A Multi-Class Intrusion Detection System for the Internet of Medical Things Based on Hybrid Deep Learning

Aisha Essa Mohammad and Amer Abdulmajeed Abdulrahman
Department of Computer Science, College of Science, University of Baghdad, 10071 Baghdad, Iraq
aaesha.eesa2201m@sc.uobaghdad.edu.iq, amer.abdulrahman@sc.uobaghdad.edu.iq

Keywords: Internet of Medical Things (IoMT), Intrusion Detection System (IDS), Deep Learning (DL), Deep Neural Network (DNN), Long Short-Term Memory (LSTM).

Abstract: The IoMT improves healthcare through smart medical devices, enabling real-time monitoring and data transmission. However, increased connectivity exposes IoMT systems to cyber threats, jeopardizing patient data confidentiality, system integrity, and availability. Traditional IDS struggle to detect sophisticated attacks, thus requiring advanced solutions. This study presents a hybrid deep learning model that integrates LSTM and DNN to improve intrusion detection in IoMT networks. The CICIoMT2024 dataset, comprising network traffic of 40 IoMT devices under 18 types of cyberattacks, was used for training and evaluation. Data preprocessing included label encoding, normalization. The LSTM component captures sequential traffic patterns, while the DNN extracts advanced features for classification. Batch normalization, dropout layers, and early stopping were implemented to improve model performance. Experimental results show that the proposed model outperforms the conventional intrusion detection system, achieving 99.6% accuracy in binary classification, 99.4% in 6-class classification, and 98.4% in 19-class classification. Compared with stand-alone models, the hybrid approach demonstrates superior accuracy and robustness. This research underscores the effectiveness of LSTM-DNN in securing IoMT networks. Future work will focus on real-time deployment, optimization of computational efficiency, and expansion of the dataset to improve cyber threat detection in medical settings.

1 INTRODUCTION

The Internet of Medical Things (IoMT) represents a transformative paradigm in healthcare, integrating interconnected medical devices, sensors, and software to enable real-time patient monitoring, data transmission, and intelligent diagnostics [16], [13]. This ecosystem significantly enhances remote patient care, chronic disease management, and clinical efficiency, ultimately improving healthcare outcomes while reducing operational costs.

However, this pervasive connectivity and the critical nature of medical data make IoMT networks a prime target for cyber adversaries [7], [8]. Security breaches in this context transcend conventional data theft, posing direct risks to patient safety through the potential compromise of data confidentiality, manipulation of vital medical information, or disruption of life-sustaining healthcare services. Ensuring the confidentiality, integrity, and availability (CIA triad) of IoMT systems is therefore not merely a technical requirement but a fundamental prerequisite for patient safety.

Traditional signature-based Intrusion Detection Systems (IDS) often fail to detect novel, sophisticated, or evolving cyberattacks tailored to IoMT protocols and device constraints. This limitation necessitates the adoption of advanced, adaptive security solutions. Deep Learning (DL) has emerged as a powerful approach, capable of autonomously learning complex patterns and anomalies from high-dimensional network traffic data. Among DL architectures, Long Short-Term Memory (LSTM) networks excel at modeling temporal dependencies in sequential data, such as network flow, while Deep Neural Networks (DNNs) are effective at extracting hierarchical features for robust classification. A hybrid model that synergistically combines these strengths presents a promising avenue for building a more resilient IDS.

Despite progress, existing research often faces challenges in handling the multi-class, imbalanced nature of IoMT attack datasets, achieving real-time efficiency on resource-constrained devices, and

generalizing across diverse network environments. To address these gaps, this study makes the following key contributions:

- We propose a novel hybrid LSTM-DNN framework specifically designed for IoMT security. This architecture leverages LSTM layers to capture sequential traffic patterns and DNN layers to perform high-level feature abstraction, thereby improving detection accuracy for both known and subtle attack vectors.
- We introduce a task-oriented preprocessing pipeline and an adaptive label mapping strategy optimized for multi-class intrusion detection, enhancing the model's flexibility and learning efficiency across different classification tasks (binary, 6-class, 19-class).
- We implement an enhanced regularization approach combining batch normalization, dropout, and a performance-based early stopping mechanism. This strategy is tailored to the variability of IoMT data to mitigate overfitting and ensure model generalizability.
- We conduct a comprehensive evaluation using the recent and relevant CICIoMT2024 dataset [5], which includes traffic from 40 devices under 18 attack types. Our model demonstrates superior performance, outperforming standalone LSTM and DNN baselines as well as other contemporary approaches reported in the literature [2], [17].

The remainder of this paper is structured as follows: Section 2 reviews related work on DL-based IDS for IoMT. Section 3 details the proposed methodology, including the dataset, preprocessing, and model architecture. Section 4 presents and discusses the experimental results and comparative analysis. Section 5 outlines the study's limitations, and Section 6 concludes the paper with directions for future work.

2 RELATED WORK

The Internet of Medical Things (IoMT) is a vastly networked network of medical equipment that both opens up new possibilities for patient care and reveals serious weaknesses. The security threats in these IoMT ecosystems have been the subject of several research [3].

Recently, several research papers have been published on analyzing deep learning-based intrusion detection systems. This research field is becoming increasingly important, as its learning and adaptability capabilities make it highly effective in dealing with an increasing number of unforeseen attacks. Since IDS are developed using deep learning techniques, they work more effectively and precisely [9].

Dadkhah et al. (2024) [5] proposed the CICIoMT2024 dataset and evaluated it using used machine learning (ML) techniques, including Logistic Regression, AdaBoost, Random Forest and Deep Neural Networks (DNN) to address the lack of operational diversity, multiple protocols, and the lack of device information in existing datasets regarding IoMT. The donation includes a practical dataset that includes 40 devices (25 real and 15 virtual) and 18 different attacks. While the dataset is based on current technology, it lacks consideration for future device advancements. Their experimental results showed that they had a strong capacity in binary classification (99.6%), a moderate capacity in 6-class classification (73.4%), and a lower degree of accuracy in 19-class classification (72.9%).

Akar et al. (2025) [2] released the L2D2 model, a custom Long Short-Term Memory (LSTM)-based architecture that was specifically designed to detect multiple classes of intrusion in IPMs. The model demonstrated excellent classification performance, achieving 100% accuracy in binary classification, 98% in 6-class, and 95% in 19-class scenarios. However, the high computational and memory costs present a significant obstacle to deploying on resource-limited IoMT devices.

H. Naeem et al. (2024) [17] Used deep learning models, CNN-BiGRU, CNN-BiLSTM, CNN-BiRNN, CNN-GRU, CNN-LSTM And CNN-SimpleRNN, used CICIOMT2024. The CICIoMT2024 dataset contains only six different traffic types: DDoS, DoS, MQTT, RECON, SPOOFING, and Benign. Contrasting with other claims, it lacks Brute Force, Phishing, or device-specific intrusion (e.g., insulin pumps). As a result, researchers should specifically align their analysis with the dataset's actual content in order to avoid distortion.

3 PROPOSED METHODOLOGY

3.1 CICIoMT2024 Dataset

The Canadian Institute for Cyber Security developed the CICIoMT2024 dataset to serve as a

comprehensive benchmark for evaluating security solutions for the Internet of Medical Things (IoMT). It includes network traffic data from 40 IoMT devices (25 real and 15 simulated) that are subject to 18 different cyber-attacks, and primarily focuses on three key protocols: (e.g., Wi-Fi, MQTT, and Bluetooth), the number of features in the CICIoMT2024 dataset is 45 features represent different network traffic characteristics, which help in analyzing normal and malicious activities [5].

The dataset is organized as follows:

Contains a csv/ folder with two subfolders:

- train/ – Data for training deep learning model.
- test/ – Data for evaluating/testing model.

The dataset used was unlabeled, so we added the labels and calculated the attack types as shown in Table 1 The dataset is divided into: (Training 80% / Testing 20%).

3.2 Data Preprocessing

Most available datasets contain unwanted elements (missing and duplicate values) that should be removed or transformed. The preprocessing step is crucial to obtain a suitable dataset [14].

Standard Deviation. It is a measure of how data is dispersed around the arithmetic mean. If the standard deviation is equal to zero or close to zero, it means the data does not add new information and may be unhelpful in analysis, so it is removed. If the standard deviation is large or above zero, the data contains useful information and is therefore studied.

The training file and test file are combined for extraction the columns that contain numerical data such as float64 and integer64 to calculate the Standard deviation value for each numerical column while avoiding non-numerical columns such as texts (object).

The Drate column was deleted because its standard deviation value is zero and does not add new information, so the number of features will be reduced from 45 to 44.

Handling with missing values and duplicate values:

1) Missing values. that are not present or cannot be accessed for a particular observation in a dataset. This can be caused by a variety of reasons, such as human error, data corruption, or system failures. Missing values can impact data analysis and machine learning models, requiring techniques like imputation (filling in missing values with estimates) or removal of incomplete records.
2) Duplicate Values. refer to repeated data entries within a dataset. These can arise due to errors in the collection of data, the combination of datasets, or problems with the system. Duplicates can lead to biased analysis and must be handled by identifying and removing redundant records while preserving necessary information.

Table 1: The splitting ratio of 80:20 for training and testing and distribution of the dataset.

Label	Train Count	Test Count	Total	Train 80%	Test 20 %
TCPIPDDoSUDP	1635956	362070	1998026	81.9	18.1
TCPIPDDoSICMP	1537476	349699	1887175	81.5	18.5
TCPIPDDoSTCP	804465	182598	987063	81.5	18.5
TCPIPDDoSSYN	801962	172397	974359	82.3	17.7
TCPIPDoSUDP	566950	137553	704503	80.5	19.5
TCPIPDoSSYN	441903	98595	540498	81.8	18.2
TCPIPDoSICMP	416292	98432	514724	80.9	19.1
TCPIPDoSTCP	380384	82096	462480	82.2	17.8
Benign	192732	37607	230339	83.7	16.3
MQTTDDoSConnectFlood	173036	41916	214952	80.5	19.5
ReconPortScan	83981	22622	106603	78.8	21.2
MQTTDoSPublishFlood	44376	8505	52881	83.9	16.1
MQTTDDoSPublishFlood	27623	8416	36039	76.6	23.4
ReconOSScan	16832	3834	20666	81.4	18.6
ARPSpoofing	16047	1744	17791	90.2	9.8
MQTTDoSConnectFlood	12773	3131	15904	80.3	19.7
MQTTMalformedData	5130	1747	6877	74.6	25.4
ReconVulScan	2173	1034	3207	67.8	32.2
ReconPingSweep	740	186	926	79.9	20.1

The dataset does not contain any missing values, but it contains duplicate values. In the training file, there are 5119 duplicate values, while in the test file, there are 2065 duplicate value, where duplicate values have been removed.

3.3 The proposed LSTM-DNN Hybrid Model

The model is implemented in the Kaggle environment, which supports Python and provides access to the tools and libraries required for data processing and training deep learning models. First, important libraries such as NumPy, Pandas for data analysis, scikit-learn for data normalization and one-hot encoding converts categorical variables into vectors of binary, which ensures that all categories are treated equally without presupposing a relationship between them [1], TensorFlow/Keras for model development and training, and Matplotlib and Seaborn for visualization and performance analysis are imported.

The CICIoMT2024dataset is loaded the data is segmented from its origin into training 80% and testing 20%. The dataset does not contain labels. We add the label and separating the input features (X) from the target labels (y). The number of classes is determined based on user input using a label mapping dictionary. The categorical labels are converted to numerical values using a label encoder and then one-hot encoded to prepare the labels for training.

To ensure stable training, the input features were normalized using a standard scaler. The data was then reshaped into the sequential format required by the LSTM model. A sequential model was created that included LSTM layers for sequential data processing and batch normalization is a method of training that facilitates the stabilization and acceleration of deep neural networks by normalizing inputs' values to each layer as shown in Figure 2. It decreases the internal rate of change, which allows for a quicker convergence and increased performance [6]. Dropout is a method of regularization that randomly turns off part of the neurons during training in order to avoid overfitting. This causes the model to learn additional, more generalized features [12],[10], were followed by fully connected DNN layers with ReLU activation. At the output layer, SoftMax activation function was used for classification.

The model was optimized using the Adam optimizer with a learning rate of 0.0001, this default is commonly used because it balances the speed of convergence and stability. A batch size of 64 [4] was chosen based on experimental testing, this size provides an effective compromise between training speed and generalizability. The training process was conducted for 50 epochs, which was sufficient for the model to converge without the signs of overfitting, as observed by monitoring the loss of validation. Early stopping, which employed a patience of 5 epochs, was employed to halt the training process if no significant improvement was observed. This prevented the unnecessary expenditure of computational resources and overfitting. These hyperparameters were derived from the initial search for a good compromise between cost and performance, which is in line with prior research in the deep learning field for the purpose of intrusion detection.

After training is complete, trends in training and validation accuracy are monitored and performance curves are analyzed to identify potential issues with data fit. Finally, the trained model is evaluated on the test dataset and saved for future predictions or deployment.

To fully assess the proposed intrusion detection model, the dataset was divided into three different scenarios: binary (2 classes), categorical (6 classes), and multiclass (19 classes). This multiple-level strategy was employed to represent different levels of difficulty with classification and to assess the model's viability in practical environments. Binary classification allows for an easy distinction between legitimate and criminal traffic. Six-class categorization enables the classification of the attack type as either a targeted or unsystematic attack. This is important for developing an appropriate response strategy. Meanwhile, the 19-classification system provides a specific explanation of the various attack vectors that are important in comprehending the specific threat in healthcare systems that are critical to safety.

The main phases of the proposed model shows in Figure 1.

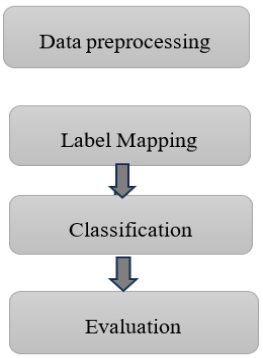

Figure 1: Main phases of the proposed model.

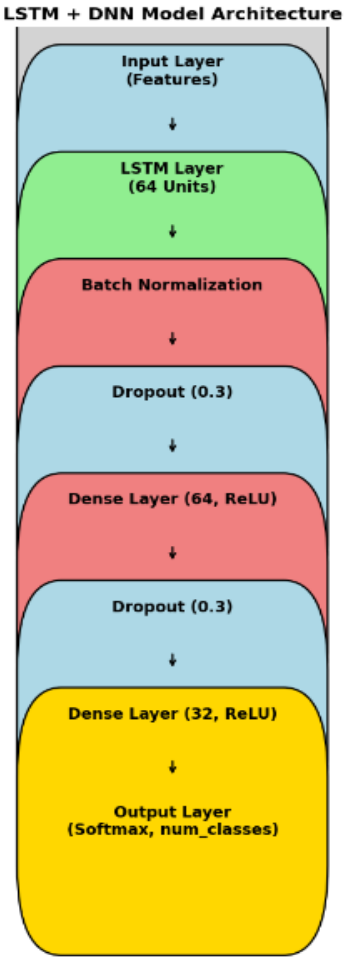

Figure 2: LSTM-DNN Hybrid Model architecture.

4 RESULTS AND DISCUSSION

The performance of the proposed LSTM- DNN Hybrid Model is rigorously evaluated on the CICIoMT2024 dataset, which includes binary, six-class, and 19-class classification tasks. The results for each task are detailed below, along with the precision, recall, and F1 -score metrics for different cyberattack types, and distribution of attack and confusion matrix for each model.

The results of this study suggest the potential of deep learning, specifically hybrid LSTM–DNN structures, in improving the safety of the internet-based medical environment (IoMT). By accurately recognizing and categorizing various attacks in real time, these models can be incorporated into smart healthcare systems that preserve patient data, ensure uninterrupted communication, and reduce the likelihood of life-threatening disruptions. This is of special importance for applications like remote patient monitoring, smart pumping devices, and connected tools that are diagnostic, these tools can directly affect the safety of patients, and the continuity of their treatment.

The proposed LSTM-DNN hybrid model is superior to traditional models like LSTM alone and DNN in multiple levels of classification. In the 6-class and 19-class scenarios, its benefits became more significant. For example, the accuracy of the proposed model was 99.47% for 6 classes and 98.43% for 19 classes, while the DNN model was drastically reduced to 73.4% and 72.9%, respectively [5]. Even the LSTM model, which is more powerful than the DNN, had accuracy (95% in 19 classes) lower case than the hybrid strategy [2]. And CNN integrated with BiGRU, BiLSTM, BiRNN, GRU, LSTM, CNN-SimpleRNN, they got an accuracy between 93% and 94% and did not work on all the data sets [17].

These enhancements are not only numerical but also structural. The LSTM layers are effective at learning the temporal patterns of traffic flow associated with typical IoMT data. However, without additional depth, they may lack the capacity to extract high-dimensional features. By combining DNN layers, the model's ability to learn more complex, abstract representations is increased, this enables the model to both generalize and differentiate between classes that are overlapping. This hybrid alliance enables the model to have a superior performance compared to its counterparts, especially in challenging, unequal, and corruptible environments.

4.1 Binary Classification Result

The proposed LSTM-DNN model had a superior performance in the binary classification task, having an accuracy of 99.64%, a precision of 99.64%, a recall of 99.64%, and an F1-score of 99.64%. These metrics demonstrate the model's effective ability to differentiate between criminals and regular traffic. The high rate of recall in particular demonstrates its effectiveness in recognizing all instances of attacks, this is crucial to healthcare applications that may be lost by a missed attack and lead to serious consequences.

The Tables 2 and 3 show the distribution of data between the attack and benign classes in the training and test sets.

Table 2: Train value count.

	Label	Count
1	Attack	6962980
2	Benign	192732

Table 3: Test value count.

	Label	Count
1	Attack	1574510
2	Benign	37607

Figure 3 shows a clear and organized distribution of data, making it easy to understand the contrast between the attack and benign categories.

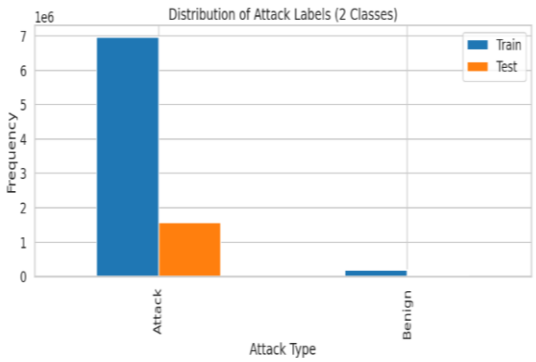

Figure 3: Distribution of training and testing samples for attack and benign classes in binary classification (2-class).

The confusion matrix in Figure 4 illustrates the model's performance in classifying Attack and Benign instances. A large number of samples are correctly classified, with 1,572,281 attack instances and 34,105 benign instances accurately identified. However, some misclassifications occur, where 2,229 attack samples are incorrectly classified as benign, and 3,502 benign samples are mistaken for attacks. This indicates strong model performance with a relatively low error rate.

Figure 5 illustrates the model's performance over epochs. In the left graph, both training and validation accuracy increase rapidly and stabilize at high values, indicating effective learning. In the right graph, training and validation loss decrease progressively, showing model improvement and error reduction over time, with no clear signs of overfitting.

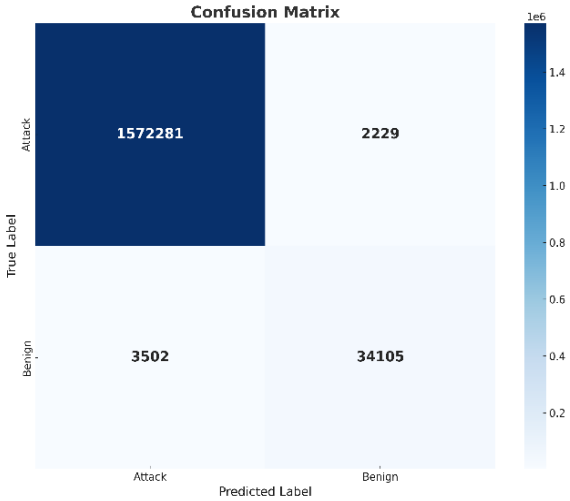

Figure 4: Confusion Matrix of attack and benign in 2-class.

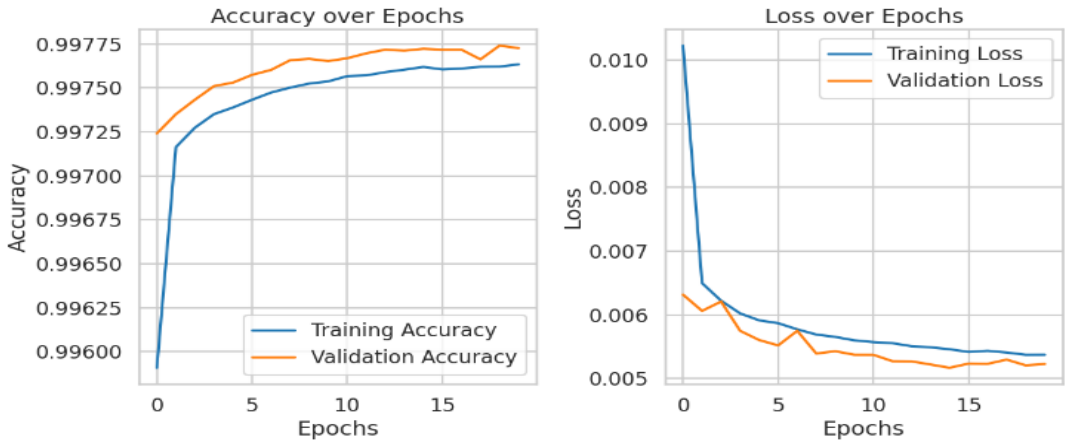

Figure 5: Accuracy and loss curves for 2-class.

4.2 Multi-Class (6-Class) Classification Results

When augmented with a 6-class classification task, the model still exhibited a reliable performance, achieving 99.47% accuracy, 99.62% precision, 99.47% recall and an F1 score of 99.53%. This demonstrates its ability to differentiate between five types of attacks and benign. Compared to the standalone LSTM and DNN models, which exhibited a decrease in performance in the presence of multiple classes, the LSTM-DNN architecture maintained a high degree of precision and generalizability.

The Tables 4 and 5 show the distribution of training and test data across various categories, such as "DDoS," "DoS," and "Benign." The numbers indicate a large sample size for "DDoS," followed by "DoS" and "MQTT," while the other categories have fewer samples. This distribution provides insight into the data volume for each category during training and testing.

Table 4: Train value count.

	Label	Count
1	DDos	1066762
2	Dos	416676
3	MQTT	63715
4	Benign	37607
5	RECON	25613
6	SPOOFING	1744

Table 5: Test value count.

	Label	Count
1	DDos	4779187
2	Dos	1805529
3	MQTT	262938
4	Benign	192732
5	RECON	99279
6	SPOOFING	16047

The chart in Figure 6 illustrates the distribution of samples across six data categories in the training and test sets. "DDoS" and "DoS" have the highest number of samples, while categories like "SPOOFING" and "RECON" have significantly fewer. This visual representation helps in understanding the prevalence of each data type in both sets.

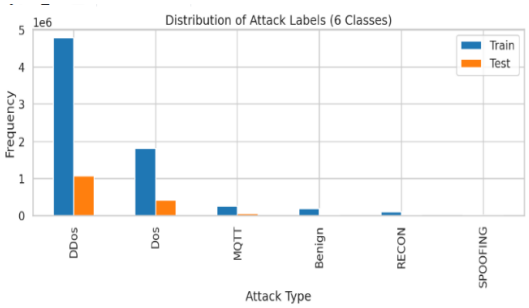

Figure 6: Distribution of attacks and benign in 6-class.

The confusion matrix in Figure 7 displays the model's performance in classifying different categories, with high values along the main diagonal indicating strong classification accuracy, particularly for "DDoS" and "DoS," where most samples were correctly classified. However, some misclassifications are present, such as "Benign" samples being classified as "SPOOFING". This suggests that the model demonstrates high efficiency in distinguishing most categories.

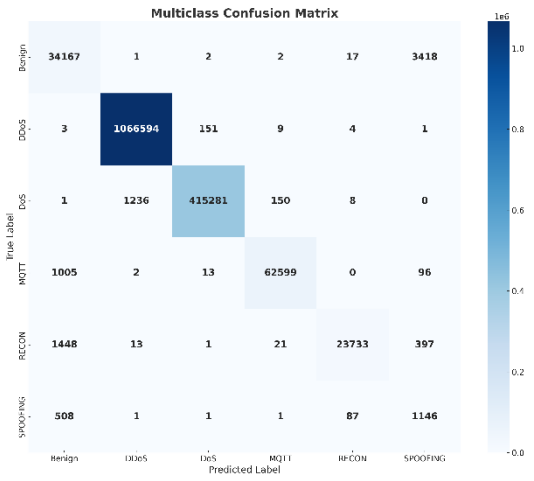

Figure 7: Confusion Matrix of attack and benign in 6-class.

The Figure 8 shows the progression of the model's accuracy and loss over epochs. In the left graph, accuracy increases rapidly, indicating effective learning. In the right graph, loss drops sharply in the early epochs and then stabilizes at a low value, reflecting improved model performance.

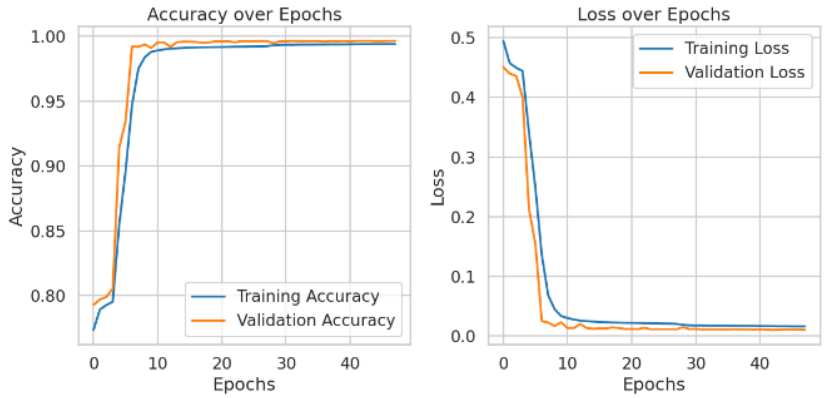

Figure 8: Accuracy and Loss Curves for 6-class.

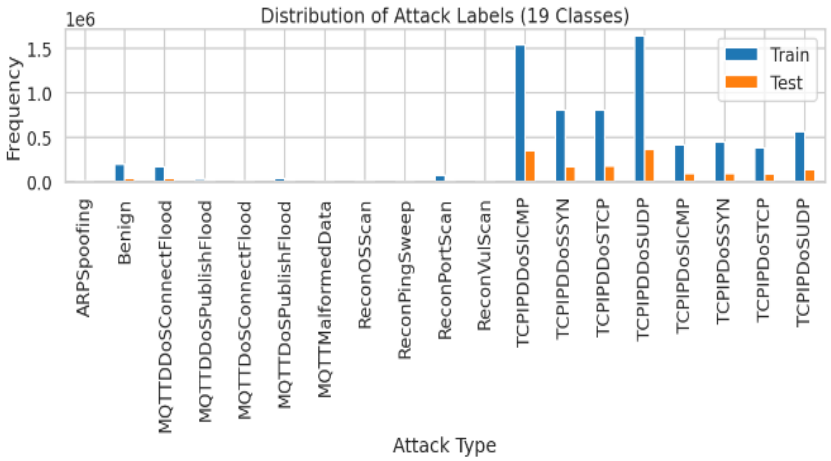

Figure 9: Distribution of attacks and benign in 19-class.

4.3 Complex Multi-Class (19-Class) Results

In the more difficult 19-class categorization task, which involves a greater degree of granularity and often has a similar nature to the attacks, the model had an accuracy of 98.43% a precision of 98.77%, a recall of 98.43% and an F1-score of 98.25%. Despite a slight decrease in performance compared to binary and 6-class tasks, the model still performed better than conventional baselines. This drop is anticipated because of the increased difficulty in differentiating the behavior of classes that are overlapped, class imbalance, and the limited amount of training data for the minority classes. Involving 18 different attack types in addition to benign attacks.

The Table 1 show the distribution of samples across various attack categories and benign data in the training and test sets. "TCPIPDDoS" attack types have the highest number of samples, while other categories like "ReconVulScan" and "MQTTMalformedData" have fewer. This distribution reflects the diversity of data used for training and testing.

Figure 9 shows the sample distribution of 19 data categories in the training and test datasets. "TCPIPDDoS" attacks have the highest number of samples, while other categories like "ARPSpoofing" and "MQTTMalformedData" have fewer. This representation helps to understand the popularity of each category in the dataset.

The confusion matrix shown in Figure 10 illustrates the model's performance in classifying 19 categories. High values along the main diagonal indicate strong classification accuracy, especially for "TCPIPDDoS" and "Benign.". These results reflect the model's efficiency in distinguishing between classes, with some limited errors.

In Figure 11 the left graph shows the gradual improvement in training and validation accuracy across epochs, while the right graph illustrates the consistent reduction in loss, indicating that model performance improves over time.

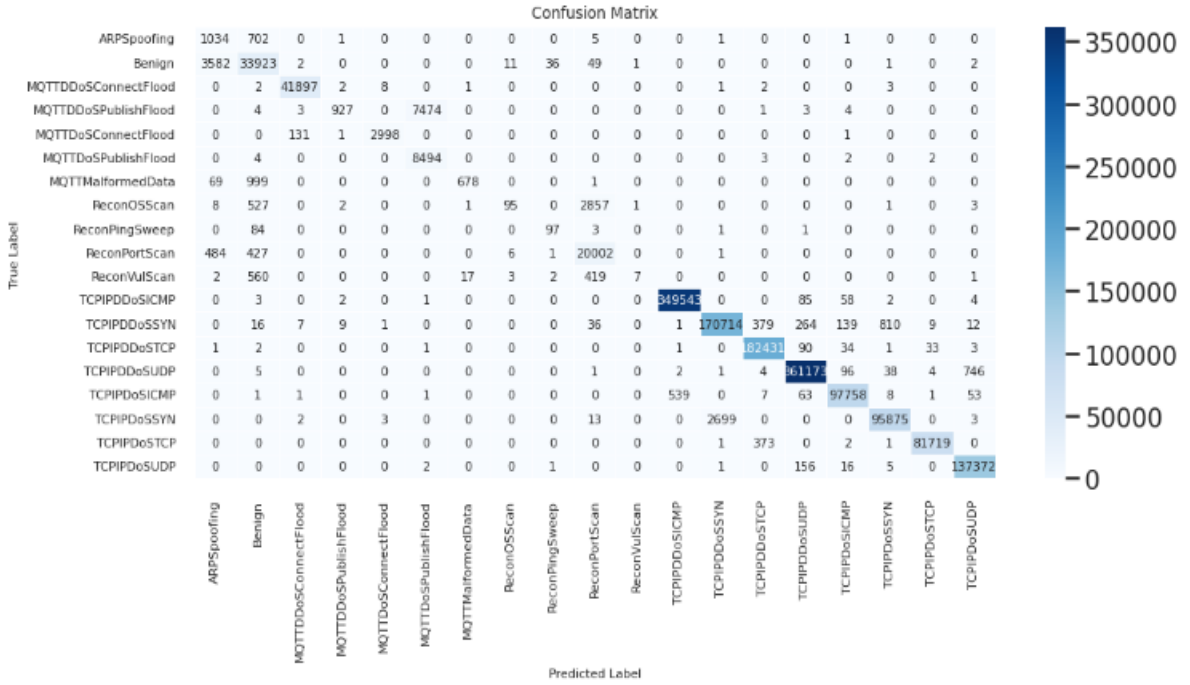

Figure 10: Confusion Matrix of attacks and benign in 19-class.

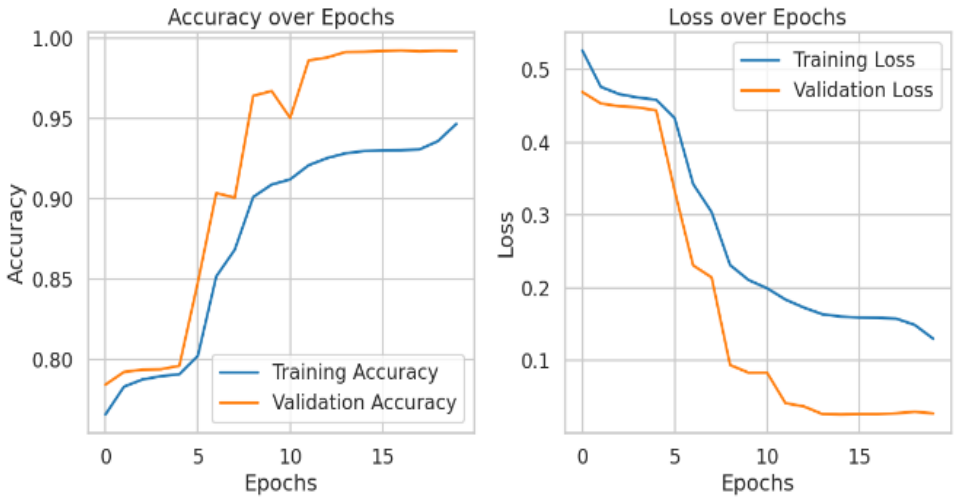

Figure 11: Accuracy and Loss Curves for 19-class.

Table 6: Performance Metrics of DL Models Across Different Classification Tasks in CICIOMT2024 dataset.

Model	Classification Task	Accuracy	Precision	Recall	F1-Score	Attack family	Count	Number of samples in category
Proposed LSTM-DNN Hybrid Model	2 class	99.6445	99.6398	99.6445	99.6415	Attack Benign	8,537,490 230,339	8,537,490 230,339
	6 class	99.4667	99.6227	99.4667	99.5310	DDos Dos MQTT RECON SPOOFING Benign	5,845,949 2,222,205 326,653 124,892 17,791 230,339	5,845,949 2,222,205 326,653 124,892 17,791 230,339
	19 class	98.4256	98.7723	98.4256	98.2460	Classifying 19 categories in Table 1	18 type of attacks 8,537,490 Benign 230,339	18 type of attacks 8,537,490 Benign 230,339
DNN [5]	2 class	99.6	95.6	94.8	95.2	Attack Benign	8,537,490 230,339	8,537,490 230,339
	6 class	73.4	72.5	69.3	66.5	DDos Dos MQTT RECON SPOOFING Benign	5,845,949 2,222,205 326,653 124,892 17,791 230,339	5,845,949 2,222,205 326,653 124,892 17,791 230,339
	19 class	72.9	64.9	55.3	52.2	classifying 19 categories in Table 1	18 type of attacks 8,537,490 Benign 230,339	8,537,490 230,339
LSTM [2]	2 class	100	100	100	100	Attack Benign	8,537,490 230,339	8,537,490 230,339
	6 class	98.0	98.0	98.0	98.0	DDos Dos MQTT RECON SPOOFING Benign	5,845,949 2,222,205 326,653 124,892 17,791 230,339	5,845,949 2,222,205 326,653 124,892 17,791 230,339
	19 class	95.0	96.0	95.0	95.0	Classifying 19 categories in Table 1	18 type of attacks 8,537,490 Benign 230,339	18 type of attacks 8,537,490 Benign 230,339
CNN-BiGRU	–	94.0	–	–	–			
CNN-BiLSTM	–	93.0	–	–	–			
CNN-BiRNN	–	94.0	–	–	–			
CNN-GRU	–	93.0	–	–	–	DDoS	5,845,949	20,000
CNN-LSTM	–	93.0	–	–	–	Normal	230,339	25,000
CNN-SimpleRNN [17]	–	93.0	–	–	–			

4.4 Evaluation Metrics

To assess the performance of the proposed model, standard classification metrics were used, including Accuracy, Precision, Recall, and the F1-score. These metrics follow their conventional definitions widely adopted in machine learning and pattern recognition research [2], [9], [14], [16].

- Accuracy reflects the overall proportion of correctly classified instances.
- Precision measures the correctness of positive predictions.
- Recall (Sensitivity) quantifies the proportion of actual positive cases correctly detected by the model.
- F1-score represents the harmonic mean of Precision and Recall, providing a balanced assessment when class distributions are uneven.

These well-established metrics are sufficient to rigorously evaluate the model's predictive performance and ensure comparability with existing approaches.

4.5 Comparative and Scientific Analysis

The outcomes of the three classification tasks are evident of the effectiveness and versatility of the proposed hybrid LSTM-DNN model. Its superior capabilities are attributed to the combination of temporal pattern recognition (via LSTM) with abstract feature extraction (via DNN). The consistent superiority of the model over standalone LSTM and DNN models, especially in terms of accuracy, demonstrates its potential for real-world security applications in the IoMT. The logical progression of decreasing performance from binary to multi-class classification is expected scientifically, as the complexity of the task and the similarity between classes increase.

Table 6 shows the performance comparison of different models on different classification tasks (2, 6, and 19 classes) based on accuracy, precision, recall, and F1 score.

The LSTM-DNN hybrid model is the best overall because it combines the advantages of LSTM and DNN, ensuring balanced performance on all classification tasks.

5 LIMITATIONS OF THE STUDY

Despite the encouraging performance of the proposed LSTM-DNN model on the CICIoMT2024 dataset, several caveats should be mentioned. First, the dataset may have a bias because of its synthetic nature and the limited number of real-world scenarios and devices represented. This may have an effect on the generalizability of the model to other IoMT environments with different traffic patterns or attacks. Second, although the model demonstrated high accuracy in off-line experiments, its effectiveness in real time intrusion detection systems has not been empirically tested and may be adversely affected by limitations such as processing delay, resource constraints, and evolving attacks. Finally, the model's performance may differ depending on the distribution of classes is balanced or novel (day zero) attacks are present in the training data. Future endeavors should investigate practical adaptive learning methods and frequent real time comparison to address these issues

6 CONCLUSIONS

This article described a hybrid LSTM-DNN deep learning model that was intended to detect and categorize cyberattacks in IoMT environments. The proposed model had a high performance that was achieved with a 99.64% accuracy for binary classification, 99.47% for 6 classes, and 98.43% for 19 classes. Other than that, it consistently demonstrated superior generalizations and consistency.). Leveraging the CICIoMT2024 dataset, which encompasses diverse cyberattack scenarios, our model demonstrated superior performance in detecting intrusions across binary, six-class, and 19-class classification tasks. The results show that the proposed LSTM-DNN hybrid model outperforms traditional machine learning techniques, achieving high accuracy, precision, recall, and F1-score.

The study underscores the importance of advanced deep learning methodologies in securing IoMT networks, which face an increasing number of cyber threats. The combination of LSTM for sequential data analysis and DNN for feature extraction effectively identifies normal and malicious traffic. Additionally, data preprocessing techniques, dynamic label mapping, and model optimization contributed to the model's robustness and generalizability.

Despite its promising performance, challenges remain regarding real-time deployment and computational efficiency. Future research will focus on optimizing model performance, integrating attention mechanisms for improved feature selection, and evaluating real-world applicability. Expanding the dataset with additional attack variations and real-

time traffic data will further enhance the reliability of intrusion detection in IoMT environments.

As cyber threats continue to evolve, ensuring the security of IoMT systems is critical to protecting patient data and healthcare infrastructure. The findings of this study contribute to the ongoing development of intelligent, automated cybersecurity solutions, paving the way for more resilient and adaptive intrusion detection systems in medical networks.

These findings confirm the model's potential for real-time deployment in healthcare settings where speed and accuracy are paramount.

ACKNOWLEDGMENT

I am deeply grateful to my supervisor, Associate Professor Dr. Amer Abdulmajeed Abdulrahman, for his unwavering support, insightful guidance, and invaluable encouragement throughout this research. My sincere gratitude is extended to the staff of the Department of Computer Science, Faculty of Science, University of Baghdad, for their continued cooperation and assistance. Their expertise and dedication have contributed significantly to the success of this study. I would also like to express my sincere gratitude to all those who have supported me intellectually and emotionally throughout this journey. Their encouragement and faith in me have been truly invaluable.

REFERENCES

[1] A. A. Abdualrahman and M. K. Ibrahem, “Intrusion detection system using data stream classification,” Iraqi J. Sci., pp. 319–328, 2021, doi: 10.24996/ijs.2021.62.1.30.

[2] G. Akar, S. Sahmoud, M. Onat, Ü. Cavusoglu, and E. Malondo, “L2D2: A Novel LSTM Model for Multi-Class Intrusion Detection Systems in the Era of IoMT,” IEEE Access, vol. 13, pp. 7002–7013, 2025, doi: 10.1109/ACCESS.2025.3526883.

[3] M. Barnett, J. Womack, C. Brito, K. Miller, L. Potter, and X. L. Palmer, “Botnets in healthcare: Threats, vulnerabilities, and mitigation strategies,” in European Conf. Cyber Warfare and Security, 2024, pp. 58–65, doi: 10.34190/eccws.23.1.2345.

[4] S. F. Behadil and N. K. Mhalhal, “Mobility prediction based on deep learning approach using GPS phone data,” Ibn Al-Haitham J. Pure Appl. Sci., vol. 37, no. 4, pp. 423–438, 2024, doi: 10.30526/37.4.3916.

[5] S. Dadkhah, E. C. P. Neto, R. Ferreira, R. C. Molokwu, S. Sadeghi, and A. A. Ghorbani, “CICIoMT2024: A benchmark dataset for multi-protocol security assessment in IoMT,” Internet Things, vol. 28, p. 101351, 2024, doi: 10.1016/j.iot.2024.101351.

[6] S. De and S. Smith, “Batch normalization biases residual blocks towards the identity function in deep networks,” Adv. Neural Inf. Process. Syst., vol. 33, pp. 19964–19975, 2020.

[7] J. M. Ehrenfeld, “Wannacry, cybersecurity and health information technology: A time to act,” J. Med. Syst., vol. 41, no. 7, p. 1, 2017, doi: 10.1007/s10916-017-0752-1.

[8] M. L. Hernandez-Jaimes, A. Martínez-Cruz, K. A. Ramírez-Gutiérrez, and E. Guevara-Martínez, “Enhancing machine learning approach based on Nilsimsa fingerprinting for ransomware detection in IoMT,” IEEE Access, 2024, doi: 10.1109/ACCESS.2024.3480889.

[9] W. F. Kamil and I. J. Mohammed, “Adapted CNN-SMOTE-BGMM deep learning framework for network intrusion detection using unbalanced dataset,” Iraqi J. Sci., pp. 4846–4864, 2023, doi: 10.24996/ijs.2023.64.9.43.

[10] W. F. Kamil and I. J. Mohammed, “Deep learning model for intrusion detection system utilizing convolution neural network,” Open Eng., vol. 13, no. 1, p. 20220403, 2023, doi: 10.1515/eng-2022-0403.

[11] D. Koutras, G. Stergiopoulos, T. Dasaklis, P. Kotzanikolaou, D. Glynos, and C. Douligeris, “Security in IoMT communications: A survey,” Sensors, vol. 20, no. 17, p. 4828, 2020, doi: 10.3390/s20174828.

[12] E. Poslavskaya and A. Korolev, “Encoding categorical data: Is there yet anything ‘hotter’ than one-hot encoding?,” arXiv preprint arXiv:2312.16930, 2023, doi: 10.48550/arXiv.2312.16930.

[13] A. D. Saleem and A. A. Abdulrahman, “Attacks detection in Internet of Things using machine learning techniques: a review,” J. Appl. Eng. Technol. Sci., vol. 6, no. 1, pp. 684–703, 2024, doi: 10.37385/jaets.v6i1.4878.

[14] K. Saurabh, S. Sood, P. A. Kumar, U. Singh, R. Vyas, O. P. Vyas, and R. Khondoker, “LBDMIDS: LSTM-based deep learning model for intrusion detection systems for IoT networks,” in 2022 IEEE World AI IoT Congress (AIIoT), 2022, pp. 753–759, doi: 10.48550/arXiv.2207.00424.

[15] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, “Efficient processing of deep neural networks: A tutorial and survey,” Proc. IEEE, vol. 105, no. 12, pp. 2295–2329, 2017, doi: 10.1109/JPROC.2017.2761740.

[16] P. Udayakumar and R. Anandan, “Evaluation of protocol-centric IDS for the IoMT leveraging ML techniques,” in 2024 IEEE World AI IoT Congress (AIIoT), 2024, pp. 546–551, doi: 10.1109/AIIoT61789.2024.10578945.

[17] H. Naeem, A. Alsirhani, F. M. Alserhani, F. Ullah, and O. Krejcar, “Augmenting Internet of Medical Things security: Deep ensemble integration and methodological fusion,” Comput. Model. Eng. Sci., vol. 141, no. 3, pp. 2185–2223, 2024.