Wi-Fi Network Intrusion Detecting using Multi-Layer Stacking Ensemble Learners

Jomari R. Edulan
Department of Computer Science, Caraga State University

Niel Sean Mikhael M. Gildo
Department of Computer Science, Caraga State University

Regien B. Nakazato
Department of Computer Science, Caraga State University

https://doi.org/10.5109/7395550

出版情報：Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES). 11, pp.415-420, 2025-10-30. International Exchange and Innovation Conference on Engineering & Sciences
バージョン：

Wi-Fi Network Intrusion Detecting using Multi-Layer Stacking Ensemble Learners

Jomari R. Edulan[1], Niel Sean Mikhael M. Gildo[1],Regien B. Nakazato[2]
[1]Student, Department of Computer Science, Caraga State University, Butuan City, Philippines
[2]Assistant Professor I, Department of Computer Science, Caraga State University, Butuan City, Philippines

Corresponding author email: rbnakazato@carsu.edu.ph

Abstract: *Wi-Fi networks are increasingly vulnerable to attacks like De-authentication, Evil Twin, Rogue Access Point, Beacon Flooding, and Krack Attacks, exploiting WPA2 and WPA3 protocol weaknesses, particularly in public hotspots vital to initiatives like the Philippines' Free Wi-Fi for All. This study introduces a Multi-Layer Stacking Ensemble model for Wi-Fi Network Intrusion Detection Systems (WNIDS) to enable robust, cross-protocol detection. The model incorporates Random Forest, Gradient Boosting, and LightGBM as base learners with a CatBoost meta-classifier. It was trained using the WPA3 dataset (Saini et al., 2022) and evaluated on a combined WPA3-AWID3 dataset. Preprocessing preserved important features while reducing the WPA3 dataset from 9.54 GB to 872 MB. On the combined dataset (100% Normal, 93% RogueAP), the model produced balanced recall and nearly flawless WPA3 attack detection performance (Accuracy: 0.9999, F1-Score: 0.999). Despite Gradient Boosting's superior accuracy (98.03%), its generality demonstrates usefulness for protecting a variety of Wi-Fi configurations.*

Keywords WNIDS, Multi-Layer Stacking, Wi-Fi Security, Cross-Protocol, Machine Learning

1. INTRODUCTION

Wi-Fi networks have become a cornerstone of global connectivity, enabling seamless wireless communication across billions of devices in homes, enterprises, and public spaces. With an estimated 20 billion Wi-Fi-enabled devices in use worldwide by 2023, Wi-Fi's ability to provide high-speed, wireless data transmission without the need for physical infrastructure has revolutionized the way information is shared and accessed. Leveraging this robust wireless capability, modern paging systems present a cost-effective and easily deployable alternative to traditional methods. These systems typically comprise strategically placed Access Points (APs) for signal distribution, centralized Paging Servers to manage user accounts and message delivery, and portable Paging Devices for reception. This wireless architecture not only ensures scalability and reduced deployment costs but also benefits from strong security protocols offered by modern Wi-Fi encryption technologies. This underscores the critical role of wireless infrastructure and design in enabling efficient, autonomous communication systems [7, 8]. However, their widespread adoption has made them prime targets for cyber-attacks, exploiting vulnerabilities in protocols like WPA2 and WPA3, especially in public hotspots where security is often lax [3]. In the Philippines, the Free Wi-Fi for All initiative, which activated 3,194 access points across 1,499 remote locations by October 2023, exemplifies the growing reliance on public Wi-Fi, yet also highlights the urgent need for robust security measures to protect users from escalating threats [1]. However, their widespread adoption has made them prime targets for cyber-attacks, exploiting vulnerabilities in protocols like WPA2 and WPA3, especially in public hotspots where security is often lax [3]. In the Philippines, the Free Wi-Fi for All initiative, which activated 3,194 access points across 1,499 remote locations by October 2023, exemplifies the growing reliance on public Wi-Fi, yet also highlights the urgent need for robust security measures to protect users from escalating threats [1]. Traditional intrusion detection systems struggle with the dynamic nature of Wi-Fi environments, particularly in mixed-protocol settings where WPA2 and WPA3 coexist, and imbalanced datasets where normal traffic vastly outnumbers attack samples [4]. This study addresses these challenges by leveraging advanced machine learning techniques, specifically multi-layer stacking ensembles, to develop a Wi-Fi Network Intrusion Detection System (WNIDS) capable of cross-protocol detection and robust performance across diverse settings. The following subsections delve into the specific vulnerabilities of Wi-Fi networks, the unique challenges in public hotspots, prior research on ML-based WNIDS, identified research gaps, and the objectives driving this study to enhance Wi-Fi security.

1.1 Wi-Fi Security Vulnerabilities

Wi-Fi networks, operating at the Data Link Layer (Layer 2) of the OSI model, are susceptible to a range of attacks due to their wireless nature and protocol vulnerabilities [3]. WPA2, introduced in 2004, relies on AES encryption but is prone to Key Reinstallation Attacks (KRACK), which exploit the 4-way handshake to decrypt traffic [3]. WPA3, released in 2018, improves security with Simultaneous Authentication of Equals (SAE) but remains vulnerable to Dragonblood attacks, including downgrade and side-channel exploits [3]. These vulnerabilities enable attacks like De-authentication, Evil Twin, Rogue Access Point, Beacon Flooding, and Krack, which compromise user data integrity and network availability, particularly in environments with open transmissions.

1.2 Challenges in Public Hotspots

Public Wi-Fi hotspots, such as those deployed under the Philippines' Free Wi-Fi for All initiative, which activated 3,194 access points across 1,499 remote locations by October 2023 [1], are especially vulnerable due to their accessibility. Blancaflor et al. (2023) demonstrated that even encrypted public Wi-Fi in the Philippines can be

compromised through De-authentication and man-in-the-middle attacks, exposing users to risks like data theft [4]. The open nature of these networks, combined with the coexistence of WPA2 and WPA3 protocols, amplifies the need for robust intrusion detection systems capable of operating across diverse settings.

1.3 Prior Work on ML-Based WNIDS

Machine learning (ML) has emerged as a powerful tool for enhancing Wi-Fi Network Intrusion Detection Systems (WNIDS), with several studies demonstrating its potential. Vaca and Niyaz (2018) applied ensemble models like Random Forest on the AWID dataset, achieving 99.096% accuracy in detecting WPA2-based attacks such as De-authentication and Evil Twin [5]. Their approach leveraged features like packet length and signal strength, but was limited to WPA2, lacking cross-protocol applicability. Saini et al. (2022) focused on WPA3, using Random Forest and Decision Trees on a WPA3 dataset, achieving up to 99.98% accuracy for attacks like Krack and Beacon Flooding [2]. However, they highlighted accuracy's inadequacy as a sole metric in imbalanced datasets, where normal traffic dominates, leading to overfitting on majority classes. Reyes et al. (2020) developed a two-stage ML model using AWID2, achieving 99.42% accuracy for impersonation attacks like Evil Twin, employing a combination of feature selection and ensemble methods [5]. Despite these advancements, most studies focus on single-protocol datasets (either WPA2 or WPA3), limiting their relevance in mixed-protocol environments, a gap this study aims to address [2, 5].

1.4 Research Gaps

Significant gaps remain in WNIDS research, particularly in addressing real-world complexities. Most studies, including those by Vaca and Niyaz (2018) and Saini et al. (2022), focus on either WPA2 or WPA3 datasets, failing to account for the coexistence of protocols in modern networks, such as public hotspots where devices may negotiate between WPA2 and WPA3 [2, 5]. This single-protocol focus limits model generalizability, as attacks often exploit protocol transitions, such as downgrade attacks in WPA3 [3]. Additionally, imbalanced datasets pose a challenge, with normal traffic often comprising over 90% of samples, leading to biased models that prioritize majority classes while misclassifying rare but critical attack types like Evil Twin or RogueAP [4]. Blancaflor et al. (2023) noted that such imbalances result in high false negatives, undermining detection reliability in public Wi-Fi settings [4]. Furthermore, the potential of multi-layer stacking ensembles, which hierarchically combine diverse ML models to improve robustness, remains underexplored in WNIDS, offering a novel opportunity to enhance cross-protocol detection and address these limitations [2].

1.5 Objectives of the Study

This study aims to designing a high-performing and generalizable Wi-Fi Network Intrusion Detection System (WNIDS) through the implementation of a Multi-Layer Stacking Ensemble model, integrating Random Forest, Gradient Boosting, and LightGBM as base learners, with a CatBoost meta-classifier to synthesize predictions, aimed at enhancing security across diverse Wi-Fi environments, including home, enterprise, and public settings such as those supported by the Philippines' Free Wi-Fi for All initiative [1]. The primary objectives include training the model on the WPA3 dataset to capture modern Wi-Fi traffic patterns and testing its generalization on a combined WPA3-AWID3 dataset, simulating mixed-protocol environments, to ensure robust cross-protocol detection [2, 6]. To achieve this, the study specifically seeks to design and rigorously test the Multi-Layer Stacking Ensemble model, evaluating its effectiveness compared to Bagging and Boosting classifiers by employing established metrics such as F1-Score, Accuracy, Precision, and Recall, while utilizing confusion matrices to comprehensively assess performance and mitigate the risk of overfitting, particularly in the context of imbalanced datasets prevalent in Wi-Fi traffic analysis [4]. Performance will be evaluated with an emphasis on robustness against minority classes like Evil Twin, aiming to provide a scalable solution for real-world deployment that addresses the limitations of single-protocol models and enhances security for users and administrators in dynamic Wi-Fi networks [2].

2. RESOURCE AND DATA

The effectiveness of a Wi-Fi Network Intrusion Detection System (WNIDS) hinges on the quality and diversity of datasets, the nature of captured attacks, and the computational resources utilized for model development and evaluation. The selection of datasets like WPA3 and AWID3 in this study ensures comprehensive coverage of both modern and legacy Wi-Fi protocols, addressing the need for cross-protocol robustness highlighted in prior research [2, 5, 6]. Hardware and software resources were chosen to balance computational efficiency with the demands of processing large-scale datasets, a challenge often noted in WNIDS literature due to the voluminous nature of network traffic data [4]. This section reviews the characteristics of the datasets used, the attack types they cover, and the hardware and software tools employed, drawing on related studies to contextualize the resources within the broader landscape of Wi-Fi security research.

2.1 WPA3 Dataset Characteristics

The WPA3 dataset by Saini et al. (2022) was used for training, capturing Wi-Fi traffic at the 2.4 GHz frequency [2]. It includes 250 attributes derived from packet captures using devices like a Linksys E8450 access point, a D-link DWA-X1850 Wi-Fi adapter, and a Netgear A6210 for monitoring. Client devices included a Samsung A7 tablet, MacBook Air, and HP laptop, ensuring diverse traffic patterns. The raw dataset size was 9.54 GB, later reduced to 872 MB through preprocessing.

2.2 AWID3 Dataset for Cross-Protocol Testing

The AWID3 dataset, an updated iteration of AWID2, was employed for cross-protocol evaluation, capturing WPA2-based attacks such as De-authentication, Evil Twin, and Rogue Access Point [6]. Developed by Chatzoglou et al. (2021), AWID3 serves as a benchmark for legacy Wi-Fi protocols, requiring preprocessing to align features like frame.len with the WPA3 dataset for consistent testing across protocols. The AWID family, including AWID and AWID2, has been widely used in

WNIDS research, with Vaca and Niyaz (2018) achieving 99.096% accuracy using Random Forest on AWID for WPA2 attacks [5]. Reyes et al. (2020) also utilized AWID2, reporting 99.42% accuracy for impersonation attacks, highlighting the dataset's relevance for evaluating impersonation-based threats like Evil Twin. However, AWID3's inclusion of more recent attack types and its focus on WPA2 make it particularly suitable for cross-protocol testing, addressing a gap in prior studies that often overlooked mixed-protocol environments [5, 6]. This dataset's comprehensive feature set and attack diversity provide a robust foundation for validating model generalization.

2.3 Attack Types

The WPA3 and AWID3 datasets collectively cover key attack types prevalent in Wi-Fi networks: De-authentication, Evil Twin, Rogue Access Point, Beacon Flooding, and Krack, all of which target Layer 2 vulnerabilities [2, 6]. De-authentication attacks forcibly disconnect users by spoofing management frames, while Evil Twin attacks impersonate legitimate networks to steal credentials, a threat Blancaflor et al. (2023) identified as rampant in Philippine public Wi-Fi [4]. Rogue Access Points introduce unauthorized entry points, and Beacon Flooding overwhelms clients with fake signals, both documented in AWID3 [6]. Krack, as detailed by Vanhoef and Piessens (2017), exploits WPA2's 4-way handshake, enabling traffic decryption [3]. These attacks reflect real-world threats, with studies like Vaca and Niyaz (2018) noting their prevalence in WPA2 networks and Saini et al. (2022) extending this to WPA3 vulnerabilities [2, 5]. The datasets' focus on these attacks ensures the WNIDS model is trained on scenarios mirroring those in public hotspots, enhancing its practical relevance [1].

2.4 Hardware Specifications

The study was conducted on an Intel Core i5 12450H laptop with 16GB DDR5 RAM, providing sufficient computational power for processing large datasets and training multiple ML models. The Intel Core i5 12450H, a 12th-generation processor, offers 8 cores (4 performance, 4 efficiency) with a max turbo frequency of 4.4 GHz, ensuring efficient handling of data-intensive tasks like preprocessing and model training.

2.5 Software Tools and Libraries

Python served as the primary programming language, leveraging scikit-learn for ensemble methods (Random Forest, Gradient Boosting), CatBoost for the meta-classifier, and LightGBM for boosting. Dask and pandas were used for data preprocessing, efficiently handling the large WPA3 dataset, while joblib facilitated model serialization for reuse during evaluation, ensuring consistency and computational efficiency. These tools align with those used in prior WNIDS research; for instance, Saini et al. (2022) also employed scikit-learn for Random Forest and Decision Trees on WPA3 data, achieving high accuracy [2]. Vaca and Niyaz (2018) utilized scikit-learn for ensemble models on AWID, supplemented by WEKA for additional analysis, a tool not used here due to its redundancy with Python-based libraries [5]. Additionally, Chatzoglou et al. (2021) used Wireshark for AWID3 packet capture, a tool that could complement preprocessing but was substituted here with pandas for its integration with ML workflows [6]. The combination of Dask and joblib addresses scalability and reproducibility, critical for handling large-scale Wi-Fi traffic data as noted in related studies [2, 5].

3. APPROACH AND METHODOLOGY

Developing a robust WNIDS requires a systematic approach to data preparation, model selection, architecture design, and evaluation. This study employed a multi-stage methodology to preprocess datasets, train diverse classifiers, and evaluate their performance across protocols. The Multi-Layer Stacking Ensemble model was designed to leverage the strengths of individual learners, ensuring adaptability to varied Wi-Fi environments. The following subsections outline the preprocessing steps, classifier selection, stacking architecture, evaluation metrics, and cross-protocol testing strategy.

3.1 Preprocessing Steps for WPA3 Dataset

The WPA3 dataset by Saini et al. (2022) was used for training, capturing Wi-Fi traffic at the 2.4 GHz frequency [2]. It includes 250 attributes derived from packet captures using devices like a Linksys E8450 access point, a D-link DWA-X1850 Wi-Fi adapter, and a Netgear A6210 for monitoring. Client devices included a Samsung A7 tablet, MacBook Air, and HP laptop, ensuring diverse traffic patterns. The raw dataset size was 9.54 GB, later reduced to 872 MB through preprocessing.

3.2 Classifier Selection and Training

Seven classifiers were trained: Random Forest, Extra Trees, CatBoost, AdaBoost, Gradient Boosting, LightGBM, and the proposed Multi-Layer Stacking Ensemble. All were trained on the WPA3 training set using default hyperparameters to establish a baseline. Random Forest and Extra Trees leverage bagging, while CatBoost, AdaBoost, Gradient Boosting, and LightGBM use boosting techniques. Models were saved as .pkl files using joblib for reuse, ensuring consistency across evaluation phases.

3.3 Multi-Layer Stacking Architecture

The stacking model combined Random Forest, Gradient Boosting, and LightGBM as base learners, with a CatBoost meta-classifier. Base learners were loaded from pre-trained .pkl files, and their predictions were used as meta-features for the meta-classifier. The stacking architecture, implemented via scikit-learn's StackingClassifier with passthrough=True, allowed original features to be passed to the meta-classifier, enhancing its ability to capture complex patterns across diverse Wi-Fi traffic.

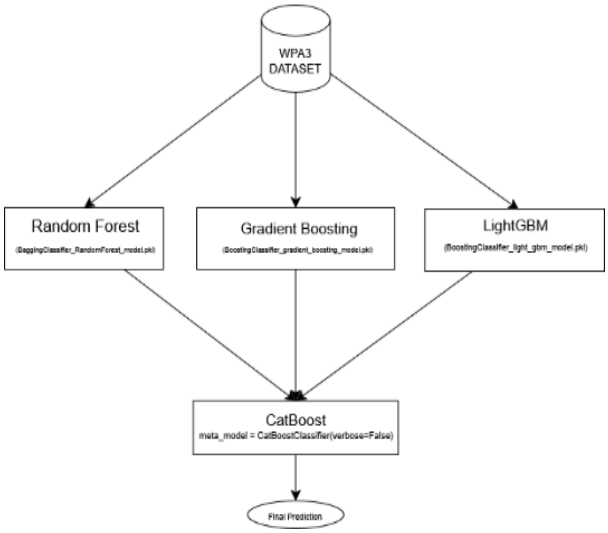

Fig. 1. Architecture of the Multi-Layer Stacking Ensemble Model.

3.4 Evaluation Metrics

To assess the effectiveness of each classifier in identifying various types of Wi-Fi network intrusions, this study utilized commonly used classification metrics. The metrics included Accuracy, Precision, Recall, and F1-Score. Accuracy is defined as the ratio of correctly predicted samples to the total number of samples and is used in general evaluation when all classes are of equal importance. However, in the domain of network intrusion detection, where false positives and false negatives have serious implications, F1-Score was used as the primary metric for comparison. The F1-Score provides a balanced measure by calculating the harmonic mean of Precision and Recall. Precision and Recall are informative on their own, the F1-Score integrates them into a single metric, which is more effective when dealing with imbalanced datasets or critical error sensitivity, as is common in intrusion detection systems. All models were initially tested on the WPA3 dataset and then on the combined WPA3 and AWID3 dataset to evaluate whether they regained performance when faced with data distributions from another Wi-Fi standard. The consistency of the F1-Score across both testing conditions served as the primary indication of a model's robustness and ability to generalize.

$$\text{Precision} = TP \times (TP + FP)^{-1}$$
$$\text{Recall} = TP \times (TP + FN)^{-1}$$
$$\text{Accuracy} = (TP + TN) \times (TP + TN + FP + FN)^{-1}$$
$$F1 = 2 \times (\text{Precision} \times \text{Recall}) \times (\text{Precision} + \text{Recall})^{-1}$$

Fig. 2. Formula for the Evaluation Metrics.

3.5 Cross Protocol Testing with AWID3

For cross-protocol evaluation, the AWID3 dataset was preprocessed to align with WPA3 [6]. Unique WPA2-specific features were removed, and normal traffic was downscaled to balance classes. Common features were retained, ensuring compatibility. The combined WPA3-AWID3 dataset simulated real-world mixed-protocol environments, testing the models' ability to generalize across WPA2 and WPA3 settings.

4. RESULTS AND DISCUSSIONS

The evaluation of the Multi-Layer Stacking Ensemble model and other classifiers revealed significant insights into their performance across WPA3 and mixed-protocol environments. The stacking model demonstrated exceptional robustness, particularly in handling imbalanced datasets and generalizing across protocols, making it a promising solution for real-world Wi-Fi security challenges.

4.1 Initial Evaluation on WPA3 Dataset

The Multi-Layer Stacking Ensemble model achieved near-perfect performance on the WPA3 test set (Accuracy: 0.9999, F1-Score: 0.9999), matching Random Forest, Extra Trees, and CatBoost (all 1.000 across metrics). Gradient Boosting scored lower (Accuracy: 0.9994, F1-Score: 0.9698), misclassifying 48 Evil Twin and 98 Krack samples, while AdaBoost (Accuracy: 0.7944, F1-Score: 0.3584) and LightGBM (Accuracy: 0.9448, F1-Score: 0.5941) struggled with imbalanced classes, showing significant misclassifications for minority attacks [2].

Table 1. Classifier Performance – WPA3 Dataset

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	1.0000	1.0000	1.0000	1.0000
Extra Trees	1.0000	1.0000	1.0000	1.0000
CatBoost	1.0000	1.0000	1.0000	1.0000
Gradient Boosting	0.9994	0.9557	0.9886	0.9698
AdaBoost	0.7944	0.3240	0.4653	0.3584
LightGBM	0.9448	0.6187	0.5777	0.5941
Multi-Layer Stacking Model	1.0000	1.0000	1.0000	1.0000

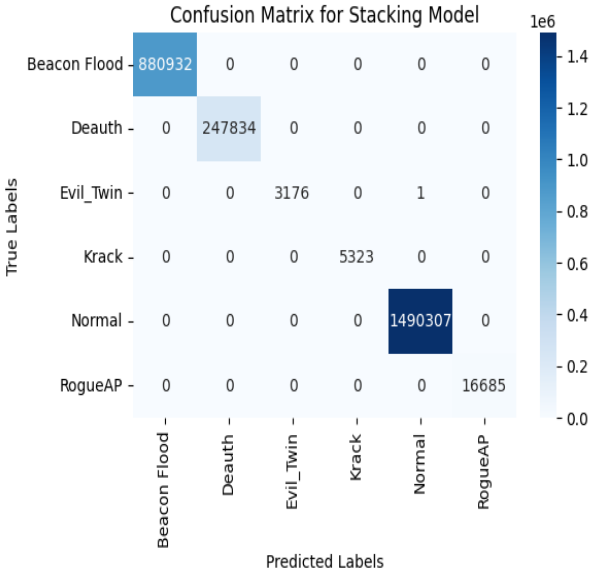

Fig. 2. Multi-Layer Stacking Model Confusion Matrix on WPA3 Dataset

4.2 Cross-Protocol Performance on Combined Dataset

On the combined WPA3-AWID3 dataset, Gradient Boosting achieved the highest accuracy (98.03%), with recalls of 90% for Evil Twin and 93% for RogueAP. The stacking model followed with 96.18% accuracy, excelling in balanced recall (100% Normal, 93% RogueAP, 30% Evil Twin), demonstrating robust cross-protocol generalization. CatBoost scored 96.54% accuracy, while Random Forest and Extra Trees hovered around 95.5%. AdaBoost (75.63%) and LightGBM (89.07%) showed significant performance drops, failing on minority classes like Evil Twin and RogueAP [6].

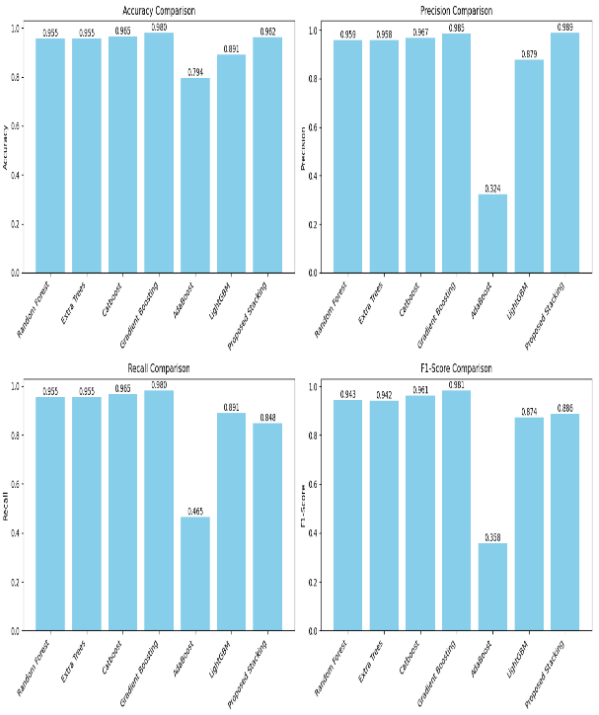

Fig. 3. Model Performance Comparison on Combined Dataset

Table 2. Classifier Performance – Combined WPA3-AWID Dataset

Model	Accuracy	Precision	Recall (Normal)	Recall (Evil Twin)
Random Forest	0.9550	1.0000	0.9000	0.2100
Extra Trees	0.9550	1.0000	0.9100	0.1000
CatBoost	0.9654	1.0000	0.9200	0.3000
Gradient Boosting	0.9803	1.0000	0.9300	0.9000
AdaBoost	0.7563	0.8200	0.7000	0.0000
LightGBM	0.8907	0.9500	0.8500	0.1000
Multi-Layer Stacking Model	0.9618	1.0000	0.9300	0.3000

4.3 Comparative Analysis of Classifiers

The stacking model's balanced performance across protocols highlights its advantage over single-method classifiers. Random Forest and Extra Trees, while effective on WPA3 (1.0000 F1-Score), showed lower Evil Twin recall (21.00% and 10.00%, respectively) on the combined dataset, indicating sensitivity to distribution shifts. Gradient Boosting's high accuracy (98.03%) was offset by less balanced recall compared to the stacking model, which maintained consistency across classes. AdaBoost's poor performance (75.63% accuracy) underscores its limitations with imbalanced data, while LightGBM's moderate scores (89.07 accuracy) suggest sensitivity to noise in rare attack classes [5].

4.4 Practical Implications

The stacking model's ability to generalize across WPA2 and WPA3 protocols makes it suitable for real-world deployment in diverse Wi-Fi settings, such as public hotspots under the Free Wi-Fi for All initiative [1]. Its high recall for Normal (100.0) and RogueAP (93.00) traffic ensures minimal false positives, critical for user experience, while its robustness to protocol variations addresses the dynamic nature of modern Wi-Fi networks [4]. This approach enhances security for users, administrators, and policymakers reliant on public Wi-Fi infrastructure.

5. SUMMARY

This study aimed to design a high performing, generalizable Wi-Fi Network Intrusion Detection System (WNIDS) using a Multi-Layer Stacking Ensemble model to enhance security across diverse Wi-Fi environments, including homes, enterprises, and public hotspots like those in the Philippines' Free Wi-Fi for All program. The research addressed vulnerabilities in the WPA2 and WPA3 protocols exploited by attacks such as De-authentication, Evil Twin, Rogue Access Point, Beacon

Flooding, and Krack, which threaten user data and network integrity. Existing WNIDS models often lack cross-protocol generalization and struggle with imbalanced datasets, where normal traffic overshadows attack samples. To tackle these issues, the study utilized the WPA3 dataset (Saini et al., 2022) for training and a combined WPA3-AWID3 dataset for testing, simulating real-world mixed-protocol scenarios.

The methodology involved preprocessing the WPA3 dataset (originally 9.54 GB, reduced to 872 MB) by removing features with over 90% missing values, eliminating zero-variance features, and standardizing data. Multiple classifiers; Random Forest, Extra Trees, CatBoost, AdaBoost, Gradient Boosting, LightGBM, and the proposed multi-layer stacking model, were trained with default hyperparameters on the WPA3 training set (67% of the data). The stacking model combined Random Forest, Gradient Boosting, and LightGBM as base learners with a CatBoost meta-classifier. Initial evaluation on the WPA3 test set showed near-perfect performance for the multi-layer stacking model, Random Forest and CatBoost, while Gradient Boosting, AdaBoost, and LightGBM had lower scores. Final evaluation on the combined WPA3-AWID3 dataset revealed Gradient Boosting with the highest accuracy (98.03%), but the stacking model excelled in balanced recall (100% Normal, 93% RogueAP, 30% Evil Twin), demonstrating robustness across protocols.

5.1 Conclusion

The Multi-Layer Stacking Ensemble model emerged as a robust and effective WNIDS solution, achieving exceptional performance on the WPA3 dataset and strong generalization on the combined WPA3-AWID3 dataset. Its architecture, leveraging diverse base learners and a dynamically trained meta-classifier, enabled it to detect complex attack patterns and adapt to mixed-protocol environments, addressing a key gap in prior single-protocol studies. The model's near-perfect metrics on WPA3 and balanced recall on the combined dataset (e.g., 100% Normal, 93% RogueAP) highlight its ability to handle imbalanced data and minimize false positives, critical for real-world deployment. While Gradient Boosting achieved the highest accuracy (98.03%) on the combined dataset, the multi-layer stacking model's stability and consistency across classes make it a practical choice for securing diverse Wi-Fi settings. The use of default hyperparameters provided a solid baseline, indicating potential for further improvement with optimization. This study validates that cross-protocol robustness is achievable through feature alignment, class balancing, and ensemble techniques, advancing WNIDS research beyond the limitations of prior work focused on WPA2 (e.g., AWID3) o WPA3 alone. The stacking model's success underscores the value of integrating multiple learning strategies to enhance detection in dynamic, threat-prone Wi-Fi networks.

5.2 Recommendations

Based on the findings of this study, several steps are proposed to refine and extend the capabilities of the Wi-Fi Network Intrusion Detection System (WNIDS) model. Future research should prioritize hyperparameter optimization by conducting systematic tuning of both the base learners and the meta-classifier to enhance detection rates, particularly for minority classes such as Evil Twin, which achieved a recall of 30% on the combined dataset, thereby potentially improving overall performance. Additionally, expanding the training dataset to incorporate emerging attack types and additional protocols would ensure the model remains effective against evolving threats and diverse network configuration. Investigating the deployment of the stacking model in real-time WNIDS frameworks is also recommended to test its scalability and responsiveness in dynamic environments like public hotspots. Furthermore, applying advanced feature selection techniques to eliminate redundant features, such as highly correlated pairs, while retaining critical Layer 2 attributes, would enhance efficiency without sacrificing accuracy. Finally, implementing k-fold cross-validation during training is advised to further validate the model's stability and reduce the risk of overfitting, thereby strengthening its reliability for practical deployment in real-world scenarios. These recommendations aim to build on the study's contribution, fostering the development of adaptive and scalable WNIDS solutions to address modern Wi-Fi security challenges.

6. REFERENCES

[1] DICT Activates Over 3,900 New Free Wi-Fi Sites Nationwide - Free Wi-Fi for All, n.d.

[2] S. Saini, et al., WPA3 Dataset for Wi-Fi Intrusion Detection, 2022.

[3] M. Vanhoef, F. Piessens, Key Reinstallation Attacks: Forcing Nonce Reuse in WPA2, Proc. ACM CCS, 2017, pp. 1313–1328.

[4] E. Blancaflor, et al., Security Vulnerabilities of Encrypted Public Wi-Fi in the Philippines, 2023.

[5] D. Vaca, Q. Niyaz, Ensemble Models for Wi-Fi Intrusion Detection on AWID Dataset, 2018.

[6] E. Chatzoglou, et al., AWID3: A Comprehensive Dataset for Wi-Fi Intrusion Detection, 2021.

[7] C. Dispo, A. Balamad, R. Mendoza, A. Demetillo, Enhanced Campus Communication Hub Through Wi-Fi Optimization and Emergency Earthquake Notifier (ECHOER), Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES), vol. 10, pp. 514-519, 2024, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University.

[8] K. Ali, Energy-Harvesting-Based Wireless Sensor Nodes for Inaccessible Environments, Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES), vol. 7, pp. 34-35, 2021, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University.