Eco-Secure SCADA: Towards Machine Learning Reliability for Green Cybersecurity

Love Allen Chijioke Ahakonye [1], Kalibbala Jonathan Mukisa [2],
Cosmas Ifeanyi Nwakanma [3], Dong-Seong Kim [2] *
[1] ICT Convergence Research Center, *Kumoh National Institute of Technology*, Gumi, South Korea
[2] IT-Convergence Engineering, *Kumoh National Institute of Technology*, Gumi, South Korea
* NSLab Co. Ltd., Gumi, South Korea, *Kumoh National Institute of Technology*, Gumi, South Korea
[3] Computer Science and Electrical Engineering, *West Virginia University*, Morgantown, 26506, MV, USA
(loveahakonye, kjonmukisa, dskim)@kumoh.ac.kr, profcosmas@gmail.com

***Abstract*—This study explores energy-efficient machine learning approaches for intrusion detection in SCADA systems, addressing the dual challenges of cybersecurity and sustainability. A comprehensive evaluation of models, including Decision Trees, Random Forests, and LightGBM, highlights their performance across SCADA, IIoT, and Edge IoT environments. Decision Trees achieve exceptional efficiency scores, with 83.85 in Edge IoT systems, demonstrating high accuracy with minimal energy consumption. Random Forests and LightGBM balance scalability, computational cost, and resilience, supporting robust deployment in resource-constrained environments. Integrating lightweight and high-performing models provides a roadmap for achieving eco-secure SCADA systems, advancing the synergy between green cybersecurity and machine learning reliability.**

***Index Terms*—Green cybersecurity, Energy, Efficiency, Reliability, SCADA, Sustainability,**

I. Introduction

Supervisory control and data acquisition (SCADA) systems are vital to critical infrastructure such as power grids, water treatment, and manufacturing [1]. Integrating operations with information technology has heightened their exposure to cyber threats. While Artificial intelligence (AI)-driven techniques significantly enhance SCADA cybersecurity by precisely detecting conventional and SCADA-specific threats [2], they often increase energy consumption [3]. Employing machine learning (ML) and deep learning (DL) enables adaptive threat detection [4], but optimizing energy usage is crucial to aligning with green cybersecurity principles, ensuring both security and sustainability [5], [6].

Green cybersecurity emphasizes secure, energy-efficient practices to minimize the carbon footprint of cybersecurity operations without compromising protection [7]. It focuses on sustainable hardware, energy-efficient data centers, optimized software, and resource-efficient intrusion detection systems (IDS). Techniques include efficient coding, automated scaling, optimized storage, AI-driven security, and cloud-native solutions [8]. Integrating AI with green cybersecurity reduces energy use by optimizing computational resources for threat detection and response while enhancing critical infrastructure resilience [9]. This ensures SCADA-managed systems, like energy grids and water networks, remain secure, efficient, and reliable.

The computational demand and energy consumption of critical industrial operations, essential to modern infrastructures, is rising to meet energy needs, contributing 2.1-3.9% of global emissions [7], with energy demands projected to rise. Reliance on fossil fuels exacerbates climate issues, affecting public health, biodiversity, and resources, highlighting the need for "green" cybersecurity in IDS, which protects critical infrastructure [10], [11]. Optimizing IDS energy consumption can reduce environmental impact and operational costs. Initiatives like improving cloud data center efficiency show the potential of green strategies to align sustainability with robust security [7]. Energy-efficient algorithms and resource optimization are vital to minimizing emissions while maintaining cybersecurity's critical role.

Traditional cybersecurity overlooks energy efficiency, posing scalability challenges for expanding systems like SCADA networks due to high energy demands. Green cybersecurity addresses these limitations by integrating AI to optimize resource usage, reduce operational costs, and align with Environmental, Social, and Governance principles [11]. By emphasizing scalable, energy-efficient solutions, it delivers robust protection against advanced threats while supporting global sustainability goals and fostering environmental security practices. A recent study highlighted inefficient cybersecurity scenarios and the necessity of energy-aware cybersecurity solutions [7]. The authors reviewed techniques for measuring and optimizing cybersecurity solutions' energy consumption and presented a Green Security taxonomy. This study leverages the foundation in [7] to investigate IDS algorithms for reliable, secure, eco-friendly intrusion detection. This study pioneers the research examining intrusion detection algorithms for sustainability and green industrial operations.

Specifically, this study focuses on the following:

1) Investigate ML and DL algorithms for energy-efficient intrusion detection, analyzing trade-offs between training and inference times across SCADA, IIoT, and Edge IoT scenarios.
2) A roadmap for Green IDS implementations, emphasizing lightweight models that achieve competitive accuracy with minimal computational overhead, aligning cybersecurity with sustainability.

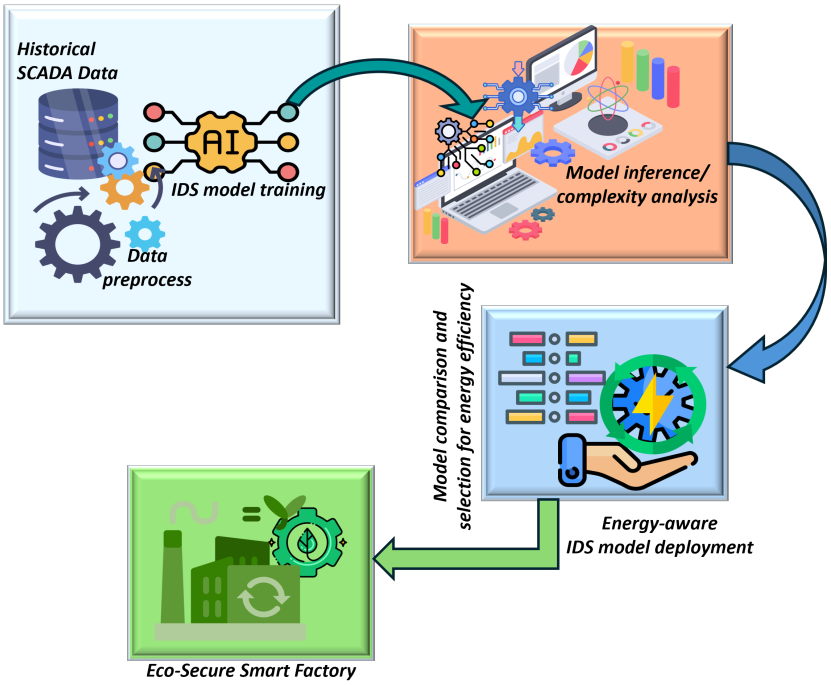

Fig. 1. The process flow of the concept of selecting an energy-efficient algorithm for the intrusion detection systems

3) A comparative framework showcasing resilient models that balance accuracy, energy efficiency, and computational cost, enabling scalable deployment in resource-constrained, green cybersecurity solutions.

The study is structured thus: introduction in Section I is followed by Section II reviewing the existing works on assessing the energy impact of security measures. Section III discusses evaluation methodology. Section IV focused on the experimentation and results. Section V concludes the study.

II. Related Works

An assessment of the energy impact of security measures has been suggested [7], and this section briefly discusses the methods utilized in contemporary IDSs to optimize energy consumption. Prior approaches to energy optimization in IDSs have focused on reducing computational overheads, communication overheads, or in-network job division [9]. They provided a simple model for evaluating the energy cost of distributed packet inspection in intrusion detection systems (IDSs). They used it to investigate energy leakage brought on by the delayed detection of malicious packets.

To balance energy consumption, false positive rates, and detection rates, a study proposed applying game theory to activate anomaly detection techniques only in anticipation of a new attack's signature [12]. In scaling mode, the lightweight anomaly detection method requires less energy to detect assaults with high detection and low false-positive rates, which is why simulation results show that it performs better than existing anomaly detection strategies.

A recent study highlighted the necessity of energy-aware cybersecurity solutions by highlighting inefficient cybersecurity scenarios [7]. The authors examined methods for calculating and optimizing the energy consumption of cybersecurity solutions based on the supposition that cybersecurity derivatives account for around 20% of global ICT emissions, translating into yearly carbon emissions of 142.5 million metric tons, or roughly 0.4% carbon emissions globally [7]. They presented a green security taxonomy and possible directions for improvement through a systemic review of current research initiatives and technical developments, laying the foundation of the methodological approach in this paper toward an efficient and reliable ML-based IDS for green cybersecurity in SCADA networks.

III. Methodology

Figure 1 is the process flow of the proposed methodology. It investigates intrusion detection algorithms, emphasizing computational energy efficiency and resource optimization, aligning with the principles of green cybersecurity. To ensure consistent and fair evaluation, benchmark intrusion detection datasets such as Edge-IIoTset [13][1], WUSTL-IIoT-2021 dataset [14][2] and ICS-SCADA [15][3] were evaluated.

These datasets are preprocessed by scaling features to a common range for unbiased training. Reconstructing the datasets from the reduced space in Equation 1 improved the computational efficiency.

$$\mathbf{X}_{\text{reconstructed}} = \mathbf{X}_{\text{reduced}}\mathbf{P}^{\top}, \tag{1}$$

[1]https://ieee-dataport.org/documents/edge-iiotset-new-comprehensive-realistic-cyber-security-dataset-iot-and-iiot-applications

[2]https://ieee-dataport.org/documents/wustl-iiot-2021

[3]https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets

where $X_{\text{reconstructed}}$ is the data reconstructed back into the original space. X_{reduced} is data in the reduced-dimensional space. $\mathbf{X}$ is the matrix of the top k eigenvectors used for projection. $\mathbf{P}^\top$ represents the transpose of the projection matrix mapping back into the original space. Dividing the dataset into training (80%) and testing (20%) subsets validates model performance.

We evaluate representative ML and DL models, including random forest (RF), pruned decision tree (DT) [1], light gradient boost (LightGBM), K-nearest neighbours (KNN), 3-Layered neural network (NN), a hybrid of convolutional neural network and long-short term memory (CNN-LSTM) [16]. Each model is configured with standard hyperparameters, and parameter tuning is performed where necessary to ensure optimal performance.

The computational demands of each IDS model were analyzed based on the total model training time (T_{total}), inference time (I_{total}), which the time taken to classify the test dataset and the total number of trainable parameters known as parameter count (P). Using the performance data collected, the following metrics are calculated as follows:

1) Normalized time per data unit: evaluates the computational time for training and inference normalized by the dataset size N. It allows fair comparisons between models trained on varying data subsets to improve computational efficiency. Training time per data unit is calculated as in Equation 2: The metric draws light on the computational efficiency of the training process. A lower T_{unit} indicates that the model can learn effectively without excessive computational expense. This is critical for SCADA systems managing massive data streams, such as sensor readings in power grids or water networks.

$$T_{\text{unit}} = \frac{T_{\text{total}}}{N}, \tag{2}$$

where T_{total} is the total training time. Inference time per data unit is as in Equation 3.

$$I_{\text{unit}} = \frac{I_{\text{total}}}{N}, \tag{3}$$

where I_{total} is the total inference time. The lower the I_{unit}, the faster the system can respond to anomalies, ensuring operational reliability and minimizing potential downtime of the SCADA system.

2) Efficiency score: normalizes computational cost with a logarithmic scale for fair comparison and incorporates a weighting factor α for customizable accuracy-cost trade-offs as in Equation 4. It reflects the non-linear relationship between computational cost and accuracy, enhancing real-world applicability.

$$E_{\text{score}} = \frac{A}{\alpha \cdot \log(1 + T_{\text{unit}} + I_{\text{unit}})}, \tag{4}$$

$\log(1+x)$ normalizes computational cost, mitigating high-cost impact and capturing diminishing accuracy returns. Adding 1 ensures positivity, while the tunable parameter $\alpha > 0$ balances accuracy and cost with smaller α favoring efficiency in resource-limited SCADA systems. E_{score} ranked the IDS models to balance detection accuracy and computational efficiency. Scalability and resource efficiency, measured via computation time, prioritized low-complexity models for resource-constrained deployments. Algorithm 1 summarizes the investigation of the computational demands of the evaluated intrusion detection algorithms. Experimentation on fixed computational resources ensured reproducibility, using Visual Studio Code, Solidity v0.8.22, and Python 3.6.13 on a system with an Intel i5-8500 CPU, 8GB RAM, and Windows 11.

Algorithm 1: Computational demand evaluation for IDS models

Require: Total training time T_{total}, Total inference time I_{total}, Dataset size N, Model accuracy A

Ensure: Computed metrics T_{unit}, I_{unit}, and E_{score}

Calculate training time per data unit:

```
T_unit = T_total / N
```

Calculate inference time per data unit:

```
I_unit = I_total / N
```

Calculate efficiency score:

```
E_score = A / (T_unit + I_unit)
```

Output `T_unit`, `I_unit`, `E_score`

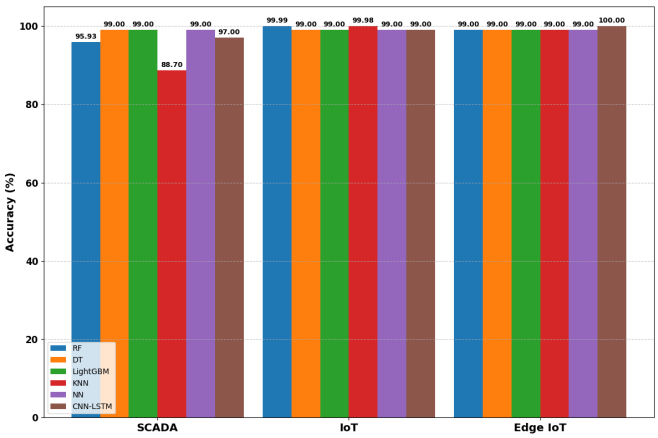

Fig. 2. Plot showing the accuracy performance of the evaluated algorithms

IV. Experimentation and Result Discussion

This study investigates machine learning models like RF, DT, LightGBM, KNN, NN, and CNN-LSTM across IoT environments (SCADA, IIoT, and Edge IoT). RF, NN, and CNN-LSTM consistently achieve near-perfect accuracy across environments, making them reliable for cybersecurity detection tasks. The accuracy performance in Figure 2 shows that DT maintains high accuracy (90% on SCADA, 99% on IIoT), albeit slightly lower than RF and NN. Deep learning models like CNN-LSTM sacrifice efficiency for high accuracy, while DT and LightGBM balance both accuracy and efficiency, making them ideal candidates for green cybersecurity solutions.

Table I compares the training and inference times of the investigated models across evaluated datasets. With the lowest training and inference times, DT demonstrates significant efficiency and is ideal for green cybersecurity in real-time and resource-constrained scenarios. RF and LightGBM show moderate training and fast inference times, balancing efficiency and performance. CNN-LSTM and KNN are computationally intensive, with the highest training and inference times, making them suitable for non-time-critical tasks. However, times increase with data complexity ($SCADA < IIoT < EdgeIIoT$), with DT and RF retaining scalability.

TABLE I
TRAINING AND INFERENCE PERFORMANCE OF THE INVESTIGATED MODELS ACROSS SCADA, IIOT, AND EDGEIIOT DATA SCENARIOS

Dataset	Model	Training Time (s/MB)	Inference Time (s/MB)
SCADA	CNN-LSTM	90.2600	1.172222
	KNN	0.8007	0.74390
	NN	7.9800	0.87930
	LightGBM	0.1510	0.00348
	RF	0.4800	0.00223
	DT	**0.0500**	**0.00085**
IIoT	CNN-LSTM	90.3680	1.15210
	KNN	1.1700	1.73924
	NN	9.4800	1.21510
	LightGBM	0.2300	0.01654
	RF	0.5600	0.00227
	DT	**0.0500**	**0.00089**
Edge IoT	CNN-LSTM	90.3960	1.17020
	KNN	1.2700	1.04103
	NN	9.9600	0.86824
	LightGBM	0.1500	0.01289
	RF	0.2100	0.00223
	DT	**0.0800**	**0.00085**

Likewise, the average training and inference performance in Figure 3 shows DT as the highest efficient with the lowest average training and inference time. This is followed by LightGBM and RF, which achieve low training and inference times, albeit slightly less efficient than DT. Efficient models like DT and RF provide a practical path for achieving energy-efficient, scalable, and sustainable green cybersecurity systems, balancing performance and environmental impact. Conversely, computationally complex models like CNN-LSTM should be selectively used, prioritizing scenarios where their complexity adds significant value.

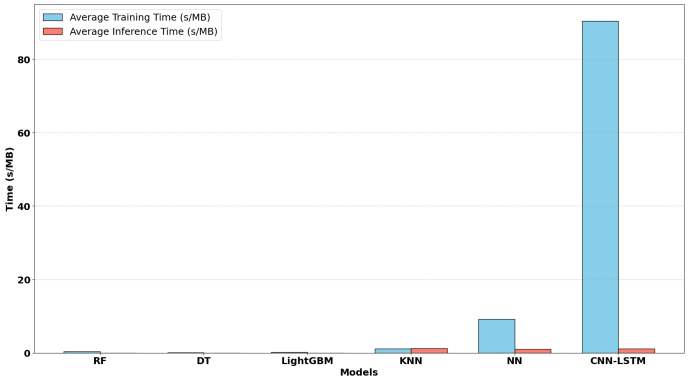

Fig. 3. Plot showing the average training and inference performance of the evaluated algorithms

A. *Efficiency Insights*

Efficiency scores highlight the model's ability to combine accuracy with computational efficiency. Efficiency scores in Figure 4 demonstrate that DT leads with 13.03 (SCADA), 19.05 (IIoT), and 83.85 (Edge IoT), indicating its robustness in delivering high accuracy at minimal energy cost. CNN-LSTM and NN exhibit poor efficiency scores due to high computational costs despite good accuracy. Although KNN demonstrates excellent scores in SCADA (25.08), there are declines in IIoT and Edge IoT due to inference inefficiencies.

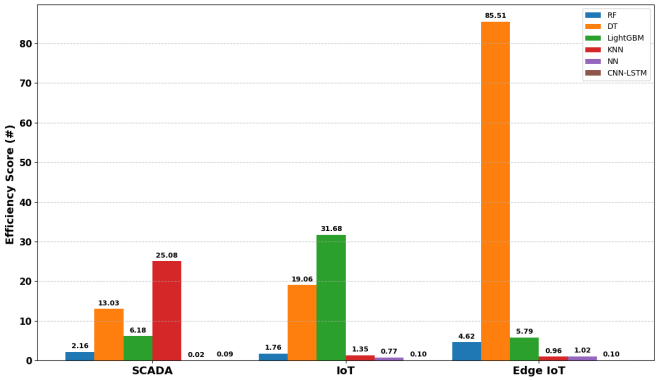

Fig. 4. Plot showing the efficiency scores of the evaluated algorithms

Energy-efficiency insights highlight that DT and LightGBM exhibit minimal normalized time per data unit for training and inference, making them ideal for low-energy SCADA environments. CNN-LSTM are computationally intensive, consuming significantly more time per data unit, indicating higher energy costs. Similarly, LightGBM and DT scale efficiently across IoT and Edge IoT, ensuring consistent performance without significant energy overheads. In contrast, NN and CNN-LSTM struggle to maintain efficiency, leading to increased energy consumption in larger datasets.

B. *Complexity Trade-off*

DT and LightGBM outperform complex models NN and CNN-LSTM, providing the best trade-off between accuracy, energy efficiency, and scalability in energy-constrained environments. NN's high parameter count and CNN-LSTM's deep architecture result in excessive training and inference times, reducing their practicality for eco-friendly SCADA systems. RF can be considered for environments where slightly higher computational costs are acceptable.

Table II shows the sizes and trainable parameters of the evaluated models. It highlights the importance of selecting models based on the trade-offs between computational complexity, energy efficiency, and accuracy. Lightweight models offer compelling solutions for eco-secure SCADA systems, while more complex neural networks may be reserved for scenarios where higher accuracy justifies the energy cost. However, models with a combined advantage of less complexity, accuracy, and significance regarding resource efficiency are most suitable for green cybersecurity.

TABLE II
TRAINABLE PARAMETERS AND SIZES OF THE EVALUATED MODELS

Model	Size / Trainable Parameters
3-layered Neural network	#10,570
CNN-LSTM	#39,529
Random Forest	4.7 MB
Pruned Decision Tree	0.05 MB
LightGBM	1.00 MB
K-nearest neighbours	1.00 MB

C. Green Cybersecurity Implications

The results highlight the effective integration of green technology into cybersecurity for SCADA and IoT systems, emphasizing energy efficiency, operational reliability, and real-time performance. Green IDSs aim to minimize energy consumption while maintaining robust protection [7], [9], addressing energy-performance-security trade-offs for resource-constrained embedded devices. Training and inference times correlate directly with energy consumption, making lightweight models like LightGBM and DT ideal for energy-constrained environments like Edge IoT. With minimal training and inference times, these models ensure fast, low-power operation critical for real-time cybersecurity applications. In contrast, resource-intensive models like CNN-LSTM, offering enhanced threat detection, are unsustainable for large-scale or distributed deployments like IIoT and Edge IoT due to their high energy demands. In systems where resources are less constrained, CNN-LSTM may be justified if its security benefits outweigh energy costs. However, LightGBM provides a balanced alternative with lower energy use and fast inference. For IIoT and Edge IoT, DT and LightGBM stand out as optimal choices, aligning with green cybersecurity principles by minimizing energy consumption and reducing the carbon footprint. Adopting such models enables the achievement of sustainability goals without compromising security. Prioritizing efficient, lightweight models is essential for balancing cybersecurity performance with environmental sustainability.

V. CONCLUSION

This study demonstrates that lightweight ML models like DT and LightGBM strike the optimal balance between accuracy, energy efficiency, and scalability, making them ideal candidates for eco-friendly SCADA cybersecurity solutions. These models consistently deliver high performance with minimal computational cost, as evidenced by their low training and inference times and superior efficiency scores across all datasets. In contrast, while achieving high accuracy, complex models such as CNN-LSTM and neural networks exhibit excessive energy demands and poor efficiency, rendering them impractical for resource-constrained environments. It emphasizes the need to prioritize efficient algorithms in green cybersecurity frameworks. Future research should explore hybrid approaches combining simplicity and efficiency with the robustness of deep learning alongside energy-aware hardware optimizations to further enhance machine learning reliability in sustainable SCADA networks.

ACKNOWLEDGMENT

This research was supported by the Priority Research Centers Program through the NRF funded by the MEST (2018R1A6A1A03024003) (50%) and by MSIT under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2025-RS-2020-II201612) (50%) supervised by the IITP.

REFERENCES

[1] L. A. C. Ahakonye, C. I. Nwakanma, J. M. Lee, and D.-S. Kim, "Agnostic CH-DT Technique for SCADA Network High-Dimensional Data-Aware Intrusion Detection System," *IEEE Internet of Things Journal*, vol. 10, no. 12, pp. 10 344–10 356, 2023.

[2] L. A. C. Ahakonye, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "SCADA Intrusion Detection Scheme Exploiting the Fusion of Modified Decision Tree and Chi-Square Feature Selection," *Internet of Things*, vol. 21, p. 100676, 2023.

[3] A. Wali and F. Alshehry, "A Survey of Security Challenges in Cloud-Based SCADA Systems," *Computers*, vol. 13, no. 4, p. 97, 2024.

[4] L. A. C. Ahakonye, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Trees Bootstrap Aggregation for Detection and Characterization of IoT-SCADA Network Traffic," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 4, pp. 5217–5228, 2024.

[5] A. Alzahrani and T. H. Aldhyani, "Design of Efficient Based Artificial Intelligence Approaches for Sustainable of Cyber Security in Smart Industrial Control System," *Sustainability*, vol. 15, no. 10, p. 8076, 2023.

[6] A. M. Koay, R. K. L. Ko, H. Hettema, and K. Radke, "Machine Learning in Industrial Control System (ICS) Security: Current Landscape, Opportunities and Challenges," *Journal of Intelligent Information Systems*, vol. 60, no. 2, pp. 377–405, 2023.

[7] S. Brudni, S. Anidgar, O. Brodt, D. Mimran, A. Shabtai, and Y. Elovici, "Green Security: A Framework for Measurement and Optimization of Energy Consumption of Cybersecurity Solutions," in *2024 IEEE 9th European Symposium on Security and Privacy (EuroS&P)*, 2024, pp. 676–696.

[8] E. Wai and C. Lee, "Seamless Industry 4.0 Integration: A Multilayered Cyber-Security Framework for Resilient SCADA Deployments in CPPS," *Applied Sciences*, vol. 13, no. 21, p. 12008, 2023.

[9] S. Roy, S. Sankaran, and M. Zeng, "Green Intrusion Detection Systems: A Comprehensive Review and Directions," *Sensors*, vol. 24, no. 17, p. 5516, 2024.

[10] A. Hassan, S. Z. Ilyas, A. Jalil, and Z. Ullah, "Monetization of the Environmental Damage Caused by Fossil Fuels," *Environmental Science and Pollution Research*, vol. 28, pp. 21 204–21 211, 2021.

[11] H. Lee, K. Calvin, D. Dasgupta, G. Krinner, A. Mukherji, P. Thorne, C. Trisos, J. Romero, P. Aldunce, K. Barret *et al.*, "IPCC, 2023: Climate Change 2023: Synthesis Report, Summary for Policymakers. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change," in *AR6 Synthesis Report Climate Change 2023*. Intergovernmental Panel on Climate Change (IPCC), 2023.

[12] H. Sedjelmaci, S. M. Senouci, and T. Taleb, "An Accurate Security Game for Low-Resource IoT Devices," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 9381–9393, 2017.

[13] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning," *IEEE Access*, vol. 10, pp. 40 281–40 306, 2022.

[14] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, "Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6822–6834, 2019.

[15] U. Adhikari, S. Pan, T. Morris, R. Borges, and J. Beave, "Industrial Control System (ICS) Cyber Attack Datasets," *Datasets used in the Experimentation.[Online]. Available: https://sites. google. com/a/uah. edu/tommy-morris-uah/ics-data-sets*, 2019.

[16] L. A. C. Ahakonye, G. C. Amaizu, C. I. Nwakanma, J. M. Lee, and D.-S. Kim, "Classification and Characterization of Encoded Traffic in SCADA Network using Hybrid Deep Learning Scheme," *Journal of Communications and Networks*, vol. 26, no. 1, pp. 65–79, 2024.