electronics

MDPI

Article

Using Ensemble Learning for Anomaly Detection in Cyber–Physical Systems

Nicholas Jeffrey [1,*], Qing Tan [2] and José R. Villar [1]

[1] Faculty of Computer Science, University of Oviedo, 33003 Oviedo, Spain; villarjose@uniovi.es
[2] Faculty of Science and Technology, Athabasca University, Athabasca, AB T9S 3A3, Canada; qingt@athabascau.ca
* Correspondence: uo292630@uniovi.es

Abstract: The swift embrace of Industry 4.0 paradigms has led to the growing convergence of Information Technology (IT) networks and Operational Technology (OT) networks. Traditionally isolated on air-gapped and fully trusted networks, OT networks are now becoming more interconnected with IT networks due to the advancement and applications of IoT. This expanded attack surface has led to vulnerabilities in Cyber–Physical Systems (CPSs), resulting in increasingly frequent compromises with substantial economic and life safety repercussions. The existing methods for the anomaly detection of security threats typically use simple threshold-based strategies or apply Machine Learning (ML) algorithms to historical data for the prediction of future anomalies. However, due to the high levels of heterogeneity across different CPS environments, minimizing the opportunities for transfer learning, and the scarcity of real-world data for training, the existing ML-based anomaly detection techniques suffer from a poor predictive performance. This paper introduces a hybrid anomaly detection approach designed to identify threats to CPSs by combining the signature-based anomaly detection typically utilized in IT networks, the threshold-based anomaly detection typically utilized in OT networks, and behavioural-based anomaly detection using Ensemble Learning (EL), which leverages the strengths of multiple ML algorithms against the same dataset to increase the accuracy. Multiple public research datasets were used to validate the proposed approach, with the hybrid methodology employing a divide-and-conquer strategy to offload the detection of certain cyber threats to computationally inexpensive signature-based and threshold-based methods using domain knowledge to minimize the size of the behavioural-based data needed for ML model training, thus achieving a higher accuracy over a reduced timeframe. The experimental results showed accuracy improvements of 4–7% over those of the conventional ML classifiers in performing anomaly detection across multiple datasets, which is particularly important to the operators of CPS environments due to the high financial and life safety costs associated with interruptions to system availability.

Keywords: anomaly detection; Cyber–Physical Systems; IoT; IIoT; machine learning; ensemble learning

check for updates

Citation: Jeffrey, N.; Tan, Q.; Villar, J.R. Using Ensemble Learning for Anomaly Detection in Cyber–Physical Systems. *Electronics* **2024**, *13*, 1391. https://doi.org/10.3390/electronics13071391

Academic Editors: Umit Karabiyik, Mamoun Alazab and Abdelkader Ouda

Received: 13 March 2024
Revised: 28 March 2024
Accepted: 4 April 2024
Published: 7 April 2024

1. Introduction

Society is currently transitioning from the third industrial revolution, IT Systems and Automation, to the so-called Industry 4.0, Cyber–Physical Systems. This transition began in 2011 as the result of an industrial research effort [1] initiated by the German government, with the objective of leveraging the pervasive availability of high-speed network connectivity and rich telemetry data to improve industrial manufacturing processes.

The impetus for Industry 4.0 was the growing availability of ubiquitous networking connectivity, allowing for manufacturing processes to be optimized through smart automation. To contrast, the earlier iterations of industrial control processes were characterized by isolated systems using basic relay logic over point-to-point serial connections with an extremely limited bandwidth. These primitive Operational Technology (OT) networks gradually evolved into the Industrial Internet of Things (IIoT), with hyper-connected sensors and actuators forming an intelligent Cyber–Physical System (CPS), an integrated environment comprised of software and real-world physical components such as sensors and actuators.

CPS environments can be viewed as the fusion of Information Technology (IT) and Operational Technology (OT) environments, each having distinct priorities regarding system security. IT networks have historically embraced the CIA (Confidentiality, Integrity, and Availability) pillars of information security to establish the organizational security stance, prioritizing each facet in a specific order of importance. In contrast, OT networks reverse this order [2,3], placing the highest importance on availability, followed by integrity, while regarding confidentiality as the least critical aspect of overall system security. This divergence largely stems from the evolution of OT from earlier Industrial Control Systems (ICSs), where availability was paramount, and integrity and confidentiality were seldom considered due to the use of trusted, independent, and isolated network connections.

As OT networks have been amalgamated with IT networks to create modern CPS environments, the persistent disparities in priorities have led to ongoing challenges that are yet to be fully addressed [4]. IT networks emphasize authentication (identifying users) and authorization (determining allowed actions), aligning roughly with the confidentiality and integrity facets of the CIA triad of information security. However, OT networks have traditionally placed such a heavy emphasis on the availability facet of the CIA triad that authentication and authorization were assumed to be valid based on physical access to the trusted and isolated OT network [5].

The historical assumption of a fully trusted and isolated environment [6] is no longer valid with the integration of IT and OT networks, exposing vulnerabilities to common network-based attacks such as Distributed Denial of Service (DDoS), Man-in-the-Middle (MitM), replay attacks, impersonation, spoofing, false data injection, etc.

This merging of IT and OT networks has tremendously increased economic activity and brought quality of life improvements, but is not without challenges. The rapid growth of IIoT has outpaced advancements in cybersecurity, with new threat models and security challenges that lack a unified framework for secure design, malware resistance, and risk mitigations.

The increased incorporation of CPS across different facets of modern society has led to a significant rise in malicious attacks by adversaries. In recent times, there has been a notable increase in incidents targeting critical civilian infrastructure, including power grids and oil pipelines. This surge in attacks can be attributed to heightened connectivity to the public internet, substantially amplifying the vulnerability of CPSs.

The common anomaly detection methods will typically employ one or two ML algorithms that have been manually selected [7,8] by the researcher, which leads to results that are influenced by the limitations of a single algorithm. Additionally, the results can vary widely by dataset due to variations in features and data distribution, leading to limited opportunities for generic solutions that can be applied to multiple CPS environments.

To contrast, Ensemble Learning (EL) utilizes multiple ML algorithms with the same dataset, which leads to improved predictive performance when compared to a single ML algorithm [9–11]. Even minor improvements in predictive performance are highly desirable in CPS environments, which are extremely intolerant of false positives and false negatives due to the potential economic and life safety consequences of incorrect predictions.

Due to the high levels of heterogeneity in CPS environments and the lack of real-world datasets with representative data, ML algorithms for performing anomaly detection for CPS have historically been very environment-specific, with significant amounts of effort in the data pre-processing stages. A generic framework for anomaly detection in CPS has proved elusive, partially due to the high levels of manual effort required.

EL is a promising strategy for performing anomaly detection in CPS, as the highly imbalanced datasets that typically introduce bias into ML algorithms can be minimized by combining multiple weak classifier models to obtain a strong model [12]. By providing different weights to each ML algorithm, the weaknesses can be minimized, and the strengths can be maximized, resulting in a hybrid group of algorithms with an overall better predictive performance. While ensemble-based learning methods predate written human history, the concept of using Ensemble Learning in computational intelligence is mere decades old, first proposed by Dasarathy and Sheela [13] in 1979. In its early years, EL struggled to find broad acceptance with researchers in anomaly detection studies due to the added complexity and high computational costs for what was a marginal improvement at best. However, with exponential improvements in microprocessor performances over the years, combined with the high costs of false positives and false negatives in CPS environments, EL has become increasingly popular as an anomaly detection strategy in this niche.

This paper proposes a novel Ensemble Learning-Based Hybrid Anomaly Detection Method comprised of signature-based detection for known threats, threshold-based metrics for the immutable physical characteristics of a CPS, combined with an ensemble-based learning model for behaviour-based anomaly detection, with the goal of improved predictive performance over those of the existing anomaly detection methods, which is demonstrated using two public research datasets (Edge-IIoTset2023 and CICIoT2023). This paper builds upon previous works [14–16] by the authors of this paper, furthering the development of a generalizable framework for threat detection in CPS environments that can be applied in a broad variety of CPS environments through the use of EL to overcome weaknesses in existing threat detection models.

The remainder of this paper is organized as follows: Section 2 provides a literature analysis of existing research covering Ensemble Learning methods for anomaly detection. Section 3 describes the experimental methodology presented in this study, Section 4 describes the experimental evaluation, while Section 5 describes the experimental results. Finally, Section 6 discusses the conclusions and opportunities for future works.

2. Related Works

Ensemble-based learning methods have existed as long as human civilization, since the first group of people learned that better outcomes to problems can be obtained from group input than individual decisions. Indeed, many of the day-to-day tasks in modern society are based on ensemble-based decisions, from democratic processes using votes to reach a consensus to the academic process of peer review to advance scientific knowledge, and many more. While anomaly detection is a common area of study in ML, there has been limited attention given to threat detection to CPS environments, and less still in the specific area of EL as a strategy for improving accuracy in threat detection to CPSs. Interested readers on this topic may find this review paper [15] worth reading.

Jeffrey et al. [16] propose a method of anomaly detection in CPS-leveraging unsupervised learning models with one-class classification algorithms as a method of compensating for the extreme scarcity of anomalous data in the commonly available research datasets. This proposed method focuses on the differences between supervised and unsupervised learning with a limited number of classification algorithms and mitigates some of the accuracy issues caused by imbalanced data classes, but suffers from limited transferability between CPS environments.

Afrifa et al. [17] start with the assumption that significant numbers of IoT devices are regularly commandeered into botnets by malicious actors to further nefarious goals that threaten global commerce. Due to resource constraints of typical IoT devices, a single compromised device may not be particularly dangerous, but a botnet comprised of thousands or millions of compromised devices can cause significant harm. A novel approach of botnet detection is proposed using Ensemble Learning to identify an individual node in a botnet and providing real-time intrusion prevention. This is a particularly interesting approach, as it focuses on Ensemble Learning as a strategy for the identification of botnet membership, rather than the more common classification task of identifying malicious vs. benign activities against an individual host.

Araya et al. [18] focus on the use of Ensemble Learning for anomaly detection in smart building energy consumption. This research is focused on improving electricity usage efficiency in commercial and industrial environments to reduce the environmental impact and is not specifically aimed at detection of malicious activity, but makes use of the same strategies for anomaly detection as IDS. Since power consumption in commercial buildings varies on an hour-by-hour basis correlating to the operational hours, as well as month-by-month fluctuations based on the seasonal requirements for heating and cooling, an anomaly detection framework based on sliding temporal windows is proposed to maximize efficiency, while minimizing false classifications due to cyclical variations. Real-world data were obtained and used to train multiple ML models using moving time windows, and then combined with an ensemble method, resulting in significantly improved predictive performance.

Yazdinejad et al. [19] propose an anomaly detection model for IIoT environments based on ensemble deep learning, utilizing Long Short-Term Memory (LTSM) and the AutoEncoder (AE) architecture to analyze time series data to identify anomalous activity. A common issue in anomaly detection in IIoT/CPS environments is highly imbalanced datasets, which affect the predictive performance of many individual ML algorithms. This research starts with the assumption that IIoT environments are highly distributed, with a large number of potentially heterogenous sensors and actuators, so approaches the problem as a big data challenge using an ensemble deep Recurrent Neural Network (RNN) model to perform pattern recognition on time series data gathered from monitoring the IIoT environment, and then classify the activity as normal or anomalous.

Saharkhizan et al. [20] further develop the concepts of using an ensemble of deep learning models to merge multiple ML models with low-level accuracy into an aggregated prediction model with higher accuracy than its component algorithms. Multiple LTSM models are trained on Modbus network traffic, and then aggregated with a decision tree to achieve higher levels of classification accuracy.

Danso et al. [21] propose an ensemble-based IDS located on the IoT gateway, avoiding any resource constraints on the IoT devices by performing passive network sniffing to collect network traffic samples, which are then used to train multiple ML models to serve as base learners, which are then used as inputs for an Ensemble Learning model to improve the predictive performance through combining the individual ML algorithms with a stacking meta-classifier to make the final predictions.

Illy et al. [22] propose a two-tiered approach to Ensemble Learning for IoT anomaly detection, with the first level used to perform a rapid normal vs. anomaly detection, and the second level performing a more detailed and time-consuming classification of the specific attack type. This hybrid approach seeks to quickly achieve a coarse-grained anomaly detection in order to provide rapid notification and response, followed up with more fine-grained detailed analysis that can still be processing after the initial alert notification. This trade-off between the detection speed and classification accuracy is implemented through the use of multiple base learners that differ in domain expertise to achieve rapid detection with a voting ensemble classifier, followed by more detailed and time-consuming attack classification analysis via a bagging ensemble classifier.

Zhao et al. [23] propose a novel Ensemble Learning algorithm for anomaly detection on smart power grids, focusing on feature matching across a federated learning environment to determine if anomalous behaviour is the result of a physical fault (i.e., power line break due to weather or other environmental conditions) or actions of a malicious actor (i.e., network-based attack). The proposed model attempts to represent the smart power grid as a state machine, with normal behaviour modeled as state transitions that are processed with multiple base classifiers in an ensemble model to detect anomalous behaviour.

Tsogbaatar et al. [24] propose a hybrid Deep Learning and Ensemble Learning framework for anomaly detection on IoT devices using Software-Defined Networking (SDN) as the management plane for observation of classification of network flows. By implementing SDN in an IoT environment, full visibility of all the network flows can be obtained, as well as rapid attack mitigation through the rate-limiting or blocking of malicious traffic by the SDN controller. The proposed Ensemble Learning framework performs deep feature extraction with stacked autoencoders, which are used as inputs to Probabilistic Neural Networks (PNNs) for anomaly detection.

Zhong et al. [25] assert that the key limitations in the existing anomaly detection methods are related to low-quality training datasets and low levels of transference to different environments. A novel model of Ensemble Learning is proposed, starting with the use of autoencoders to extract features from raw data, and then training the autoencoders with a small amount of labeled data. The trained autoencoders are then used to train a Long Short-Term Memory (LTSM) model, which then uses different weights in multiple base classifiers to perform the final classification.

Zheng et al. [26] propose a method of anomaly detection called a Manifold regularization-based deep convolutional autoencoder (MR-DCAE) to identify unauthorized terrestrial radio broadcasts on licensed frequencies. While the proposed implementation has limited applicability outside its defined problem domain of illicit radio frequency usage, the underlying deep learning model is particularly robust with limited data and offers a degree of generalizability that is particularly useful in the highly diverse domain of CPSs.

Chen et al. [27] utilize a deep learning model based on a sliding-window convolutional variational autoencoder (SWCVAE) that focuses on multivariate time series data, which is commonly used for logging sensor and actuator readings in CPS environments. The focus of the proposed model is aimed at optimizing predictive maintenance scheduling for industrial robots based on hardware component failures rather than cyber threats, but is desirable from a system availability perspective to the operations of industrial CPS environments.

Yu et al. [28] propose a novel method of anomaly detection through combining the Ensemble Learning methods of convolutional variational autoencoders, resulting in a mixture-of-experts model. While extremely high accuracy levels can be obtained for vector datasets, the performance suffers on matrix datasets, resulting in limited generalizability in complex CPS environments.

A notable gap in the existing literature is the availability of robust and scalable methods to update the trained models as new data become available. The existing methods typically require retraining the entire model if the dataset experiences shifts in distribution over time, but this method is relatively time-consuming. Certain ensemble models such as online boosting or online random forests can perform incremental learning methods without retraining on the entire dataset, but can be susceptible to significant accuracy loss if the data distribution drifts excessively from the original dataset. As available computational capacity continues to increase, and if sufficiently large datasets are available, more complex deep learning models such as autoencoders may be able to provide greater scalability and improved accuracy.

This paper intentionally focuses on improving the predictive accuracy, even at the expense of additional computational time due to the unique motivators of the operations of CPS environments for extreme availability, making false positives or false negatives more costly than in IT-only networks. This paper focuses on EL as a method for improving accuracy, but it should be recognized that EL methods do have their own limitations, which are summarized as follows: Hyperparameter tuning for multiple base classifiers can increase the model training time and/or require increased computational resources. Additionally, the researcher must carefully select the base classifiers appropriate for the dataset, with sufficiently diverse base classifiers to be able to maximize strengths and minimize weaknesses, while being careful to avoid overfitting or amplifying the model bias. Interpretability is another challenge, as the multiple base classifiers may lead to nonintuitive or difficult-to-understand decisions when compared to those of single classification algorithms.

3. Materials and Methods

CPSs face unique challenges with using ML for anomaly detection, as there is frequently a large amount of available data that shows normal activity, but a lack of real-world data showing anomalous or malicious activity [29].

The lack of training data for anomalous activity has frequently been addressed by the use of artificially generated research datasets, which have varying degrees of fidelity to real-world environments. Due to the scarcity of training data, the operators of CPS environments are particularly concerned about predictive accuracy, as false positives or false negatives can have significant financial and/or quality of life impacts, especially in the case of critical national infrastructure, such as power grids and water treatment facilities.

For this reason, the operators of CPS environments are particularly interested in improving the predictive accuracy, even to the extent of gaining small incremental improvements that operators of traditional IT networks may dismiss as a too-costly trade-off in computational requirements with diminishing returns.

The approach proposed in this paper operates under the assumption that IT traffic and OT traffic are sufficiently distinct to warrant a hybridized threat detection strategy. Signature-based filters within the IT network are employed to screen out any malicious traffic originating from the Internet or potentially adversarial computers located on the IT network, as well as threshold-based methods to detect excessive IT resource utilization (CPU, RAM, etc.). Consequently, the security measures for CPS on OT networks should specifically target unforeseen behaviours that relate to the physical components of the CPS, which will include some threshold-based detection strategies for immutable physical characteristics of the CPS, such as temperature and pressure operating tolerances. These signature-based and threshold-based strategies will be addressed via a traditional IDS, while this paper focuses on performing behaviour-based anomaly detection in the OT portion of the CPS, with the proposed approach leveraging Ensemble Learning methods to achieve higher predictive performance than traditional ML classifiers. Additional details on the methodology used within the Ensemble Learning step are shown in Figure 1 and Algorithm 1.

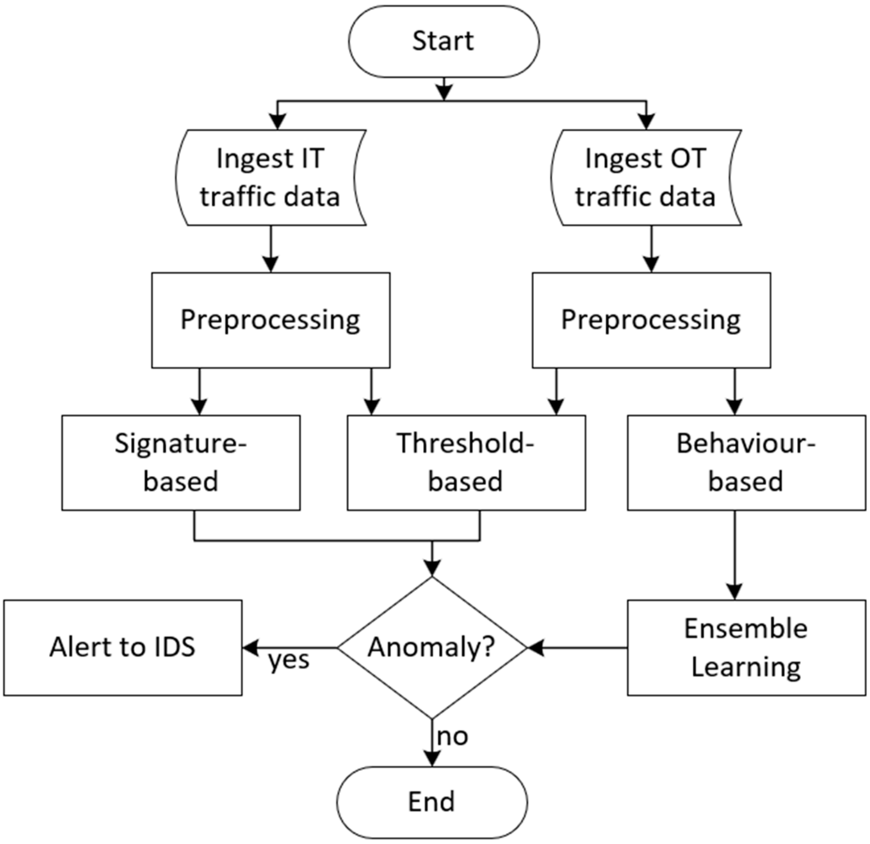

Figure 1. Logical workflow of threat detection approach. IT/OT traffic is ingested and preprocessed to eliminate unwanted or redundant data. Signature-based and threshold-based metrics are not added to an ML model, but sent directly to a rules engine, with normal traffic dropped, and anomalous traffic sent to an IDS for exception alerting. Behaviour-based metrics are fed to a learning model for deeper analysis, with traffic classified as anomalous forwarded to an IDS for exception alerting.

Algorithm 1. Pseudocode showing the algorithm used to combine multiple base classifiers into Ensemble Learning models.

```
FOR each base_classifier in [LR, NB, SVM, KNN, MLP]
    Initialize base_classifier
    Perform hyperparameter optimization
END FOR
FOR each ensemble_method in [voting, stacking, bagging, boosting]
    Combine optimized base_classifiers
    Calculate predictive performance
END FOR
Choose best ensemble_method
Make prediction
```

The diversity of endpoint devices in OT networks pose a challenge for accurate anomaly detection compared to the more uniform nature of IT networks, which often consist of large estates of near-identical Windows/Linux/etc., systems. To contrast, OT networks exhibit significant variations across different organizations. These differences extend to bespoke hardware specifically tailored for individual Cyber–Physical System (CPS) installations, introducing complexity to the signature-based and threshold-based detection strategies.

In response to this challenge, behaviour-based ML algorithms become a valuable tool for training models to comprehend the normal behaviour within the OT segment of the CPS, facilitating the identification of anomalous activities that deviate from the established norms in a particular environment.

It should be noted that certain anomaly detection classes in this paper intentionally exclude ML, relying on a Host-based Intrusion Detection System (HIDS) for IT components. Additionally, the physical components of the CPS, such as 5y3 operating temperature, pressure, vibration, and frequency of actuator duty cycles, often adhere to defined operational tolerances. These tolerances are typically dictated by life safety regulations and remain immutable, enabling swift and accurate validation through a straightforward threshold-based anomaly detection strategy. This approach bypasses the time-consuming and costly training of an ML model for threats related to these characteristics.

This paper concentrates on the behavioural characteristics of the OT portions of the CPS, focusing on Ensemble Learning as an effective means for improving predictive performance of ML algorithms for anomaly detection in CPS environments.

Furthermore, in this study, we assert that the problem of anomaly detection in CPS places a much higher value on predictive accuracy than the traditional IT networks, even at the cost of higher computational requirements due to the significant economic impact and life safety concerns related to CPS environments.

The traditional binary classification models are disadvantaged by highly imbalanced datasets, typically leading researchers to generate a balanced dataset by undersampling the normal class, thereby inducing a bias on the models. Additionally, it may lead to learning models with poor generalization capabilities, rendering them incapable of facing cyber-attacks that were not present in the training data. However, combining multiple traditional binary classification models into an Ensemble Learning model maximizes the strengths and minimize the weaknesses of multiple binary classifiers, leading to higher predictive accuracy that is so sought after in CPS environments.

This paper will make use of the following traditional classification models as base classifiers: Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Multi-Level Perceptron (MLP). These particular base classifiers [30] were selected due to their high diversity, which improves the predictive performance in the subsequent ensemble models through maximizing the strengths and minimizing the weaknesses of each base classifier. For example, LR and MLP complement each other because LR excels at capturing the linear patterns in data, while MLP excels at capturing the non-linear patterns. NB was selected due to its resistance to bias in datasets containing irrelevant or minimally correlated features, while SVM and KNN complement each other due to differing strengths in highly dimensional datasets. These base classifiers will be aggregated and combined into Voting/Stacking/Bagging/Boosting ensemble classifiers in an attempt to improve the predictive performance to a level greater than can be achieved by the base classifiers. A brief description of each classification algorithm [30] is provided in Table 1.

Table 1. Summary of classification algorithms.

Base Classifiers	
LR	Logistic Regression is a supervised learning classification algorithm that employs a linear model for binary classification, which uses a logistic function to predict the probability of a data point belonging to one of the available classes.
NB	There are multiple variations of Naïve Bayes, all of which are based on Bayes' theorem. This paper employs the Bernoulli variation of NB, a supervised learning classification algorithm commonly used for binary or multiclass classification. A unique feature of Bernoulli NB is the assumption that features are conditionally independent, making it particularly well suited for binary classification tasks.
SVM	Support Vector Machine is a supervised learning algorithm used for both classification and regression tasks by finding a hyperplane to separate the data into multiple classes. SVM is particularly useful in highly dimensional datasets that are common in CPS environments.
KNN	K-Nearest Neighbor is a supervised learning algorithm used for both classification and regression tasks and works by assigning a data point to a particular class based on the labels of its nearest neighbors. This algorithm is commonly used due to its simplicity and low computational requirements.
MLP	Multi-Level Perceptron is a supervised learning algorithm that performs both classification and regression tasks using a neural network consisting of multiple layers of nodes (perceptrons) with non-linear activation functions. MLP is particularly effective in detecting complex patterns in data that can be missed by other algorithms.
Ensemble Classifiers	
Voting	Voting is a simple ensemble method where multiple individual models (classifiers or regressors) are trained independently, and their predictions are combined to make a final decision. In a majority voting scheme (hard voting), the class predicted by the majority of models is chosen. In weighted voting (soft voting), models may contribute different weights to their predictions.
Stacking	Stacking, or stacked generalization, involves training multiple diverse models and combining their predictions using a meta-model (also known as a blender or meta-classifier). The base models' predictions serve as input features for the meta-model. Stacking aims to leverage the strengths of different models and improve overall predictive performance.
Bagging	Bagging (Bootstrap aggregating) is an ensemble technique where multiple instances of the same learning algorithm are trained on different subsets of the training data, created through bootstrap sampling (random sampling with replacement). The final prediction is obtained by averaging (for regression) or voting (for classification) over the predictions of individual models.
Boosting	Boosting is an ensemble method that focuses on sequentially training weak learners (models slightly better than random guessing) and giving more weight to instances that are misclassified by previous models. The final prediction is a weighted combination of the weak learners' predictions, which minimizes the weaknesses and maximizes the strengths of the base classifiers to achieve higher accuracy than would be possible through other ensemble methods such as voting.

4. Experimental Evaluation

4.1. Description of Datasets

This study intentionally makes use of multiple datasets from diverse environments, with the goal of showing that EL is an effective method of anomaly detection across a broad range of CPS environments. Each of the selected datasets have sufficient popularity in academic research, making it possible to compare single ML classification models against the EL methods to validate their increased predictive performance through the use of multiple algorithms.

As is typical for anomaly detection in CPS environments, each of the available datasets have imbalanced classes, which causes overfitting and inaccurate correlations with many

ML algorithms, which is a shortcoming that the use of EL is intended to alleviate. A summary of the selected public datasets is shown in Table 2 and described in further detail in the subsequent sections.

Table 2. Summary of datasets.

Dataset	Rows	Columns	Normal Data	Anomaly Data
Edge-IIoTset2023	2,219,201	63	85.9%	14.1%
CICIot2023	2,867,734	46	38.3%	61.7%

4.2. Edge-IIoTset2023

The Edge-IIoTset2023 dataset [31] was developed in 2023 as a comprehensive dataset that can be used to develop and accurately validate IoT/IIoT security solutions. The dataset was collected from a sophisticated seven-layer testbed, including more than 10 IoT devices, IIoT-based Modbus flows, and 14 IoT and IIoT protocol-related attacks. The attacks include Mirai-udpplain, MITM-ArpSpoofing, DNS_Spoofing, Recon-PingSweep, Recon-PortScan, Recon-OSScan, Recon-HostDiscovery, XSS, CommandInjection, VulnerabilityScan, Backdoor_Malware, BrowserHijacking, DictionaryBruteForce, SqlInjection, and Uploading_attack.

This dataset was generated from the direct observation of a controlled testbed environment, with normal activity and attack activity generated by the researchers in a controlled environment, which may result in limited fidelity to IIoT environments directly connected to the less-predictable public internet.

Prior to pre-processing, the dataset contained ~22.2 million lines, of which ~20.3 million lines were comprised of DDoS-related attacks, which severely skewed the balance of class distribution. Since DDoS attacks can be considered a threshold-based metric, they are best addressed via perimeter firewalls rather than in an ML model, so the DDoS entries were dropped from this dataset in order to focus the Ensemble Learning model on behavioural aspects of the CPS.

After pre-processing, this dataset contains 2,291,201 rows and 63 columns, with 85.9% normal data and 14.1% anomaly data. To aid in reproducibility efforts by future researchers, the dataset pre-processing steps are provided at https://github.com/nickjeffrey/ensemble_learning (accessed 5 April 2024).

4.3. CICIoT2023

The CICIoT2023 dataset [32] was developed in 2023 by the Canadian Institute for Cybersecurity in cooperation with the University of New Brunswick to generate a research dataset for large-scale attacks in IoT environments. To provide contrast to the previous dataset, this dataset focuses more heavily on the cyber portions than the physical portions of a CPS environment, providing 33 different cyber-attacks against 105 different IoT devices. The different attacks can be broadly classified into these categories: DDoS, DoS, Recon, Web-based, Brute Force, spoofing, and Mirai.

This dataset was generated from the direct observation of a real IoT environment, but the attack scenarios were generated by the researchers in a controlled environment, which may result in limited fidelity to IoT environments directly connected to the less-predictable public internet.

Prior to pre-processing, the dataset contained ~45.6 million lines, of which ~43.8 million lines were comprised of DDoS-related attacks, which severely skewed the balance of class distribution. Since DDoS attacks can be considered a threshold-based metric, they are best addressed via perimeter firewalls rather than in an ML model, so the DDoS entries were dropped from this dataset in order to focus the Ensemble Learning model on behavioural aspects of the CPS.

After pre-processing, this dataset contains 2,867,734 rows and 46 columns, with 61.7% anomaly data, and 38.3% normal data. To aid in reproducibility efforts by future researchers, the dataset pre-processing steps are provided at https://github.com/nickjeffrey/ensemble_learning (accessed 5 April 2024).

While this dataset is still imbalanced, it should be noted that the imbalance is in the opposite direction of the previous dataset, with this dataset having a minority of the data in the normal class. The two public research datasets in this paper were intentionally chosen because of their differences in order to show the broad applicability of Ensemble Learning as a method of improving the predictive performance.

4.4. Experiment Setup

Due to the imbalanced nature of the datasets, it is expected that typical ML classification models such as LR/NB/SVM/KNN/MLP will suffer from some degree of overfitting, thus reducing the model's accuracy.

This study intentionally uses multiple public research datasets from diverse sources and with dissimilar features to demonstrate the broad applicability of EL methods.

After balancing the data classes via random undersampling of the majority class, in order to minimize model bias, a portion of each class was reserved for final validation. After the data were resampled to balance the classes, they were normalized, and a 10-fold cross validation was carried out to minimize the distribution bias.

The final validation stage includes all the data that have not been used in training and testing; this is an imbalanced dataset containing instances from normal traffic and from anomalies. The aim of this validation stage is to compare the behaviour of the different modeling techniques included in this comparison, so conclusions could be extracted.

In order to avoid drawing conclusions from biased data, the whole procedure was repeated 10 times from the balanced dataset to the final validation stage, with the obtained partial results aggregated for comparison of different learning models.

Due to the diverse nature of CPS environments, hyperparameter optimization is particularly important when training multiple conventional ML algorithms on different datasets. For this reason, the proposed model makes no assumptions about parameter settings for each of the base classifiers and instead employs a Grid Search Cross-Validation technique to tune each model parameter to find the optimal parameter settings for a particular dataset. The specific hyperparameter tuning details for each of the base classifiers is expected to vary by dataset due to the heterogeneous nature of CPS environments. For example, one dataset may perform better using the SVM base classifier with an RBF kernel, while another dataset may use a sigmoid kernel. The dataset-specific parameter settings used in the proposed model were programmatically determined and are provided in the Supplementary Materials at https://github.com/nickjeffrey/ensemble_learning (accessed 5 April 2024).

To measure the quality of each individual model, the accuracy, Sensitivity, Specificity, Geometric Mean, Precision, Recall, and F1-score measurements will be used. The accuracy will give indications of the performance on the balanced dataset, while Sensitivity (True Positive rate) and Specificity (True Negative rate) will help in the final validation stage, where the data will be clearly imbalanced. The Geometric Mean provides a balanced measure of Sensitivity and Specificity. Precision denotes the proportion of correct attack classes to the total amount of predicted attack results. Recall denotes the proportion of proper attack classifications relative to the overall count of all the samples that should have been identified as attacks. The F1-score is the Harmonic Mean between Precision and Recall. The equations used to calculate these metrics are shown below. The following abbreviations are used in the equations below: True Positive (TP), True Negative False Positive (FP), and false negative (FN).

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}), \tag{1}$$

$$\text{Sensitivity} = \text{TP}/(\text{TP} + \text{FN}), \tag{2}$$

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP}) \tag{3}$$

$$\text{Geometric Mean} = \text{sqrt}\ (\text{Sensitivity} \times \text{Specificity}) \tag{4}$$

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{5}$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{6}$$

$$\text{F1-score} = (2 \times \text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall}) \tag{7}$$

The above steps are performed for each of the individual ML classification models, and the results are used as inputs for an EL model that will use voting/stacking/bagging/boosting models to further improve predictive performance, as illustrated in Figure 2. It is worthwhile to note that a recurring challenge in anomaly detection for CPS is the lack of comprehensive research datasets with fidelity to real-world environments. Due to the high heterogeneity across CPS environments, the authors are not suggesting one particular EL method to be universally superior to the others. Instead, this proposed model automatically tests different EL methods (voting, stacking, bagging, and boosting) and selects the method that provides the highest accuracy for a particular dataset to further the goal of a generalizable model for anomaly detection across diverse CPS environments.

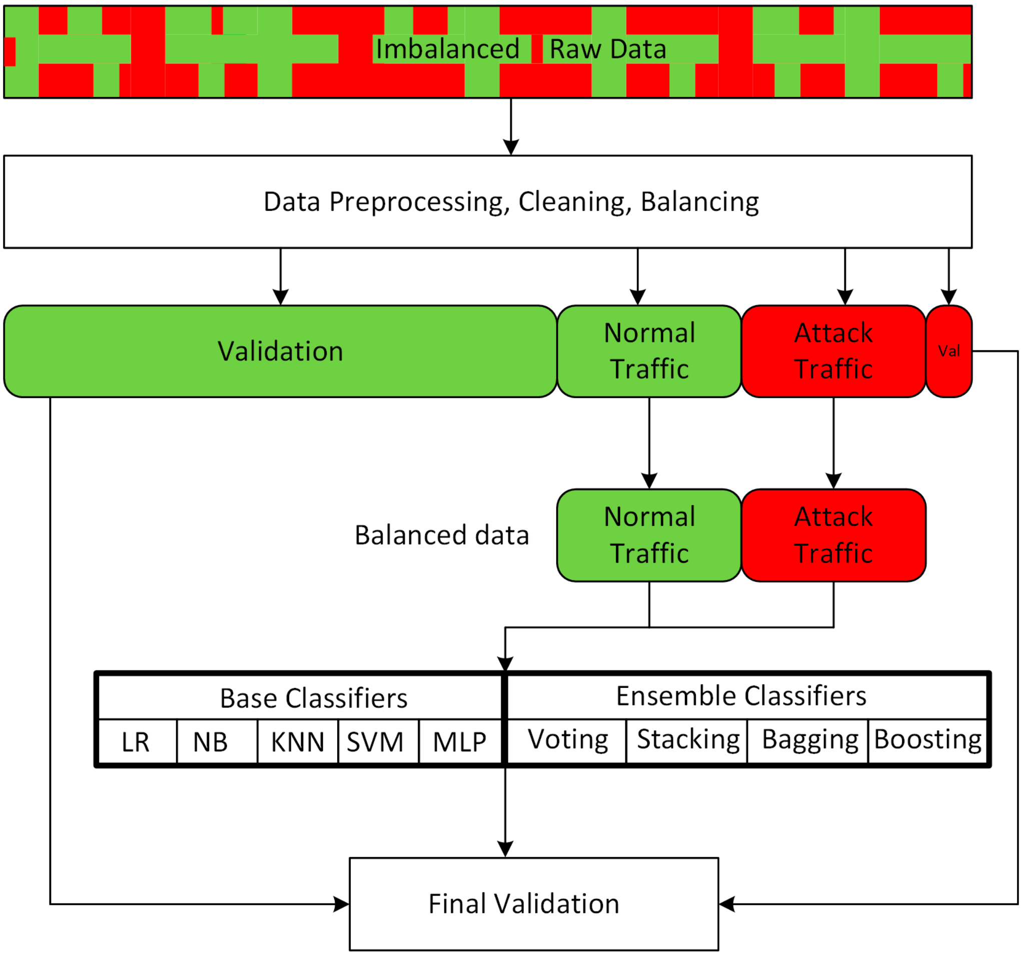

Figure 2. The experimental setup followed in this research. The raw data must undergo preprocessing, cleaning, and balancing to minimize the model bias. After separating the data into training and validation samples, the balanced portion of the normal (green) and anomaly (red) dataset was used for training and testing the base classification models and the ensemble models. The remaining data were used in the final validation, and the resulting predictive performances of the base and ensemble classifiers were compared.

5. Results

The initial hypothesis of this paper was that EL could provide an improved predictive performance of threats to CPSs when compared to those of the traditional ML classification models. This hypothesis was confirmed by the experiments for both of the public research datasets employed in this paper.

Of all the ensemble classifiers tested in this paper, the results were similar for both of the tested datasets, indicating the general applicability of the original hypothesis across diverse CPS environments. The Voting and Stacking ensemble methods showed minor accuracy improvements of 3–4% over those of the traditional classification models for either dataset, which is to be expected due to the lack of extreme variation or diversity in the accuracy levels of the base classifiers.

The Bagging ensemble method also shows minor partial improvements, with a ~3% accuracy increase for the Edge-IIoTset2023 dataset and ~5% accuracy increase for the CICIoT2023 dataset due to the differences in data variance in the underlying datasets, making the Bagging method slightly less able to generalize across multiple CPS environments. This very slight improvement is small enough that we consider the Bagging ensemble method to have no significant improvement over the traditional classification models.

The Boosting ensemble method was the most effective across both datasets, with ~10% improvement in the mean accuracy for the Edge-IIoTset2023 dataset and ~8% improvement in the mean accuracy for the CICIoT2023 dataset. This is to be expected, as the Boosting classifier is designed to train multiple weak learners, and then adjust the weights of the base classifiers to minimize the weaknesses and maximize the strengths, resulting in higher accuracy levels than can be obtained from the base classifiers. The detailed experimental results are shown in Tables 3–6. The results are visualized in Figures 3–6.

Table 3. Cross-validation scores for the Edge-IIoTset2023 dataset. Best mean score is indicated in bold type.

Fold	Individual Classifiers					Ensemble Classifiers			
	LR	NB	SVM	KNN	MLP	Voting	Stacking	Bagging	Boosting
1	0.8251	0.7174	0.8323	0.8282	0.8395	0.8364	0.8344	0.8251	0.9369
2	0.8199	0.7019	0.8302	0.8219	0.824	0.8313	0.8292	0.8219	0.9224
3	0.8458	0.7536	0.8582	0.8385	0.8478	0.8602	0.8561	0.8478	0.9275
4	0.8344	0.7246	0.8478	0.8385	0.852	0.8509	0.8416	0.8354	0.9369
5	0.8313	0.7091	0.8458	0.8437	0.8489	0.8468	0.8458	0.8302	0.9369
6	0.8219	0.7267	0.8416	0.8427	0.8427	0.8437	0.8416	0.823	0.9369
7	0.8282	0.7122	0.8406	0.8416	0.8427	0.8416	0.8551	0.8282	0.9224
8	0.8199	0.7091	0.8292	0.8188	0.8282	0.8333	0.8313	0.8178	0.9141
9	0.8292	0.7298	0.8571	0.8716	0.8551	0.8551	0.8613	0.8292	0.9369
10	0.8261	0.7277	0.8323	0.8395	0.8375	0.8385	0.8416	0.8251	0.9379
Mean	0.8282	0.7212	0.8415	0.8385	0.8418	0.8438	0.8438	0.8284	**0.9308**
StdDev	0.0074	0.014	0.0101	0.0139	0.0095	0.009	0.0103	0.008	0.0082

Table 4. Summary of performance metrics for the Edge-IIoTset2023 dataset. Best results are indicated in bold type.

Metric	Individual Classifiers					Ensemble Classifiers			
	LR	NB	SVM	KNN	MLP	Voting	Stacking	Bagging	Boosting
Accuracy	0.8733	0.7699	0.8875	0.8687	0.8789	0.8885	0.9013	0.8852	**0.9601**
Sensitivity	0.7515	0.6209	0.7537	0.7713	0.7544	0.7584	0.7298	0.7643	**0.8772**
Specificity	0.9188	0.8254	0.9374	0.905	0.9253	0.937	0.9652	0.9304	**0.991**
GeoMean	0.8309	0.7159	0.8406	0.8355	0.8355	0.843	0.8393	0.8432	**0.9324**
Precision	0.8722	0.7767	0.8855	0.8698	0.8774	0.8866	0.9004	0.8838	**0.9606**
Recall	0.8733	0.7699	0.8875	0.8687	0.8789	0.8885	0.9013	0.8852	**0.9601**
F1-score	0.8727	0.7728	0.886	0.8692	0.878	0.8871	0.8981	0.8843	**0.9594**

Table 5. Cross-validation scores for the CICIoT2023 dataset. Best mean score is indicated in bold type.

Fold	Individual Classifiers					Ensemble Classifiers			
	LR	**NB**	**SVM**	**KNN**	**MLP**	**Voting**	**Stacking**	**Bagging**	**Boosting**
1	0.8629	0.7821	0.8817	0.8555	0.8811	0.8845	0.8845	0.8629	0.9226
2	0.8641	0.7901	0.8851	0.8595	0.8999	0.8942	0.897	0.8663	0.9357
3	0.8669	0.7867	0.884	0.8714	0.9016	0.8942	0.8965	0.8663	0.9403
4	0.8475	0.7843	0.8953	0.8651	0.8896	0.8902	0.8958	0.8463	0.926
5	0.8429	0.7615	0.8799	0.8526	0.889	0.8856	0.8902	0.8435	0.938
6	0.8662	0.7894	0.8924	0.8645	0.8981	0.893	0.8987	0.8657	0.9334
7	0.8668	0.786	0.8896	0.8657	0.8976	0.8987	0.8987	0.8674	0.9323
8	0.8714	0.7849	0.8964	0.8571	0.8964	0.9038	0.9004	0.8708	0.9402
9	0.8617	0.7883	0.8896	0.8634	0.9112	0.8987	0.9067	0.864	0.9391
10	0.8645	0.8059	0.8919	0.8668	0.905	0.897	0.905	0.8651	0.9345
Mean	0.8615	0.7859	0.8886	0.8622	0.8969	0.894	0.8973	0.8618	**0.9342**
StdDev	0.0086	0.0102	0.0054	0.0055	0.0082	0.0057	0.0061	0.0087	0.0056

Table 6. Summary of performance metrics for the CICIoT2023 dataset. Best results are indicated in bold type.

Metric	Individual Classifiers					Ensemble Classifiers			
	LR	**NB**	**SVM**	**KNN**	**MLP**	**Voting**	**Stacking**	**Bagging**	**Boosting**
Accuracy	0.8506	0.77	0.8763	0.847	0.8931	0.8842	0.8925	0.8939	**0.9319**
Sensitivity	0.7979	0.7174	0.8212	0.7863	0.8728	0.8415	0.865	0.8664	**0.9162**
Specificity	0.9357	0.8549	0.9562	0.945	0.926	0.953	0.9369	0.9382	**0.9574**
GeoMean	0.864	0.7831	0.8903	0.862	0.899	0.8955	0.9002	0.9016	**0.9366**
Precision	0.8717	0.7981	0.8961	0.8721	0.8997	0.8984	0.9011	0.9024	**0.9353**
Recall	0.8506	0.77	0.8763	0.847	0.8931	0.8842	0.8925	0.8939	**0.9319**
F1-score	0.8527	0.7732	0.878	0.8492	0.8942	0.8856	0.8937	0.895	**0.9324**

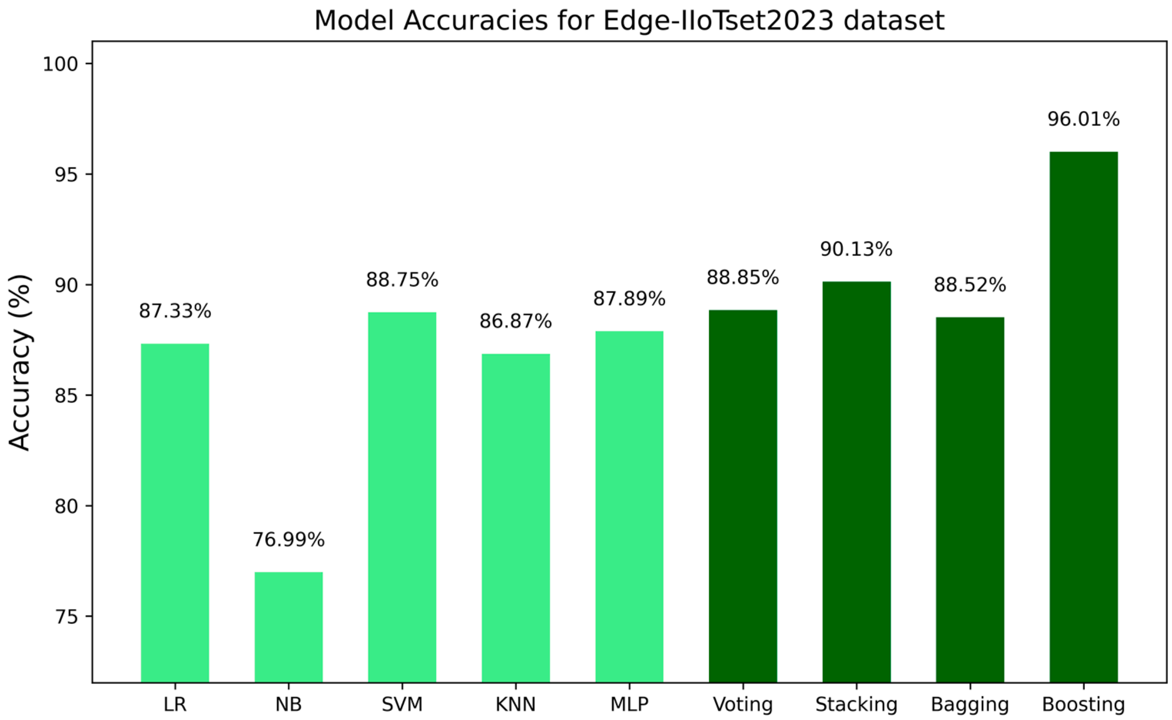

Figure 3. Experimental results for Edge-IIoTset2023 dataset, showing model accuracy for each base classifier and ensemble model.

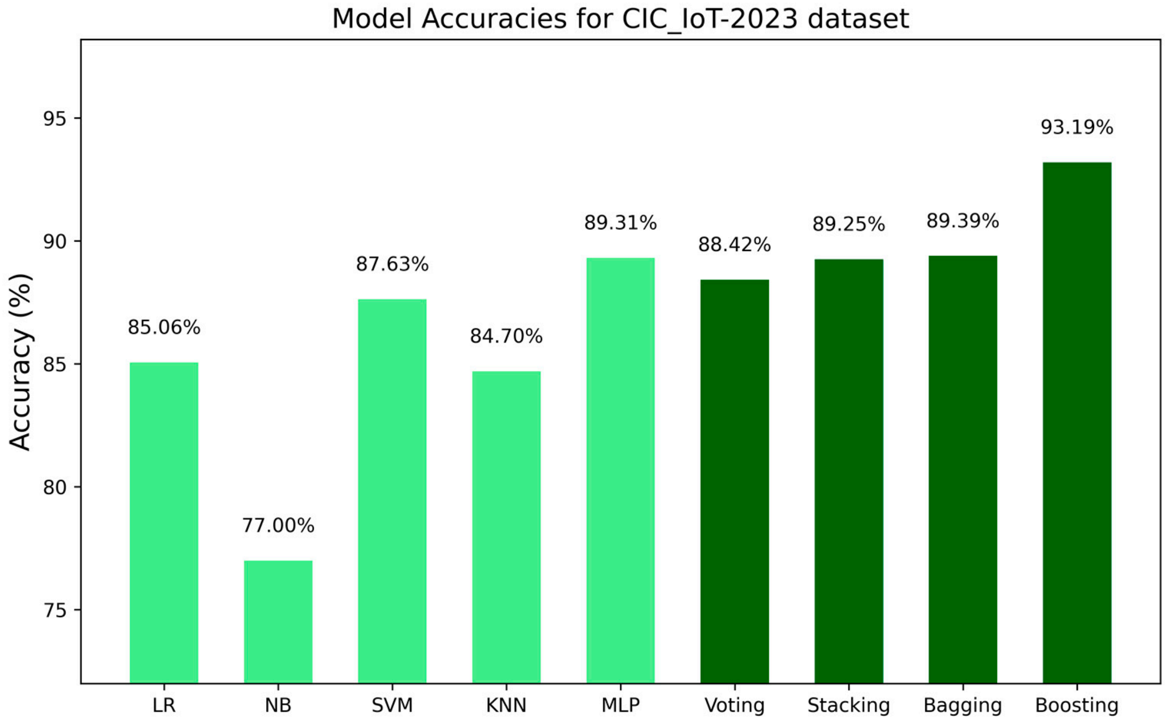

Figure 4. Experimental results for CIC_IoT_2023 dataset, showing model accuracy for each base classifier and ensemble model.

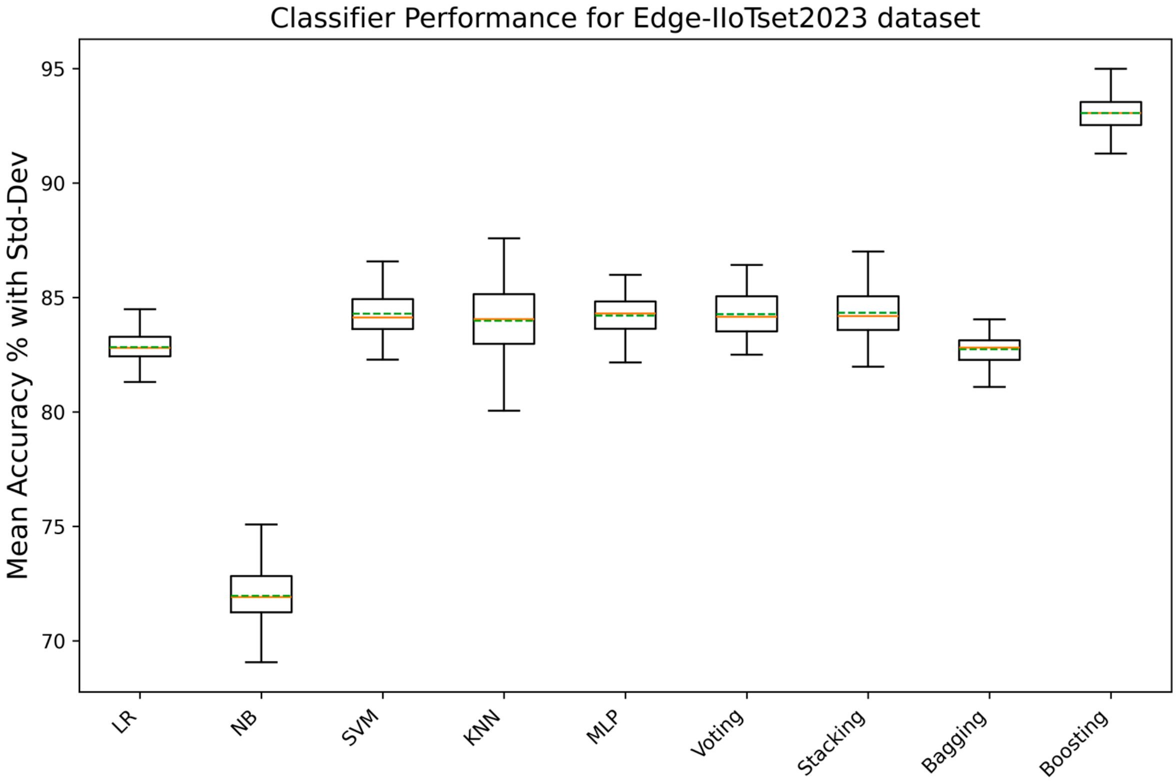

Figure 5. Experimental results for Edge-IIoTset2023 dataset, showing mean accuracy and standard deviation for 10-fold cross-validation for each base classifier and ensemble model.

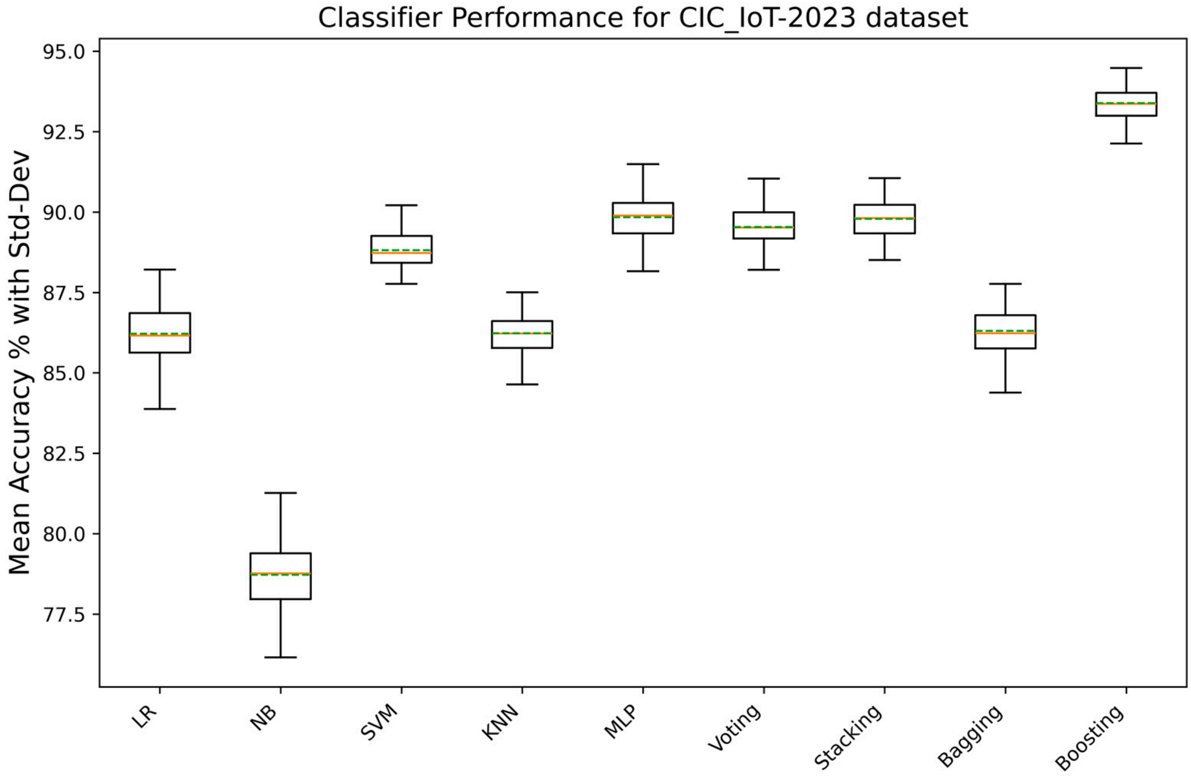

Figure 6. Experimental results for CIC_IoT_2023 dataset, showing mean accuracy and standard deviation for 10-fold cross-validation for each base classifier and ensemble model.

6. Conclusions and Future Works

Threat detection in CPSs poses unique challenges, with the differing security postures of IT and OT networks making it difficult to provide a unified threat detection strategy. This paper proposes a hybrid methodology that leverages signature-based detection strategies for known threats, threshold-based detection strategies for the immutable properties of the CPS, and the combination of multiple ML algorithms with Ensemble Learning for the behaviour-based detection of anomalies, with the goal of providing higher accuracy via EL than that which can be achieved with the traditional ML methods.

This paper details experiments with multiple classification algorithms against imbalanced datasets, which are then combined into an Ensemble Learning model, starting from the hypothesis that because malicious activity makes up a very small portion of the datasets due to its natural scarcity, the traditional ML classifiers suffer due to the imbalanced nature of the datasets, so greater accuracy can be obtained by combining multiple classification algorithms into an EL model. The results show that EL improves the predictive performance across multiple datasets by 4–7% above that of the highest-performing base classifier, which is particularly important to the operators of CPSs due to the high financial and life safety costs associated with interruptions to system availability.

The future works include continued investigations into increasing accuracy through the use of more complex learning models, including deep learning for large and heterogeneous environments, decision threshold tuning to minimize misclassification in imbalanced datasets, and the further development of complementary detection methodologies that combine ML algorithms for OT networks with signature-based and threshold-based detection strategies for the IT components of CPS. For many cybersecurity use cases, the occasional false negative is acceptable, but CPS environments are particularly intolerant of false negatives due to the low risk/high impact consequences of an interruption to critical CPS infrastructure, making further investigation and threshold tuning a worthy avenue for future works.

Supplementary Materials: To aid in reproducibility efforts by future researchers, the dataset pre-processing steps are provided at https://github.com/nickjeffrey/ensemble_learning (accessed 5 April 2024).

Author Contributions: Conceptualization, N.J., Q.T. and J.R.V.; methodology, N.J., Q.T. and J.R.V.; software, N.J.; validation, N.J., Q.T. and J.R.V.; formal analysis, N.J., Q.T. and J.R.V.; investigation, N.J., Q.T. and J.R.V.; resources, N.J.; data curation, N.J.; writing—original draft preparation, N.J.; writing—review and editing, N.J., Q.T. and J.R.V.; visualization, N.J.; supervision, Q.T. and J.R.V.; project administration, N.J.; funding acquisition, Q.T. and J.R.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been funded by the Spanish Ministry of Economics and Industry, grant PID2020-112726RB-I00; the Spanish Research Agency, grant PID2023-146257OB-I00; the Missions Science and Innovation project MIG-20211008 (INMERBOT); by the Principado de Asturias, grant SV-PA-21-AYUD/2021/50994; and by the Council of Gijón through the University Institute of Industrial Technology of Asturias, grants SV-21-GIJON-1-19, SV-22-GIJON-1-19, SV-22-GIJON-1-22, and SV-23-GIJON-1-09. Finally, this research has also been funded by Fundación Universidad de Oviedo via grants FUO-23-008 and FUO-22-450.

Data Availability Statement: This article makes use of public research datasets that are available from their respective authors using the links provided in the References section of this article. To aid in reproducibility efforts by other researchers, the Ensemble Learning methods and procedures described in this paper are available in jupyter notebook format at https://github.com/nickjeffrey/ensemble_learning (accessed 5 April 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kagermann, H.; Wahlster, W. Ten Years of Industrie 4.0. *Sci* **2022**, *4*, 26. [CrossRef]
2. Ashibani, Y.; Mahmoud, Q.H. Cyber physical systems security: Analysis, challenges and solutions. *Comput. Secur.* **2017**, *68*, 81–97. [CrossRef]
3. Yaacoub, J.-P.A.; Salman, O.; Noura, H.N.; Kaaniche, N.; Chehab, A.; Malli, M. Cyber-physical systems security: Limitations, issues and future trends. *Microprocess. Microsyst.* **2020**, *77*, 103201. [CrossRef] [PubMed]
4. Rakas, S.V.B.; Stojanovic, M.D.; Markovic-Petrovic, J.D. A Review of Research Work on Network-Based SCADA Intrusion Detection Systems. *IEEE Access* **2020**, *8*, 93083–93108. [CrossRef]
5. Stout, W.M. Toward a Multi-Agent System Architecture for Insight & Cybersecurity in Cyber-Physical Networks. In Proceedings of the 2018 International Carnahan Conference on Security Technology (ICCST), Montreal, QC, Canada, 22–25 October 2018; pp. 1–5. [CrossRef]
6. Altunay, H.C.; Albayrak, Z.; Özalp, A.N.; Çakmak, M. Analysis of Anomaly Detection Approaches Performed Through Deep Learning Methods in SCADA Systems. In Proceedings of the 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 11–13 June 2021; pp. 1–6. [CrossRef]
7. Boateng, E.A.; Bruce, J.W. Unsupervised Machine Learning Techniques for Detecting PLC Process Control Anomalies. *J. Cybersecur. Priv.* **2022**, *2*, 220–244. [CrossRef]
8. Boateng, E.A.; Bruce, J.W.; Talbert, D.A. Anomaly Detection for a Water Treatment System Based on One-class Neural Network. *IEEE Access* **2022**, *10*, 115179–115191. [CrossRef]
9. Cagnini, H.E.L.; Das Dôres, S.C.N.; Freitas, A.A.; Barros, R.C. A survey of evolutionary algorithms for supervised ensemble learning. *Knowl. Eng. Rev.* **2023**, *38*, e1. [CrossRef]
10. Xu, S.; Qian, Y.; Hu, R.Q. Data-Driven Edge Intelligence for Robust Network Anomaly Detection. *IEEE Trans. Netw. Sci. Eng.* **2019**, *7*, 1481–1492. [CrossRef]
11. Vasan, D.; Alazab, M.; Venkatraman, S.; Akram, J.; Qin, Z. MTHAEL: Cross-Architecture IoT Malware Detection Based on Neural Network Advanced Ensemble Learning. *IEEE Trans. Comput.* **2020**, *69*, 1654–1667. [CrossRef]
12. Liu, L.; Wu, X.; Li, S.; Li, Y.; Tan, S.; Bai, Y. Solving the class imbalance problem using ensemble algorithm: Application of screening for aortic dissection. *BMC Med. Informatics Decis. Mak.* **2022**, *22*, 82. [CrossRef]
13. Dasarathy, B.; Sheela, B. A composite classifier system design: Concepts and methodology. *Proc. IEEE* **1979**, *67*, 708–713. [CrossRef]
14. Jeffrey, N.; Tan, Q.; Villar, J.R. Intrusion Detection and Prevention in Industrial Internet of Things: A Study. In Proceedings of the International Joint Conference 16th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2023) 14th International Conference on EUropean Transnational Education (ICEUTE 2023), Salamanca, Spain, 5–7 September 2023; Volume 748, pp. 37–48. [CrossRef]
15. Jeffrey, N.; Tan, Q.; Villar, J.R. A Review of Anomaly Detection Strategies to Detect Threats to Cyber-Physical Systems. *Electronics* **2023**, *12*, 3283. [CrossRef]

16. Jeffrey, N.; Tan, Q.; Villar, J.R. A hybrid methodology for anomaly detection in Cyber–Physical Systems. *Neurocomputing* **2023**, *568*, 127068. [CrossRef]
17. Afrifa, S.; Varadarajan, V.; Appiahene, P.; Zhang, T.; Domfeh, E.A. Ensemble Machine Learning Techniques for Accurate and Efficient Detection of Botnet Attacks in Connected Computers. *Eng* **2023**, *4*, 650–664. [CrossRef]
18. Araya, D.B.; Grolinger, K.; ElYamany, H.F.; Capretz, M.A.; Bitsuamlak, G. An ensemble learning framework for anomaly detection in building energy consumption. *Energy Build.* **2017**, *144*, 191–206. [CrossRef]
19. Yazdinejad, A.; Kazemi, M.; Parizi, R.M.; Dehghantanha, A.; Karimipour, H. An ensemble deep learning model for cyber threat hunting in industrial internet of things. *Digit. Commun. Netw.* **2023**, *9*, 101–110. [CrossRef]
20. Saharkhizan, M.; Azmoodeh, A.; Dehghantanha, A.; Choo, K.-K.R.; Parizi, R.M. An Ensemble of Deep Recurrent Neural Networks for Detecting IoT Cyber Attacks Using Network Traffic. *IEEE Internet Things J.* **2020**, *7*, 8852–8859. [CrossRef]
21. Danso, P.K.; Neto, E.C.P.; Dadkhah, S.; Zohourian, A.; Molyneaux, H.; Ghorbani, A.A. Ensemble-based Intrusion Detection for Internet of Things Devices. In Proceedings of the 2022 IEEE 19th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET), Marietta, GA, USA, 19–21 December 2022; pp. 34–39. [CrossRef]
22. Illy, P.; Kaddoum, G.; Moreira, C.M.; Kaur, K.; Garg, S. Securing Fog-to-Things Environment Using Intrusion Detection System Based On Ensemble Learning. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, 15–18 April 2019; pp. 1–7. [CrossRef]
23. Zhao, H.; Li, C.; Yin, X.; Li, X.; Zhou, R.; Fu, R. Ensemble Learning-Enabled Security Anomaly Identification for IoT Cyber–Physical Power Systems. *Electronics* **2022**, *11*, 4043. [CrossRef]
24. Tsogbaatar, E.; Bhuyan, M.H.; Taenaka, Y.; Fall, D.; Gonchigsumlaa, K.; Elmroth, E.; Kadobayashi, Y. DeL-IoT: A deep ensemble learning approach to uncover anomalies in IoT. *Internet Things* **2021**, *14*, 100391. [CrossRef]
25. Zhong, Y.; Chen, W.; Wang, Z.; Chen, Y.; Wang, K.; Li, Y.; Yin, X.; Shi, X.; Yang, J.; Li, K. HELAD: A novel network anomaly detection model based on heterogeneous ensemble learning. *Comput. Netw.* **2019**, *169*, 107049. [CrossRef]
26. Zheng, Q.; Zhao, P.; Zhang, D.; Wang, H. MR-DCAE: Manifold regularization-based deep convolutional autoencoder for unauthorized broadcasting identification. *Int. J. Intell. Syst.* **2021**, *36*, 7204–7238. [CrossRef]
27. Chen, T.; Liu, X.; Xia, B.; Wang, W.; Lai, Y. Unsupervised Anomaly Detection of Industrial Robots Using Sliding-Window Convolutional Variational Autoencoder. *IEEE Access* **2020**, *8*, 47072–47081. [CrossRef]
28. Yu, Q.; Kavitha, M.S.; Kurita, T. Mixture of experts with convolutional and variational autoencoders for anomaly detection. *Appl. Intell.* **2020**, *51*, 3241–3254. [CrossRef]
29. Tama, B.A.; Lim, S. Ensemble learning for intrusion detection systems: A systematic mapping study and cross-benchmark evaluation. *Comput. Sci. Rev.* **2020**, *39*, 100357. [CrossRef]
30. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2012**, *12*, 2825–2830. [CrossRef]
31. Ferrag, M.A.; Friha, O.; Hamouda, D.; Maglaras, L.; Janicke, H. Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning. *IEEE Access* **2022**, *10*, 40281–40306. [CrossRef]
32. Neto, E.C.P.; Dadkhah, S.; Ferreira, R.; Zohourian, A.; Lu, R.; Ghorbani, A.A. CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment. *Sensors* **2023**, *23*, 5941. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.