



e-ISSN: 2278-8875

p-ISSN: 2320-3765

International Journal of Advanced Research

in Electrical, Electronics and Instrumentation Engineering

Volume 15, Issue 7, July 2026



Impact Factor: 9.867

☎ 9940 572 462

☎ 6381 907 438

✉ ijareeie@gmail.com

@ www.ijareeie.com



A Machine Learning Framework with ADASYN Tomek Balancing and PSO for Intrusion Detection in Wireless Sensor Network

S.A. Arunmozhi, S. Sheela Angel

Associate Professor, Dept. of ECE, Saranathan College of Engineering College, Trichy, Tamil Nadu, India

PG Student [Communication Systems], Dept. of ECE, Saranathan College of Engineering College, Trichy, Tamil Nadu, India

ABSTRACT: A Wireless Sensor Network (WSN) consists of numerous low power nodes used to monitor, collect, and transfer data across the network. However, they are vulnerable to various attacks such as TDMA (Time Division Multiple Access), blackhole, flooding, and grayhole. This work presents an intelligent machine learning based Intrusion Detection System (IDS) using a Particle Swarm Optimization (PSO) Light Gradient Boosting Machine (LGBM) for efficient and accurate attack detection. To improve the performance, data preprocessing techniques such as feature selection using embedded LightGBM importance and data balancing through a modified ADASYN Tomek approach are employed in our proposed work. Class imbalance and preventing bias towards majority classes are handled in this method. The proposed IDS are evaluated on the WSN-DS dataset, achieving 99.8% accuracy with high precision, recall, and F1 score. A real time Dash based web dashboard is also developed for live attack detection and visualization. The results confirm that the PSO optimized LightGBM model offers a lightweight and highly accurate solution for real-time WSN intrusion detection.

KEYWORDS: WSN, Machine Learning, IDS, LightGBM, PSO, modified ADASYN, Tomek

I. INTRODUCTION

WSNs are widely utilized in applications such as environmental monitoring, healthcare systems, military surveillance, and industrial automation, making them extremely important. Despite their benefits, WSNs are very susceptible to security risks due to their open wireless communication channel, decentralized design, and severe resource restrictions. Because of these vulnerabilities, WSNs are a desirable target for a wide range of attacks, such as denial of service, blackhole, grayhole, flooding, and scheduling assaults, all of which have the potential to significantly reduce network performance and data reliability.

IDS are essential in protecting WSNs, traditional IDS strategies based on predetermined rules and signatures are inadequate to deal with the changing and dynamic nature of modern attacks because they are inflexible and need a lot of manual feature engineering [1]. To overcome these restrictions, machine learning and deep learning approaches have been extensively used to automatically learn sophisticated attack patterns and increase detection accuracy [5], [10].

The imbalanced nature of intrusion datasets in WSNs, where normal traffic occurrences exceed attack samples by a large margin, is a significant issue for IDSs that use ML; with such asymmetry classifiers are frequently biased and unable to properly identify minority attack categories. To address this problem and improve classification accuracy, a number of studies have used data resampling methods, including SMOTE, SMOTE Tomek, and ADASYN [2], [7]. Optimal feature selection and hyperparameter tuning are also necessary for data imbalance to attain high detection accuracy while keeping computational overhead low, which is ideal for resource constrained WSN environments. In order to optimize IDS models and enhance convergence speed and robustness, metaheuristic and swarm intelligence based optimization methods have been successfully used [4].

This paper presents an intelligent intrusion detection system for WSN that combines the PSO-optimized LightGBM classifier with the ADASYN Tomek data balancing technique. The objective of this method is to solve class imbalance, automatically optimize model parameters, and attain high detection accuracy at a low computational cost. Using the WSN-DS dataset, the model performs effectively and proving that it is capable of real time intrusion detection in WSN.



II. RELATED WORK

WSNs have become a significant technology for monitoring and collecting data in various civilian and industrial applications. However, the open nature of communication and the limited resources of WSNs render them vulnerable to numerous security threats and intrusions. Boahen et al. introduced a comprehensive intrusion detection system that combines unsupervised feature learning and deep neural networks to minimize human input and enhance detection accuracy on standard datasets [1]. Babu et al. put forth a hybrid metaheuristic feature selection method integrated with a modified ADASYN technique to tackle class imbalance and boost intrusion detection effectiveness in wireless sensor networks [2]. Almomani et al. presented the WSN-DS dataset, which is specifically tailored for intrusion detection systems within WSNs, facilitating through assessment of DoS attacks [3]. Arunmozhi et al. suggested a swarm intelligence based method for routing and attack detection aimed at recognizing black hole attacks while optimizing energy efficiency and network performance [4]. Salmi and Oughdir analyzed several deep learning based models for intrusion detection in preventing DoS attacks in WSN and showed enhanced performance compared to conventional machine learning techniques [5]. Birahim et al. introduced a PSO based explainable ensemble machine learning framework utilizing SMOTE Tomek balancing to improve the detection accuracy and interpretability with SHAP and LIME methods [6]. Talukder et al. created an ML based intrusion detection framework using SOMTE Tomek resampling to address data imbalance and attain high accuracy in various WSN intrusion scenarios [7]. Elsadig put forward a lightweight machine learning based IDS employing decision trees and Gini feature selection to identify DOS attacks in WSNs with minimal processing overhead [8]. Mittal et al. concentrated on energy efficiency and anomaly detection in hybrid WSNs by merging neural networks with routing frameworks and SVM driven intrusion detection techniques [9]. Sadia et al. devised a machine learning based IDS for Wi-Fi enabled WSNs using CNN models and feature reduction strategies to enhance detection accuracy and decrease false alarm frequencies [10]. Park and colleagues presented a graph based approach for network intrusion detection and classification that boosts attack detection precision while providing clarity and resilience against attack variations [11]. Behiry and Aly proposed a hybrid approach to feature reduction integrating PCA, SVD, and deep learning methods to improve accuracy in intrusion detection across various benchmark datasets [12]. Tan et al. tackled class imbalance in WSN intrusion detection by implementing SMOTE with Random Forest classifiers, showing enhanced detection accuracy following the oversampling process [13]. Liu et al. recommended a multitasking deep learning framework that combines anomaly detection, clustering, and classification to effectively manage imbalanced intrusion datasets [14]. Yadav and Kumar utilized recurrent neural networks to identify sequential attack patterns in WSNs, which led to improved accuracy for real time intrusion detection [15]. Zamry et al. proposed an efficient anomaly detection method using incremental PCA and one class SVM to lessen computational demands while preserving high detection accuracy [16]. Kanharo et al. developed a deep learning framework aimed at real time intrusion detection in IOT enables networks, achieving high detection rates across several benchmark datasets [17]. Shirazi et al. introduced a data synthesis method based on an adversarial auto encoder to bolster machine learning resilience and accuracy in detecting evolving cyber threats [18]. Venkataramani et al. suggested a secure routing mechanism for detecting black hole attacks to enhance packet delivery and diminish delays in mobile ad hoc networks [19]. Osanaiye et al. revealed a multi filter feature selection strategy for cluster based WSNs that lowers feature dimensionality while enhancing both intrusion detection precision and energy efficiency [20].

This work presents an intrusion detection system for WSN that combines a particle swarm optimization tuned LightGBM classifier with ADASYN Tomek balancing. The modified ADASYN method prioritizes challenging minority samples, while Tomek Links eliminates overlapping instances, resulting in an unbiased training dataset. The experimental result shows the model provides high detection accuracy with minimal computational cost, making it ideal for real time WSN security applications.

III. PROPOSED WORK

This work proposes an intelligent intrusion detection framework for WSN that integrates data balancing, optimization, and machine learning to attain high detection accuracy despite resource limitations. Fig. 1 shows the overall architecture of the proposed system.

1) Dataset Description

The proposed system uses the WSN-DS dataset, which has 19 features that describe network, routing, and energy related aspects, and it was generated using WSN implemented based on the LEACH protocol and runs in the NS-2 environment. The dataset includes various attack scenarios to replicate realistic intrusion behaviors in WSN, like blackhole attacks, where malicious nodes discard all incoming packets rather than forwarding them; grayhole attacks,



which consist of selective packet dropping to avoid detection; TDMA attacks, where nodes transmit beyond their designated timeslots, resulting in channel disruption; and flooding attacks, where excessive generation of packets causes network congestion.

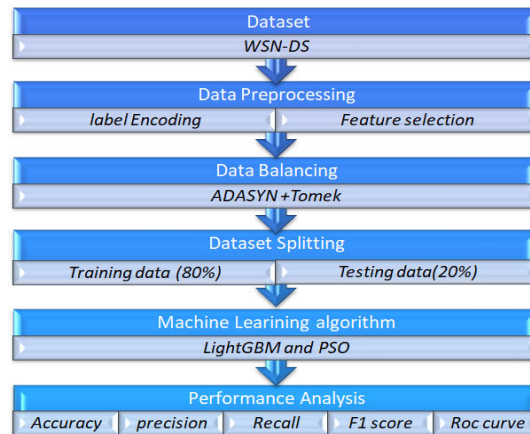


Fig. 1 Proposed System Flow Diagram

2) Data Preprocessing and Feature Selection

Data preprocessing is essential for preparing the dataset for effective training and evaluation of models. In the dataset, categorical labels are transformed into numerical values through a label encoding method. Effective preprocessing enhances learning efficiency, minimizes bias, and boosts detection accuracy. Furthermore, feature selection is conducted using LightGBM's built-in feature importance mechanism based on the cumulative information gain across all decision tree splits. This approach reduces dimensionality and decreases computational complexity.

3) Data Balancing

The ADASYN Tomek data balancing technique is utilized to solve the class imbalance in datasets. The ADASYN method creates synthetic samples for minority attack classes through a distance weighted approach, and the weight of each minority sample is calculated in equation (1).

$$r_i = \frac{\text{Weighted sum of majority neighbors}}{\text{Weighted sum of all neighbors}} \quad (1)$$

The neighbor weights are calculated using

$$\text{weight} = \frac{1}{\text{distance} + \epsilon} \quad (2)$$

In equation (2), a small constant ϵ is added to prevent division by zero. This weighting system places a greater emphasis on nearby neighbors, facilitating the creation of higher quality synthetic samples in overlapping and complex areas and enhancing the learning of uncommon attack patterns. The Tomek links method is used to eliminate borderline and noisy samples after oversampling. If two samples from distinct classes are mutual nearest neighbors, this indicates an ambiguous region. Eliminating such pairs improves class separability and prevents overlap between synthetic minority samples and majority class data. As a result, the dataset achieved a nearly uniform class distribution after the application of modified ADASYN and Tomek Links, leading to balanced learning and enhanced intrusion detection capabilities.

The dataset used in this work is divided into 80% for training the model to properly recognize normal and assault traffic patterns and 20% for testing on unseen data.

4) Machine Learning Algorithm

The proposed intrusion detection system utilizes a hybrid machine learning framework that combines LightGBM with PSO. This machine learning algorithm aims to attain high classification accuracy while ensuring computational



efficiency that is appropriate for resource limited WSNs. LightGBM is responsible for traffic classification, and PSO fine tunes its hyperparameters to improve detection performance.

i. Light Gradient Boosting Machine (LightGBM) Algorithm

LightGBM represents a sophisticated ensemble learning algorithm that is founded on gradient boosting decision trees. In contrast to conventional boosting techniques, LightGBM employs a histogram-based learning approach, which transforms continuous feature values into discrete bins, thus minimizing memory consumption and enhancing training speed. Furthermore, it implements a leaf-wise tree growth strategy with depth limitations, allowing the model to decrease loss more effectively and attain greater accuracy. The model progressively builds decision trees, with each subsequent tree rectifying the mistakes made by the preceding ensemble. The prediction is derived by consolidating the outputs from all trees, which guarantees enhanced robustness and diminished variance.

ii. Particle Swarm Optimization (PSO) Algorithm

In the context of PSO, each particle signifies a potential solution within the search space. In this work, each particle is associated with a distinct set of hyperparameters for LightGBM, which include the learning rate, the number of estimators, the maximum depth, and the number of leaves. The fitness function is established as the classification accuracy, which is equivalent to minimizing the error rate. Each particle modifies its velocity and position according to its personal best solution (pbest) and global best solution (gbest). Through a series of iterative updates, PSO effectively explores the optimal configuration of hyperparameters that maximizes detection accuracy while simultaneously reducing computational costs. This process negates the need for manual tuning and improves the speed of convergence.

IV. RESULT AND DISCUSSION

The experimental results of the proposed intrusion detection system employing the WSN-DS dataset are presented and analyzed in this section. Hyperparameter optimization of the LightGBM classifier using PSO was carried out following data preprocessing and class balancing utilizing the modified ADASYN Tomek method. The evaluation of the proposed model is assessed using performance metrics.

Fig. 2 shows the original distribution of attack classes in the dataset, where the left chart represents the categorical attack labels, and the right chart shows their corresponding numeric encoding after label encoding. The visualization clearly indicates that the dataset is highly imbalanced; therefore, a data balancing technique is required to ensure fair learning across all classes and improve the detection rate of minority attacks.

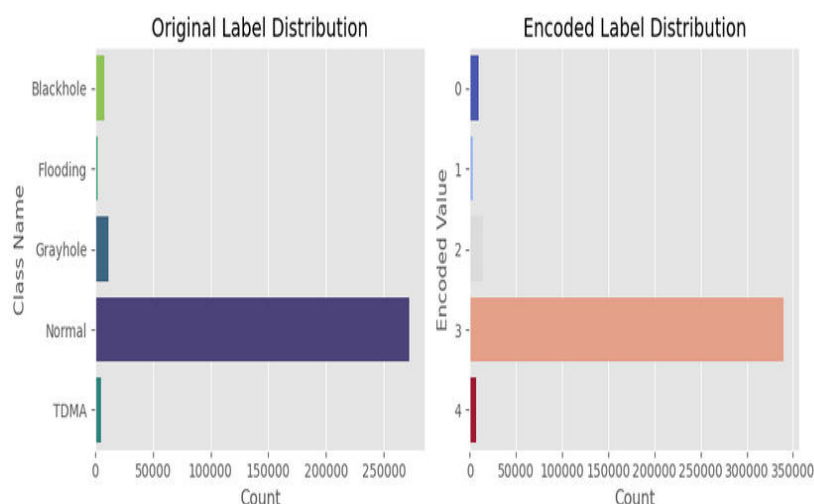


Fig. 2 Original and Encoded Label Distributions in the Dataset

In Fig. 3, the final class distribution obtained after applying the modified ADASYN Tomek Links technique is presented. The bar chart clearly demonstrates that all five attack classes now contain nearly equal numbers of samples. This indicates that the modified ADASYN method successfully generated synthetic samples for the minority classes using a distance weighted interpolation formula in equation (1), while Tomek Links effectively removed borderline and overlapping instances, ensuring cleaner class boundaries. The resulting balanced dataset eliminates the dominance of



the Normal class observed in the original distribution, enabling the model to learn from all attack categories with equal representation.

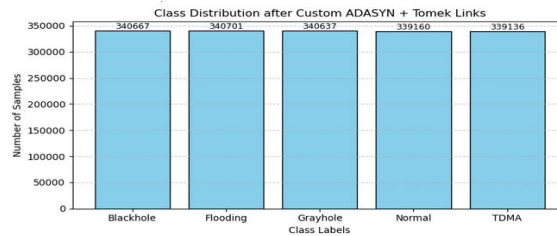


Fig. 3 Class distribution using Modified ADASYN Tomek

Performance Analysis

The performance metrics are used to evaluate the effectiveness of the proposed IDS in distinguishing normal and malicious in WSNs. These metrics are derived from four basic parameters: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

Accuracy: It shows how well the intrusion detection model works overall. It finds the correct predictions out of all samples. A high score means the system spots a normal and attacks well in WSN.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Precision: It checks the share of true attacks among all alerts the model raises. A high value means few false alarms. This saves energy and stops wrong actions in WSN.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Recall (Sensitivity): It tracks how well the model finds real attacks. It is the true positive rate. This boosts network safety and trust.

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

F1 Score: It blends precision and recall as a balanced average. It works best with uneven datasets. This keeps the model fair in spotting attacks.

$$F1_Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (6)$$

Confusion Matrix: It lays out all results in detail and lists true positives, true negatives, false positives, and false negatives.

The overall classification accuracy attained by the proposed PSO optimized LightGBM model is shown in Fig. 4. Fig. 4(a) shows the classification report for the LightGBM model, which attained an accuracy of 98.9%. In contrast, Fig. 4(b) shows the classification report of the PSO optimized LightGBM model, which achieved a significantly improved accuracy of 99.8% with perfect precision, recall, and F1 scores across all classes.

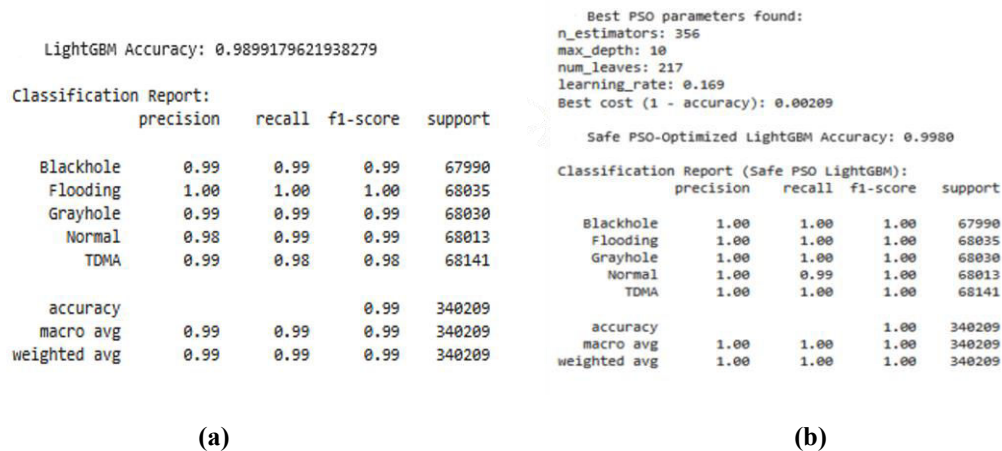


Fig. 4 Classification report, (a) LightGBM, (b) PSO optimized LightGBM

Fig. 5 illustrates the confusion matrices for both the baseline LightGBM and a PSO optimized LightGBM model. As depicted in Fig. 5(a), the baseline LightGBM model demonstrates strong classification performance; however, there are slight misclassifications noted among the blackhole, grayhole, and TDMA attack categories. Conversely, Fig. 5(b) shows that the PSO optimized LightGBM models significantly minimize misclassification across all categories, achieving nearly flawless classification with very few false positives and negatives. This enhances the effectiveness of PSO-optimized hyperparameter tuning for improving class separability and the overall reliability.

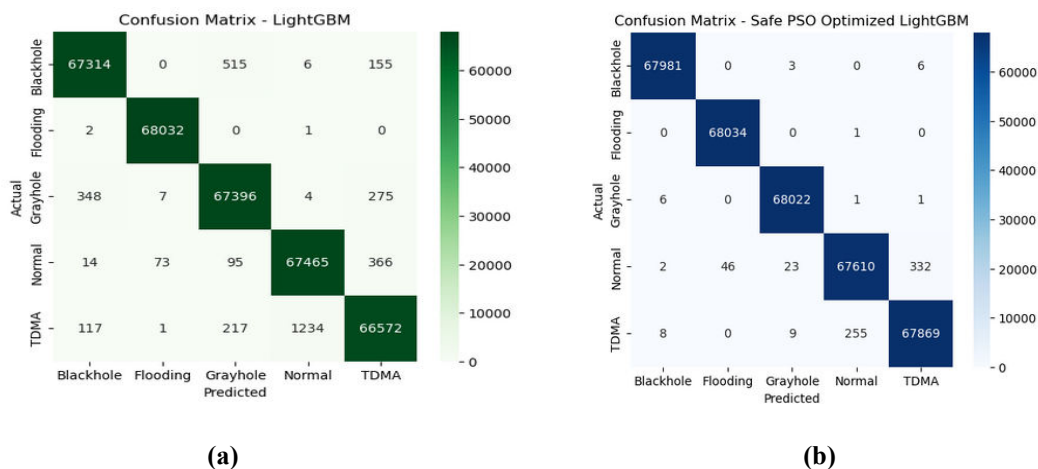


Fig. 5 Confusion Matrix: (a) LightGBM, (b) PSO-optimized LightGBM

Table 1 Comparison of IDS Method

References	Method	Accuracy	Training time and complexity
[5]	DNN	97.07%	High
[10]	CNN	97%	High
[13]	Random Forest	93.5%	Medium
[16]	SVM	98%	Medium
Proposed Work	PSO optimized LGBM	99.8%	Low



The performance evaluation of the proposed method against other intrusion detection techniques is illustrated in Table 1. It is evident that the proposed method attains superior accuracy compared to conventional machine learning and deep learning methods. Furthermore, the model exhibits decreased training duration and computational complexity, attributed to its effective histogram based learning and optimized hyperparameter tuning through PSO. Although deep learning models have competitive results, they necessitate considerably high computational resources. Conversely, the proposed method provides a more favorable balance between accuracy, efficiency, and appropriateness for real time intrusion detection in a WSN environment.

4.2 Web Application Results

The effectiveness of the proposed PSO optimized LightGBM intrusion detection system is further validated through a Dash based web application. The output screens of the developed web interface, which enables users to input WSN sensor feature values and obtain real time intrusion detection results. As shown in Fig. 6, the system accurately identifies normal network traffic when suitable feature values of legitimate activity are provided. The interface indicates the predicted class as normal along with a confidence score, validating the model's ability to distinguish between traffic patterns and malicious actions. Fig. 7 shows the detection of a blackhole attack.

In each instance, the system accurately identifies the correct attack type and reports the corresponding confidence level, demonstrating the reliability and effectiveness of the proposed model. These observed outputs confirm the robustness and reliability of the proposed model under real time operating conditions and suitable for practical deployment in WSN security monitoring applications.

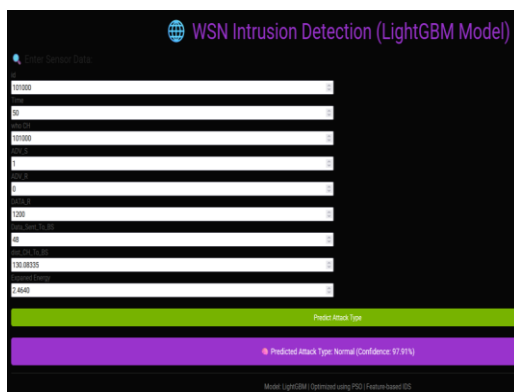


Fig. 6 Normal traffic detected in web app

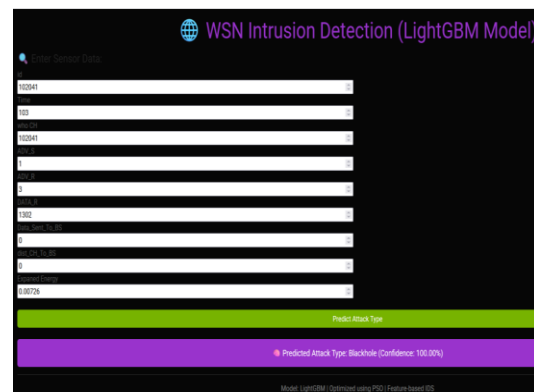


Fig. 7 Blackhole attack detected in web app

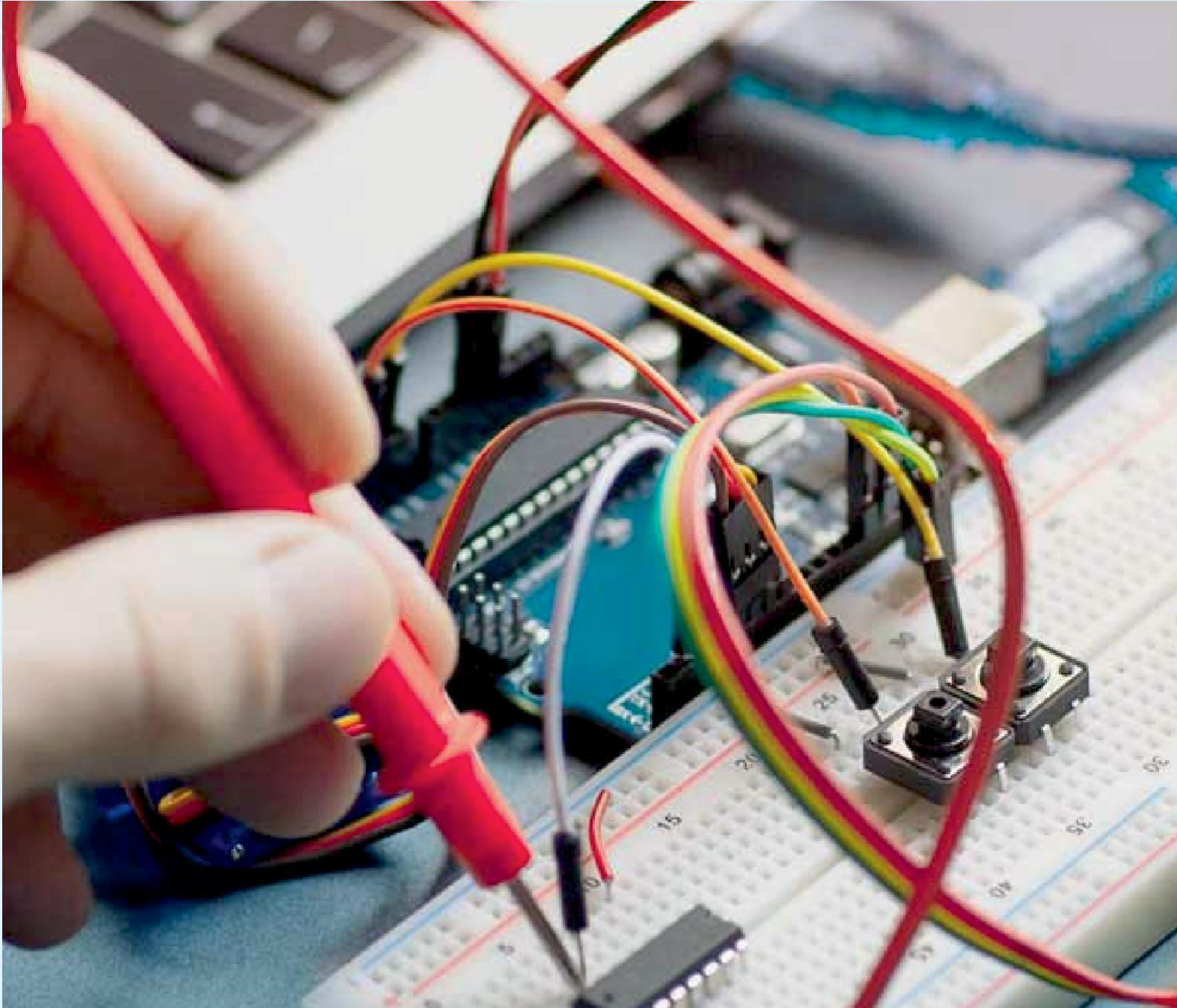
V. CONCLUSION

This paper presented a lightweight and effective intrusion detection system for WSNs that employs a modified ADASYN Tomek data balancing and a PSO-optimized LightGBM classifier approach. The proposed balancing strategy effectively handled class imbalance and noise, resulting in better identification of minority attacks and reduced model bias. A LightGBM classifier optimized using PSO is employed to enhance attack pattern learning and improve overall detection reliability. Experimental results demonstrate a detection accuracy of 99.8%, with near-perfect precision and recall for all attacks. Real time applicability was further confirmed by interactive and confidence based intrusion prediction using a web application built using the DASH framework. The proposed system offers a dependable and energy-efficient security solution for WSNs. Future work includes extending the framework towards the direction of IoT and edge-based deployment as well as integrating explainable intrusion detection.



REFERENCES

- Boahen, E. K., Frimpong, S. A., Ujakpa, M. M., Sosu, R. N. A., Larbi-Siaw, O., Owusu, E., Appati, J. K., & Acheampong, E. (2022). A deep multi-architectural approach for online social network intrusion detection system. In Proceedings of the IEEE World Conference on Applied Intelligence and Computing (AIC) (pp. 919–924). <https://ieeexplore.ieee.org/document/9848865>
- Babu, K. S. (2025). Optimizing Feature Selection in Imbalanced Intrusion Detection System Using Hybrid Metaheuristics Algorithms for Wireless Sensor Networks. *Cluster Computing*, 28, Article 575. <https://doi.org/10.1007/s10586-025-05248-6>
- Almomani, I., Al-Kasasbeh, B., & Al-Akhras, M. (2016). WSN-DS: A Dataset for intrusion detection systems in wireless sensor networks. *Journal of Sensors*, 2016, Article 4731953. <https://doi.org/10.1155/2016/4731953>
- Arunmozhi, S. A., Rajeswari, S., & Venkataramani, Y. (2023). Swarm Intelligence Based Routing with Black Hole Attack Detection in MANET. *Computer Systems Science & Engineering*, 44(2), 183–194. <https://doi.org/10.32604/csse.2023.024340>
- Salmi, S., & Oughdir, L. (2023). Performance evaluation of deep learning techniques for DoS attacks detection in wireless sensor network. *Journal of Big Data*, 10(1), 17. <https://doi.org/10.1186/s40537-023-00692-w>
- Birahim, S. A., Paul, A., Rahman, F., Islam, Y., Roy, T., Hasan, M. A., Haque, F., & Chowdhury, M. E. H. (2025). Intrusion Detection for Wireless Sensor Network Using Particle Swarm Optimization Based Explainable Ensemble Machine Learning Approach. *IEEE Access*, 13, 13142–13158. <https://doi.org/10.1109/ACCESS.2025.3528341>
- Talukder, M. A., Sharmin, S., Uddin, M. A., Islam, M. M., & Aryal, S. (2024). MLSTL-WSN: Machine Learning Based Intrusion Detection Using SMOTE Tomek in WSNs. *International Journal of Information Security*, 23(3), 2139–2158. <https://doi.org/10.1007/s10207-024-00833-z>
- Elsadig, M. A. (2023). Detection of denial-of-service attack in wireless Sensor networks: A lightweight machine learning approach. *IEEE Access*, 11, 83537–83552. <https://doi.org/10.1109/ACCESS.2023.3303113>
- Mittal, M., de Prado, R. P., Kawai, Y., Nakajima, S., & Muñoz Expósito, J. E. (2021). Machine learning techniques for energy efficiency and anomaly detection in hybrid wireless sensor networks. *Energies*, 14(11), 3125. <https://doi.org/10.3390/en14113125>
- Sadia, H., Farhan, S., Haq, Y. U., Sana, R., Mahmood, T., Bahaj, S. A. O., & Rehman, A. (2024). Intrusion detection system for Wireless Sensor Networks: A Machine Learning Based Approach. *IEEE Access*, 12, 52565–52582. <https://doi.org/10.1109/ACCESS.2024.3380014>
- Park, S. B., Jo, H. J., & Lee, D. H. (2023). G-IDCS: Graph-based Intrusion detection and classification system for CAN protocol. *IEEE Access*, 11, 39213–39227. <https://doi.org/10.1109/ACCESS.2023.3268519>
- Behiry M. H., & Aly, M. (2024). Cyberattack detection in wireless sensor networks using a hybrid feature reduction technique with AI and machine learning methods. *Journal of Big Data*, 11(1). <https://doi.org/10.1186/s40537-023-00870-w>
- Tan, X., Su, S., Huang, Z., Guo, X., Zuo, Z., Sun, X., & Li, L. (2019). Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm. *Sensors*, 19(1), Article 203. <https://doi.org/10.3390/s19010203>
- Liu, Q., Wang, D., Jia, Y., Luo, S., & Wang, C. (2022). A multi-task based deep learning approach for intrusion detection. *Knowledge Based Systems*, 238, Article 107852. <https://doi.org/10.1016/j.knosys.2021.107852>
- Yadav, A., & Kumar, A. (2022). Intrusion detection and prevention using RNN in WSN. In S. Smys, V. E. Balas, & R. Palanisamy (Eds.), *Inventive computation and information technologies* (Vol. 336). Springer. https://doi.org/10.1007/978-981-16-6723-7_40
- Zamry, N. M., Zainal, A., Rassam, M. A., Alkhamash, E. H., Ghaleb, F. A., & Saeed, F. (2021). Lightweight anomaly detection scheme using incremental principal component analysis and support vector machine. *Sensors*, 21(23), 8017. <https://doi.org/10.3390/s21238017>
- Kandhro, I. A., Alanazi, S. M., Ali, F., Kehar, A., Fatima, K., Uddin, M., & Karuppayah, S. (2023). Detection of real-time malicious intrusions and attacks in IoT empowered cybersecurity infrastructures. *IEEE Access*, 11, 9136–9148. <https://doi.org/10.1109/ACCESS.2023.3238664>
- Shirazi, H., Muramudalige, S. R., Ray, I., Jayasumana, A. P., & Wang, H. (2023). Adversarial auto-encoder data synthesis for enhancing machine learning based phishing detection algorithms. *IEEE Transactions on Services Computing*, 1–13. <https://doi.org/10.1109/TSC.2023.3234806>
- Arunmozhi, S. A., & Venkataramani, Y. (2012). Black Hole Attack Detection and Performance Improvement in Mobile Ad-Hoc Network. *Information Security Journal: A Global Perspective*, 21(3), 145–153. <https://doi.org/10.1080/19393555.2011.652291>
- Osaniye, O., Ogundile, O., Aina, F., & Periola, A. (2019). Feature Selection for Intrusion Detection System in a Cluster Based Heterogeneous Wireless Sensor Network. *Facta Universitatis, Series: Electronics and Energetics*, 32(3), 315–330.



INNO  SPACE
SJIF Scientific Journal Impact Factor



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



International Journal of Advanced Research

in Electrical, Electronics and Instrumentation Engineering

 9940 572 462  6381 907 438  ijareeie@gmail.com



www.ijareeie.com

Scan to save the contact details