International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

AN ENHANCED INTRUSION DETECTION FRAMEWORK USING SMOTETOMEK AND RANDOM FOREST FOR WIRELESS SENSOR NETWORKS

Bikash Kalita[1], Satyajit Sarmah[2*], Deepjyoti Kalita[3], Surajit Medhi[4], Manas Pratim Kalita[5]

[1,2*,5]Gauhati University, Guwahati, Assam, India

EMail:[1] bikax99@gmail.com,[2] ss@gauhati.ac.in, [5] manaspratimkalita4@gmail.com

[3]Mangaldai College, Mangaldai, Assam, India

E-Mail: [3] deepjyoti111@gmail.com

[4]B. Borooah College, Guwahati, Assam, India

E-Mail: [4] surajitmdh@gmail.com

Abstract

The widespread use of Wireless Sensor Networks has become an unavoidable necessity in many critical situations. Simultaneously, making them prime targets for network attacks such as Grayhole, Blackhole, TDMA, and Flooding attacks. Traditional Intrusion Detection Systems often struggle with different problems, such as class imbalance, that lead to poor attack detection accuracy, especially for minority class attacks. To address this issue, this study proposes an advanced network threat detection system framework integrating SMOTETomek with Random Forest for efficient performance in the analysis and classification of network threats. The study uses the WSN-DS dataset for training and evaluation of the methods. The SMOTETomek technique is employed to mitigate class imbalance by oversampling minority attack instances and removing noisy overlapping samples, ensuring a more balanced and representative dataset. The Random Forest technique, which is more familiar for its ensemble learning capabilities, is used for classification. The experimental results demonstrate a high performance, where the overall accuracy is 99.38%, with per-class F1-scores exceeding 97.62%, outperforming many existing IDS techniques. Therefore, the proposed framework can be a more advanced and effective solution in enhancing cybersecurity in WSNs.

Keywords: WSN; Network Threats; Machine Learning; SMOTETomek; Cybersecurity; Ensemble Learning.

INTRODUCTION

Wireless Sensor Networks (WSNs) are a very crucial application in various domains, including smart cities, healthcare monitoring, industrial automation, and military applications (Majid et

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

al., 2022). For the distributed and resource-constrained architecture of WSN, it is highly susceptible to security breaches(Keerthika & Shanmugapriya, 2021), such as Grayhole, Blackhole, TDMA, and Flooding attacks, which can disrupt network integrity and performance of the system. Analysis and detection of such network threats is a significant challenge for many network threat detection systems due to class imbalance in real-world network data (Le et al., 2024), where certain attack types occur less frequently than others.

Network Threat Detection Systems (NTDS) mainly rely on Machine Learning (ML) and Deep Learning (DL) models to classify network anomalies. However, class imbalanced datasets often make classifiers to be biased to the majority classes, resulting in poor recall for minority attack types. Many existing techniques frequently use oversampling and undersampling techniques to address this class imbalance issue (Werner de Vargas et al., 2023), but they either introduce redundant data or lose valuable information about the network traffic.

To overcome these challenges, this study introduces a robust NTDS framework that integrates SMOTETomek with Random Forest (RF). The SMOTETomek technique is a hybrid data balancing technique that synthetically generates new minority class samples using Synthetic Minority Over-sampling Technique (SMOTE), and on the other hand, the noisy instances are removed via Tomek links (Swana et al., 2022). This advanced approach ensures a more reliable and balanced dataset, which improves classification performance. Random Forest is a robust and widely used ensemble learning method, enhances model strength, reduces overfitting, and improves generalisation (Thomas & Kaliraj, 2024).

The proposed method uses the WSN-DS dataset for the experiment, which is a benchmark dataset for WSN. The experiment demonstrates a high classification performance, confirming the effectiveness of the SMOTETomek-based balancing strategy. The key contributions of this paper are:

- An advanced network threat detection framework leveraging SMOTETomek and Random Forest for enhanced attack classification.
- A novel approach for handling the class imbalance issue in WSN-DS, ensuring improved recall and precision for minority attack classes.
- Comparative analysis across many existing works demonstrates that the proposed method is superior in performance compared to traditional IDS models.

RELATED WORK

Threat detection in WSN is a major area of extensive research, with various ML and DL technique proposed to enhance security mechanisms. However, class imbalance and model robustness remain critical challenges for researchers that affect the performance of the models. This section reviews some existing literature based on IDS, data balancing techniques, and ensemble learning models for WSN security.

1.1 Intrusion Detection Systems in WSNs

Several IDS have been developed to identify network threats in WSNs. Traditional IDS

Received: August 02, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

approaches include:

Signature-based IDS: This type of system detects known attacks using predefined patterns but struggles with novel threats (Ahmed et al., 2025) (Khraisat et al., 2019).

Anomaly-based IDS: Anomaly based system uses statistical models and machine learning techniques to identify deviations from normal network behaviour (Khraisat et al., 2019).

Hybrid IDS: Hubrid models combine signature-based and anomaly-based detection for improved accuracy (Maseno et al., 2022).

Various supervised and unsupervised machine learning techniques have been widely used in anomaly-based IDS (Liu & Lang, 2019). However, these models often suffer poor performance on minority attack classes when trained on imbalanced datasets.

1.2 Class Imbalance in IDS Models

The imbalanced attack distributions in many datasets like WSN-DS dataset, severely impact models' performance (Patel et al., 2020). Several approaches have been explored to address this issue.

Oversampling: This technique increases the instances of minority classes using duplication of the original features or synthetic data generation.

Undersampling: This technique reduces majority class samples to balance the dataset, but this may lead to loss of some critical information.

Hybrid Methods: Techniques like SMOTETomek combine oversampling and undersampling for optimal balancing of the dataset.

Different studies have shown that SMOTETomek outperforms traditional oversampling or undersampling methods by eliminating redundant samples and reducing noise (Or, 2025) (Zhao et al., 2024), making it an ideal choice for improving IDS robustness.

1.3 Ensemble Learning for Intrusion Detection

Random Forest is a strong ensemble learning method, and has gained popularity for threat detection due to its:

- High accuracy and robustness in handling large feature sets (Farnaaz & Jabbar, 2016).
- Ability to reduce overfitting through bagging and random feature selection (Hasan et al., 2014).
- Efficient handling of imbalanced datasets when combined with preprocessing techniques like SMOTETomek (Hairani et al., 2023).

Recent studies have demonstrated that Random Forest model has achieved superior intrusion detection performance compared to traditional classifiers such as SVM and KNN (Uwazie et al., 2024). However, when used with imbalanced datasets, its effectiveness diminishes, which demands the necessity of the use of balancing techniques like SMOTETomek.

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

1.4 State-of-the-Art IDS Models

Many researchers have explored different models, primarily deep learning-based IDS solutions, such as:

CNN-LSTM models: Extract spatial-temporal attack patterns but require significant computational resources (Salmi & Oughdir, 2022).

Autoencoders: Detect anomalies through feature reconstruction but may struggle with class imbalance (Choi et al., 2019).

Hybrid deep learning models: Combine multiple architectures (Nazir et al., 2024) but are complex to implement in real-world WSN environments.

Though deep learning-based models achieve high detection rates, they may lack interpretability and require higher computational power. In contrast, Random Forest with SMOTETomek provides an interpretable, computationally efficient, and highly accurate IDS solution, making it more practical for real-world WSN security applications.

1.5 Research Gaps and Contributions

Despite the advancements in IDS models, robust solutions for class imbalance remain an explorative area. Many studies rely on basic oversampling or deep learning models, which are either ineffective or computationally expensive. These gaps are considered to be tackled in this study by:

- Proposing a hybrid SmoteTomek-Random Forest IDS, ensuring optimal class balance and superior attack classification.
- Demonstrating robust performance on the WSN-DS dataset, surpassing existing IDS models.
- Providing a computationally efficient and interpretable IDS framework, suitable for real-world WSN applications.

This literature review establishes the foundation for the proposed SMOTETomek-Enhanced Random Forest IDS, which is detailed in Section 3 (Methodology).

METHODOLOGY

This section presents the proposed SMOTETomek-integrated Random Forest IDS for detecting security threats in WSN. This method integrates SMOTETomek for handling class imbalance and Random Forest for robust classification. The training and evaluation were conducted using the WSN-DS dataset, which contains labeled network traffic data representing both normal and attack scenarios. Algorithm 1 is designed to input the WSN-DS dataset and produce performance results of the method used in this study.

Algorithm 1 SMOTETomek-RF method

1: **Input:** Dataset D with features X and target y

2: **Output:** Trained RF model and performance metrics

3: **Step 1: Data Preprocessing**

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

I: Load dataset D and define feature matrix X and target labels y

II: Remove redundant attributes and handle missing values

III: Encode each categorical label into a numerical value.

4: **Step 2: Handling Class Imbalance with SMOTETomek**

SMOTE:

I: For every minority attack sample x_i, find its k-nearest neighbors.

II: Select one neighbor x_j and generate a synthetic sample as:

$$x_{\text{new}} = x_i + \lambda \cdot (x_j - x_i),\ \lambda \sim U(0,1)$$

III: Repeat until the class distribution is balanced.

5: **Tomek Links for Cleaning Overlap**

I: A pair of samples (x_i,x_j) forms a Tomek link if:

$$\forall x_k,\ d(x_i,x_j) < d(x_i,x_k),\ d(x_i,x_j) < d(x_j,x_k)$$

where $d(x,y)$ is the Euclidean distance.

II: Remove the redundant samples from the majority class in each Tomek link.

III: Prepare the final balanced dataset $(X_{\text{resampled}},y_{\text{resampled}})$.

6: **Step 3: Train-Test Split**

I: Split $(X_{\text{resampled}}, y_{\text{resampled}})$ into training and test sets :

(Xtrain, Xtest, $y$train, $y$test) ← train test split (Xresampled, $y$resampled, 70%−30%)

7: **Step 4: Train Random Forest Model**

I: Start RF technique with n estimators:

RF ← RF Classifier (n *estimators* = 100, random state = 42)

II: Train the model on balanced dataset:

RF.fit (Xtrain, $y$train)

8: **Step 5: Model Prediction**

I: Predict class labels on test set:

$y$pred ← RF.predict (Xtest)

II: Compute class probabilities:

$y$pred proba ← RF.predict proba (Xtest)

9: **Step 6: Performance Evaluation**

I: Compute Accuracy, Precision, Recall, and F1-score.

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

1.6 Dataset and Preprocessing

The WSN-DS dataset consists of multiple network traffic features and five distinct labels: Normal, Grayhole, Blackhole, TDMA, and Flooding. To prepare the dataset for model training, the following data preprocessing steps have been considered.

Feature Selection: In this step, the non-relevant attributes, such as ID and Time, were removed.

Missing Values: Here, any incomplete records or null attributes were eliminated to ensure data integrity.

Label Encoding: Each attack and normal class were encoded numerically for compatibility with machine learning models.

1.7 Addressing Class Imbalance with SMOTETomek

To counteract the class imbalance issue, the dataset was balanced using SMOTETomek, a combination of Smote that generates synthetic data features for minority classes and Tomek links that remove redundant instances of majority class samples, refining the decision boundary.

The SMOTE technique generates synthetic instances along the vector between a minority class sample and one of its k-nearest neighbors:

$x_{\text{new}} = x_i + \lambda \cdot (x_j - x_i), \lambda \sim U(0,1)$xnew=xi+λ·(xj−xi),λ∼U(0,1)

where:

- $x_i$xi is a randomly selected minority sample,
- $x_j$xj is its k-nearest neighbor,
- λλ is a random number between 0 and 1.

This process increases the representation of attack categories, ensuring that the model learns attack patterns effectively.

The Tomek links refine the dataset by identifying pairs of nearest-neighbor samples from different classes:

$$\forall x_k, d(x_i, x_j) < d(x_i, x_k), d(x_i, x_j) < d(x_j, x_k)$$

- where $d(x, y)$ is the Euclidean distance.

The majority-class sample in each pair is removed, preventing misclassification at class boundaries.

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

1.8 Training the Random Forest Model

Once the original dataset WSN-DS dataset was balanced, it was split into training (70%) and testing (30%) subsets. The RF classifier was then trained on the resampled dataset and finally tested and evaluated the model.

RESULTS AND ANALYSIS

The performance evaluation of the proposed SMOTETomek enhanced Random Forest Intrusion Detection System on the WSN-DS dataset has been presented in this section. The performance of the proposed method is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. The results shows its ability to effectively identify various network security threats. Moreover, some visual analysis, such as confusion matrix, ROC–AUC curves are employed to illustrate the model's discriminative capability across different classes.

The original WSN-DS dataset has a significant class imbalance, where normal traffic instances are far more prevalent than other attack categories. The Fig. 1 depicts the class distribution of the WSN-DS dataset before applying the SMOTETomek technique, clearly showing the dominance of the normal class. After implementing SMOTETomek technique, the dataset achieved a more balanced distribution, as shown in Fig. 2, ensuring equitable representation of all classes and thus improving model generalization and detection accuracy. Also, the Fig. 3 depicts the feature distribution of the WSN-DS dataset showing the changes before and after applying SMOTETomek technique.

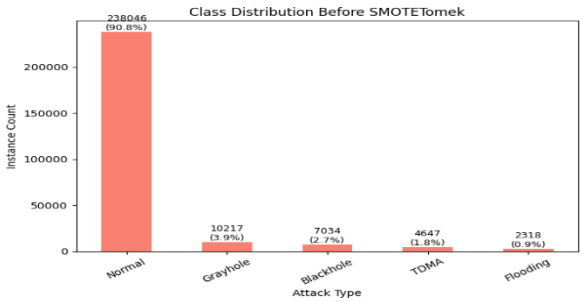


Fig. 1. Class Distribution of the WSN-DS dataset before applying SMOTETomek.

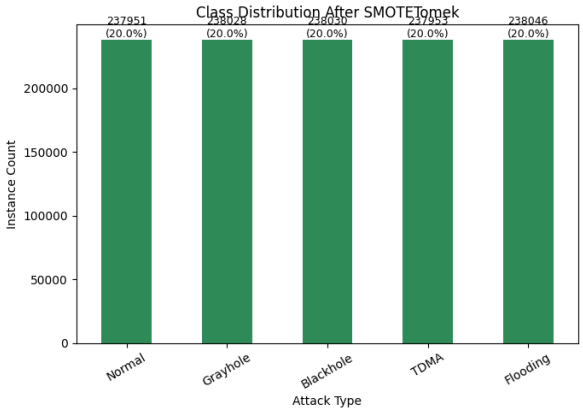


Fig. 2. Class Distribution of the WSN-DS dataset after applying SMOTETomek.

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

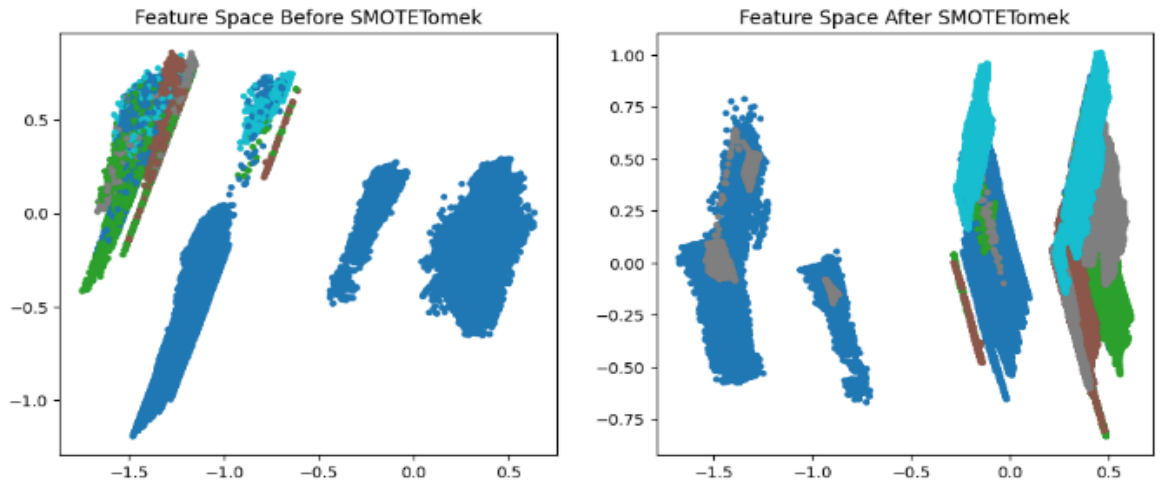


Fig. 3. Feature distribution of the WSN-DS dataset showing the changes before and after applying SMOTETomek.

To get a better understanding of the characteristics of the dataset, pairwise feature relationships were analyzed, as illustrated in Fig. 4, revealing inter-feature correlations and separability trends between the classes. Furthermore, the most influential features of the dataset identified by the enhanced Random Forest model are also shown in Fig. 5. This figure indicates that certain attributes play a crucial role in differentiating between several types of attacks and normal behaviour of the network traffic.

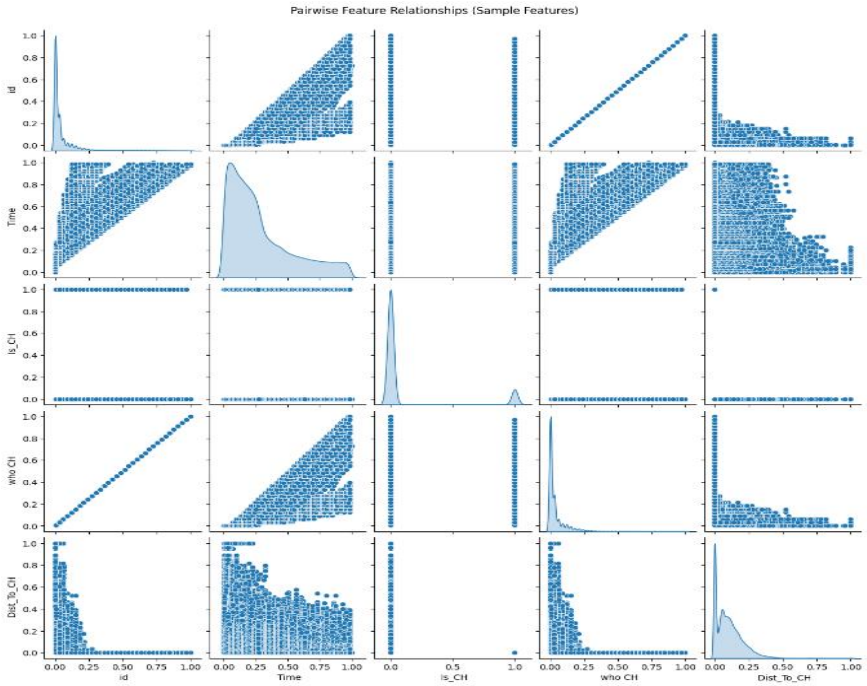


Fig. 4. Pairwise feature relationship of the WSN-DS dataset.

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

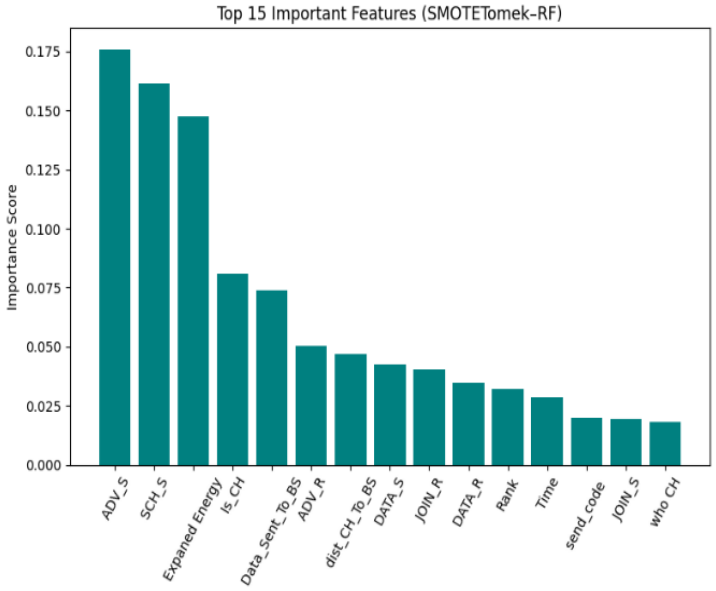


Fig. 5. The most significant features considered in the WSN-DS dataset.

The classification performance of the proposed SMOTETomek-RF model is summarized in Table 1, which presents class-wise accuracy, precision, recall, and F1-score. The model demonstrated exceptional performance across all categories, achieving consistently high metric values.

Table 1. Performance Evaluation Metrics for Different Attack Classes and Normal Traffic of the SMOTETomek-Random Forest Model on the WSN-DS Dataset

Class	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
Normal	99.63	99.86	99.79	99.82
Grayhole	99.62	97.96	98.84	98.40
Blackhole	99.63	98.49	99.67	99.08
TDMA	98.40	96.67	93.32	9497
Flooding	99.63	92.68	99.30	95.87

Table 2 shows that the proposed model achieved an overall accuracy of 99.38%, precision of 97.53%, recall of 98.18%, and an F1-score of 97.62%, reflecting its strong capability to classify both majority and minority classes accurately. The per-class F1-scores also indicate the strength of the model, with excellent performance even for complex attack types.

The confusion matrix in Fig. 6 provides further insights into the model's prediction performance, showing the proportion of properly and inaccurately categorized instances for each class. The strong diagonal values highlight the model's high detection accuracy with

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

minimal misclassification errors.

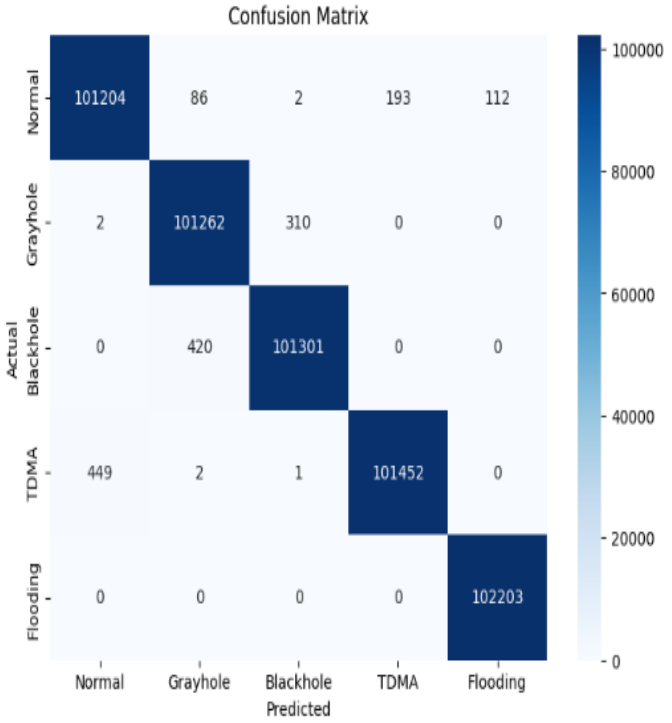
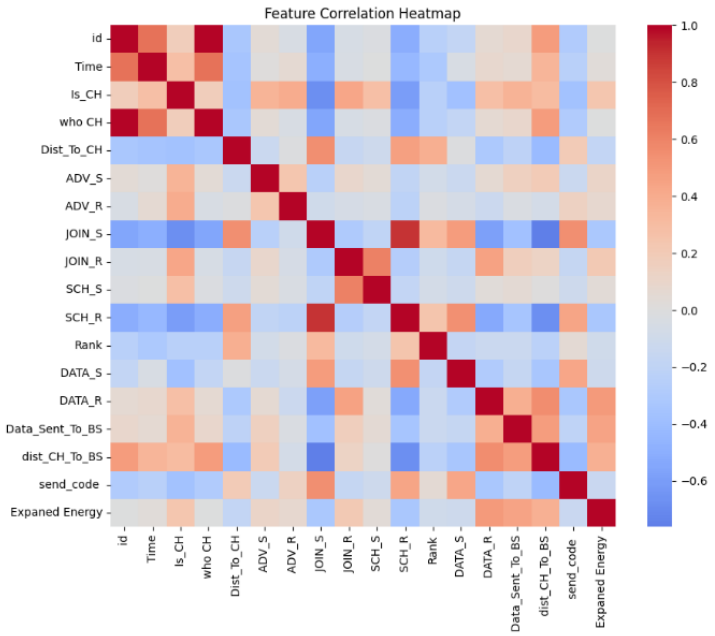


Fig. 6. Confusion matrix for the SMOTETomek-RF model on the WSN-DS dataset.

Additionally, feature correlation within the dataset was examined, as illustrated in Fig. 7, to analyze redundancy and interdependence among input variables. This analysis supports the selection of relevant features, thereby enhancing model interpretability and reducing overfitting risks.



International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Fig. 7. Feature correlation heatmap of the WSN-DS dataset.

To evaluate discriminative performance, the ROC-AUC curve were plotted for each attack class, as shown in Fig. 8 The curves closely approach the top-left corner, confirming that the proposed model achieved near-perfect AUC scores for all categories. This indicates superior capability in distinguishing between normal and malicious traffic.

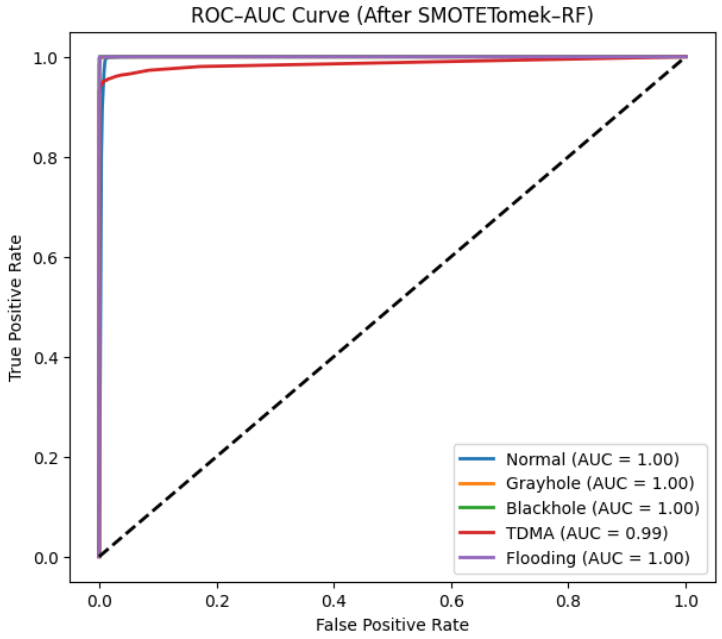


Fig. 8. ROC-AUC Curve for Each Attack Class of the WSN-DS dataset.

Finally. In order to verify the effectiveness of the suggested model, a performance comparison with many existing methods was carried out. Table 2 compares the SMOTETomek-RF model with conventional ML and DL methods such as Decision Tree (DT), and CNN–LSTM, as reported in prior studies (Kasongo & Sun, 2020; Salmi & Oughdir, 2022).

Table 2. Comparison of Proposed SMOTETomek-RF model with Existing Models

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
SVM (Kasongo & Sun, 2020)	93.45	91.78	90.56	91.16
Decision Tree (Kasongo & Sun, 2020)	95.67	94.23	93.89	94.06
CNN-LSTM (Salmi & Oughdir, 2022)	94.4	95.9	92.2	-
Proposed	99.38	97.53	98.18	97.62

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SMOTETomek-RF				

From Table 2, it can be concluded that the proposed SMOTETomek-Random Forest model significantly outperforms existing traditional and hybrid IDS approaches. This remarkable improvement in all evaluation metrics demonstrates the model's robustness, stability, and high detection capability for both frequent and rare intrusion types in Wireless Sensor Networks.

CONCLUSION AND FUTURE WORK

The proposed framework effectively addressed the class imbalance problem in the WSN-DS dataset through the integration of the SMOTETomek resampling technique, which improves the model's learning capability and led to superior classification outcomes across all attack categories. This method achieved an overall accuracy of 99.38%, precision of 97.53%, recall of 98.18%, and F1-score of 97.63%, which shows robustness over conventional approaches such as SVM, Decision Tree, and CNN–LSTM models.

Future research will focus on deploying the model in real-time WSN environments to evaluate its adaptability under dynamic network conditions, energy constraints, and communication overhead. Additionally, the integration of deep representation learning and attention-based hybrid models could further enhance feature extraction and decision accuracy.

REFERENCES

1. Ahmed, U., Nazir, M., Sarwar, A., Ali, T., Aggoune, E. H. M., Shahzad, T., & Khan, M. A. (2025). Signature-based intrusion detection using machine learning and deep learning approaches empowered with fuzzy clustering. *Scientific Reports 2025 15:1*, *15*(1), 1–33. https://doi.org/10.1038/s41598-025-85866-7
2. Choi, H., Kim, M., Lee, G., & Kim, W. (2019). Unsupervised learning approach for network intrusion detection system using autoencoders. *Journal of Supercomputing*, *75*(9), 5597–5621. https://doi.org/10.1007/s11227-019-02805-w
3. Farnaaz, N., & Jabbar, M. A. (2016). Random Forest Modeling for Network Intrusion Detection System. *Procedia Computer Science*, *89*, 213–217. https://doi.org/10.1016/j.procs.2016.06.047
4. Hairani, H., Anggrawan, A., & Priyanto, D. (2023). Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link. *International Journal on Informatics Visualization*, *7*(1), 258–264. https://doi.org/10.30630/JOIV.7.1.1069
5. Hasan, Md. A. M., Nasser, M., Pal, B., & Ahmad, S. (2014). Support Vector Machine and Random Forest Modeling for Intrusion Detection System (IDS). *Journal of Intelligent Learning Systems and Applications*, *06*(01), 45–52.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

https://doi.org/10.4236/jilsa.2014.61005

6. Kasongo, S. M., & Sun, Y. (2020). Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/S40537-020-00379-6
7. Keerthika, M., & Shanmugapriya, D. (2021). Wireless Sensor Networks: Active and Passive attacks - Vulnerabilities and Countermeasures. *Global Transitions Proceedings*, *2*(2), 362–367. https://doi.org/10.1016/J.GLTP.2021.08.045
8. Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity*, *2*(1), 1–22. https://doi.org/10.1186/S42400-019-0038-7/FIGURES/8
9. Le, T. T. H., Shin, Y., Kim, M., & Kim, H. (2024). Towards unbalanced multiclass intrusion detection with hybrid sampling methods and ensemble classification. *Applied Soft Computing*, *157*, 111517. https://doi.org/10.1016/J.ASOC.2024.111517
10. Liu, H., & Lang, B. (2019). Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey. *Applied Sciences 2019, Vol. 9, Page 4396*, *9*(20), 4396. https://doi.org/10.3390/APP9204396
11. Majid, M., Habib, S., Javed, A. R., Rizwan, M., Srivastava, G., Gadekallu, T. R., & Lin, J. C. W. (2022). Applications of Wireless Sensor Networks and Internet of Things Frameworks in the Industry Revolution 4.0: A Systematic Literature Review. *Sensors 2022, Vol. 22, Page 2087*, *22*(6), 2087. https://doi.org/10.3390/S22062087
12. Maseno, E. M., Wang, Z., & Xing, H. (2022). A Systematic Review on Hybrid Intrusion Detection System. *Security and Communication Networks*, *2022*(1), 9663052. https://doi.org/10.1155/2022/9663052
13. Nazir, A., He, J., Zhu, N., Qureshi, S. S., Qureshi, S. U., Ullah, F., Wajahat, A., & Pathan, M. S. (2024). A deep learning-based novel hybrid CNN-LSTM architecture for efficient detection of threats in the IoT ecosystem. *Ain Shams Engineering Journal*, *15*(7), 102777. https://doi.org/10.1016/J.ASEJ.2024.102777
14. Or, B. (2025). *Improving Requirements Classification with SMOTE-Tomek Preprocessing*. http://arxiv.org/abs/2501.06491
15. Patel, H., Singh Rajput, D., Thippa Reddy, G., Iwendi, C., Kashif Bashir, A., & Jo, O. (2020). A review on classification of imbalanced data for wireless sensor networks. *International Journal of Distributed Sensor Networks*, *16*(4). https://doi.org/10.1177/1550147720916404/ASSET/IMAGES/LARGE/10.1177_1550147720916404-FIG2.JPEG
16. Salmi, S., & Oughdir, L. (2022). CNN-LSTM Based Approach for Dos Attacks Detection in Wireless Sensor Networks. *IJACSA) International Journal of Advanced Computer Science and Applications*, *13*(4), 835–842. www.ijacsa.thesai.org
17. Swana, E. F. ;, Doorsamy, W. ;, Bokoro, P. T., Link, S., Fezeka Swana, E., Doorsamy, W.,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

& Bokoro, P. (2022). Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset. *Sensors 2022, Vol. 22, Page 3246*, *22*(9), 3246. https://doi.org/10.3390/S22093246

18. Thomas, N. S., & Kaliraj, S. (2024). An Improved and Optimized Random Forest Based Approach to Predict the Software Faults. *SN Computer Science*, *5*(5), 1–18. https://doi.org/10.1007/S42979-024-02764-X/FIGURES/21
19. Uwazie, E. C., Obiniyi, A. A., Olalere, M., & Achi, P. N. (2024). Comparison of Random Forest, K-Nearest Neighbor, and Support Vector Machine Classifiers for Intrusion Detection System. *International Conference on Science, Engineering and Business for Driving Sustainable Development Goals, SEB4SDG 2024*. https://doi.org/10.1109/SEB4SDG60871.2024.10629939
20. Werner de Vargas, V., Schneider Aranda, J. A., dos Santos Costa, R., da Silva Pereira, P. R., & Victória Barbosa, J. L. (2023). Imbalanced data preprocessing techniques for machine learning: a systematic mapping study. *Knowledge and Information Systems*, *65*(1), 31–57. https://doi.org/10.1007/S10115-022-01772-8/TABLES/7
21. Zhao, C., Yan, Z., Sun, X., & Wu, M. (2024). Enhancing aspect category detection in imbalanced online reviews: An integrated approach using Select-SMOTE and LightGBM. *International Journal of Intelligent Networks*, *5*, 364–372. https://doi.org/10.1016/J.IJIN.2024.10.002

Received: August 02, 2025