International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

WSN Security Enhancement using Machine Learning Application

Kush Patel, Indroneil Sinha Roy
Electronics and Communication Engg. Department,
Babu Banarasi Das Northern India Institute of Technology, Lucknow, India
kush4504@gmail.com, indroneilroy79@gmail.com

Abstract: Wireless Sensor Networks (WSNs) are highly susceptible to routing-based attacks such as wormhole, Sybil, and blackhole attacks, which significantly degrade network performance, localization accuracy, and Quality of Service (QoS). This paper proposes a secure and energy-efficient framework for the detection and localization of malicious nodes using hybrid machine learning techniques. The proposed approach integrates anchor-assisted localization with optimized intrusion detection mechanisms to accurately estimate node positions and identify abnormal behavior. Advanced preprocessing, feature engineering, and clustering-based labeling are employed to enhance classification performance, while hyperparameter tuning and Bayesian optimization improve model efficiency and detection accuracy. The framework is evaluated using benchmark intrusion detection datasets, and performance is measured in terms of accuracy, precision, recall, localization error, and network lifetime. Simulation results demonstrate that the proposed model achieves high detection accuracy with reduced false positives and significantly lower localization error compared to conventional methods. Additionally, the approach minimizes communication overhead and energy consumption, thereby improving the overall reliability, security, and scalability of WSNs.

Keywords: WSNs, Machine Learning, Localization, Intrusion Detection, Energy Efficiency.

1. Introduction:

The identification and localization of malicious nodes in Wireless Sensor Networks (WSNs) has garnered a significant amount of attention due to the fact that it plays a crucial part in determining the durability, reliability, and overall performance of the system. Generally speaking, localization strategies rely on anchor nodes that have specified coordinates in order to estimate the positions of unknown sensors. Nevertheless, getting precise positioning within a confined initialization duration continues to be a challenging endeavor. In the presence of routing-based attacks, such as wormhole, Sybil, blackhole, and replay assaults, this problem becomes much more serious. These attacks disrupt the functionality of routing and misuse network resources, which ultimately leads to a reduction in the accuracy of localization and a degradation of Quality of Service (QoS).

In order to address these problems, a safe architecture that incorporates both localization and routing attack detection is provided. This framework makes use of hybrid machine learning algorithms that are optimized. The solution that has been developed focuses on determining the appropriate distances between nodes, ensuring that sensors are placed precisely, and enabling reliable communication throughout the network.

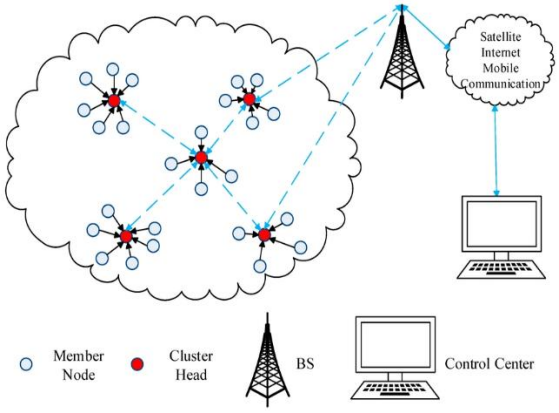

Figure1. Illustration of working scenario of wireless sensor network.

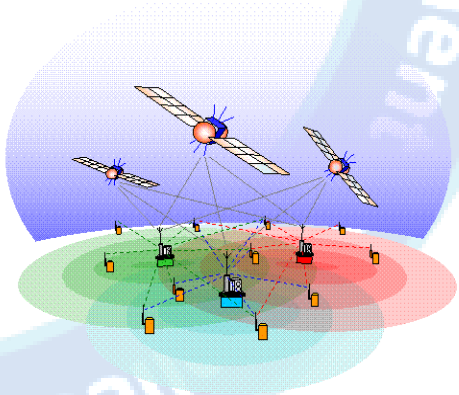

Figure 2. Localization in WSN for finding the position of sensor nodes.

In order to evaluate the effectiveness of detecting malicious nodes as well as the performance of the localization process, intrusion detection datasets are utilized. Furthermore, a sophisticated model that is based on machine learning is utilized in order to categorize nodes into the categories of regular and harmful operations. Based on the results of the simulation, it can be concluded that the framework that has been proposed is

capable of achieving a high level of detection accuracy while also greatly improving localization performance with little localization error.In the context of Wireless Sensor Networks (WSNs), the term "localization" refers to the process of establishing the spatial placements of sensor nodes once they have been deployed. Due to the fact that these nodes are frequently dispersed in a random manner and do not make use of integrated GPS units due to constraints regarding cost, size, and energy consumption, the precise coordinates of these nodes are typically unknown. Localization is a critical function in wireless sensor networks (WSNs) because, despite this, it is essential to have information of the positions of nodes in order to provide effective network operation, accurate data interpretation, routing optimization, performance evaluation of coverage, and event detection.

Detecting and accurately localizing the source of routing-based attacks in Wireless Sensor Networks (WSNs) is the primary objective of the proposed approach. This is accomplished through the utilization of a hybrid distance-vector-based methodology. This is due to the fact that traditional localization techniques continue to be vulnerable to threats such as wormhole, blackhole, and Sybil attacks. The framework that has been proposed provides a thorough picture of the system architecture and specifies the simulation design. It also highlights the important operational steps that are involved in ensuring secure localization and attack detection within the network.

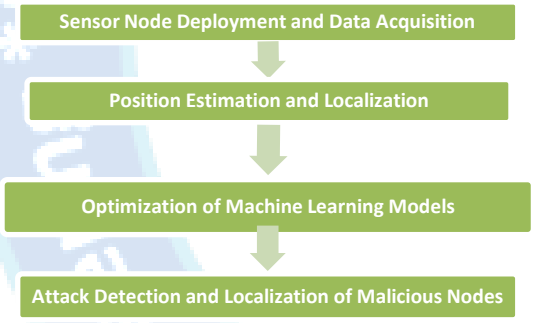

Figure 3. Workflow block diagram for proposed framework

The fundamental objective of this project is to design and build a framework that is safe, accurate, and efficient in terms of energy consumption for the purpose of using Wireless Sensor Networks for the identification of hostile nodes and for localization. For the purpose of achieving this, a number of particular goals have been established.

An anchor-assisted localization mechanism that is capable of reliably predicting the positions of randomly dispersed sensor nodes under practical environmental conditions is the first aim that needs to be accomplished. Detecting and localizing routing-based assaults, such as wormhole, Sybil, and blackhole attacks, is the second target. This will be accomplished by merging localization information with machine learning-based intrusion detection algorithms.

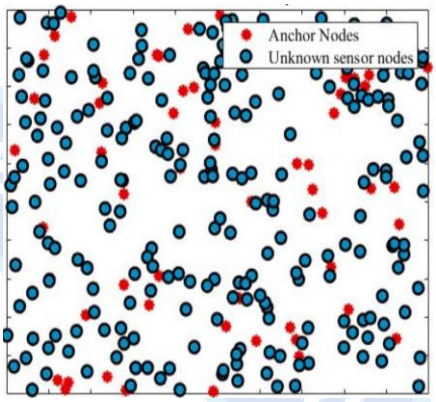

Figure 4. Node distribution map without an attack in wireless sensor networks

Performing extensive feature engineering and data preprocessing, which includes the application of sampling techniques, is another key purpose. The goal is to improve the quality and effectiveness of the dataset that is used for classification. In addition, the research is centered on optimizing hybrid machine learning classifiers by means of hyperparameter tuning and Bayesian optimization. The goal of this research is to enhance detection accuracy while simultaneously reducing the number of false positives. In conclusion, the framework's objective is to reduce the amount of energy consumption, communication overhead, and localization errors as much as possible, all while extending the lifetime of the network as a whole and preserving a high level of Quality of Service.

A robust and secure localization framework that is capable of reliably predicting the positions of randomly placed sensor nodes while maintaining resilience against routing-based assaults such as wormhole, Sybil, and blackhole attacks is the goal of this project. The goal is to develop a machine learning-based detection mechanism that is optimized for identifying malicious nodes. This will be accomplished by using feature extraction, clustering-based labeling, and hyperparameter optimization. This will ensure that the detection accuracy is high while few false positives are produced. By decreasing communication overhead, eliminating localization errors, and optimizing energy consumption through efficient data aggregation, intelligent routing, and secure localization procedures, the goal is to improve the overall efficiency and lifetime of the network.

2. Literature Review:

The literature reveals a strong trend toward integrating machine learning, energy-efficient protocols, and security mechanisms in WSNs and IoT systems. While significant progress has been made in anomaly detection, localization, and secure communication, challenges such as scalability, computational

overhead, and real-time adaptability still require further research.

(A) Review on Machine Learning Integration Schemes in WSN and IoT

Machine learning (ML) has emerged as a transformative tool in enhancing the intelligence, adaptability, and efficiency of wireless sensor networks (WSNs) and Internet of Things (IoT) systems. Several studies emphasize the role of ML in predictive maintenance, anomaly detection, and smart decision-making. For instance, Schwendemann et al. [1] presented a comprehensive survey on ML techniques for predictive maintenance in industrial systems, highlighting their effectiveness in fault detection and condition monitoring. Similarly, Gouda et al. [2] proposed a deep variational autoencoder (VAE)-based unsupervised approach for outlier detection in IoT environments, demonstrating improved detection accuracy without labeled data.

In the context of smart cities, Sharma et al. [10] reviewed ML applications in WSNs, identifying their role in optimizing resource utilization and enabling real-time analytics. Additionally, hybrid anomaly detection techniques combining clustering and ML models have been explored by Ahmad et al. [16], showing enhanced detection of irregular patterns in network behavior. These contributions collectively establish that ML significantly improves network intelligence, reduces human intervention, and enhances system robustness in dynamic environments.

(B) Review on Energy Efficiency and Secure Communication in WSNs

Energy efficiency and secure communication remain critical challenges in WSNs due to limited node resources and vulnerability to attacks. Masdari [3] proposed an energy-efficient clustering mechanism combined with congestion control using mobile sinks, which significantly prolongs network lifetime. Similarly, Sangeetha [4] introduced a heuristic path search algorithm to mitigate congestion, improving data transmission reliability.

Security is another vital concern addressed through lightweight cryptographic techniques. Khashan et al. [7] developed an automated lightweight encryption scheme that ensures secure communication while maintaining low energy consumption. Supporting this, Rana et al. [9] provided a detailed survey on lightweight cryptography for IoT systems, emphasizing its suitability for resource-constrained environments. Furthermore, Ahmad et al. [8] analyzed the combined effect of clustering and encryption on network lifetime, demonstrating a trade-off between security and energy efficiency.

Advancements in wireless communication technologies, particularly 5G, have also contributed to secure data transmission. Chen et al. [5] explored machine learning-based physical layer authentication for 5G systems, enhancing resistance against spoofing attacks. These studies highlight that integrating energy-aware protocols with lightweight security mechanisms is essential for sustainable and secure WSN operations.

(C) Review on Localization and Security Threat Detection in WSNs

Localization and security threat detection are crucial for maintaining the reliability and integrity of WSNs. Accurate node localization improves routing efficiency and network management. Robinson et al. [12] proposed a machine learning-based 3D localization algorithm, achieving higher accuracy compared to traditional methods. Similarly, Chen et al. [13] introduced the CWDV-Hop algorithm, which integrates distance weighting and optimization techniques to enhance localization precision.

Security threats such as Sybil and wormhole attacks pose significant risks to WSNs. Giri et al. [14] addressed secure localization under Sybil attacks using an information-theoretic approach, improving detection accuracy. Singh et al. [15] utilized artificial neural networks (ANNs) to detect wormhole attacks, demonstrating the effectiveness of ML in identifying malicious behavior. Additionally, Farjamnia et al. [18] improved the DV-Hop localization method to detect wormhole attacks, further strengthening network security.

Moreover, Hasan et al. [17] proposed an ANN-based secured node detection technique, which enhances the identification of compromised nodes. These studies collectively indicate that integrating ML techniques with localization algorithms significantly improves both positional accuracy and security resilience in WSNs.

3. Methodology:

The framework starts with the random deployment of sensor nodes across a specific sensing region, together with a restricted number of anchor nodes whose placements are predefined. This is followed by the deployment of certain anchor nodes. To estimate the locations of unknown nodes, anchor-assisted localization is utilized. This technique takes into account distance measurements as well as information regarding connectedness. A number of issues, including signal attenuation, interference, and noise, are taken into consideration during the design of this process so that it can perform efficiently in realistic environmental conditions. The base for secure communication and the efficient identification of malicious nodes is provided by accurate localization which serves as the foundation.

Following the phase of localization, the framework adds methods for data transmission and aggregation in order to collect data pertaining to network traffic and routing. Following this, techniques from the field of feature engineering are utilized in order to extract relevant characteristics that distinguish malicious activity from normal node behavior. In order to address the issue of class imbalance and enhance the generalization capabilities of the model, proper sampling procedures are utilized during the preprocessing step.

For the purpose of identifying malicious nodes, hybrid machine learning classification models are constructed and trained with the dataset that has been processed. The optimization of model parameters is accomplished through the utilization of hyperparameter tuning and Bayesian optimization approaches in order to improve performance. In addition, K-means clustering is utilized for cluster-based labeling and binary classification, which enables the separation of normal and

malignant nodes in an effective manner. The utilization of clustering in conjunction with supervised learning results in an increase in detection accuracy and a decrease in the number of false positives.

Following the identification of malicious nodes, the precise positions of these nodes are established by utilizing the computed localization data. Consequently, this makes it possible for the network to isolate or mitigate compromised nodes, which in turn prevents future disruptions to the routing process and improves the overall security of the network. In order to maximize detection accuracy and network lifetime, the suggested framework is meant to minimize localization error, reduce communication overhead, and optimize energy consumption. Additionally, it is designed to minimize energy consumption.

Comprehensive simulations employing benchmark intrusion detection datasets and performance indicators such as localization error, detection accuracy, energy consumption, and network lifetime are used to test the effectiveness of the proposed system. These simulations are used to validate the effectiveness of the system. In light of the findings, it appears that the framework that has been proposed has the potential to exceed the existing methods of attack detection and localization by offering a solution that is all-encompassing, secure, and efficient in terms of energy consumption for wireless sensor networks.

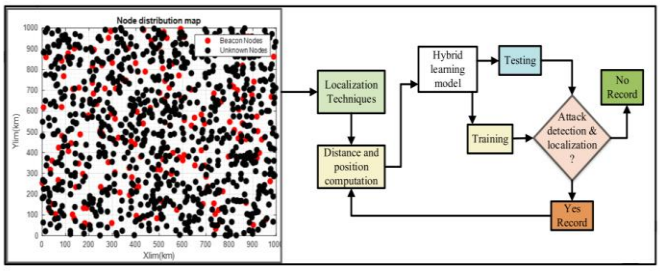

Figure 5. Proposed secured localization scheme block diagram to detect and localization of malicious node in WSNs.

4. Results

This section provides an overview of the simulation setup, as well as the assessment measures that were utilized in order to evaluate the effectiveness of the suggested framework. In order to assist the localization of unknown nodes, it is assumed that sensor nodes and the sink node would remain stable after deployment. Anchor nodes, on the other hand, are considered mobile. Each sensor node is given a one-of-a-kind identity, in addition to the positional and energy-related characteristics that are associated with it. In addition to being located outside of the clustering region, the sink node is accessible to all of the other nodes there. Within a sensing area measuring 1000 × 1000 m2, the network is comprised of nodes that have been randomly deployed. These nodes include 60 beacon nodes, 240 unknown nodes, and 35 malicious nodes across the network. There is a communication range of 250 meters between the beacon nodes and the sensor nodes also. The Distance Vector hop localization algorithm is utilized in order to ascertain the locations of Sybil nodes. In Table 1, a comprehensive listing of the simulation parameters. The network design and simulation are carried out using MATLAB R2019a on a 64-bit Windows PC that is outfitted with an Intel® Core i5 processor that operates at 2.20 GHz and 128 GB of random access memory (RAM). Through the utilization of benchmark datasets, signal processing toolboxes are utilized for the purpose of data analysis and performance evaluation [8].

The dataset known as CICIDS2017 is utilized for the purpose of attack classification. Metrics such as localization error (LE), average localization error (ALE), localization accuracy of unknown nodes, and confusion matrix analysis are utilized in order to assess the effectiveness of the framework that has been proposed. Additional metrics, such as the average localization error (ALE) and the average localization accuracy (ALA), are also taken into consideration in order to conduct a full evaluation of the effectiveness of the system.

Table 1: Parameters data table

Simulation parameters:	Parameters Values:
Software	MATLAB
Deployment	Random
Number of nodes	300
Simulation area	1000x1000 m^2
Protocol	Routing and clustering
Number of beacons	60
Transmission Radius	250 m
Unknown nodes	240
Model	Regular
Malicious nodes	35%

A number of important metrics, including accuracy, detection rate, precision, and recall, are utilized in order to assess the performance of the framework that has been proposed. In addition, the average localization error (ALE) is utilized in order to evaluate the performance of the localization process. To get the ALE, first add up all of the localization errors (LE) of all of the unknown nodes, and then divide that total by the total number of nodes that are unknown. In this context, the term "localization error" refers to the disparity between the estimated and actual positions of sensor nodes that contain unknown information.

Based on the findings of the simulation, it has been determined that anchor nodes have a higher level of connectedness and a greater number of surrounding nodes in comparison to ordinary sensor nodes. When compared to the standard model, the whole network demonstrates an average connectedness of 61, with each anchor node being connected to around three nearby nodes. In addition, an error map was produced in order to represent nodes inside the network whose coordinates could not be determined with precision. It is clear that the suggested method is highly robust, as evidenced by the fact that the average localization error obtained with sixty beacon nodes and

two hundred and forty unknown nodes is 0.1908. However, in comparison to the method that was described [19, 20], which achieved an average localization error of 0.37 by utilizing an iterative optimization-based localization strategy, the proposed approach achieves a substantially lower level of error. Based on these findings, it is clear that secure localization-based network planning and simulation has the potential to successfully detect rogue nodes by properly calculating their placements. The establishment of a robust and secure localization method is what allows this to be performed. It is also demonstrated by the results of the simulation that range-free localization approaches are able to estimate the positions of unknown sensor nodes in an effective manner. By analyzing the positions of nodes, beacon nodes are able to more effectively identify malicious or rogue nodes, which ultimately results in an improvement in the overall security, scalability, and reliability of wireless sensor networks.

Figure 6. Comparison of localization error

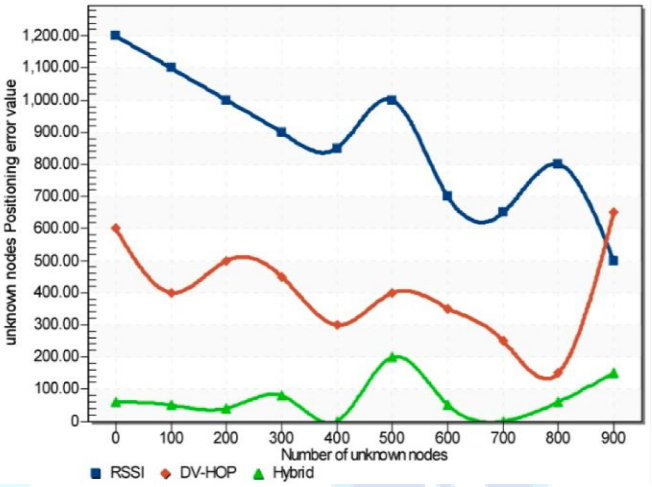

Figure 7. Error observed in finding unknown nodes

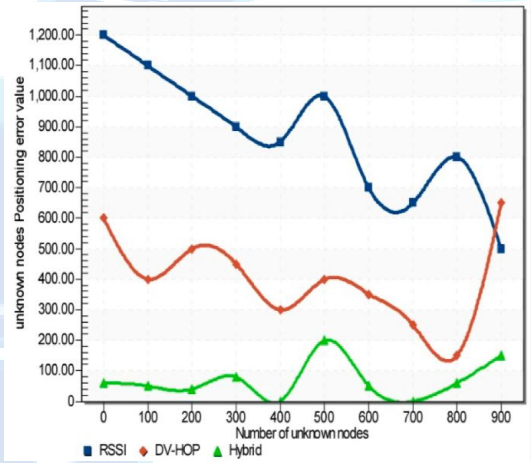

Table 2. Comparison of hybrid machine learning models using NSL-KDD dataset for attack detection

ML Classifiers	Performance evaluation metrics						
	Validation	Accuracy	Precision	Recall	F1-score	ROC	Running time
NB	83.94	83.71	90.35	83.71	85.68	99.90	0.025
DT	87.94	88.10	88.34	88.10	88.07	99.91	0.22
XGB	99.16	99.34	99.32	99.34	99.32	99.91	1.83
RF	98.66	99.46	99.44	99.46	99.45	99.93	0.17
DT-XGB	99.57	99.80	99.80	99.80	99.80	99.90	1.24
RF-XGB	99.57	99.80	99.80	99.80	99.80	99.85	1.44
RF-DT	99.57	99.79	99.79	99.79	99.79	99.90	0.327

5. Conclusion:

This study presents a comprehensive and secure framework for enhancing localization accuracy and detecting malicious nodes in Wireless Sensor Networks using hybrid machine learning techniques. By integrating anchor-based localization with optimized classification models, the proposed system effectively addresses critical challenges such as routing attacks, localization errors, and energy constraints. The use of clustering-based labeling, feature engineering, and hyperparameter optimization significantly improves detection performance while reducing false positives.Simulation results confirm that the proposed approach outperforms existing methods in terms of detection accuracy, localization precision, and network efficiency. The reduction in average localization error and improved identification of malicious nodes demonstrate the robustness of the system under realistic network conditions. Furthermore, the framework ensures efficient resource utilization by minimizing communication overhead and energy consumption, thereby extending network

lifetime. Despite these advancements, challenges such as scalability in large-scale deployments and real-time implementation remain open for future research. Further work can explore deep learning-based adaptive models and real-time deployment strategies to enhance system responsiveness and applicability in dynamic IoT environments.

References

[1] S. Schwendemann, Z. Amjad, and A. Sikora, ‘‘A survey of machinelearning techniques for condition monitoring and predictive maintenance of bearings in grinding machines,’’ Comput. Ind., vol. 125, Feb. 2021, Art. no. 103380.
[2] W. Gouda, S. Tahir, S. Alanazi, M. Almufareh, and G. Alwakid, ‘‘Unsupervised outlier detection in IoT using deep VAE,’’ Sensors, vol. 22, no. 17, p. 6617, Sep. 2022.
[3] M. Masdari, ‘‘Energy efficient clustering and congestion control in WSNs with mobile sinks,’’ Wireless Pers. Commun., vol. 111, no. 1, pp. 611–642, Mar. 2020.
[4] G. Sangeetha, ‘‘A heuristic path search for congestion control in WSN,’’ in Industry Interactive Innovations in Science, Engineering and Technology. Cham, Switzerland: Springer, 2018.
[5] S. Chen, H. Wen, J. Wu, J. Chen, W. Liu, L. Hu, and Y. Chen, ‘‘Physicallayer channel authentication for 5G via machine learning algorithm,’’Wireless Commun. Mobile Comput., vol. 2018, pp. 1–10, Oct. 2018.
[6] R. I. Bhopal, ‘‘Evolution of fifth generation technology in wireless communication,’’ 2023.
[7] O. A. Khashan, R. Ahmad, and N. M. Khafajah, ‘‘An automated lightweight encryption scheme for secure and energy-efficient communication in wireless sensor networks,’’ Ad Hoc Netw., vol. 115, Apr. 2021, Art. no. 102448.
[8] R. Ahmad, E. A. Sundararajan, and T. Abu-Ain, ‘‘Analysis the effect of clustering and lightweight encryption approaches on WSNs lifetime,’’ in Proc. Int. Conf. Electr. Eng. Informat. (ICEEI), Oct. 2021, pp. 1–6.
[9] M. Rana, Q. Mamun, and R. Islam, ‘‘Lightweight cryptography in IoT networks: A survey,’’ Future Gener. Comput. Syst., vol. 129, pp. 77–89, Apr. 2022.
[10] H. Sharma, A. Haque, and F. Blaabjerg, ‘‘Machine learning in wireless sensor networks for smart cities: A survey,’’ Electronics, vol. 10, no. 9, p. 1012, Apr. 2021.
[11] R.-F. Liao, H. Wen, J. Wu, F. Pan, A. Xu, Y. Jiang, F. Xie, and M. Cao, ‘‘Deep-learning-based physical layer
[12] Y. H. Robinson, S. Vimal, E. G. Julie, K. Lakshmi Narayanan, and S. Rho, “Three-dimensional manifold and machine learning-based localization algorithm for wireless sensor networks,” Wireless Personal Communications, vol. 127, no. 1, pp. 523–541, 2022.
[13] J. Chen, W. Zhang, Z. Liu, R. Wang, and S. Zhang, “CWDV-Hop: A hybrid localization algorithm with distance-weight DV-Hop and CSO for wireless sensor networks,” IEEE Access, vol. 9, pp. 380–399, 2021, doi: 10.1109/ACCESS.2020.3045555.
[14] A. Giri, S. Dutta, and S. Neogy, “An information-theoretic approach for secure localization against Sybil attack in wireless sensor networks,” Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 10, pp. 9491–9497, 2021.
[15] M. M. Singh, N. Dutta, T. R. Singh, and U. Nandi, “A technique to detect wormhole attacks in wireless sensor networks using artificial neural networks,” in Proceedings of the Springer Conference, Singapore, 2021.
[16] B. Ahmad, W. Jian, Z. A. Ali, S. Tanvir, and M. S. A. Khan, “Hybrid anomaly detection using clustering for wireless sensor networks,” Wireless Personal Communications, vol. 106, no. 4, pp. 1841–1853, 2019, doi: 10.1007/s11277-018-5721-6.
[17] B. Hasan, S. Alani, and M. A. Saad, “Secured node detection technique based on artificial neural networks for wireless sensor networks,” International Journal of Electrical and Computer Engineering, vol. 11, no. 1, pp. 536–544, 2021, doi: 10.11591/ijece.v11i1.pp536-544.
[18] G. Farjamnia, Y. Gasimov, and C. Kazimov, “An improved DV-Hop localization method for detecting wormhole attacks in wireless sensor networks,” Journal of Sensors, vol. 9, no. 1, pp. 1–24, 2020.
[17] Sahay, S., Anurag Banoudha, and Raghawendra Sharma. "On the use of ANFIS for Ground Water Level Forecasting in an Alluvium Area." International Journal of Research and Development in Applied Science and Engineering 2, no. 1 (2012): 1-7.
[18] Sahay, S., Anurag Banoudha, and Raghawendra Sharma. "Comparative study of soft computing techniques for ground water level forecasting in a hard rock area." International Journal of Research and Development in Applied Science and Engineering 4, no. 1 (2013): 1-6.