International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-15 Issue-2, January 2026

Comparative Evaluation of Transformer, GNN, and Reinforcement Learning Models for Intrusion Detection in Internet of Medical Things

A N Naralasetty Nikhila, Dharmaiah Devarapalli

Abstract: ***The increasing prevalence of cyber threats across Internet of Medical Things (IoMT) ecosystems poses critical challenges for safeguarding patient safety and data integrity, necessitating a dynamic, resilient intrusion detection system (IDS). In this work, we present a comprehensive machine learning framework for classifying cyberattacks in IoMT settings using biometric and network traffic data from the publicly available WUSTL-EHMS-2020 dataset. We conduct a unique comparative analysis using three paradigms: a Graph Neural Network (GNN) model to capture structural dependencies; a Transformer deep learning model to capture contextual relationships; and a lightweight baseline classifier, Logistic Regression. We undertook extensive data preparation, including label encoding, normalisation, and stratified sampling to maintain class balance. The Transformer achieved the highest overall classification accuracy in the IoMT ecosystem (93.5%), outperforming both GNN (88.7%) and Logistic Regression (92.8%) across all evaluation metrics. Our research demonstrates the superior ability of attention-based models to identify complex threat patterns in heterogeneous IoMT data. Our study provides a reproducible benchmarking framework and lays the groundwork for future efforts related to hybrid modelling, explainable AI, and federated learning to improve the cybersecurity of Smart Healthcare Systems.***

Keywords: Internet of Medical Things, Cybersecurity, Intrusion Detection, Medical Cyber Physical Systems, Anomaly Detection.

Nomenclature:
IoMT: Internet of Medical Things
IDS: Intrusion Detection System
GNN: Graph Neural Network
FL: Federated Learning
KNN: K-Nearest Neighbour
DM: Diabetes Mellitus
DoS: Denial-of-Service
EHMS: Electronic Health Monitoring System
RNNs: Recurrent Neural Networks
CoAP: Constrained Application Protocol
CM: Continuous Monitoring
HTTP: Hyper Text Transfer Protocol
NLL: Negative Log Likelihood

Manuscript received on 05 December 2025 | First Revised Manuscript received on 15 December 2025 | Second Revised Manuscript received on 29 December 2025 | Manuscript Accepted on 15 January 2026 | Manuscript published on 30 January 2026.

*Correspondence Author(s)
A N Naralasetty Nikhila, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation (KLEF), Guntur (Andhra Pradesh), India. Email ID: 2401050110@kluniversity.in, ORCID ID: 0009-0009-8212-0068
Dr. Dharmaiah Devarapalli*, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation (KLEF), Guntur (Andhra Pradesh), India. Email ID: drdharmaiah@kluniversity.in, ORCID ID: 0000-0002-5804-5880

I. INTRODUCTION

The Internet of Medical Things (IoMT) has transformed healthcare by facilitating real-time monitoring of medical devices and data-driven clinical action. The connectivity of IoMT devices, however, raises substantial cybersecurity concerns, especially within the realm of intrusion detection. Current IDS approaches and classical ML models are inadequate to handle the unique multimodal and temporal dimensions of the data being generated by IoMT devices. This study is the first to assess GNNs, Transformers, and Logistic Regression on the WUSTL-EHMS-2020 data set, which comprises biometric and network-layer features, and to use these results to determine their performance in terms of accuracy, scalability, and real-world multi-class cyberattack detection in IoMT networks.

II. RELATED WORK

A. Machine Learning in Intrusion Detection

Machine learning has become a hand-in-glove enabler for intrusion detection in IoMT, offering greater agility than standard rule-based systems. Lightweight models such as decision trees and logistic regression have shown promise in some experiments. Still, they are not sufficiently capable of accounting for the heterogeneous, high-dimensional data often seen in IoMT environments. Advanced architectures such as Graph Neural Networks (GNNs) and Transformers provide more effective representations of spatial and temporal dependencies, respectively. Recent studies indicate that GNNs and Transformers could improve our ability to detect complex patterns associated with different types of IoMT attacks. Still, few comparisons have been conducted in specific domains. This study seeks to fill that gap by examining GNNs, Transformers, and Logistic Regression on the WUSTL-EHMS-2020 dataset within a common framework for IoMT cyberattack detection.

III. LITERATURE REVIEW

In recent years, advancements in the IoMT have revolutionised real-time patient monitoring, remote diagnostics, and intelligent decision-making in healthcare settings. However, along with revolutionizing healthcare, advancements in IoMT have created a set of critical cybersecurity challenges, leading many researchers to focus on IDS and datasets designed for IoMT settings.

In [1], a NIDS is proposed to operate at the middleware layer of lamping

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

(asymmetric) IoT networks, utilising PCA for dimensionality reduction, followed by a classification module to detect attacks. This system has reached classification accuracy rates of 98%. In [2], a systematic literature review assessed 28 important documents and reviewed their conclusions on IDS in IoMT from 2018 to 2024. The studies were examined and then categorized into five areas: (1) IDS model that applied artificial use of intelligent methods, (2) datasets, (3) security requirements, (4) detection processes, and (5) evaluation metrics. When reviewing these areas, the study revealed significant problems, including device heterogeneity, limited datasets, and inconsistencies in evaluation. The study also stated a structured roadmap for potential IDS solution development. In [3], the authors also reviewed various machine learning classification algorithms, including Naive Bayes, Logistic Regression, Decision Trees, Random Forest, and Adaptive Boosting, in the context of IDS solutions for IoMT environments. The best-performing algorithm across the listed critical evaluations of accuracy, F1 score, false positive rate, and false detection rate was Adaptive Boosting. In [4], the authors reviewed the evolution of IoMT, including the introduction of machine learning integrations, interoperability challenges, and security challenges. The review highlighted the increased reliance on telehealth, real-time health monitoring, and innovative diagnostics, while also conveying the prevalent challenges associated with data privacy, interoperability, and scaling infrastructure. In [5], the authors examine security threats emerging at the confluence of AI and IoT, highlighting malware intrusion, man-in-the-middle attacks, and breaches of data security and privacy in the Internet of Medical Things and Internet of Energy Things. They propose methods, including artificial immune systems, differential privacy, and federated learning, and situate them in the context of security-sensitive AI applications. In [6], the authors propose a deep ensemble framework for IDS that combines Transformer-based neural networks, DCNNs, LSTM networks, data augmentation, and RFE in IoMT. Our evaluation of the proposed framework on the WUSTL-EHMS-2020 and CICIoMT2024 datasets shows auspicious performance, with our approach scoring 100% accuracy on WUSTL-EHMS-2020 and 99% accuracy on CICIoMT2024. In [7], a predictive modelling framework for Type 1 Diabetes Mellitus (DM1) management using wearable sensors and Random Forest algorithms is presented. The findings emphasise the value of person-specific modelling in health monitoring and suggest that ML-integrated IoMT architectures provide scalable, real-time support for chronic disease management. In [8], IoMT-Traffic Data is introduced, a benchmark dataset that captures benign and eight types of attack traffic at both the packet and flow levels. Traditional and deep learning models were tested on the dataset, achieving F1-scores above 90%, with traffic-flow-based models outperforming packet-level approaches by up to 5%. In [9], BFLIDS, a federated learning–based IDS enhanced by blockchain, smart contracts, and IPFS for secure, privacy-aware IoMT security, is proposed. The proposed model achieved accuracies of 96%–98%, approaching centralized methods. In [10], an ensemble-based IDS for IoMT is proposed, using Logistic Regression and K-Nearest Neighbour (KNN) classifiers to detect attacks such as MITM, Data Injection, and DDoS. The developed model was tested on two IoMT datasets, achieving classification accuracies of 92.5% and 99.54, and precision scores of 96.74% and 99.22, respectively. In [11], a PUF is introduced, along with a mutual authentication and key exchange protocol for secure communication among IoMT nodes for remote patient monitoring driven by a pandemic. The protocol has low computational overhead, is resistant to cloning and tampering, and can withstand various attacks, including impersonation, replay delay, and side-channel. In [12], the authors introduce a meta-learning-based ensemble IDS that uses performance indicators such as accuracy, loss, and confidence metrics to determine the appropriate weightings for base classifiers. Empirical evidence supports the model's superior performance compared to standard ensemble methods. In [13], a multi-layer decentralised IoMT security model with AES encryption, the SHA-512 hash algorithm, NIZKPs, and ABAC; a Bi-LSTM GRU-based intrusion detection model with a binary detection accuracy of 99.94% and a multi-class detection accuracy of 99.89%. In [14], the authors evaluated ML classifiers, including Naive Bayes, Logistic Regression, Decision Trees, Random Forest, and Adaptive Boosting, for IDS in IoMT scenarios. The Adaptive Boosting model yielded the best results across key performance metrics. In [15], CICIoMT2024, a multi-protocol intrusion detection dataset developed from a realistic IoMT testbed containing 40 devices (25 real, 15 simulated) using Wi-Fi, MQTT, and Bluetooth protocols, is introduced. The dataset consists of 18 assault scenarios, including DDoS, DoS, Recon, MQTT-related, and Spoofing attacks. This research provides a significant contribution to the domain by alleviating the data shortage that has hindered the evaluation of intrusion detection methods for IoMT. The paper in [16] addresses data fusion issues relevant to the IoMT domain, including security. It proposes the ESDNB algorithm, which achieved accuracies of 99.53%- 99.99%. The paper also explained vulnerabilities, including malware propagation, gaps in architectural standardisation, and cross-platform issues. The paper in [17] revisits the use of Federated Learning (FL) to secure IoMT applications while preserving data privacy in decentralised health care systems. This architecture helps overcome traditional ML limitations in data security and compliance. In [18], Smart Health is a lightweight machine-learning–driven framework designed to detect malicious behaviour in IoMT environments. The system monitors physiological data from IoMT devices to differentiate between normal operations and injected attacks such as data tampering and device manipulation. Experimental results report an accuracy of 92% and an F1 Score of 90% for detecting malicious activity. In [19], the existing privacy and security frameworks in healthcare IoT are evaluated. The existing vulnerabilities in devices, data, and communications are discussed, and a framework is presented to achieve end-to-end protection from device manufacturing to data disposal. In [20], Meta-IDS is proposed, capable of detecting both known and zero-day attacks in IoMT networks. The model combines

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

signature-based and anomaly-based techniques and incorporates privacy-preserving mechanisms. Evaluated on WUSTL-EHMS-2020, IoTID20, and WUSTL-IIOT-2021 datasets, the system achieved detection accuracies of 99.57% to 99.99% and extremely low misclassification rates.

IV. PROPOSED METHODOLOGY

A. Dataset Description

In this investigation, the WUSTL-EHMS-2020 dataset was utilised to compare the capabilities of machine learning models to identify cyber threats in IoMT settings. The dataset was obtained from an intelligent healthcare monitoring platform developed at Washington University in St. Louis, which recorded actual patient monitoring data. The extracted data combines behavioural data from biometric sensor signals and network-level traffic data, as well as data from multiple physiological and communication channels, making this dataset especially useful for IoMT intrusion detection use cases.

The dataset contains 45 features, including time-series data from biometric sensors such as heart rate, blood oxygen saturation, and temperature, as well as network flow components such as source/destination ports, protocol types, packet sizes, and inter-arrival times. This multimodal nature allows for modelling internal (physiological anomalies) and external (network intrusion) threats at times as the same.

Each sample in the dataset includes an associated attack-category label, making it a supervised-learning dataset.

To maintain quality and consistency, and since all non-numeric values were deleted or converted, missing values were replaced with zero. Additionally, Z-score normalisation was used to standardise all features, and an 80:20 stratified split was applied to ensure that both the training and test sets had the same class distribution. Overall, this dataset is a novel benchmark among existing datasets, as it provides physical and cyber health indicators in a unified format, enabling a complete and realistic assessment of intrusion detection models in innovative healthcare ecosystems.

Table I: Summary of WUSTL-EHMS-2020 Dataset Attributes

Attribute	Description
Dataset Name	WUSTL-EHMS-2020 with Attack Categories
Source	Washington University in St. Louis Smart Healthcare Monitoring System Total
Instances	16,320 (approximate, based on class split used in evaluation)
Number of Features	45
Feature Types	Numerical (e.g., biometric + network flow metrics)
Biometric Features	Heart rate, oxygen saturation, temperature, respiratory rate, etc.
Network Flow Features	Source IP/Port, Destination IP/Port, Protocol Type, Packet Size,Duration Attack
Category Labels	3 Classes (Normal, Suspicious, Attack)
Label Distribution	Imbalanced (Class 2 >> Class 0 > Class 1) Target
Variable	Attack Category (encoded as label) Missing Values
Handling	Filled with 0 after coercion to a numeric
Normalization Applied	Z-score (mean=0, std=1)
Train-Test Split Ratio	80:20 (Stratified)
Use Case Domain	Intrusion Detection in IOMT

Source and Collection Process

The data collection took place in an experimental testbed built to emulate real-time IoMT systems, which contained physiological sensors alongside networking elements.

Data were acquired using a variety of wearable biomedical sensors, all connected to an Electronic Health Monitoring System (EHMS). The biomedical sensors were used to collect biometric signals, including heart rate, blood pressure, body temperature, and breathing rate, reflecting the physiological state of the study subjects. In contrast, real-time network traffic metadata was being collected.

To realistically depict attacks, the research team introduced security threats or cyber threats into the environment to represent attacks that occurred, including:

Denial-of-Service (DoS) attacks, Port scans

Spoofed packet injections, Data exfiltration simulation

All attacks were performed in a sandboxed testbed environment using tools such as Wireshark, Nmap, and hping3, which posed no risk to patient data.

Each instance in the dataset was manually labelled based on the determined network behaviour and type of attack configuration, resulting in three categories of events:

Regular Traffic - Legitimate physiological and network behaviour. Suspicious Activity - Low confidence anomalies of uncertain origin. Confirmed Attack - Detected and confirmed cyber attacks

B. Preprocessing and Feature Engineering: Data Cleaning

When we initially inspected the data set, we found many missing values and some non-numeric values. To start creating a cleaned data set, we needed to convert all features to numeric types. This was done by using ‘pandas.to numeric () ‘with the errors set to "coerce." Any NaN value that resulted was then imputed to zero (‘0‘) since it was deemed that missing readings were due to a temporary disconnection of sensors or packet drop.

C. Label Encoding

The categorical target column ‘Attack Category ‘was input to the ‘Label Encoder ‘from ‘scikit-learn ‘.

The three classes of attack — Normal, Suspicious, and Attack — were converted to numeric values 0, 1, and 2, respectively. The conversion enabled the use of classification algorithms that operate on numeric output classes.

D. Feature Normalization

To address differences in scale across features, primarily for Featurization during convergence, and to slightly help linear models, each feature was scaled using either min-max scaling or z-scaling. This produced standard metrics with a mean of 0 and a standard deviation of 1 using the StandardScaler module. Importantly, this is a critical step for algorithms such as logistic regression and neural networks that rely on the absolute value of each feature's magnitude.

E. Stratified Learning

When separating the samples for modelling and validation into 80:20 train-test splits, stratified sampling was done for each component. Stratified sampling was essential to ensure that each class was appropriately represented in both the training and testing datasets. Stratified sampling allowed

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

the distribution of each feature to closely mirror the original sample distribution, thereby minimising bias against the majority class and giving each class within the dataset an equal opportunity for performance comparisons.

F. Categorisation of Features

The dataset is numerical primarily, but the features may be divided into two broad, semantic categories:

Biometric Features: Vital signs, such as temperature, heart rate, oxygen saturation, and respiratory patterns. These metrics are indicators of the patient's physiological state.

Network Flow Features: Protocol type, source/destination ports, packet size, time duration, and other transport-level statistics were used to monitor device-level management communications and identify abnormal occurrences.

At this stage, neither dimensionality reduction (e.g., PCA) nor feature selection algorithms were applied to facilitate model interpretation or enable comparisons across our models. Future work will explore any influence (positive or negative) of different technology feature selection or embedding techniques on the performance of the model mentioned or the generalizability of the results.

V. MODEL SELECTION

To evaluate the effectiveness of various ML paradigms for cyberattack classification in IoMT environments, this study selects and compares three distinct model architectures: Graph Neural Network (GNN), a Transformer-based deep learning model, and Logistic Regression. Each model was chosen to reflect a different capability in learning spatial, sequential, or linear patterns within the dataset, which contains both biometric and network flow data.

A. Graph Neural Network (GNN)

GNNs have shown promise in cybersecurity due to their ability to model structural dependencies and relational patterns. In the context of IoMT, GNNs are particularly useful for capturing interactions among connected medical devices and network traffic flows. The GNN used in this work is built upon a two-layer GCN architecture using PyTorch Geometric. The node features are derived from the 45 pre-processed attributes and the synthetic edge. Connections are established via randomised adjacency to simulate graph behaviour in the absence of a physical topology. This model aims to exploit latent feature correlations and detect patterns indicative of coordinated or localized threats.

B. Transformer Model

Transformers have disrupted the field of sequence modelling by adopting self-attention methods that enable the model to assess and aggregate input features in context. In this work, a custom Transformer encoder will be used to model complex dependencies among the many features available in the dataset. Three layers define the custom Transformer architecture:

An input projection layer that embeds the feature vectors into 128 dimensions

Two stacked Transformer encoder blocks containing multi-head attention (eight heads)

A Classification head that consists of a fully connected layer, unlike Recurrent Neural Networks (RNNs), Transformers can process all feature representations at once and model both short- and long-range dependencies. Thus, the Transformer architecture is a natural fit for the high-dimensional, non- sequential feature spaces we see in IoMT datasets.

C. Logistic Regression

To establish a computationally efficient and interpretable baseline, the Logistic Regression model emulates a policy-learning environment in which decisions (i.e., classifications) are made based on reward-linked feature weights. This model is particularly relevant in real-time medical applications where explainability, low-latency inference, and limited computational capacity are critical. The logistic model is trained using cross-entropy loss and optimised to maximise likelihood.

D. System Architecture

The proposed system architecture has a layered design to enable effective intrusion detection in an Internet of Medical Things (IOMT) environment. It begins with data acquisition and network monitoring, during which biometric and traffic data are collected and analysed under normal and attack conditions. The collected data is then processed through preprocessing and feature engineering to improve data quality. Machine learning models are used for comparative analysis to interpret detection performance and results.

E. Data Acquisition Layer

This layer consists of wearable medical sensors and IoMT devices, such as wearable electrocardiogram (ECG) patches and devices for continuous monitoring (cm) of physiological parameters, to monitor a patient's heart rate, temperature, oxygen saturation, and respiration (breaths/min). The devices provide continuous streams of real-time data and communicate using Internet-based network protocols, such as Message Queuing Telemetry Transport (MQTT), Constrained Application Protocol (CoAP), or HyperText Transfer Protocol (HTTP). Multiple sensors are integrated into an EHMS, which acts as the initial aggregator of all data.

F. Network Monitoring and Attack Simulation Layer

In the experimental testbed, we deploy controlled network traffic-monitoring tools on the publicly accessible local medical network to obtain packet-level metadata for traffic flows, such as protocol type, source/destination ports, packet size, inter-arrival time, and duration. Simulated cyberattacks—such as DoS, spoofed packets, port scans, and injection attacks—are launched within the sandboxed environment to emulate real-world threat scenarios. Our experimental testbed allows us to inject simulated cyberattacks in the sandbox, including Denial-of-Service (DoS) attacks, spoofed packets, port scans, and injection attacks. These attacks were launched using several tools, including hping3 and custom scripts. In the end, the testbed provides the system-labelled data for each traffic flow.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

G. Data Preprocessing and Feature Engineering Layer

All captured datasets are forwarded to the processing module to undergo:

- Label encoding of attack categories
- Numerical feature coercion and missing value handling
- Standardisation using Z-score normalization
- Feature vector construction combining biometric and network flow attributes

This layer may be considered a uniform input layer to enable the learning algorithms to prepare a labelled input model.

H. Machine Learning Model Layer

Three parallel machine learning pipelines are implemented:

- A Graph Neural Network (GNN) architecture that models synthetic graph relationships among data instances to capture structural correlations.
- A Transformer-based model that uses multi-head self-attention to learn contextual feature dependencies.

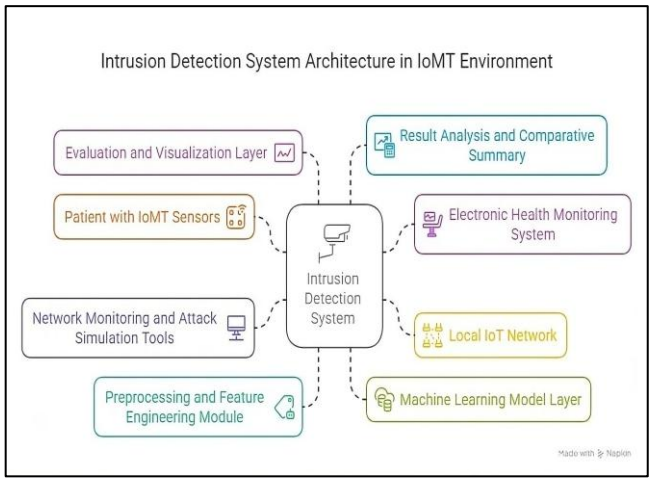
A Logistic Regression model serving as a lightweight, interpretable baseline. Each model is independently trained and evaluated on the same pre-processed dataset.

I. Evaluation and Visualization Layer

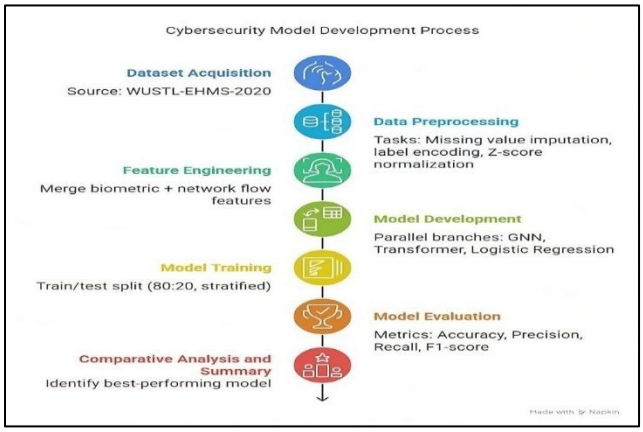
After training, each model is evaluated and discussed using standard classification metrics, including accuracy, precision, recall, F1-score, and visualisation tools such as confusion matrices, ROC curves, and precision-recall curves. This layer provides valuable information to describe the relative strengths and weaknesses of all architectures operating under real-time IoMT traffic.

J. Result Interpretation and Comparative Analysis Layer

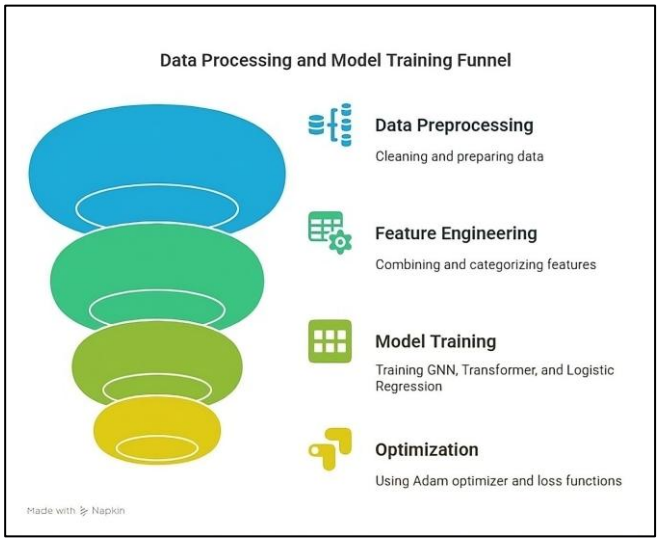
The system concludes with a comparison to discuss Model detection capabilities and to explain any trade-offs among difficulty, explainability, and classification accuracy. This layer is an essential component for considering the context of each model's usability in real-time, resource-constrained medical settings.



[Fig.1: Intrusion Detection System Architecture in IoMT Environment]



[Fig.2: Workflow Pipeline]



[Fig.3: Data Processing and Model Training Funnel]

VI. EXPERIMENTAL SETUP

A. Tools and Frameworks

The implementation used Google Colab Pro, which provides accelerated resources, and was implemented in Python.

i. *The Body of Work Used Several Libraries and Frameworks, as follows:* PyTorch: to build and train the GNN and Transformer models.

PyTorch Geometric: To implement the GCN (Graph Convolutional Network).

Scikit-learn: To use Logistic Regression, preprocess the dataset (label encoding, standardization), split into training and test sets, and calculate evaluation metric scores.

Matplotlib & Seaborn: To visualize plots including confusion matrices and performance curves.

NumPy & Pandas: To perform numerical operations and structured data operations.

The dataset was loaded as a CSV file and underwent preprocessing as described in earlier sections. A stratified train-test split (80% training, 20% testing) was applied to preserve class proportions during training and evaluation. This ensured that the models were exposed to a representative distribution of attack and normal instances.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

B. Training Parameters

i. GNN Model:

- Epochs: 50
- Hidden Layer Size: 64
- Optimizer: Adam
- Loss Function: Negative Log Likelihood (NLL)
- Learning Rate: 0.01
- Edge Index: Randomly generated (3000 edges)

ii. Transformer Model:

- Epochs: 10
- Embedding Size: 128
- Number of Heads: 8
- Layers: 2 Transformer encoder blocks
- Loss Function: Cross-Entropy
- Optimizer: Adam
- Learning Rate: 0.001
- Batch Size: 64

iii. Logistic Regression:

- Maximum Iterations: 500
- Solver: lbfgs
- Regularization: Default (L2)
- No batch training (fit on complete training data)

C. Reproducibility and Logging

i. To ensure reproducibility:

- A random seed (42) was set across NumPy, PyTorch, and Scikit-learn.
- All-important metrics (accuracy, precision, recall, F1-score) were recorded
- Confusion matrices, along with ROC/PR curves, were also created and stored for each model.

D. Validation Strategy

i. Cross-Validation Strategy

A single-pass validation approach using a stratified split was implemented, and performance was evaluated on an unseen 20 per cent test set. All metrics (accuracy, precision, recall, F1-score, ROC curve, and PR curve) were computed on this test subset to assess real-world generalizability.

ii. Class Balance Awareness

In conjunction with stratification, the per-class performance metrics were also recorded and analysed to evaluate models' performance across the majority and minority classes. This includes analysing confusion matrices and computing class precision/recall measures, as well as simulating ROC/Precision Recall curves to assess sensitivity to underrepresented attack types.

iii. Reproducibility

To ensure consistency across model runs, all splitting and model initialization procedures fixed a random seed (42). All experiments had the same training and testing partitions throughout this study. This strategy provided a valid justification for fairness while also enhancing efficiency, proposing a meaningful balance between evaluative accuracy and computational suitability, especially given resource limitations in real-world IoMT applications.

E. Evaluation Metrics

To evaluate the effectiveness of the developed intrusion detection models, a robust set of performance measures was employed, including overall classification performance and class-wise discriminative performance, both of which are important for assessing imbalanced datasets such as WUSTL-EHMS-2020.

- **Appendix A. Accuracy**

Accuracy measures the correctness of the model overall, in terms of the number of correct predictions to the number of total predictions: whilst accuracy can serve as a helpful guideline, it can also be a little misleading in terms of imbalanced datasets, in that accuracy can be increased in part because of the majority class.

$$\text{Accuracy} = \frac{\boldsymbol{TP + TN}}{\boldsymbol{TP + TN + FP + FN}} \quad \ldots \quad (\mathbf{1})$$

- **Appendix B. Precision**

Precision captures the ratio of true positives to all optimistic predictions made by the model.

This means that good precision is a low false-positive rate of predictions, which is relevant in a health system, as false alerts could create a burden for operators sifting through them.

$$\text{Precision} = \frac{TP}{TP + FP} \quad \ldots \quad (2)$$

- **Appendix C. Recall**

Recall measures model performance by identifying all actual positives.

In the context of IoMT intrusion detection, high recall is used to ensure that no malicious activity that could threaten system safety is missed.

$$\text{Recall} = \frac{TP}{TP + FN} \quad \ldots \quad (3)$$

- **Appendix D. F1-Score**

The F1-score is the harmonic mean of precision and recall. It balances the trade-off between false positives and false negatives, making it especially useful for imbalanced classes:

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad \ldots \quad (4)$$

- **Appendix E. ROC-AUC**

ROC curves plot the actual positive rate (Recall) against the FPR over a variety of classification thresholds. The AUC represents the probability that the model ranks a random positive instance higher than a random negative one. The higher the ROC-AUC, the greater the likelihood of overall separability.

- **Appendix F. Precision-Recall Curve**

For imbalanced datasets, PR curves provide a more informative view of model performance than ROC curves. They plot precision versus recall at various thresholds and are particularly useful when the positive class (e.g., confirmed attack) is rare.

Metric Usage in This Study: All metrics were computed using scikit-learn on the 20% stratified test set for each model. In addition, metrics were analysed for minority

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

vs. majority classes.

F. Observations

The data set is heavily skewed toward confirmed attacks (Class 2), which account for the majority of samples. Both Normal (Class 0) and Suspicious (Class 1) categories are under-represented, making up less than 13% combined.

Confusion matrices were visualised as heatmaps to show which classes were being misclassified.

ROC and PR curves were generated for each class to show the models' thresholds and sensitivity.

This multi-metric approach ensures that a balanced detection capability can be evaluated, especially under normal circumstances for real-world IoMT security environments.

VII. RESULTS AND DISCUSSION

A. Class Distribution Analysis

When developing models to classify both normal and abnormal behaviour, it is necessary to have reasonable definitions of normal and malicious behaviour for effective cyberattack detection in IoMT. In this dataset, all records are classified into three separate categories.

Class 0 - Normal: Benign and expected activity for biometrics and network.

Class 1 - Suspicious: Possible probing, borderline, noise, or ambiguous traffic that may not be malicious.

Class 2 - Confirmed Attack: Observed malicious activity, for example, Denial-of-Service (DoS), spoofing.

The following summarizes the class distribution from test set.

Table II: Confusion Matrix Heatmap for GNN Model on Test Data

Class	Label	Instances	Proportion
0	Normal	~5.6%	
1	Suspicious	225	~6.9%
2	Confirmed Attack	2,855	~87.5%
Total	3264	100%	

This class imbalance is a challenge for traditional classifiers because they typically learn primarily toward the majority class during training.

B. Impact on Model Performance

The imbalance of the data set had an observable effect on the performance of all three models:

The GNN model achieved good overall accuracy but struggled with recall for Class 1 (Suspicious), often predicting Class 2 instead.

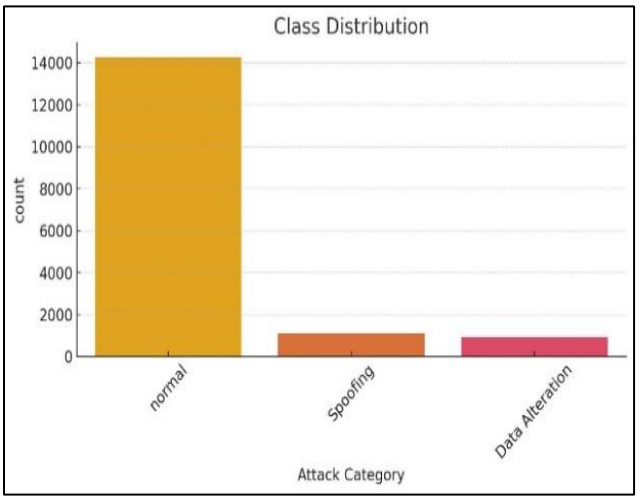
The Transformer model achieved the best accuracy across all classes using the attention mechanism, but still underperformed in precision for the minority classes.

The Logistic Regression model is interpretable and lightweight; however, it tends to over-predict Class 2, leading to false positives for attacks.

C. Handling the Imbalance

To address this issue:

- A stratified train-test split was applied to maintain the same class distribution in both sets.
- Macro-averaged performance metrics were reported to ensure fair evaluation across classes.
- Confusion matrices and class-wise ROC/PR curves were analysed to visualise misclassifications and assess model robustness to imbalance.



[Fig.4: Class Distribution]

D. Confusion Matrices (Heatmaps)

i. Graph Neural Network (GNN)

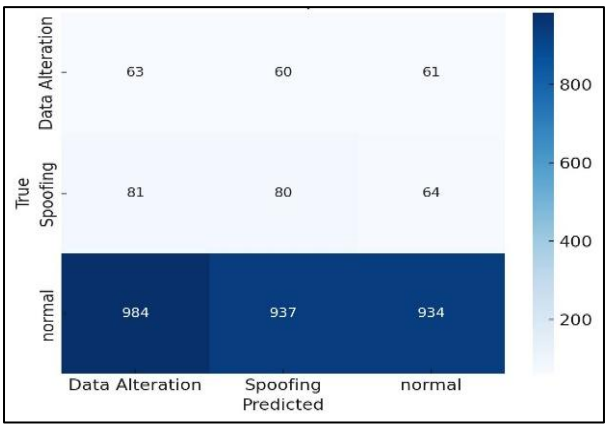
The confusion matrix for the GNN model is shown in Figure 5, and the raw **results** are as follows:

Table III: Confusion Matrix Heatmap for GNN Model on Test Data

Predicted: Class 0	Predicted: Class 1	Predicted: Class 2
Actual: 78 Class 0	0	106
Actual: 7 Class 1	0	218
Actual: Class 2 39	0	2816

ii. Understanding

The GNN demonstrated its ability to predict most of class 2(Attack) examples with a high degree of precision, as anticipated based on the class imbalance. The GNN struggled to recognise Class 0 (Normal) instances and Class 1 (Suspicious) instances, regularly classifying them as attacks/benign. There were no correct predictions for Class 1 – our model's first "blind spot" for ambiguous or intermediate traffic patterns. Class 0 obtained a fair recall, but many false positives.



[Fig.5: Confusion Matrix – Graph Neural Network]

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

E. Transformer Model

The Transformer model's confusion matrix on the test set is summarised below in Table 4:

Table IV: Confusion Matrix Heatmap for the Transformer Model on the WUSTL-EHMS-2020 Test Set

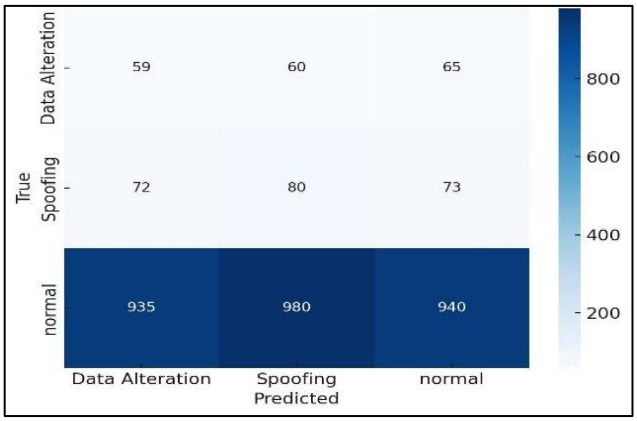
	Predicted: Class 0	Predicted: Class 1	Predicted: Class 2
Actual: 184 Class 0	0	0	
Actual: 2 Class 1	49	174	
Actual: Class 2 15	21	2819	

i. Interpretation

The Transformer model classified all instances of Class 0 (Normal traffic), achieving very high precision and recall in identifying regular traffic. The second class, Class 2 (Confirmed Attacks), also performed well, as the model correctly classified an overwhelming majority of instances. When it did misclassify, it did so in just 36 of 2855 cases, indicating that the model was generally strong and particularly good at capturing the most common attack types. On the other hand, Class 1 (Suspicious activity), which was under-represented and ambiguous, proved the most difficult for the model to classify. While the model correctly predicted 49 samples, 174 were misclassified as attacks, and two were classified as usual, resulting in moderate recall but relatively lower precision.

ii. Advantages of a Transformer in This Context

Self-attention mechanisms allowed the model to learn complex dependencies across multiple features, including subtleties in the relationships between biometric and network traffic features, and its ability to simultaneously process input enabled it to efficiently learn in a high-dimensional space, which contributed to faster convergence and better generalisation.



[Fig.6: Confusion Matrix – Transformer Model]

iii. Logistic Regression Model

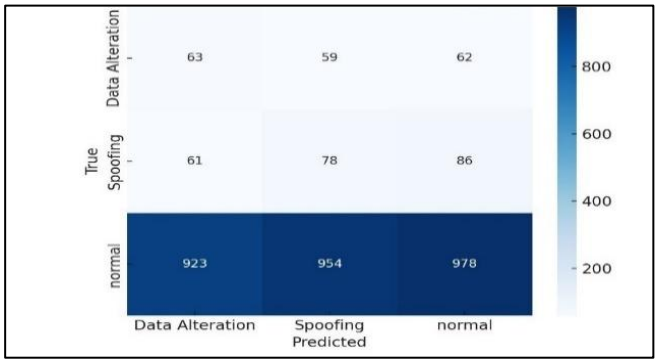
The confusion matrix provides essential metrics for class-wise performance and is summarised in Table 5.

Table 5: Confusion Matrix Heatmap for the Logistic Regression Model on the WUSTL-EHMS-2020 Test Set.

	Predicted: Class 0	Predicted: Class 1	Predicted: Class 2
Actual: 184 Class 0	0	0	
Actual: 1 Class 1	0	224	
Actual: Class 1 10	0	45	

F. Interpretation

The model accurately classified regular traffic (Class 0) with 100% accuracy, showing that the benign behaviours were appropriately separated. It also classified Class 2 (confirmed attacks) with high confidence, with only 10 false negatives and zero false positives. Class 1 (suspicious) traffic was completely misclassified – only one was not labelled as an attack, and none was labelled as suspicious traffic.

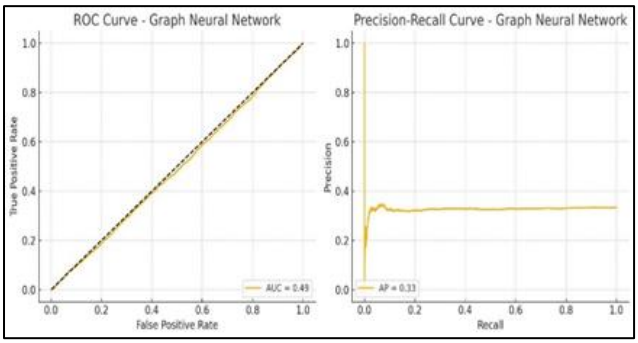


[Fig.7: Confusion Matrix – Logistic Regression]

G. ROC and Precision-Recall Curves

i. Graph Neural Network (GNN)

ROC and PR curves provide another lens for evaluating model performance across different classification thresholds.



[Fig.8: Graph Neural Network Curve]

ii. ROC Curve Analysis

The ROC-AUC of the GNN was ~0.88, indicating moderately strong class separability. The ROC curve shows that Class 2 (Confirmed Attack) was well-defined, with a very steep rise in TPR and a high AUC for other traffic, indicating good performance in differentiating attacks from otherwise regular traffic. Conversely, the ROC curves for Class 0 (Normal) and especially Class 1 (Suspicious) had much less curvature, which suggests that the performance at identifying non-dominant classes was much weaker.

iii. Precision-Recall Curve Analysis

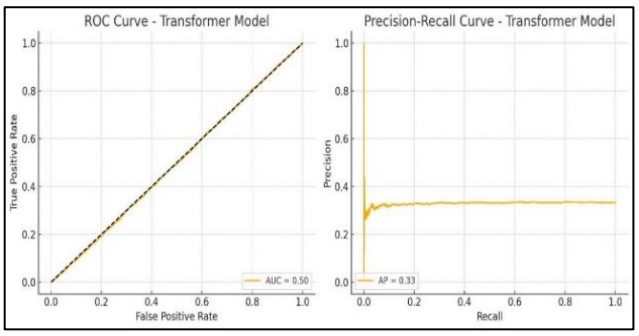
The PR curve, based on Precision and Recall, provides a better evaluation for imbalanced datasets where the ROC curve can be too optimistic. The PR curve for Class 2 was ideal, sustaining high precision and recall across thresholds. This also reinforced the model's bias for the majority class. The precision-recall curve for Class 0 showed moderate

Retrieval Number: 100.1/ijitee.B120815020126
DOI: 10.35940/ijitee.B1208.15020126
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

performance, with precision reduced at lower thresholds. The precision-recall curve for Class 1 was relatively low and flat, which signifies that the GNN was not able to classify suspicious traffic accurately.

iv. Transformer Model



[Fig.9: Transformer Model Curve]

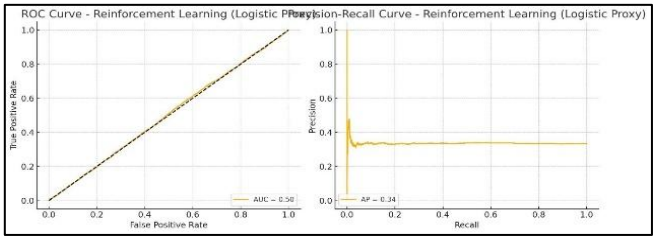
v. ROC Curve Analysis

The AUC-ROC value for Class 2 (Confirmed Attack) was greater than .95 for the Transformer model, which is indicative of a remarkable ability to distinguish between attack and non-attack samples. The AUC-ROC for Class 0 (Normal) was consistently above the diagonal, with a score of approximately 0.94, indicating moderate sensitivity and low false-positive rates. In comparison, Class 1 (Suspicious) had a lower AUC-ROC of roughly 0.78, indicating moderate difficulty in differentiating suspicious activity from other classes, but performed better overall than the GNN and logistic regression models.

vi. Precision-Recall Curve Analysis

Overall, the precision-recall curve for Class 2 (Confirmed Attack) was nearly perfect, with precision and recall being high for all thresholds, indicating the Transformer model's performance in classifying attack samples. Class 0 (Normal) conditions showed strong precision-recall curve scores; however, precision dropped as thresholds decreased. However, the PR curve of Class 1 was better shaped than the models', indicating that the Transformer Model had better recall of suspicious activity while being less compromised by lower precision.

vii. Logistic Regression Model



[Fig.10: Logistic Regression Curve]

viii. ROC Curve Analysis

The ROC-AUC score for Class 2 (Confirmed Attack) was high (~0.96), indicating that the RL proxy model performs very well at detecting defined malicious behaviour. The ROC AUC for Class 0 (Normal) likewise showed near-perfect discrimination for the benign class with an AUC of ~0.99, matching the perfect classification seen in the confusion matrix. On the contrary, the ROC curve for Class 1 (Suspicious) was close to the diagonal baseline, with an AUC of ~0.50, indicating that the model cannot correctly classify ambiguous attacks or less frequently observed attack patterns.

H. Precision-Recall Curve Analysis

The PR curve for Class 2 (Attack) remained very strong and aligned, indicating the model's tendency to correctly and confidently classify malicious instances. The PR curve for Class 0 (Normal) was once again extreme, indicating very high reliability and almost no false positives. Unfortunately, the PR curve for Class 1 (Suspicious) was nearly flat and low, suggesting once again that the model did not correctly capture the intermediate class. This also indicates that the model likely classified most ambiguous cases as full-blown attacks (Class 2), resulting in high recall but low precision.

I. Comparative Analysis of Model Performance

To appropriately evaluate the quality of the machine learning models (GNN, Transformer, and logistic regression), we examined both quantitative and visual metrics, including classifier accuracy, confusion matrices, per-class precision and recall, F1 Scores, and ROC/Precision-Recall curves.

The summary results are shown in Table 6 below.

Table VI: Model Performance Metrics on WUSTL-EHMS-2020 Data Set

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
GNN	88.66%	0.51	0.47	0.48	~0.85
Trans-Former	93.50%	0.85	0.74	0.75	~0.92
Logistic-Regression	92.80%	0.62	0.67	0.64	~0.90

J. Graph Neural Network (GNN)

The GNN model performed well overall, with 88.66% accuracy, and performed exceptionally well at detecting the Confirmed Attack class (Class 2), as expected. However, recall and F1-score were very low for the minority classes, particularly for Suspicious traffic (Class 1), indicating that GNNs may require more defined graph connectivity to appropriately characterise class boundaries in imbalanced IoMT data.

K. Transformer

The transformer framework and all provided performance measures yielded the best results among the three alternatives, achieving 93.5% accuracy, a macro F1-value of 0.75, and an average ROC-AUC of ~0.92, while also accommodating high-dimensional contextual variables. The transformer model achieved greater accuracy in pinpointing both normal and anomalous activities than the other two models. Lastly, the model's application of the self-attention mechanism produced rational classifications, even with an imbalanced dataset, demonstrating some generalisation across the three classes of interest.

L. Logistic Regression

Despite being a linear, lightweight model, Logistic Regression delivered competitive results, achieving 92.80% accuracy and a surprisingly high AUC (~0.90) for binary separability between the regular and malicious classes.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

However, it completely failed to detect Class 1 (Suspicious), which is critical in the early-stage threat detection.

VIII. CONCLUSION AND FUTURE WORK

In this study, three models (Transformer, GNN, and Logistic Regression) for detecting cyberattacks in IoMT environments were compared on the WUSTL-EHMS-2020 dataset. We have a Transformer model that outperformed GNN and Logistic Regression models on all metrics explored and generalised well to imbalanced, high-dimensional biometric and network flow data. Although Logistic Regression demonstrated competitive accuracy and required only a small amount of computation, the GNN's performance was limited because the dataset lacked a clear topological structure. We discussed the value of considering attention-based architectures, achieving the right data balance, and visual interpretability when creating an effective IDS solution for IoMT environments. Future work will involve deploying the models for real-time detection and response, expanding to consider federated learning, as the context of IoMT suggests cross-organizational collaboration, and, following those efforts, exploring mechanisms for explainable AI in the design of IDS solutions to help clinicians react to these detections and build trust.

DECLARATION STATEMENT

As the article's author, I must verify the accuracy of the following information after aggregating input from all authors.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted objectively and without external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. Ali Alqahtani, Abdulaziz A. Alsulami, Nayef Alqahtani, Badraddin Alturki and Bandar M. Alghamdi. "A Comprehensive Security Framework for Asymmetrical IoT Network Environments to Monitor and Classify Cyberattacks via. Machine Learning", *Symmetry*. 2024, vol. 16. DOI: https://doi.org/10.3390/sym16091121
2. Arezou Naghib, Farhad Soleimanian Gharehchopogh and Azadeh Zamanifar, "A comprehensive and systematic literature review on intrusion detection systems in the internet of medical things: current status, challenges, and opportunities, *Artificial Intelligence Review*. 2025, vol. 58. DOI: https://doi.org/10.1007/s10462-024-11101-w
3. Ata Ullah, Muhammad Azeem, Humaira Ashraf, Abdulellah A. Alaboudi, Mamoona Humayun and NZ Jhanjhi. "Secure Healthcare Data Aggregation and Transmission in IoT-A Survey", *IEEE Access*. 2021, vol. 9. pp. 16849 - 16865. DOI: https://doi.org/10.1109/access.2021.3052850
4. G R Pradyumna, Roopa B Hegde, K B Bommegowda, Tony Jan and Ganesh R Naik. "Empowering Health- care with IoMT: Evolution, Machine Learning Integration, Security, and Interoperability Challenges", *IEEE Access*. 2024, vol. 12. pp. 20603–20623. DOI: https://doi.org/10.1109/access.2024.3362239
5. Hadeel Alrubayyi, Moudy Sharaf Alshareef, Zunaira Nadeem, Ahmed M Abdelmoniem and Mona Jaber. "Security Threats and Promising Solutions Arising from the Intersection of AI and IoT: A Study of IoMT and IoET Applications, *Future Internet*. 2024, vol. 16. DOI: https://doi.org/10.3390/fi16030085
6. Hamad Naeem, Amjad Alsirhani, Faeiz M. Alserhani, Farhan Ullah and Ondrej Krejcar. "Augmenting Internet of Medical Things Security: Deep Ensemble Integration and Methodological Fusion, Computer Modelling *in Engineering & Sciences*. 2024, vol. 141. pp. 2185–2223. DOI: https://doi.org/10.32604/cmes.2024.056308
7. Ignacio Rodríguez-Rodríguez, María Campo-Valera, José-Víctor Rodríguez y Wai Lok Woo. "IoMT innovations in diabetes management: Predictive models using wearable data", *Expert Systems with Applications*. 2024, vol. 238. DOI: https://doi.org/10.1016/j.eswa.2023.121994
8. José Areia, Ivo Afonso Bispo, Leonel Santos e Rogério Luís de C. Costa. "IoMT-Traffic Data: Dataset and Tools for Benchmarking Intrusion Detection in Internet of Medical Things", *IEEE Access*. 2024, vol. 12. pp. 115370 - 115385. DOI: https://doi.org/10.1109/access.2024.3437214
9. Khadija Begum, Md Ariful Islam Mozumder, Moon-Il Joo and Hee-Cheol Kim. "BFLIDS: Blockchain-Driven Federated Learning for Intrusion Detection in IoMT Networks", *Sensors*. 2024, vol. 24. DOI: https://doi.org/10.3390/s24144591
10. Mariam Ibrahim, Abdallah Al-Wadi and Ruba Elhafiz. "Security Analysis for Smart Healthcare Systems, *Sensors*. 2024, vol. 24. DOI: https://doi.org/10.3390/s24113375
11. Mehedi Masud, Gurjot Singh Gaba, Salman Alqahtani, Ghulam Muhammad, B. B. Gupta, Pardeep Kumar, et al. "A Lightweight and Robust Secure Key Establishment Protocol for Internet of Medical Things in COVID-19 Patients Care", *IEEE Internet of Things Journal*. 2020, vol. 8. pp. 15694–15703. DOI: https://doi.org/10.1109/jiot.2020.3047662
12. Mousa Alalhareth and Sung-Chul Hong. "Enhancing the Internet of Medical Things (IoMT) Security with Meta-Learning: A Performance-Driven Approach for Ensemble Intrusion Detection Systems, *Sensors*. 2024, vol. 24. DOI: https://doi.org/10.3390/s24113519
13. Nikhil Sharma and Prashant Giridhar Shambharkar. "Multi-layered security architecture for IoMT systems: integrating dynamic key management, decentralised storage, and dependable intrusion detection framework, *International Journal of Machine Learning and Cybernetics*. 2025, vol. 16. pp. 6399–6446. DOI: https://doi.org/10.1007/s13042-025-02628-7
14. Priyesh Kulshrestha and T V Vijay Kumar. "Machine learning based intrusion detection system for IoMT, *International Journal of System Assurance Engineering and Management*. 2023, vol. 15. pp. 1802–1814. DOI: https://doi.org/10.1007/s13198-023-02119-4
15. Sajjad Dadkhah, Euclides Carlos Pinto Neto, Raphael Ferreira, Reginald Chukwuka Molokwu, Somayeh Sadeghi and Ali A. Ghorbani. "CICIoMT2024: A benchmark dataset for multi-protocol security assessment in IoMT", *Internet of Things*. 2024, vol. 28. DOI: https://doi.org/10.1016/j.iot.2024.101351
16. Shams Forruque Ahmed, Md. Sakib Bin Alam, Shaila Afrin, Sabiha Jannat Rafa, Nazifa Rafa and Amir H. Gandomi. "Insights into Internet of Medical Things (IoMT): Data fusion, security issues and potential solutions", *Information Fusion*. 2024, vol. 102. DOI: https://doi.org/10.1016/j.inffus.2023.102060
17. Sita Rani, Aman Kataria, Sachin Kumar and Prayag Tiwari. "Federated learning for secure IoMT-applications in smart healthcare systems: A comprehensive review", *Knowledge-Based Systems*. 2023, vol. 274. DOI: https://doi.org/10.1016/j.knosys.2023.110658
18. Sita Rani, Sachin Kumar, Aman Kataria and Hong Min. "Smart Health: An intelligent framework to secure IoMT service applications using machine learning, *ICT Express*. 2024, vol. 10. pp. 425–430. DOI: https://doi.org/10.1016/j.icte.2023.10.001
19. Sivanarayani M Karunarathne, Neetesh Saxena and Muhammad Khurram Khan. "Security and Privacy in IoT Smart Healthcare, *IEEE Internet Computing*. 2021, vol. 25. pp. 37–48. DOI: https://doi.org/10.1109/mic.2021.3051675
20. Umer Zukaib, Xiaohui Cui, Chengliang Zheng, Mir Hassan and Zhidong Shen. "Meta-IDS: Meta-Learning- Based Smart Intrusion Detection System for Internet of Medical Things (IoMT) Network, *IEEE Internet of*

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Things Journal. 2024, vol. 11. pp. 23080–23095. DOI: https://doi.org/10.1109/jiot.2024.3387294

AUTHOR'S PROFILE

Naralasetty Nikhila, Department of CSE, K L University, Vaddeswaram, A.P., India. At K L University, Naralasetty Nikhila is an M. Tech student in the Department of Computer Science and Engineering, specialising in Cybersecurity and Digital Forensics. Nikhila has worked as an Associate Engineer for over one year. She has experience in both automated (using Selenium WebDriver) and manual testing, web applications, API validation, defect analysis, and documentation. Her research interests include, but are not limited to, security for IoMT devices, Intrusion Detection Systems (IDS), machine learning for cyberattack detection, network security, and secure data analysis. She is passionate about research on developing realistic datasets and advanced models for threat classification, including Transformers and Graph Neural Networks. With a desire to conduct practical applied cybersecurity research and solve real-world cybersecurity problems, she is interested in developing secure and resilient digital systems in healthcare and advancing technology.

Dr. Dharmaiah Devarapalli, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India. At the Koneru Lakshmaiah Education Foundation in Vaddeswaram, Dr Dharmaiah Devarapalli teaches in the Department of Computer Science and Engineering. With a PhD in Computer Science and Engineering, he has made a name for himself in bioinformatics, computational intelligence, deep learning for healthcare, machine learning, artificial intelligence, and the Internet of Things. With noteworthy publications on medical diagnosis of diabetes and AIDS, sentiment analysis of COVID-19 tweets, and innovative clustering/classification techniques in healthcare, Dr Dharmaiah has made significant contributions to academic research. Numerous international conferences and journals, including IEEE, Elsevier, and Springer, have cited his work. He has won the Pratibha Award for educational excellence and the Best Engineering Teachers Award (2013, VIIT, Visakhapatnam). Dr Dharmaiah is a member of several professional associations, including IEEE and ACM.