Hybrid Feature-Aware Stacking for Intrusion Detection in 5G eMBB Networks

Md. Monirul Islam, and Seong Ho Jeong*
Department of Information and Communications Engineering
Hankuk University of Foreign Studies
Seoul, South Korea
monirul@hufs.ac.kr, and shjeong@hufs.ac.kr*

***Abstract*—Intrusion detection in 5G networks is increasingly challenged by high-volume traffic, slice-specific behaviors, and real-time latency constraints. Existing Intrusion Detection System (IDS) studies primarily focus on legacy datasets and conventional classifiers, lacking analysis on emerging 5G traffic and fusion-level ensemble behavior. This paper presents a hybrid feature–aware stacking framework in the eMBB slice of the 5G-SliciNdd dataset. A three-stage hybrid feature selection method named mutual information, variance thresholding, and random importance is used to retain features contributing to 90% cumulative importance, enabling dimensionality reduction without accuracy loss. We analyze three fusion strategies for stacked learning: (i) hard-label fusion, (ii) probability-level fusion, and (iii) a hybrid fusion combining both outputs. Eight traditional machine learning (ML) models are benchmarked, and the 5 top performers are integrated into the stacked ensemble. Experiments show that the proposed framework achieves 99.14% accuracy with an ultra-low inference delay of 0.0019 ms per sample, demonstrating its suitability for real-time IDS in 5G edge environments. Unlike prior IDS studies, this work provides the first comparative fusion analysis on 5G-SliciNdd traffic, highlighting the benefits of feature-aware stacking for high-speed networks.**

Keywords—Stacking ensemble learning, intrusion detection system, feature selection, meta-learner, 5G.

I. Introduction

The increasing adoption of 5G networks introduces new challenges in securing high-speed traffic, particularly in enhanced Mobile Broadband (eMBB) slices where high throughput and low latency coexist. Conventional intrusion detection systems (IDS) relying on signature matching or single-model classifiers struggle under these conditions due to high-dimensional feature spaces, evolving attack patterns, and real-time inference demands [1].

Most prior IDS studies evaluate classical ML or DL models on traditional datasets such as NSL-KDD, UNSW-NB15, and CICIDS-2017, limiting their relevance to modern 5G environments. These datasets do not capture slice-specific behaviors, flow structures, or the protocol-level nuances present in 5G service-based architectures. Furthermore, existing ensemble-based IDS typically employ a single stacking strategy without analyzing the fusion behavior of different prediction types (labels vs probabilities) [2], [3]. This limit understanding of how ensemble fusion affects robustness, latency, and predictive stability in high-speed 5G workloads.

To address these gaps, we propose a hybrid feature–aware stacking framework tailored in 5G eMBB traffic. This study delivers the first fusion-level comparative analysis on 5G-SliciNdd, evaluating hard-label, probability-based, and hybrid stacking strategies. A hybrid feature-selection pipeline combining mutual information, variance filtering, and tree-based importance with a 90% cumulative selection rule is designed specifically for eMBB traffic. The framework achieves an ultra-low inference time of 0.0019 ms per sample, demonstrating its suitability for microsecond-level 5G edge deployments. We further benchmark eight traditional ML models on the eMBB slice, establishing a new baseline for IDS research on 5G-SliciNdd. Overall, results confirm improved robustness, as stacking consistently reduces error variance across base models even when its accuracy does not surpass the top-performing individual classifier.

The rest of the paper is structured as follows: Section II reviews related work; Section III outlines methodology; Section IV presents performance analysis; and Section V concludes with future directions.

II. Related Work

Recent studies have extensively explored hybrid feature selection and stacked ensemble learning for intrusion detection. In [4], the authors introduced MI-Boruta, a filter–wrapper hybrid integrating mutual information with Boruta, combined with a stacked ensemble of Random Forest (RF), CatBoost, and XGBoost, with an Multi-Layer Perceptron (MLP) meta-learner. Evaluated on UNSW-NB15 and CICIDS2017, the model achieved 95.34% and 99.92% accuracy. In [5], the authors proposed IGRF-RFE, merging Information Gain, Random Forest filtering, and RFE, reducing UNSW-NB15 features from 42 to 23 and improving MLP accuracy from 82.25% to 84.24%.

In IoT anomaly detection, the study [6] developed a hybrid IDS using 5 classical ML obtaining 92.52% and 92.92% accuracy on CICIoT2023 and Aposemat IoT-23 datasets. In [7], the authors proposed a general stack-ensemble IDS combining multiple base learners into a meta-model, achieving 98.7% accuracy across heterogeneous traffic and log datasets.

Ensemble models combining RF, Decision Tree (DT), and K-Nearest Neighbors (KNN) with a Logistic Regression meta-learner were explored in [8], obtaining 96.16% training and 97.95% testing accuracy on UNSW-NB15. In [9], the authors evaluated a meta-classification-based stacking IDS on UNSW-NB15 and UGR'16, reporting 97% accuracy on real-time data and 94% on emulated traffic. In [10], the authors proposed a deep learning based intrusion detection based on 5G-SliciNdd datasets achieving 98.50% accuracy.

III. Proposed Methodology

Fig. 1 illustrates the overall workflow of the proposed intrusion detection framework. The process begins with the eMBB slice of the 5G-SliciNdd dataset, followed by preprocessing that includes categorical encoding, missing-value handling, and train–test splitting. A hybrid feature-selection pipeline combining mutual information, variance thresholding, and tree-based importance is then applied, retaining features contributing to 90% cumulative importance.

Eight ML classifiers are trained on the selected features and evaluated using 3-fold cross-validation, after which the top five models are selected for stacking. The stacking layer implements hard-label, probability-level, and hybrid fusion strategies to generate the final normal/attack classification.

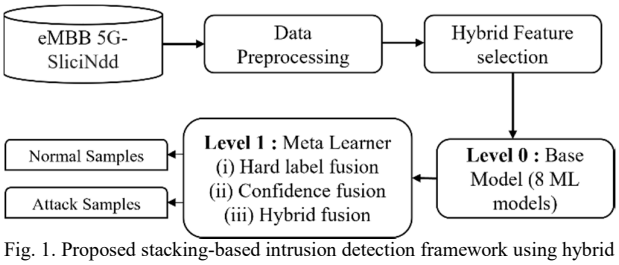


Fig. 1. Proposed stacking-based intrusion detection framework using hybrid feature selection on 5G-SliciNdd data

A. Dataset and Preprocessing

The study uses flow-level eMBB traffic from the 5G-SliciNdd dataset, containing 52 features that capture temporal, statistical, and protocol-level characteristics of normal and malicious flows [11]. Categorical attributes (Proto, sDSb, dDSb, Cause, State, sVid, dVid, Label) are transformed using label encoding, and redundant columns—such as previously generated prediction fields—are removed to avoid data leakage. Numerical fields are checked for missing or infinite values; infinities are converted to NaN and imputed with zero. The data is then split into 80:20 training and testing sets using stratified sampling to preserve class balance.

B. Feature Selection

A hybrid feature-selection strategy is used to reduce dimensionality and enhance model efficiency. It combines mutual information to capture nonlinear feature–target dependency, variance thresholding to remove low-variance attributes, and tree-based importance to assess impurity reduction contributions. The normalized scores from all three methods are aggregated and ranked, and features contributing to 90% cumulative importance are retained to balance compactness and relevance. Fig. 2 shows the selected features.

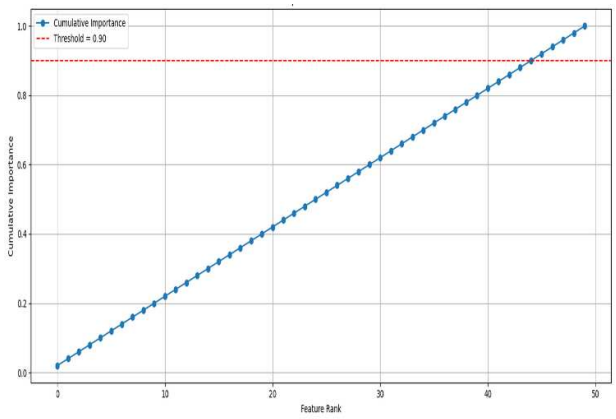


Fig. 2. Cumulative normalized feature importance on eMBB slice of the 5G-SliciNdd dataset

C. Base Model Learning

Eight supervised ML models: DT, RF, Extra Trees (ET), XGBoost (XGB), LightGBM (LGBM), CatBoost (CAT), and AdaBoost (ADA) are trained on the selected feature set to form the base learner pool, leveraging their complementary strengths in non-linearity handling, robustness, and interpretability. Performance is evaluated using 3-fold cross-validation, with mean accuracy serving as the selection metric. The top five models (RF, ET, CAT, LGBM, XGB) are chosen for stacking to control ensemble complexity while maintaining high predictive power.

D. Stacking Ensemble Learning

To improve performance beyond individual models, a stacking ensemble is employed in which outputs from level-0 base models serve as inputs to a level-1 meta-learner. Three variants are examined: hard-label fusion using discrete predictions, confidence fusion using probability scores, and hybrid fusion combining both to capture decision outcomes and associated uncertainty. XGB is used as the meta-learner due to its scalability and strong ability to model complex interactions.

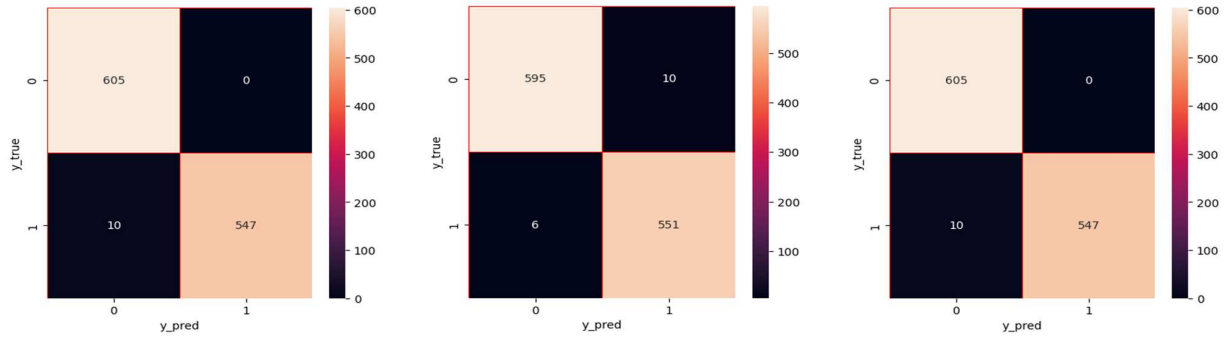
IV. Performance Analysis

The proposed intrusion detection framework is implemented in Python 3.10 using the Google Colaboratory environment, leveraging its cloud-based GPU runtime. Experiments are performed on the eMBB slice of the 5G-SliciNdd dataset, and performance is evaluated using accuracy (Acc), precision (Pre), recall (Re), F1-score (F1), training time (TT), and testing time per sample (TTPS). Confusion matrices are also examined to provide detailed insight into classification correctness for both the base models and the stacked ensemble.

TABLE I. Base Model Performance Comparison

Model Name	*Acc (%)*	*Pre (%)*	*Re (%)*	*F1 (%)*	*TT(s)*	*Test Time PS (ms)*
RF	99.14	99.15	99.14	99.14	0.9905	0.0134
ET	99.14	99.15	99.14	99.14	0.6213	0.0261
CAT	99.14	99.15	99.14	99.14	10.128	0.0053
LGBM	99.40	99.40	99.40	99.40	0.5539	0.0108
XGB	99.23	99.23	99.23	99.23	1.5596	0.0225

Table 1 compares the performance of the five top base classifiers: RF, ET, CAT, LGBM, and XGBoost using Acc, Pre, Re, F1, TT, and TTPS. All models exhibit strong classification capability, achieving above 99.13% accuracy. LGBM delivers the best results overall, attaining 99.40% accuracy along with the highest precision, recall, and F1-score. XGB performs competitively at 99.23% accuracy, though with slightly higher training and inference times. CAT matches the accuracy of RF and ET but records the highest training time (10.12 s), potentially limiting its applicability in latency-sensitive scenarios. In terms of computational efficiency, LGBM shows the fastest training time (0.55 s) and a low TTPS (0.0108 ms), making it an excellent choice for rapid and accurate intrusion detection. Overall, the consistently strong performance of these models supports their integration into a stacking ensemble to enhance robustness and generalization.



(a) (b) (c)

Fig. 3. Confusion matrices for stacked of (a) hard fusion, (b) confidence fusion, and (c) hybrid fusion

Fig. 3 shows the confusion matrices of the stacked ensemble model, where class 0 represents normal traffic and 1 denotes attacks. The hybrid approach correctly classified 605 normal and 547 attack samples, with only 10 attacks misclassified and no false alarms on normal traffic; the traditional approach produced identical results. The confidence-based method correctly identified 595 normal and 551 attack samples, with 16 attacks misclassified and again no normal samples labeled as attacks. The strong diagonal dominance across all matrices confirms reliable separation of normal and malicious flows, highlighting the robustness of the preprocessing and feature-selection pipeline.

TABLE II. STACKED MECHANISM PERFORMANCE COMPARISON

Stacked Mechanism	*Acc (%)*	*Pre (%)*	*Re (%)*	*F1 (%)*	*TT(s)*	*TTPS (ms)*
Hard	99.14	99.15	99.14	99.14	0.0309	0.0019
Confisdence	98.62	98.63	98.62	98.62	0.1761	0.0062
Hybrid	99.14	99.15	99.14	99.14	0.4756	0.0147

Table II further compares the three stacking strategies using Acc, Pre, Re, F1, TT, and TTPS. Hard fusion provides the best overall balance, achieving 99.1394% accuracy, strong precision/recall/F1, the lowest training time (0.0309 s), and the fastest inference (0.0019 ms/sample). Hybrid fusion attains identical predictive performance but requires higher computation (0.4756 s training, 0.01473 ms/sample inference), making it suitable when accuracy outweighs efficiency. Confidence fusion shows slightly lower accuracy (98.6231%) but maintains moderate cost and offers probabilistic interpretability. Overall, hard and hybrid fusion outperform the confidence method, with hard fusion emerging as the most efficient; in rich feature spaces like eMBB, ensembles mainly improve robustness and variance reduction rather than raw accuracy.

V. CONCLUSION

This paper presented a hybrid feature aware stacking framework for intrusion detection in 5G eMBB networks. We demonstrated a three-stage hybrid feature selection pipeline, evaluated three fusion mechanisms for stacked learning, and established new performance baselines on the 5G-SliciNdd dataset. The system achieves microsecond-level inference, making it suitable for real-time deployment at 5G edge nodes. Future work will explore slice-adaptive learning and online models for dynamic 5G environments.

ACKNOWLEDGMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (RS-2024-00436887) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation). This work was supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE)(P0028596).

REFERENCES

[1] *Recommendation ITU-R M.2160-0 (11/2023)*, Framework and overall objectives of the future development of IMT for 2030 and beyond.

[2] A. G. Ayad, N. A. Sakr, and N. A. Hikal, "A hybrid approach for efficient feature selection in anomaly intrusion detection for IoT networks.," *J. Supercomput.*, vol. 80, no. 19, pp. 26942–26984, Dec. 2024, doi: 10.1007/s11227-024-06409-x.

[3] "A hybrid feature selection and aggregation strategy-based stacking ensemble technique for network intrusion detection | Applied Intelligence.

[4] A. M. Alsaffar, M. Nouri-Baygi, and H. M. Zolbanin, "Shielding networks: enhancing intrusion detection with hybrid feature selection and stack ensemble learning," *J. Big Data*, vol. 11, no. 1, p. 133, Sept. 2024, doi: 10.1186/s40537-024-00994-7.

[5] "IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset | Journal of Big Data | Full Text." Accessed: Nov. 17, 2025.

[6] H. B. Akande, A. L. Imoize, T. C. Adeniran, C.-C. Lee, and J. B. Awotunde, "RF-FLIDS: A Novel Hybrid Intrusion Detection Model for Enhanced Anomaly Detection in IoT Networks," *Secur. Priv.*, vol. 8, no. 3, p. e70041, 2025, doi: 10.1002/spy2.70041.

[7] V. Jain, A. Sharma, and R. Malik, "Cybersecurity Intrusion Detection Techniques Using Stack Ensemble Machine Learning," in *2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS)*, Oct. 2024, pp. 1–6. doi: 10.1109/ICBDS61829.2024.10837054.

[8] A. Almomani *et al.*, "Ensemble-Based Approach for Efficient Intrusion Detection in Network Traffic," *Intell. Autom. Soft Comput.*, vol. 37, no. 2, pp. 2499–2517, 2023, doi: 10.32604/iasc.2023.039687.

[9] S. Rajagopal, P. P. Kundapur, and K. S. Hareesha, "A Stacking Ensemble for Network Intrusion Detection Using Heterogeneous Datasets," *Secur. Commun. Netw.*, vol. 2020, no. 1, p. 4586875, 2020, doi: 10.1155/2020/4586875.

[10] V. K. Gugueoth, "Enhanced Security Attack Detection and Prevention in 5G Networks Using CD-GELU-CNN and FMLRQC with HDFS-ECH-KMEANS," in *2024 8th International Conference on Computer, Software and Modeling (ICCSM)*, July 2024, pp. 36–43. doi: 10.1109/ICCSM63823.2024.00015.

[11] S. S. Samarakoon *et al.*, "5G-NIDD: A Comprehensive Network Intrusion Detection Dataset Generated over 5G Wireless Network." IEEE DataPort. doi: 10.21227/XTEP-HV36.