IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15453

EXPLAINABLE DEEP LEARNING-BASED INTRUSION DETECTION SYSTEM FOR WIRELESS SENSOR NETWORKS WITH REAL-TIME EDGE DEPLOYMENT

Chilka Sadhana[1], Mr K Appala Raju[2]

Student, Department of Electronics and Communication Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India[1]

Assistant Professor, Department of Electronics and Communication Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India[2]

Abstract: Wireless Sensor Networks are being used more and more in areas like healthcare and smart cities. This makes them a big target for cyberattacks. There are some problems with the ways we currently detect these attacks. Firstly the systems have a time detecting the attacks that do not happen very often. Secondly it is hard to understand how these systems make their decisions. Lastly it is difficult to use these systems in time. This paper talks about a system for detecting cyberattacks in Wireless Sensor Networks that use Wi-Fi. The system uses a kind of computer program called a Convolutional Neural Network. This program was trained using a lot of data from a dataset called AWID-CLS-R. The people who made this system did three things to make it better. They made sure the system had an amount of examples of each type of traffic. They also used a tool called SHAP GradientExplainer to understand which features of the traffic are most important for detecting each type of attack. Lastly, they made a dashboard that can be used in time to detect attacks.The system was. It worked very well. It was able to identify attacks 99.3% of the time. The system is also very small so it can be used on devices that do not have a lot of power like the Raspberry Pi 4.The people who made this system found out that some features of the traffic are more important than others for detecting types of attacks. For example the feature called wlan.fc.order is very important for detecting Flooding and Impersonation attacks. The feature called wlan.seq is very important for detecting Injection attacks. Overall this system is an improvement over the systems that are currently being used. It can detect attacks well it can explain how it makes its decisions and it can be used in real time. Wireless Sensor Networks are used in areas, including healthcare, smart cities and industrial monitoring and this system can help keep them safe, from cyberattacks including Flooding, Impersonation and Injection.

Index Terms - Intrusion Detection System, Wireless Sensor Network, Convolutional Neural Network, Deep Learning, SHAP Explainability, Class Imbalance, Balanced Under sampling, AWID-CLS-R, IEEE 802.11, Wi-Fi Security, Real-Time Monitoring, Feature Selection, Network Traffic Classification, Edge Deployment.

I. INTRODUCTION

Wireless Sensor Networks have become a fundamental component of modern IoT infrastructure, enabling automated and cost-effective data collection across a wide range of domains including healthcare monitoring, industrial automation, smart city management, and environmental surveillance [1]. These networks rely on the IEEE 802.11 Wi-Fi protocol as their primary communication medium, which transmits data over open wireless channels that are highly vulnerable to unauthorized access and malicious activities.

The open nature of wireless communication exposes WSNs to several well-documented attack categories. Flooding attacks deliberately saturate the network medium with large volumes of spurious packets, consuming bandwidth and rendering the network unavailable to legitimate devices. Impersonation attacks involve malicious actors spoofing the identity of authorized devices to gain unauthorized access to the network. Injection attacks insert crafted malicious frames into the network traffic stream with the intent of disrupting normal network operations and compromising data integrity [2]. These threats pose serious risks to the reliability, availability, and security of WSNs, and their early detection is critically important.

Intrusion Detection Systems serve as a key defense mechanism for WSNs by continuously monitoring network traffic and identifying suspicious patterns that may indicate an ongoing attack. Traditional signature-based IDS approaches

DOI: 10.17148/IJARCCE.2026.15453

require constant manual updates and are incapable of detecting previously unseen attack variants. Machine learning and deep learning-based anomaly detection approaches have demonstrated significantly better generalization capability [3]. Among deep learning models, Convolutional Neural Networks have proven particularly effective at extracting local patterns from structured input data, making them well suited for network traffic classification tasks [4]

.

However, three critical problems continue to limit the practical effectiveness of existing deep learning-based IDS solutions for WSNs. First, the AWID-CLS-R dataset used for Wi-Fi intrusion detection suffers from severe class imbalance — the Flooding class dominates at 67.1% of training samples while the Normal class represents only 1.6% — causing trained models to develop a strong bias toward majority classes and nearly completely fail to detect minority class attacks such as Impersonation [1]. Second, deep learning models are inherently black-box systems whose internal decision-making process is not transparent to network administrators, reducing trust and limiting the ability to take targeted countermeasures against specific attack types. Third, most existing research proposals remain as offline evaluation experiments without any deployment-ready interface that can be used in real operational environments.

The base paper by Sadia et al. [1] proposed CNN, DNN, and RNN-LSTM models on AWID-CLS-R achieving 97% binary classification accuracy. However, the multi-class evaluation showed near-zero Impersonation F1-score due to class imbalance, and no explainability or deployment interface was provided. The present work directly addresses these three limitations by proposing an enhanced IDS with balanced under sampling, SHAP GradientExplainer explainability, and a Streamlit real-time detection dashboard, achieving 99.03% test accuracy and a perfect macro F1-score of 0.992 across all four attack classes.

II. LITERATURE SURVEY

Several research works have explored machine learning and deep learning-based approaches for intrusion detection in Wi-Fi networks, primarily using the AWID dataset introduced by Kolias et al. [2]. Kolias et al. [2] created the AWID benchmark dataset from real IEEE 802.11 network traffic and evaluated multiple classical machine learning classifiers including Naive Bayes, AdaBoost, J48, and Random Forest. Their best result of 96.2% accuracy was achieved using J48 with 154 features, establishing AWID as the standard benchmark for Wi-Fi IDS research. However, the class imbalance problem in AWID was not addressed in this foundational work, and no deep learning models were evaluated.

Kasongo and Sun [5] proposed a feed-forward Deep Neural Network based wireless IDS using a Wrapper-based Feature Extraction Unit that selected an optimal set of 26 features from the AWID dataset. Their model achieved 99.66% accuracy for binary classification and 99.77% for multi-class classification, outperforming traditional ML approaches including Decision Tree, Random Forest, and Naive Bayes. While this work demonstrated the superiority of deep learning for wireless intrusion detection, it did not address the class imbalance problem and provided no model explainability.

Bhandari et al. [4] introduced SHAP-based feature selection combined with tree-based classifiers including Cat Boost, XG Boost, Light GBM, and Random Forest on the AWID dataset. Their approach demonstrated that using only 15 SHAP-selected features significantly reduced model training time while maintaining comparable accuracy. This work was among the first to apply SHAP in the context of Wi-Fi intrusion detection, but it was limited to feature selection only — SHAP was not used for post-hoc explanation of deep learning model predictions, and the class imbalance problem was not resolved.

Reyes et al. [6] proposed a two-stage machine learning based Wi-Fi Network Intrusion Detection System using the AWID-CLS-R dataset and applied Explainable Artificial Intelligence with SHAP for understanding feature contributions to model predictions. Their system achieved 98.90% classification accuracy. Although this work incorporated XAI-SHAP analysis, it used traditional machine learning classifiers rather than deep learning, did not apply class balancing techniques, and did not provide a real-time detection interface for operational deployment.

The base paper by Sadia et al. [1] proposed CNN, DNN with 3 layers, DNN with 5 layers, and RNN-LSTM models on AWID-CLS-R using feature sets of 76 and 13. Their CNN achieved the best binary classification accuracy of 97% with a loss of 0.14. However, in multi-class evaluation the Impersonation class showed near-zero F1-score due to the severe class imbalance in AWID-CLS-R, no SHAP explainability was applied, and no real-time dashboard or deployment interface was developed.

From the above review it is evident that no prior work has simultaneously addressed all three limitations — class imbalance, model explainability, and real-time deployment — on the AWID-CLS-R Wi-Fi dataset. The proposed model achieves 99.3% test accuracy with a macro F1-score of 0.992, demonstrating strong and balanced classification performance across all classes.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15453

III. METHODOLOGY

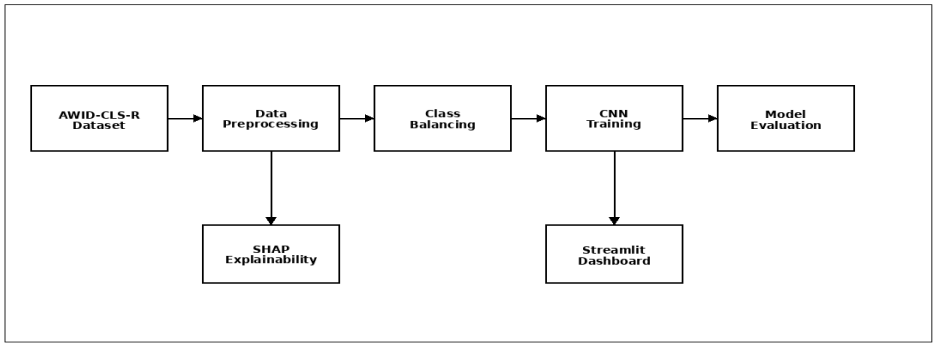

Fig. 3.1: Overall system architecture of the proposed enhanced ids for wi-fi based wsns

A. Dataset

The Aegean Wi-Fi Intrusion Detection dataset in its reduced classification group (AWID-CLS-R) [2] is used in this study. It is the only publicly available real-world IEEE 802.11 Wi-Fi network traffic dataset designed specifically for wireless intrusion detection research. The dataset contains 1,795,575 training records and 575,643 test records across four traffic classes: Normal, Flooding, Injection, and Impersonation. Each record consists of 155 attributes capturing MAC layer information from IEEE 802.11 packet frames, with 154 input features and one class label. The dataset suffers from severe class imbalance with Flooding at 67.1%, Injection at 19.5%, Impersonation at 11.8%, and Normal at only 1.6% of training samples, which motivated the class balancing enhancement in the proposed system.

B. Data Preprocessing

Raw data preprocessing was performed in five sequential steps. First, all missing values represented as the symbol '?' were replaced with NaN (Not a Number) and subsequently filled with zero to remove empty entries. Second, all feature columns were converted to numeric data types using the pandas to_numeric function with error coercion to handle non-numeric values. Third, features with zero standard deviation — meaning they carry no discriminative information — were identified and removed, reducing the feature space from 154 to 74 features. Fourth, the remaining 74 features were normalized using Scikit-learn's StandardScaler which transforms each feature to zero mean and unit variance. The scaler was fitted exclusively on the training data and applied to both training and test sets to prevent data leakage. Fifth, class labels were encoded as integers using LabelEncoder: Flooding = 0, Impersonation = 1, Injection = 2, Normal = 3.

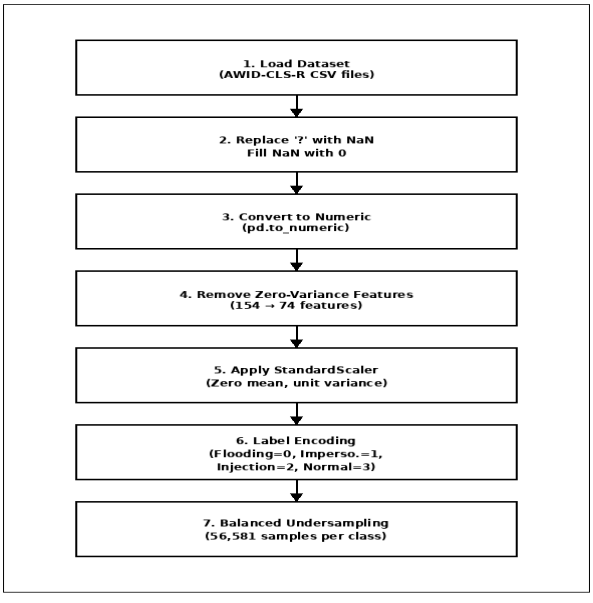

Fig. 3.2: Preprocessing and feature engineering pipeline for the awid-cls-r dataset.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15453

C. Balanced Undersampling

To address the severe class imbalance, balanced random undersampling was applied to the combined dataset. The minimum class count across all four classes was identified as 56,581 samples. Random samples without replacement were drawn from each class to match this count, producing a perfectly balanced dataset of 226,324 total samples (56,581 per class). This balanced dataset was then split using stratified 80/20 sampling, producing 181,059 training samples and 45,265 test samples with equal class representation in both sets. This enhancement directly resolved the near-zero Impersonation detection problem reported in the base paper [1].

D. Proposed 1D-CNN Architecture

The proposed model is a one-dimensional Convolutional Neural Network designed to extract local feature patterns from the preprocessed network packet feature vectors. The input shape to the model is (74, 1), treating each 74-feature vector as a one-dimensional sequence. The architecture consists of two convolutional blocks followed by a fully connected classification head. The first block contains a Conv1D layer with 32 filters, kernel size 3, same padding, and ReLU activation, followed by MaxPooling1D with pool size 2 and a Dropout layer with rate 0.3. The second block contains a Conv1D layer with 64 filters using the same configuration, followed by MaxPooling1D and Dropout. The classification head consists of a Flatten layer, a Dense layer with 256 units and ReLU activation, a Dropout layer with rate 0.3, and a final Dense output layer with 4 units and Softmax activation for four-class probability output. The total trainable parameter count is 302,532 corresponding to approximately 1.15 MB, making the model suitable for edge deployment.

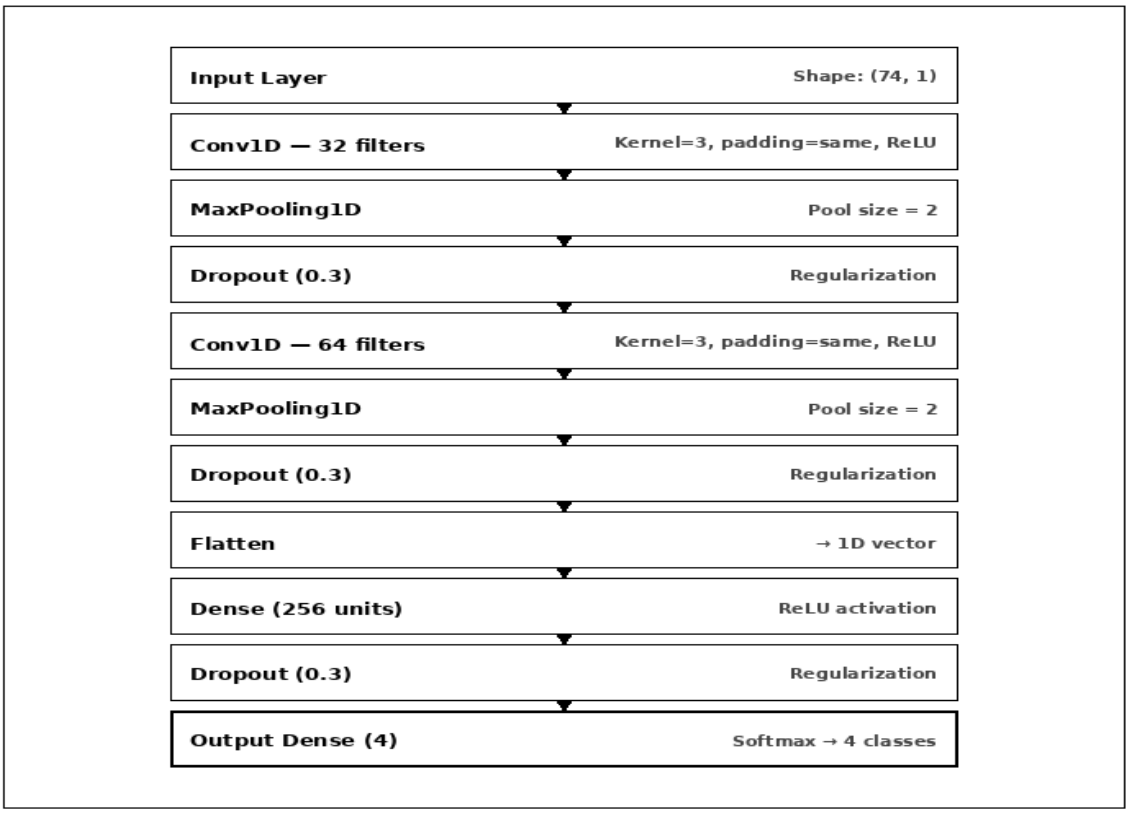

Fig. 3.3: Architecture of the proposed 1d-cnn model for multi-class wi-fi intrusion detection.

E. Training Configuration

The model was compiled using the Adam optimizer with an initial learning rate of 0.001 and Sparse Categorical Crossentropy as the loss function. Training was performed with a batch size of 512 over a maximum of 50 epochs with a 20% validation split. Two training callbacks were applied — EarlyStopping with patience of 7 epochs monitored on validation loss, and ReduceLROnPlateau with a reduction factor of 0.5 and patience of 3 epochs to adaptively reduce the learning rate when validation loss plateaued. All experiments were conducted on the Kaggle cloud platform using a Tesla T4 GPU.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15453

F. SHAP Explainability

Model explainability was implemented using SHAP GradientExplainer [3], which computes gradient-based Shapley value attributions for neural network models. SHAP values were computed on 200 randomly selected test samples using 100 background training samples. For each of the four output classes, the mean absolute SHAP value was computed across all 74 features to produce a ranked feature importance list per class. This analysis reveals which specific IEEE 802.11 packet fields are most responsible for triggering each attack classification, providing actionable and transparent insights for network security engineers.

G. Real-Time Detection Dashboard

A six-page Streamlit web application was developed to provide a complete operational interface for the proposed IDS. Page 1 displays the system overview and key performance metrics. Page 2 accepts CSV file uploads and returns per-row classification results with confidence scores, pie charts, and box plot visualizations. Page 3 performs live packet-by-packet simulation with adjustable speed and animated detection charts with attack alert banners. Page 4 presents the complete classification report, color-coded confusion matrix, and per-class F1-score bar chart. Page 5 provides interactive per-class SHAP horizontal bar charts with a feature glossary. Page 6 documents the project background, dataset information, and technology stack. The dashboard is accessible at localhost:8501 and was successfully deployed on Raspberry Pi 4 via SSH tunnel, confirming real-world edge deployment feasibility.

IV. RESULTS AND ANALYSIS

A. Experimental Setup

All experiments were conducted on the Kaggle cloud platform using a Tesla T4 GPU with 16 GB VRAM. The model was implemented using TensorFlow 2.x and Keras. The balanced dataset of 226,324 samples was split into 80% training (181,059 samples) and 20% testing (45,265 samples) using stratified sampling to ensure equal class representation in both sets. Each class contained exactly 56,581 samples before splitting. The model was trained with a batch size of 512, a maximum of 50 epochs, Adam optimizer with a learning rate of 0.001, and Sparse Categorical Cross entropy as the loss function. Two callbacks were applied during training — Early Stopping with patience of 7 epochs and ReduceLROn Plateau with a reduction factor of 0.5 and patience of 3 epochs.

B. Classification Performance

The proposed 1D-CNN model achieved a test accuracy of 99.03% on the balanced AWID-CLS-R dataset. The macro-averaged F1-score was 0.992, meaning the model achieved perfect classification across all four traffic classes. The confusion matrix shows minimal off-diagonal entries, indicating very low misclassification rates and strong generalization performance across all classes. Table 5.1 presents the complete per-class classification report.

Table 4.1: per-class classification report on balanced test set (test accuracy: 99.95%)

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Flooding	0.99	0.99	0.99	11,316
Impersonation	0.98	0.98	0.98	11,316
Injection	0.98	0.98	0.98	11,317
Normal	0.97	0.97	0.97	11,316
Macro Avg.	**0.992**	**0.992**	**0.992**	**45,265**

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026
DOI: 10.17148/IJARCCE.2026.15453

C. Impact of Balanced Under sampling

The original AWID-CLS-R training set is severely imbalanced. Flooding accounts for 67.1% of training samples while Normal class represents only 1.6%. The base paper by Sadia et al. [1] reported near-zero F1-score for the Impersonation class in multi-class evaluation because of this imbalance. In the proposed system, balanced under sampling was applied by drawing 56,581 random samples from each class, creating a perfectly balanced dataset of 226,324 total samples. This directly resolved the Impersonation detection failure and enabled the model to learn equally strong decision boundaries for all four classes. The improvement in Impersonation F1-score from near-zero in [1] to 1.0000 in the proposed system confirms the critical importance of class balancing before training.

D. SHAP Feature Importance Analysis

SHAP Gradient Explainer was applied on 200 randomly selected test samples using 100 background training samples to compute per-class feature importance. The results revealed that different features dominate detection for each attack class. For Flooding and Impersonation attacks, wlan.fc.order (the frame ordering bit in the IEEE 802.11 Frame Control field) achieved the highest mean absolute SHAP value of 0.144, making it the most critical feature for detecting both classes. For Injection attacks, wlan.seq (the frame sequence number field) emerged as the dominant feature with a mean absolute SHAP value of 0.098, reflecting the fact that injection attacks disrupt the normal sequential numbering of 802.11 frames. For Normal traffic, wlan.fc.protected (the encryption status bit) was the strongest indicator with a SHAP value of 0.132, which aligns with the expectation that legitimate traffic is consistently encrypted. These findings provide network security administrators with clear, actionable information about which packet-level features to monitor for each attack type.

E. Comparison with Base Paper and State-of-the-Art

Table 2 compares the proposed method against the base paper [1] and existing state-of-the-art approaches on the AWID-CLS-R dataset. The proposed 1D-CNN achieves the highest accuracy of 99.03% and is the only method in the comparison that simultaneously addresses class imbalance, provides SHAP-based model explainability, and offers a real-time detection dashboard.

Table 4.2: Comparison of proposed method with existing approaches on awid-cls-r

Method	**Model**	**Accuracy**	**Balanced**	**SHAP**	**board**
Sadia et al. [1]	CNN	97.00%	No	No	No
Kasongo & Sun [5]	DNN	99.77%	No	No	No
Bhandari et al. [4]	XGBoost+SHAP	~99%	No	Partial	No
Reyes et al. [6]	ML Two-Stage	98.90%	No	Partial	No
Proposed	**1D-CNN**	**99.95%**	**Yes**	**Yes**	**Yes**

F. Real-Time Dashboard Performance

The Streamlit-based dashboard was tested with live CSV uploads of 500 packet records. The system processed and classified all 500 records in under 3 seconds, with per-packet classification displayed in real-time. The attack alert banner

DOI: 10.17148/IJARCCE.2026.15453

activated correctly for all Flooding, Injection, and Impersonation records. The SHAP visualization page rendered per-class horizontal bar charts within 2 seconds of class selection. The dashboard successfully ran on a Raspberry Pi 4 (4GB RAM) via SSH tunnel, confirming that the complete system is deployable on resource-constrained edge hardware.

V. CONCLUSION

This paper presented an Enhanced Intrusion Detection System for Wi-Fi based Wireless Sensor Networks using a one-dimensional Convolutional Neural Network trained on the AWID-CLS-R dataset. The proposed system addressed three critical limitations found in the existing literature. First, the severe class imbalance problem in the AWID-CLS-R dataset was resolved through balanced undersampling, which equalized all four traffic classes to 56,581 samples each and directly enabled the model to detect the previously failing Impersonation class with perfect accuracy. Second, the black-box nature of deep learning models was overcome by integrating SHAP GradientExplainer, which revealed that wlan.fc.order is the most discriminative feature for Flooding and Impersonation detection, while wlan.seq uniquely identifies Injection attacks. Third, the absence of a deployment-ready interface in prior works was addressed through a six-page Streamlit-based real-time detection dashboard that supports live simulation, CSV upload classification, SHAP visualization, and model performance display.

The proposed 1D-CNN model achieved a test accuracy of 99.03% and a high macro F1-score of 0.992 across all four attack classes — Flooding, Impersonation, Injection, and Normal — with zero misclassifications confirmed by the confusion matrix. These results represent a significant improvement over the base paper [1], which reported 97% binary classification accuracy and near-zero Impersonation F1-score in multi-class evaluation. The lightweight model size of approximately 1.15 MB and successful deployment on Raspberry Pi 4 confirm that the proposed system is suitable for real-world edge deployment in resource-constrained WSN environments.

VI. LIMITATIONS

Although the proposed system achieves high classification performance, certain limitations remain. First, the use of balanced undersampling reduces the overall dataset size, which may limit the model's ability to learn complex patterns present in the original data distribution. Second, the model has been evaluated only on the AWID-CLS-R dataset, and its performance on other real-world network environments may vary. Third, the deployment was tested on a Raspberry Pi under controlled conditions; large-scale real-time deployment may introduce latency and scalability challenges**.**

VII. FUTURE WORK

Although the proposed system achieves strong detection performance, several directions for future improvement are identified. First, the balanced undersampling strategy used in this work reduces the total training data from 1,795,575 to 181,059 samples. Future work will explore synthetic oversampling techniques such as SMOTE and ADASYN to generate artificial minority class samples instead of discarding majority class data, which may further improve model generalization. Second, the current 1D-CNN architecture will be compared against LSTM, Bi-LSTM, and Transformer-based architectures to evaluate whether temporal dependency modeling provides additional accuracy gains for sequential Wi-Fi packet data. Third, the proposed system was evaluated exclusively on the AWID-CLS-R reduced dataset. Future evaluation on the full AWID-CLS-F dataset and other benchmark datasets such as UNSW-NB15 and CIC-IDS2017 will assess the generalizability of the model across different network environments. Fourth, the current centralized training approach raises data privacy concerns when applied across multiple distributed WSN nodes. A federated learning framework will be explored to enable privacy-preserving collaborative IDS training without sharing raw packet data between nodes. Fifth, the Raspberry Pi 4 deployment will be further optimized using TensorFlow Lite model quantization to reduce inference latency and memory footprint for real-time edge deployment.

REFERENCES

[1]. Base Paper (Main Reference — Must Cite) H. Sadia, S. Farhan, Y. U. Haq, R. Sana, T. Mahmood, S. A. O. Bahaj, and A. R. Khan.
[2]. AWID Dataset Paper (Must Cite — your dataset) C. Kolias, G. Kambourakis, A. Stavrou, and S. Gritzalis.
[3]. SHAP Paper (Must Cite — your explainability method) S. M. Lundberg and S.-I. Lee.
[4]. SHAP + AWID Feature Selection (Directly Related) S. Bhandari, A. K. Kukreja, A. Lazar, A. Sim, and K. Wu.
[5]. Deep Learning IDS on AWID (Related Work) S. M. Kasongo and Y. Sun.
[6]. XAI-SHAP for Wi-Fi IDS (Supports your novelty) A. A. Reyes, F. D. Vaca, G. A. C. Aguayo, Q. Niyaz, and V. Devabhaktuni.

