Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 4, 2026

A TRUST-BASED ENSEMBLE MACHINE LEARNING FRAMEWORK FOR INTRUSION DETECTION IN MEDICAL INTERNET OF THINGS ENVIRONMENTS

[1]Muhammad Bilal Abid
[1]Department of Advanced Internetwork Engineering, Glasgow Caledonian University
imbilalabid@gmail.com

Keywords
Medical Internet of Things (IoMT); Intrusion Detection; Ensemble Learning; Mutual Information; SMOTE; WUSTL-EHMS; Trust-Based Security

Article History
Received on 27 Feb, 2026
Accepted on 31 March, 2026
Published on 02 April, 2026

Corresponding Author:

Abstract
The high rate of Internet of Medical Things (IoMT) devices spread in clinical settings has provided critical attack surfaces that cannot be countered by conventional security measures. This paper suggests a trust-based collective machine learning framework of intrusion detecting in healthcare internet of things networks. The framework uses Mutual Information (MI) to generate dimensionality reduction by using 45 input features to generate 34 discriminative features, and then an ensemble classifier is used with hard-voting, comprising K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM). The suggested model is tested on the WUSTL-EHMS-2020 data in a real-world healthcare monitoring benchmark with 16,318 network flow and patient biometric records and benchmarked with three individual baseline classifiers. The ensemble had a 95 percent accuracy, specificity of 0.99, and sensitivity (recall) of 0.65 on the minority attack class and AUC of 0.82, which is better than all the individual baselines. The training set was used only and Synthetic Minority Over-Sampling Technique (SMOTE) was used to reduce class imbalance. The findings indicate that the framework in question is capable of making reliable differentiation between malicious and regular data related to network operation, hence justifying the implementation of reliable, secure IoMT systems in mission-critical healthcare settings.

1. Introduction

The Internet of Medical Things (IoMT) has revolutionized the care of patients with the ability to constantly monitor them remotely, transmit biometric data in real-time, and implement clinical decision support. The wearable vital-signs monitor, smart infusion pump, and remote patient monitoring gateway devices are now integrated into clinical processes at scale. Nonetheless, this connectivity lures vulnerable patient information and life-dependent device functionality to an increasing number of cyber attacks such as man-in-the-middle, data injection, and network spoofing.

Traditional perimeter-based security models do not suit the heterogeneous, resource-constrained IoMT environment, where devices can regularly operate through open wireless networks, and with weak authentication protocols. Regulations like HIPAA and GDPR have stringent data protection requirements, yet IoMT networks do not have the computational resources to execute the cryptographic heavyweight protocols. Another opportunity involves machine learning (ML)-based anomaly detection: learning the behavioural baseline of a device or a network, the ML models can detect anomalies that may signal malicious activity without using any fixed signature databases.

Even though there is some improvement in this regard, the current literature has significant shortcomings. Single-classifier-based models are often prone to distributional shift; others do not consider the imbalance of different classes in attack datasets, inflating accuracy artificially and obscuring poor recall of the minority attack class. There are few publications that use both network traffic characteristics and patient biometrics indicators a dual-modality framework that better represents the actual threat surface of the IoMT.

The present work addresses these gaps through the following primary contributions:

- A trust-based security framework for IoMT that integrates both network flow and patient biometric data streams to characterize normal device behavior and detect intrusions.
- A Mutual Information (MI)-based feature selection pipeline that reduces the original 45-feature space to 34 discriminative features, improving model efficiency and generalization.
- A hard-voting ensemble classifier combining KNN, Random Forest, and SVM, achieving 95% accuracy and an AUC of 0.82 on the WUSTL-EHMS-2020 benchmark surpassing standalone baselines and the best prior result on this dataset (92.98% by an ANN, (Hady et al., 2020)).
- Application of SMOTE solely to the training partition, preserving the natural class distribution of the test set and yielding a more realistic estimate of real-world detection performance.

2. Related Work

There is a high research interest in the intersection of machine learning and IoMT security in the last five years. (Hady et al., 2020) created the WUSTL-EHMS-2020 testbed and tested SVM, RF, KNN, and ANN classifiers, achieving an overall maximum accuracy of 90.42% with ANN creating the benchmark that is directly comparable to the current work. However, they did not use feature scaling in their study and the long latency that they reported makes real-time deployment difficult.

(Gupta, 2024) presented a tree-based network intrusion detector model with a 92.85-accuracy with a dataset of IoMT using a variety of dimensionality reduction methods. Recent ensemble-based strategies are not also assessed in the study, and sensitivity-specificity tradeoffs that are essential in healthcare applications are not analyzed. (Astillo et al., 2021) designed a misbehavior detection system, specification-based (SMDAps), that is specific to artificial pancreas devices, with an Auroc score of over 99.9% using KNN and SVM, but the authors do not show cross-heterogeneous IoMT deployments.

(Singh et al., 2023) proposed a Dew-Cloud hierarchical federated learning architecture using HLSTM

components, and obtained 99.31 percent training accuracy on the NSL-KDD dataset. Although the federated paradigm has privacy-protective advantages, its use is restricted by the use of high-availability cloud infrastructure, which cannot be utilized in clinical environments with limited bandwidth resources. An Empirical Intelligent Agent that uses a Swarm-Neural Network with 99.5% accuracy on the ToNIoT dataset was proposed by (Nandy et al., 2021); however, the cost of this method is quite high, and is not applied to feature selection on high-dimensional IoMT data.

From a trust architecture perspective, (Al-Hamadi & Chen, 2017) described a trust-based decision-making method of health IoT systems based on Bayesian inference, whereas (Abou-Nassar et al., 2020) presented a blockchain-based DITrust chain of sustainable healthcare IoT trust management. Both methods do not combine the ML-based anomaly detection with a multi-modal (network + biometric) feature set.

The literature, therefore, demonstrates three gaps that have persisted: (i) absence of ensemble methods used explicitly to the WUSTL-EHMS benchmark with dual-modality characteristics; (ii) insufficient coverage of the issue of the class imbalance in attack detection; and (iii) inadequate coverage of the integration of trust-based frameworks and data-driven anomaly detection. The current work is particularly meant to discuss all the three.

Table 1: ***Comparative Summary of Related Work***

Author / Study	Dataset	Best Model	Accuracy	Key Limitation
(Hady et al., 2020)	WUSTL-EHMS-2020	ANN	90.42%	No feature scaling; high prediction latency
(Gupta, 2024)	IoMT (custom)	Tree Classifier	92.85%	No ensemble; limited algorithm coverage
(Astillo et al., 2021)	Artificial Pancreas	KNN / SVM	99.5% AUROC	Single device scope; no generalization
(Singh et al., 2023)	NSL-KDD	HLSTM (Federated)	99.31%	High infrastructure requirements
(Nandy et al., 2021)	ToNIoT	Swarm-NN	99.5%	High compute; no feature selection
Proposed Framework	WUSTL-EHMS-2020	Ensemble (KNN+RF+SVM)	95.00%	Single dataset; future cross-validation planned

3. Methodology

3.1 Dataset Description

Data WUSTL-EHMS-2020 was generated on a custom-built Enhanced Healthcare Monitoring System (EHMS) testbed by the researchers of Washington University in St. Louis. The testbed is a simulation of a clinical IoT system that includes medical sensors, a gateway node, a network of routers and switches, and an intrusion detection system console. Three typical types of attack were carried out against this testbed, which included man-in-the-middle (MITM), network spoofing, and data injection.

In the dataset, there are 16318 labeled records, 14272 normal (Class 0), and 2046 attack (Class 1), and 44 features. The network flow metrics include thirty-five features (e.g., source/destination counts of bytes, packet rates, jitter, loss), eight capture patient biometric metrics (temperature, SpO2, pulse rate, systolic/diastolic blood pressure, heart rate, respiration rate, ST segment), and a single feature with the value of

the binary class label of the attacker based on MAC address. This two-modality format is what separates WUSTL-EHMS-2020 among strictly network-based IoT datasets and qualifies it to be the only data format that can be used in the current trust-based paradigm.

Table 2: ***WUSTL-EHMS-2020 Dataset Summary***

Property	Value
Total Records	16,318
Normal Samples (Class 0)	14,272
Attack Samples (Class 1)	2,046
Total Features (raw)	44
Network Flow Features	35
Patient Biometric Features	8
Label Feature	1 (binary: MAC-derived)
Attack Types	MITM, Spoofing, Data Injection
Dataset Size	4.4 MB

3.2 Data Preprocessing

A four-stage pipeline was used as preprocessor. To begin with, the columns that had no informative value were eliminated (Dir and Flgs). This left the feature count at 43. Second, there was no need to use missing value imputation since the dataset has no null values. Third, six categorical values (SrcAddr, DstAddr, Sport, SrcMac, DstMac, Attack Category) were turned into numeric values through label encoding whereby each unique categorical value was assigned a unique integer. Fourth, every numerical characteristic was normalized to zero mean and unit variance (Z-score normalization) so that distance sensitive algorithms, like KNN and SVM, are not overwhelmed with high-valued features. Z-score normalization can also maintain the original data distribution besides making models converge quickly.

3.3 Feature Selection via Mutual Information

Mutual Information (MI) quantifies the statistical dependence between each feature X and the binary label Y according to the equation:

$$I(X; Y) = \Sigma_X \Sigma_Y p(x,y) \log[p(x,y) / p(x)p(y)]$$

A score of zero on the MI scale suggests that a feature and the target are statistically independent; scores above zero suggest that a feature is predictive. MI was calculated on all 43 preprocessed features and features with MI scores less than positive were removed. This process left 34 features eliminating 11 features that did not add any discriminative information. The set of features in MI-selected features is the last input of all classifiers, which provides the similarities of evaluation conditions among models.

3.4 Class Imbalance Handling

The dataset exhibits a pronounced class imbalance: the attack class constitutes only 12.5% of the total samples. To counter the subsequent bias in favor of the majority type, Synthetic Minority Over-sampling Technique (SMOTE) was only used to the training split following the 80:20 train-test split. SMOTE uses interpolating between a selected instance of minority and either one of its k nearest neighbours (k=5 default) in feature space to generate a balanced training distribution. Using SMOTE solely on training data eliminates data leakage and creates performance on test-sets that is

representative of the inherent distribution of classes in practical use.

3.5 Model Architecture

3.5.1 Baseline Classifiers

As baselines, three individual classifiers were applied. K-Nearest Neighbors (KNN) classifies the instances by the majority of the k=2 closest data points in Euclidean feature space non-parametric instance-based classification, local data density sensitive. Logistic Regression (LR) uses L-BFGS to optimize a linear decision boundary in log-odds space (max_iter=1000), which is appropriate to patterns that are linearly separable. The Support Vector Machine (SVM) is trained to learn a maximum-margin hyperplane on a kernel-induced feature space with a radial basis function (RBF) kernel with regularization parameter C=1.0 that allows non-linear separation by classes.

3.5.2 Proposed Ensemble Model

The model proposed here uses hard-voting VotingClassifier which is used to combine the predictions of three component estimators: KNN (n 2), Random Forest (RF, default hyperparameter: 100 trees, Gini impurity) and SVM (RBF kernel, C=1.0). In hard voting, a single vote is cast by each constituent classifier to a class label; the most popular class label is used as the ensemble prediction. This strategy exploits the complementary inductive biases of the three models instance-based proximity (KNN), ensemble tree-based variance reduction (RF), and geometric margin maximization (SVM) to yield a more robust decision boundary than any single model.

The ensemble mitigates the vulnerability of each of the individual classifiers: KNN is vulnerable to irrelevant features but invulnerable to linear separability; SVM is able to utilize high-dimensional spaces effectively, but is computationally intensive; RF is resistant to noise and overfitting but may be slow with large datasets. This is because by combining them with majority voting, every weakness of them is reduced and empirical findings prove that this combination is synergistic.

4. Experimental Setup

All experiments were implemented in Python 3.10 by the scikit-learn 1.3 library, executed on Google Colaboratory with an Intel Xeon CPU backend and 12 GB RAM. The data was imported to Google drive in the CSV format. The pipeline preprocessing, feature selection, SMOTE, training, and evaluation were performed as a sequential workflow which could be reproduced. The dataset was divided into stratified 80:20 train-test split (random state=42). It resulted in 13,054 training samples and 3,264 test samples. The training set had balanced classes after SMOTE. The four models (KNN, LR, SVM, Ensemble) were trained and tested in the same preprocessing conditions to make them comparable.

Accuracy, per-class, recall (sensitivity) and F1-score, specificity and AUC-ROC were used to evaluate performance. Each model was calculated on the confusion matrix and ROC curve. Given the clinical context in which false negatives (missed attacks) carry greater operational risk than false positives, sensitivity on the attack class (Class 1) and AUC are treated as primary indicators alongside overall accuracy.

5. Results and Discussion

5.1 Baseline Model Performance

Table 3 shows the per-class precision, recall and F1-score of each of the baseline classifiers. KNN (k=2) had an accuracy of 86 percent and the minority-class recall was poor (0.54) with an AUC value of 0.72. Although KNN is computationally inexpensive, it is susceptible to the curse of dimensionality as well as the unequal distribution of tests, and performance is the poorest of all the models. The Logistic Regression model had 90% accuracy, though high specificity (0.96), its boundary between the minority and majority class with the attack-class recall of 0.46 and the AUC of 0.71 suggest that it is hard to differentiate the minority class from the majority one. SVM achieved a high accuracy of 91 and specificity of 0.97, but low attack-class recall of 0.45 with AUC of 0.78 as a result of the inclination of

margin-maximizing classifiers to lose sensitivity on low-represented classes.

5.2 Proposed Ensemble Model Performance

The suggested ensemble performed 95 percent overall accuracy which is an improvement of 4-percentage-point improvement over the highest single baseline (SVM, 91 percent). More importantly, the ensemble increased the accuracy of the attack classes (0.69) (SVM) to (0.93), F1-score (0.55) to (0.77) and AUC (0.78) to (0.82). Specificity was 0.99, which implies that only 1-percent of normal traffic was incorrectly identified as malicious a clinically significant attribute that restricts alert fatigue. The sensitivity of 0.65 is a trade-off: it means that 1/3 of attack cases were not detected, and the future work will solve this issue by using deep learning augmentation. However, the fact that the overall profile of high accuracy, close to perfect specificity and a significantly better precision and F1 on the attack class, indicates that the ensemble is a materially more credible classifier than any of its constituents.

Table 3: *Per-Class Classification Results for All Models*

Model	Class	Precision	Recall	F1-Score	Support
KNN	0 (Normal)	0.93	0.90	0.92	2,848
	1 (Attack)	0.45	0.54	0.49	416
LR	0 (Normal)	0.92	0.97	0.94	2,848
	1 (Attack)	0.66	0.46	0.54	416
SVM	0 (Normal)	0.92	0.97	0.95	2,848
	1 (Attack)	0.69	0.46	0.55	416
Ensemble (Proposed)	0 (Normal)	0.95	0.99	0.97	2,848
	1 (Attack)	0.93	0.65	0.77	416

Table 4: *Aggregate Performance Comparison*

Model	Accuracy	AUC	Sensitivity (Attack)	Specificity
K-Nearest Neighbors (KNN)	0.86	0.72	0.54	0.90
Logistic Regression (LR)	0.90	0.71	0.46	0.96
Support Vector Machine (SVM)	0.91	0.78	0.45	0.97
Proposed Ensemble Model	0.95	0.82	0.65	0.99

5.3 Analysis and Discussion

The gains of the performance of the ensemble can be explained by two complementary mechanisms. First, hard-voting utilizes the diversity of models: in the case where KNN and RF both accurately classify a borderline attack example that SVM erroneously classifies, the majority vote yields the appropriate result. Second, the MI based feature selection pipeline filters out noisy features that singly degrade the performance of SVM and KNN to create a less contaminated decision boundary of each piece.

It is of special interest that the precision of the attack-classes (0.69 with SVM and 0.93 with ensemble) has been improved. The precision on the attack class in a clinical IDS implementation is high and this limits the false alarm, a factor that has been identified to cause

security fatigue among clinical personnel. The fact that 0.99 is very specific supports the idea that the model imposes very little disruptive impact on the normal functionality of the devices - which is not a trivial demand of life-critical IoMT infrastructure.

The attack-class recall of 0.65 indicates the difficulty inherent to finding minority-class instances in a hard-voting scheme: all three classifiers must agree which would reduce sensitivity. Soft voting in terms of probability would sacrifice specificity in favor of higher recall and be a logical extension of future research. However, at 0.65, the attack recall is significantly greater than standalone SVM (0.45) and LR (0.46), a true ensemble gain and not just an increase in accuracy. The proposed ensemble improves the results of WUSTL-EHMS-2020 by 2 percentage points (95% vs. 92.98) in accuracy and at the same time, eliminates the issue of imbalance in the classes by applying SMOTE and dimensional reduction by means of MI scoring. This implies that ensemble methods well-designed can compete with deep models on mid-scale IoMT data, at a fraction of the computational energy.

6. Conclusion

This paper presented a trust-based ensemble machine learning framework for intrusion detection in IoMT environments. The framework combines the feature selection based on Mutual Information and balancing of classes based on SMOTE with a hard-voting ensemble of KNN, Random Forest, and SVM classifiers. The proposed model scored 95 percent accuracy, 0.99 specificity, 0.65 sensitivity on attack-class, and an AUC of 0.82 on the WUSTL-EHMS-2020 benchmark, which is higher than all single baseline classifiers, as well as the highest recorded results on this dataset. The dual-modality feature set (network flow + patient biometrics) and the principled treatment of class imbalance are key design choices that differentiate this framework from prior work.

The results demonstrate the viability of ensemble ML approaches for real-time intrusion detection in resource-sensitive clinical IoT networks. By correctly classifying malicious traffic with high precision and near-perfect specificity, the framework provides a practical foundation for trustworthy, audit-compliant IoMT security systems that align with HIPAA and GDPR requirements.

7. Future Work

Several directions are identified for extending this work. First, the ensemble will be evaluated on additional IoMT benchmarks (e.g., ToNIoT-medical, CICIoT2023) to assess generalization across heterogeneous device profiles and attack taxonomies. Second, soft-voting and stacking meta-learning strategies will be investigated to improve attack-class recall without sacrificing specificity. Third, deep learning architectures particularly LSTM and Transformer-based models will be explored for sequential anomaly detection across network traffic streams. Fourth, k-fold cross-validation will be applied to obtain more statistically robust performance estimates. Finally, a lightweight model variant targeting embedded gateway hardware (Raspberry Pi, ESP32) will be prototyped to demonstrate real-time feasibility within clinical IoT infrastructure.

References

Abou-Nassar, E. M., Iliyasu, A. M., El-Kafrawy, P. M., Song, O. Y., Bashir, A. K., & El-Latif, A. A. A. (2020). DITrust Chain: Towards Blockchain-Based Trust Models for Sustainable Healthcare IoT Systems. *IEEE Access*, *8*(9106335), 111223–111238. https://doi.org/10.1109/ACCESS.2020.2999468

Al-Hamadi, H., & Chen, I. R. (2017). Trust-Based Decision Making for Health IoT Systems. *IEEE Internet of Things Journal*, *4*(5), 1408–1419. https://doi.org/10.1109/JIOT.2017.2736446

Astillo, P. V., Jeong, J., Chien, W.-C., Kim, B., Jang, J., & You, I. (2021). *SMDAps: A Specification-based Misbehavior Detection System for Implantable Devices in Artificial Pancreas System*. https://www.semanticscholar.org/paper/SMDAps%3A-A-Specification-based-Misbehavior-Detection-

Astillo-Jeong/bf4f05239e355b563facb875bd4f6750a7087e53

Gupta. (2024). Stochastic Analysis of Two Non-Identical Units System Model Subject to Inspection Policy. *Journal of Information Systems Engineering & Management*. https://doi.org/10.52783/jisem.v9i4.158

Hady, A. A., Ghubaish, A., Salman, T., Unal, D., & Jain, R. (2020). Intrusion Detection System for Healthcare Systems Using Medical and Network Data: A Comparison Study. *IEEE Access*, *8*, 106576–106584. https://doi.org/10.1109/ACCESS.2020.3000421

Nandy, S., Adhikari, M., Ayoub, M., Menon, V., & Verma, S. (2021). An Intrusion Detection Mechanism for Secured IoMT Framework Based on Swarm-Neural Network. *IEEE Journal of Biomedical and Health Informatics*, *PP*. https://doi.org/10.1109/JBHI.2021.3101686

Singh, P., Gaba, G. S., Kaur, A., Hedabou, M., & Gurtov, A. (2023). Dew-Cloud-Based Hierarchical Federated Learning for Intrusion Detection in IoMT. *IEEE Journal of Biomedical and Health Informatics*, *27*(2), 722–731. https://doi.org/10.1109/JBHI.2022.3186250