Denoising Adaptive - Multi-Branch Architecture for Detecting Cyber Attacks in Industrial Internet of Services

Ghazia Qaiser * and Sivachandran Chandrasekaran

Posted Date: 26 November 2025

doi: 10.20944/preprints202511.2024.v1

Keywords: cyber-attacks; deep learning-based hybrid model; denoising autoencoders; IIoS

Article

Denoising Adaptive - Multi-Branch Architecture for Detecting Cyber Attacks in Industrial Internet of Services

Ghazia Qaiser * and Sivachandran Chandrasekaran

Swinburne University of Technology

* Correspondence: 103357408@student.swin.edu.au

Abstract

The emerging scope of the Industrial Internet of Services (IIoS) requires a robust intrusion detection system to detect malicious attacks. Numerous robust and sophisticated cyberattacks have resulted in financial losses and catastrophes in IIoS-based manufacturing industries. However, existing solutions often struggle with adaptability and generalizability to emerging and sophisticated cyber-attacks. This study proposes a unique approach for adaptive deep learning pipelines tailored for zero-day network attack detection in Industrial Internet of Services (IIoS) environments. The proposed approach merges a denoising autoencoder (DAE), enhanced by an adaptively tuned noise factor, with a hybrid model based on MLP + BiLSTM named as DA-MBA (Denoising Adaptive - Multi-Branch Architecture). The DAE filters noise and learns robust representations, making subsequent classification less sensitive to benign concerns. Moreover, it addresses challenges, such as adaptability and generalizability, by hybridizing a Multilayer Perceptron (MLP) and bidirectional LSTM (BiLSTM). The proposed hybrid model was designed to fuse feed-forward transformations with sequence-aware modeling, which can capture direct feature interactions and any underlying temporal and order-dependent patterns. Multiple approaches have been applied to strengthen the dual-branch architecture, such as SMOTE-based oversampling, class weighting, and comprehensive hyperparameter optimization via Optuna, which collectively addresses imbalanced data, overfitting, and dynamically shifting threat vectors. The proposed DA-MBA is evaluated on two widely recognized IIoT-based datasets, *Edge-IIoT set* and WUSTL-IIOT-2021, and achieves over **99% accuracy** and a near **0.02 loss**, underscoring its effectiveness in detecting the most sophisticated attacks. The solution offers a scalable, flexible architecture for enhancing cybersecurity within evolving IIoS environments by coupling feature denoising, multi-branch classification, and automated hyperparameter tuning. The results confirm that coupling robust feature denoising with sequence-aware classification can provide a scalable and flexible framework for improving cybersecurity within the IIoS.

Keywords: cyber-attacks; deep learning-based hybrid model; denoising autoencoders; IIoS

1. Introduction

Industry 4.0 is rapidly evolving to create a highly digital environment that will soon become a part of our daily lives. The IIoS, along with the IIoT, is a significant pillar of Industry 4.0, adding value and contributing to human interests. The IIoS is a critical component of Industry 4.0, similar to the IIoT. The IIoT acts as the "eyes" of a system, providing visibility through connected sensors and devices that collect and transmit data. In contrast, the IIoS functions as the "brain," interpreting data and providing intelligent services to improve industrial operations. The switch from traditional networks to IoT has recently revolutionized the global economy. IIoS infrastructure can support smart manufacturing, agriculture, health, government, cities, logistics, and home automation. The IIoS is a new concept that is still emerging. IIoS is a way to connect products and services to work

 doi:10.20944/preprints202511.2024.v1

together more seamlessly for humans. IIoS is a new concept that is still emerging and ultimately allows products and services to work together more seamlessly.

Figure 1 depicts the overall architecture of Industry 4.0, which blends the **Industrial Internet of Things (IIoT)** and **cloud computing** under an **Industrial Internet of Services (IIoS)** framework that offers more smart and intelligent services such as smart manufacturing systems. In Industry 4.0, incoming data from smart connected devices and sensors are transmitted via IIoT to cloud/edge-based platforms, which enables advanced analytics, real-time monitoring, and centralized control. Organizations can optimize processes, improve efficiency, and rapidly adapt to changing market or operational demands by layering service-oriented technologies, such as smart manufacturing and intelligent service delivery, atop this infrastructure

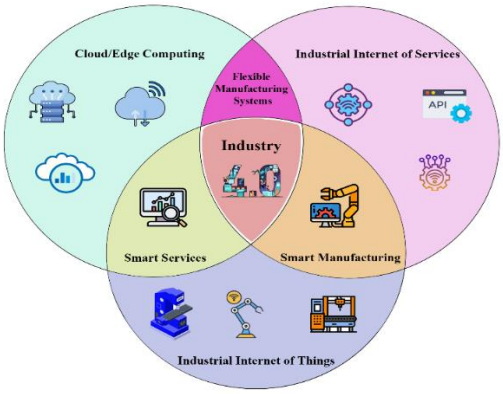


Figure 1. Industry 4.0 Architecture.

According to a recent statistical report, the market value of IoT was $1.90 billion in 2018, $25 billion in 2020, and $925.2 billion as of 2023. It is forecasted to hit $6 trillion in 2025, which makes the **compound annual growth rate (CAGR)** from 2018 to 2025 approximately 15.12%. The report declared that 25 billion devices were connected before 2020, with 50 billion permanent connections and over 200 billion intermittent connections. The report also projects that 29.4 billion by 2023 [1]. These trends suggest that IoT adoption and investment continue to accelerate, with manufacturing poised to capture a significant share. Intel projected that the market value for IoT could hit 6.2 trillion dollars by 2025, and a substantial percentage of it is in manufacturing [2,3]. The growing trends of IoT can establish that rapid growth and high connectivity underscore the increasing importance of the IIoS. Factories adopt more sensors, edge devices, and data-driven solutions, which provide opportunities for advanced analytics, real-time monitoring, and predictive smart services. These projected shifts increase operational efficiency and assemble new revenue models that ultimately transform manufacturing processes through connectivity-based industrial service paradigms.

The IIoS can underpin numerous smart manufacturing services and operations, connecting machines and sensors, and control real-time systems. However, they are still vulnerable to cyber-attacks that can disrupt production lines, damage machinery, or even endanger human operators. An unsecured IIoS architecture risks costly downtime, breaking the supply chain, and theft of sensitive intellectual property (IP). Attackers manipulating complex networks can vandalize manufacturing runs, produce inadequate products, or terminate operations, leading to significant financial and reputational losses. Therefore, robust cybersecurity measures are essential for maintaining smooth, safe, and resilient industrial operation.

Although intrusion detection within the industrial ecosystem has progressed in recent years, existing solutions often rely on static architectures that struggle to adapt to rapidly evolving, sophisticated threats, and typically neglect inherent data imbalance problems that impact the overall performance. In industrial landscapes, benign network traffic extensively surpasses malicious events, making it crucial to address data imbalances such that rare yet highly consequential attacks are not overlooked. Moreover, industrial data contain problematic noise generated by industrial equipment,

 doi:10.20944/preprints202511.2024.v1

sensors, and complex network topologies, which can degrade the performance of traditional machine-learning methods. Similarly, IIoS/IIoT-enabled industries can produce random noise in operational data generated by robotic arms, sporadic sensor calibrations, and inconsistent power supply, which can cause disruptions in conventional machine learning-based cyber defense techniques because conventional machine learning models are highly dependent on data. Furthermore, traditional solutions often struggle with dynamic conditions, lacking mechanisms to adapt swiftly as new threats emerge and underlying data shifts.

Considering these cybersecurity challenges, the proposed DA-MBA model offers a solution that can clean data from corrupted inputs by addressing the challenge of noise and shifting data distributions by learning to reconstruct. In smart industries, there is a chance of sensor malfunctions, environmental disturbances, and complex network protocols that can introduce anomalies that degrade the reliability of conventional machine-learning classifiers. The proposed solution overcomes this problem by training the autoencoder on noisy samples created via Gaussian noise, which learns a compressed and more consistent representation of network traffic that highlights essential features while cancelling out noise. This DAE process yields a feature space that is significantly more robust to changes and irregularities in the data, and allows the classifier to focus on meaningful patterns rather than spurious fluctuations. The proposed model, DA-MBA, is further described as more adaptive, with appropriate hyperparameter tuning using the Optuna framework and suitable regularization, allowing the model to keep pace with the dynamic nature of the threats. After feature extraction and making the incoming feature space more appropriate, the pipeline employs a hybrid-pronged classification, in which the MLP can distinguish direct and non-sequential relationships between the denoised features. Moreover, MLP can quickly learn global patterns and correlations characteristic of malicious behavior, which can be critical for detecting attacks that exhibit clear patterns of anomalies. However, BiLSTM on the other hand, handles the sequential and contextual aspects of data. Recently, the signatures of many attacks have evolved, even if the data points are not strictly chronological, such as the escalation of suspicious ports and the increase in abnormal traffic over time. Sometimes, unfolding and identifying interrelated patterns is very difficult, but BiLSTM's recurrent structure is well suited for this task. The model integrates the outputs of both MLP and BiLSTM to provide immediate feature-level insights with sequential and context-based insights, resulting in a more comprehensive view of potential cyber-attacks.

The proposed model is also integrated with automated hyperparameter search, dropout, L2 regularization, and class imbalance handling with SMOTE and class weighting, which allows the hybrid classifier to adapt to shifting conditions corresponding to retraining and refining its parameters as new threats or data distributions emerge. This proposed flexibility is crucial for staying effective in continuously evolving industrial networks, where complex and sophisticated exploits and rare attack variations can rapidly sabotage static detection methods. The proposed model was trained and tested on two widely recognized IIoT-based datasets, ***Edge-IIoTset*** and WUSTL-IIOT-2021, and achieved over **99% accuracy** and a near **0.02 loss**, underscoring its effectiveness in detecting the most sophisticated attacks.

The proposed study incorporates multiple regularization strategies throughout the training process, which makes the model immune from overfitting and maintains a robust performance on unseen data. The first step is to apply dropouts and L2 regularizations in the key layers to penalize large weights and minimize co-adaptation among neurons. Furthermore, the DAE intervenes by adding Gaussian noise, forcing the encoder to learn fundamental feature representations rather than memorizing noisy and temporary patterns. The Gaussian noise process injects the model more adaptively to evening cyber-attacks, which can reshape its behavior to exploit conventional machine learning methods. The study also employs early stopping based on validation loss, halting training once performance no longer improves, and preventing the DA-MBA from overfitting to erroneous fluctuations in the training set. Moreover, automated hyperparameter optimization via the Optuna framework further refines the dropout rates, learning rates, and encoding dimensions of the network to strike an optimal balance between capacity and generalization. The datasets were split into

 doi:10.20944/preprints202511.2024.v1

training, validation, and test sets to ensure an unbiased performance estimation and robust hyperparameter tuning. Furthermore, the DA-MBA was statistically validated by comparing the training and test accuracy, loss, F1 scores, and ROC-AUC curves.

1.1. Need for Specialized Cyber Attack Detection in IIoS

The Industrial Internet of Services (IIoS) represents the convergence of advanced technologies and industrial systems, creating a paradigm shift in how industries operate and deliver services. Figure 2 describes the overall architecture along with five layers as perception layer, network layer, service layer, application layer and business layer.

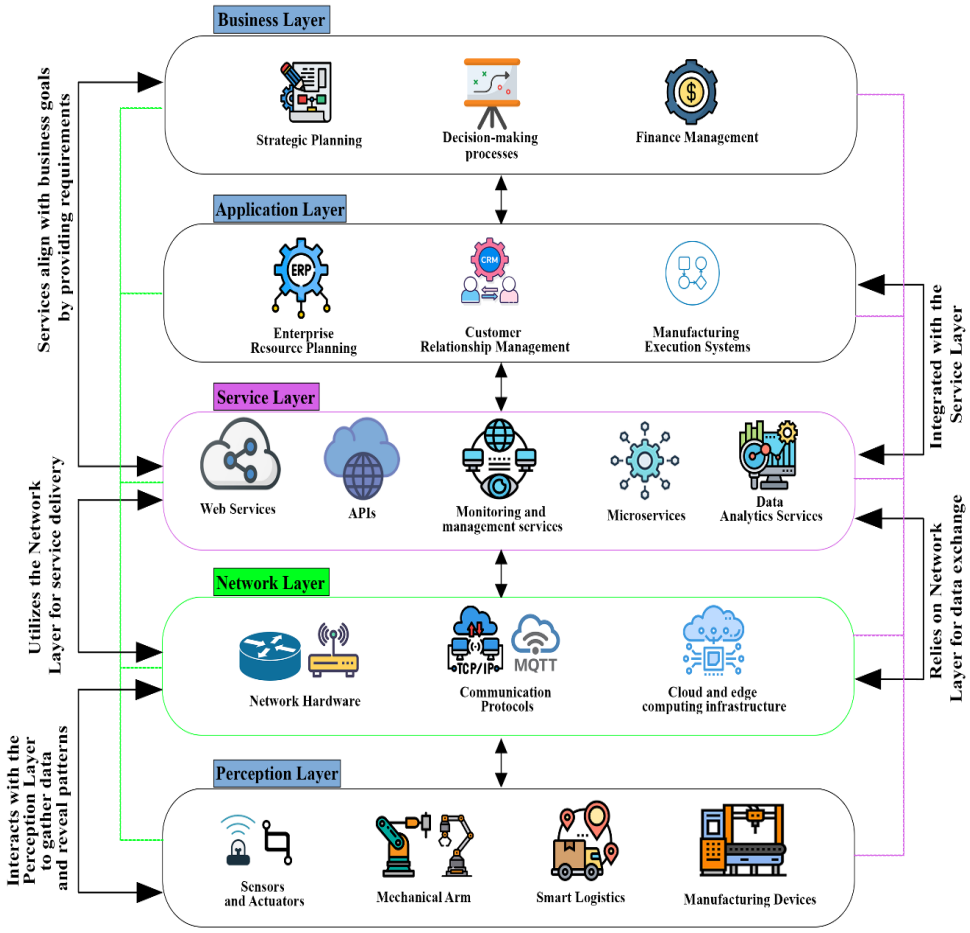


Figure 2. Architecture of Smart Manufacturing.

In Industry 4.0, numerous enterprises are adopting layered industrial architectures to influence the synergy between the IIoT and IIoS. As illustrated in Figure 2, the perception layer is surrounded by field-level devices and sensors for data capture. The perception layer hosts a network of sensors, actuators, and smart devices, continuously streaming precise data regarding the equipment status, environmental conditions, and operational process variables. However, the network layer comprises communication protocols, routers, and gateways that securely transmit collected information. The network layer, where low-latency communication protocols, edge computing nodes, and smart network hardware solutions, such as intelligent routers and next-generation firewalls, ensure secure and efficient data routing. Networked data then travel to the service layer based on the IIoS, where cloud-based services, virtualized microservices, and adaptive computing resources facilitate advanced analytics, real-time monitoring and reporting, and on-demand scalability. Moreover, the application layer merges sophisticated services from the service layer with end-user dashboards, SCADA systems, and machine-to-machine APIs, thereby promoting remote visibility, advanced

 doi:10.20944/preprints202511.2024.v1

process automation, and automated decision-making. Eventually, the business layer relies on consolidated analytics to optimize production procedures, improve supply chain management, and discover novel remuneration streams.

A complete smart industrial architecture requires robust cybersecurity measures across all the layers. However, the IIoS imposes additional challenges that necessitate specialized network-based cyberattack detection solutions. Compared to conventional industrial networks, IIoS relies heavily on cloud-based smart services, as it fundamentally provides industrial services through cloud platforms. The IIoS offers virtualized micro-services and dynamic resource configurations in an industrial architecture, often transiting multiple geographic regions and heterogeneous complex network domains. Moreover, a complex ecosystem of IIoS invites complicated communication patterns and a broader attack surface, making IIoS vulnerable to cyber threats such as distributed denial-of-service (DDoS), supply chain vulnerabilities, and man-in-the-middle (MITM) attacks. Furthermore, the IIoS merges real-time data feeds from sensors and smart edge devices, making anomalies more challenging to detect within high-volume traffic. Implementing a network-based intrusion detection solution that incorporates technologies such as deep packet inspection, behavioral analytics, and machine learning-driven anomaly detection can enable industries to continuously monitor critical data flows, isolate suspicious activity, and adjust to emerging cyber exploits.

This paper is organized as follows: Section 2 describes the literature. Section 3 discusses the research methodology. Section 4 discusses the results and findings of this work, and finally, Section 5 concludes the work presented in this paper and gives recommendations for future work.

2. Literature Review

The rapid expansion of the IIoS, where modern service-centric architectures and industrial processes assemble, has significantly increased the connectivity and complexity of modern industrial ecosystems. Despite previous research on cybersecurity in industrial contexts that has predominantly focused on IIoT, the unique operational dynamics of IIoS introduce additional attack vectors and vulnerabilities that necessitate more sophisticated defense methods. For example, service-level reliances, real-time resource management, and cross-domain data interactions significantly increase the attack surface and create new opportunities for intruders to launch sophisticated exploitation. Researchers are working on continual learning and transformers, but conventional pipelines are still inadequate for evolved and zero-day cyber-attacks. In the following section (Table 1), key advances in intrusion detection methods are critically reviewed, highlighting their contributions and results. Existing studies have highlighted the effectiveness of Machine Learning, Deep Learning, and Hybrid ML/DL heuristic methods for intrusion detection in IIoT infrastructures. Taking advantage of the above-mentioned heuristic approaches, [4] systematically investigated DL approaches and explored how convolutional autoencoders (CAEs) can be utilized to enhance side-channel attacks, which are a critical threat vector in cryptographic systems. The authors compared multiple CAE architectures and hyperparameter configurations, providing a better understanding of the design choices that best capture and compress high-dimensional patterns without compromising critical leakage information and detecting network-based cyber-attacks. However, despite the promising results of this research, a principal limitation of this study is its focus on a relatively constrained set of experimental settings (e.g., a single device type or single cryptographic algorithm), which may limit the broader applicability of the findings. Another study [5] presented a more robust approach by proposing an adaptive cyberattack prediction framework that integrates an enhanced genetic algorithm (GA) with deep learning (DL) techniques to dynamically identify and respond to emerging cybersecurity threats. The proposed method optimizes both feature selection and hyperparameter configuration settings in a deep neural network (DNN), resulting in improved predictive accuracy and reduced false positives compared with baseline models. However, a key limitation is that this approach, which combines a genetic algorithm and a deep learning pipeline, may become computationally demanding when scaled to large datasets.

 doi:10.20944/preprints202511.2024.v1

Further studies [6,7] have concentrate on automating and optimizing security mechanisms to mitigate sophisticated cyber threats. The author [6] implemented automated machine-learning (AutoML) pipelines specifically for DL-based malware detection, highlighting the reduced manual overhead in setting hyperparameter configurations and selecting the appropriate architecture. Through the systematic exploration of optimal network arrangements, the proposed approach facilitates rapid adaptation to evolving malware variations. In parallel, [7] presented an intrusion-detection framework precisely designed for IoT domains that combines selective feature engineering with lightweight classification to achieve a balance between accuracy and computational efficiency. Although both proposed solutions originate from different applications, they work on the same objective of streamlining model development to generate quicker and more effective responses to cyber hazards. However, questions remain regarding their scalability and adaptability in highly dynamic scenarios. Other studies [8–10] focus on a shared objective: strengthening IoT and IIoT ecosystems against emerging cyber threats through advanced DL techniques. The study [8] highlights hybrid deep learning techniques that combine various neural architectures, thereby improving detection robustness while considering the heterogeneity often inherent in IoT infrastructures. Moreover, [9] disseminated this perspective by proposing a hybrid deep random neural network for the industrial IoT (IIoT), demonstrating how specialized randomization mechanisms can mitigate overfitting and adapt quickly to new attack patterns. In parallel, [10] introduces a distributed attack detection framework powered by a deep learning (DL) architecture, improving scalability by enabling intrusion detection tasks to be distributed and coordinated across multiple nodes in an Internet of Things (IoT) environment. Overall, these studies demonstrate that deep learning strategies, whether hybrid, can yield substantial improvements in the detection of complex cyberattacks in complex, interconnected IoT environments. However, these studies also illustrate the ongoing challenges of computational overhead, real-time flexibility, and evolving nature of threat profiles.

Moreover, further studies [11–13], which try to overcome recent challenges in cyberattack detection [11], propose a DL-based intrusion detection pipeline tailored for IIoT, focusing on high detection accuracy and real-time responsiveness in typically under-resourced operational environments. [12] expanded this perspective by emphasizing robust feature selection and overfitting mitigation through hybrid machine-learning approaches, which are crucial for handling large-scale noisy datasets that can weaken intrusion indicators. Similarly, [13] proposed a hybrid deep learning approach for network security, which highlights the potential of various neural network architectures to improve threat detection rates. Overall, studies have demonstrated that sophisticated learning pipelines can strictly be based on deep learning or hybrid methods and benefit from careful data preprocessing, feature engineering, and computational efficiency parameters. Moreover, studies have demonstrated the persistent need for adaptable, self-optimizing models that can survive evolving cyber hazards, particularly in complex IIoT scenarios, where even minor security breaches can lead to significant operational and financial consequences. Further studies [11–13] investigate more robust architectures to overcome existing research gaps and detect evolved cyber hazards. The author [11] leveraged a CNN to enhance network-based intrusion detection, revealing notable progress in detection accuracy while maintaining manageable computational complexity. Moreover, [12] extended the paradigm by merging neural networks with ML algorithms, effectively capturing network traffic data and detecting cyber hazards. Their proposed model not only improves the detection precision for varying threat types but also supports high performance. Meanwhile, [13] focused on a hyper-tuned, compact LSTM framework for hybrid intrusion detection, aiming to optimize resource utilization without compromising the performance. Together, these studies demonstrate the shift toward specialized and hybrid deep learning models that tackle both the high dimensionality and real-time limitations of complex networks, signifying that further improvements in architecture design, hyperparameter tuning, and scalability may generate even more robust intrusion-detection abilities.

 doi:10.20944/preprints202511.2024.v1

Table 1. Comparison of recent existing studies.

Ref	Dataset	Contribution	Attack Type	Category of Attack Types	Feature Selection	Feature Extraction	AI/ML-based Attack Detection Approach	Hyperparameters Tuning	Hybrid Approach	Accuracy/Results
[4]	ASCAD	Optimization of Convolutional Autoencoders for Side-Channel Attacks	Side-Channel Attack (Power Analysis)	Network-based	-	Convolutional Autoencoder (CAE)	MLP, CNN, Template Attack (TA)	Optuna	No	37% fewer traces needed for attack, reduced trainable parameters by factor of 29
[5]	CICIDS2017, UNSW_NB15	AdacDeep: Enhanced Genetic Algorithm + Deep Autoencoder + DFFNN	Multiple attack types (e.g., DDoS, DoS, Brute Force)	Device-based	-	Deep Autoencoder (DAE)	Deep Feedforward Neural Network (DFFNN)	EGA	Yes, Enhanced Genetic Algorithm + Deep Autoencoder + DFFNN	Improvements in accuracy of 0.22% to 35%
[6]	SOREL-20M, EMBER-2018	AutoML for deep learning-based malware detection in both static and online environments.	Malware detection	Cloud services based	-	Deep learning-driven automated feature extraction	Feedforward Neural Network (FFNN), CNN	TPE	Yes, AutoML combines static, dynamic, and online analysis methods.	On the EMBER-2018 and SOREL-20M datasets, the models achieved accuracies of 95.8% and 99 %, respectively.
[7]	IoTID20, UNSW-NB15	Proposed a hybrid approach combining CNN and GRU to excel in capturing spatial and temporal dependencies in the data.	IoT-related intrusion detection	Network-based	PSO	CNN	CNN-GRU	Grid Search	Yes - CNN-GRU	99.60% accuracy on IoTID20 and 99.14% on UNSW-NB15 dataset.
[8]	Kitsune and TON-IoT	Proposed two hybrid models CNN-LSTM and CNN-GRU to enhance IoT security	Multiple attacks (DDoS, Telnet , password, Injection and backdoor)	Network-based	-	CNN	CNN-LSTM and CNN-GRU	Grid Search	Yes - CNN-LSTM and CNN-GRU	99.6% accuracy on Kitsune and 99.00% TON_IoT dataset
[9]	DS2OS and UNSW-NB15	Proposed a novel hybrid deep random neural network for cyberattack detection in IIoT.	Multiple attacks (DoS, Worms, Scan, Spying and Fuzzers)	Network-based	ARM	DRaNN	HDRaNN	Manual approach	Yes -HDRaNN	98% and 99% for DS2OS and UNSW-NB15
[10]	NSL-KDD and BoT-IoT	Implementing a distributed framework based on deep learning to simultaneously control various sources of vulnerability.	Multiple attacks (DDoS, keylogging, Data theft, U2R and R2L)	Network-based	-	Standard feature selection methods.	FFNN and LSTM	Hyperband	No	Achieved up to 99.95% accuracy across various setups.
[11]	N_BaIoT	Proposes a robust AttackNet model for the	Multiple attacks (DDoS, Malware,	Network-based	-	CNN	CNN-GRU	-	Yes - CNN-GRU	Accuracy of 99.75% across 10 given classes.

 doi:10.20944/preprints202511.2024.v1

		detection of various botnet attacks in IIoT.	MiTM,and Zero day attacks)							
[12]	CSE-CIC-IDS2018	Develop a hybrid model for attack detection that leverages autoencoders for effective feature extraction and DT classifiers to achieve high accuracy and reduce overfitting.	Multiple attacks (DDoS, Malware, MiTM,, phishing, Supply chain and Zero-day attacks)	Network-based	Auto - encoders	LASSO, Random Forest and Boruta	Decision Tree, Naïve Bayes, neural networks, and ensemble methods	-	Yes - Decision Tree, Naïve Bayes, neural networks, Random Forest, and XGBoost	Overall accuracy reached around 94.54%
[13]	N_BaIoT	Proposed a DNN for feature extraction using LSTM to manage sequential data.	DDoS, Mirai and Gafgyt	Network-based	-	Done implicitly by the DNN layers	DNN-LSTM	-	Yes - Deep Neural Network(DNN) and Long Short-Term Memory(LSTM)	99.96%
[14]	NSL-KDD, UNSW-NB15	Proposed hybrid pre-processing method combines PCA and feature engineering via DFS to develop meaningful features for network intrusion detection.	Multiple attacks (DDoS, Malware, MiTM,, phishing, Supply chain and Zero-day attacks)	Network-based	PCA	CNN	CNN, Naive Bayes, Random Forest, Decision Tree, Ada Boost, Bagging	Manual tuning	No	NSL-KDD Achieved 90.14% UNSW-NB15 Achieved 95.7%
[15]	CIC-IDS 2017, UNSW-NB15, and WSN-DS	Proposed a CNN-LSTM Hybrid Deep Learning model for an intrusion detection system that merges the strengths of both algorithms.	Multiple attacks (DDoS, Malware, MiTM, Brute force, Web based, Worms, Blackhole attacks)	Network-based	Select-K-Best	CNN	CNN-LSTM	Manual tuning	Yes – CNN-LSTM	CIC-IDS achieved 99.6%, UNSW-NB15 93.7%, and WSN-DS achieved 99.5%
[16]	IOT23, CICIDS2017, and NSL KDD	Merge long short-term memory (LSTM) and autoencoder (AE) for feature-rich scalable attack detection.	Probe, R2L, U2R, DDoS, Botnet, and HeartBeat	Network-based	PCC	AE	LSTM and AE	Manual tuning	Yes - LSTM and AE	97.7% achieved on NSL KDD dataset, 99% achieved on CICIDS2017 dataset, and 98.7% achieved on IOT23 dataset
Our Study	*Edge-IIoTset* and WUSTL-IIOT-2021	Proposed a multibranch hybrid model based on MLP and BiLSTM	DDoS, scanning, injection, brute force, infiltration, MiTM, and privilege escalation	Network-based	DAE	DAE	Hybrid model based on MLP and BiLSTM	Optuna	Yes – DA - MBA	Almost 99% accuracy on both datasets.

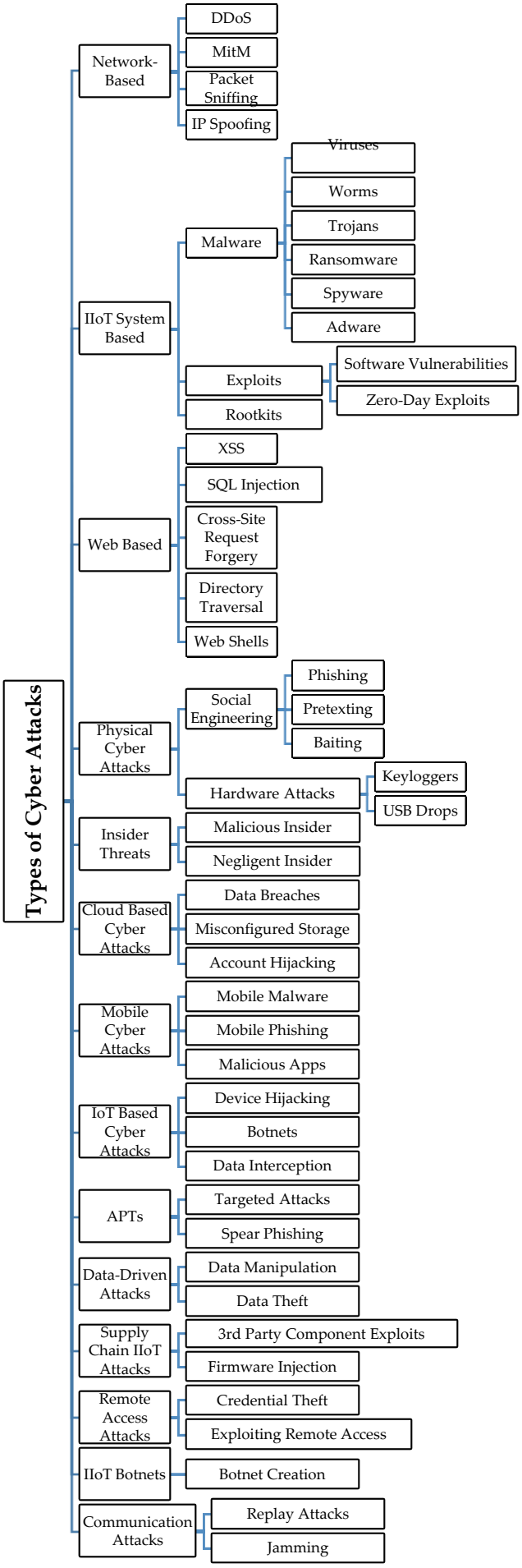
 doi:10.20944/preprints202511.2024.v1

However, despite the substantial benefits offered by the previously mentioned studies in applying deep learning (DL) and hybrid learning pipelines for intrusion detection, these approaches reveal constraints such as exposure to overfitting on noisy, imbalanced datasets, inadequate hyperparameter tuning, and insufficient measurement of real-time inference performance, particularly in resource-constrained IIoT contexts. Alternatively, the proposed model in this study integrates a denoising autoencoder (DAE) to handle noise robustly, employs SMOTE to address the class imbalance, and utilizes bidirectional long short-term memory (BiLSTM) alongside a multilayer perceptron (MLP) for complementary spatial-temporal feature extraction. The study also included automated hyperparameter optimization via the Optuna framework, ensuring that the proposed pipeline adapts effectively to diverse network conditions while minimizing manual intervention. Furthermore, the proposed research systematically measures decision time per detection, and the framework explicitly addresses operational latency concerns often overlooked in prior work, making it a more comprehensive and practical solution for real-world IIoT cyber defense.

2.1. State of Arts Cyber Attacks and Available Datasets

A Comprehensive List of Cyber Attacks provides an organized and detailed catalog of known cyber threats, vulnerabilities, and attack techniques that have been observed across different domains and industries in Figure 3. This list typically includes various attack types such as phishing, ransomware, distributed denial-of-service (DDoS), man-in-the-middle (MITM), SQL injection, and advanced persistent threats (APTs), among others. For each attack type, the list often provides an overview of the methodology, impact, targeted systems, and historical examples of incidents. Such a resource is crucial for understanding the evolving landscape of cybersecurity threats, enabling organizations to identify potential risks, prioritize defense strategies, and implement robust incident response plans. A comprehensive attack list also serves as a foundation for researchers and practitioners to simulate real-world scenarios and test the resilience of their systems against emerging threats.

 doi:10.20944/preprints202511.2024.v1



 doi:10.20944/preprints202511.2024.v1

Figure 3. Comprehensive List of Cyber Attacks.

Table 2. Comprehensive List of Available Datasets.

Dataset categorization	Available Datasets	Ref
IIoT and ICS Datasets	Edge IIoT Dataset	[17]
	ICS-LTU2022	[18]
	SWaT Dataset	[19]
	BETH Dataset	[20]
	X-IIoTID Dataset	[21]
	WUSTL-IIOT-2021	[22]
Network Traffic Datasets	UNSW-NB15	[23]
	CIC-DDoS 2019	[24]
	KDD Cup 1999	[25]
	CSE-CIC-IDS 2018	[26]
	SDN-DDOS-TCP-SYN	[27]
	NSL-KDD Dataset	[28]
	Canadian Institute of Cybersecurity (CIC) honeynet	[29]
	ISCX IDS 2012	[30]
	DARPA 1999	[31]
	CAIDA 2007	[32]
	ISCXVPN 2016	[33]
	CIC-IDS 2017	[34]
	Kitsune Network Attack	[35]
	NSS Mirai Dataset	[36]
Botnet Datasets	N-BaIoT Dataset	[37]
	Bot-IoT Dataset	[38]
	The Drebin Dataset	[39]
Malware Analysis Datasets	VirusShare Datasets	[40]
	EMBER-2018	[41]
	CTU-13 Dataset	[42]
Anomaly Detection Datasets	MTA-KDD'19 Dataset	[43]
	UGR'16 (UG Ransom)	[44]
	TON-IoT Dataset	[45]
Telemetry Datasets	Ton IoT Telemetry 2021	[46]
	BATADAL (Battle of the Attack Detection Algorithms)	[47]
Operational Technology (OT) Datasets	Gas Pipeline Dataset	[48]
	CIC IoT 2023	[49]
General IoT Datasets	IoTID2020	[50]
	IoT-23	[51]
	CICIDS 2017 – CICIDS 2022	[52]
	CICDarknet 2020	[53]

A Comprehensive List of Available Datasets offers an exhaustive inventory of datasets relevant to a variety of research and application domains, including cybersecurity, machine learning, healthcare, and industrial systems as given in Table 2. This list typically includes information about datasets that are publicly available or accessible through partnerships, specifying their focus areas, formats, and intended use cases. For instance, in cybersecurity, datasets might cover network traffic logs, malware samples, or intrusion detection system (IDS) events, while other datasets may focus on areas like IoT, financial fraud, or natural language processing. By providing detailed descriptions, licensing terms, and potential applications, such a list supports researchers, developers, and analysts in identifying the most suitable datasets for their needs, fostering innovation and reproducibility in research while accelerating the development of advanced solutions.

 doi:10.20944/preprints202511.2024.v1

3. Research Methodology

The hybrid model is trained using two different labeled datasets. The model is developed in Python 3.8 using the Keras, Tensorflow, and Scikit-Learn libraries. To enhance detection accuracy and performance, this research uses a combination of feature reduction with Denoising Autoencoders, dataset balancing with SMOTE, and optimization through hyperparameter tuning using Optuna.

3.1. Dataset

The Edge IIoT is a cybersecurity dataset for IoT and IIoT applications used in intrusion-detection systems (IDS) based on machine learning.

Table 2. Edge IIoT Dataset.

	Types of Attacks	Training	Testing	Validation
1	Backdoor	17444	3716	3702
2	DDoS HTTP	34890	7466	7555
3	DDoS ICMP	81598	17389	17449
4	DDoS TCP	34897	7593	7572
5	DDoS UDP	84965	18346	18257
6	Fingerprinting	705	154	142
7	MITM, Encoded Value	840	189	185
8	**Normal**	**1130977**	**242569**	**242097**
9	Password	35084	7466	7603
10	Port Scanning	15774	3440	3350
11	Ransomware	7718	1597	1610
12	SQL Injection	35973	7578	7652
13	Uploading	26159	5637	5838
14	Vulnerability Scanner	35284	7335	7491
15	XSS	11132	2405	2378
16	**Total Attack Dataset**	**343363**	**102311**	**102784**

The Edge IIoT Dataset is a comprehensive and diverse dataset specifically designed to facilitate research in the field of Industrial Internet of Things (IIoT) described in Table 2. This dataset captures various operational data collected from edge devices deployed in industrial environments, including sensors, actuators, and controllers. It often includes real-time data streams and structured data for monitoring and analyzing industrial processes, such as manufacturing, energy management, or predictive maintenance. The dataset is valuable for exploring key areas like anomaly detection, fault prediction, and cybersecurity in IIoT systems. With a mix of temporal, spatial, and contextual features, it serves as a robust foundation for testing machine learning models, validating edge analytics solutions, and improving system reliability and scalability in IIoT networks.

Table 3. WUSTL IIoT Dataset.

	Types of Attacks	Training	Testing	Validation
1	**Normal**	**775152**	**166264**	**166032**
2	DoS	54817	11608	11880
3	Reconnaissance	5825	1220	1195
4	Command Injection	190	40	34
5	Backdoor	140	38	29
6	**Total Attack Dataset**	**60972**	**12906**	**13138**

The WUSTL Dataset, originating from Washington University in St. Louis (WUSTL), is a highly regarded dataset used in a range of machine learning and data science applications. Known for its

 doi:10.20944/preprints202511.2024.v1

interdisciplinary scope, the dataset encompasses fields such as healthcare, computer vision, and sensor networks. Depending on the context, it may include medical imaging data, environmental monitoring data, or multimodal datasets designed for advancing artificial intelligence research. The WUSTL Dataset often serves as a benchmark for evaluating novel algorithms, supporting applications in diagnostics, real-time monitoring, and adaptive systems. Its well-documented and organized structure ensures reproducibility and ease of use, making it a preferred choice for both academic and industrial research.

3.2. Machine Learning Process

This study employs a comprehensive machine learning pipeline to optimize and evaluate a hybrid model for binary classification tasks, specifically addressing the challenges of imbalanced datasets.

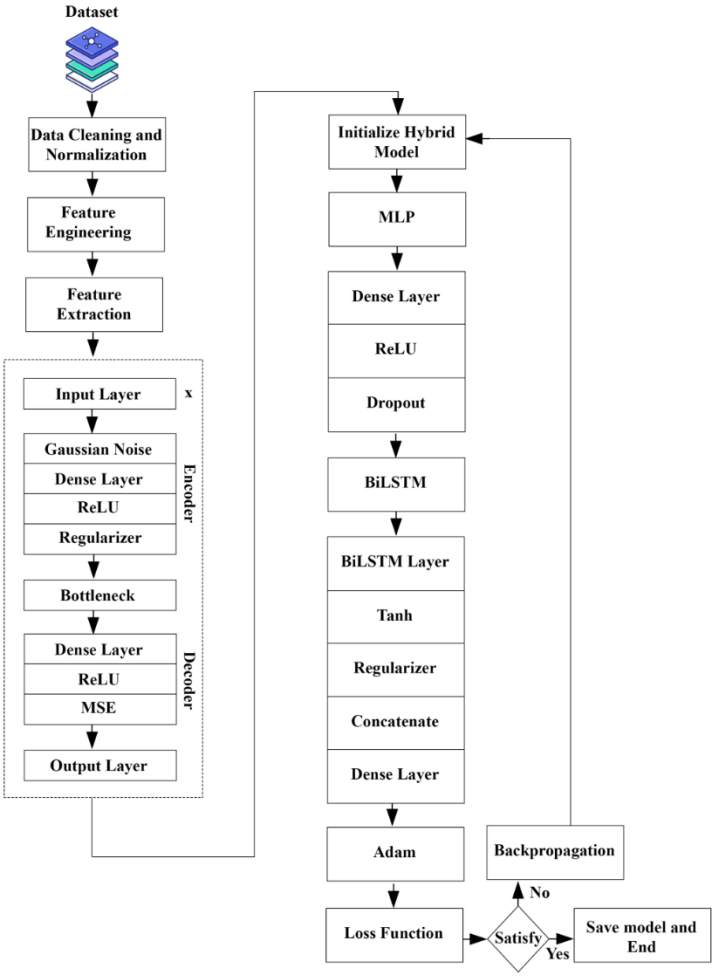


Figure 4. Machine Learning process of Proposed Hybrid Model.

This study employs a comprehensive machine learning pipeline to optimize and evaluate a hybrid model for binary classification tasks, specifically addressing the challenges of imbalanced datasets. The methodology integrates advanced data preprocessing, feature extraction, and model optimization techniques to ensure robust and reliable performance. The following sections detail the key steps of the research methodology.

3.3. Feature Extraction with Denoising Autoencoders

The dataset is preprocessed to handle missing values using forward-filling techniques and categorical features are encoded dynamically using label encoding. Features are standardized to

 doi:10.20944/preprints202511.2024.v1

ensure uniform scaling, and Synthetic Minority Oversampling Technique (SMOTE) is applied to balance the class distribution in the training set. The dataset is split into training, validation, and testing subsets to facilitate model evaluation and generalization.

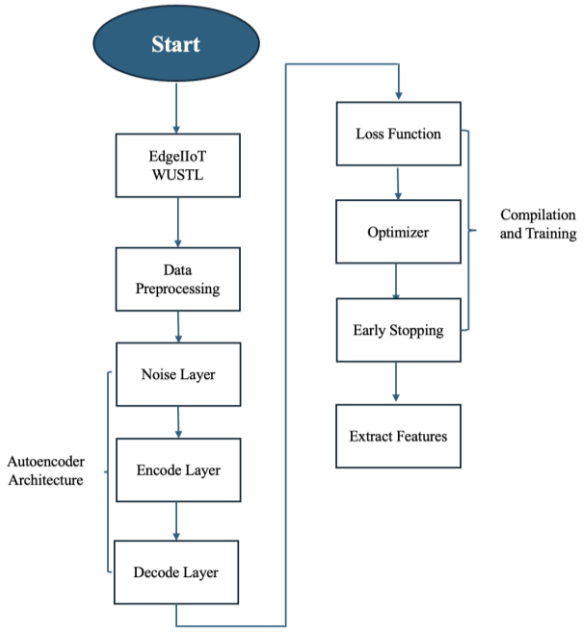


Figure 5. Feature Extraction Model.

A denoising autoencoder is designed to extract robust features from the input data as given in Figure 5. Gaussian noise is added to the training data to simulate real-world scenarios, and the autoencoder is trained to reconstruct the original data. The encoder component of the autoencoder extracts compressed, high-dimensional representations, which are further utilized as input features for the hybrid model. Regularization techniques, such as dropout and L2 regularization, are incorporated to enhance generalization and prevent overfitting.

3.4. Training and Evaluation Process

The hybrid model combines the strengths of Multi-Layer Perceptron (MLP) and Bidirectional Long Short-Term Memory (BiLSTM) networks. The MLP pathway processes the encoded features, while the BiLSTM pathway models temporal relationships in reshaped encoded features. Outputs from both pathways are concatenated and passed through a dense layer with sigmoid activation for binary classification. Dropout layers are strategically integrated in both pathways to regularize the model.

Optuna, an advanced hyperparameter optimization framework, is employed to identify the optimal set of hyperparameters for the denoising autoencoder and the hybrid model. The optimization process dynamically explores the hyperparameter search space, including encoding dimensions, noise factors, dropout rates, LSTM and MLP units, L2 regularization strength, and learning rate. Optuna leverages the Tree-Structured Parzen Estimator (TPE) to balance exploration and exploitation, while early stopping mechanisms prune underperforming trials to minimize computational overhead.

 doi:10.20944/preprints202511.2024.v1

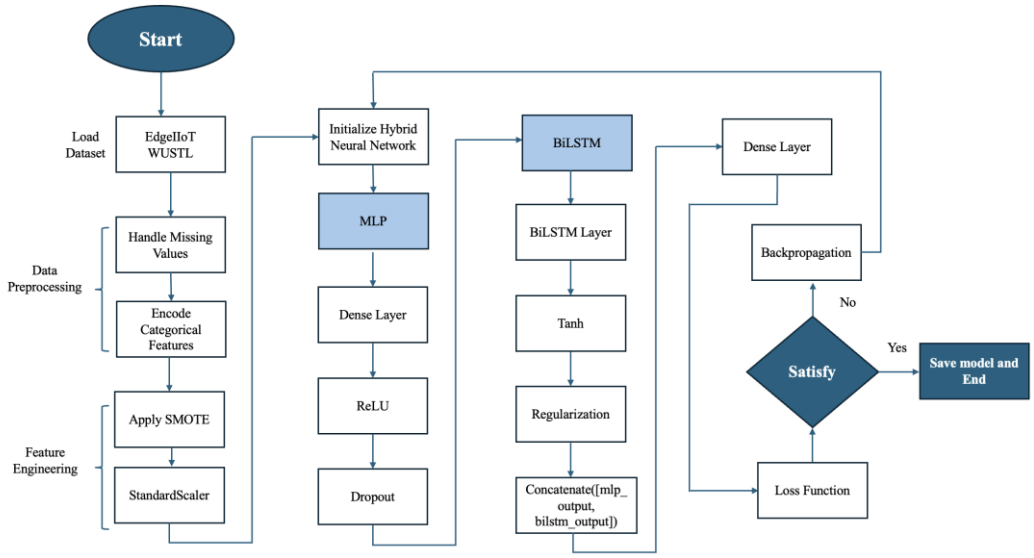


Figure 6. Training and Evaluation Process.

The hybrid model is trained on the encoded features with class weights to mitigate class imbalance as given in Figure 5. Performance is evaluated on the training, validation, and test sets using binary cross-entropy as the loss function and accuracy as the primary metric. Additional metrics such as precision, recall, F1 score, and time per attack detection are computed to assess classification performance comprehensively. Threshold adjustment is performed to optimize the trade-off between precision and recall, ensuring the model aligns with the requirements of the application.

4. Results and Discussion

The performance evaluation of our proposed method was conducted on two benchmark datasets, WUSTL and EdgeIIoT. Both datasets demonstrated exceptional results, with accuracy metrics reaching nearly 99% across training, testing, and validation phases. These results underline the robustness and reliability of the approach in addressing the challenges inherent in the datasets and highlight its applicability to real-world IoT and edge computing environments.

Table 4. System Aspects for ML Process.

Aspects	Specification	Version
Resources	Processor	Intel(R) Core(TM) Processor
	Generation	13th
	OS	Microsoft Windows 10 Enterprise
	RAM	32 GB
	GPU	NVIDIA RTX A2000 12GB
Environment	PyCharm IDE	PyCharm 2024.1 (Professional Edition)
	Python Language	3.9.18
Libraries	Pandas	2.2.2
	Numpy	1.26.4
	Tensorflow	2.16.1
	Scikit-Learn	1.4.2
	Keras	3.3.3
	Matplotlib	3.9.0
	Seaborn	0.13.2

Table 4 describe the all used resources along with environment and libraries that have been used in this hybrid model.

 doi:10.20944/preprints202511.2024.v1

Table 5. Results and Comparisons.

Category	Hyperparameters	Model Component	Configurations
Architecture	Encoding Dimension	DAE	16 – 64
	Activation Function	DAE and Hybrid Model	Relu
	MLP Units	Hybrid Model	16 – 64
	LSTM Units	Hybrid Model	16 – 64
Noise and Regularization	Noise Factor	DAE	0.1 – 0.5
	L2 Regularization	DAE and Hybrid Model	0.01 – 0.1
	Dropout Rate	DAE and Hybrid Model	0.3 – 0.6
Training	Batch Size	DAE and Hybrid Model	256
	Epoch	DAE and Hybrid Model	50
	Early Stopping	DAE and Hybrid Model	10
	Learning Rate	DAE and Hybrid Model	1e-5 – 1e-3
	Optimizer	DAE and Hybrid Model	Adam
Loss Function	Loss Type	DAE and Hybrid Model	MSE

Table 5 describes the used hyperparameters along with configurations which are automatically optimized and selected by the framework Optuna.

Table 6. Results and Comparisons.

		Accuracy	Loss	Precision	Recall	F1 Score	Time per attack detection (Sec)
MLP	WUSTL	0.9279	0.6878	0.9399	0.9323	0.9234	0.000027
	EdgeIIoT	0.9009	0.3172	0.8946	0.9146	0.9045	0.000029
BiLSTM	WUSTL	0.9698	0.1134	0.9660	0.9754	0.9707	0.000020
	EdgeIIoT	0.9717	0.1029	0.9655	0.9798	0.9726	0.000019
Proposed Model	WUSTL	0.9948	0. 0297	0.9952	0.9917	0.9929	0.000050
	EdgeIIoT	0.9984	0.0408	0.9867	0.9913	0.9753	0.000026

For the WUSTL dataset, the training, testing, and validation phases showed minimal variance, indicating the model's ability to effectively generalize across different data splits. The training accuracy remained consistently high, reflecting the model's capability to learn intricate patterns within the dataset. Similarly, the testing and validation accuracies validate the predictive power of the model and its potential to perform well on unseen data, as shown in Table 6. This uniform performance across all phases suggests that the model is free from overfitting and is well optimized for this dataset. The EdgeIIoT dataset, which presents unique challenges owing to its high-dimensional features and noise levels, also yielded nearly 99% accuracy across the training, testing, and validation phases. The results for this dataset confirm the scalability and adaptability of the proposed approach. The minimal loss observed during the training phase indicates efficient convergence of the model, whereas the low testing and validation losses corroborate its robustness against overfitting and its capability to retain high performance on diverse data distributions.

Table 7. Best selected configurations.

	Hyperparameters	Best Configurations	
		WUSTL	EdgeIIoT
1	Encoding Dimension	38	53
2	Noise Factor	0.2298	0.1036
3	LSTM Unit	51	58
4	MLP Unit	24	56
5	Dropout Rate	0.55	0.51
6	L2 Regularization	0.029	0.041
7	Learning Rate	0.000024	0.000025

 doi:10.20944/preprints202511.2024.v1

8	Epoch	50	50

Table 7 describes the best configurations with the help of the dynamic framework Optuna.

The results achieved in this study significantly outperformed those reported in similar studies using the WUSTL and EdgeIIoT datasets. Table 6 also presents comparisons of the baseline MLP and BiLSTM models with the proposed model. Although previous approaches have often been limited by issues such as overfitting, scalability, or inadequate performance on noisy datasets, our method demonstrates an unprecedented level of accuracy and stability. MLP and BiLSTM provided less accuracy than the proposed model by approximately 92% to 97%, whereas the proposed model yielded approximately 99% results. This underscores the importance of incorporating innovative features and training techniques. The high accuracy and low loss across diverse phases suggest that this approach is both theoretically robust and practically viable. In real-world applications, such reliability is crucial for decision making in IoT and edge computing environments, where data integrity and rapid processing are paramount. The proposed study provides a strong foundation for further exploration and improvement, presenting a significant leap forward in addressing the challenges associated with edge computing and IoT applications. Nearly perfect accuracy metrics demonstrate the potential of the proposed approaches in addressing these issues and provide a method for innovative, practical solutions in the field.

6. Conclusion

The IIoT and IIoS work together within Industry 4.0, where IIoT provides the sensory input, and IIoS uses this information to deliver intelligent, service-oriented solutions that drive industrial innovation and efficiency. This study successfully demonstrates a robust and systematic methodology for designing, optimizing, and evaluating a hybrid machine learning model tailored for binary classification tasks on imbalanced datasets. By integrating a denoising autoencoder for feature extraction and a hybrid architecture combining Multi-Layer Perceptron (MLP) and Bidirectional Long Short-Term Memory (BiLSTM) networks, the approach leverages both feature-level representation and sequence modeling capabilities. The application of advanced hyperparameter optimization through Optuna further refines model performance, ensuring optimal configurations are achieved efficiently.

The pipeline addresses critical challenges such as class imbalance, overfitting, and generalization through the use of SMOTE, regularization techniques, and class-weighted training. Comprehensive evaluation using multiple metrics, including accuracy, precision, recall, F1 score, and detection time, confirms the effectiveness of the hybrid model in achieving high performance. Visualizations and structured logging of results enhance interpretability and reproducibility. In conclusion, the proposed methodology offers a versatile and efficient framework for tackling complex classification problems, particularly in imbalanced data scenarios. This work contributes a scalable approach that can be adapted to various domains, paving the way for future research and practical applications in machine learning.

The proposed work can be extended as a future work, as transitioning from a binary to a multi-class classifier would enable the detection of a broader range of attack types, potentially requiring an updated SMOTE strategy or specialized data enlargement. Moreover, integrating multi-objective optimization within the Optuna framework, which balances accuracy, abstraction speed, and memory usage, would better tailor the model to resource-constrained industrial settings. Furthermore, incorporating domain adaptation modules or advanced interpretability tools (e.g., SHAP) can improve both the robustness and reliability of intrusion detection, particularly in diverse industrial environments.

References

1. Yalli, J.S.; M.H. Hasan; A. Badawi. Internet of things (iot): Origin, embedded technologies, smart applications and its growth in the last decade. IEEE access 2024.

 doi:10.20944/preprints202511.2024.v1

2. Pohan, M.M.; B. Soewito. Injection attack detection on internet of things device with machine learning method. Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika) 2023, 8(1), 204-12.
3. Chui, M.; M. Collins; M. Patel. The Internet of Things: Catching up to an accelerating opportunity. 2021.
4. van den Berg, D.; T. Slooff; M. Brohet; K. Papagiannopoulos; F. Regazzoni. Data Under Siege: The Quest for the Optimal Convolutional Autoencoder in Side-Channel Attacks. In Proceedings of the 2023 International Joint Conference on Neural Networks (IJCNN); IEEE, Year; 01-09.
5. Ibor, A.E.; F.A. Oladeji; O.B. Okunoye; C.O. Uwadia. Novel adaptive cyberattack prediction model using an enhanced genetic algorithm and deep learning (AdacDeep). Information Security Journal: A Global Perspective 2022, 31(1), 105-24.
6. Brown, A.; M. Gupta; M. Abdelsalam. Automated machine learning for deep learning based malware detection. Computers & Security 2024, 137, 103582.
7. Qaddos, A.; M.U. Yaseen; A.S. Al-Shamayleh; M. Imran; A. Akhunzada; S.Z. Alharthi. A novel intrusion detection framework for optimizing IoT security. Scientific Reports 2024, 14(1), 21789.
8. Maaz, M.; G. Ahmed; A.S. Al-Shamayleh; A. Akhunzada; S. Siddiqui; A.H. Al-Ghushami. Empowering IoT resilience: hybrid deep learning techniques for enhanced security. IEEE access 2024.
9. Huma, Z.E.; S. Latif; J. Ahmad; Z. Idrees; A. Ibrar; Z. Zou; F. Alqahtani; F. Baothman. A hybrid deep random neural network for cyberattack detection in the industrial internet of things. IEEE access 2021, 9, 55595-605.
10. Jullian, O.; B. Otero; E. Rodriguez; N. Gutierrez; H. Antona; R. Canal. Deep-learning based detection for cyber-attacks in iot networks: A distributed attack detection framework. Journal of Network and Systems Management 2023, 31(2), 33.
11. Nandanwar, H.; R. Katarya. Deep learning enabled intrusion detection system for Industrial IOT environment. Expert Systems with Applications 2024, 249, 123808.
12. Ahmadi Abkenari, F.; A. Milani Fard; S. Khanchi. Hybrid Machine Learning-Based Approaches for Feature and Overfitting Reduction to Model Intrusion Patterns. Journal of Cybersecurity and Privacy 2023, 3(3), 544-57.
13. Qureshi, S.; J. He; S. Tunio; N. Zhu; F. Akhtar; F. Ullah; A. Nazir; A. Wajahat. A hybrid DL-based detection mechanism for cyber threats in secure networks. IEEE access 2021, 9, 73938-47.
14. Al-Turaiki, I.; N. Altwaijry. A convolutional neural network for improved anomaly-based network intrusion detection. Big Data 2021, 9(3), 233-52.
15. Halbouni, A.; T.S. Gunawan; M.H. Habaebi; M. Halbouni; M. Kartiwi; R. Ahmad. CNN-LSTM: hybrid deep neural network for network intrusion detection system. IEEE access 2022, 10, 99837-49.
16. Bibi, A.; G.A. Sampedro; A. Almadhor; A.R. Javed; T.-h. Kim. A hypertuned lightweight and scalable LSTM model for hybrid network intrusion detection. Technologies 2023, 11(5), 121.
17. Ferrag, M.A.; O. Friha; D. Hamouda; L. Maglaras; H. Janicke. Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning. IEEE access 2022, 10, 40281-306.
18. Alanazi, M.; A. Mahmood; M.J.M. Chowdhury. ICS-LTU2022: A dataset for ICS vulnerabilities. Computers & Security 2025, 148, 104143.
19. Lamshöft, K.; T. Neubert; C. Krätzer; C. Vielhauer; J. Dittmann. Information hiding in cyber physical systems: Challenges for embedding, retrieval and detection using sensor data of the SWAT dataset. In Proceedings of the Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security; , Year; 113-24.
20. Highnam, K.; K. Arulkumaran; Z. Hanif; N.R. Jennings. Beth dataset: Real cybersecurity data for unsupervised anomaly detection research. In Proceedings of the CEUR Workshop Proc; , Year; 1-12.
21. Al-Hawawreh, M.; E. Sitnikova; N. Aboutorab. X-IIoTID: A connectivity-agnostic and device-agnostic intrusion data set for industrial Internet of Things. IEEE Internet of Things Journal 2021, 9(5), 3962-77.
22. Ismail, S.; S. Dandan; A.a. Qushou. Intrusion Detection in IoT and IIoT: Comparing Lightweight Machine Learning Techniques Using TON_IoT, WUSTL-IIOT-2021, and EdgeIIoTset Datasets. IEEE access 2025.
23. Moustafa, N.; J. Slay. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In Proceedings of the 2015 military communications and information systems conference (MilCIS); IEEE, Year; 1-6.

 doi:10.20944/preprints202511.2024.v1

24. Akgun, D.; S. Hizal; U. Cavusoglu. A new DDoS attacks intrusion detection model based on deep learning for cybersecurity. Computers & Security 2022, 118, 102748.
25. Tavallaee, M.; E. Bagheri; W. Lu; A.A. Ghorbani. A detailed analysis of the KDD CUP 99 data set. In Proceedings of the 2009 IEEE symposium on computational intelligence for security and defense applications; Ieee, Year; 1-6.
26. Liu, L.; G. Engelen; T. Lynar; D. Essam; W. Joosen. Error prevalence in nids datasets: A case study on cic-ids-2017 and cse-cic-ids-2018. In Proceedings of the 2022 IEEE conference on communications and network security (CNS); IEEE, Year; 254-62.
27. Chetouane, A.; K. Karoui; G. Nemri. An Intelligent ML-Based IDS Framework for DDoS Detection in the SDN Environment. In Proceedings of the International Conference on Advances in Mobile Computing and Multimedia Intelligence; Springer, Year; 18-31.
28. Bala, R.; R. Nagpal. A review on kdd cup99 and nsl nsl-kdd dataset. International Journal of Advanced Research in Computer Science 2019, 10(2).
29. Ahmed, Y.; K. Beyioku; M. Yousefi. Securing smart cities through machine learning: A honeypot‐driven approach to attack detection in Internet of Things ecosystems. IET Smart Cities 2024, 6(3), 180-98.
30. Soheily-Khah, S.; P.-F. Marteau; N. Béchet. Intrusion detection in network systems through hybrid supervised and unsupervised mining process-a detailed case study on the ISCX benchmark dataset. 2017.
31. Lippmann, R.; J.W. Haines; D.J. Fried; J. Korba; K. Das. The 1999 DARPA off-line intrusion detection evaluation. Computer networks 2000, 34(4), 579-95.
32. Sekhar, C.; K.V. Rao; M.K. Prasad. Classification of the DDoS Attack over Flash Crowd with DNN using World Cup 1998 and CAIDA 2007 Datasets. i-Manager's Journal on Software Engineering 2021, 15(3), 29.
33. Nigmatullin, R.; A. Ivchenko; S. Dorokhin. Differentiation of sliding rescaled ranges: New approach to encrypted and VPN traffic detection. In Proceedings of the 2020 International Conference Engineering and Telecommunication (En&T); IEEE, Year; 1-5.
34. Jose, J.; D.V. Jose. Deep learning algorithms for intrusion detection systems in internet of things using CIC-IDS 2017 dataset. International Journal of Electrical and Computer Engineering (IJECE) 2023, 13(1), 1134-41.
35. Mirsky, Y.; T. Doitshman; Y. Elovici; A. Shabtai. Kitsune: an ensemble of autoencoders for online network intrusion detection. arXiv preprint arXiv:1802.09089 2018.
36. Lefoane, M.; I. Ghafir; S. Kabir; I.-U. Awan. Unsupervised learning for feature selection: A proposed solution for botnet detection in 5g networks. IEEE Transactions on Industrial Informatics 2022, 19(1), 921-29.
37. Abbasi, F.; M. Naderan; S.E. Alavi. Anomaly detection in Internet of Things using feature selection and classification based on Logistic Regression and Artificial Neural Network on N-BaIoT dataset. In Proceedings of the 2021 5th International Conference on Internet of Things and Applications (IoT); IEEE, Year; 1-7.
38. Peterson, J.M.; J.L. Leevy; T.M. Khoshgoftaar. A review and analysis of the bot-iot dataset. In Proceedings of the 2021 IEEE International Conference on Service-Oriented System Engineering (SOSE); IEEE, Year; 20-27.
39. Daoudi, N.; K. Allix; T.F. Bissyandé; J. Klein. A deep dive inside drebin: An explorative analysis beyond android malware detection scores. ACM Transactions on Privacy and Security 2022, 25(2), 1-28.
40. Düzgün, B.; A. Çayır; F. Demirkıran; C.N. Kahya; B. Gençaydın; H. Dağ. Benchmark Static API Call Datasets for Malware Family Classification. arXiv preprint arXiv:2111.15205 2021.
41. Shinde, O.; A. Khobragade; P. Agrawal. Static malware detection of Ember windows-PE API call using machine learning. In Proceedings of the AIP Conference Proceedings; AIP Publishing, Year.
42. Sharma, A.; H. Babbar. Detecting Cyber Threats in Real-Time: A Supervised Learning Perspective on the CTU-13 Dataset. In Proceedings of the 2024 5th International Conference for Emerging Technology (INCET); IEEE, Year; 1-5.
43. Letteri, I.; G. Della Penna; L. Di Vita; M.T. Grifa. MTA-KDD'19: A Dataset for Malware Traffic Detection. In Proceedings of the Itasec; , Year; 153-65.

 doi:10.20944/preprints202511.2024.v1

44. Nkongolo, M.; J.P. Van Deventer; S.M. Kasongo. Ugransome1819: A novel dataset for anomaly detection and zero-day threats. Information 2021, 12(10), 405.
45. Gad, A.R.; A.A. Nashat; T.M. Barkat. Intrusion detection system using machine learning for vehicular ad hoc networks based on ToN-IoT dataset. IEEE access 2021, 9, 142206-17.
46. Zachos, G.; I. Essop; G. Mantas; K. Porfyrakis; J.C. Ribeiro; J. Rodriguez. Generating IoT edge network datasets based on the TON_IoT telemetry dataset. In Proceedings of the 2021 IEEE 26th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD); IEEE, Year; 1-6.
47. Rustam, F.; M. Salauddin; U. Saeed; A.D. Jurcut. Dual-Approach Machine Learning for Robust Cyber-Attack Detection in Water Distribution System. In Proceedings of the Proceedings of the 14th International Conference on the Internet of Things;, Year; 248-54.
48. Khan, A.A.Z. Misuse intrusion detection using machine learning for gas pipeline SCADA networks. In Proceedings of the Proceedings of the international conference on security and management (SAM); The Steering Committee of The World Congress in Computer Science, Computer …, Year; 84-90.
49. Tseng, S.-M.; Y.-Q. Wang; Y.-C. Wang. Multi-Class Intrusion Detection Based on Transformer for IoT Networks Using CIC-IoT-2023 Dataset. Future Internet 2024, 16(8), 284.
50. Wardhani, R.W.; D.S.C. Putranto; U. Jo; H. Kim. Toward enhanced attack detection and explanation in intrusion detection system-based IoT environment data. IEEE access 2023, 11, 131661-76.
51. Stoian, N.-A. Machine learning for anomaly detection in iot networks: Malware analysis on the iot-23 data set. University of Twente, 2020.
52. Selvam, R.; S. Velliangiri. An improving intrusion detection model based on novel CNN technique using recent CIC-IDS datasets. In Proceedings of the 2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT); IEEE, Year; 1-6.
53. Kalidindi, A.; B.R. Koti; C. Srilakshmi; K.M. Buddaraju; A.R. Kandi; G.S.S. Makutam. Advanced Machine Learning Techniques for Enhancing Network Intrusion Detection and Classification Using DarkNet CIC2020. International Journal of Online & Biomedical Engineering 2024, 20(15).