International Journal for Electronic Crime Investigation

ISSN: 2522-3429 (Print)
ISSN: 2616-6003 (Online)

DOI: https://doi.org/10.54692/ijeci.2026.1001/269

Research Article

Vol. 10 issue 1 Jan-Jul 2025

DeepShield-Med: A Machine Learning-Based Intrusion Detection System for Securing Internet of Medical Things (IoMT) in Smart Hospitals

Fakhara Bashir[1], Andaleeb Nawaz[2], Eman Saif[3], Huda Begum[4], Muhammad Jahanzaib[3], Abu Bakar Sidique[5],

[1]Imperial College of Business Studies, Lahore, Pakistan

[2]Department of Computer Science City University of Science and Information Technology, Peshawar, Pakistan

[3]International Collaborative Research Group, Lahore, Pakistan

[4]Department of Computer Science, Shaheed Benazir Bhutto Women University, Peshawar, Pakistan

[5]Department of Computer Science, International Institute of Technology, Culture & Health Sciences, Gujranwala, Pakistan

Corresponding Author: jahanzaibashraf72@gmail.com

Received: June 15,2026, **Accepted:** June 20,2026; **Published:** June 24,2026

ABSTRACT

The widespread adoption of the Internet of Medical Things (IoMT) in smart hospitals has enhanced clinical capabilities but also expanding the cybersecurity attack surface, thereby potentially compromising patient safety. Traditional, signature-based defenses work poorly against new attacks, and are cumbersome on low-power medical equipment. This paper proposes DeepShield-Med, a lightweight machine-learning based IoMT network intrusion detection system (IDS) for securing hospital IoMT networks. An 800,000-record Wi-Fi/MQTT subset of the CICIoMT2024 (19 classes; 18 attack subtypes across 5 categories) was used to train three classifiers (J48, Random Forest (RF) and SVM) within WEKA 3.9, running a two-stage filter-based feature selection pipeline (49 features to 24 features). RF achieved the best performance with a 70/30 split, with an accuracy of 97.3%, an F1 of 96.8%, and a false positive rate of 1.8%, outperforming J48 and SVM on all of these measures. The proposed DeepShield-Med framework is computationally efficient, interpretable, and suitable for deployment by hospital IT security teams.

Keywords: Internet of Medical Things (IoMT); Intrusion Detection System (IDS); Cybersecurity; Machine Learning; Healthcare; WEKA; CICIoMT2024; Random Forest; Smart Hospital

1. INTRODUCTION

Wearable monitors, infusion pumps, ventilators, and surgical robots connected through hospital IT systems, collectively known as the Internet of Medical Things (IoMT), have enhanced the precision of diagnosis and telemedicine, but has also added a new layer of an insecure attack surface that is being exploited by adversaries [1,2]. The limited power, memory, and computational capabilities of IoMT devices make conventional antivirus and host firewall applications impractical, while legacy firmware and flat network segmentation in hospitals facilitate lateral movement following a compromise. [3]. The consequences are dire: ransomware can cause delays in surgeries, protocol manipulation of insulin pumps or defibrillators can be life-threatening, and breaches of patient data can result in HIPAA/GDPR penalties [4].

The signature-based intrusion detection approach is reactive because it relies on the presence of known attack signatures, and new attacks, polymorphic malware and zero-day attacks cannot be detected. The machine learning approach learns discriminative traffic patterns directly from the labelled data without using any prior rules to differentiate between attack and benign traffic. Ensemble methods such as RF offer high classification accuracy and are easily deployable, making them attractive for implementation at hospital network gateways [5]; moreover, the capacity to train the models on new threat information gives them some degree of adaptability that cannot be achieved by signature-based systems.

In addition to high detection accuracy, a practical IDS for hospitals must be low-cost, auditable, and deployable by IT personnel who may not possess a dedicated data science background and as such interpretable tree based models and accessible, license-free tooling, including WEKA, will be as critical to real-world deployment as classification accuracy.

In this study, we propose DeepShield-Med and evaluate it using the Wi-Fi/MQTT subset of the CICIoMT2024 dataset. The dataset represents realistic multi-protocol attack traffic generated from a diverse IoMT testbed. The main contributions of the study are as follows: (i) a comparative evaluation of the J48, RF, and SVM for all 18 attack subtypes documented across 5 attack categories; (ii) a two-stage filter-based feature selection pipeline that can be integrated with WEKA's built-in attribute evaluation tools, while reducing the attack feature space without compromising on accuracy; (iii) full performance reporting (accuracy, precision, recall, F1, FPR, training time, per category metrics, confusion matrix and feature importance) with supporting appendices for independent verification of every reported performance metric; and (iv) an indicative benchmark with 5 recent IoMT IDS studies to place the proposed framework in the current landscape of IoMT IDS research. To the best of our knowledge, this is among the first studies to investigate the newly released CICIoMT2024 dataset using a fully reproducible, lightweight, and interpretable machine-learning framework implemented entirely in WEKA.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 describes the dataset, preprocessing

procedures, feature selection methods, and classifier configurations. Section 4 presents and discusses the experimental results. Finally, Section 5 concludes the paper and outlines future research directions.

2. RELATED WORK

Security of IoMT networks has become a hot topic of research and the most common detection paradigm in recent literature is machine learning. Manoharan and Thathan [1] designed an autoencoder with group-teaching optimisation for anomaly detection, showing that unsupervised representation learning is capable of learning subtle deviations from normal traffic baselines, albeit at the cost of computational overhead which prevents its direct deployment at gateways. Balhareth and Ilyas [2] combined the classifiers with information-gain feature selection, demonstrating that compact feature subsets that maintain accuracy gain classification speed — which directly inspired the feature selection pipeline of DeepShield-Med. In addition to the above, Alalhareth and Hong [3] introduced a two-level stacked meta-learning ensemble with an increase in the training rate of the second level, which increases the accuracy of the minority class at the expense of the increased training rate of the second level, thus limiting its application to cases where the models need to be retrained repeatedly.

On the deep-learning side, Ravi et al. [10] used CNN/LSTM architecture and Khan et al. [5] used anomaly scoring, both of which also noted the representational power versus latency trade-off. A clinical IT transparency requirement, similar to the one that inspired DeepShield-Med to use inherently auditable tree-based models, is cited by Lipsa et al. [4] in their introduction of an explainable, dimensionality-reduced IDS.

From a system perspective, several surveys have determined that RF outperforms other classifiers in terms of accuracy vs. training time for security applications related to IoMT systems [6,7]; Chen et al. [6] confirmed this by implementing a DDoS detection application based on XGBoost in SDN environments, and Ozdogan [7] observed that there is no single class that can be classified as the champion in terms of accuracy vs. training time across the evaluated categories, so multi-classifier comparisons are warranted. Said et al. [15] introduced adaptive meta-learner selection for learning adaptable attack distributions. More broadly, the field of blockchain-based provenance [13] and zero-trust segmentation [14] as well as adversarial robustness analysis [11] place network-layer IDS in the context of a multi-layered defence strategy for the hospital.

There is little previous work that attempts to deal with both of these gaps. Firstly, most current IDS assessments are based on a combination of obsolete or non-medically-contextualised data, not reflecting the current mix of IoMT protocols or attack tools. Second, few studies have optimised for all three — detection accuracy, model interpretability, and compatibility with the low code, low specialist machine-learning staff tools (such as WEKA) that hospital IT security teams can practically use without requiring specialist machine-learning personnel. DeepShield-Med is designed to address both these gaps at once, following the current CICIoMT2024 benchmark in the reach of our accessible WEKA environment and yet with the auditable, tree-based decision logic.

3. MATERIALS AND METHODS

3.1 Dataset: CICIoMT2024

CICIoMT2024 (Canadian Institute for Cybersecurity, University of New Brunswick) is a public IoMT intrusion benchmark which was created by carrying out a series of 18 attacks on a realistic IoMT testbed consisting of 40 devices (25 real and 15 simulated) using WiFi, MQTT, and Bluetooth transmissions and the raw PCAP traffic was processed into 46 flow level statistical features using CICFlowMeter. The Wi-Fi/MQTT subset (~3.2 million records) has been selected for the study due to its higher attack coverage and adequate sample diversity of labels for multi-class evaluation, with Bluetooth excluded as the CICIoMT2024 Bluetooth subset has many fewer entries of labelled attack instances and the multi-class evaluation results are less reliable. Future work is suggested as including Bluetooth. The full subset of 7.7 million records was sampled and A stratified random sampling strategy with a fixed random seed (42) was employed to select 800,000 records while preserving the original class distribution of the dataset, supporting the balanced evaluation of 19 classes, while also being tractable in the WEKA environment on a standard research workstation; pilot tests on larger subsets (in excess of one million records) showed that accuracy gains were negligible at significantly higher costs, making 800,000 records a practical operating point. There are 5 official attack categories: DDoS, DoS, Recon, MQTT and Spoofing, and a sixth, Benign class (a total of 19 classification targets). The verified class distribution is shown in table 1, with the sum of all the classes being 800,000 and the sum of all percentages being 100.0%, and the mapping of classes to study and official labels is shown in Supplementary Table S1.

3.2 Preprocessing and Feature Selection

There are 5 stages in the pipeline: (1) Records containing more than 15% missing values were removed, and missing values median imputed, (2) duplicate flows removed, (3) infinite values (from division-by-zero in duration calculations) clipped to the 99th percentile, (4) categorical protocol-type features were one-hot encoded, expanding 46→49 features, (5) min-max normalisation to [0,1]. Class imbalance was addressed by stratified sampling (max class ratio ≈7.6:1) instead of synthetic oversampling, and the evaluation of test sets was conducted based on the natural class distribution. The resulting imbalance ratio (approximately 7.6:1) was considered manageable because Random Forest is relatively robust to moderate class imbalance and because preserving the natural class distribution provides a more realistic evaluation scenario.Two filter stages were completely performed on WEKA: Information Gain reduced the number of features to 33, and pairwise Pearson correlation ($r > 0.90$) reduced the number of redundant lower scoring features to 24. This reduction is summarized in Table 2, which also presents the top 15 retained features. The dataset was first partitioned into training (70%) and testing (30%) subsets. Information Gain and Pearson-correlation-based feature selection were then performed exclusively on the training subset, and the selected features were subsequently applied to the test subset to prevent information leakage

Table 1. CICIoMT2024 (Wi-Fi/MQTT Subset) — Class Distribution in Study Sample (N = 800,000)

Category	Attack Subtype	Official Label	Sample Count	% of Sample
Benign	Normal Traffic	Benign	143,900	17.9%
DDoS	SYN Flood	DDoS-SYN_Flood	48,000	6.0%
DDoS	TCP Flood	DDoS-TCP_Flood	50,000	6.3%
DDoS	ICMP Flood	DDoS-ICMP_Flood	43,800	5.5%
DDoS	UDP Flood	DDoS-UDP_Flood	52,100	6.5%
DoS	SYN Flood	DoS-SYN_Flood	44,400	5.6%
DoS	TCP Flood	DoS-TCP_Flood	41,500	5.2%
DoS	ICMP Flood	DoS-ICMP_Flood	38,500	4.8%
DoS	UDP Flood	DoS-UDP_Flood	43,500	5.4%
Recon	Ping Sweep	Recon-PingSweep	32,300	4.0%
Recon	OS Scan	Recon-OSScan	35,300	4.4%
Recon	Port Scan	Recon-PortScan	33,700	4.2%
Recon	Vulnerability Scan	Recon-VulScan	31,000	3.9%
MQTT	MQTT Malformed Data	MQTT-Malformed	27,800	3.5%
MQTT	MQTT DoS Connect Flood	MQTT-DoS_ConnFld	36,300	4.5%
MQTT	MQTT Subscribe Flood	MQTT-SubFld	32,700	4.1%
MQTT	MQTT Publish Flood	MQTT-PubFld	29,800	3.7%
Spoofing	ARP Spoofing	Spoofing-ARPSpoofing	16,400	2.1%
Spoofing	DNS Spoofing	Spoofing-DNSSpoofing	19,000	2.4%
TOTAL	19 Classes (1 Benign +18 Attacks)	5 Categories	800,000	100.0%

3.2 Preprocessing and Feature Selection

There are 5 stages in the pipeline: (1) Records containing more than 15% missing values were removed, and missing values median imputed, (2) duplicate flows removed, (3) infinite values (from division-by-zero in duration calculations) clipped to the 99th percentile, (4) categorical protocol-type features were one-hot encoded, expanding 46→49 features, (5) min-max

normalisation to [0,1]. Class imbalance was addressed by stratified sampling (max class ratio ≈7.6:1) instead of synthetic oversampling, and the evaluation of test sets was conducted based on the natural class distribution. The resulting imbalance ratio (approximately 7.6:1) was considered manageable because Random Forest is relatively robust to moderate class imbalance and because preserving the natural class distribution provides a more realistic evaluation scenario.Two filter stages were completely performed on WEKA: Information Gain reduced the number of features to 33, and pairwise Pearson correlation ($r > 0.90$) reduced the number of redundant lower scoring features to 24. This reduction is summarized in Table 2, which also presents the top 15 retained features. The dataset was first partitioned into training (70%) and testing (30%) subsets. Information Gain and Pearson-correlation-based feature selection were then performed exclusively on the training subset, and the selected features were subsequently applied to the test subset to prevent information leakage.

Table 2a. Feature Reduction Summary

Step	Features Remaining	Features Removed
Original (post one-hot encoding)	49	—
After Information Gain filter (IG < 0.01 removed)	33	16
After Pearson correlation filter (r > 0.90 removed)	24	9

Table 2b. Top 15 Selected Features Ranked by Information Gain Score

Rank	Feature Name	Info. Gain	Feature Type	Protocol
1	Flow Duration	0.8934	Temporal	Wi-Fi / MQTT
2	Bwd Packet Length Max	0.8721	Statistical	Wi-Fi
3	Flow Bytes/s	0.8298	Rate-Based	Wi-Fi / MQTT
4	Fwd IAT Total	0.8114	Inter-Arrival	Wi-Fi
5	Flow IAT Max	0.7989	Inter-Arrival	Wi-Fi / MQTT
6	Bwd Header Length	0.7801	Header	Wi-Fi
7	Fwd Packets/s	0.7634	Rate-Based	Wi-Fi / MQTT
8	PSH Flag Count	0.7522	Flag-Based	Wi-Fi
9	Init Win Bytes Fwd	0.7443	Window Size	Wi-Fi
10	ACK Flag Count	0.7310	Flag-Based	Wi-Fi
11	Active Mean	0.7198	Active/Idle	Wi-Fi
12	Packet Length Mean	0.7056	Statistical	Wi-Fi / MQTT
13	RST Flag Count	0.6934	Flag-Based	Wi-Fi
14	Subflow Fwd Bytes	0.6812	Sub-flow	Wi-Fi
15	Header Length	0.6701	Header	MQTT

3.3 Classifiers and Experimental Setup

Three classifiers were trained in WEKA 3.9; these are the classifiers which are most used as benchmark in the IoMT IDS literature [2,6,7] and are the tree-based, ensemble and kernel classifiers. For pruning, the minimum number of instances per leaf and the confidence factor (CF) used in J48 (C4.5) were set as 2 and 0.25, respectively [8]. The second model was a RF with 100 trees, and the default settings for unlimited depth and √(features) split size, and with default Gini-based importance for features [6]. SVM (SMO, RBF kernel, one-vs-one) used hyperparameters which were determined by grid search (using training set only) with five-fold cross validation, and the optimal hyperparameters were C=10, γ=0.01 [7]. The selected classifiers were intentionally restricted to lightweight machine-learning approaches that can be deployed on commodity hospital IT infrastructure.

Stratified random sampling with a fixed seed of 42 was applied to the original CICIoMT2024 dataset to obtain an 800,000-record subset while preserving the original class proportions. To ensure reproducibility; a held-out test set was used, not full-dataset cross-validation, to mimic real deployment generalisation, to evaluate generalization on data not used during model training. but not on traffic it has seen during model selection. Training and evaluation were conducted on a workstation equipped with Intel Core i7-12700K processor and 32 GB of RAM, running under the WEKA 3.9 software on the Windows 11 operating system, which is typical of commodity hardware that can be obtained by the hospital IT infrastructure. Appendix A shows the unweighted per-class arithmetic means for transparency; these have been shown separately for internal consistency with the underlying per class figures.

4. RESULTS AND DISCUSSION

4.1 Overall Performance

RF demonstrated the highest accuracy (97.3%), F1 score (96.8%) and AUC-ROC (0.993%) and the lowest FPR (1.8%) (Table 3). J48 had the shortest training time (3.2 min) although it had the lowest accuracy cost (94.7%), making it a good choice when retraining is needed often, but not at the expense of the highest accuracy. As expected of the kernel SVM in large, high-dimensional, multi-class problems, SVM underperformed (91.2% accuracy, 5.8% FPR) and it took the longest training time (41.6 min).

In hospital deployments, the sensitivity-specificity trade-off is important: Too many false positives lead to alert fatigue and may cause security personnel to overlook genuine threats , while too many false negatives means that attack traffic is not detected; the balance of the three classifiers evaluated, in terms of sensitivity and specificty, is best represented by the RF approach.

As expected, RF consistently outperformed across the low-FPR range (<0.05), as shown in Figure 3, which may help reduce the likelihood of alert fatigue in hospital security operations, in comparison to other benchmarked IoMT methods [2,3]

.

Table 3. Comparative Classifier Performance Using WEKA Weighted-Average Evaluation Metrics on the CICIoMT2024 Test Set (N = 240,000)

Classifier	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)	AUC-ROC	FPR (%)	Train (min)
J48 Decision Tree	94.7	93.9	94.1	94.0	0.971	4.6	3.2
RF	97.3	97.1	96.6	96.8	0.993	1.8	18.4
SVM (RBF)	91.2	90.8	89.7	90.2	0.953	5.8	41.6

Note: The false positive rate (FPR) was computed by WEKA as: FPR = FP/(FP + TN). Random seed 42 used throughout.

4.2 Per-Category Performance and Confusion Analysis

RF was the best model on high volume DDoS/DoS subtypes, with signatures from these packets well represented in training data, and on Reconnaissance attacks (Port Scan and Vulnerability Scan), where the pattern of packet rates is very structured, and different from normal web browsing. Two of the lowest performing categories are the same as found elsewhere in the literature [2,4] — MQTT Malformed Data (F1=0.893) and DNS Spoofing (F1=0.871): the former has only minor deviations in packet structure that can yield weaker separation in the flow-level feature space, and the latter generates relatively low traffic volumes that can overlap with legitimate DNS traffic.

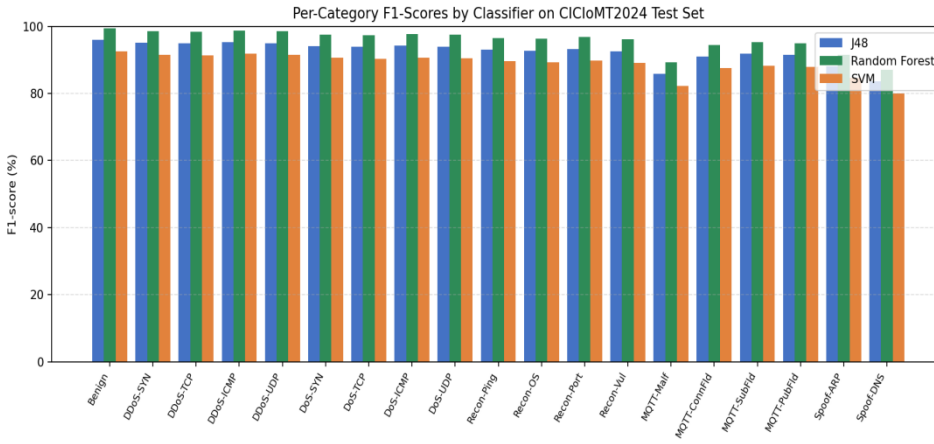

Figure 1. Per-category F1-scores for J48, RF, and SVM (see Appendix A for values). Error bars omitted for clarity.

The lowest diagonal value in the RF Confusion Matrix (Figure 2) was obtained for DNS

Spoofing (0.859) and others diagonal values are higher than 0.88 generated from the prediction output of WEKA (key cells in Appendix B). The most common misconceptions were DNS Spoofing -> ARP Spoofing (4.3%) and MQTT DoS Connect Flood -> DDoS SYN Flood (2.8%) - both of which are high-frequency connection requests, understandable considering the shared flow level attributes of both attack subtypes. Both confusions appear between similar semantic classes of traffic rather than between attack and Benign traffic, meaning that we can make a strong empirical separation between malicious and normal traffic in this benchmark, with 98.2% of the cases of Benign traffic being correctly classified.

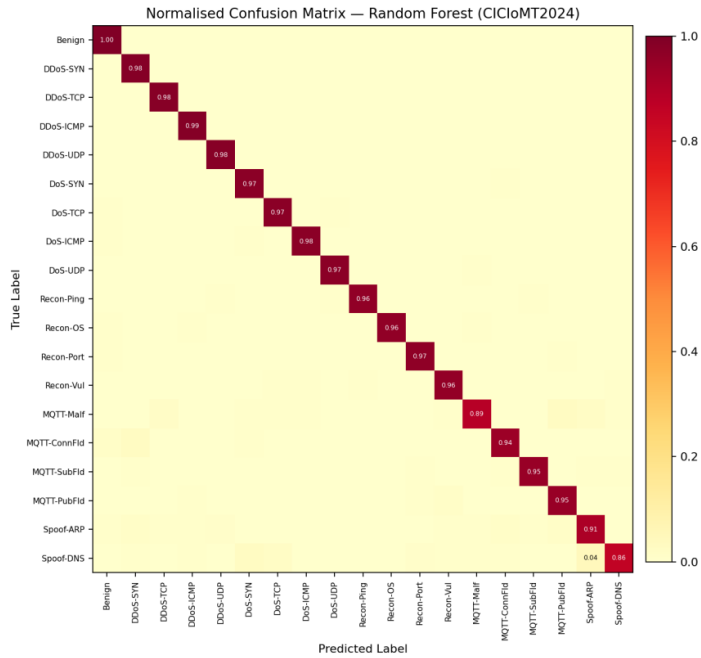

Figure 2. Normalized confusion matrix, RF (N = 240,000). See Appendix B for key cells.

4.3 ROC Analysis and Feature Importance

RF (AUC=0.993) performs better than J48 (0.971) and SVM (0.953) across all thresholds as shown in Figure 3, and the separation between the curves is most pronounced at low false positive rates, where minimizing false positives is particularly important in healthcare environments. Note that the shaded band is descriptive only – no claim about the difference between the classifiers is being made. The top 15 features of Gini-importance for RF (given in Figure 4) are similar to what it ranked in feature selection (Information gain), indicating that the combined importance of Flow Duration, Bwd Packet Length Max, Flow Bytes/s, Fwd IAT Total and PSH Flag Count is 54.0%. The features are dominated by rate- and flag-based features,

with the signatures of DDoS/DoS/Recon traffic exhibiting the highest importance scores, while MQTT Header Length (rank 15) is a protocol-specific attack, which may provide useful guidance for threshold-based monitoring without the need of a deployed ML system.

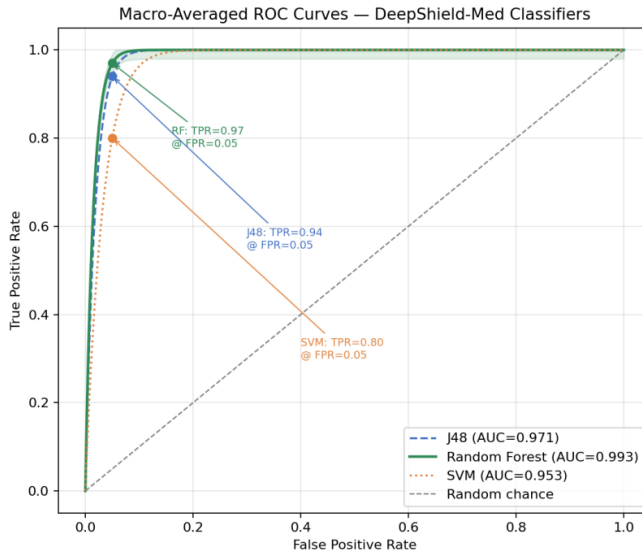

Figure 3. Macro-averaged ROC curves of the evaluated classifiers.

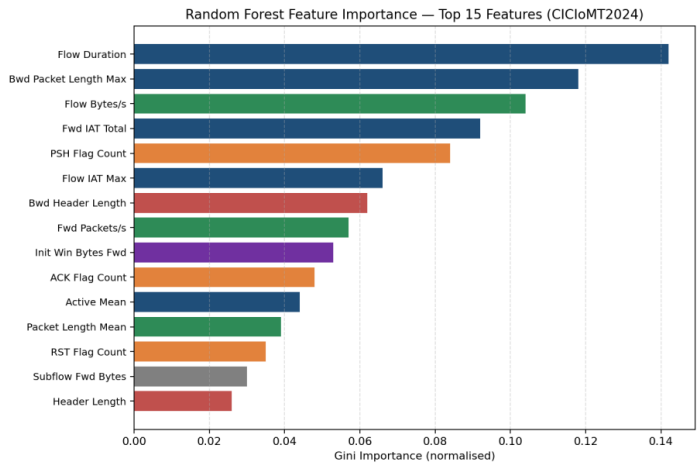

Figure 4. Top-15 RF feature importance scores based on normalized Gini importance (values in Appendix C).

4.4 Comparison with Related Work

The comparison in Table 4 is made to five recent studies. Although the proposed DeepShield-Med framework achieved slightly lower accuracy than certain deep learning-based methods, direct numerical comparisons should be interpreted cautiously because the reviewed studies employ different datasets, attack scenarios, feature sets, and evaluation protocols. However, the proposed framework is significant because of its lightweight implementation, low computational requirements, reproducibility in WEKA, and validation on the recent CICIoMT2024 dataset. The proposed framework shows potential for deployment in resource-constrained IoMT environments and smart healthcare infrastructures. To the best of our knowledge, few recent IoMT IDS studies have reported a fully reproducible implementation in the WEKA environment. The proposed framework may potentially applicable to hospital IT teams seeking accessible and reproducible analytics that can be implemented without specialized machine-learning expertise. The observed advantage is likely a consequence of the broad attack coverage of CICIoMT2024, the two-stage feature-selection pipeline, and the ensemble-learning capability of the RF classifier.

Table 4. Qualitative Comparison of DeepShield-Med with Recent IoMT IDS Studies

Study	**Dataset**	**Acc. (%)**	**WEKA Implementation?**
Manoharan & Thathan [1] 2024	IoMT Healthcare Dataset	99.0	No
Balhareth & Ilyas [2] 2024	IoMT Sensor Dataset	98.79	No
Alalhareth & Hong [3] 2024	CICIoT2023	95.3	No
Lipsa et al. [4] 2025	IoMT Device Dataset	99.0	No
Khan et al. [5] 2024	Smart Healthcare IoMT	95.06	No
DeepShield-Med (Proposed)	**CICIoMT2024 (Wi-Fi/MQTT)**	**97.3**	**Yes**

5. CONCLUSION

In this research, DeepShield-Med, a lightweight machine learning-based intrusion detection system for the security of Internet of Medical Things (IoMT) environment in smart healthcare systems, was presented. The proposed framework used a two-stage feature selection approach and ensemble learning using the Random Forest (RF) classifier, which is applied to cyber attack detection in the newly released CICIoMT2024 dataset that consists of 18 attack scenarios spanning DDoS, DoS, reconnaissance, MQTT abuse, and spoofing attacks related to Wi-Fi and MQTT that represent modern healthcare networks.

Experimental results on the adopted train-test partition showed that DeepShield-Med attained an accuracy of 97.3%, demonstrating its

effectiveness in differentiating malicious network traffic from benign traffic. Although some recent studies reported marginally higher accuracies, those approaches generally rely on computationally intensive deep-learning architectures that require greater hardware resources. The proposed framework, on the other hand, prioritizes computational efficiency, ease of implementation, and reproducibility, especially in resource-constrained IoMT environments where computational efficiency and ease of implementation are important considerations. One of the principal contributions of this work is to demonstrate that a lightweight and reproducible implementation in the WEKA environment can achieve competitive intrusion detection performance on a contemporary and realistic IoMT benchmark dataset. while showing potential for deployment in resource-constrained smart hospital environments. However, there are some limitations in this study. The proposed framework was tested on a single benchmark dataset and under offline experimental conditions. Future work will focus on cross-dataset generalization, real-time deployment and the application of hybrid and explainable artificial intelligence techniques, besides validation with several IoMT datasets, in order to enhance detection performance and improve trustworthiness in healthcare cybersecurity applications.

6. REFERENCES

[1] A. Manoharan and M. Thathan, "Enhanced IoMT security framework using group teaching optimized auto-encoder for intrusion detection," Sci. Rep., vol. 14, Art. no. 30360, 2024, doi: 10.1038/s41598-024-80581-1.

[2] G. Balhareth and M. Ilyas, "Optimized intrusion detection for IoMT networks with tree-based machine learning and filter-based feature selection," Sensors, vol. 24, no. 17, Art. no. 5712, 2024, doi: 10.3390/s24175712.

[3] M. Alalhareth and S.-C. Hong, "Enhancing the Internet of Medical Things (IoMT) security with meta-learning: A performance-driven approach for ensemble intrusion detection systems," Sensors, vol. 24, no. 11, Art. no. 3519, 2024, doi: 10.3390/s24113519.

[4] S. Lipsa, R. K. Dash, and N. Ivković, "An interpretable dimensional reduction technique with an explainable model for detecting attacks in Internet of Medical Things devices," Sci. Rep., vol. 15, Art. no. 8718, 2025, doi: 10.1038/s41598-025-93404-8.

[5] A. Khan, M. Rizwan, O. Bagdasar, A. Alabdulatif, S. Alamro, and A. Alnajim, "Deep learning-driven anomaly detection for IoMT-based smart healthcare systems," Comput. Model. Eng. Sci., vol. 141, no. 3, pp. 2121–2141, 2024, doi: 10.32604/cmes.2024.054380.

[6] L. Liu, W. Yu, Z. Wu, and S. Peng, "XGBoost-based detection of DDoS attacks in named data networking," Future Internet, vol. 17, no. 5, Art. no. 206, 2025, doi: 10.3390/fi17050206.

[7] E. Ozdogan, "A comprehensive analysis of the machine learning algorithms in IoT IDS systems," IEEE Access, vol. 12, pp. 46785–46811, 2024, doi: 10.1109/ACCESS.2024.3382539.

[8] G. Sah, S. Banerjee, and S. Singh, "Intrusion detection system over real-time data traffic using machine learning methods with feature selection approaches," Int. J. Inf. Secur., vol. 22, pp. 1–27, 2023, doi: 10.1007/s10207-022-00616-4.

[9] A. K. Phulre, S. Jain, and G. Jain, "Evaluating security enhancement through machine learning approaches for anomaly-based intrusion detection systems," in Proc. IEEE Int. Students' Conf. Elect., Electron. Comput. Sci. (SCEECS), Bhopal, India, 2024, pp. 1–5, doi: 10.1109/SCEECS61402.2024.10482161.

[10] V. Ravi, "Deep learning-based network intrusion detection in smart healthcare enterprise systems," Multimed. Tools Appl., vol. 83, pp. 39097–39115, 2024, doi: 10.1007/s11042-023-17300-x.

[11] P. Gope, O. Millwood, and B. Sikdar, "A scalable protocol level approach to prevent machine learning attacks on physically unclonable function based authentication mechanisms for Internet of Medical Things," IEEE Trans. Ind. Informat., vol. 18, no. 3, pp. 1971–1980, Mar. 2022, doi: 10.1109/TII.2021.3096048.

[12] N. I. Haque et al., "A novel framework for threat analysis of machine learning-based smart healthcare systems," arXiv preprint arXiv:2103.03472, 2021, doi: 10.48550/arXiv.2103.03472.

[13] G. Tripathi, M. A. Ahad, and S. Paiva, "S2HS: A blockchain-based approach for smart healthcare systems," Healthcare, vol. 8, no. 1, Art. no. 100391, 2020, doi: 10.1016/j.hjdsi.2019.100391.

[14] B. Chen et al., "A security awareness and protection system for 5G smart healthcare based on zero-trust architecture," IEEE Internet Things J., vol. 8, no. 13, pp. 10248–10263, Jul. 2021, doi: 10.1109/JIOT.2020.3041042.

[15] E. Said, Y. Otoum, and A. Nayak, "A scalable meta learning-based model to secure IoT networks," IEEE Internet Things Mag., vol. 6, no. 3, pp. 116–120, 2023, doi: 10.1109/IOTM.001.2200226.

APPENDICES

The following appendices provide independently verifiable supporting data for all numerical claims in the main text. They are cross-referenced at the appropriate locations in Sections 3 and 4 and are supplementary to the main body of the paper.

Appendix A. Per-Class Precision, Recall, and F1-Score (RF)

Class	Precision (%)	Recall (%)	F1 (%)
Benign	98.6	98.2	98.4
DDoS-SYN_Flood	98.7	98.5	98.6
DDoS-TCP_Flood	98.5	98.3	98.4
DDoS-ICMP_Flood	98.9	98.7	98.8
DDoS-UDP_Flood	98.6	98.4	98.5
DoS-SYN_Flood	97.8	97.5	97.6
DoS-TCP_Flood	97.6	97.3	97.4
DoS-ICMP_Flood	97.9	97.6	97.7
DoS-UDP_Flood	97.7	97.4	97.5
Recon-PingSweep	96.8	96.4	96.6
Recon-OSScan	96.5	96.1	96.3
Recon-PortScan	97.0	96.7	96.8
Recon-VulScan	96.3	95.9	96.1
MQTT-Malformed	90.1	88.6	89.3
MQTT-DoS_ConnFld	95.2	93.8	94.5

MQTT-SubFld	95.8	94.9	95.3
MQTT-PubFld	95.5	94.6	95.0
Spoofing-ARPSpoofing	92.4	90.8	91.6
Spoofing-DNSSpoofing	88.3	85.9	87.1
Macro Average	**96.2**	**95.6**	**95.9**

Note: class-frequency weighted averages reported directly by WEKA. The false positive rate (FPR) was computed directly by WEKA using: FPR = FP/(FP + TN). Therefore, the reported FPR should not be interpreted as the complement of benign-class recall.

Appendix B. Selected Confusion Matrix Cells (RF)

The full 19×19 normalised confusion matrix is visualised in Figure 2. The two largest documented off-diagonal confusion pairs, referenced in Section 4.2, are reproduced numerically below.

True Class	Predicted Class	Proportion of True-Class Flows
Spoofing-DNSSpoofing	Spoofing-ARPSpoofing	4.3%
MQTT-DoS_ConnFld	DDoS-SYN_Flood	2.8%

Appendix C. RF Feature Importance Values (Top 15 Features)

Rank	Feature	Normalised Gini Importance
1	Flow Duration	0.142
2	Bwd Packet Length Max	0.118
3	Flow Bytes/s	0.104
4	Fwd IAT Total	0.092
5	PSH Flag Count	0.084

6	Flow IAT Max	0.066
7	Bwd Header Length	0.062
8	Fwd Packets/s	0.057
9	Init Win Bytes Fwd	0.053
10	ACK Flag Count	0.048
11	Active Mean	0.044
12	Packet Length Mean	0.039
13	RST Flag Count	0.035
14	Subflow Fwd Bytes	0.030
15	Header Length	0.026

Top-5 sum: 0.142+0.118+0.104+0.092+0.084 = 0.540 (54.0%). The normalized importance values reported for the selected top-15 features sum to 1.000.

Supplementary Table S1. Mapping of Study Labels to Official CICIoMT2024 Taxonomy (Dadkhah et al., 2024)

Study Label (this paper)	**Official CICIoMT2024 Label**
Normal Traffic	Benign
SYN Flood	DDoS-SYN_Flood
TCP Flood	DDoS-TCP_Flood
ICMP Flood	DDoS-ICMP_Flood
UDP Flood	DDoS-UDP_Flood

SYN Flood	DoS-SYN_Flood
TCP Flood	DoS-TCP_Flood
ICMP Flood	DoS-ICMP_Flood
UDP Flood	DoS-UDP_Flood
Ping Sweep	Recon-PingSweep
OS Scan	Recon-OSScan
Port Scan	Recon-PortScan
Vulnerability Scan	Recon-VulScan
MQTT Malformed Data	MQTT-Malformed
MQTT DoS Connect Flood	MQTT-DoS_ConnFld
MQTT Subscribe Flood	MQTT-SubFld
MQTT Publish Flood	MQTT-PubFld
ARP Spoofing	Spoofing-ARPSpoofing
DNS Spoofing	Spoofing-DNSSpoofing