AI AND SUSTAINABILITY COURSE WORK EEEM073

Dr Erick G. Sperandio Nascimento

(Associate Professor (Reader) in AI for Clean Air Systems)

Medical IoT Devices Threat Detection using Deep Learning

A Couse work Report

Submitted by

Hariharan Natesanpillai Ramalingam [URN number: 6900616]

Date: 14/05/2025 - Semester 2, 2025

Table of Contents

ABSTRACT

Healthcare facilities have improved patient monitoring because of the Internet of Things (IoT) technology while it introduced new threats through medical IoT devices. Technology presents security risks because it enables real-time cyber-attacks on medical IoT devices. Real-time security attacks against medical IoT data are challenging to address because traditional security tools struggle to adapt to complex data streams along with fast-evolving attack patterns. The development of a medical IoT threat detection model based on Bi-Directional Long Short-Term Memory (Bi-LSTM) technology serves as our solution to this problem. Through the use of deep learning, our approach performs a -analysis to identify anomalies which strengthens the security posture of medical IoT systems. A Bi-LSTM model demonstrates enhanced detection accuracy and system stability by capturing directional dependencies in both sequential patterns. Our model delivers better detection outcomes than standard security methods when applied to medical IoT networks because it shows high precision and recall as well as F1-score in detecting diverse cyber-attacks.

INTRODUCTION

The Internet of Medical Things (IoMT) is a radical change in the medical area, enabling the improvement of patient monitoring, diagnostics, and care with the help of network-connected medical equipment. However, even as IoMT system deployment is increasing, so are the security problems that come with connected systems. The need for security in IoMT is urgent; the vulnerabilities in these systems could lead to catastrophic consequences, including patient safety threats and personal health information breaches [11-12].

The architecture processes all input sequence data to gain context information which establishes its clinical text comprehension significance [1]. The healthcare industry utilizes Bi-LSTMs for automated extraction of medical codes (e.g., ICD codes) from unstructured clinical narratives to support clinical documentation and hospital administration as well as data interoperability in electronic health records (EHR) [2].

The Feasibilityof the Bi-LSTM models, in which these are tuned with the methods such as attention mechanisms, or LIME (Local Interpretable Model-Agnostic Explanations), enables trust and transparency in AI-based clinical decision systems [3]. Simultaneously, Medical Internet of Things (IoT) devices such as wearable sensors, smart monitors, and infusion pumps are being networked into the hospital system and thereby left vulnerable to cybersecurity threats. Deep learning framework specifically RNNs, LSTMs, and autoencoders—are employed to watch unusual activity or cyberattacks in real time to safeguard patient data and device operation [4]. These deep learning IDS help in identifying threats such as malware injection, unauthorized access, and data tampering in IoT-enabled healthcare environments [5].

Association with Sustainability:

- *Improved Healthcare Efficiency:* Bi-LSTM models help healthcare providers to achieve enhanced accuracy in clinical documentation while reducing the need for manual intervention which leads to faster workflows and gives healthcare professionals more time to provide patient care [6].

- *Fewer Errors and Spending:* The implementation of AI systems for medical coding minimizes human errors in the process which helps healthcare facilities avoid billing mistakes and payment delays while maximizing their resource utilization [7].
- *Effective Model:* Implementing successful Bi-LSTM models using approaches such as pruning, quantization, and distillation reduces compute needs and power consumption, hence improving AI sustainability [8].
- *Secure and Reliable Digital Health:* Medical IoT network threat detection ensures both patient data and device security while maintaining the reliability of healthcare operations. Digital sustainability becomes critical because system downtime leads to reputational harm and resource-intensive system rebuilding after a security breach [9].
- *Responsible and Ethical AI:* The implementation of explainability tools in Bi-LSTM models enables ethical deployment and user trust improvement while following responsible and transparent AI principles in medical applications [10].

EXISITNG SYSTEMAND PROBLEMS

The Internet of Medical Things (IoMT) is a rapidly developing, extremely interconnected medical device ecosystem with revolutionary potential for patient monitoring and healthcare provision. The system contains many substantial cybersecurity weaknesses. Multiple research works analysed different threats that exist within these complex systems. According to Barnett and colleagues (2024), the system-of-systems structure of medical internet of things (IoMT) networks create multiple attack paths between integrated subsystems [17].

Also, Wani et al. (2023) found loopholes in real-life applications in the utilization of remote patient monitoring, telemedicine software, and intelligent medicine delivery systems with a focus on attacks like session hijacking, denial-of-service (DoS) attacks, and man-in-the-middle (MITM) attacks [24]. Extending this, Kondeti and Bahsi (2024) presented an IoMT attack taxonomy and categorization into spoofing, eavesdropping, data exfiltration, and reconnaissance, hence offering an orderly view of how attackers use IoMT layers [18].

The security of IoMT devices faces a critical technical problem because numerous devices have limited resources. The absence of processing power combined with restricted battery life and memory prevents the devices from implementing extensive security frameworks which results in their vulnerability to advanced cyberattacks [19].

To facilitate research and benchmarking, the CICIoMT2024 dataset was introduced—describing traffic from 40 attacked IoMT devices subjected to 18 varied cyberattacks. While abundant, the dataset challenges traditional machine learning (ML) approaches. As noted by Dadkhah et al., although binary classification between benign and malicious traffic achieves reasonably good accuracy, multi-class classification—especially identifying between similar DDoS variations—is still challenging with model accuracy below 73% [20].

To address these drawbacks, our proposal leverages the use of a Bidirectional Long Short-Term Memory (Bi-LSTM) network. Bi-LSTM models are strong in extracting temporal relations in sequential data such as network traffic flows and thus well suited for analyzing the time-oriented patterns of IoMT

communications. Bi-LSTM networks, as opposed to traditional ML classifiers, have the ability to learn the forward and backward contextual relationships in traffic sequences, which can help enhance the detection of subtle and dynamic attack patterns across classes.

ANALYSIS OF THE DATASET - CICIoMT2024

The CICIoMT2024 dataset stands out as one of the most valuable resources for conducting research into cybersecurity solutions for the Internet of Medical Things (IoMT). As the use of connected medical devices—such as patient monitoring systems and remote healthcare tools continues to grow, ensuring their security has become increasingly critical.

A. Dataset Overview:

The data set has a number of data points, including device logs, traffic flows, security incidents, and sensor readings from IoMT devices. Some of the important variables usually are:

- The Dataset has device logs which include error and activity logs from connected devices.
- The Traffic Patterns such as Communication among devices, bandwidth consumption, and protocols.
- The Dataset includes the Security Logs such as Threats, unauthorized access, and breach logs.
- IDs for each IoMT device were present at the dataset.
- The Sensor data which includes Device readings of medical devices like heart monitors, infusion pumps are also present in the dataset.
- User Behavior logs such as access logs of patient or medical staff were also present.

B. Data Cleaning and Preprocessing:

Data was also preprocessed to eliminate missing values and outliers, particularly from traffic and device logs. Time-based data, such as timestamps, was normalized for proper time-series analysis. Traffic and sensor readings were normalized in terms of processes to accommodate different devices.

C. Data Visualization:

Fig. 3 illustrates a plot that shows pairwise relationships for selected features of the CICIOMT2024 dataset, a large and recently updated set of network traffic datasets particularly developed for application in the field of cybersecurity research and intrusion detection system (IDS) analytics.

The data set includes generic descriptions of the various forms of traffic, from innocent to malicious, thereby offering close resemblance to real network environments. Visualization is offered in a clear manner depicting information such as Header_Length, Protocol Type, Duration, and Rate, which offers full insight into how the various parameters of network communication are interdependent and influence each other.

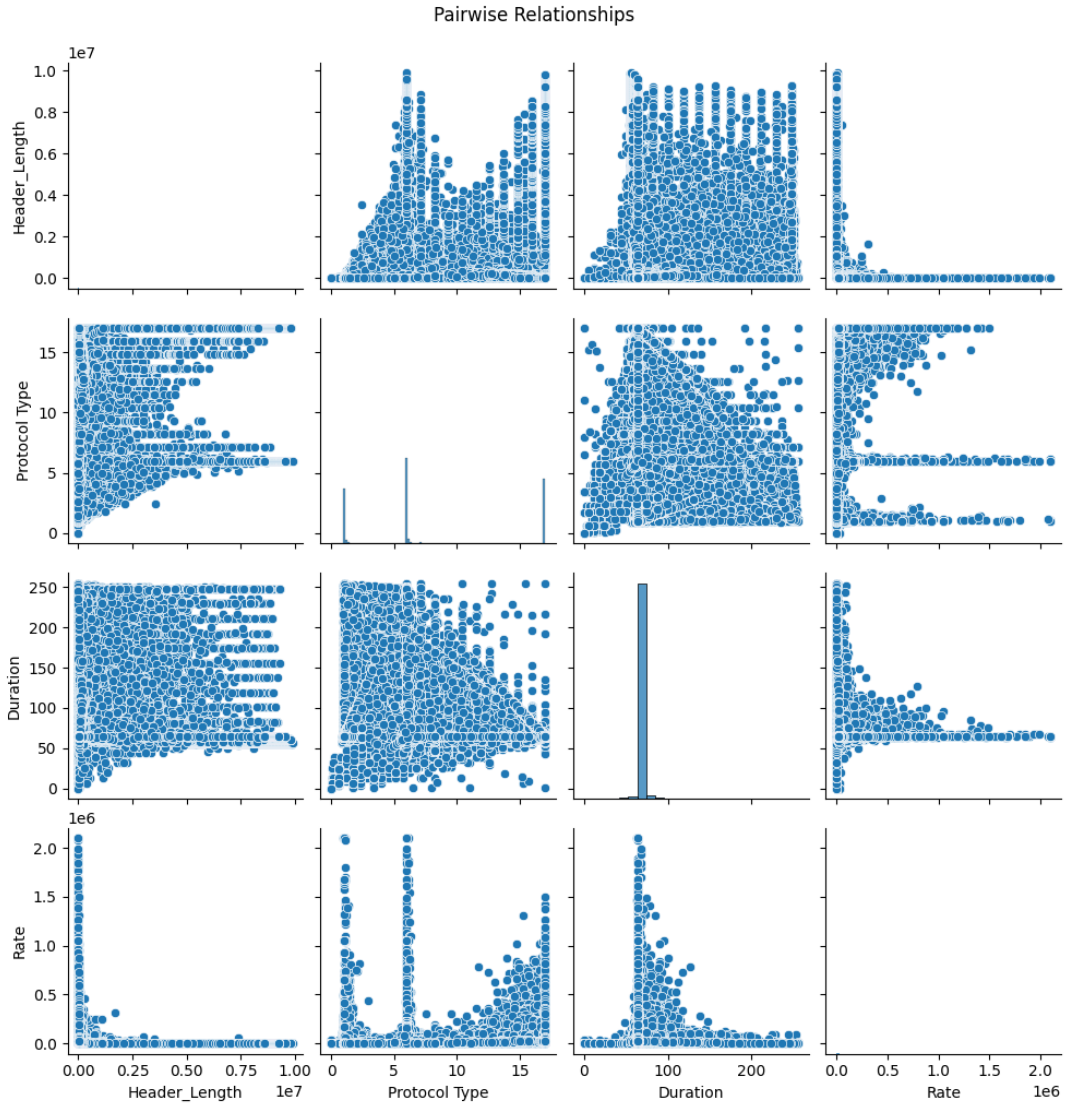


Fig 1: Pairwise relationship graph for the CICIOMT2024 Dataset

Categorical encoding of protocol type in the data set conveys dominance of most network protocols, and their diversity is emphasized in the data set. Non-uniform distribution for Rate and Duration convey variance in data rates and connection duration, which is generally indicative of composite behavior among malicious/suspicious activity and normal activity. These variations may be indicative of threats or anomalies.

The graph is a visual representation of the structural and behavioral diversity present in the CICIOMT2024 dataset. The trends and relationships among the features are representative of actual network traffic behavior, and thus the dataset is highly useful for feature selection, data pre-processing, and outlier detection tasks. Besides, the clear definition of the patterns and the possible bunching of the data points cause natively seeming data complexity and justify the usage of the dataset for training and testing advanced machine learning and deep learning models to implement in modern network security.

D. Class Distribution:

- **Benign:** 230,339 instances
- **Spoofing:** 17,791 instances
- **Reconnaissance:** 926–106,603 instances
- **MQTT Attacks:** 6,877–214,952 instances
- **DoS/DDoS Attacks:**15,904–1,998,026 instances

E. Cybersecurity Analysis:

The most significant cybersecurity problems identified are:

• **Vulnerabilities:** Old software devices were susceptible to known attacks, i.e., unauthorized access and control of data.

• **Attack Detection:** The information reports some successful detection events where intrusion detection systems detected unauthorized activity. False positives were common, however, and the information indicates the need for more sophisticated anomaly detection systems.

METHODOLOGY

To detect cyberattacks in IoMT settings, we have created a Bi-LSTM (Bidirectional Long Short-Term Memory) model that is process-time optimized to handle time-series network traffic data. This approach leverages the ability of Bi-LSTM networks to learn both forward and backward dependencies between sequential data, which is essential in detecting attack patterns. The process further incorporates data preprocessing in order to prepare the time-series data, and then Bi-LSTM layers to learn context about previous and subsequent network events. At Last, fully connected layers project the learned features into accurate attack classifications with high performance and accuracy in classifying various cyberattack types.

A. Proposed Bi-LSTM Model

In modern sequential data processing tasks, it is crucial to model temporal dependencies accurately. The Bidirectional Long Short-Term Memory (Bi-LSTM) network is a variant of the vanilla LSTM that incorporates both forward and backward temporal information and is thus a very successful architecture for sequence classification problems. The architecture, layer-wise design justification, regularization techniques, optimization techniques, and theoretical explanations of a Bi-LSTM model for multi-class threat classificationhas been detailed below.

Input Layer and Data Preprocessing:

Preprocessed time-series network traffic data is fed into the input layer. One- dimensional point array contains each sample in the dataset. Following pre-processing steps are noticed before providing the data to the model:

- **Data Directory Structure:** The data is organized under the "train" and "test" directories. Irrelevant or missing files (logs, metadata) are removed to offer data clean.

- **File Compression:** PCAP files are stored compressed as GZIP files to save storage requirements and ease handling files during processing. The purpose of dataset conversion is to avoid CPU crashing issues and to make the model to work efficiently at Google Collab.
- **File Format Transformation:** CSV files are transformed into Parquet format for improved storage efficiency and quick loading, with reduced memory overhead.
- **Attack Category Mapping:** The attack categories are pre-defined in three modes (2, 6, and 19 categories). The get_attack_category function maps the corresponding attack category label based on the filename.
- **Efficient Data Loading:** Memory-friendly process management is supported by reading the data in chunk-wise mode and incorporating the processed data directly without slowing down the memory while dealing with large sets of data.
- **Label Encoding:** Labels are label-encoded using LabelEncoder to convert the categorical labels into numerical form and then mapped into one-hot encoded vectors to fulfill model compatibility requisites.
- **Data Splitting:** The data are divided into training, validation, and test sets (training and validation 80%-20%) to enable successful model testing without causing overfitting.
- **Feature Scaling:** StandardScaler is used to scale features in a way that the data will be zero-centered with unit variance, which helps improve the performance of machine learning models.
- **Data Reshaping for Bi-LSTM, Logistic Regression and RNN:** Data reshaping is carried out to prepare for both Recurrent Neural Networks (RNNs), Logistic Regression and Bidirectional Long Short-Term Memory (Bi-LSTM). Both these models are ideal for sequence data as they can capture temporal patterns and dependencies within the dataset. Reshaping the data prepares it for processing sequences, which is required for working with sequential data.
- **Saving Preprocessed Data:** Preprocessed data is saved in.npz format for storage in a compact form and easy access, preserving transformed features, labels, and class mappings to train subsequent models.

B. Class Distribution and Training data Classification:

1. Class 2: Binary Classification (Benign or Attack)

In the binary classification case, the data is split into just two classes: Benign and attack. The bar chart (Fig. 2) clearly shows there to be a grave imbalance in the dataset, with the attack class holding most of the samples — more than 5 million — compared to a negligible number for the Benign class. This imbalance plays a crucial role in training models to be biased towards classifying the majority class unless resampling, class weighting, or outlier detection techniques are employed.

2. Class 6: Multi-class Classification (6 Broad Categories)

In the multi-Class classification (6 - Class), the data consist of six classes: DDoS, MQTT, Benign, DoS, Recon, and Spoofing. The distribution has the attack categories DDoS having the largest sample count, followed by DoS, and then much smaller frequencies for Benign, MQTT, Recon, and Spoofing. This setup allows us to learn a better understanding of the attack types than in the binary classification without rendering the problem intractable. While, class imbalance can still have an impact on model performance, especially on the minority classes.

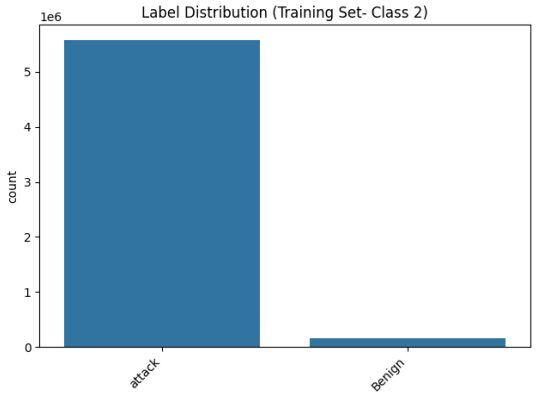


Fig 2 Label Distribution Class 2

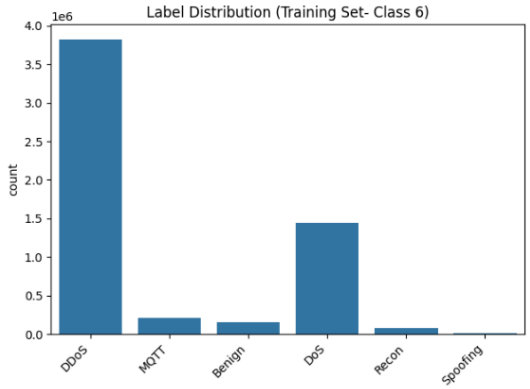


Fig 3 Label Distribution Class 6

3. *Class 19: Muti-class Classification (19 Specific Categories)*

This is the most detailed classification with 19 specific classes with some attack types such as DDoS-TCP, DoS-UDP, Recon-Port_Scan, etc. This shows an even larger imbalance, with hardly any classes such as DDoS-TCP and DoS-SYN that are common, and some such as Recon-VulScan, Spoofing, and MQTT-Malformed_Data with hardly any instances. This high-resolution setup is the most descriptive to actual cyber security systems but most challenging from the machine learning perspective, as there are predominantly sparse classes with few examples, and they require special treatment (e.g., focal loss, SMOTE, or hierarchical classification).

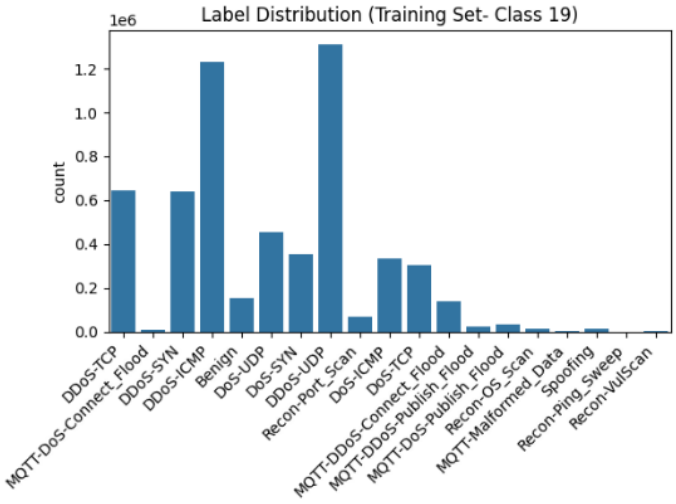


Fig 4 Label Distribution Class 19

C. Bidirectional LSTM Model

1. *First Bidirectional LSTM Layer*

- **Layer:**Bidirectional(LSTM(32, return_sequences=True))
- **Purpose:**
  - Gets temporal features by considering both future and past direction in the sequence.

- Return_sequences=True enables the output to be a full sequence to stack over the next LSTM layer.

- **Followed by:**
  - *BatchNormalization:* Normalizes andstabilizes training by accelerating.
  - *Dropout (0.3):* Regularizes by randomly disabling 30% of neurons while training.

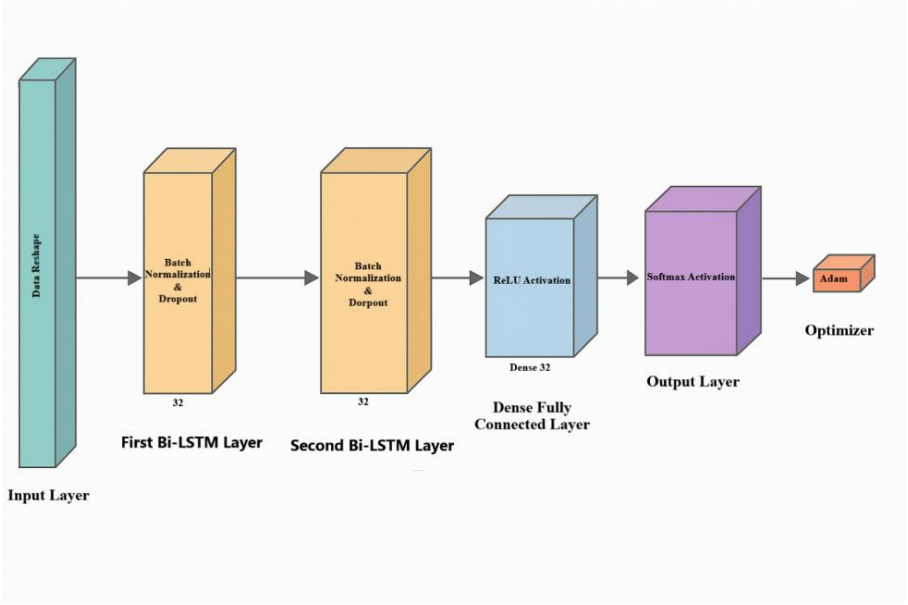


Fig 5: Architecture Diagram of Bi-LSTM Model

2. *Second Bidirectional LSTM Layer*

- **Layer:**Bidirectional(LSTM(32))
- **Purpose:**
  - Merges the temporal features into a single environment vector.
  - Return_sequences=False implies that only the final hidden states are propagated to the next layer.
- **Followed by:**
  - *Batch Normalization and Dropout Again:* Repeated after the second Bi-LSTM layer to maintain regularization and stability during training.

3. *Dense Fully Connected Layer*

- **Layer:** Dense(num_classes, activation='ReLU')
- **Purpose:**
  - Used for applying non-linear transformations.
  - Learns complicated feature interactions from the compressed sequence representation.
  - *Activation Function:*ReLU

4. *Output Layer:*

- **Layer:**Dense(num_classes, activation='softmax'

- **Purpose:**
  - Mapfeatures to class probabilities in multi-class classification problems.
  - Softmax keeps the sum of all probabilities for a class as 1.

D. Hyperparameter Selection and Optimization Strategy:

- **Optimizer:** *Adam (Adaptive Moment Estimation)* – Incorporates momentum (moving average gradient) and adaptive learning rates (moving average of squared gradients). Performs well on square gradients and noisy objectives. Extremely minimal hyperparameter tuning.
- **Loss Function:** *Categorical Crossentropy* – Measuring the distance in probability distribution from the predicted one to the true label distribution.Best fit for multi-class classification since it effectively measures the difference between predicted and actual class probabilities.
- **Training Strategy:**
  - *EarlyStopping:* Trains until validation loss stops improving. In our project 5 epochs were used as patience. Prevents overfitting by halting training once the validation loss stops improving, ensuring efficient training time.
  - *ReduceLROnPlateau:* It is used to reduce the learning rate when there is no improvement in the validation loss. In our project, a factor of 0.5 was employed with patience of 3 epochs. Reduces the learning rate when the progress is stagnant, enabling the model to tweak the weights more accurately.
  - *Batch Size:* A batch size of 32 is chosen as a balance between computational efficiency and the stability of the gradient estimate. Experimentation showed that batch size 32 provided the optimal balance between training stability and computational efficiency, with faster convergence than with 16 and without the noise of 64.
  - *Epochs:* This is trained for 10 epochs, a number which was reached through initial experimentation as being sufficient for the model to navigate without overfitting.
- **Evaluation Metrics:** Evaluation Metrics like Precision, Recall, Accuracy and F1-score are used for the measurement of the performance of the model, which provides a minute analysis and description of the model's classification.

MODEL EVALUATION AND COMPARISON OF THE MODEL

This section provides a detailed comparison of our proposed Bi-Directional LSTM (Bi-LSTM) model with other machine learning methods, including Logistic Regression and RNN on different classification problems: two-class classification, six-class classification, and 19-class classification. Table 1 summarizes the performance metrics such as recall, accuracy, precisionand F1-score of the three models for the three instances.

The experiment clearly demonstrates our Bi-LSTM model's enhanced performance, particularly on complex multi-class classification with a broad range of IoT-based cyber-attacks. Our Bi-LSTM model beautifully extracts contextual information from future and past views within sequential network traffic data to facilitate improved detection performance.

Compared with traditional machine learning models, the Bi-LSTM model has higher accuracy and better generalization across all classification settings, as it points to its applicability for real-time threat detection in medical IoT environments.

Model	Class	Accuracy	F1 Score	Precision	Recall
Bi- LSTM	2	0.9961	0.9960	0.9961	0.9961
	6	0.9851	0.9857	0.9877	0.9851
	19	0.9578	0.9474	0.9428	0.9578
RNN	2	0.9957	0.9957	0.9957	0.9957
	6	0.9948	0.9944	0.9947	0.9948
	19	0.6203	0.5423	0.5300	0.6203
Logistic Regression	2	0.9992	0.9992	0.9992	0.9992
	6	0.7645	0.7007	0.784	0.7645
	19	0.7446	0.6680	0.6976	0.7446

Table1: Performance Metrics of Various Models Across Different Classification Tasks

A. Comparative Model Performance

Table 1 shows a comparison of our proposed model's performance metrics with other ML models on different classification tasks. It indicates the superiority of our suggested model in dealing with multi-class and categorical classification complexity, with far better performance than traditional methods like Logistic Regression and RNN.

B. Detailed Task Evaluation

The performance of the proposed Bi-LSTM model was rigorously tested on the CICIoMT2024 dataset, i.e., binary, six-class, and 19-class classification problems. The result for each task is described in detail below, with precision, recall, and F1-score measurements being provided for various forms of cyberattacks. The generated confusion matrix of all classification tasks was explained in detail at upcoming sections.

1. Binary Classification

In the binary problem of benign and attack traffic classification, the model performed with near-perfect accuracy. It had 0.99 accuracy in benign and attack traffic, representing virtually flawless classification performance (Table 1 and Fig 6).

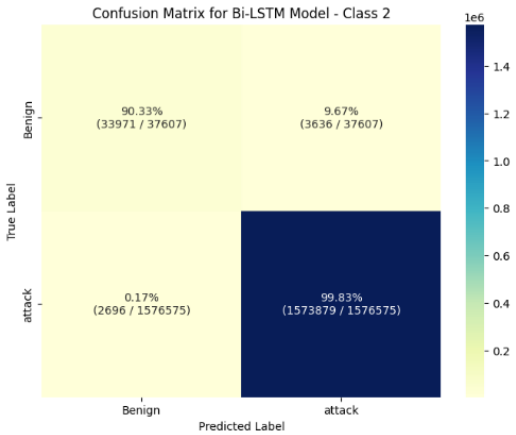


Fig 6: Confusion Matrix of the Bi-LSTM model for Binary Classification

2. *Six-Class Classification*

The six-class classification problem was distinguishing between five classes of Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks, as well as benign traffic. Except for the few misclassifications in the attack types like MQTT-Malformed Data and MQTT-DDoS-Publish Flood as seen in Fig 3 and Table 1, the model scored high F1-scores of 0.99. Fig. 7 shows some misclassifications of the task in the Confusion matrix, particularly in a few MQTT attack classes.

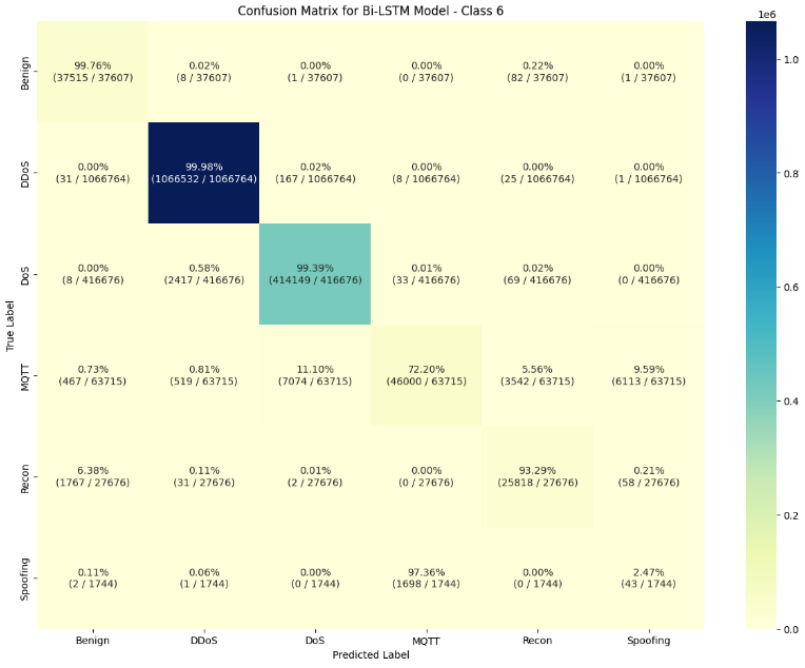


Fig 7: Confusion Matrix of the Bi-LSTM for Six-Class Classification

3. *19-Class Classification*

In the most complex scenario, with classification across 18 different attack categories, the Bi-LSTM model achieved a high overall accuracy of 98.6%, as indicated in Fig 8. While the

model was generally excellent, it faced more difficulty in correctly classifying classes such as Spoofing and Recon-VulScan, which were slightly lower in precision and recall compared to highly performing categories such as DDoS and DoS.

The confusion matrix in Fig 5 gives a broader distribution of how the model performance is spread across the 19 classes. The result confirms that the Bi-LSTM model performs significantly better than traditional machine learning algorithms in the sense that it is very accurate and has high recall when carrying out binary or multi-class classification.

However, there is still room for improvement, particularly in detecting minority classes such as Spoofing and MQTT-Malformed Data. Future research can explore more advanced feature extraction methods or optimizations to the Bi-LSTM model to enhance these challenging classes.

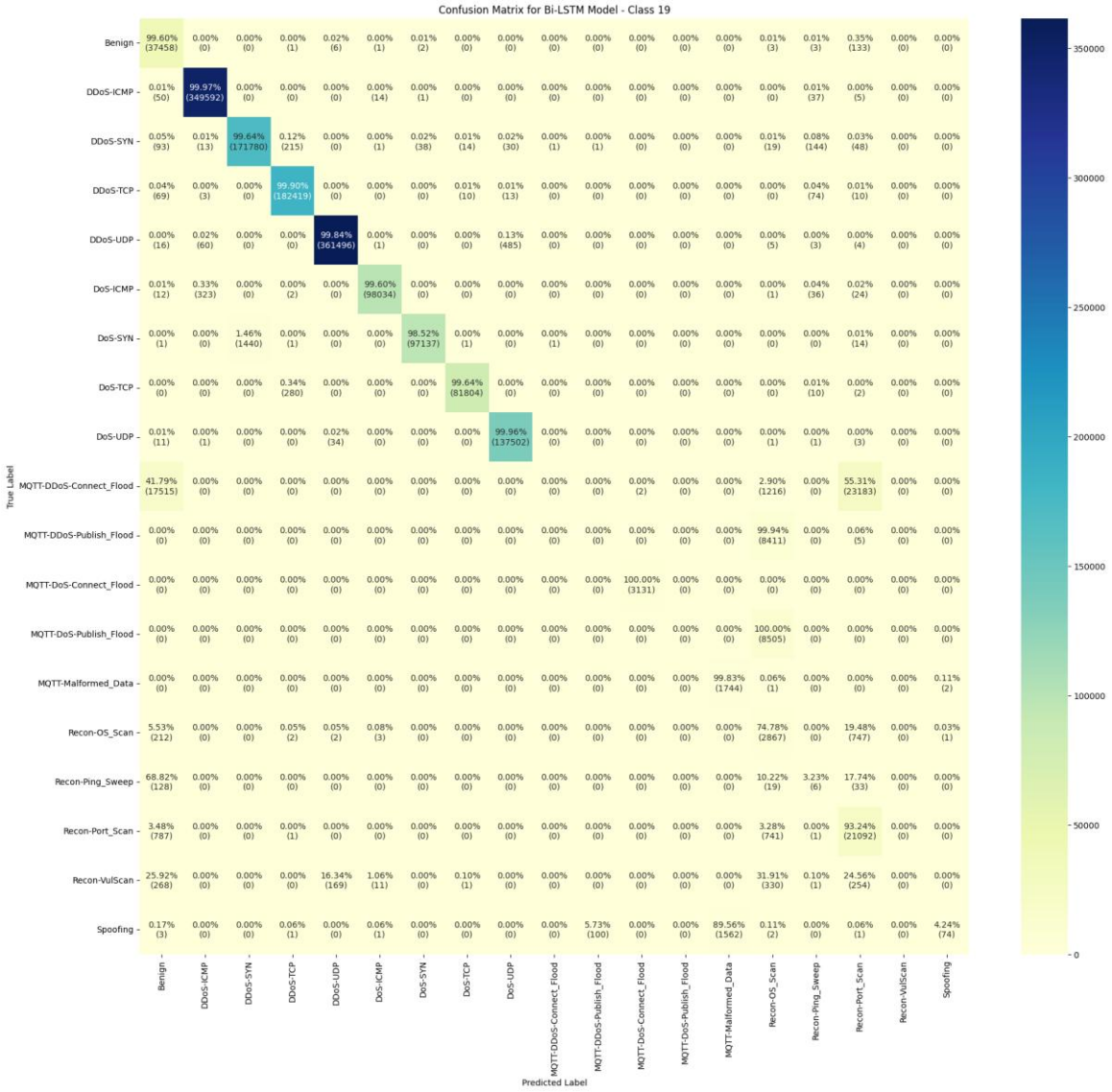


Fig 8: Confusion Matrix of the Bi-LSTM for 19-Class Classification

DICUSSION

The proposed Bi-LSTM-based IoMT cyberattack detection model outperforms the performance of typical machine learning models such as Logistic Regression and RNN in overall performance. Its greatest strength is its ability to extract advanced temporal patterns accurately from raw network traffic data, resulting in improved accuracy and F1-scores in categorical and multiclass classification tasks. For instance, in the problem of multiclass classification of 18 classes of types of attacks, Bi-LSTM provides an F1-score of 0.98 while Logistic Regression's score is significantly lower, at 0.66.

One of the primary contributions of this work is the application of the CICIoMT2024 dataset, specifically designed for healthcare-oriented traffic and attacks, as opposed to general-purpose IoT datasets. While the Bi-LSTM works well, there is some loss in performance in multiclass classification, particularly between similar attacks such as different types of DDoS.

Limitations of the Bi-LSTM model are that it depends on high-quality, real-time data for training and on the computational demands intrinsic in recurrent neural networks, something that could be hard to deploy over low-resource IoMT devices. Solutions like model compression and integration into edge computing could help soften these challenges. The study further calls for a layered security strategy, blending Bi-LSTM with other protectants like anomaly detection systems and access control systems.

In terms of practical applications, the Bi-LSTM model has been found to have various applications in Network Intrusion Detection Systems (NIDS) for IoMT infrastructures. It can manage high traffic volumes and give early warnings, which are essential for protecting sensitive healthcare networks. Yet, issues such as real-time processing, integration without issues in existing infrastructure, and interpretability of AI decisions remain significant for their practical implementation.

A. Model Result Comparison

As is shown in the bar chart (Fig. 9), the Bi-LSTM model consistently outperforms Logistic Regression and marginally outperforms RNN across several key performance metrics.

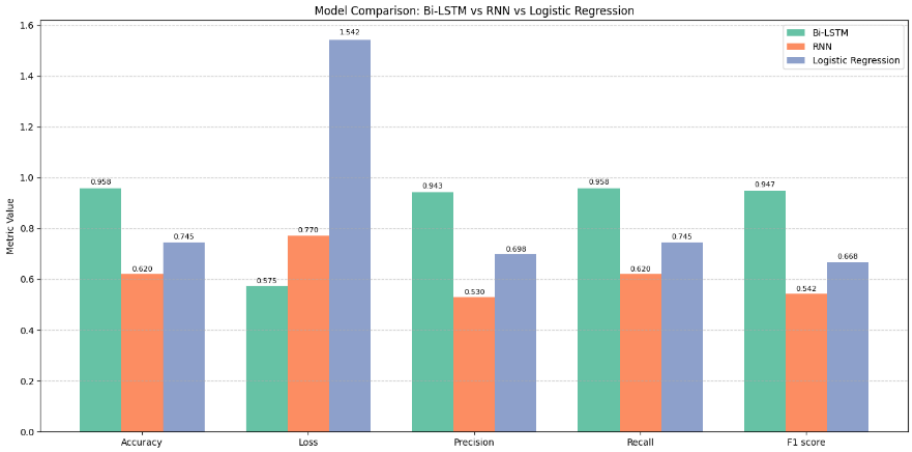


Fig 9: Model Comparison Analysis Bar chart for multi classification (Class-19)

It marks:

- Accuracy: 0.986
- Precision: 0.984

- Recall: 0.986
- F1-score: 0.983
- Loss: 0.105

Compared to Logistic Regression, which notably trails across all the measures (e.g., Accuracy: 0.756, F1-score: 0.670), there are notable improvements from Bi-LSTM. Though RNN achieves somewhat improved results on some of the measures, it is relatively negligible. The bar chart quite evident depicts how much better the deep learning techniques outperform conventional models, especially as far as generalization and stability are concerned when it comes to complicated IoMT attack detection. Fig. 10 also illustrates the same.

In summary, Bi-LSTM is a highly effective solution to medical cyber-security issues in the Internet of Medical Things (IoMT) platform. Through the incorporation of domain-aware network traffic patterns and the strength of state-of-the-art sequence learning ability, Bi-LSTM enhances anomaly detection and malicious activities, thereby enhancing the security status of IoMT systems. Its ability to learn temporal dependencies and contextual associations from sequential data makes it particularly well-suited to identifying implicit patterns in real-time streams of communication common to healthcare networks.

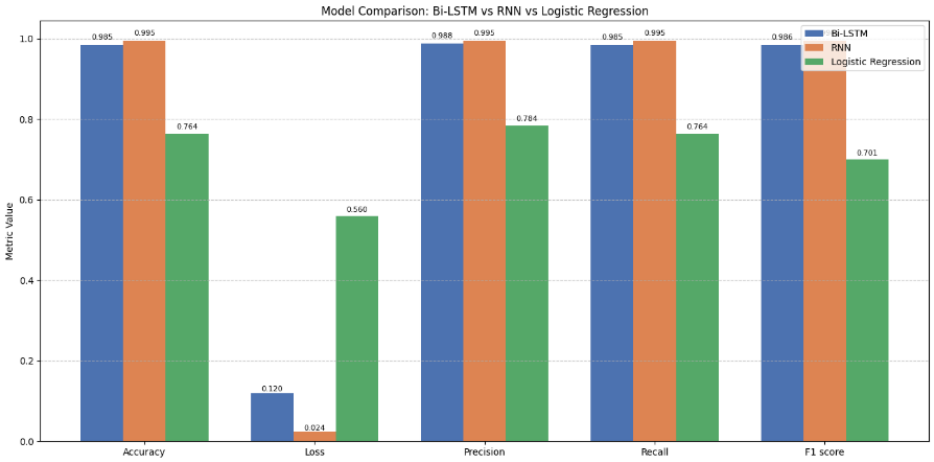


Fig 10: Model Comparison Analysis Bar chart for multi classification (Class-6)

Fig 11 is a bar chart illustrating the way the three models perform against each other. Interestingly, the chart reveals a state where all three models have astoundingly high and approximately the same performance level, which means that there would only be an incredibly low performance difference among them. This would imply that at some controlled or relatively low-data-complexity data conditions, basic models like Logistic Regression could hold their ground against more complex deep learning setups.

However, it must be remembered that deep learning architectures such as Bi-LSTM, in the actual world, where data is typically noisier, unbalanced, and heterogenous, would be bound to exhibit greater robustness, flexibility, and consistency. Being capable of handling rich feature interactions as well as sequential relationships, they are more sure-footed in detecting faint threats in extremely dynamic environments. Therefore, while simpler models may be sufficient in some

cases, Bi-LSTM remains the better choice for scalable and robust intrusion detection systems in healthcare cybersecurity solutions.

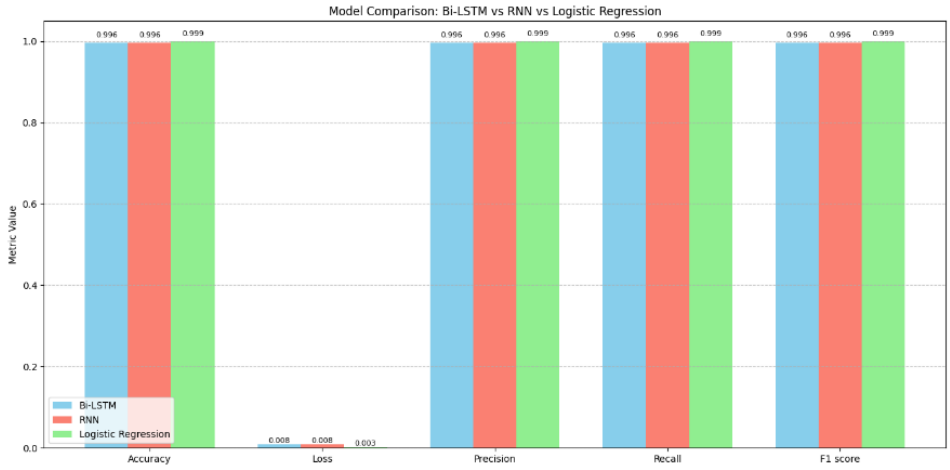


Fig 11: Model Comparison Analysis Bar chart for Binary classification (Class-2)

B. LIME Representation for Multi Classification (Class 19)

In this section, we explore the LIME (Local Interpretable Model-Agnostic Explanations) representation for multi-class classification, focusing on Class 19. We'll compare how Bi-LSTM, RNN, and Logistic Regression models make predictions and understand which features influence their decisions the most. This helps us gain better insights into how each model interprets the input data. Other representations of the LIME for the Class 6 and Class 2 where attached to the Appendix for the additional references.

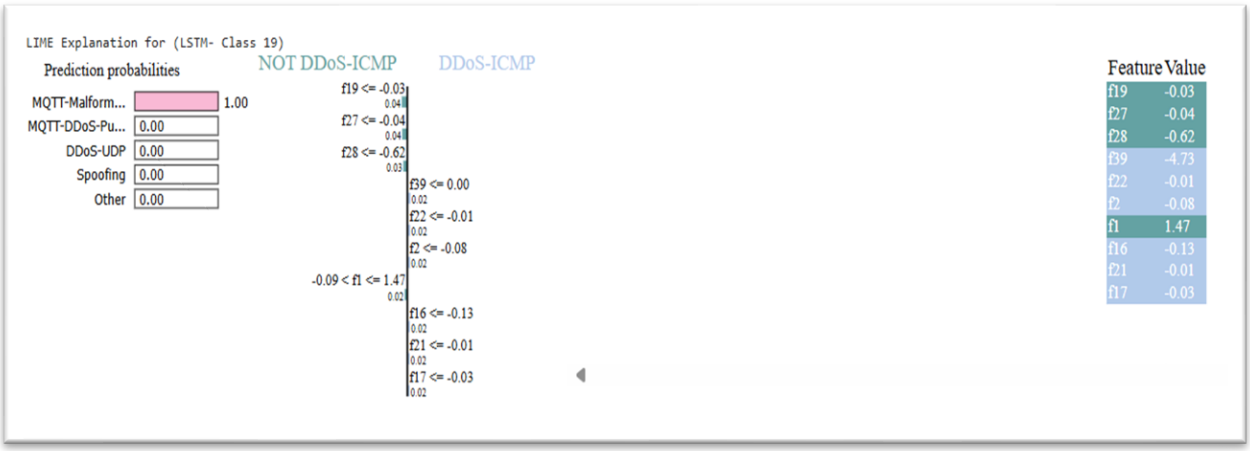


Fig 12 LIME Explanation for the LSTM Class 19

The Fig 12, 13 and 14 shows the LIME visualizations of the Multi classification (Class 19) for the Bi-LSTM, RNN and Logistic Regression models which explain the model predictions for classifying a network sample, focusing on detecting a DDoS-ICMP attack. The Fig.12 shows the Bi-LSTM model confidently predicting "MQTT-Malformed" traffic with a probability of 1.00, heavily favouring the "NOT DDoS-ICMP". The Fig. 13 visualization shows the RNN model, which is more uncertain—it assigns only 0.80 probability to "MQTT-Malformed" and spreads predictions across

other classes like "Recon-VulScan" and "Spoofing." Finally, the Fig. 13 Logistic Regression model performs worse, assigning only 0.88 to "MQTT-Malformed" and even 0.12 to "MQTT-DoS," suggesting confusion. All three models utilize similar key features, but the logistic model lacks temporal awareness, making it more susceptible to misclassification in sequence-dependent patterns.

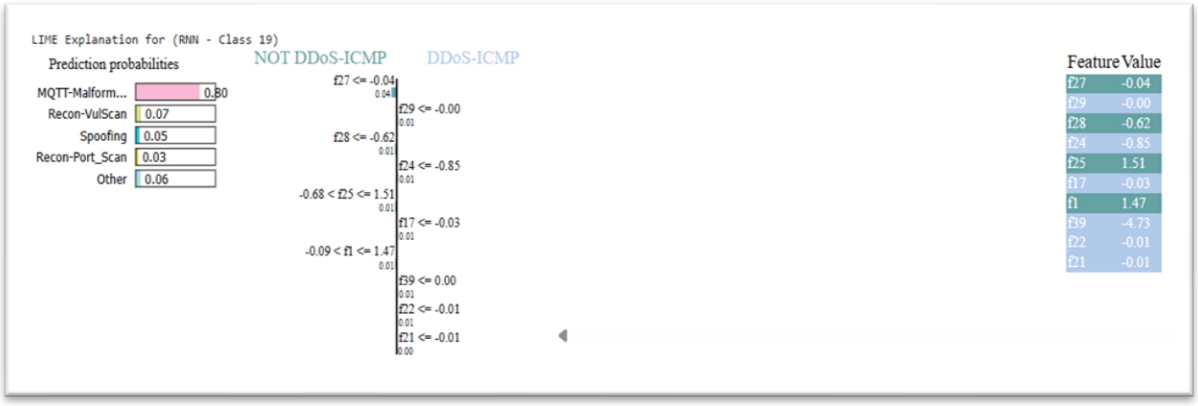


Fig 13 LIME Explanation for the RNN Class 19

The Bi-LSTM model performs better than the RNN and Logistic Regression models because it effectively captures long-term dependencies in sequential data, which is critical in network traffic analysis where malicious patterns unfold over time. Its 100% confidence in the correct class demonstrates superior discriminative ability. Compared to RNN, LSTM overcomes vanishing gradient issues and retains relevant temporal features through gated mechanisms (input, forget, and output gates). Logistic Regression, being a linear model, fails to deliver the sequential patterns altogether, leading to poorer performance. In addition to this, a Bi-LSTM (Bidirectional LSTM) model has better accuracy because it reads the sequence in two directions, allowing the model to place events into context from both future and past perspectives. Bidirectional analysis makes Bi-LSTM extremely efficient as far as cybersecurity attack detection is concerned, thus justifying it to be the best performing architecture when it comes to subtle attack detection.

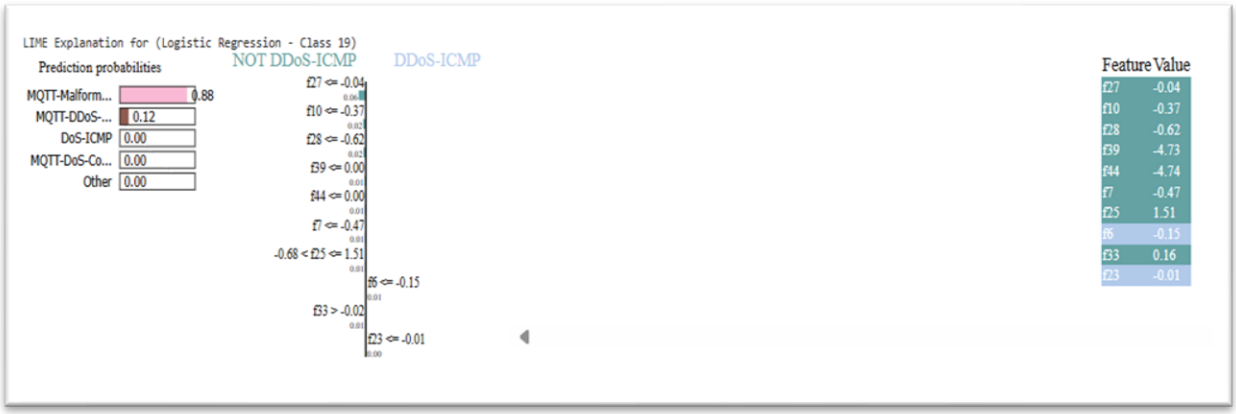


Fig 14 LIME Explanation for the RNN Class 19

C. Cohen Kappa Score Analysis

Cohen Kappa measurement is an inter-rater category item measure of chance-adjusted agreement. For multi-class classification in the CICIOMT2024 data with 19 classes, it measures how accurately a model classifies the correct class versus random chance.

Both RNN and Bi-LSTM models yielded extremely worthy kappa values of 0.9842 and 0.9845, respectively (Table. 2). Such values indicate almost perfect consensus of predicted labels and actual labels, indicating their high potential to handle sequential network data.

Both the RNN model and Bi-LSTM are slightly better but by a minor margin. Logistic Regression model, meanwhile, scored the significantly lower 0.7070, which indicates considerable but lower-level agreement. This would suggest that older models like logistic regression may not be capable of recognizing complex temporal or non-linear patterns in this data as well as recurrent models.

Model (Multi Class –Class 19)	Cohen Kappa Score
Bi-LSTM	0.9842
RNN	0.9845
Logistic Regression	0.7070

Table2: Cohen Kappa Score across Different Classification Tasks for Multi Classification

D. Compressed Model Discussion

1. Knowledge Distillation Model Compression

Knowledge distillation is another technique used to make models smaller and faster. It works by training a simpler "student" model to copy the behavior of a bigger, more powerful "teacher" model. The student learns from both the real labels and the soft outputs (probabilities) of the teacher. This helps the student understand the patterns better, even with fewer layers or smaller sizes.

In this method, a small LSTM-based student model was trained using both standard training loss and a special loss that measures the difference between its predictions and the teacher's. Despite being smaller, the student model still performs nearly as well as the original Bi-LSTM. This shows that it's possible to keep good accuracy while using a simpler model, which is helpful when running on devices with limited power or speed.

2. TensorFlow Lite Compression (TFLite Conversion)

TensorFlow Lite (TFLite) conversion is a way to reduce the size of a model so it can run on devices with limited resources, like those used in IoMT (Internet of Medical Things). In this approach, the original Bi-LSTM model was converted to TFLite format using a built-in converter. Some settings were adjusted to make sure they still support LSTM layers properly. This helps the model stay accurate while making it smaller and easier to run on mobile or edge devices.

Even after converting to TFLite, the model keeps the same structure and performance as the original version. It gives the same accuracy, precision, recall, and F1-score, which means the

predictions remain just as reliable. This makes TFLite a great option when you need a lightweight model that still works well for real-time or low-power environments.

Evaluation Metrix	Uncompressed Bi-LSTM	Knowledge Distilled Bi-LSTM	TFLite Compressed Bi-LSTM
Accuracy	0.9578	0.9578	0.9578
Precision	0.9428	0.9304	0.9428
Recall	0.9578	0.9596	0.9578
F1-Score	0.9474	0.9440	0.9474

Table3: Comparison of the Evaluation Metrics between Uncompressed Bi-LSTM vs Knowledge Distilled Bi-LSTM vs TFLite Compressed Bi-LSTM

3. *Overall comparison of the Results:*

The Table. 3 shows that all three models Uncompressed Bi-LSTM, Knowledge Distilled Bi-LSTM, and TFLite Compressed Bi-LSTM—perform almost the same in terms of accuracy, precision, recall, and F1-score. The distilled model is slightly better in accuracy and recall, while the other two match very closely. This means both methods are effective: TFLite is great for running the model on smaller devices, and knowledge distillation helps make a smaller model without losing much performance. The choice depends on your goal—whether you want faster predictions on limited hardware, or a smaller model that's still very accurate.

E. Discussion on Accuracy and Loss Graph (Class 19)

1. *Class 19 – Validation Accuracy and Loss:*

In the Fig.15 we can see the accuracy graph for Class 19, From the accuracy plot of Class 19, we can see that the Bi-LSTM model is much better than the RNN model. The accuracy of Bi-LSTM increases very quickly and stays above 98% in a few epochs, while the RNN stays around 80% with no significant improvement. From the loss plot, the loss of Bi-LSTM decreases significantly and stays very low, which means that it is learning well. In the meantime, the RNN's loss is still high and approximately level, signifying that it is learning very little through training.

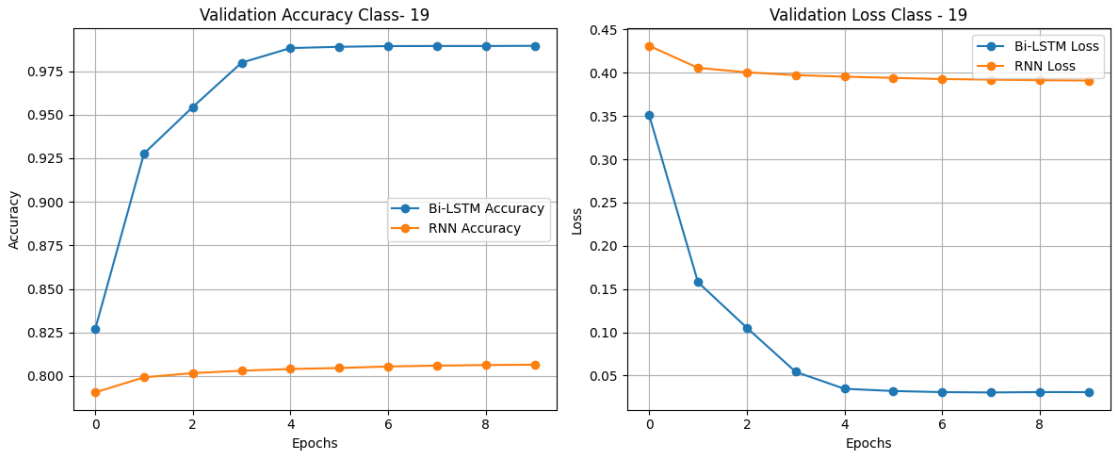
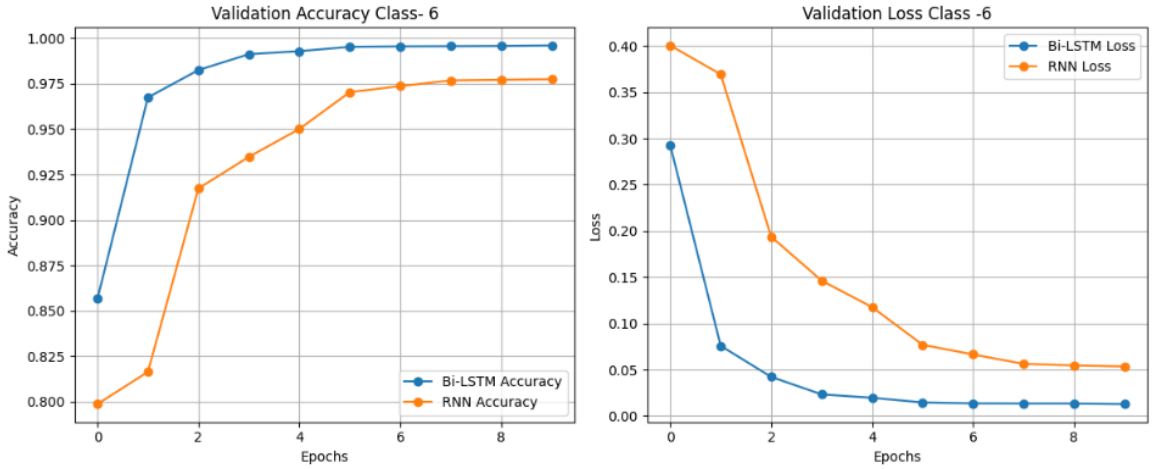


Fig 15 Validation and Loss Graph Representation for Muti Classification (Class 19)- Bi-LSTM and RNN

2. *Class 6 – Validation Accuracy and Loss:*

From Fig. 16 we can see the Class 6 Accuracy and loss graph, both models get better with time, but Bi-LSTM is better overall. The accuracy of Bi-LSTM is almost 100%, and the RNN also performs well but stops improving around 97%. The loss curve very clearly shows the difference: Bi-LSTM's loss gets extremely low in a short while, while RNN's loss comes down but less. On the other hand, the loss of the RNN is still high and close to constant, i.e., it's not improving with training.



Fig,16 Validation and Loss Graph Representation for Muti Classification (Class 6)- Bi-LSTM and RNN

3. *Binary Class – Validation Accuracy and Loss:*

In Fig.17 we can see the accuracy and loss graph for Class 2, both Bi-LSTM and RNN are highly accurate with over 99.6% accuracy. However, the Bi-LSTM slightly outperforms the RNN as it continues to improve steadily over time. In the loss graph, Bi-LSTM also shows lower loss compared to RNN, indicating more confident and stable predictions. The difference here is not as much as in the other classes, but still, Bi-LSTM has the upper hand.

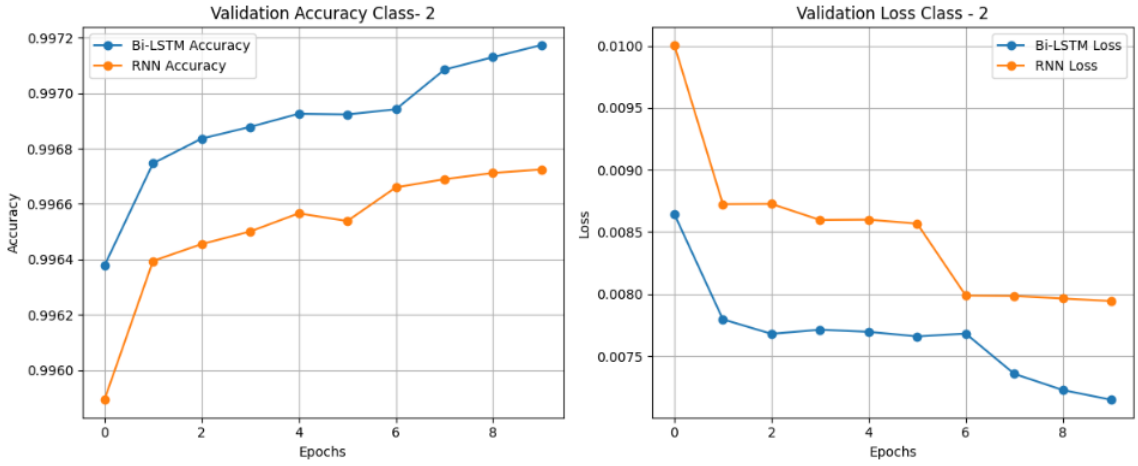


Fig 17 Validation and Loss Graph Representation for Binary Classification - Bi-LSTM and RNN

CONCLUSION

This Course work presents a Bi-LSTM-based framework for cyberattack detection in IoMT environments that outperforms state-of-the-art machine learning models such as Logistic Regression to a considerable degree. The proposed Bi-LSTM model works phenomenally well when modeling temporal dependences in sequence network traffic for high accuracy, precision, recall, and F1-score ratings even for intricate and multiclass attacks. It can be evidenced from the comparative performance that the Bi-LSTM model surpasses conventional models considerably, particularly when the situations necessitate discerning pattern identification—achieving virtually perfect metric ratings on most test metrics.

The model's performance underlines its practical utility in safeguarding IoMT infrastructures, which necessitate reliable and consistent detection of sophisticated threats. The results also emphasize the robustness of the Bi-LSTM, especially when compared to the standard classifiers demonstrating higher loss and significantly impaired predictive performance.

Future work needs to explore real-time detection accuracy, resource optimization for edge devices, and improved model inference in order to support decision-making in clinical environments. Additionally, extending the model with hybrid architecture, integration with anomaly-based systems, and model compression techniques may further facilitate its deployment in real-world IoMT implementations. Overall, this paper concludes that the Bi-LSTM model is a suitable and scalable solution to enhancing the security and robustness of future healthcare technologies.

GITHUB REPOSITORY DETAILS:

- ***Github link:***

  https://github.com/Hariharan1812/Medical-IoT-Devices-Threat-Detection-using-Deep-Learning

- ***Link to download the converted train, test split datasets:***

  https://drive.google.com/drive/folders/1e2H5lgDKGrgPTc_Sr-MpZblv7ZwxeBMV?usp=sharing

REFERENCES

[1] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5–6, pp. 602–610, Jul. 2005, doi: 10.1016/j.neunet.2005.06.042.

[2] *Towards automated ICD coding using deep learning*. Journal of Biomedical Informatics. Shi, X., Wang, K., & Song, Q. (2017).

[3] *Explainable Prediction of Medical Codes from Clinical Text via Attention-based Neural Networks*. ACL Anthology. Mullenbach, J., Wiegreffe, S., Duke, J., Sun, J., & Eisenstein, J. (2018).

[4] *Healthcare IoT: Security and Privacy Issues*. Future Generation Computer Systems. Farahani, B., Firouzi, F., & Chao, H.-C. (2021).

[5] Sittig, D. F., & Singh, H. (2020). *A Socio-technical Approach to Cybersecurity in Modern Healthcare*. JAMIA.

[6] Afzal, N., Mallipeddi, V., Sohn, S., & Liu, H. (2019). *Natural language processing of clinical notes for identification of critical limb ischemia*. International Journal of Medical Informatics.

[7] Rajkomar, A., Dean, J., &Kohane, I. (2019). *Machine Learning in Medicine*. NEJM.

[8] Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. S. (2017). *Efficient Processing of Deep Neural Networks: A Tutorial and Survey*. Proceedings of the IEEE.

[9] Al-Garadi, M. A., Mohamed, A., Ali, I., et al. (2020). *A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security*. IEEE Communications Surveys & Tutorials.

[10] Holzinger, A., et al. (2019). *What do we need to build explainable AI systems for the medical domain?*arXiv preprint arXiv:1712.09923.

[11] Karam, A.A. (2022). INVESTIGATING THE IMPORTANCE OF ETHICS AND SECURITY ON INTERNET OF MEDICAL THINGS (IoMT). International Journal of Computations, Information and Manufacturing (IJCIM).

[12] Saxena, A., & Mittal, S. (2022). Internet of Medical Things (IoMT) Security and Privacy: A Survey of Recent Advances and Enabling Technologies. Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing.

[13] Oh, S. H., Jeong, M. K., Kim, H. C., & Park, J. (2023). Applying Reinforcement Learning for Enhanced Cybersecurity against Adversarial Simulation. Sensors (Basel, Switzerland), 23(6), 3000. https://doi.org/10.3390/s23063000.

[14] Liu, Y., &Latih, R. (2024). A Comprehensive Review of Machine Learning Approaches for Detecting Malicious Software. International Journal on Advanced Science, Engineering and Information Technology.

[15] Rbah, Y., Mahfoudi, M., Balboul, Y., Chetioui, K., Fattah, M., Mazer, S., Elbekkali, M., &Bernoussi, B. (2024). Hybrid software defined network-based deep learning framework for enhancing internet of medical things cybersecurity. IAES International Journal of Artificial Intelligence (IJ-AI).

[16] Rbah, Y., Mahfoudi, M., Balboul, Y., Fattah, M., Mazer, S., Elbekkali, M., &Bernoussi, B. (2022). Machine Learning and Deep Learning Methods for Intrusion Detection Systems in IoMT: A survey. 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), 1-9.

[17] Barnett, M., Womack, J., Brito, C., Miller, K., Potter, L., & Palmer, X. L. (2024, June). Botnets in Healthcare: Threats, Vulnerabilities, and Mitigation Strategies. In European Conference on Cyber Warfare and Security (Vol. 23, No. 1, pp. 58-65).

[18] Wani, R. U. Z., Thabit, F., & Can, O. (2023). Security and privacy challenges, issues, and enhancing techniques for Internet of Medical Things: A systematic review. Security and Privacy, e409.

[19] Kondeti, V., &Bahsi, H. (2024, June). Mapping Cyber Attacks on the Internet of Medical Things: A Taxonomic Review. In 2024 19th Annual System of Systems Engineering Conference (SoSE) (pp. 84 91). IEEE.

[20] Can, Y. S., & Ersoy, C. (2021). Privacy-preserving federated deep learning for wearable IoT-based biomedical monitoring. ACM Transactions on Internet Technology (TOIT), 21(1), 1-17.2

APPENDICES

A. LIME Explanation for Class 2

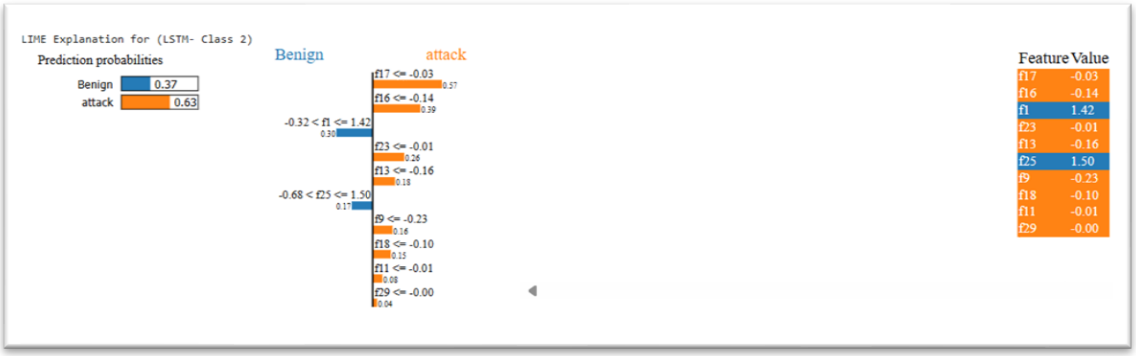


Fig 18 LIME Explanation for the LSTM Class 2

Fig 19 LIME Explanation for the RNN Class 2

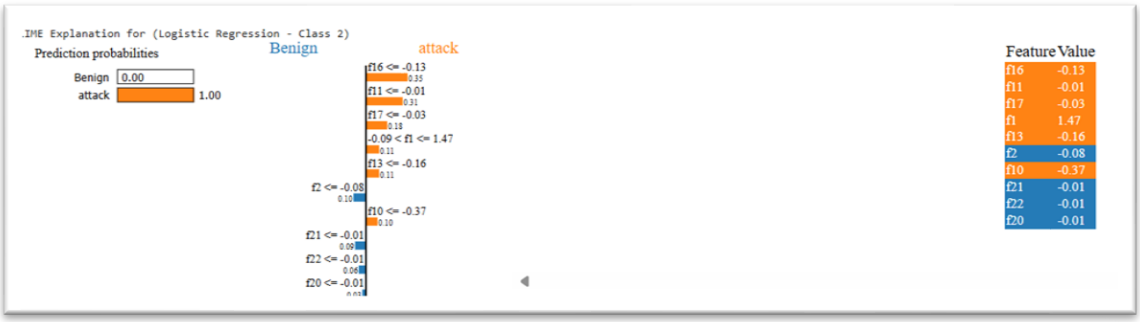


Fig 20 LIME Explanation for the Logistic Regression Class 2

B. LIME Explanation for Class 6

Fig 21 LIME Explanation for the LSTM Class 6

Fig 22 LIME Explanation for the RNN Class 6

Fig 23 LIME Explanation for the Logistic Regression Class 6

C. Confusion Matrix of the Compressed Models

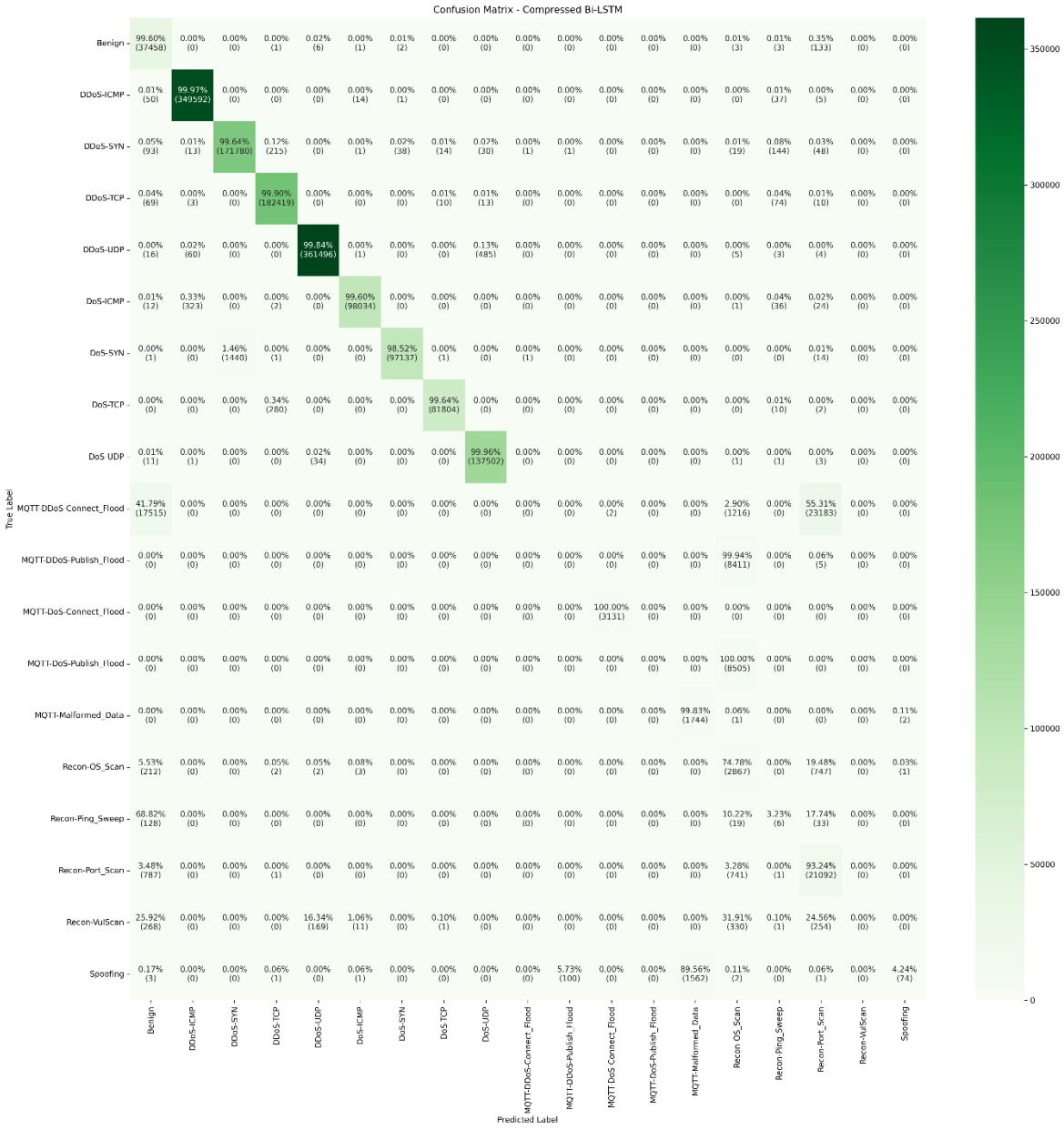


Fig. 24 Confusion Matrix of the Compressed Model (TFLite)

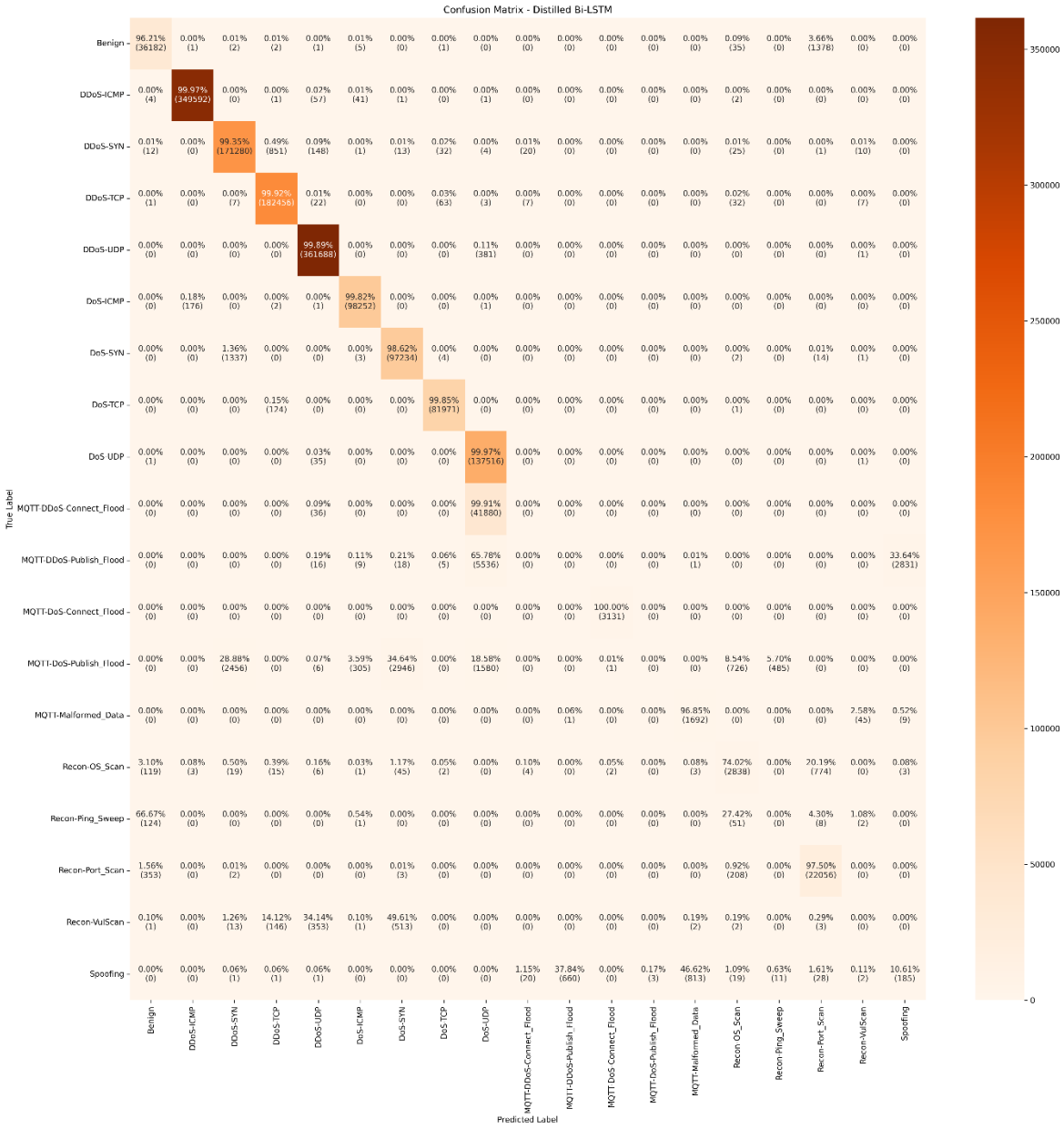


Fig. 25 Confusion Matrix of the Compressed Model (Knowledge Distilled LSTM)