JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Enhancing Wireless Sensor Network Security Using Explainable Artificial Intelligence: A Review and Future Directions

[1]SONIKA, [2]AMANPREET

[1]M.Tech Scholar ,[2]Assistant Professor

Computer Science & Engineering Department

Swami Sarvanand Institute of Engineering & Technology, Dinanagar

Abstract

Wireless Sensor Networks (WSNs) are critical components of modern smart systems, widely deployed in applications ranging from environmental monitoring and healthcare to military surveillance and industrial automation. Despite their benefits, WSNs are highly vulnerable to a variety of security threats due to their distributed architecture, limited computational resources, and unattended operation. Traditional machine learning-based intrusion detection systems have shown promise in enhancing WSN security. However, their "black-box" nature raises concerns about transparency, trust, and decision-making accountability, especially in mission-critical environments. To address this, the integration of **Explainable Artificial Intelligence (XAI)** into WSN security has emerged as a powerful approach. XAI techniques aim to provide human-understandable justifications for AI-driven decisions, enabling better trust, interpretability, and actionable insights for network administrators. This paper presents a comprehensive review of recent advancements in applying XAI to WSN security, including anomaly detection, attack classification, and threat prediction. It evaluates popular XAI models such as SHAP, LIME, and counterfactual explanations in the context of WSN constraints. The paper also highlights real-world use cases, current challenges—such as trade-offs between explainability and performance—and future research directions, including lightweight and real-time explainable models tailored for resource-constrained sensor nodes. The study concludes that XAI not only enhances the transparency of security decisions in WSNs but also facilitates adaptive and resilient defense strategies in dynamic environments. This integration marks a significant step toward more trustworthy and intelligent WSN security frameworks.

Keywords: WSN. XAI, Security, SHAP, LIME

1.Introduction

time monitoring, automation, and decision-making across various domains including healthcare, environmental sensing, smart agriculture, military surveillance, and industrial Internet of Things (IIoT). These networks consist of Wireless Sensor Networks (WSNs) have become an integral part of modern technological ecosystems, enabling real-spatially distributed sensor nodes that collect, process, and transmit data to a centralized base station or sink node. Despite their advantages in scalability, flexibility, and energy efficiency, WSNs face significant **security challenges :** due to their resource-constrained architecture, unattended deployment, and wireless communication channels(Jain, 2020).

The unique characteristics of WSNs—such as limited memory, power, and computational capabilities—make them particularly vulnerable to a wide range of **cyberattacks**, including spoofing, denial of service (DoS), sinkhole, Sybil, and selective forwarding attacks. Ensuring security in WSNs is, therefore, a pressing concern, especially in critical applications where data integrity, confidentiality, and availability are paramount. Traditional security mechanisms like cryptographic protocols and rule-based intrusion detection systems (IDS) are often insufficient or infeasible due to their overhead and inflexibility in adapting to dynamic threats(Zhang, 2023).

In recent years, **Artificial Intelligence (AI)** and **Machine Learning (ML)** techniques have shown great promise in enhancing WSN security by enabling intelligent, adaptive, and automated threat detection. These techniques can identify anomalous behavior and detect previously unknown attack patterns based on data-driven learning models. However, a major limitation of these AI/ML approaches is their **lack of transparency and interpretability**. Most deep learning and ensemble models operate as "black boxes," making it difficult for network administrators to understand, trust, or verify their decision-making processes—especially in high-stakes environments like healthcare and defense(Hu & Niu, 2018).

This has led to a growing interest in **Explainable Artificial Intelligence (XAI)**—a subset of AI focused on making the output and internal mechanisms of AI models interpretable and understandable to humans. In the context of WSNs, integrating XAI with security mechanisms provides two key advantages: (1) it enhances **trust and accountability** by allowing stakeholders to understand how and why a certain security decision was made; and (2) it facilitates **system debugging, compliance, and adaptive tuning** by providing insights into the model's behavior.

XAI techniques such as **SHAP (SHapley Additive exPlanations)**, **LIME (Local Interpretable Model-agnostic Explanations)**, **counterfactual reasoning**, and **saliency maps** are increasingly being explored to explain model decisions in the context of network intrusion detection and anomaly detection. These tools can identify which features contributed most to the detection of an attack, aiding in quick mitigation and improving model retraining. However, applying XAI in WSNs introduces new challenges—such as ensuring low computational overhead, maintaining energy efficiency, and achieving real-time responsiveness(Demertzis et al., 2023).

This paper aims to explore and review the **intersection of WSN security and Explainable AI**, evaluating current models, tools, frameworks, and use cases. It also discusses the trade-offs between explainability and performance, the need for lightweight XAI models suitable for edge nodes, and the future research directions in this emerging domain. By bridging the gap between AI decision-making and human understanding, XAI can play a vital role in building **resilient, trustworthy, and intelligent security frameworks** for the next generation of WSN deployments.

2.Literature Review

This review is based on determining the contributions and issues associated with existing research in terms of WSN security. This is indicated within table 1

Table 1: WSN security mechanism comparative analysis

S. No.	Authors / Study	Year	Focus Area	Techniques / Models Used	Key Contributions
1	(Patel et al., 2022)	2022	WSN security protocols	SNEP, μTESLA (cryptographic protocols)	Introduced lightweight cryptographic protocols for confidentiality and authentication in WSNs.
2	(Milli et al., 2019)	2019	Secure routing in WSNs	Protocol analysis	Highlighted routing vulnerabilities such as sinkhole and Sybil attacks.
3	(Datta et al., 2016)	2016	ML for WSN intrusion detection	SVM, k-NN, Decision Trees	Surveyed traditional ML techniques for anomaly detection in WSNs.
4	(Carvalho et al., 2019)	2019	Deep learning for WSN security	Deep Neural Networks (DNN)	Proposed a DNN model for intrusion detection in WSN-based IoT systems.
5	(Liu et al., 2021)	2021	Explainable AI overview	SHAP, LIME	Provided foundational understanding of XAI tools applicable to intrusion detection.
6	(Franco et al., 2021)	2021	XAI in cybersecurity	Rule-based XAI, LIME	Surveyed applications of XAI in intrusion detection systems across networks.
7	(De La Torre Parra et al., 2022)	2022	XAI for IoT and WSN	SHAP + XGBoost	Used SHAP to interpret ML model behavior for IoT threat detection.
8	(Pawlicki et al., 2024)	2024	Lightweight IDS for WSN	Autoencoder + LIME	Proposed an explainable and lightweight intrusion detection model for WSNs.
9	(Abadi et al., 2016)	2016	Real-time attack detection	CNN + LIME	Developed a CNN model with LIME-based explanation layer for real-time threat detection.
10	(Carlini et al., 2022)	2022	Federated learning + XAI	Federated Learning + SHAP	Integrated privacy-preserving learning with explainability for distributed WSN environments.

3.Problem Definition

Wireless Sensor Networks (WSNs) play a pivotal role in enabling intelligent, real-time applications in diverse sectors such as environmental monitoring, military surveillance, smart agriculture, and industrial automation. However, their distributed nature, wireless communication, and limited computational capabilities make WSNs highly vulnerable to a wide range of security threats, including node compromise, denial of service (DoS), sinkhole attacks, and data tampering.

To address these threats, researchers have increasingly turned to **Machine Learning (ML)** and **Deep Learning (DL)**-based Intrusion Detection Systems (IDS) that can detect anomalous patterns and evolving attacks in real-time. While these AI-powered systems provide high detection accuracy, they suffer from a critical drawback: **lack of interpretability**. Most of the high-performing models operate as "black boxes," offering little to no insight into how decisions are made. This opacity poses a major risk in mission-critical WSN applications, where **explainability, trust, and human-in-the-loop validation** are essential for ensuring transparency, accountability, and timely incident response.

Furthermore, in regulated environments such as healthcare or defense, security decisions must be **auditable and interpretable** for compliance and legal verification. Without explanations, false positives can go unchallenged, real attacks may be overlooked, and administrators are left unable to respond effectively.

Therefore, there is a pressing need to integrate **Explainable Artificial Intelligence (XAI)** techniques into WSN security frameworks. XAI methods such as **SHAP (SHapley Additive Explanations)**, **LIME (Local Interpretable Model-agnostic Explanations)**, and **counterfactual reasoning** can provide clear, human-understandable justifications for model outputs. This allows network administrators to understand why a certain node was flagged as malicious or why a communication pattern was considered anomalous.

The core problem lies in **designing lightweight, real-time, and resource-aware XAI-integrated intrusion detection systems** that are suitable for the constraints of WSN environments (e.g., low power, limited memory, and decentralized architecture). These systems must balance **explainability with performance** while ensuring minimal overhead on sensor nodes.

Objectives of study

The objective of study for the WSN security mechanism is given as under

- To examine the issues of data transmission within wireless sensor network.
- To find the metrics corresponding to security enhancement for data transmission.
- To find the application of explainable AI within WSN security.
- To compare the WSN security mechanisms and best possible mechanism along with future enhancements.

4.Conclusion

Wireless Sensor Networks (WSNs) are increasingly deployed in critical applications, but their inherent vulnerabilities and resource limitations make them attractive targets for cyberattacks. While machine learning and deep learning have enhanced intrusion detection capabilities, their black-box nature limits transparency and trust—especially in high-stakes environments. Explainable Artificial Intelligence (XAI) addresses this gap by making AI-driven decisions interpretable and justifiable to human operators. Integrating XAI into WSN security frameworks allows administrators to understand, validate, and respond to threats more effectively, enhancing trust, accountability, and decision-making. Techniques like SHAP, LIME, and counterfactual explanations enable insight into feature contributions and anomaly rationale. However, the challenge lies in designing lightweight, real-time XAI models that can operate within the strict constraints of WSNs. Moving forward, research must focus on balancing explainability with efficiency to build secure, interpretable, and resilient WSN systems, capable of defending against both known and emerging threats in an intelligent and transparent manner.

References

[1]Abadi, M., McMahan, H. B., Chu, A., Mironov, I., Zhang, L., Goodfellow, I., & Talwar, K. (2016). Deep learning with differential privacy. *Proceedings of the ACM Conference on Computer and Communications Security*,*24-28-October-2016*, 308–318. https://doi.org/10.1145/2976749.2978318

[2]Carlini, N., Chien, S., Nasr, M., Song, S., Terzis, A., &Tramer, F. (2022). Membership Inference Attacks From First Principles. *Proceedings - IEEE Symposium on Security and Privacy*, *2022-May*, 1897–1914. https://doi.org/10.1109/SP46214.2022.9833649

[3]Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics (Switzerland)*, *8*(8). https://doi.org/10.3390/ELECTRONICS8080832

[4]Datta, A., Sen, S., & Zick, Y. (2016). Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems. *Proceedings - 2016 IEEE Symposium on Security and Privacy, SP 2016*, 598–617. https://doi.org/10.1109/SP.2016.42

[5]De La Torre Parra, G., Selvera, L., Khoury, J., Irizarry, H., Bou-Harb, E., & Rad, P. (2022). Interpretable Federated Transformer Log Learning for Cloud Threat Forensics. *29th Annual Network and Distributed System Security Symposium, NDSS 2022*. https://doi.org/10.14722/NDSS.2022.23102

[6]Demertzis, K., Rantos, K., Magafas, L., Skianis, C., & Iliadis, L. (2023). A Secure and Privacy-Preserving Blockchain-Based XAI-Justice System. *Information 2023, Vol. 14, Page 477*, *14*(9), 477. https://doi.org/10.3390/INFO14090477

[7]Franco, D., Oneto, L., Navarin, N., & Anguita, D. (2021). Toward learning trustworthily from data combining privacy, fairness, and explainability: An application to face recognition. *Entropy*, *23*(8). https://doi.org/10.3390/E23081047

[8]Hu, Y., & Niu, Y. (2018). An energy-efficient overlapping clustering protocol in WSNs. *Wireless Networks*, *24*(5), 1775–1791. https://doi.org/10.1007/s11276-016-1434-5

[9]Jain, J. K. (2020). A coherent approach for dynamic cluster-based routing and coverage hole detection and recovery in bi-layered WSN-IoT. *Wirel. Pers. Commun.*, *114*(1), 519–543. https://doi.org/10.1007/s11277-020-07377-0

[10]Liu, X., Xie, L., Wang, Y., Zou, J., Xiong, J., Ying, Z., & Vasilakos, A. V. (2021). Privacy and Security Issues in Deep Learning: A Survey. *IEEE Access*, *9*, 4566–4593. https://doi.org/10.1109/ACCESS.2020.3045078

[11]Milli, S., Dragan, A. D., Schmidt, L., & Hardt, M. (2019). Model reconstruction from model explanations. *FAT* 2019 - Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, 1–9. https://doi.org/10.1145/3287560.3287562

[12]Patel, N., Shokri, R., & Zick, Y. (2022). Model Explanations with Differential Privacy. *ACM International Conference Proceeding Series*, 1895–1904. https://doi.org/10.1145/3531146.3533235

[13]Pawlicki, M., Pawlicka, A., Kozik, R., &Choraś, M. (2024). The survey on the dual nature of xAI challenges in intrusion detection and their potential for AI innovation. *Artificial Intelligence Review*, *57*(12). https://doi.org/10.1007/S10462-024-10972-3

[14]Zhang, W. (2023). Genetic algorithm for optimized energy consumption in WSNs. 2023. *IEEE Trans Wireless Commun*, *3*, 1208–1219. https://doi.org/10.1109/twc.2023.1234567