Cybersecurity and Intrusion Detection in Big Data's Wireless Sensor Networks:A Survey

Naima Samout
ISET Gafsa Tunisia
SETIT Lab
ENIG Tunisia

Thouraya Gouasmi
University of Gafsa,
ISSATG, Tunisia

Nejah Nasri
SETIT Laboratory
Sfax, Tunisia

***Abstract*—Wireless Sensor Networks are becoming more and more crucial to the advancement of numerous technologies, particularly when combined with Big Data platforms. Although this connection has a lot of potential, there are challenging security challenges as well. Even though WSNs have been the subject of a lot of research, the security requirements for WSNs functioning in Big Data environments have not yet been thoroughly examined. It is also a crucial use of IoT, allowing sensors to exchange a variety of data. However, because of its inherent unreliability and natural surroundings, such a network is susceptible to numerous types of attacks, including insider attacks. Intrusion detection systems (IDSs) are commonly used in WSNs to protect against insider assaults by putting in place the right procedures and techniques. However, sensors may produce too much data in the big data era, which could reduce the efficiency of WSN computing. An overview of the security concerns and difficulties facing WSNs in the big data era is provided in this study. In order to improve the detection of insider threats and the overall security posture of WSNs, a literature review on cybersecurity IDS on WSN in the context of big data is finally suggested. It highlights advancements in IDS methodologies, including federated learning, machine learning, deep learning, and big data techniques.**

***Index Terms*—Cybersecurity, WSN, Big Data, IDS, Intrusion Detection, Machine Learning, Deep Learning, Federated Learning**

I. Introduction

Wireless Sensor Networks are networks made up of autonomous, spatially dispersed sensor nodes that cooperate to track various environmental or physical factors, including motion, pressure, temperature, humidity, and others [1]. Through wireless communication, these nodes transmit data to a server or central processing unit for analysis. An other definition, a WSN is a network of devices, known as nodes, that are able to sense their surroundings and use wireless networks to transmit the data they have collected from the monitored field [2]. WSNs have a wide range of applications across various fields. In environmental monitoring, they are utilized for tracking climate changes, managing natural resources, and detecting pollution levels. In healthcare, they support patient monitoring and facilitate telemedicine In addition, WSNs play a crucial role in industrial automation, smart agriculture, and military surveillance. They have garnered considerable interest, particularly as they have been integrated with different technologies like Big Data, putting them as crucial components of ubiquitous computing [3]. Despite the large amount of research on WSNs, there is still a notable gap in the security challenges that arise when these networks are used within Big Data era. Without consideration of security requirements, the integration of WSNs on Big Data systems can result in serious vulnerabilities. The vast amount of data generated by WSNs in Big Data contexts presents many challenges, including the potential for data breaches, unauthorized access, and data manipulation, all of which can compromise the overall efficiency and performance of the system. It is impossible to overestimate the significance of strong cybersecurity measures in WSNs, particularly in light of the volume, diversity, and velocity of data connected to big data environments. Sensitive data can be compromised in the absence of proper protection, resulting in breaches that could seriously affect operational integrity, privacy, and safety. IDSs are essential for protecting WSNs against possible online attacks. IDS must be able to process massive amounts of data effectively in the setting of big data while guaranteeing that the network's performance is not adversely affected. To help understanding the current security problems affecting those networks, we review in this paper their characteristics, threats, vulnerabilities and attacks and examine several security solutions to protect them. In particular, we survey the newest literature by focusing our attention on intrusion detection in WSN in the context of big data. The paper is organized as follows.

Section II presents a synthesis of threats and attacks targeting big data's WSNs. In Section III, we present a literature survey on Cybersecurity IDS on WSN in the era of Big Data, focusing on those that exploit machine learning, deep learning, federated learning and hybrid learning with Big data technologies. Finally, Section IV draws some conclusions.

II. Synthesis of Threats on Big Data WSNs

With the responsibility of gathering data in realtime, WSNs play a crucial role in big data analytics. Their function, however, exposes them to a variety of cyberattacks. Every attack compromises the availability, quality, and integrity of data, with cascading consequences on the Big Data technologies that depend on the information gathered. A summary of significant risks and implications in relation to big data is provided here.

Blackhole Attack:

A malicious node that poses as a trustworthy network node can capture and discard all data packets without sending them to their intended recipient in a blackhole attack. According to

N Panda, M Supriya [4], this attack has serious implications for WSN integration with Big Data systems, since incomplete or absent data can seriously disrupt analytics and decision-making procedures.

Selective Forwarding Attack:

The infected node forwards certain packets while dropping others in this attack. Shiyao Luo, al. have discussed how selective forwarding results in biased data collecting, which is especially harmful in Big Data applications where accuracy is crucial. The authors of [7] study analyzes the models of selective forwarding attacks and proposes an abnormal node detection method, which includes a node behavior measurement scheme and trust-value evaluation mechanism.

Sinkhole Attack:

When a compromised node incorrectly offers an ideal route to the base station, it attracts traffic from other nodes and subsequently drops or manipulates the data. This is known as a sinkhole attack. According to Omar,A [8]the main target of a sinkhole attack is to alter the topology of the network so that all the traffic in the network is redirected through the malicious node. The gray hole attack is an enhanced form of the black hole attack, in which the hacked node will only drop packets with a specific probability in order to avoid detection and exclusion from the network .

Data Flooding Attack:

The network is overwhelmed with pointless data packets as a result of this assault, exhausting the memory and bandwidth of the sensor nodes. M Dener [9] draw attention to the grave repercussions of data flooding in Big Data WSNs. It commonly causes transmission delays and keeps legitimate data from being processed by the system. The main goal of the flooding assault is to steal power by using up a significant quantity of network and battery bandwidth.

Sybil Attack:

A Sybil attack confuses the network and interferes with functions like routing, voting, and data aggregation by having an opponent create numerous false identities. According to Karthikeyan,M, et al [10], Sybil attacks can distort Big Data analysis by producing erroneous data aggregation points, which can result in conclusions that are not correct.

Jamming Attack:

A jamming attack is a form of denial-of-service (DoS) attack where malicious nodes transmit interference signals to obstruct communication between legitimate network nodes. This type of attack can severely impair the reliability of wireless sensor networks, leading to significant delays or the loss of essential sensor data [12]. Such disruptions pose a considerable threat to time-sensitive Big Data applications, where the timely and accurate collection of data is critical for informed decision-making and system responsiveness.

Replay Attack:

Replay attacks involve capturing legitimate network packets and retransmitting them at later times, misleading the network into processing outdated data as if it were current. When outdated information is processed, it can skew results, leading to incorrect analyses and potentially flawed decision-making. As highlighted by Elsaeidy et al. [10], the use of outdated data not only distorts analytics but can mask real trends or create false ones, reducing trust in the system's outputs.

Node Replication Attack:

An adversary can obtain unauthorized access to a network by impersonating a legitimate node in a node replication attack, which poses a major security risk. An attacker can get access to the network and perhaps alter or compromise data by copying a node's login credentials. According to M. Jamshidi, M. Esnaashari, A. M. Darwesh, and M. R. Meybodi. [11]these kinds of assaults compromise the legitimacy and dependability of data collecting procedures, presenting serious threats to Big Data systems that depend on safe and reliable data.

Routing Table Overflow Attack:

This attack focuses on a node's routing table in a wireless network and is frequently called a routing table overflow or routing table poisoning attack [12]. By corrupting the routing table, the attacker causes legitimate routes to be ignored, overwritten, or deleted. As a result, when a node attempts to send data, it may be unable to find a valid or optimal route to forward the data. In some cases, the node may end up using faulty or inefficient paths, or even no path at all, leading to data loss or misrouting.

III. Literature survey On Cybersecurity IDS on wsn in the Context of Big Data

Due to the vast amount, speed, and diversity of data produced by Wireless Sensor Networks, intrusion detection presents special difficulties where conventional security measures might not be sufficient. The limitations of WSN nodes, such their short battery life, bandwidth, and processing power, must be balanced with the handling of Big Data by Intrusion Detection Systems. Beyond processing enormous datasets, big data plays a role in WSN intrusion detection by facilitating deeper insights via machine learning, anomaly detection, and adaptive approaches. Big Data, however, also makes data security, scalability, and privacy more difficult—problems that are critical for WSNs functioning in dispersed and resource-constrained situations. This survey delves into the current and recent research on intrusion detection in WSN security, focusing on issues and solutions relevant to Big Data contexts.

Machine Learning Methods for Cyber Attacks Detection in Big Data WSN

One of the most important areas of research in WSN security is the application of machine learning algorithms. In this section, we provide a summary of a number of algorithms and group them into different categories, including supervised, unsupervised, reinforcement learning, deep learning, and federated learning. This classification offers a comprehensive overview with those that do not use big data techniques.

- **Supervised Machine Learning Methods**

Decision Trees (DT) are a popular supervised learning technique in intrusion detection systems (IDS) because they are easy to understand, effective for lightweight sensor nodes, and simple. These are very helpful for establishing precise cutoff

points for data classification. An ensemble of decision trees called Random Forest (RF) increases accuracy and robustness, which makes it appropriate for identifying assaults in dynamic network settings like sinkholes or blackholes. A Hadoop-based automatic intrusion detection system designed for extremely fast massive data environments was reported in this paper [36]. REPTree, Support Vector Machine, Random Forest Tree, Naïve Bayes, J48, and Conjunctive Rule classifiers are five well-known machine learning techniques that they employed. With over 99% true positive (TP) rates and less than 0.001% false positives (FP), the results demonstrated that REPTree and J48 were the most effective classifiers. The Spark-Chi-SVM intrusion detection model, which can handle big data, was introduced by the researchers.They used KDD99 to train and test the model. The experiment showed that the model has high performance and reduces the false positive rate and 99.55% accuracy. The Spark Big Data platform, which can handle and analyze data quickly, was employed in the suggested model. The enormous dimensionality of big data makes classification more difficult and time-consuming. Therefore, the researchers utilized SVMWithSGD to categorize data into normal or attack and ChiSqSelector to choose associated features in the suggested model [37].

In [21], authors presented a method that uses the Random Forest algorithm to train a classifier for intrusion detection after addressing dataset imbalance with the Synthetic Minority Oversampling Technique (SMOTE). The KDD Cup 99 dataeset was used for simulations, which showed that the Random Forest approach outperformed other algorithms with an accuracy of 92.39%. Moreover, the Random Forest classifier's accuracy increased to 92.57% by using SMOTE to oversample the minority samples. This suggests that the suggested approach improves intrusion detection performance and successfully resolves class imbalance issues. According to Thanthrige et al. [22], Chi-squared statistics, information gain, and feature reduction strategies assisted in speed of classification and detection accuracy. Using machine learning models such as Random Tree, OneR, Random Forest, AdaBoost. All of these methods achieved an accuracy of greater than 95%. Rezvi et al [24] applied a number of classification algorithms, such as KNN, Naöve Bayes, Logistic Regression, Support Vector Machine (SVM), and Artificial Neural Network (ANN), to the WSN-DS dataset and evaluated how well they identified different kinds of DoS attacks using a data mining approach. According to the analysis, ANN had the highest accuracy (98.56%), closely followed by KNN (98.4%).

- **Unsupervised Machine Learning Methods**

Without being pre-trained on particular threats, the intrusion detection system use methods in unsupervised machine learning to spot patterns in network data. For intrusion detection in Big Data era, the cluster machine learning technique was employed by Ferhat et al. [38] To identify whether network traffic is an attack or a typical one, the authors employed the k-Means approach in the machine learning libraries on Spark. The authors of this suggested approach did not choose the relevant characteristics using a feature selection technique.The KDD Cup 1999 is used for testing and training in the suggested approach and attained an exceptional accuracy of 99.99% . Using unsupervised clustering, M Zakariah et al. [27] introduced various classifiers to define the violation by filtering the information. The NSL-KDD dataset was used. The decision tree classifier was used in its creation, but it is limited to binary classification. In 2024, [25] suggested a novel Stochastic Machine Learning-Based Attack Detection System for WSNs that combines Gaussian Mixture Models (GMMs) and Hidden Markov Models (HMMs). In order to reduce dimensionality in WSN datasets while maintaining essential routing properties, the system employed Principal Component Analysis. The exceptional performance of the system was demonstrated by the experimental results, which indicated that a configuration of three HMMs and four GMMs achieved an accuracy of 94. 55%. This strategy offered a viable way to improve the security of WSNs.

Deep-Learning Methods

Deep learning, a subfield of machine learning, has grown in popularity and been applied recently to intrusion detection; studies show that deep learning works much better than traditional methods. Deep learning for WSN is necessary to create effective intrusion detection and prevention systems. The enormous dimensionality of large data and the dynamic nature of WSN systems have presented issues that have been addressed by a variety of architectures and methodologies. So,in the Big Data era, many deep learning IDS used in WSN are cited in this section. CNNs have provenopular option for intrusion detection and feature extraction.

Wang [30] suggested a detection system based on convolutional neural network and their models as they applyed deep learning to IDS. The study analyzes these models' performance on important datasets including UNSW-NB15 and KDD Cup 99 and talks about how well they capture temporal and spatial aspects. The study also discusses issues with real-time performance, computational complexity, and model interpretability. It also makes recommendations for future research areas, such as model optimization, lightweight architectures, and enhanced interpretability. Several attacks were recognized and categorized using the CNN model developed by Kaur [31]. The efficacy of their methodology was assessed using the CICIDS-2017 and CICIDS-2018 datasets. Their model separated attacks into a number of groups. In the context of IDS, Aleesa [35] investigated the use of DL models for classification. They used the UNSW-NB15 to perform their analyses. In particular, the study evaluated the performance of three different kinds of neural network models: ANN, RNN, and DNN. To increase its accuracy, they employed min-max normalization after employing data cleaning methods, such as managing missing and category variables. When accuracy was used to gauge efficiency, the ANN, RNN, and DNN produced, respectively, 97.89%, 85.4%, and 95.9% for multilabel classification and 99.26%, 85.42%, and 99.22% for binary classification. Additionally, in [39], the authors had

suggested a deep learning method that uses the Bidirectional LSTM Recurrent Neural Network (BLSTM RNN) to enable intrusion detection in IoT networks. Seven measures have been used to assess the model: accuracy, precision,FAR, detection time, recall, f1-score, and miscalculation rate. The accuracy of the suggested model was a remarkable 95.7%. However, only one dataset has been used to assess the suggested model. Furthermore, the model was not evaluated by comparison with other models of a similar nature. Chuanlong Yin [41] proposed a deep learning approach for intrusion detection using recurrent neural networks(RNN-IDS). They analysed the multi-classification of the RNN-IDS model based on the NSL-KDD dataset and The experimental tests demonstrated that RNN-IDS performs better than conventional machine learning classification techniques in both binary and multiclass classification, and that it is very appropriate for developing a classification model with high accuracy to 99.81%. Furthermore,Vimalkumar [40] created a big data framework for intrusion detection utilizing classification techniques including naïve Bayes, DNN, SVM, random forest, and decision trees. Accuracy, recall, false rate, specificity, and prediction time are the measures utilized for assessment. One application for Apache Spark is as a platform for applying big data analytics to intrusion detection in smart grids. According to their claims, the DNN algorithm achieves the best accuracy on the unprocessed dataset. Although the accuracy is less than 80%, the DNN model achieved the maximum accuracy. Furthermore, when compared to other models, the DNN forecast time is longer.

Federated-Learning Methods

It's a machine learning technique in which several nodes or devices work together to train a model without directly exchanging data. Instead, just the model updates or summaries are shared and aggregated; the model is trained locally on each node. To put it another way, ML/DL's features allow it to be trained over a number of devices and servers using decentralized data and iterations [44]. Federated learning has emerged as a compelling method for real-time cyber threat detection in Big Data WSNs, where the volume and diversity of data can be daunting. WSNs are perfect candidates for FL because of their size and decentralized structure, which allows for the use of the rich data produced by WSN nodes while addressing the issue of data privacy. Multi-View Federated-based Learning for Intrusion Detection (MV-FLID) is a proposal by Dinesh Chowdary et al. [14]. This provides the most unique prediction and can learn from many data views. When compared to KNNs, DT, and SVM, the Deep Feature Embedding Learning (DFEL) model [33] yields an F1 score of 99.14%. The technique known as Transient Search Optimization (TSO) keeps the stages of exploration and exploitation in balance. The most widely used IoT datasets, such as KDD99, NSL-KDD, and CICIDS-2017, are used to evaluate the model. It outperforms a number of current methods in terms of accuracy. For instance, in order to identify hacked IoT devices in the network, the authors of [34] suggested a self-learning anomaly detection system that uses a FL technique. The suggested model is constructed using Gated Recurrent Units (GRUs) and Long Short Term Memory (LSTM). With a 98.2% accuracy rate, it can identify 95.6% of attacks in 257 ms with a lower false alarm rate. In 2024, Khan R [42] proposed a federated-learning-based anomaly detection for IoT security attacks which detects distributed sensor faults using Long Short-Term Memory (LSTM) networks. Data privacy is maintained by this decentralized method, which enables several clients (sensors) to train local models on their data and only communicate model updates to a central server. The experimental findings show that FedLSTM detects sensor defects with high accuracy (94.45%) while protecting the privacy of the data. Zhuo Chen [43] suggested a federated learning-based technique for wireless edge network intrusion detection. Data privacy is ensured by the technique, which trains machine learning models across dispersed edge devices without requiring raw data sharing. The results of the study demonstrate that the federated learning-based intrusion detection system identifies network intrusions with good accuracy, roughly 99.28%, and efficiency while significantly decreasing communication overhead when compared to traditional centralised models. By preserving decentralized model training across edge devices, the suggested approach not only solves privacy issues but also guarantees strong intrusion detection.

Hybrid-Learning Methods

The hybrid technique, which combines two or more of the earlier types, is the last type of machine learning methods. They are designed to reduce the FP rate from unknown attackers and increase known intrusion detection rates. Research and analysis did not cover the full range of pure anomaly detection approaches, particularly in large data wsn sectors, because the majority of strategies were hybrid. The more sophisticated data distribution of the infiltration patterns is difficult for the single-learning model technique to understand. To greatly enhance the hybrid performance of machine learning-based intrusion detection systems Research continues on deep, fererated, machine, and big data learning.

The study [26] offered a novel approach using a networking chatbot and an advanced deep recurrent neural network framework called Long Short Term. The Apache Spark framework memory. Detection rates are higher than those of traditional IDS, and false positives are decreasing by 10% compared to typical learning models, according to experimental validation.

Bukhari et al. [47] presented an advanced solution for detecting intrusions in wireless sensor networks. They proposed a method that combines federated learning (FL) with a specialized machine learning architecture involving Separable Convolutional Neural Networks (SCNNs) and Bidirectional Long Short-Term Memory (Bi-LSTM) networks. This results in the highest accuracy (99.7%) for the suggested approach (FL-SCNN-Bi-LSTM) for the WSN-DS Dataset. Other classifiers, such as CNN-Bi-LSTM, KNN, SVM, RF (Random Forest), NN (Neural Network), and LightGBM, also performed well and offered high prediction accuracy. With the

given dataset, the recommended model operated efficiently while maintaining data privacy. On CICISD-2017 Dataset , the proposed model demonstrated remarkable performance. Specifically, the model obtained an an accuracy of 99.93% when employing the method of Federated Learning. The model without FL had somewhat worse accuracy,whith 99.9%, then the model with FL in turn. Indra . A hybrid intrusion detection model combining two machine learning techniques, Random Forest and XGBoost, was proposed by Farah Jemili [45]. These algorithms were selected because intrusion detection tasks require them to be more effective. They, first, divided the dataset into training and testing sets in order to train the hybrid model. After using the training set to train the Random Forest algorithm, they used the testing set to get its predictions. Next, they fed the XGBoost algorithm with the Random Forest method's predictions. Finally, they use three datasets—N-Balot, NSL-KDD, and CICIDS2017—to evaluate the effectiveness of the suggested methodology. On the N-BaIoT dataset, the Random Forest model performed admirably, with accuracy of 96.31%. However, the decision tree only produced an accuracy of 95.74%, which is comparable to the outcomes of the NSL-KDD (94.97%) and CICIDS2017 (95.86%) dataset analyses.Additionally, combining XGBoost and Random Forest resulted in even higher accuracy of 96.39%, almost matching the findings of the NSL-KDD (98.21%) and CICIDS datasets. This confirms that the combined model performs better than the model that is only based on the decision tree. In 2024, M Sajid et al. [46] presented and tested an hybrid approach that , for intrusion detection, a combination of machine learning and deep learning approaches is used. In this study, three different machine and deep learning approaches were used: XGBoost, CNN, and LSTM. With the CIC-IDS 2017 and WSN DS datasets, they used the CNN-LSTM model; with the UNSW NB15 and NSL KDD datasets, they used the XGBoost-LSTM model.The results indicated that the CNN-LSTM model was marginally outperformed by XGBoost with LSTM. These results also shown that, compared to other approaches, the hybrid models performed better, ran at maximum efficiency, and achieved 98.40% accuracy. The authors Talukder [49], in 2024, combined six machine learning methods to the SMOTE-Tomek algorithm to produce a balanced dataset, which enhanced WSN intrusion detection.Additionally, in order to enhance the performance of the suggested technique for WSNs and provide data balancing efficiency, they carried out two distinct experiments on a WSN-DS dataset: one using SMOTETomek-Link and the other without. Overall, all of the machine learning algorithms' accuracy rates increased when the SMOTETomek-Link approach was included with a 99.78% accuracy rate in binary classification and 99.92% in multiclass classification.

IV. Conclusions and Perspectives

In the context of big data systems, cybersecurity in WSNs has grown in importance. These networks produce enormous volumes of data, necessitating strong and effective defenses against changing threats. In order to monitor and protect WSNs, intrusion detection systems are essential, particularly when dealing with the complexity of big data environments. This study offers a thorough analysis of the state of cybersecurity in WSNs today, looking at how IDS might be used to improve network security. Our overview is followed by a thorough state-of-the-art survey of intrusion detection and cybersecurity in WSNs in the big data era. An examination of security needs are presented first, followed by a summary of WSN-specific threats and attacks in big data situations. Lastly, a survey of the literature on cybersecurity intrusion detection systems for WSNs in the big data setting is addressed.

References

[1] R. Verdone, D. Dardari, G. Mazzini, and A. Conti, Wireless sensor and actuator networks: technologies, analysis and design. Academic Press, 2010.

[2] F. Rabeb, F. Faleh, and others, 'An extensive comparison among DSDV, DSR and AODV protocols in wireless sensor network', in *Proc. Int. Conf. Education and e-Learning Innovations (ICEELI), 2012.

[3] E. Barka, C. A. Kerrache, R. Hussain, N. Lagraa, A. Lakas, and S. H. Bouk, 'A trusted lightweight communication strategy for flying named data networking', Sensors, vol. 18, no. 8, p. 2683, 2018.

[4] S. Balakrishna and M. Thirumaran, 'Semantic interoperability in IoT and big data for health care: a collaborative approach', in Handbook of data science approaches for biomedical engineering, Elsevier, 2020, pp. 185–220.

[5] S. Ali, M. A. Khan, J. Ahmad, A. W. Malik, and A. ur Rehman, 'Detection and prevention of Black Hole Attacks in IOT and WSN', in 2018 third international conference on fog and mobile edge computing (FMEC), 2018, pp. 217–226.

[6] N. Panda and M. Supriya, 'Blackhole attack prediction in wireless sensor networks using support vector machine', in Advances in Signal Processing, Embedded Systems and IoT: Proceedings of Seventh ICMEET-2022, Springer, 2023, pp. 321–331.

[7] N. M. Alajmi and K. Elleithy, 'A new approach for detecting and monitoring of selective forwarding attack in wireless sensor networks', in 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT), 2016, pp. 1–6.

[8] A. A. R. A.-C. Omar, B. Soudan, and Others, 'A comprehensive survey on detection of sinkhole attack in routing over low power and Lossy network for internet of things', Internet of Things, vol. 22, p. 100750, 2023.

[9] A. M. Eid, A. B. Nassif, B. Soudan, and M. N. Injadat, 'IIoT network intrusion detection using machine learning', in 2023 6th International Conference on Intelligent Robotics and Control Engineering (IRCE), 2023, pp. 196–201.

[10] A. A. Elsaeidy, N. Jagannath, A. G. Sanchis, A. Jamalipour, and K. S. Munasinghe, 'Replay attack detection in smart cities using deep learning', IEEE Access, vol. 8, pp. 137825–137837, 2020.

[11] M. Jamshidi, M. Esnaashari, A. M. Darwesh, and M. R. Meybodi, 'Using time-location tags and watchdog nodes to defend against node replication attack in mobile wireless sensor networks', International Journal of Wireless Information Networks, vol. 27, no. 2, pp. 102–115, 2020.

[12] S. Lata, S. Mehfuz, and S. Urooj, 'Secure and reliable wsn for internet of things: Challenges and enabling technologies', IEEE Access, vol. 9, pp. 161103–161128, 2021.

[13] I. Almomani, B. Al-Kasasbeh, and M. Al-Akhras, 'WSN-DS: a dataset for intrusion detection systems in wireless sensor networks', Journal of Sensors, vol. 2016, no. 1, p. 4731953, 2016.

[14] S. Choudhary and N. Kesswani, 'Analysis of KDD-Cup'99, NSL-KDD and UNSW-NB15 Datasets using Deep Learning in IoT', Procedia Computer Science, vol. 167, pp. 1561–1573, 2020.

[15] G. Kocher and G. Kumar, 'Machine learning and deep learning methods for intrusion detection systems: recent developments and challenges', Soft Computing, vol. 25, no. 15, pp. 9731–9763, 2021.

[16] R. Zuech, T. M. Khoshgoftaar, and R. Wald, 'Intrusion detection and big heterogeneous data: a survey', Journal of Big Data, vol. 2, pp. 1–41, 2015.

[17] S. More, M. Idrissi, H. Mahmoud, and A. T. Asyhari, 'Enhanced Intrusion Detection Systems Performance with UNSW-NB15 Data Analysis', Algorithms, vol. 17, no. 2, p. 64, 2024.

[18] M. H. Kabir, M. S. Rajib, A. S. M. T. Rahman, M. M. Rahman, and S. K. Dey, 'Network intrusion detection using unsw-nb15 dataset: Stacking machine learning based approach', in 2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), 2022, pp. 1–6.

[19] Z. Chkirbene, S. Eltanbouly, M. Bashendy, N. AlNaimi, and A. Erbad, 'Hybrid machine learning for network anomaly intrusion detection', in 2020 IEEE international conference on informatics, IoT, and enabling technologies (ICIoT), 2020, pp. 163–170.

[20] M. H. Behiry and M. Aly, 'Cyberattack detection in wireless sensor networks using a hybrid feature reduction technique with AI and machine learning methods', Journal of Big Data, vol. 11, no. 1, p. 16, 2024.

[21] X. Tan et al., 'Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm', Sensors, vol. 19, no. 1, p. 203, 2019.

[22] U. S. K. P. M. Thanthrige, J. Samarabandu, and X. Wang, 'Machine learning techniques for intrusion detection on public dataset', in 2016 IEEE Canadian conference on electrical and computer engineering (CCECE), 2016, pp. 1–4.

[23] M. Revathi, V. V. Ramalingam, and B. Amutha, 'A machine learning based detection and mitigation of the DDOS attack by using SDN controller framework', Wireless Personal Communications, pp. 1–25, 2022.

[24] M. A. Rezvi, S. Moontaha, K. A. Trisha, S. T. Cynthia, and S. Ripon, 'Data mining approach to analyzing intrusion detection of wireless sensor network', Indonesian Journal of Electrical Engineering and Computer Science, vol. 21, no. 1, pp. 516–523, 2021.

[25] A. R. A. Moundounga and H. Satori, 'Stochastic machine learning based attacks detection system in wireless sensor networks', Journal of Network and Systems Management, vol. 32, no. 1, p. 17, 2024.

[26] K. Al Jallad, M. Aljnidi, and M. S. Desouki, 'Big data analysis and distributed deep learning for next-generation intrusion detection system optimization', Journal of Big Data, vol. 6, no. 1, p. 88, 2019.

[27] M. Zakariah, S. A. AlQahtani, A. M. Alawwad, and A. A. Alotaibi, 'Intrusion Detection System with Customized Machine Learning Techniques for NSL-KDD Dataset', Computers, Materials and Continua, vol. 77, no. 3, 2023.

[28] H. Sadia et al., 'Intrusion Detection System for Wireless Sensor Networks: A Machine Learning Based Approach', IEEE Access, vol. 12, pp. 52565–52582, 2024.

[29] A. Ahmim, L. Maglaras, M. A. Ferrag, M. Derdour, and H. Janicke, 'A novel hierarchical intrusion detection system based on decision tree and rules-based models', in 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), 2019, pp. 228–233.

[30] H. Wang, Z. Cao, and B. Hong, 'A network intrusion detection system based on convolutional neural network', Journal of Intelligent Aand Fuzzy Systems, vol. 38, no. 6, pp. 7623–7637, 2020.

[31] G. Kaur, A. H. Lashkari, and A. Rahali, 'Intrusion traffic detection and characterization using deep image learning', in 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), 2020, pp. 55–62.

[32] V. Mothukuri, P. Khare, R. M. Parizi, S. Pouriyeh, A. Dehghantanha, and G. Srivastava, 'Federated-learning-based anomaly detection for IoT security attacks', IEEE Internet of Things Journal, vol. 9, no. 4, pp. 2545–2554, 2021.

[33] A. Fatani, M. Abd Elaziz, A. Dahou, M. A. A. Al-Qaness, and S. Lu, 'IoT intrusion detection system using deep learning and enhanced transient search optimization', IEEE Access, vol. 9, pp. 123448–123464, 2021.

[34] T. D. Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan, and A.-R. Sadeghi, 'DÏoT: A federated self-learning anomaly detection system for IoT', in 2019 IEEE 39th International conference on distributed computing systems (ICDCS), 2019, pp. 756–767.

[35] A. Aleesa, M. Younis, A. A. Mohammed, and N. Sahar, 'Deep-intrusion detection system with enhanced UNSW-NB15 dataset based on deep learning techniques', Journal of Engineering Science and Technology, vol. 16, no. 1, pp. 711–727, 2021.

[36] Kumar, S., Singh, K.J. (2024). Intrusion Detection System Using Supervised Machine Learning. In: Swain, B.P., Dixit, U.S. (eds) Recent Advances in Electrical and Electronic Engineering. ICSTE 2023. Lecture Notes in Electrical Engineering, vol 1071. Springer, Singapore,2024.

[37] S. M. Othman, F. M. Ba-Alwi, N. T. Alsohybe, and A. Y. Al-Hashida, 'Intrusion detection model using machine learning algorithm on Big Data environment', Journal of big data, vol. 5, no. 1, pp. 1–12, 2018.

[38] F. Karataş and S. A. Korkmaz, 'Big Data: controlling fraud by using machine learning libraries on Spark', International Journal of Applied Mathematics Electronics and Computers, vol. 6, no. 1, pp. 1–5, 2018.

[39] B. Roy and H. Cheung, 'A deep learning approach for intrusion detection in internet of things using bi-directional long short-term memory recurrent neural network', in 2018 28th international telecommunication networks and applications conference (ITNAC), 2018, pp. 1–6.

[40] K. Vimalkumar and N. Radhika, 'A big data framework for intrusion detection in smart grids using apache spark', in 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017, pp. 198–204.

[41] C. Yin, Y. Zhu, J. Fei, and X. He, 'A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks', IEEE Access, vol. 5, pp. 21954–21961, 2017.

[42] R. Khan, U. Saeed, and I. Koo, 'FedLSTM: A Federated Learning Framework for Sensor Fault Detection in Wireless Sensor Networks', Electronics, vol. 13, no. 24, 2024.

[43] Z. Chen, N. Lv, P. Liu, Y. Fang, K. Chen, and W. Pan, 'Intrusion Detection for Wireless Edge Networks Based on Federated Learning', IEEE Access, vol. 8, pp. 217463–217472, 2020.

[44] S. Agrawal et al., 'Federated Learning for intrusion detection system: Concepts, challenges and future directions', Computer Communications, vol. 195, pp. 346–361, 2022.

[45] F. Jemili, R. Meddeb, and O. Korbaa, 'Intrusion detection based on ensemble learning for big data classification', Cluster Computing, vol. 27, no. 3, pp. 3771–3798, 2024.

[46] M. Sajid et al., 'Enhancing intrusion detection: a hybrid machine and deep learning approach', Journal of Cloud Computing, vol. 13, no. 1, p. 123, 2024.

[47] S. M. S. Bukhari et al., 'Secure and privacy-preserving intrusion detection in wireless sensor networks: Federated learning with SCNN-Bi-LSTM for enhanced reliability', Ad Hoc Networks, vol. 155, p. 103407, 2024.

[48] H. Xie, Z. Yan, Z. Yao, and M. Atiquzzaman, 'Data Collection for Security Measurement in Wireless Sensor Networks: A Survey', IEEE Internet of Things Journal, vol. 6, no. 2, pp. 2205–2224, 2019.

[49] Talukder, Md Alamin, et al. "MLSTL-WSN: machine learning-based intrusion detection using SMOTETomek in WSNs." International Journal of Information Security 23.3 (2024): 2139-2158.