Towards a SCADA Cybersecurity Model with Machine Learning Algorithms

Takudzwa Vincent Banda[1[0009-0003-2900-5518]], Dewald Blaauw[2[0002-0002-2188-5931]] and Bruce Watson[3[0000-0003-0511-1837]]

Centre for AI Research (CAIR), School of Data Science & Computational Thinking and Department of Information Science, Stellenbosch University

[1] Stellenbosch University, South Africa
tadiwanashebanda74@gmail.com
[2] Stellenbosch University, South Africa
dnblaauw@sun.ac.za
[3] Stellenbosch University, South Africa
bruce@bruce-watson.com
http://www.bruce-watson.com

Abstract. The escalating threat landscape on South Africa's critical infrastructure cyberattacks demands innovative solutions that harness the synergy between AI and cybersecurity. This research delves into the intersection of these domains, utilising machine learning algorithms to bolster Supervisory Control and Data Acquisition (SCADA) networks of such critical infrastructures. These networks, pivotal for controlling and monitoring critical infrastructures, collect and process sensitive data. The approach melds machine learning's predictive prowess with cybersecurity imperatives, resulting in a resilient model capable of preemptively identifying and thwarting cyber threats—an aspect often lacking in existing models. To achieve this, a comprehensive methodology involves the emulation of cyberattacks through a Kali Linux machine integrated into a simulated SCADA network. Leveraging Wireshark, network traffic is collected for machine learning insights. The solution emerges in the form of a dual-tasked ensemble supervised machine learning model, combining the power of a Multi-Layer Perceptron (MLP) Neural Network and Extreme Gradient Boosting (XGBoost). This synergy equips the model to decipher subtle patterns and anomalies within network traffic. With an average accuracy of 99.60% and a detection rate of 99.48%, the model effectively discriminates between normal and suspicious states, while proactively preventing malicious packet infiltration. Through PowerShell, the model enforces firewall rules - an embodiment of AI's adaptive governance - endowing it with a dual role as both an Intrusion Detection and Prevention Model. Rigorous testing on previously unseen data validates the model's performance, boasting 99.19% accuracy and 98.95% detection rates. Benchmarked against existing models, the proposed solution excels in precision, accuracy, and recall/detection rates. This cybersecurity model stands as a robust defence for computer-based, data-driven networks, underscoring the vital role of machine learning and AI in fortifying the security of SCADA networks and critical infrastructure.

Keywords: Machine Learning, Ensemble model, SCADA, Cybersecurity, Cyberattacks, MLP, XGBoost, Network Traffic, Wireshark, and GNS3.

1 Introduction

In an era defined by data-driven decision-making, the role of Artificial Intelligence (AI) models in fortifying security cannot be overstated. These models possess the exceptional ability to analyse extensive datasets, preemptively identifying anomalies and potential threats before they escalate. In this landscape, data-driven AI emerges as the linchpin—a formidable ally poised to redefine the security paradigm of Supervisory Control and Data Acquisition (SCADA) networks, which are essential for controlling and monitoring critical infrastructure by relying on data and information technologies (see Fig. 1) [1].

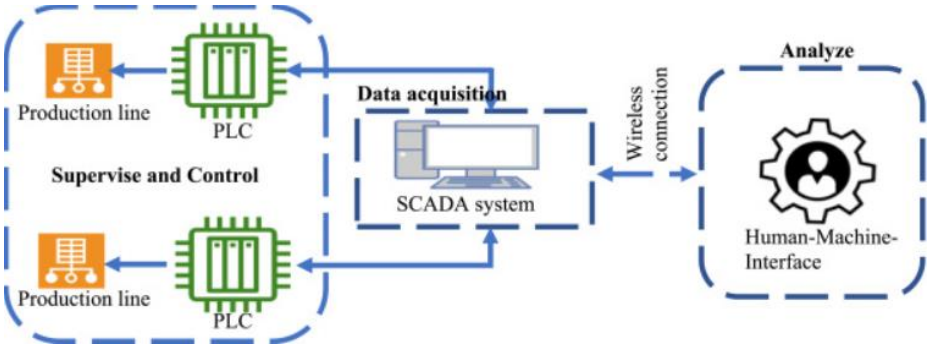

Fig. 1. SCADA Architecture

The imperative to safeguard critical infrastructure against relentless cyber threats resonates not only in South Africa but also on a global scale, where these threats pose significant risks to public health, safety, and economic well-being [2]. Examples of such attacks include the Stuxnet virus attack on Iranian nuclear centrifuges [3], cyberattacks on Ukraine's power grid by Russia through SCADA sub-stations [4], SCADA attacks on Venezuela power grid causing blackouts [5], at-tacks on power, oil, and gas utilities in the United States [6], ransomware attacks on energy companies in Portugal, Pakistan, and Italy [7] - [9], and cyberattacks on the Grand Ethiopian Renaissance Dam [10]. Johannesburg's City Power in South Africa experienced a ransomware incident, which led to disruptions in a significant portion of the utility's applications and networks, with the potential to affect around 250,000 customers [11].

The central challenge lies in the current limitations of data-driven AI cybersecurity models. While these models are proficient in detecting cyber threats within SCADA networks, they often fall short in their capacity to proactively prevent these threats. This critical gap leaves vital infrastructure vulnerable to potentially devastating attacks.

The primary objective of this study is to bridge this gap by harnessing the transformative power of data-driven AI models. The aim is to develop an innovative model that not only excels in detecting cyber threats within data-driven SCADA networks but also possesses the capability to prevent these threats in real-time. This is achieved through the seamless integration of the predictive prowess of the Multi-layer Perceptron (MLP) with the robustness of eXtreme Gradient Boost (XGBoost), thereby redefining the standard of cyber threat defence.

The paper will explore how data-driven AI models and machine learning techniques can be optimally deployed to enhance the security of SCADA networks.

The rest of the paper unfolds as follows: Section 2 conducts a comprehensive literature review. Section 3 outlines the research design, data collection, and the proposed ensemble algorithm. Section 4 provides findings and the results. Section 5 provides analysis and discussion of the results obtained. Section 6 presents concluding observations, while Section 7 outlines recommendations and avenues for further research.

2 Literature Review

This section provides a review of existing literature related to the application of Artificial Intelligence (AI) and machine learning in enhancing the security of Supervisory Control and Data Acquisition (SCADA) networks. SCADA networks are crucial for managing critical infrastructure, yet they face complex challenges in distinguishing normal traffic from potential cyber threats and identifying specific attack types.

2.1 AI in SCADA networks security

Supervisory Control and Data Acquisition (SCADA) networks are indispensable for managing critical infrastructure across sectors like power grids and transportation [12]. These data-centric networks face a perplexing issue—distinguishing normal traffic from suspicious activities is intricate [13]. Additionally, pinpointing specific cyber threats like Distributed Denial of Service (DDOS), Denial of Service (DoS), False Data Injection Attack (FDIA), and Man-in-the-Middle At-tacks (MitMA) remains elusive [14], [15].

Debates persist on the efficacy of existing traffic delineation methods. [16] advocate for improved detection mechanisms, emphasising the need for precision in anomaly detection to strengthen SCADA network security. In contrast, authors such as [17], [18] argued that anomaly detection solutions suffice, with a shift toward proactive prevention. This divergence highlights the complexity and demands innovation.

Data-driven Artificial Intelligence (AI) integration revolutionises SCADA net-work security. These AI models excel at real-time analysis, proactively identifying and preventing anomalies within SCADA networks [19]. Their adaptability and learning capacity make them ideal for safeguarding critical systems [20]. [21] assert their future significance, citing superiority in threat detection and mitigation. This shows that data-driven AI models in SCADA protection revolve around effectiveness and feasibility.

2.2 Previous Studies

Authors in [22] proposed the use of Decision Trees (DT) and k-Nearest Neighbour (KNN) algorithms, enhanced with ensemble classifiers like AdaBoost, Gradient Boost, and XGBoost. Their models achieved outstanding performance metrics of 99% F1-score, recall, accuracy, and precision after trained and tested using ORNL Electric Testbed dataset, outperforming Gaussian Naïve Bayes (GNB) and Random Forest (RF) classifiers. This research emphasises the significance of performance-enhancing techniques and the importance of evaluating multiple algorithms for fair comparisons.

Authors in [23] focused on Intrusion Detection Systems (IDS) for smart grid SCADA and IT components. They proposed Long Short-Term Memory (LSTM) and Feedforward Neural Network (FNN) algorithms on MODBUS datasets, achieving an average accuracy of 99.76%, detection rate of 99.57%, and F1 score of 99.68%. This study demonstrated the effectiveness of these algorithms in identifying various network intrusion cyberattacks, with FNN-LSTM performing well in both correlated and uncorrelated attacks.

Authors in [24] proposed Kernel Principal Component Analysis (KPCA) and Support Vector Data Description (SVDD) as one-class classifiers for intrusion detection in smart grid SCADA systems. Their model achieved an accuracy rate of 93.91% and a detection rate of 93.6% on KDD Cup 99 dataset. While effective, the study suggested further optimisation of free parameters to reduce the computational time.

Authors in [25] proposed cyberattack detection techniques based on temporal pattern recognition using Artificial Neural Networks (ANN) and Hidden Markov Models (HMM). They utilised Cyber-Gym dataset and real SCADA system at Ben-Gurion University (BGU) for their experiment. Proposed algorithms achieved high precision, accuracy, F1-score, and recall of 98.1%, 98.9%, 98.2%, and 98.0%. However, the study highlighted the need for contextual features related to time of day to further enhance the algorithm's performance.

Authors in [26] applied a hybrid deep learning approach combining Gated Recurrent Unit (GRU) and Convolutional Neural Network (CNN) to combat Distributed Denial of Service (DDoS) attacks on smart grid SCADA communication infrastructure. They used a benchmark cybersecurity dataset from the Canadian Institute of Cybersecurity Intrusion Detection System (CICIDSS2017) that closely aligned with real-world scenarios. Their models achieved remarkable performance metrics with an accuracy of 99.7%, precision of 98.1%, recall of 99.9%, and F1-score of 98.9%.

Existing, and proposed machine learning approaches predominantly emphasise the detection and classification of cyberattacks, falling short of providing proactive real-time defence mechanisms. To bridge the gap, the study presents the development of the model using machine learning algorithms that not only identify threats but also function as robust firewalls, thwarting suspicious data packets and safeguarding system integrity.

3 Method

This section outlines the methodology for developing a data-driven AI model for cybersecurity in SCADA networks using the Python programming language. It covers experimental setup, data collection, preprocessing, model development, training, testing, hyperparameter settings, and model evaluation for detection tasks. The section also details machine learning insights extraction from network traffic data, crafting of a dynamic firewall rule to the model for prevention tasks, and model evaluation for prevention tasks.

3.1 Research Tools

To ensure a robust and comprehensive approach, a selection of tools was thoughtfully employed in this study. Graphical Network Simulator-3 (GNS3) served as the open-source platform for emulating the SCADA network. VMWare Workstation Player provided the essential virtual machine environment, enabling the hosting and operation of GNS3 [27]. Kali Linux played a pivotal role in simulating cyberattacks, allowing for the assessment of network vulnerabilities and testing resilience [28]. Wireshark, functioning as a network traffic analyser, captured essential network traffic data for subsequent analysis and model training and testing [29].

The Python programming language, in conjunction with libraries like NumPy, SciPy, SciKit-Learn, Pandas, Matplotlib, and Seaborn, formed the backbone of all machine learning tasks and data visualization [30]. The Subprocess Module executed firewall rules and recorded their outcomes, while Joblib was used for loading and saving machine learning models and pre-processing objects, ensuring efficient model management. The logging module was systematically deployed to record crucial network packet information captured by the machine learning model, and Jupyter Notebook, as described in [30], provided an interactive environment for code writing and execution in various machine learning tasks and data visualisation.

3.2 Experiment Setup

A practical example within a smart grid SCADA configuration was chosen for data collection and analysis. While centred on SCADA networks in general, this approach harnesses the specific characteristics of the smart grid context to vali-date and showcase the AI-driven methodology, ensuring broader applicability to diverse SCADA systems. For a comprehensive guide on establishing the smart grid SCADA network topology, the work of [31] provides valuable insights.

The simulated smart grid SCADA network was designed in GNS3 [32]. The network architecture encompassed the control centre, substation, and field area. Devices within the control centre, such as Control PCs, SCADA servers, and MTUs, were consistently interacting with PLCs and RTUs in the substation. Sensors in the field area were crucial by ever capturing essential grid status data. This data underwent processing within the substation before being presented in the control centre for analysis.

This design underscored SCADA's data-driven essence. Additionally, a network intruder, Kali Linux was connected to the network to perform cyberattacks.

3.3 Data Collection

The data collection phase involved capturing real-time network packets traffic through Wireshark. Captured packet features included IP addresses, MAC addresses, ports, time to live, flags, ICMP code, and ICMP type. IP addresses are numerical labels facilitating device communication on a network; MAC addresses are unique identifiers at the hardware level for network interfaces; ports are logical endpoints directing network traffic to specific applications; Time to Live (TTL) is a value in packet headers determining the maximum network hops; flags are conveying control and status information in packet headers; ICMP code provides specific network condition details within ICMP packets; and ICMP type is a category of the purpose or function of ICMP messages in network communication.

Data was labelled into two classes: "Normal", representing regular network traffic behaviour, and "Suspicious", representing network attacks traffic intended to disrupt normal operations. The normal network traffic involved routine activities within the SCADA network. These activities spanned the data exchange, data presentation, inter-device communication, network interaction with other smart grid domains and the outside world, seamless server access, control, and monitoring. Conversely, the suspicious network traffic encompassed a variety of simulated attacks traffic that is meant to disrupt normal network activities, including Denial of Service (DoS) attacks, Distributed Denial of Service (DDoS) attacks, Man in the Middle (MitMA), and False Data Injection Attack (FDIA). DoS and DDoS attacks were targeted to disrupt network traffic on SCADA devices in the control centre, substation, and field such as servers, PLC, MTU and sensor. MitMA and FDIA were carried to intercept communication network traffic between devices in different sections of the SCADA such as MTU in substation and sensors in the field.

The resulting datasets, encompassing training and testing dataset along with un-seen data, formed the basis for model development. Training and testing dataset, before split, featured 12 input features and a total of 63,845 data points, and previously unseen data featured 27 919 data points. These datasets with "Normal" and "Suspi-cious" labels laid the groundwork for robust model training and evaluation. (See Fig.2) depicts the snippet of the dataset.

	Time	Src_ip	Dst_ip	Src_mac	Dst_mac	Protocol	Src_port	Dst_port	T_to_live	Flags	ICMP_type	ICMP_code	Traffic_Class
0	0.021	10.10.10.4	10.10.10.9	0c:85:46:d8:00:00	c2:71:a5:be:36:82	TCP	42974	21	64	2	0	0	Normal
1	0.000	10.10.10.9	10.10.10.4	c2:71:a5:be:36:82	0c:85:46:d8:00:00	TCP	21	42974	64	18	0	0	Normal
2	0.002	10.10.10.4	10.10.10.9	0c:85:46:d8:00:00	c2:71:a5:be:36:82	TCP	42974	21	64	16	0	0	Normal
3	0.001	10.10.10.9	10.10.10.4	c2:71:a5:be:36:82	0c:85:46:d8:00:00	FTP	21	42974	64	24	0	0	Normal
4	0.010	10.10.10.4	10.10.10.9	0c:85:46:d8:00:00	c2:71:a5:be:36:82	TCP	42974	21	64	16	0	0	Normal

Fig. 2. Sample dataset - Network traffic

3.4 Data Preprocessing

Data preprocessing encompassed the transformation of the real-time network packet traffic dataset to make it suitable for machine learning analysis. Guided by AI principles, this process aimed to address inherent imperfections within raw data, ranging from noisy and inconsistent values to the challenge of missing in-formation [33]. The goal was to shape the dataset into a format conducive for machine learning algorithms to extract insights and identify patterns, forming the basis for strengthening SCADA networks.

1. Data Cleaning

A significant observation unveiled missing values, notably in ICMP protocols where Source Port and Destination Port fields lay vacant. To safeguard the essence of meaningful instances indispensable for training, a pragmatic approach emerged, to replace missing values with zeros. This deliberate choice bolstered the consistency of the dataset, casting aside hindrances that might obstruct the algorithm's understanding. Acknowledging the spectrum of methodologies to tackle missing values, the study opted for zero substitution, recognising its aptitude for nurturing our intended outcomes.

2. Data Transformation

i. Encoding Categorical Variables

Categorical variables- the essence of network attributes - demanded numerical translations. A strategic selection manifested in hash encoding, especially for IP addresses and MAC addresses. This technique translated intricate strings into an intelligible numeric representation, steering towards elevated model performance and accuracy. Meanwhile, the Protocol column witnessed the magic of one-hot encoding. This generated binary components for each protocol, introducing a fresh dimension to algorithm's perception of categorical richness.

ii. Normalisation of Numerical Features

Variability among numerical features required harmonisation through normalisation. This practice brought features to a common scale, circumventing scenarios where any single attribute might disproportionately impact model learning. By fostering consistency, normalisation unlocked the model's capacity to capture patterns from the multifaceted dataset.

iii. Taming Class Imbalance

Class imbalance was an issue on the training set, which means there was a significant disparity between the number of instances in the "Normal" (28 911 instances) and "Suspicious" (15 780) traffic classes. This disparity could potentially lead to biases in the

model's performance, as it might be more inclined to classify instances into the majority class. To harness the full potential of our model and ensure its effectiveness in detecting both "Normal" and "Suspicious" traffic, this conundrum of class imbalance was addressed.

To mitigate the class imbalance, various methods were explored, including Random Oversampling, Synthetic Minority Over-Sampling Technique (SMOTE), and ADASYN (Adaptive Synthetic Sampling). These techniques are designed to balance the representation of the minority class (in this case, "Suspicious" traffic) to ensure that the model receives sufficient training data for both classes.

SMOTE emerged as the most effective solution. SMOTE involves generating synthetic instances of the minority class by interpolating between existing instances. This method not only balanced the class representation but also enhanced the model's performance in detecting and classifying "Suspicious" traffic instances accurately. By addressing the class imbalance issue with SMOTE, it ensured that our model could effectively handle the complexity of network traffic patterns and provide robust results.

3.5 Feature Extraction

The paper did not implement a separate feature extraction step because the Wireshark tool, which we used for capturing network packet data, was already configured to provide comprehensive and essential features of network packets. These features included attributes such as IP addresses, MAC addresses, ports, time to live, flags, ICMP code, and ICMP type. The configuration of Wireshark ensured access to the most relevant attributes for the analysis without the need for additional feature extraction. This approach allowed us to work with the raw packet data, which contained the necessary information for machine learning tasks.

3.6 Model Development

Model Selection

Four distinct models were meticulously chosen to address the task at hand: Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and an ensemble consisting of Multi-Layer Perceptron (MLP) and XGBoost. These selections were made based on their aptitude for handling non-linear variables and intricate data.

XGBoost, as a performance-enhancing algorithm, was selected to create ensemble model because it is excellent in handling complex, non-linear data, and its ability to generalise well on unseen data. In addition, XGBoost is well known for its regularisation techniques, and its efficient computation.

Performance assessment of these models revolved around evaluating their accuracy metrics on both the testing data and previously unseen data. While all models demonstrated commendable performance on the testing data, the ultimate choice was significantly influenced by their efficacy in handling unseen data. The ensemble model, a fusion of MLP and XGBoost, notably outperformed the others on unseen data,

solidifying its selection as the preferred model for fortifying SCADA networks against cyber threats.

Hyperparameter Tuning and Model Robustness

In our pursuit of harnessing the full potential of the machine learning algorithms, meticulous hyperparameter tuning was conducted. The hyperparameters of the XGBoost and MLP classifiers were optimised to strike a balance between model complexity and overfitting while maximising performance.

For the XGBoost classifier, a comprehensive exploration of key hyperparameters was undertaken. The learning rate, a critical factor influencing model convergence, was fine-tuned within a range of [0.01, 0.3]. This range was carefully chosen to ensure the models explored various step sizes during optimisation. The number of estimators, which impacts the tree structure and ensemble size, was optimised over a range of values from 100 to 1000, with increments of 100. This wider range allowed us to encompass different trade-offs between complexity and model performance, making it more adaptable to the nuances of our dataset.

In the case of MLP, the selection of hidden layer sizes [50, 100, 50] was based on the need for flexibility in modelling both simple and complex data relationships, accommodating various network traffic patterns. An iteration limit of 1000 was chosen to ensure convergence to an optimal solution, particularly when dealing with complex patterns. The choice of the ReLu activation function was made to handle non-linearities effectively. A small alpha value of 0.0001 was introduced for regularisation to prevent overfitting, and a constant learning rate was used to ensure stable and controlled training. Using a validation fraction of 0.1 allowed to monitor model performance during training.

The optimisation of these hyperparameters was conducted through an extensive automatic grid search combined with cross-validation. The grid search involved exploring a range of values for each hyperparameter, ensuring that the study considered a wide spectrum of configurations. This deliberate approach steered the models away from overfitting tendencies and paved the way for capturing intricate patterns in the complex network traffic data.

The specific range or values considered in the grid search were tailored to the characteristics of our dataset and the algorithms used, ensuring that we fine-tuned the models for optimal performance. The hyperparameter settings are presented in Table 1.

Table 1. Final hyperparameter setting values for optimal performance.

Parameter	Quantity
N estimators	100
Learning rate (XGBoost) (XGBoost)	0.1
Hidden layers	50, 100, 50

Iteration limit	1000
Activation function	ReLu
Validation fraction	0.1
Learning rate (MLP)	Constant
Alpha	0.0001

The Power of Ensemble: Voting Classifier

The culmination of model development materialised in the form of a Voting Classifier. The Voting Classifier exploited the individual strengths of each model, combining their diverse decision-making strategies to produce a more robust and accurate prediction. The fusion of the ensemble model, using soft voting, allowed the model to consider the weighted average of predicted probabilities from XGBoost and MLP, mitigating biases from either model. While other ensemble techniques such as Bagging and Boosting were considered, the Voting Classifier emerged as the ideal choice, underscoring its adaptability to the intricacies of network traffic patterns.

Model Training, Testing and Evaluation for Detection Tasks

The ensemble model (MLP and XGBoost) underwent rigorous training on a training set (70%), learning intricate patterns present in SCADA network traffic. This model was then tested on a separate testing set (30%) and further tested for its generalisation capabilities on unseen data to emulate real-world scenarios. Its performance was evaluated using essential metrics including accuracy, F1 score, precision, recall, and F2-score [34]. In addition, decision boundary was used to provide more insights regarding the model decision making.

3.7 Model Firewall rules for Prevention Tasks

The integration of model's detection insights with real-time security measures in firewall rules was implemented using Python for model's prevention tasks.

The following is the overview of how the process was implemented:

1. Load Ensemble Model and Auto Preprocessor: Once the model was trained and tested for detection tasks, it was saved and loaded using Joblib library, together with its automatic preprocessor.

2. Automatic Preprocessing of Unseen Data: Unlabelled unseen data was loaded into the model and preprocessor using the Joblib library. This was representing real time incoming traffic.

3. Anomaly Scoring: Anomaly scores quantify packet abnormality, aiding in identifying suspicious behaviour. A range of tests were performed, adjusting the threshold up and down, and evaluated the model's performance in blocking packets. Anomality threshold of 0.5 was set. A threshold of 0.5 performed well in identifying potential threats while maintaining an acceptable level of false positives.

4. Model Decision: High anomaly scores (above 0.5) signal potential threats.

5. Dynamic Firewall Rules: Used source IP, destination IP, source port, destination port, ICMP type, and ICMP code as important features. The model triggers real-time firewall rules via PowerShell and Python. Suspicious traffic is blocked, tailored to packet attributes. Subprocess module was used to execute PowerShell scripts for creating firewall rule.

6. Logging: Logging module was used to log different types of messages during packet processing. A log file is created to store the log entries for later review and analysis.

Model Evaluation for Prevention Tasks

To foster evaluation, transparency and provide actionable insights, a created log file that captures and documents the firewall's actions, cataloguing both allowed and denied packets, was reviewed to see whether the model was able to correctly deny or allow packets based on its prediction.

4 Findings and Results

4.1 Ensemble Model - MLP and XGBoost

In the pursuit of fortifying SCADA networks against cyber threats, a comprehensive exploration of machine learning algorithms was undertaken to unveil intricate patterns and results are displayed in Table 2.

While DT, RF, and SVM displayed competence in managing non-linear relationships and intricate data, their distinctive characteristics posed limitations in this context. DT, proficient in capturing elementary patterns, faltered when confronted with the intricate nuances of network traffic, resulting in comparatively lower performance. Despite the ensemble nature of RF, their performance on unseen data was underwhelming. DT and RF models suggests that these models were overfitting the training data, therefore, manifesting in a significant gap between their testing and unseen data performance. The formidable generalisation ability of Support Vector Machines was overshadowed by their inability to unveil the concealed subtleties of network traffic.

In contrast, the proposed ensemble model of Multi-Layer Perceptron (MLP) and XGBoost, harnessing MLP's potential to unravel intricate patterns and XGBoost's

gradient boosting prowess. This amalgamation resulted in a robust performance on testing data and a commendable performance on previously unseen data.

Table 2. Model selection - performance results

Model	Accuracy on Testing set (%)	Accuracy on Unseen data (%)
DT	99.41	37.81
RF	99.58	38.72
SVM	97.77	89.67
MLP and XGBoost	100	99.19

4.2 Model's Decision Making and Decision Boundary

Decision boundaries (see Fig.3) provides valuable insights into the behaviour of the ensemble model and its ability to distinguish between "Normal" and "Suspi-cious" network traffic instances and provide a deeper understanding of the model's decision-making process.

In the case of the testing set, the observed round-shaped data points represent regions where the model can effectively classify instances as "Normal" traffic. The crescent-shaped red points overlapping the boundary of the blue points indicate instances that are more challenging to classify, possibly situated at the inter-face between classes. This suggests the model's ability to differentiate subtle patterns in network traffic data, but also highlights instances that require more intricate decision-making.

For unseen data, the persistence of similar behaviour underscores the model's consistent performance in different scenarios. The red points on a different side of the circle indicate the model's adaptability to variations in the data distribution while maintaining its decision-making prowess.

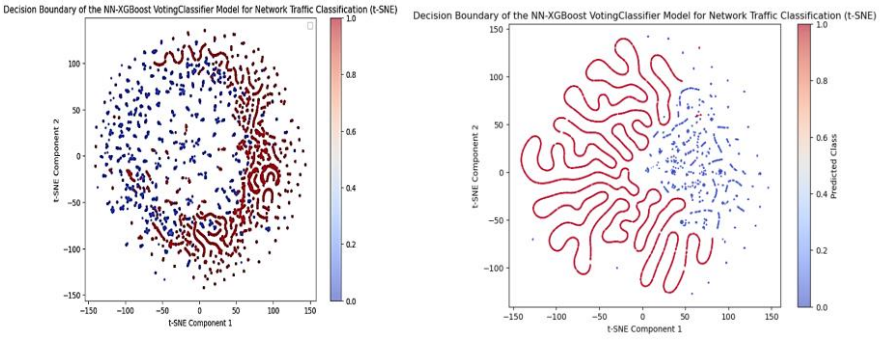

Fig. 3. Decision boundaries on testing set (left) and unseen data (right)

4.3 Feature Importance and Firewall Rules

The model's role as an AI-driven firewall is exemplified by its discerning identification of important intricate network features, as revealed in the comprehensive log file (see Fig.4). By meticulously recording denied and allowed packets, encompassing important features identified by the model like source and destination IPs, ports, ICMP types, and codes, the firewall showcases its aptitude for precise classification. Notably, the model excels in distinguishing and blocking packets exhibiting suspicious or malicious behaviours, while seamlessly permit-ting legitimate communication. This performance underscores the model's efficacy in fortifying the SCADA network against a diverse range of cyber threats.

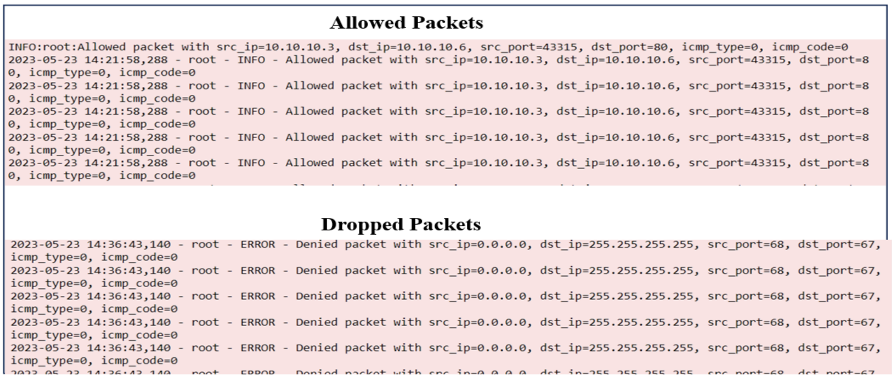

Fig. 4. Log file analysis – Model's prevention tasks

4.4 Performance Evaluation

Table 3 summarises the model's key performance metrics. Averaging performance metrics was a good practice to offer a well-rounded evaluation of a model's performance. It provided a more informative and robust assessment of the model's capabilities.

Table 3. Model's performance results

Metrics (%)	Testing set (%)	Unseen data (%)	Average Score (%)
Accuracy	100.00	99.19	99.60
Precision	100.00	99.34	99.67
Detection/Recall	100.00	98.95	99.48
F1 Score	100.00	100.00	100.00
F2 score	100.00	98.69	99.35

4.5 Comparison against existing ensemble models

To further evaluate the performance of the proposed ensemble model for SCADA network traffic analysis, results model's results were compared with those of existing ensemble models. Table 4 presents a summary of the comparison based on the accuracy and detection rates achieved by each model.

Table 4. Model performance against existing models

Algorithms Approach	Accuracy (%)	Detection (%)
Kernel PCA and SVDD algorithms in [24]	93.9	93.6
Ensemble LSTM and FNN algorithms in [23]	99.8	99.6
Combining supervised and unsupervised learning		
Hybrid ANN and HMM in [25]	98.6	98.0
Hybrid GRU and CNN algorithm in [26]	99.7	99.9
Performance-enhanced algorithms in [27]	99.0	99.0
Proposed Ensemble Model (MLPNN and XGBoost)	99.60	99.48

It is evident from the comparison that the proposed ensemble model achieves competitive accuracy and detection rates, outperforming some existing models in both aspects. Additionally, the model showcases robustness by achieving high detection rates without relying solely on a high accuracy rate.

5 Analysis and Discussion

The ensemble model, which combines the strengths of Multi-Layer Perceptron (MLP) and XGBoost, offered an effective solution for dealing with the intricate challenges of SCADA network traffic analysis. MLP, with its deep learning capabilities, exceled at uncovering intricate patterns within complex network traffic data. Its multiple hidden layers and nonlinear activation functions made it adept at capturing both low-level and high-level features, enhancing its ability to detect anomalies that may be overlooked by other methods. On the other hand, XGBoost, a gradient boosting algorithm, demonstrated exceptional performance in classification tasks. It iteratively trained a sequence of decision trees, each correcting the errors of the previous one, leading to improved predictive power. This boosting technique reduced bias and enhanced accuracy.

The Voting Classifier combined the decision-making strategies of both MLP and XGBoost, resulted in a more robust prediction process. This amalgamation mitigated biases and enhanced the model's resilience in handling the complexity of network traffic patterns. This is evident as the model achieved 99.19% accuracy on unseen data, along with near-perfect precision and F1 scores, and an effective F2 score for anomaly

detection, demonstrates its capability to identify both malicious and legitimate network traffic accurately.

When integrated as a firewall, the model marks a significant advancement in SCADA network security. Its dynamic analysis and packet filtering based on predictions hold immense potential for safeguarding critical infrastructures. The firewall's accurate distinction between normal and anomalous behaviours is particularly advantageous in the intricate landscape of SCADA networks.

6 Conclusion

The study demonstrates the effectiveness of data-driven AI in detecting and preventing cyberattacks in a SCADA network. The proposed model, combining Multi-layer Perceptron (MLP) and XGBoost, achieved high average accuracy and detection rates of 99.60% and 99.48% respectively, outperforming individual models. It effectively addressed the challenge of imbalanced classification through pre-processing and oversampling. Ensemble machine learning algorithms mitigated biases and enhanced the model's resilience to the complexity of network traffic patterns, and adaptability to variations in the data distribution while maintaining its decision-making prowess. The model's integration with data-driven firewall rules showcased its capability to identify and block malicious packets, enhancing SCADA network security.

7 Recommendations and Future work

In the realm of recommendations and future work, there lies a critical pathway to elevate the prowess of AI in the domain of SCADA networks. Primarily, this journey demands the augmentation of data collection by encompassing pivotal protocols like MODBUS and SMTP. An expansive dataset, encompassing an array of features and attack scenarios, shall further enhance the model's depth. Beyond the conventional, an evolved firewall strategy, transcending the confines of anomaly scores, emerges as a strategic imperative. In partnership with industry stakeholders, real-world validation stands as a cornerstone, sharpening practicality. This trajectory demands real-time vigilance to proactively detect threats, adaptive learning techniques for ongoing evolution, and the unearthing of profound insights through advanced machine learning paradigms such as deep learning. Yet, the heart of progress lies in addressing research gaps through meticulous scenario simulations and diversified data exploration, culminating in a fortified front of AI-driven cybersecurity within SCADA networks.

References

1. Borenius, S., Gopalakrishnan, P., Bertling, L., and Kantola, R.: Expert-Guided Security Risk Assessment of Evolving Power Grids. Energies, 15(9), p. 3237 (2022).

2. Allen, K.: Critical infrastructure attacks: Why South Africa should worry, ISS Africa. [Online]. Available at: https://issafrica.org/iss-today/critical-infrastructure-attacks-why-south-africa-should-worry, [Assessed: 28 March 2023]. (2021).
3. Langner, R.: Stuxnet: Dissecting a cyberwarfare weapon. IEEE Security and Privacy, 9(3), pp. 49-51 (2011).
4. Robert M., and Michael, J.: Analysis of the Cyber-Attack on the Ukrainian Power Grid. Electricity Information Sharing and Analysis Center (E-ISAC), Tech. Rep. (2016).
5. BBC: Venezuela blackout: Power cuts plunge country into darkness. [Online]. Available at: https://www.bbc.com/news/world-latin-america-49079175, [Assessed: 05 April 2023]. (2019).
6. Greenberg, A. 2020: Iranian hackers have been 'password-spraying' the US grid [Online] Available at: https://www.wired.com/story/iran-apt33-us-electric-grid/, [Assessed: 24 April 2023].
7. Paganini, P.2020.: Netwalker ransomware operators leaked files stolen from K-Electric. [Online]. Available at: https://securityaffairs.co/wordpress/109000/hacking/k-electric-net-walker-data-leak.html, [Assessed: 15 April 2023]
8. Ilascu, I.: Power company Enel Group suffers Snake ransomware attack. [Online]. Available at: https://www.bleepingcomputer.com/news/security/power-company-enel-group-suffers-snake-ransomware-attack/ [Accessed: 5 April 2023] (2020).
9. Osborne, C.: Energy company EDP confirms cyberattack, Ragnar Locker ransomware blamed. [Online]. Available at: https://www.zdnet.com/article/edp-energy-confirms-cyberattack-ragnar-locker-ransomware-blamed, (2020).
10. Markos, Y.: Cyber Security Challenges that Affect Ethiopia's National Security. Available at SSRN 4190146 (2022).
11. Panettieri, J.2019.: Ransomware attack rocks city power, Johannesburg, South Africa, MSSP Alert. Available at: https://www.msspalert.com/cybersecurity-breaches-and-attacks/ransomware/city-power-johannesburg-south-africa, [Assessed: 24 April 2023].
12. Kim, Y., Hakak, S., and Ghorbani, A.: Smart Grid security: Attacks and defence techniques. IET Smart Grid (2022)
13. Faheem, M., Shah, S.B.H., Butt, R.A., Raza, B., Anwar, M., Ashraf, M.W., Ngadi, M.A., and Gungor, V.C.: Smart Grid communication and information technologies in the perspective of Industry 4.0: Opportunities and challenges. Computer Science Review, 30, pp. 1-30 (2018).
14. Essaaidi, M.: An overview of Smart Grid cyber-security state of the art study. In 2015 3rd International Renewable and Sustainable Energy Conference (IRSEC), pp. 1-7 (2015).
15. Singh, N.K., Mahajan, V., Aniket, A., Pandya, S., Panchal, R., Mudgal, U., and Bhatt, M.: Identification and prevention of cyber-attack in Smart Grid communication network. In 2019 International conference on information and communications technology (ICOIACT), pp. 5-10 (2019).
16. Ashrafuzzaman, M., Das, S., Chakhchoukh, Y., Shiva, S., and Sheldon, F.T.: Detecting stealthy false data injection attacks in the Smart Grid using ensemble-based machine learning. Computers and Security, 97, p. 101994 (2020).
17. Diovu, R.C., and Agee, J.T.: A cloud-based openflow firewall for mitigation against DDoS attacks in Smart Grid AMI networks. In 2017 IEEE PES PowerAfrica, pp. 28-33 (2017).

18. Eltayieb, N., Elhabob, R., Hassan, A., and Li, F.: An efficient attribute-based online/offline searchable encryption and its application in cloud-based reliable Smart Grid. Journal of Systems Architecture, 98, pp. 165-172 (2019).
19. Tufail, S., Batool, S., and Sarwat, A.I.: False data injection impact analysis in ai-based smart grid. In SoutheastCon 2021, pp. 01-07 (2021). IEEE.
20. Saha, S.S., Gorog, C., Moser, A., Scaglione, A., and Johnson, N.G.: Integrating hardware security into a blockchain-based transactive energy platform. In 2020 52nd North American Power Symposium (NAPS), pp. 1-6 (2021).
21. Zhang, H., Wang, J., and Ding, Y.: Blockchain-based decentralized and secure keyless signature scheme for Smart Grid. Energy, 180, pp. 955-967 (2019).
22. Efiong, J.E., Akinyemi, B.O., Olajubu, E.A., Aderounmu, G.A., and Degila, J.: Cyber SCADA Network Security Analysis Model for Intrusion Detection Systems in the Smart Grid. In Advances in Intelligent Systems, Computer Science and Digital Economics IV, pp. 481-499 (2023).
23. Gao, J., Gan, L., Buschendorf, F., Zhang, L., Liu, H., P. Li, Dong, X., and Lu, T.: Omni SCADA intrusion detection using deep learning algorithms (2019).
24. Nader, P., Honeine, P., and Beauseroy, P.: Intrusion detection in SCADA systems using one-class classification. In Proc. 21st Eur. Signal Process. Conf. (EUSIPCO), pp. 1-5 (2013).
25. Kalech, M.: Cyberattack detection in SCADA systems using temporal pattern recognition techniques. Comput. Secur., pp. 225-238 (2019).
26. Diaba, S.Y., and Elmusrati, M.: Proposed algorithm for Smart Grid DDoS detection based on deep learning. Neural Networks, 159, pp. 175-184 (2023).
27. VMWare. [Online] Available at: https://www.vmware.com/, [Assessed: 20 April 2023].
28. Cesar, P., and Pinter, R.: Some ethical hacking possibilities in Kali Linux environment. Journal of Applied Technical and Educational Sciences, 9(4), pp. 129-149 (2019).
29. Wireshark. [Online] Available at: https://www.wireshark.org/, [Assessed: 25 April 2023].
30. Raschka, S.: Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow (2018).
31. Stouffer, K., Falco, J. and Scarfone, K.: Guide to industrial control systems (ICS) security. NIST special publication, 800(82), pp.16-16 (2011).
32. Banda, T.V.: Towards a Supervised Machine Learning Algorithm for Cyberattacks Detection and Prevention in a Smart Grid Cybersecurity System, Stellenbosch University (2023).
33. Mishra, P., Biancolillo, A., Roger, J.M., Marini, F., and Rutledge, D.N.: New data preprocessing trends based on ensemble of multiple preprocessing techniques. TrAC Trends in Analytical Chemistry, 132, p. 116045 (2020).
34. Vujović, Ž.: Classification model evaluation metrics. International Journal of Advanced Computer Science and Applications, 12(6), pp. 599-606 (2021).