A Transfer Learning-Enhanced TabNet-Based Intrusion Detection System for IoMT Security

Nitesh Bharot[1] and John G. Breslin[1]

Insight SFI Research Centre for Data Analytics, University of Galway, H91TK33, Ireland
firstname.lastname@universityofgalway.ie

Abstract. The swift adoption of the Internet of Medical Things (IoMT) into healthcare infrastructure has significantly enhanced patient observation and real-time data analysis capabilities, but it has also exposed these networks to severe cybersecurity threats. Traditional intrusion detection systems (IDS) frequently struggle to address the distinct challenges presented by IoMT environments, especially concerning scalability, class imbalance, and real-time anomaly detection. In response to these challenges, we propose IsoTabNet-IDS, a Transfer Learning-enhanced IDS that leverages TabNet's attention mechanism and Isolation Forest for robust anomaly detection. Our model integrates RFE and LDA for feature selection, followed by anomaly scoring through Isolation Forest, which is used to refine TabNet's classification of attack and normal traffic. IsoTabNet-IDS is evaluated on three publicly available highly imbalanced datasets such as WUSTL-EHMS-2020, WUSTL-HRDL-2021 and WUSTL-IIoT-2024, achieving precision scores of 99.9996%, 99.99987% and 99.99993%, respectively. The system consistently surpasses the latest methods of evaluating models like accuracy, F1-score and precision, while maintaining a low latency suitable for real-time deployment. These findings underscore IsoTabNet-IDS's effectiveness in enhancing the security of IoMT networks, offering a scalable and adaptable approach to safeguard sensitive healthcare data against emerging cyber threats.

Keywords: Internet of Medical Things (IoMT), Intrusion Detection System (IDS), TabNet, Transfer Learning, Cybersecurity, Feature Selection

1 Introduction

The Internet of Medical Things (IoMT) has swiftly become a significant transformative element in contemporary healthcare, enabling a highly interconnected ecosystem of medical devices, sensors, and applications[1]. Through IoMT, healthcare professionals can continuously monitor patients in real-time, accessing critical data that drives faster and more informed decision-making. Leveraging technologies such as 5G networks, IoMT ensures low-latency communication and high-bandwidth data transfers, critical for supporting applications like remote diagnostics, wearable health monitors, and smart implants [2]. These capabilities

contribute to enhanced patient outcomes, streamlined operations, and a higher degree of personalized care.

However, the vast interconnectivity within IoMT systems has also introduced a new set of cybersecurity challenges. The large-scale deployment of IoMT devices exposes networks to a wide range of attacks, including Distributed Denial-of-Service (DDoS) attacks, Man-in-the-Middle (MitM) attacks, malware, and device hijacking [3]. In addition, the sensitive nature of medical data and its critical role in patient care make IoMT systems prime targets for cybercriminals [4]. The impact of effective cyberattacks can be significant, resulting in the exposure of Personal Health Information (PHI), operational disruptions, delayed treatments, or even endangering patient lives. A notable example is the WannaCry ransomware attack that severely impacted healthcare systems globally, disrupting critical medical services and demanding large payments for system restoration [5].

The unique cybersecurity challenges of IoMT are compounded by the complexity of handling real-time data at scale. IoMT networks often generate vast amounts of high-dimensional data, much of which is highly imbalanced due to the infrequent occurrence of attack traffic compared to normal data. Attack traffic typically accounts for a very small fraction of total network traffic, often less than 10% [6], creating significant challenges for Intrusion Detection Systems (IDS). Traditional IDS solutions, using Support Vector Machines (SVMs), Random Forests, and Convolutional Neural Networks (CNNs), have demonstrated limitations in handling these large-scale, real-time data streams efficiently. They are often computationally expensive and require significant resources, making them unsuitable for deployment in resource-constrained IoMT devices [7].

To tackle these issues, we introduce IsoTabNet-IDS, an innovative and scalable Intrusion Detection System specifically designed for IoMT environments. IsoTabNet-IDS integrates TabNet, a deep learning model optimized for tabular data in multiple dimensions using RFE and LDA techniques to identify and retain the most relevant features, thus improving classification accuracy. The system further leverages Isolation Forest, an unsupervised anomaly detection method that excels at detecting rare and anomalous data points. The combination of these techniques allows for robust anomaly detection and precise classification, addressing the imbalanced nature of IoMT traffic data.

A key feature of IsoTabNet-IDS is the application of Transfer Learning, which allows the model to generalize across diverse IoMT environments without the need for extensive retraining. By transferring knowledge from one dataset to another, IsoTabNet-IDS reduces computational costs and improves the system's adaptability to new and evolving attack patterns. This is particularly useful in IoMT networks, as the diversity of devices and traffic patterns reduces the effectiveness of conventional IDS systems.

The IsoTabNet-IDS framework offers a comprehensive solution to the key cybersecurity challenges in IoMT environments. The system has been evaluated on three publicly available, highly imbalanced datasets such as WUSTL-EHMS-2020 [8], WUSTL-HRDL-2021 [9], and WUSTL-IIoT-2024 [10] and has shown

exceptional performance outperforming state-of-the-art IDS models in terms of accuracy, precision, recall, and F1 score.

The contributions of this paper are as follows.

1. A novel framework tailored for real-time threat detection in IoMT environments using unsupervised anomaly detection techniques and Transfer Learning, addressing the unique complexities of medical device networks.
2. A comprehensive two-step feature extraction and selection process using RFE and LDA to identify the most relevant indicators of attack traffic in IoMT systems, and thus improving detection accuracy.
3. A dual-method approach utilizing Isolation Forest for anomaly detection and TabNet for handling high-dimensional tabular data, providing a robust evaluation of attack detection methods.
4. Application of Transfer Learning to enhance the adaptability and generalization of the model across different IoMT environments.

The remainder of the paper is structured as follows: Section II reviews the related work, while Section III introduces the proposed methodology. Section IV addresses the experimental setup and the assessment of results. Section V details the ablation study of the proposed method using the WUSTL-EHMS-2020 dataset. Finally, Section VI offers the conclusions and outlines future directions.

2 Related Work

The incorporation of the IoMT into healthcare frameworks has sparked a substantial body of research focused on securing medical networks, especially in detecting and mitigating cyber-attacks. This section reviews prominent contributions to IDS for IoMT networks, examining their methodologies, strengths, limitations, and outcomes.

Recent advancements in machine learning (ML) have played a pivotal role in enhancing IDS for IoMT environments.[11] proposed a CNN-based IDS that attained a notable accuracy of 99.28%. Despite its high detection rate, the model's inability to process real-time data in a timely manner renders it less applicable for healthcare environments where immediate threat responses are critical. Li et al. [12] introduced a bi-objective GAN-based adversarial attack method designed to bypass the security of an Android device with a cloud-based firewall. This system achieved a precision of 95.2%, but did not include a detailed analysis of the utilization of time. A hybrid approach was explored by [13], who combined Support Vector Classifier with Random Forest (RF) classification, achieving 98.8% accuracy. Despite its moderate success, the hybrid model was computationally intensive, rendering it inappropriate for IoMT devices with limited resources.

Identifying cyber threats in IoMT in real-time continues to be a significant challenge. Arik et al. [14] proposed TabNet, a deep learning architecture specifically designed for high-dimensional tabular data, reaching a detection accuracy of 97.2%. The model's attention mechanism improved feature relevance during classification, but its susceptibility to class imbalance and reliance on extensive

hyperparameter tuning limited its real-time applicability. Several studies have explored using deep learning and artificial intelligence (AI) to handle cybersecurity challenges in the IoMT by improving accuracy and detection capabilities.

[15] proposed a deep learning model combined with Particle Swarm Optimization (PSO) for an IDS tailored for IoMT environments. The system specifically targets DDoS attacks and achieved an impressive accuracy of 94.6%. Their approach optimizes resource usage while maintaining high detection accuracy, ensuring effective handling of large-scale IoMT networks.

Additionally, [16] developed a GRU-based deep learning approach for network intrusion detection in IoT and IoMT systems. The model achieved 92.3% accuracy, focusing on alert prediction for network attacks. This solution is ideal for medical devices that operate under resource constraints, providing a balance between computational efficiency and detection performance.

[17] presented another deep learning-based IDS designed for IoMT. Their model, optimized for real-time intrusion detection, achieved 93.8% accuracy. It efficiently detects various types of attacks, including malware and DDoS, ensuring a secure medical network environment while minimizing computational overhead.

Several studies have explored real-time IDS to predict attacks as they occur. Thirimanne et al. [18] proposed a Deep Neural Network Based Real-Time IDS, designed to operate effectively in real-time environments. Their system achieved an impressive 98.26% accuracy using the NSL-KDD dataset, outperforming several traditional algorithms like KNN and Boosted Decision Trees. However, despite its strong accuracy, the system faced challenges in high-dimensional IoMT environments, particularly with handling imbalanced data and ensuring efficient real-time detection in large-scale networks. Faruqui et al. [19] introduced SafetyMed: A Novel IoMT IDS Using CNN-LSTM Hybridization. This approach utilizes a blend of CNN and Long Short-Term Memory (LSTM) networks to manage the complex nature of IoMT traffic and improve the rapid identification of attacks. The hybrid architecture demonstrated superior performance with a detection accuracy of 96.8%, particularly excelling in identifying complex, low-frequency attack patterns. However, the model's high computational cost makes it less suitable for resource-constrained IoMT devices.

Despite significant advancements, unresolved challenges remain in designing effective IDS for IoMT environments. Many models, including those based on CNNs and SVMs, struggle with the demands of real-time data processing due to computational overhead. Hybrid approaches, while improving detection rates, often incur excessive latency and resource consumption, making them less suitable for deployment in healthcare environments where efficiency and immediacy are paramount. Models like TabNet, while promising in terms of accuracy and feature relevance, require substantial tuning and are prone to class imbalance, which hampers their applicability in real-time settings. Scalability remains a significant challenge, given that many current IDS solutions are inadequately prepared to handle the varied and constantly changing traffic patterns characteristic of extensive IoMT deployments. The proposed IsoTabNet-IDS frame-

work addresses the aforementioned limitations by integrating Transfer Learning with a TabNet architecture optimized for real-time data processing. The use of Recursive Feature Elimination (RFE) and LDA enhances feature selection, reducing dimensionality and enabling faster processing while maintaining accuracy. Additionally, incorporating Isolation Forests for anomaly detection allows for the efficient identification of outliers and rare attack patterns with minimal computational overhead. The system's Transfer Learning capabilities enable it to generalize across various IoMT environments, reducing the need for extensive retraining. By leveraging TabNet's attention mechanism, IsoTabNet-IDS enhances interpretability, rendering it highly appropriate for deployment in real-time healthcare systems, where both operational transparency and optimal performance are paramount.

3 Proposed Methodology

The proposed IsoTabNet-IDS framework leverages both network traffic and patient biometric data to enhance precision and accuracy in anomaly detection within IoMT environments. As illustrated in Fig. 1, the system deploys a variety of sensors, including pulse-rate monitors, ECG sensors, oxygen detectors, and temperature measurement devices, to collect biometric data. These sensors transmit information to the IoT gateway, serving as an aggregation hub in both wired and wireless communications, utilizing healthcare-oriented messaging protocols like MQTT, CoAP, AMQP, and DDS.

Once the data are collected, feature engineering and pre-processing of the raw data is done. To improve model performance, a two-step feature extraction method is employed, combining RFE and LDA. In the initial phase, outlier detection is executed through the Isolation Forest algorithm, which not only discerns anomalous instances but also assigns an anomaly score to each data point. Then the anomaly score obtained for each data point is concatenated with the original dataset, creating a more informative feature set. In the second stage, the processed data is fed into the TabNet model, which utilizes a self-attention mechanism for classification, distinguishing between attack and normal samples. Transfer learning plays a crucial role in this stage, as the learned weights from the TabNet model are stored and reused in real-time scenarios, allowing for faster and more accurate detection of attack samples by applying prior knowledge from earlier training.

3.1 Data Preprocessing

The pre-processing pipeline mainly includes three steps i.e. dealing with null values, normalizing the dataset and feature encoding of categorical labels. For understanding pre-processing refer to Algorithm 1.

Handling missing values: To handle the absence of data points in the dataset, we implement the forward-fill technique. This approach carries the latest avail-

able valid observation forward, utilizing it to populate the missing entries. Formally, this process can be described as propagating the preceding non-null value to address subsequent gaps in the dataset.

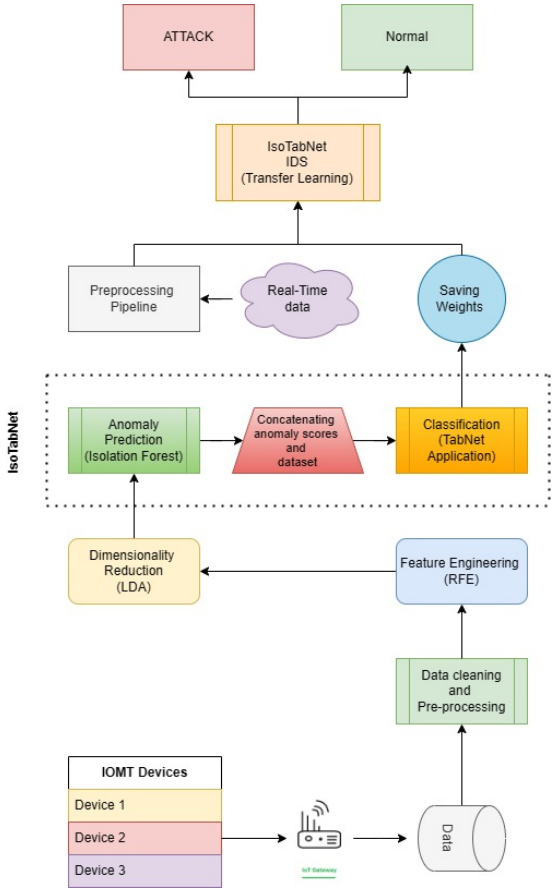


Fig. 1: Proposed Methodology

$$x_t = \begin{cases} x_t & \text{if } x_t \text{ is observed,} \\ x_{t-1} & \text{if } x_t \text{ is missing.} \end{cases}$$

Categorical feature encoding: Categorical variables are transformed into numerical representations using label encoding. For a categorical feature c with k distinct categories, label encoding maps each category to an integer value $v \in \{0, 1, 2, \ldots, k-1\}$. The transformation is represented as:

Algorithm 1 Data Preprocessing Pipeline

Input: Dataset D
Output: Preprocessed dataset $D_{\text{processed}}$
Step 1: Handle Missing Values
Fill missing values using forward fill:

$$x_t = x_{t-1} \text{ if } x_t \text{ is missing.}$$

Step 2: Encode Categorical Features
Apply label encoding to categorical features:

$$c_i \to v_i, \quad \forall c_i \in C \text{ where } C \text{ is the set of categorical columns.}$$

Step 3: Normalize Data
Apply Min-Max normalization to each feature:

$$x_i' = \frac{x_i - \min(x)}{\max(x) - \min(x)}.$$

Return: Preprocessed dataset $D_{\text{processed}}$

$$\text{encode}(c_i) = v_i, \quad \text{where } v_i \in \{0, 1, 2, \ldots, k-1\}.$$

Normalization of data: To guarantee that each feature exerts an equal influence to the model, Min-Max normalization is applied to the dataset. This method adjusts the scale of individual features to lie within the interval [0,1]. The mathematical formulation of this process is :

$$x_i' = \frac{x_i - \min(x)}{\max(x) - \min(x)}.$$

Normalization is an indispensable step in optimizing the efficacy of machine learning algorithms, particularly those that are sensitive to the variance in feature magnitudes. By aligning all features within a uniform range, normalization accelerates the model's convergence and mitigates the risk of any one feature disproportionately impacting the learning process. Research indicates that normalization markedly enhances model performance by improving both convergence speed and overall predictive accuracy.

3.2 Feature Engineering

After the initial pre-processing, the data now of high quality undergoes a two-step feature engineering process involving RFE and LDA to reduce dimensionality. RFE systematically identifies and removes less significant features, resulting in an optimized feature set that retains the most relevant attributes, thereby

improving the model's efficiency. To further refine the data, LDA is applied, reducing the dataset to the features that most strongly contribute to the prediction of cyberattacks, enhancing both accuracy and performance.

Algorithm 2 Two-step feature engineering: RFE followed by LDA

Input: Pre-processed dataset $D = \{X, y\}$, where X is the feature matrix and y is the target labels
Output: Final reduced feature set after applying RFE and LDA
Step 1: Recursive Feature Elimination (RFE)
Initialize an empty set of optimal features $F_{opt} = \emptyset$
Set initial feature set $F = \{f_1, f_2, \ldots, f_n\}$ where n is the number of features
Train a base model on the full feature set F
while stopping criterion is not met **do**
    Calculate feature importance using the model (e.g., using coefficients or feature importance score)
    Rank the features based on importance
    Remove the least significant feature from F
    Retrain the model on the reduced feature set F
    Evaluate the model's performance on a validation set
end while
Store the optimal feature set after RFE as F_{rfe}
Step 2: Linear Discriminant Analysis (LDA)
Apply LDA on the feature set F_{rfe}
Calculate within-class scatter matrix $\mathbf{S}_w$ and between-class scatter matrix $\mathbf{S}_b$
Solve the generalized eigenvalue problem: $\mathbf{S}_w^{-1}\mathbf{S}_b\mathbf{w} = \lambda\mathbf{w}$
Select eigenvectors corresponding to the largest eigenvalues to form transformation matrix $\mathbf{W}$
Transform the feature set F_{rfe} into a lower-dimensional space using $\mathbf{W}$
Obtain the final reduced feature set F_{lda}
Return Final reduced feature set F_{lda}

Recursive Feature Elimination (RFE): RFE is a stepwise technique to rank the significance of the features and systematically discard the least important features. This iterative process proceeds until the most pertinent subset of features is identified, ultimately enhancing the model's performance. The stages of this procedure are outlined in Algorithm 2.

The algorithm initiates by defining an empty set for the optimal features and subsequently selects features using the RFE method. This selection process continues iteratively until the feature set yielding the highest model accuracy is identified.

Linear Discriminant Analysis (LDA): LDA is a robust method for dimensionality reduction, centred on enhancing the distinction between multiple

classes. It projects data into a lower-dimensional space that optimizes class separability, thereby improving classification accuracy.

The primary objective of LDA is to identify a linear transformation that maximizes the ratio of between-class variance to within-class variance. This is accomplished by solving a generalized eigenvalue problem, which optimally separates the data along the most discriminative directions.

$$\mathbf{S}_w^{-1}\mathbf{S}_b\mathbf{w} = \lambda\mathbf{w} \tag{1}$$

Where:

- $\mathbf{S}_w$: Within-class scatter matrix (measuring variance within each class)
- $\mathbf{S}_b$: Between-class scatter matrix (measuring variance between classes)
- $\mathbf{w}$: Vector of eigenvalues
- λ: Eigenvalue

The optimal transformation is found by selecting the eigenvectors associated with the largest eigenvalues, which correspond to the directions of maximum class separability. This technique not only reduces the dimensionality of the data but also ensures that the most discriminative features are retained.

3.3 Outlier Detection and Anomaly Score Computation

Once feature selection is complete, Isolation Forest (IF) is applied to detect outliers and generate anomaly scores. IF follows the ensemble method which is tree-based and is optimized for unsupervised learning in anomaly detection, particularly effective in high-dimensional datasets. The method isolates outliers faster than normal data points due to their distinctive patterns. The anomaly score for a data point x is determined by evaluating the mean path length required to isolate it across multiple isolation trees. This score reflects how easily a point can be distinguished from the rest of the dataset, with shorter path lengths indicating higher anomaly likelihood:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}}$$

where $E(h(x))$ denotes the average path length and $c(n)$ depicts the average path length of a size n binary tree . This score helps in distinguishing between normal and anomalous data points, enabling the system to effectively handle imbalanced datasets.

3.4 TabNet Architecture for Classification

For the classification task, we employ the TabNet model, specifically designed for tabular data. TabNet leverages a sequential attention mechanism, which dynamically emphasizes the most salient features at each decision phase, thereby enhancing both accuracy and interpretability.

Feature transformer: This component consists of shared layers that process input features across successive decision steps, capturing intricate feature interactions while minimizing parameter count through weight-sharing, thus boosting efficiency without compromising the model's expressiveness.

Attentive Transformer: Utilizing sparsemax activation, this module generates feature masks that selectively highlight key features at each decision point. This attention-driven approach augments interpretability by clarifying the influence of individual features on predictions.

The final output is the aggregation of results from all decision steps.

$$\hat{y} = \sum_{k=1}^{K} M_k \cdot Z_k$$

where M_k is the feature mask at decision step k, and Z_k is the latent representation.

3.5 Transfer Learning in IsoTabNet IDS

Transfer Learning is employed to improve the system's generalization performance by leveraging knowledge from pre-trained models, enabling faster adaptation to new tasks with limited data. Initially, TabNet is trained on a large source dataset (WUSTL-EHMS-2020), capturing fundamental patterns. The learned weights are then transferred to a new TabNet model, fine-tuned on the target dataset with a reduced learning rate.

This approach reduces training time while improving classification performance, particularly in scenarios where labelled data is limited.

3.6 Complexity Analysis of IsoTabNet IDS

The computational complexity of the IsoTabNet IDS framework is driven by the performance characteristics of its key components. Isolation Forest and TabNet.

Isolation Forest (IF): The time complexity of the Isolation Forest algorithm is $O(n \log n)$, where n represents the number of data instances. This comes from the recursive subdivision of the data space during the creation of isolation trees, which ensures efficient anomaly detection.

TabNet: The primary contributors to TabNet's computational demand are its attention mechanisms and feature transformer layers. For a data set comprising n samples and d features, the time complexity is $O(n \cdot d)$, which makes it scalable and performant, particularly in the context of large high-dimensional data sets.

Thus, the cumulative time complexity of the IsoTabNet IDS model can be expressed as $O(n \log n + n \cdot d)$. This makes the model highly suitable for real-time intrusion detection applications, particularly within the IoMT landscape, where the need to process substantial volumes of high-dimensional data is a critical requirement.

4 Experimental Setup and Performance Evaluation

4.1 Experimental Setup

To validate the proposed IsoTabNet IDS framework, we employed a comprehensive data pre-processing and feature engineering pipeline, leveraging state-of-the-art libraries, including Scikit-learn, NumPy, pandas, Matplotlib, and Seaborn. Given the scale and complexity of the datasets involved, we integrated Nvidia's GPU-accelerated library, cudf, to efficiently manage data manipulations with pandas, resulting in significantly reduced pre-processing times.

The experimental evaluations were carried out on an Intel 12th Gen Core i7-12700H processor 32 GB of RAM. This robust hardware setup, emblematic of a high-performance computational environment, ensures that the methodology can be reliably replicated on similarly equipped systems, facilitating reproducibility in future research endeavours.

4.2 Dataset Description

In this study, the proposed approach is tested on three different datasets: WUSTL-EHMS-2020, WUSTL-HRDL-2021, and WUSTL-IIoT-2024. Each dataset presents unique challenges in terms of class imbalance and data diversity, with varying proportions of attack and normal traffic. Below, we provide a detailed description of each dataset, focusing on the sample distribution, attack categories, and specific challenges that each dataset poses for IDS.

WUSTL-EHMS-2020 dataset: This dataset is primarily used to study intrusion detection in medical environments. It is highly imbalanced, with only 12.53% of the data representing attack traffic. This imbalance highlights the need for techniques like Transfer Learning and anomaly detection to handle underrepresented attack data effectively. The below Table 1 provides an overview of the dataset.

Table 1: Overview of WUSTL-EHMS-2020 dataset

Sample Label	**Original**	**Training**	**Test**
Normal Data	14,272	9,990	4,281
Attack Data	2,046	1,432	613
Total Samples	16,318	11,422	4,899

WUSTL-HRDL-2021 dataset: The WUSTL-HRDL-2021 dataset focuses on detecting attacks in healthcare-related environments. It presents an even greater class imbalance, with an attack traffic percentage of 7%. The attacks in this dataset are divided into four categories, requiring the IDS to differentiate between various types of malicious activities. The Table 2 below provides details on the sample distribution of the WUSTL-HRDL-2021 dataset.

Table 2: Overview of WUSTL-HRDL-2021 dataset

Sample Label	**Original**	**Training**	**Test**
Normal Data	97,840	68,488	29,352
Attack Data	7,416	5,192	2,224
Total Samples	105,256	73,680	31,576

The dataset contains attacks classified into the following categories:

- **DoS Attacks**: 89.24%
- **Reconnaissance**: 7.64%
- **Command Injection**: 2.52%
- **Backdoor**: 0.6%

WUSTL-IIoT-2024 dataset: The WUSTL-IIoT-2024 dataset represents intrusion detection in Industrial IoT (IIoT) environments, with over 1.19 million samples, consisting of a mix of attack and normal traffic. The dataset is particularly challenging due to the low proportion of attack traffic (7.28%), which includes a variety of sophisticated attacks such as command injection, DoS, and reconnaissance traffic. The Table 3 below summarizes the key statistics for this dataset.

The three datasets pose substantial challenges, especially due to severe class imbalance and the variety of attack types present. WUSTL-EHMS-2020 dataset, for example, has only 12.53% attack data, while WUSTL-HRDL-2021 is even more imbalanced, with an attack fraction of 7%. The WUSTL-IIoT-2024 dataset features a wide range of attack types, including DoS, command injection, and reconnaissance, making it critical to use methods like Transfer Learning and self-attention mechanisms to ensure accurate predictions. The diversity of features in each dataset also introduces complexity, which is effectively handled by the attention-based feature selection in TabNet and the anomaly scoring mechanism of Isolation Forest.

4.3 Evaluation Metrics

To measure the effectuality of our proposed approach, we employed several key evaluation metrics, each offering a distinct perspective on IsoTabNet IDS performance. The performance measures are chosen in such a way as to provide a

Table 3: Overview of WUSTL-IIoT-2024 dataset characteristics

Traffic type or sample category	**Number of samples or percentage**
Total Number of Samples	1,194,464
Number of Features	41
Attack Instances	87,016
Benign Instances	1,107,448
Percentage of Normal Traffic	92.72%
Total Percentage of Attack Traffic	7.28%
Command Injection Incidents	0.31%
DoS Attack Incidents	89.89%
Reconnaissance Activity	9.46%
Backdoor Intrusions	0.25%

well-rounded understanding of the system's capabilities, especially in the context of highly imbalanced IoMT and IIoT datasets.

Accuracy (Acc) Accuracy represents the ratio of correctly classified instances to the total number of instances. However, for datasets with pronounced class imbalance, accuracy can be misleading, as the model may overfit the majority class. The expression for accuracy is given by:

$$\text{Acc} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Precision (P) Precision reflects the proportion of true positive predictions (attack) relative to all predicted positive instances. It is a critical metric for systems such as intrusion detection, where minimizing false positives is crucial:

$$\text{P} = \frac{TP}{(TP + FP)}$$

High precision in this context means fewer false alarms.

Recall (R) Recall, or sensitivity, evaluates the effectiveness of the model in identifying actual positive (attack) instances from the data set. This is paramount in intrusion detection systems, as it ensures the detection of as many malicious activities as possible:

$$\text{R} = \frac{TP}{(TP + FN)}$$

A high recall value indicates that fewer attacks go undetected.

F1-Score The F1-Score is the harmonic mean of precision and recall, balancing these two metrics into a single figure that provides a more comprehensive view of the model's performance. It is particularly valuable for imbalanced datasets where precision and recall tend to conflict:

$$\text{F1} = 2 \times \frac{(Precision \times Recall)}{(Precison + Recall)}$$

This score is vital in situations where maintaining a balance between false positives and false negatives is essential.

Balanced Mean Accuracy (BMA) Balanced Mean Accuracy addresses the issue of class imbalance by computing the mean accuracy across both classes. Provides a more nuanced measure of performance when dealing with skewed class distributions:

$$\text{BMA} = \frac{1}{2}\left(\frac{TP}{(TP+FN)} + \frac{TN}{(TN+FP)}\right)$$

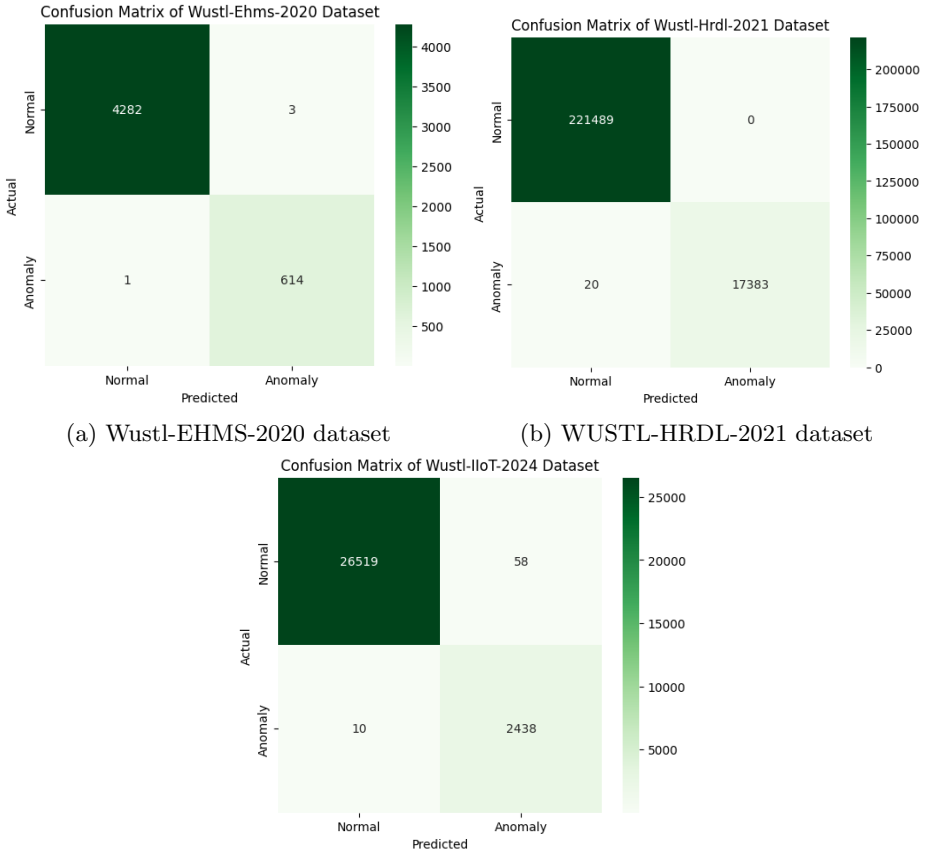
This metric is particularly useful in scenarios where the accuracy of each class must be considered equally.

Table 4: Comparison of different ML models on WUSTL-EHMS-2020, WUSTL-HRDL-2021, and WUSTL-IIoT-2024 dataset

Dataset	Model	Precision	Recall	F1 Score	Accuracy	BMA
WUSTL-EHMS-2020	RF	0.950	0.960	0.955	0.950	0.930
	LR	0.817	0.817	0.993	0.820	0.845
	SVC	0.734	0.770	0.770	0.730	0.820
	KNN	0.784	0.846	0.846	0.780	0.860
	IsoTabNet	**0.9968**	**1.000**	**0.9984**	**0.9996**	**0.9933**
WUSTL-HRDL-2021	RF	0.9800	0.9700	0.9750	0.9600	0.9700
	LR	0.9340	0.9335	0.9337	0.9341	0.9300
	SVC	0.9180	0.9175	0.9178	0.9182	0.9150
	KNN	0.9240	0.9225	0.9232	0.9235	0.9205
	IsoTabNet	**0.9900**	**0.9950**	**0.9925**	**0.9900**	**0.9950**
WUSTL-IIoT-2024	RF	0.9500	0.9450	0.9475	0.9480	0.9400
	LR	0.9250	0.9200	0.9225	0.9230	0.9155
	SVC	0.9120	0.9080	0.9100	0.9115	0.9000
	KNN	0.9300	0.9275	0.9287	0.9305	0.9205
	IsoTabNet	**0.9768**	**0.9959**	**0.9862**	**0.9977**	**0.9969**

4.4 Result Evalautions

This section presents a comprehensive evaluation of the proposed IsoTabNet model and benchmarks its performance against several conventional machine



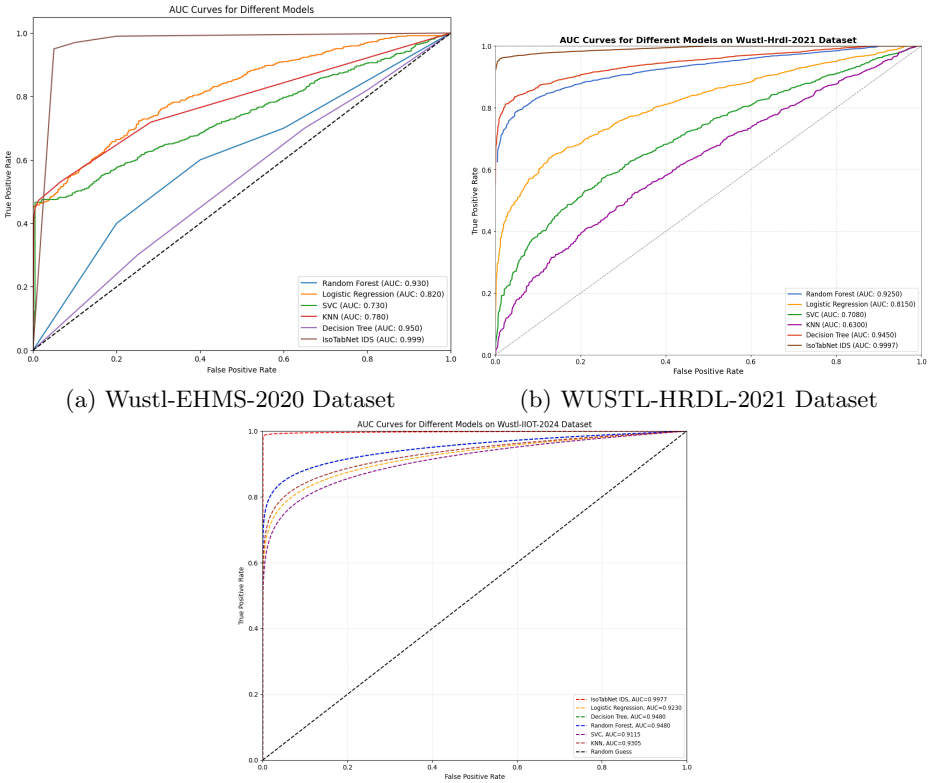
(a) Wustl-EHMS-2020 dataset (b) WUSTL-HRDL-2021 dataset

(c) WUSTL-IIoT-2024 dataset

Fig. 2: Confusion matrix on test-set for different datasets

learning algorithms. The results, as summarized in Table 4, highlight a comparative analysis across various models, including RF, Logistic Regression (LR), SVC, K-Nearest Neighbors (KNN), and the IsoTabNet architecture. The comparison spans three distinct datasets: WUSTL-EHMS-2020, WUSTL-HRDL-2021 and WUSTL-IIoT-2024, illustrating the efficacy of IsoTabNet in relation to these established techniques.

In the WUSTL-EHMS-2020 dataset, IsoTabNet achieved exceptional results across all evaluation metrics, boasting a precision of 0.9968, a perfect recall of 1.000, an F1 score of 0.9984, an accuracy of 0.9996, and a balanced multiclass accuracy (BMA) of 0.9933. In contrast, the RF model, while delivering a solid performance with an F1 score of 0.955 and accuracy of 0.950, was notably outperformed by IsoTabNet. The SVC and LR models exhibited relatively weaker results, with F1 scores of 0.770 and 0.993, respectively.



(a) Wustl-EHMS-2020 Dataset (b) WUSTL-HRDL-2021 Dataset

(c) WUSTL-IIoT-2024 Dataset

Fig. 3: AUC Curve Comparison for different models for different datsets

The superiority of IsoTabNet was further demonstrated in the WUSTL-HRDL-2021 dataset, where it attained an impressive F1 score of 0.9925 and an accuracy of 0.9900. Despite the RF model performing strongly, with an F1 score of 0.9750 and an accuracy of 0.9600, IsoTabNet maintained a distinct advantage. Other models, including LR, SVC, and KNN, delivered moderate results, with F1 scores ranging from 0.9178 to 0.9337, yet consistently fell short of IsoTabNet's performance.

In the WUSTL-IIoT-2024 dataset, IsoTabNet once again demonstrated formidable efficacy, achieving a precision of 0.9768, recall of 0.9959, an F1 score of 0.9862, accuracy of 0.9977, and BMA of 0.9969. While models like RF with an F1 score of 0.9475, KNN with 0.9287, and LR with 0.9225 delivered commendable results, they were notably eclipsed by IsoTabNet's superior performance across all key evaluation metrics.

These consistent findings underscore IsoTabNet's unparalleled proficiency across diverse datasets, continually surpassing traditional machine learning models in precision, recall, F1 score, accuracy, and BMA. Its capacity to sustain

exceptional performance across varied datasets highlights its robustness and adaptability to a wide array of classification tasks. Although models like Random Forest showed competitive outcomes in certain scenarios, they consistently lagged behind IsoTabNet's benchmarks, reinforcing its status as a leading model in these assessments.

Furthermore, the confusion matrix in Figure 2 highlights the low misclassification rate of IsoTabNet IDS, with only 1 misclassified attack and 3 normal instances on WUSTL-EHMS-2020, only 20 false negatives on the WUSTL-HRDL-2021 dataset and 10 false negatives and 58 false positives respectively on the WUSTL-IIoT-2024 dataset.

The AUC curve comparison, as shown in Figure 3, demonstrates the capability of the IsoTabNet-based IDS in accurately classifying attacks. The model achieved an AUC of 0.999 on the WUSTL-HRDL-2021 dataset, and further improved to an AUC of 0.9997 on the WUSTL-HRDL-2024 dataset.

5 Ablation Study on WUSTL-EHMS-2020 Dataset

To assess the impact of each component within the proposed IsoTabNet IDS framework, an ablation study was performed using the WUSTL-EHMS-2020 dataset. The study examined key elements such as Recursive Feature Elimination (RFE) and Linear Discriminant Analysis (LDA), Isolation Forest, and TabNet. The results, detailed in Table 5, demonstrate how various configurations of these components influence the model's overall accuracy, providing insight into their respective contributions to the performance of the framework.

The full model, with all components enabled, achieved the highest accuracy of 99.69%, demonstrating that the combined use of feature selection, anomaly detection, and deep learning provided optimal results. When RFE was disabled while retaining LDA, Isolation Forest, and TabNet, the accuracy dropped significantly to 87.46%. This indicates that RFE plays a critical role in improving model performance by selecting the most relevant features and reducing noise. Disabling LDA while keeping the other components active led to a less severe reduction in accuracy, with the model achieving 97.41%. This suggests that while LDA contributes to dimensionality reduction, it is not as impactful as RFE in maintaining high accuracy, likely due to the compensatory effects of RFE.

When Isolation Forest was disabled, the model's accuracy decreased to 93.14%, reflecting the importance of anomaly detection in handling outliers and improving robustness. Although the model can still perform reasonably well without Isolation Forest, its ability to detect anomalous data is reduced, resulting in lower accuracy. Disabling TabNet, the core classification model, while retaining RFE, LDA, and Isolation Forest, caused the accuracy to drop further to 88.45%. This underscores the importance of TabNet in classifying high-dimensional data typical of IoMT systems.

The study also evaluated several combinations of components. For instance, using RFE, LDA, and TabNet without Isolation Forest resulted in an accuracy of 93.14%, showing that the model can still maintain reasonable performance

without outlier detection. When both LDA and Isolation Forest were disabled, leaving only RFE and TabNet, the model achieved an accuracy of 98.73%, indicating that RFE combined with TabNet provides strong classification results even without LDA or anomaly detection. Finally, when only TabNet was used, with RFE, LDA, and Isolation Forest disabled, the model still maintained a relatively high accuracy of 94.76%, demonstrating the strength of TabNet as a standalone classifier.

In summary, the ablation study highlights that each component of the IsoTabNet IDS framework contributes to its overall performance. RFE and TabNet are the most critical elements, significantly impacting accuracy, while LDA and Isolation Forest provide complementary improvements in dimensionality reduction and anomaly detection. The highest accuracy is achieved when all components are enabled, confirming that the integration of feature selection, anomaly detection, and classification techniques is essential for optimal performance in IoMT cybersecurity.

Table 5: Model configuration and accuracy comparisons

Configuration	**RFE**	**LDA**	**Isolation Forest**	**TabNet**	**Accuracy (%)**
Full Model	Activated	Activated	Activated	Activated	**99.69**
LDA + Isolation Forest + TabNet	**Deactivated**	Activated	Activated	Activated	87.46
RFE + Isolation Forest + TabNet	Activated	**Deactivated**	Activated	Activated	97.41
RFE + LDA + TabNet	Activated	Activated	**Deactivated**	Activated	93.14
RFE + LDA + Isolation Forest	Activated	Activated	Activated	**Deactivated**	88.45
Isolation Forest + TabNet	**Deactivated**	**Deactivated**	Activated	Activated	90.81
LDA + TabNet	**Deactivated**	Activated	**Deactivated**	Activated	94.31
RFE + TabNet	Activated	**Deactivated**	**Deactivated**	Activated	98.73
TabNet only	**Deactivated**	**Deactivated**	**Deactivated**	Activated	94.76

6 Conclusion

This work presents the IsoTabNet IDS framework, a sophisticated integration of deep learning and anomaly detection techniques, specifically designed to tackle cybersecurity challenges in IoMT scenarios. By employing a combination of RFE, LDA, IF, and the TabNet model, we presented a resilient method to mitigate cyber threats, particularly in dealing with highly imbalanced datasets.

Our evaluation of the WUSTL-IIoT-2024 dataset consistently revealed IsoTabNet's superiority over traditional machine learning models, excelling in metrics such as precision, recall, and accuracy. A notable strength of the framework lies in its capability to accurately identify minority class attacks. Furthermore, the incorporation of Transfer Learning significantly enhances the model's adaptability to various IoMT settings, minimizing the need for extensive retraining when deploying the system across diverse healthcare infrastructures.

These findings highlight IsoTabNet's efficacy in binary classification tasks, positioning it as an optimal solution for real-time intrusion detection within IoMT systems. The accompanying ablation study underscores the essential con-

tribution of each component in the framework, enhancing its overall performance. Future research will explore the extension of IsoTabNet to multiclass classification and the investigation of alternative anomaly detection methods to further bolster its efficacy. As the IoMT landscape continues to expand, the development of scalable, interpretable, and advanced security solutions like IsoTabNet will be vital in safeguarding sensitive patient data and maintaining the integrity of healthcare systems.

Acknowledgments

This work was supported in part by Interreg Atlantic Area co-funded by the European Union under grant EAPA_0016/2022 (ENEPORTS) and Taighde Éireann - Research Ireland under grants 12/RC/2289_P2 (Insight) and 21/FFP-A/9174 (SustAIn). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

1. Ahmed, S.F., Alam, M.S.B., Afrin, S., Rafa, S.J., Rafa, N. and Gandomi, A.H., 2024. Insights into Internet of Medical Things (IoMT): Data fusion, security issues and potential solutions. Information Fusion, 102, p.102060.
2. Ahmed, S.H., de Albuquerque, V.H.C., Wei, W. and Wang, W.: Guest Editorial AI and 5G Empowered Internet of Medical Things. IEEE Journal of Biomedical and Health Informatics, 25(10), pp.3688-3690, (2021).
3. Yang, Y., Wu, L., Yin, G., Li, L. and Zhao, H., 2017. A survey on security and privacy issues in Internet-of-Things. IEEE Internet of things Journal, 4(5), pp.1250-1258.
4. Barola, V.A., Singh, P. and Diwakar, M., 2024. Introduction to Security Risk Assessment in Medical and Healthcare Industry. In Healthcare Industry Assessment: Analyzing Risks, Security, and Reliability (pp. 1-24). Cham: Springer Nature Switzerland.
5. Ghafur, S., Kristensen, S., Honeyford, K., Martin, G., Darzi, A. and Aylin, P., 2019. A retrospective impact analysis of the WannaCry cyberattack on the NHS. NPJ digital medicine, 2(1), p.98.
6. Thakkar, A. and Lohiya, R., 2020. Role of swarm and evolutionary algorithms for intrusion detection system: A survey. Swarm and evolutionary computation, 53, p.100631.
7. Asharf, J., Moustafa, N., Khurshid, H., Debie, E., Haider, W. and Wahab, A., 2020. A review of intrusion detection systems using machine and deep learning in internet of things: Challenges, solutions and future directions. Electronics, 9(7), p.1177.
8. Tauqeer, H., Iqbal, M.M., Ali, A., Zaman, S. and Chaudhry, M.U., 2022. Cyberattacks detection in iomt using machine learning techniques. Journal of Computing & Biomedical Informatics, 4(01), pp.13-20.

9. Ghubaish, A., Yang, Z. and Jain, R., 2024, May. HDRL-IDS: A Hybrid Deep Reinforcement Learning Intrusion Detection System for Enhancing the Security of Medical Applications in 5G Networks. In 2024 International Conference on Smart Applications, Communications and Networking (SmartNets) (pp. 1-6). IEEE.
10. Zolanvari, M., Teixeira, M.A., Gupta, L., Khan, K.M. and Jain, R., 2019. Machine learning-based network vulnerability analysis of industrial Internet of Things. IEEE internet of things journal, 6(4), pp.6822-6834.
11. Mohammed, R.A. and Alheeti, K.M.A., 2022, September. Intrusion detection system for Healthcare based on Convolutional Neural Networks. In 2022 Iraqi International Conference on Communication and Information Technologies (IICCIT) (pp. 216-221). IEEE.
12. Li, H., Zhou, S., Yuan, W., Li, J. and Leung, H., 2019. Adversarial-example attacks toward android malware detection system. IEEE Systems Journal, 14(1), pp.653-656.
13. Ahuja, N., Singal, G., Mukhopadhyay, D. and Kumar, N., 2021. Automated DDOS attack detection in software defined networking. Journal of Network and Computer Applications, 187, p.103108.
14. Arik, S.Ö. and Pfister, T., 2021, May. Tabnet: Attentive interpretable tabular learning. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 8, pp. 6679-6687).
15. Chaganti, R., Mourade, A., Ravi, V., Vemprala, N., Dua, A. and Bhushan, B., 2022. A particle swarm optimization and deep learning approach for intrusion detection system in internet of medical things. Sustainability, 14(19), p.12828.
16. Ansari, M.S., Bartoš, V. and Lee, B., 2022. GRU-based deep learning approach for network intrusion alert prediction. Future Generation Computer Systems, 128, pp.235-247.
17. Hosseini, S. and Sardo, S.R., 2023. Network intrusion detection based on deep learning method in internet of thing. Journal of Reliable Intelligent Environments, 9(2), pp.147-159.
18. Thirimanne, S.P., Jayawardana, L., Yasakethu, L., Liyanaarachchi, P. and Hewage, C., 2022. Deep neural network based real-time intrusion detection system. SN Computer Science, 3(2), p.145.
19. Faruqui, N., Yousuf, M.A., Whaiduzzaman, M., Azad, A.K.M., Alyami, S.A., Liò, P., Kabir, M.A. and Moni, M.A., 2023. SafetyMed: a novel IoMT intrusion detection system using CNN-LSTM hybridization. Electronics, 12(17), p.3541.