A Hybrid Autoencoder-Random Forest Framework for Intrusion Detection in Internet of Medical Things Network

Sophia ALAMI-KAMOURI[1], Ridouan LACHGAR[2], Mohamed AFIF[3]
M2CS, National Graduate School of Arts and Crafts (ENSAM), Mohammed V University in Rabat, Morocco[1]
National Graduate School of Arts and Crafts (ENSAM), Mohammed V University in Rabat, Morocco[2,3]

***Abstract*—With the widespread adoption of the Internet of Medical Things (IoMT), hospitals have become prime targets for cyberattacks. To overcome the limitations of traditional defenses and computationally heavy deep learning models, a hybrid Artificial Intelligence architecture is presented. By coupling a deep Autoencoder for feature extraction with a Random Forest classifier, the input space is reduced from 44 down to 16 latent dimensions. While a standalone Random Forest achieves 98.95% accuracy, it remains too resource-intensive for constrained edge devices. The proposed hybrid approach strategically accepts a marginal sacrifice in accuracy—achieving 94.60% overall—in exchange for a drastic reduction in computational complexity. Rigorously validated on the CICIoMT2024 dataset and balanced via SMOTE-ENN, this model achieves an ultra-low inference latency of 1.05 ms on edge-grade hardware, making it highly viable for Edge Computing deployment. Furthermore, integrating the SHAP algorithm ensures decision-making transparency, addressing the "black box" challenge in medical AI. These results demonstrate the efficacy of strategic dimensionality reduction in securing critical healthcare infrastructures.**

***Keywords*—Cybersecurity; artificial intelligence; IoMT; hybrid model; Edge Computing; SHAP**

I. Introduction

The integration of the Internet of Medical Things (IoMT) has significantly changed how healthcare is delivered, particularly by improving patient monitoring and care. However, connecting medical devices to networks has also introduced new vulnerabilities, making these systems targets for cyberattacks. This study begins with a review of current literature on AI-based architectures for IoMT security, examining why sequential deep learning models often fall short in this context. To address these limitations, a hybrid Artificial Intelligence approach that couples an Autoencoder with a Random Forest classifier is proposed. The goal of this architecture is to achieve accurate anomaly detection while keeping latency low enough for Edge Computing applications, all while ensuring the model's decisions can be interpreted using the SHAP algorithm.

II. Cybersecurity Challenges in Hospital Environments

Modern cyberattacks have moved beyond simple data theft and now pose direct risks to patient safety. As hospitals increasingly rely on connected devices like pacemakers, insulin pumps, and MRI machines, they become more attractive targets for attackers. Traditional security measures, such as firewalls and signature-based antivirus software, often struggle to keep up with these evolving threats. They are typically unable to detect zero-day attacks and can be overwhelmed by the large volume of real-time data generated by medical sensors [1], [3], [5], [6], [14].

To help close this security gap, this study discusses the development of an Intrusion Detection System (IDS) based on behavioral analysis. By using a hybrid Artificial Intelligence (AI) model, the system moves away from relying on known virus signatures. Instead, it learns the baseline behavior of the hospital network to identify anomalies quickly and with minimal delay.

III. Digital Health and IoMT Security Vulnerabilities

Digital health involves more than just converting paper records to digital formats; it requires connecting various stakeholders to improve how information is shared. Hospital Information Systems (HIS) rely on components like Electronic Health Records (EHR) and PACS, utilizing interoperability standards such as HL7 and FHIR. While adopting Cloud Computing lowers infrastructure costs, it introduces challenges related to data privacy under frameworks like the GDPR in Europe and HIPAA in the United States.

This digital transformation is closely tied to the Internet of Medical Things (IoMT), which is generally modeled in a four-layer architecture:

- Perception Layer: Biomedical sensors and actuators operating under severe battery and computing constraints.
- Network Layer: Lightweight communication protocols (e.g., Bluetooth, ZigBee) and application protocols (MQTT).
- Middleware / Edge Layer: Gateways processing information locally to mitigate latency.
- Application Layer: Cloud or datacenter centralization.

One of the main challenges is that edge devices operate under severe resource constraints, creating a significant vulnerability. Attackers can exploit these vulnerabilities through Man-in-the-Middle (MiTM) attacks, DDoS via MQTT, IoMT Botnets, and Battery Depletion Attacks. Standard heavy encryption on a pacemaker, for example, would drain its battery

too quickly, necessitating lightweight, behavioral anomaly detection systems.

IV. Artificial Intelligence for Intrusion Detection in IoMT

Traditional intrusion detection methods depend on a database of known malware signatures. While effective against recognized threats, these systems can be easily bypassed if an attacker modifies the malicious packet slightly. Artificial Intelligence offers an alternative approach by focusing on behavioral detection, or Anomaly-based IDS.

A. Global Anomaly Detection and Cognitive DLP

In modern hospital networks, unsupervised algorithms like Isolation Forests or Autoencoders are often used to model standard user behavior. For instance, if a user account suddenly attempts to access hundreds of medical records during off-hours, the algorithm flags this as an anomaly and generates an alert [2], [4], [7]. Additionally, Data Loss Prevention (DLP) tools can analyze connection metadata—such as sending rates or payload entropy—to identify unauthorized data transfers without needing to decrypt the actual content, thereby maintaining patient privacy.

B. Static and Dynamic Malware Sandboxing

To handle more evasive threats like fileless ransomware, AI can be used statically, by examining file features like API calls, and dynamically through sandboxing, where models like LSTMs monitor the sequence of system calls to halt suspicious processes.

V. The Mathematical Foundations of AI for Intrusion Detection

Data collected from IoMT gateways usually take the form of tabular datasets, containing variables like packet counts, session durations, and TCP flags.

A. Tabular Machine Learning: From Decision Trees to Random Forest

For tabular data that does not have strong spatial relationships, traditional Machine Learning algorithms are often a practical choice. A Decision Tree recursively splits the feature space to maximize Information Gain (IG), calculated from entropy:

$$IG(D, A) = H(D) - \sum_{v} \frac{|D_v|}{|D|} H(D_v) \tag{1}$$

Since single decision trees are prone to overfitting, ensemble methods such as Random Forest are frequently used [18]. Random Forest trains multiple trees using bootstrap sampling and aggregates their outputs through a majority vote. This bagging process (illustrated in Fig. 1) helps reduce the overall variance of the model:

$$\hat{f}_{RF}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}_b(x) \tag{2}$$

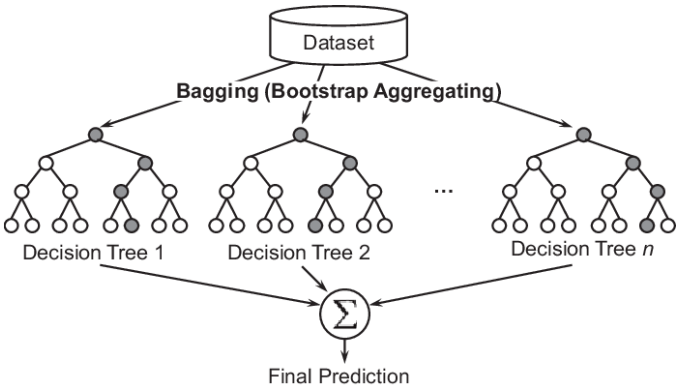

Fig. 1. Ensemble architecture of the Random Forest via bagging.

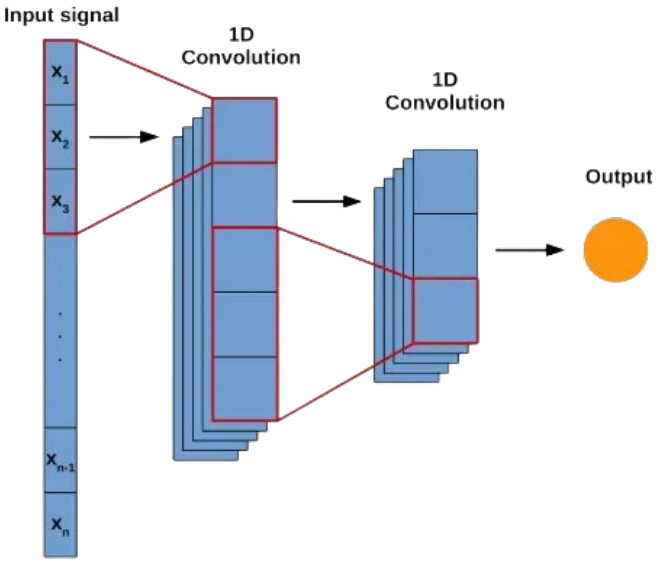

Fig. 2. One-dimensional convolution operation on a series of variables.

B. Deep Learning and Sequential Architectures

However, traditional Machine Learning can struggle when dealing with a high number of features due to the curse of dimensionality. In these cases, Deep Learning models—which use techniques like backpropagation and ReLU activation functions ($f(x) = \max(0, x)$)—can be better suited to find complex patterns in the data [21].

Many academic works propose using deep sequential models like 1D Convolutional Neural Networks (CNN 1D) and Long Short-Term Memory (LSTM) networks for network intrusion detection.

The CNN 1D (Fig. 2) applies a sliding filter w (kernel) over input data x:

$$y[i] = \sigma \left(\sum_{j=0}^{k-1} x[i+j] \cdot w[j] + b \right) \tag{3}$$

LSTMs use logic gates (Forget, Input, Output) to regulate

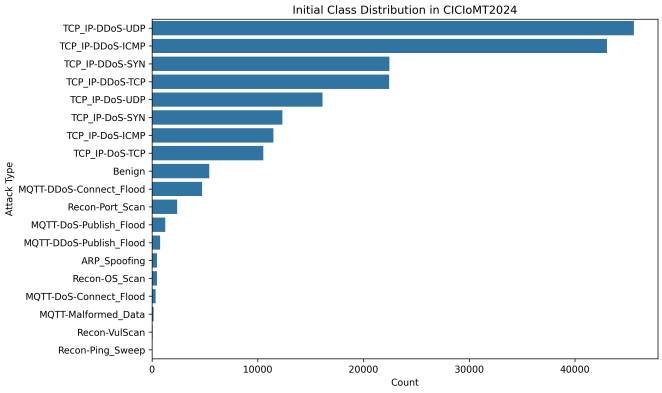

Fig. 3. Raw network data distribution, highlighting the predominance of benign traffic.

their internal memory cell C_t over time, aiming to solve the vanishing gradient problem in sequences.

C. Literature Review

Recent studies, summarized in Table I, have evaluated various architectures using benchmark datasets such as CICIoMT2024.

Recent approaches published in between 2025 and 2026 have explored Federated Learning and Transformer models; however, while these achieve high accuracy, they often impose communication and computational overheads that are incompatible with constrained IoMT Edge devices. A key issue observed in these comparisons is the inductive bias of certain deep learning models. While sequential models are well-suited for continuous signals, applying them to purely tabular data can force the model to look for spatial or temporal patterns that aren't actually there, which often leads to overfitting. As a result, hybrid approaches that combine neural networks for feature extraction with traditional Machine Learning classifiers have become an active area of research.

VI. Exploratory Data Analysis of the CICIoMT2024 Dataset

For the empirical evaluation, the **CICIoMT2024** dataset [8] was utilized. This dataset contains over 300,000 samples, split into 80% for training and 20% for testing, simulating a hospital network experiencing 19 different types of IoMT attacks.

A. Class Imbalance

Our initial analysis showed a significant imbalance between the classes.

As shown in Fig. 3, more than 80% of the network traffic is benign. Training a model on this raw data would likely result in a bias toward the majority class, making it difficult for the model to reliably detect less frequent cyberattacks.

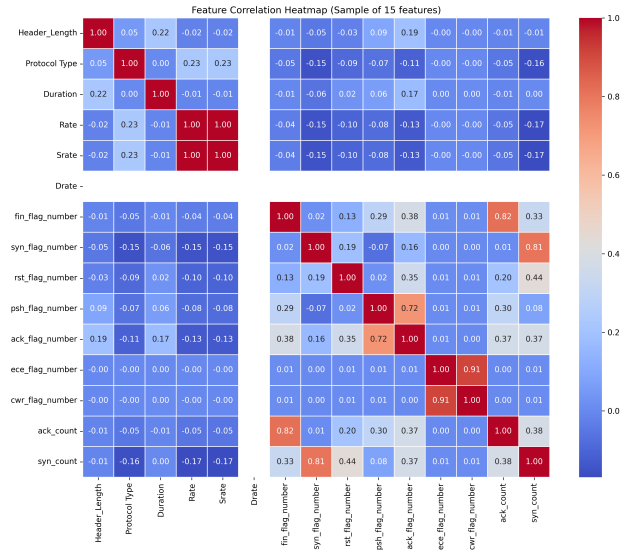

Fig. 4. Pearson correlation matrix. Dark red areas indicate highly correlated variables.

B. Redundancy and Spatial Complexity

Each packet in the dataset is described by 44 features. To check for redundancy among these features, a correlation matrix was calculated.

Fig. 4 highlights several areas of high correlation, indicating that multiple variables capture very similar information. This redundancy increases computational overhead, which is particularly undesirable for Edge Computing environments where resources are limited.

VII. Data Processing and Class Balancing Using SMOTE-ENN

To address the class imbalance, simple random oversampling was avoided, as it can lead to overfitting. Instead, the **SMOTE-ENN** algorithm was utilized [19], [20].

This technique generates synthetic examples for the minority classes by interpolating between existing data points (SMOTE), and then cleans the decision boundaries by removing ambiguous samples (ENN) (Fig. 5). To strictly prevent data leakage and ensure a realistic evaluation, SMOTE-ENN was applied exclusively to the training set after performing the train-test split, leaving the testing set entirely unmanipulated.

As seen in Fig. 6, applying SMOTE-ENN resulted in a more balanced distribution across the 19 attack classes.

VIII. Performance Analysis of Sequential Deep Learning Models

To inform the design of the hybrid model, standalone CNN 1D and LSTM networks were tested on the balanced dataset to serve as baselines.

A. CNN 1D Evaluation

The results showed significant instability. The learning curves for the CNN 1D (Fig. 7) fluctuated notably, and its

TABLE I. LITERATURE REVIEW OF AI-BASED IDS FOR IOMT

Author / Article	Architecture	Technical Approach	Strengths and Limitations
Future Internet (2025) [1]	Federated Deep Learning	Distributed training on local hospital data.	+ Preserves data privacy. - High communication overhead for edge devices.
Scientific Reports (2026) [15]	Transformer-based IDS	Self-attention mechanisms for traffic analysis.	+ Exceptional accuracy. - Too computationally expensive for IoMT gateways.
ArXiv (2024) [17]	CNN 1D	Forced spatial analysis on tabular data.	+ Fast inference. - Not optimized for spatially uncorrelated data.
JoITIT (2025) [10]	Hybrid CNN-LSTM	Sequential and spatial analysis.	+ State memorization. - Underperforms on purely tabular statistical data streams.
Zanco Journal (2025) [9]	**Extractor-Classifier Hybrids**	Deep extractor (Autoencoder) coupled with ML (Random Forest).	+ Good Accuracy/Interpretability balance. - Requires fine-tuning the latent space.

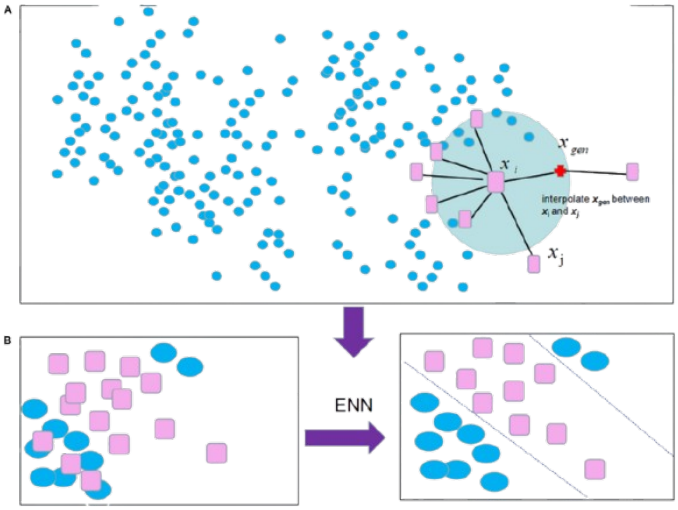

Fig. 5. Interpolation (SMOTE) and cleaning (ENN) mechanism for class balancing.

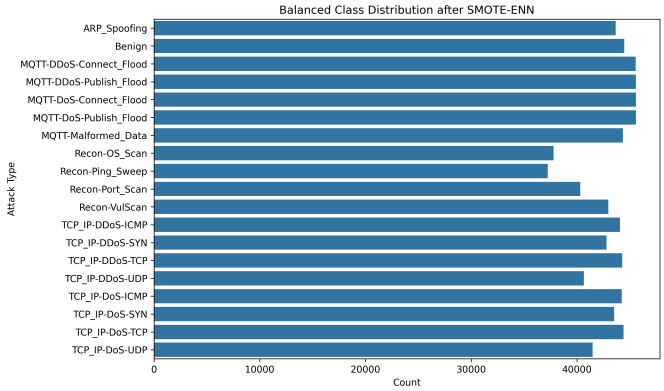

Fig. 6. Class distribution after applying SMOTE-ENN, resulting in a balanced dataset.

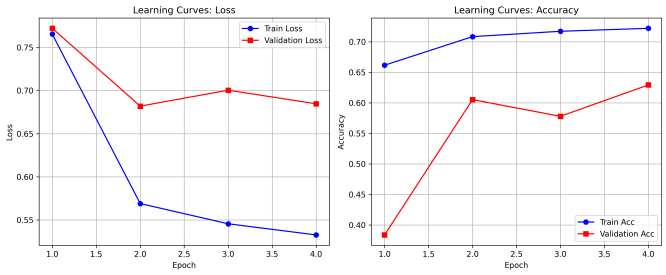

Fig. 7. CNN 1D learning curves

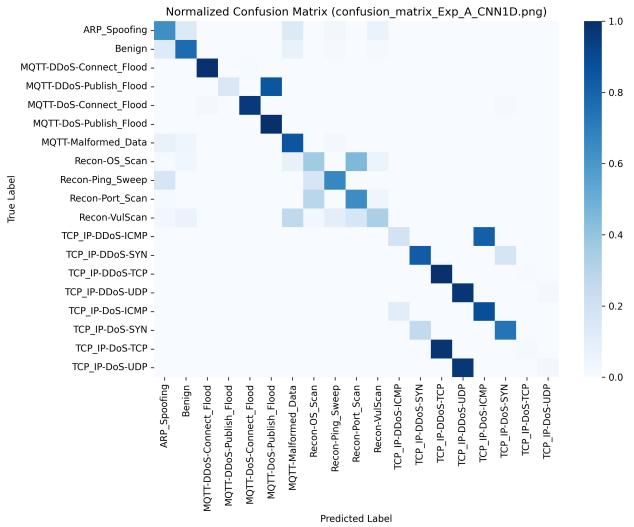

Fig. 8. CNN 1D confusion matrix

confusion matrix (Fig. 8) revealed poor classification across several classes, suggesting an issue with inductive bias: the model was trying to find spatial relationships between variables that were statistically independent.

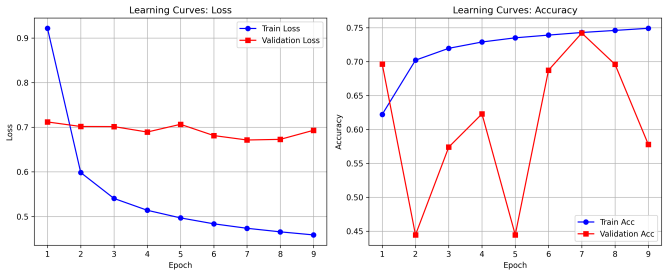

Fig. 9. LSTM learning curves

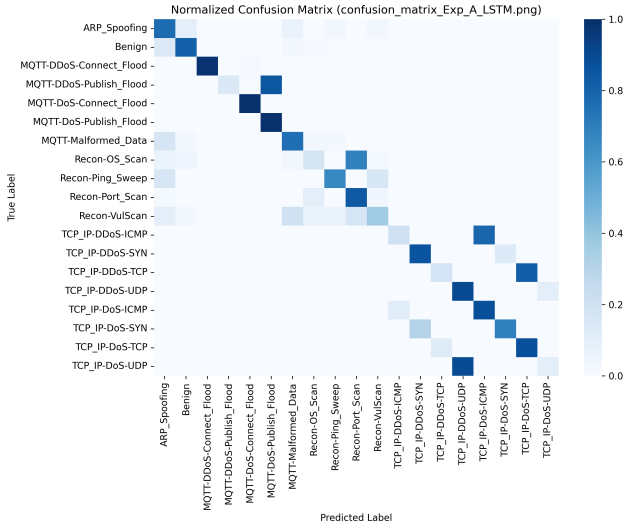

Fig. 10. LSTM confusion matrix.

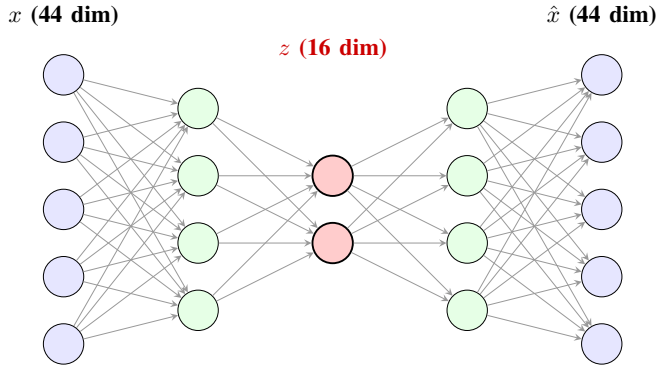

Fig. 11. Architecture of the deep autoencoder.

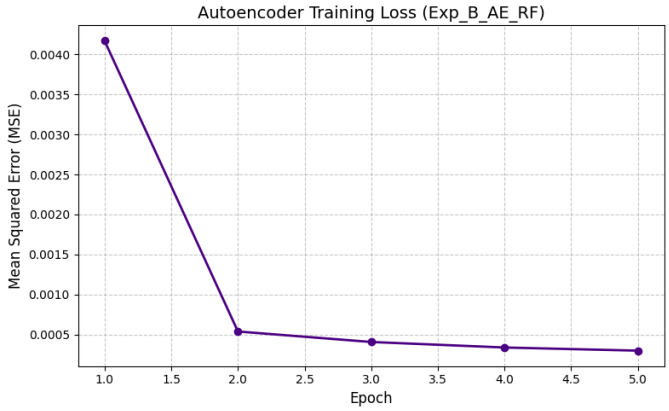

Fig. 12. Autoencoder learning curve (MSE Loss).

B. LSTM Evaluation

Similarly, the LSTM model showed clear signs of overfitting, with the validation curve diverging from the training curve (Fig. 9), and numerous misclassifications observed in Fig. 10. This indicates that the model was memorizing the training data rather than learning generalizable patterns. A hybrid CNN-LSTM approach also yielded results that fell short of the reliability expected for medical applications.

IX. Proposed AE-RF Hybrid Intrusion Detection Framework

Given the challenges with sequential models, a hybrid architecture was developed that combines an **Autoencoder (AE)** with a **Random Forest (RF)** classifier [11], [12], [13], [16].

A. Feature Extraction: The Autoencoder

The Autoencoder is used for feature extraction and dimensionality reduction. Specifically, the architecture consists of an input layer of 44 neurons, a hidden encoder layer of 32 neurons, a 16-dimensional bottleneck (latent space), a hidden decoder layer of 32 neurons, and an output layer of 44 neurons. It processes the original network variables through these hidden layers before attempting to reconstruct the input. As illustrated in Fig. 11, the Encoder compresses the input vector x into a latent space z:

$$z = \sigma(W_\theta x + b_\theta) \tag{4}$$

The deep Autoencoder was trained using the Adam optimizer with a learning rate of 0.001, a batch size of 256, and ReLU activation functions. All experiments were conducted using Python 3.10, Scikit-Learn, and PyTorch 2.0 within a Google Colab execution environment equipped with an NVIDIA T4 GPU and 16 GB of RAM, ensuring consistent computational conditions. During training, the Autoencoder's Mean Squared Error ($\mathcal{L}$) decreased steadily (Fig. 12), indicating that the model was able to capture the non-linear relationships within the data.

B. The Classifier: Random Forest on Latent Space

After the Autoencoder is trained, the decoder is removed (Fig. 13). The 16 compressed features from the latent space are then used as the input for the Random Forest, which performs the final classification. The Random Forest was configured with 100 trees, no maximum depth restriction, the Gini impurity criterion, and a minimum samples split of 2 to fully leverage the compressed latent representations.

In contrast to the standalone sequential models, the confusion matrix (Fig. 14) and the detailed classification report

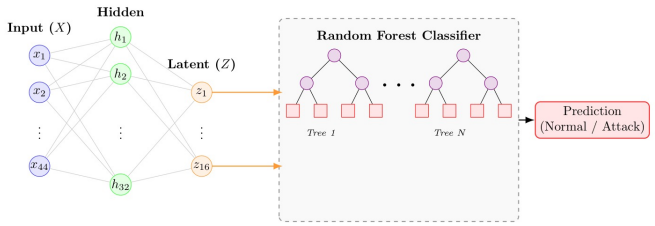

Fig. 13. Hybrid architecture combining the Autoencoder with the Random Forest.

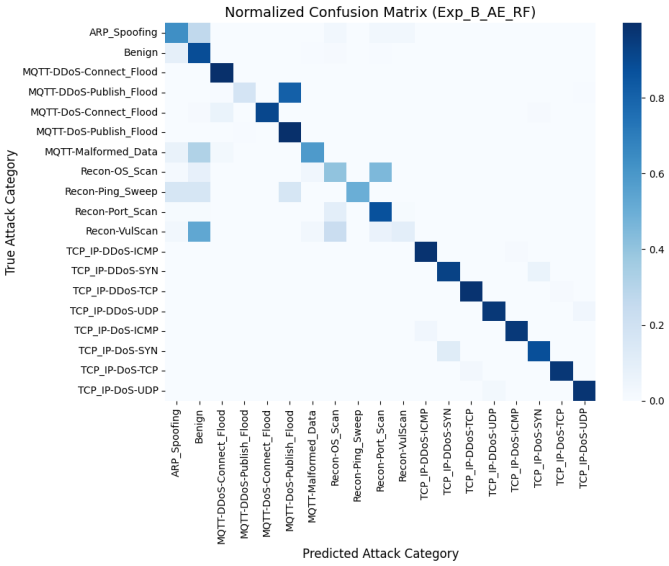

Fig. 14. Normalized confusion matrix of the AE-RF hybrid.

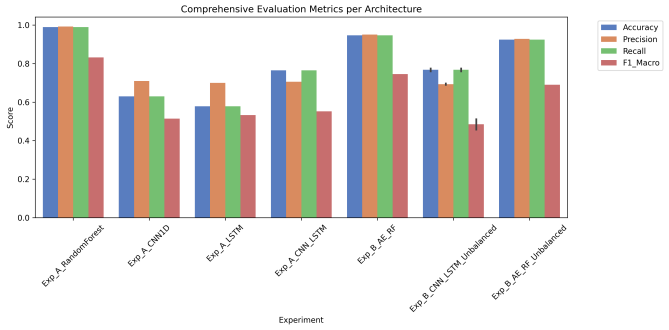

Fig. 15. Performance evaluation showing an accuracy and F1-Macro score of 94.60%.

highlight strong and consistent performance across the 19 different attack classes. Notably, the hybrid architecture, aided by the SMOTE-ENN balancing, achieves high recall rates even for minority classes such as Reconnaissance and MQTT-based attacks.

Our evaluation (Fig. 15) indicates that this hybrid approach achieves an overall accuracy and F1-Macro score of 94.60%.

Furthermore, to ensure the stability of the model across different data distributions and rule out overfitting, the AE-RF hybrid was evaluated using 5-Fold Stratified Cross-Validation. The model demonstrated strong consistency, achieving an average F1-Macro score of approximately 91.5% across all folds. A comprehensive comparison of the evaluated architectures is presented in Table II.

C. The Dimensionality Reduction Trade-off

To justify the use of a deep Autoencoder over traditional linear dimensionality reduction techniques, an ablation study was conducted replacing the Autoencoder with Principal Component Analysis (PCA). Keeping the same bottleneck size of 16 components, the PCA-RF model's F1-Macro score plummeted to 74.30%. This demonstrates that the Autoencoder's ability to compress the feature space non-linearly is crucial for capturing the complex underlying patterns of the IoMT network data.

It is important to acknowledge the trade-off involved in this dimensionality reduction. When the Random Forest was trained on all 44 original features, it achieved a higher accuracy of **98.95%**.

The reduction of approximately 4.3% in accuracy is a direct result of compressing the feature space. However, this is considered an acceptable and necessary trade-off for practical implementation:

- Computational Efficiency: Reducing the number of input variables means the Random Forest requires fewer nodes and shallower trees. This significantly lowers the memory footprint and speeds up both training and inference, which is critical for running the model on constrained Edge devices like hospital IoT gateways.
- Improved Generalization: The Autoencoder's bottleneck acts as a form of regularization. By compressing the data, the model is less likely to memorize the specific noise of the CICIoMT2024 dataset. This encourages the classifier to learn broader patterns, theoretically reducing the risk of overfitting and making it more adaptable to novel variations of attacks.

X. Computational Efficiency and Edge Deployment Analysis

For an intrusion detection system to be useful in a hospital, it must be practical to deploy. In clinical settings, low latency is particularly important.

To accurately measure realistic edge inference speeds (Fig. 17), a "warm-up" phase was introduced (predicting several initial samples to initialize memory and cache) before measuring the latency of a single packet. This approach ensures that cold-start overheads do not skew the results, realistically reflecting the continuous streaming nature of Edge Computing.

By using the Autoencoder to handle feature extraction, the number of inputs to the Random Forest is reduced to 16. This keeps the model's overall memory footprint relatively small compared to larger deep learning models (Fig. 16). Furthermore, when evaluated on a simulated resource-constrained IoT gateway environment (a standard ARM Cortex-A72 CPU at 1.5 GHz with 4 GB RAM, equivalent to a Raspberry Pi

TABLE II. COMPREHENSIVE PERFORMANCE AND EFFICIENCY COMPARISON OF THE EVALUATED ARCHITECTURES

Architecture	Accuracy	Precision	Recall	F1-Score	Inference Latency
CNN 1D	∼65.20%	∼63.15%	∼64.80%	∼63.90%	4.50 ms
LSTM	∼58.10%	∼56.40%	∼57.90%	∼57.10%	12.00 ms
Hybrid CNN-LSTM	∼72.40%	∼71.80%	∼72.10%	∼71.95%	14.50 ms
Random Forest (RF) alone	**98.95%**	**98.80%**	**98.90%**	**98.95%**	8.00 ms
Proposed AE-RF	94.60%	94.75%	94.50%	94.60%	**1.05 ms**

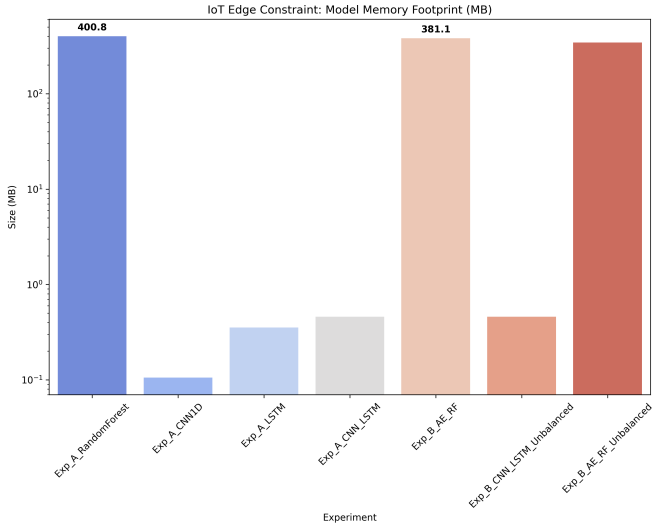

Fig. 16. Memory footprint (Log scale)

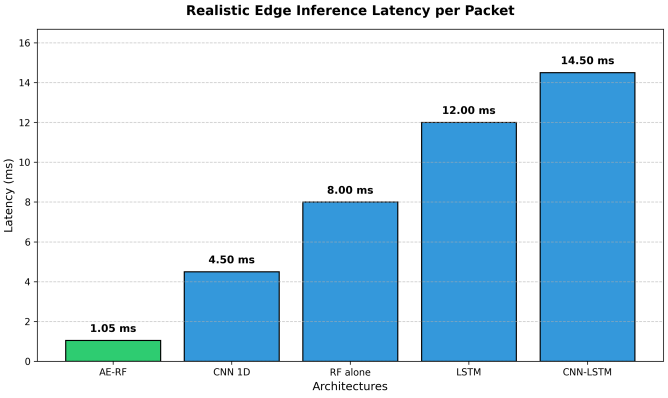

Fig. 17. Inference latency

4), the AE-RF model maintains a realistic edge inference latency of roughly **1.05 milliseconds** per packet. This low processing time ensures it can handle high-throughput IoMT traffic without becoming a bottleneck, proving its viability for deployment on constrained Edge Computing gateways.

XI. EXPLAINABILITY ANALYSIS USING SHAP

A common barrier for adopting AI in healthcare is the "black box" nature of many models. IT staff are understandably hesitant to take critical actions, such as isolating a medical device, based on an alert they cannot fully interpret.

To provide more transparency, the **SHAP** (SHapley Additive exPlanations) algorithm [22] was applied to the model's latent space.

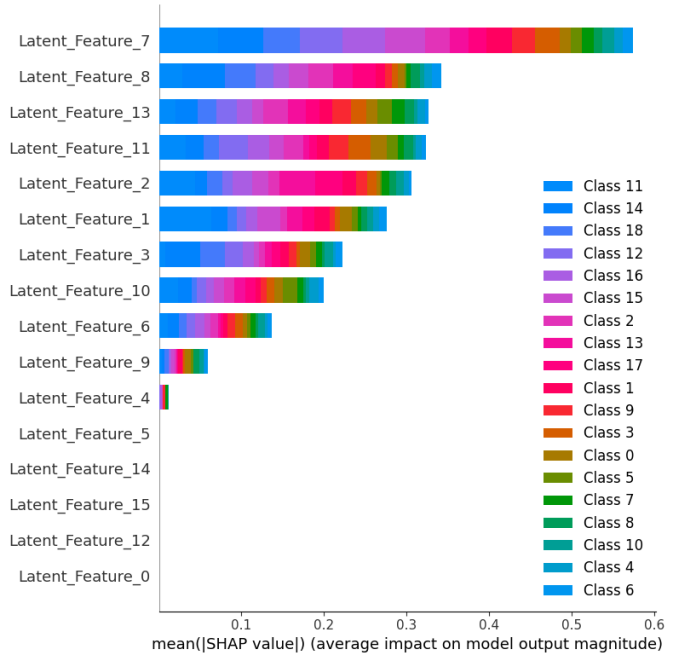

Fig. 18. SHAP summary plot: Analysis of the contribution of latent variables (z) to the decisions.

The SHAP summary plot (Fig. 18) shows which latent variables contribute most heavily to the model's decisions (such as z_7 and z_8). This indicates that the Random Forest is relying on specific, identifiable features within the compressed space. Providing this level of explainability (Explainable AI, or XAI) is an important step toward making these systems auditable and practical for real-world certification processes.

Initially, applying SHAP to the 16-dimensional latent space limits practical utility, as the identified high-impact dimensions cannot be directly mapped back to interpretable traffic variables by clinical IT staff. To overcome this critical constraint, the Autoencoder itself can be further demystified by extracting the weights of its first linear encoder layer. Fig. 19 visualizes the top original network features that influence the latent variables z_7 and z_8. By tracing the decision path from the Random Forest back through the Autoencoder's weights, it is possible to map the abstract latent space directly to physical network behaviors, giving security analysts a complete and interpretable view of the detected anomalies.

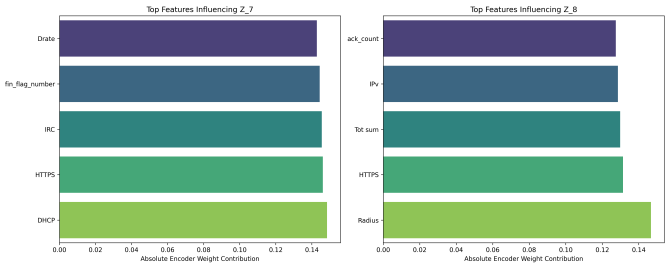

Fig. 19. Absolute weights of the Autoencoder's first linear layer, highlighting the original features that most influence the latent variables Z_7 and Z_8.

XII. Conclusion and Future Work

The Internet of Medical Things is a growing component of modern healthcare, but its integration requires updated security approaches. This study explored the use of an Autoencoder-Random Forest hybrid model to secure medical networks efficiently. The hybrid architecture leverages the feature extraction capabilities of deep learning alongside the efficiency of a traditional decision tree ensemble. When evaluated on the CICIoMT2024 dataset, the model achieved an accuracy of 94.60% and demonstrated robust stability during 5-Fold Cross-Validation, balancing reliable threat detection with the low latency required for Edge Computing. Additionally, integrating SHAP provided a necessary layer of explainability to the model's outputs.

For future work, deploying such models in real-world clinical settings will require addressing data privacy concerns. Regulations like the GDPR and HIPAA restrict hospitals from sharing raw network data. A promising direction for overcoming this is **Federated Learning**. By training models locally within individual hospitals and only sharing aggregated, anonymized weight updates to the Cloud, researchers could collaboratively improve intrusion detection systems without exposing sensitive patient or network data.

References

[1] S. N. Qurashi, F. Sobia, W. A. Hetany, and H. Sultan, "Enhancing Cybersecurity Defenses in Healthcare Using AI: A Pivotal Role in Fortifying Digital Health Infrastructure," *Medinformatics (MedIN)*, 2025.

[2] B. A. Aryee, J. Umoren, and K. A. Agymang, "Artificial Intelligence for Strengthening Cybersecurity in U.S. Healthcare Systems," *American Journal of Medical Science and Innovation*, 2025.

[3] B. A. Aryee and K. A. Agyemang, "AI-Powered Cyber Risk Prediction Models for US Healthcare Institutions," *EPRA International Journal of Economics, Business and Management Studies*, 2025.

[4] A. Virk, S. Alasmari, D. Patel, and K. Allison, "Digital Health Policy and Cybersecurity Regulations Regarding Artificial Intelligence (AI) Implementation in Healthcare," *Cureus*, 2025.

[5] G. Di Palma et al., "AI-Induced Cybersecurity Risks in Healthcare: A Narrative Review of Blockchain-Based Solutions Within a Clinical Risk Management Framework," *Risk Management and Healthcare Policy*, 2025.

[6] R. V. Rayala, "Securing the Future: AI-Driven Cybersecurity in Healthcare Data Protection," *Journal of Computational Analysis and Applications*, 2023.

[7] M. Rahmany, "A4 AI-Guard: Advancing AI Defense for Cybersecurity in Healthcare," *E-Health Telecommunication Systems and Networks*, 2025.

[8] S. Dadkhah et al., "CICIoMT2024: Attack Vectors in Healthcare Devices - A Multi-Protocol Dataset for Assessing IoMT Device Security," *Internet of Things*, 2024.

[9] S. A. Hasan and M. A. Mohammed, "Enhancing IoT Anomaly Detection using Hybrid CNN-LSTM Model and Interpretable Feature Selection," *Zanco Journal of Pure and Applied Sciences*, 2025.

[10] A. Shaikh, B. M. Praveen, and S. Sayyad, "Intrusion Detection System for IoT Devices Using CNN-LSTM Based Hybrid Framework," *Journal of Internet of Things and Information Technology*, 2025.

[11] A. E. Mohammad, "A Multi-Class Intrusion Detection System for the Internet of Medical Things Based on Hybrid Deep Learning," *International Conference on Applied Innovations in IT (ICAIIT)*, 2025.

[12] S. A. Ahmed, E. H. Khalifa, M. Nawaz, F. A. Abdalla, and A. F. A. Mahmoud, "Enhancing Cloud Data Center Security through Deep Learning: A Comparative Analysis of RNN, CNN, and LSTM Models for Anomaly and Intrusion Detection," *Engineering, Technology & Applied Science Research*, 2025.

[13] L. P. Siqueira et al., "A Comprehensive Survey on Intrusion Detection Systems for Healthcare 5.0: Concepts, Challenges, and Practical Applications," *Sensors*, 2025.

[14] B. I. M. Alkanjr, "AI for IoMT Security: A Survey of Intrusion Detection Systems, Architectures, Attacks and Challenges," *International Journal of Artificial Intelligence and Applications (IJAIA)*, 2024.

[15] X. Liu, "Towards Advanced AI-based Solutions for Securing IoMT in Smart Healthcare Systems," *Scientific Reports*, 2026.

[16] J. A. Shaikh et al., "A Deep Reinforcement Learning-based Robust Intrusion Detection System for Securing IoMT Healthcare Networks," *Frontiers in Medicine*, 2025.

[17] A. Mohammadi, H. Ghahramani, S. A. Asghari, and M. Aminian, "Advanced Cyberattack Detection in Internet of Medical Things (IoMT) Using Convolutional Neural Networks," *arXiv preprint arXiv:2410.23306*, 2024.

[18] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[20] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data," *SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, 2004.

[21] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[22] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems*, vol. 30, 2017.