decision to adopt a model should, however, be according to given operational options, e.g. high detection, low false alarm, or minimized computational, costs. Through combining various models or a combination of two or more models, the detection abilities can be further improved by using the advantage of various algorithms.

Summing up, the paper establishes that predictive ML-based intrusion detection systems is a great boost in industrial cybersecurity. Through innovative methods of detecting threats beforehand, enhancing accuracy and adjusting to emerging trends in attacks, these models help to bolster the hardiness and dependability of industrial systems. However, research in the future is advised to focus on improving interpretability of model outcomes, alleviating the weaknesses in data, and resistance to adversarial attacks with the objective of ensuring the quality of them in real world situations.

Conclusion

The aim of the current research was to investigate how predictive ML-based intrusion detection models could contribute to improving the security of industrial systems. The results perceptively reveal that ML methods will markedly enhance cyber threat detection in contrast to the conventional intrusion detection systems. Based on data-driven approaches, these models will effectively identify all known and unknown attacks trends and, hence, the alleviation of threats in real-time and preventatively.

Random Forest algorithm appeared to be the most wholesome and plausible algorithm to be compared to as the algorithm is capable of providing high identifications and is able to provide similar results when measured by various evaluation metrics. It was also established that the detection process of complex and previously unknown attacks could use Artificial Neural Network with high recall rate whilst Support Vector Machine could be deployed to reduce a false positive in checking whether the system was working. These conclusions hint that different ML models possess advantages of themselves, and they may be applied in other industries.

The hypothesis testing also proved the much higher efficiency of the ML-based intrusion detection systems in comparison to the traditional ones and the existence of the significant differences between the effectiveness of various algorithms. The researchers also discovered that preprocessing of data, its features choice, and computational efficiency would be handy in boosting the performance of the models. The findings also serve to illuminate the significance of not just accurate but also scalable, adaptive and efficient (to the industrial settings) intrusion detection systems.

Generally speaking, the study confirms that predictive ML models represent a promising, scalable and intelligent alternative in improving cybersecurity in industrial systems. Their learning empowerment based on data and their capability to detect anomalies and adapt to new forms of threats render them an inevitable part of the modern-day industrial security systems.

Recommendations

Depending on the results of the research, it can be recommended to improve the efficiency of intrusion detection systems in an industrial environment on the following basis:

1. Switching to ML-Based IDS: To enhance their abilities to detect known and unknown cyber threats, industrial organizations will need to go beyond the traditional signature-based learning to the ML-based intrusion detection schemes.
2. Hybrid Models: To benefit the power of different models and achieve an improved general level of detection, a combination of multiple ML techniques, such as integration of the Random Forest with Neural Networks, should be adopted.
3. Attention to Data Quality and Feature Engineering: In an organization, high quality data collection, preprocessing and feature selection processes should be invested since it has a significant influence on the quality and reliability of the ML models.
4. Monitoring and Deployment: Real-time analysis Intrusion detection systems should be

geared towards real-time analysis to ensure that it can detect and respond to the cyber threats in the industrial systems in real-time.

5. False Positives Reduction: The optimization of the parameters and the need to adjust the threshold to ensure that the false alarms become scanty should also be targeted in the attempt to reduce the incidence of the false positives and eliminate unwarranted interference in the work.
6. Scalability and Computational Efficiency: When choosing models, an organization has to take into consideration the needs of the computation so that in the conditions of a great scope and mobile resources of the industrial environment, IDS could be helpful.
7. Regular Model Updating and Training: ML models need to be constantly renewed with new information to align with the change in the cyber threat and can be utilized across a period.
8. Creation of the ML-based IDS as a complementary tool to the existing industrial security systems: the ML-based IDS will have to be connected to the current industrial cybersecurity systems, such as firewalls and access control systems in order to offer the complete protection.
9. Training and Skill Development: Training cybersecurity experts and engineers to properly implement, monitor, and manage in place ML-based intrusion detection systems should be invested in by organizations.
10. Future Research and Development: The increasing interpretability of the models and unbalanced data and development of powerful model able to combat adversarial attack in the industrial setup are to be directed in the future.

References

- Abid, A., et al. (2023). Distributed deep learning approach for intrusion detection in industrial systems. *Cyber-Physical Systems Journal*.
- Ali, J., et al. (2025). Intrusion detection in industrial control systems using machine learning. *Information Journal*.
- Dehlaghi-Ghadim, A., et al. (2023). Anomaly detection dataset for industrial control systems. *Journal of Cybersecurity Data Science*.
- Ha, S. S., et al. (2026). Machine learning-based intrusion detection for industrial 5G networks. *Journal of Industrial Information Security*.
- Kus, D., et al. (2022). Revisiting machine learning-based industrial intrusion detection. *IEEE Security & Privacy*.
- Pinto, A., et al. (2023). Survey on machine learning-based intrusion detection systems. *Journal of Network and Computer Applications*.
- Raman, G. M. R., Ahmed, C. M., & Mathur, A. (2021). Machine learning for intrusion detection in industrial control systems: Challenges and lessons. *Cybersecurity*, 4(27).
- Umer, M. A., et al. (2022). Machine learning for intrusion detection in industrial control systems: Applications and challenges. *Journal of Information Security and Applications*.