JOURNAL
COMPUTER SCIENCES INSTITUTE

JCSI 36 (2025) 232–242
Received: 2 November 2024
Accepted: 30 April 2025

Classification of Cyber Attacks in IoMT Networks Using Deep Learning: A Comparative Study

Klasyfikacja ataków cybernetycznych w sieciach IoMT z wykorzystaniem głębokiego uczenia się: badanie porównawcze

Asif Rahman Rumee*

Department of Computer Science and Engineering, Jashore University of Science and Technology, Jashore-7408, Bangladesh

Abstract
The Internet of Medical Things (IoMT) is transforming healthcare through enhanced remote monitoring and real-time data exchange, but it also introduces significant cybersecurity challenges. This study evaluates various deep learning architectures - Feedforward Neural Networks (FNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), Long Short-Term Memory networks (LSTM), and Bi-LSTM - for classifying cyber-attacks in IoMT networks. Utilizing the ECU-IoHT dataset, the Bi-LSTM and CNN models demonstrated superior performance, achieving 60% and 60% accuracy, 75% and 86% precision, 60% and 60% recall, and 63% and 61% F1-score, respectively. These results highlight the effectiveness of Bi-LSTM and CNN in enhancing cybersecurity measures within the IoMT, underscoring their potential to safeguard connected medical devices.

Keywords: cyber-attack classification; deep learning; RNN; Bi-LSTM

Streszczenie
Internet Rzeczy Medycznych (IoMT) rewolucjonizuje opiekę zdrowotną dzięki zaawansowanemu zdalnemu monitorowaniu i wymianie danych w czasie rzeczywistym, ale jednocześnie wprowadza istotne wyzwania związane z cyberbezpieczeństwem. W niniejszym badaniu oceniono różne architektury głębokiego uczenia - sieci neuronowe z propagacją wsteczną (FNN), sieci konwolucyjne (CNN), sieci rekurencyjne (RNN), jednostki rekurencyjne z bramkowaniem (GRU), sieci z długą i krótką pamięcią (LSTM) oraz dwukierunkowe LSTM (Bi-LSTM) - pod kątem klasyfikacji cyberataków w sieciach IoMT. Wykorzystując zbiór danych ECU-IoHT, modele Bi-LSTM i CNN wykazały lepszą wydajność, osiągając odpowiednio 60% i 60% dokładności, 75% i 86% precyzji, 60% i 60% odtworzenia oraz 63% i 61% wyniku F1. Wyniki te podkreślają skuteczność Bi-LSTM i CNN w zwiększaniu środków cyberbezpieczeństwa w ramach IoMT, podkreślając ich potencjał w zakresie ochrony podłączonych urządzeń medycznych.

Słowa kluczowe: klasyfikacja cyberataków; głębokie uczenie; RNN; Bi-LSTM

*Corresponding author
Email address: arrumee@gmail.com (A. R. Rumee)

Published under Creative Common License (CC BY 4.0 Int.)

1. Introduction

1.1. Literature Review

The Internet of Medical Things (IoMT) is rapidly transforming healthcare [1], enabling capabilities such as remote patient monitoring, real-time diagnostics, and enhanced data interoperability. IoMT devices range from wearable health trackers to sophisticated hospital equipment interconnected via cloud-based platforms, improving patient outcomes and operational efficiencies [2]. However, this technological evolution also introduces significant cybersecurity challenges, as interconnected medical devices become prime targets for cyber-attacks [3]. These attacks threaten not only sensitive patient data but also the reliability and functionality of critical healthcare services, necessitating robust security frameworks tailored to IoMT-specific vulnerabilities [4].

A comprehensive understanding of the cyber threats targeting IoMT systems is essential for developing effective defense mechanisms. Djenna and Saïdouni [4] classified various types of cyber-attacks specific to IoT-based healthcare systems, highlighting prevalent threats such as distributed denial-of-service (DDoS) attacks, ransomware, and data breaches. Their findings underscored the need for advanced security solutions to mitigate evolving attack strategies. Similarly, Adeniyi et al. [5] provided an in-depth review of attack vectors and machine learning (ML) techniques for classifying cyber threats in IoMT. They emphasized the increasing sophistication of cyber-attacks and the necessity of adaptive security systems capable of evolving with the dynamic threat landscape.

Deep learning (DL) has emerged as a promising approach to enhance cybersecurity in IoMT networks. Abbas et al. [6] evaluated multiple deep learning models for cyber-attack detection in IoT environments, demonstrating their effectiveness in achieving high detection accuracy and precision. Arnob et al. [7] further explored neural networks for cyber-attack classification, emphasizing their potential in improving IoMT security. Additionally, Alrefaei and Ilyas [8] introduced an ensemble

deep learning model for multi-class cyber-attack classification, showcasing the advantages of hybrid approaches that integrate multiple algorithms to optimize detection rates.

Recent advancements in cybersecurity research have also explored novel techniques to enhance IoMT security. For example, federated learning, a decentralized ML approach, has been proposed as a means to improve privacy-preserving cyber threat detection in IoMT networks [10]. By training models locally on IoMT devices and aggregating updates at a central server, federated learning reduces the risk of exposing sensitive medical data during transmission. Furthermore, blockchain technology has been investigated for securing IoMT communications, providing immutable and transparent audit trails that can prevent unauthorized data tampering and mitigate insider threats [11].

Despite these advancements, challenges remain in fully securing IoMT networks. One key limitation is the lack of comprehensive evaluations comparing different deep learning architectures for cyber-attack detection. Existing studies often focus on individual models [9] without systematically analyzing their performance across various metrics such as latency, false positive rates, and adaptability to novel attack patterns. Moreover, the integration of artificial intelligence (AI)-driven security mechanisms within resource-constrained IoMT devices poses computational and energy efficiency challenges [12]. Addressing these issues requires the development of lightweight AI models optimized for real-time threat detection and response.

Another critical aspect of IoMT cybersecurity is regulatory compliance and standardization. With diverse medical devices operating within heterogeneous environments, establishing unified security protocols is essential to ensure interoperability and protection against cyber threats [13]. Current regulatory frameworks, such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR), mandate stringent data protection measures; however, their enforcement within IoMT ecosystems remains complex due to the evolving nature of cyber threats and technological advancements [14].

Overall, while significant progress has been made in leveraging machine learning and deep learning for IoMT cybersecurity, gaps remain in systematically evaluating their effectiveness across diverse attack scenarios. Future research should focus on developing adaptive and lightweight security models, integrating blockchain for data integrity, and establishing standardized cybersecurity protocols tailored to IoMT environments. By addressing these challenges, the healthcare industry can enhance the security and resilience of IoMT systems, ensuring reliable and safe healthcare services.

1.2. Purpose of the Research

This research aims to conduct a thorough comparative analysis of different deep learning architectures - specifically Feedforward Neural Networks (FNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), Long Short-Term Memory networks (LSTM), and Bidirectional LSTM (Bi-LSTM) - for classifying cyber attacks in IoMT networks. By utilizing the ECU-IoHT dataset, this study will evaluate the performance of each model based on key metrics, including accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC curve.

The novelty of this research lies in the systematic and comprehensive comparison of multiple deep learning architectures for IoMT cybersecurity, which has not been extensively explored in existing literature. While previous studies have primarily focused on individual models [9] or small-scale comparisons, this study uniquely evaluates a diverse range of architectures under a unified experimental framework. Additionally, the research incorporates advanced performance metrics and real-world attack scenarios from the ECU-IoHT dataset to ensure practical relevance.

Furthermore, this study aims to bridge the gap between theoretical advancements and real-world cybersecurity applications by identifying the most effective deep learning model for IoMT security. The insights gained will contribute to the development of more robust and adaptable cybersecurity measures tailored to the unique challenges of IoMT environments.

1.3. Research Areas and Hypothesis

The research will focus on the following key areas:

- **Cybersecurity in IoMT**: Understanding the vulnerabilities and attack vectors specific to interconnected medical devices.
- **Deep Learning Techniques**: Analyzing various architectures and their applicability for cyber attack classification.
- **Performance Metrics**: Evaluating models based on established metrics to determine their effectiveness.

The central hypothesis of this study is that Bidirectional LSTM (Bi-LSTM) will outperform other deep learning models in terms of accuracy, precision, recall, and F1-score for classifying cyber attacks within IoMT networks. This hypothesis is grounded in existing literature that suggests recurrent and bidirectional architectures have advantages in capturing sequential patterns in data, which may be particularly beneficial for detecting complex cyber threats.

In essence, this research aims to advance the understanding of cyber attack classification in IoMT by providing a systematic evaluation of various deep learning approaches, ultimately contributing to improved cybersecurity frameworks in healthcare environments. The subsequent sections will outline the methodology employed in this study, present the results of the comparative analysis, and discuss the implications of the findings for both academic research and practical applications in IoMT security.

2. Materials and Methods

2.1. Research Object

The primary focus of this study is the classification of cyber attacks within Internet of Medical Things (IoMT) networks, leveraging various deep learning architectures. The research object includes both the datasets and the models employed for analysis. The ECU-IoHT dataset, a representative collection of cyber attack scenarios specifically designed for IoMT environments, serves as the foundational dataset. This dataset comprises various attack types, including denial of service, smurf attack, arp spoofing, and nmap port scan, all critical threats to connected medical devices. The dataset is structured to provide a comprehensive view of attack patterns, which is essential for training and validating the deep learning models.

2.2. Test Method

This study adopts a **comparative research design**. The objective is to systematically evaluate the performance of different deep learning architectures—Feedforward Neural Networks (FNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), Long Short-Term Memory networks (LSTM), and Bidirectional LSTM (Bi-LSTM)—in the context of cyber attack classification. The models will be assessed based on multiple performance metrics, including accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC curve.

The research methodology comprises the following key steps:

1. **Data Preprocessing**: Preparing the ECU-IoHT dataset for training and testing, which includes data normalization, handling missing values, and encoding categorical variables.
2. **Model Training**: Training each deep learning architecture on the preprocessed dataset, employing appropriate hyperparameter tuning to optimize performance.
3. **Model Evaluation**: Assessing the trained models on a separate test set to gauge their classification capabilities against unseen data.
4. **Comparative Analysis**: Analyzing and comparing the results from each model using the defined performance metrics to draw conclusions about their effectiveness in classifying cyber-attacks.

2.3. Conducted Research

2.3.1. Data Collection

The ECU-IoHT dataset is publicly available and provides a rich set of features that simulate real-world cyber attack scenarios in IoMT environments. The dataset contains multiple instances of both normal and attack traffic, classified into distinct categories. Each instance includes features such as timestamps, source and destination IP addresses, packet sizes, and payload characteristics, all of which are crucial for accurately modeling and predicting attack behaviors.

2.3.2. Data Preprocessing

Prior to model training, the data undergoes several preprocessing steps:

- **Normalization**: StandardScaler was used to standardize the features by removing the mean and scaling them to unit variance, resulting in a distribution with a mean of 0 and a standard deviation of 1. This normalization is essential for ensuring that all features contribute equally to model training. As a result, the models are better positioned to learn and generalize from the data, improving classification accuracy.
- **Encoding**: Categorical variables are converted into numerical format using label encoding, allowing them to be incorporated into the models.
- **Data Splitting**: The dataset is divided into training (80%), and testing (20%) subsets to ensure that models are trained and evaluated on distinct sets of data, minimizing overfitting.

2.3.3. Model Development

Each deep learning architecture is implemented using Python, PyTorch, and Scikit-learn libraries. The architecture for each model is as follows:

- **Feedforward Neural Network (FNN)**: A simple neural network architecture [15] with an input layer, two hidden layers (64 and 32 neurons respectively) utilizing the SELU activation function, and an output layer designed for multi-class classification, employing CrossEntropyLoss as the loss function and the Adam optimizer for training.
- **Convolutional Neural Network (CNN)**: This model utilizes convolutional layers to extract features from the input data [16], featuring two convolutional layers followed by a pooling layer to reduce dimensionality, and finishing with two fully connected layers. It uses CrossEntropyLoss as the loss function and the Adam optimizer for training.
- **Recurrent Neural Network (RNN)**: A sequential model designed to process time-series data, capturing temporal dependencies in the dataset [17]. The RNN architecture features an input layer with seven features per time step, a single hidden RNN layer with 64 units, and an output layer consisting of one neuron. It employs CrossEntropyLoss as the loss function and the Adam optimizer for training.
- **Gated Recurrent Units (GRU)**: An enhancement over RNN, GRUs utilize gating mechanisms to control information flow, making them more effective for longer sequences [18]. The GRU model consists of an input layer with seven features for each time step, one GRU layers each with 64 units, and an output layer with 5 neurons. It utilizes CrossEntropyLoss as the loss function and the Adam optimizer for training.
- **Long Short-Term Memory networks (LSTM)**: Similar to GRUs but with a more complex architecture that includes input, forget, and output gates, LSTMs are capable of capturing long-range depend-

encies in data [19]. The LSTM model is structured with an input layer containing seven features for each time step, a single LSTM layer with 64 units, and an output layer featuring five neurons. It employs CrossEntropyLoss as the loss function and the Adam optimizer for training.

- **Bidirectional LSTM (Bi-LSTM)**: This architecture processes the input data in both forward and backward directions, allowing the model to leverage context from both past and future states, enhancing its ability to classify attacks effectively [20]. The Bi-LSTM model is structured with an input layer that includes seven features for each time step, one LSTM layer with 64 units, and an output layer consisting of five neurons. It uses CrossEntropyLoss as the loss function and the Adam optimizer for training.

2.3.4. Hyperparameter Tuning

Hyperparameter tuning is conducted using techniques such as grid search and random search to find the optimal settings for each model. Key hyperparameters include the number of hidden layers, number of neurons in each layer, learning rate, and batch size.

2.3.5. Model Evaluation

Once trained, each model is evaluated using the same test set. The evaluation metrics used to assess model performance include:

- **Accuracy**: The ratio of correctly predicted instances to the total instances.
- **Precision**: The ratio of true positive predictions to the total positive predictions made by the model.
- **Recall**: The ratio of true positive predictions to the total actual positives.
- **F1-score**: The harmonic mean of precision and recall, providing a single score that balances both metrics.
- **Confusion Matrix**: A matrix that summarizes the performance of the classification model, showing the true vs. predicted classifications.
- **ROC-AUC Curve**: A graphical representation of the model's ability to discriminate between classes across different threshold values.
- **Precision-Recall Curve**: A graphical representation of the trade-off between precision and recall for different threshold values, useful in imbalanced datasets where the positive class is of primary interest.

2.4. Summary

This section outlines the materials and methods used in this comparative study of deep learning architectures for classifying cyber attacks in IoMT networks. Through careful selection of the ECU-IoHT dataset, a systematic approach to data preprocessing, and the development of multiple deep learning models, this research aims to identify the most effective techniques for enhancing cybersecurity in connected healthcare environments. The subsequent sections will present the results of the comparative analysis and discuss the implications of the findings for both academia and industry.

3. Results

3.1. Overview of Model Performance

The comparative analysis of deep learning architectures - Feedforward Neural Networks (FNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), Long Short-Term Memory networks (LSTM), and Bidirectional LSTM (Bi-LSTM) - was conducted to assess their effectiveness in classifying cyber attacks in IoMT networks using the ECU-IoHT dataset. Each model was trained and evaluated based on accuracy, precision, recall, F1-score, confusion matrix, ROC-AUC curve, and Precision-Recall Curve. The results are summarized in Figure 1 and Table 1.

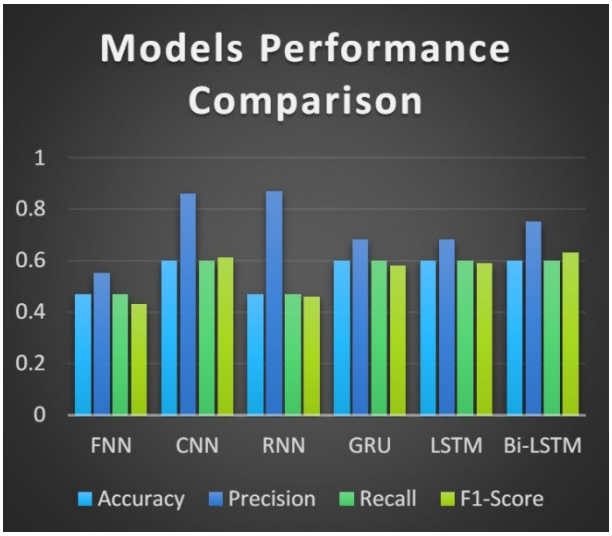

Figure 1: Comparison of Model Performance

Table 1: Summary of Model Performance Metrics

Metric	FNN	CNN	RNN	GRU	LSTM	Bi-LSTM
Accuracy	0.47	0.60	0.47	0.60	0.60	0.60
Precision	0.55	0.86	0.87	0.68	0.68	0.75
Recall	0.47	0.60	0.47	0.60	0.60	0.60
F1-Score	0.43	0.61	0.46	0.58	0.59	0.63

3.2. Detailed Performance Analysis

Before analyzing individual model performance, it is essential to outline the characteristics of the classifier, input data, and the features used for assessment.

3.2.1.Classifier and Input Data

The assessed deep learning models—Feedforward Neural Network (FNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Gated Recurrent Units (GRU), Long Short-Term Memory Networks (LSTM), and Bidirectional LSTM (Bi-LSTM)—were trained and tested using the **ECU-IoHT dataset**,

a publicly available benchmark dataset for cyber attack classification in IoMT networks.

The dataset consists of labeled network traffic data collected from IoMT devices, including normal and attack traffic. Each record represents a network packet or session with multiple extracted features. The **classification task** involved distinguishing between **benign** and **malicious** traffic and categorizing attacks into specific types.

3.2.2.Input Features and Data Format

The dataset includes a mix of **temporal, and protocol-based features**, such as:

- **Network Traffic Features**: Packet size, flow duration, protocol type, and timestamp
- **Behavioral Features**: Frequency of requests, connection states
- **Security Indicators**: Presence of known attack signatures

Each record in the dataset is structured as a feature vector, with numerical values normalized to improve model convergence. The input format consists of a structured dataset with 111207 samples × 9 features, where sample represents the number of records feature represents the number of extracted attributes.

The dataset was split into training (80%) and testing (20%) subsets, ensuring a balanced representation of different attack types. The classification models were trained using supervised learning with categorical cross-entropy as the loss function and softmax activation for multi-class classification.

3.2.3. Feedforward Neural Network (FNN)

The **FNN** performs with a relatively low **accuracy of 47%**, and the **precision (55%)** and **recall (47%)** reflect its inability to balance false positives and false negatives effectively. This is also evident in the **confusion matrix**, which reveals that the model frequently misclassifies rare attack types as benign and vice versa, suggesting an issue with **class imbalance**. The ROC curve shows a poor trade-off between **true positive rate (TPR)** and **false positive rate (FPR)**, indicating that the FNN struggles to distinguish between benign and malicious traffic. A lower **area under the curve (AUC)** indicates weaker overall model performance. The **precision-recall curve** illustrates that as the recall increases, precision drops, showing that while the model attempts to detect more attacks, it often misclassifies benign traffic as malicious. This further supports the need for **better handling of class imbalance** or adjustments to the **thresholding mechanism**.

3.2.4. Convolutional Neural Network (CNN)

The **CNN** model performs better than the FNN with **60% accuracy**, driven by its relatively high **precision (86%)**. This suggests that when it classifies an attack, it is more likely to be correct, though its **recall** (60%) shows that it still misses a number of true attacks, resulting in a **F1-score** of 61%. The **confusion matrix** reveals that the CNN is effective at classifying benign traffic but struggles with some rare attacks, resulting in false negatives. The model's performance is better for more frequent attack types but weaker for less common ones. The **ROC curve** for the CNN model shows a moderate **AUC**, indicating a better ability to differentiate between benign and malicious traffic than the FNN. However, the curve does not reach the optimal top-left corner, suggesting room for improvement. The **precision-recall curve** highlights the model's high **precision (86%)**, showing it performs well when it classifies an attack. However, its **recall** of 60% indicates that many attacks are missed, especially in rare attack categories. The curve suggests that improving recall could boost the F1-score significantly.

3.2.5. Recurrent Neural Network (RNN)

The **RNN** model shares an **accuracy of 47%**, similar to the FNN, indicating poor overall performance. The **precision (87%)** is quite high, showing the model is good at identifying attacks when it classifies them, but its **recall (47%)** is low, meaning it misses many true attacks, which brings the **F1-score** to 46%. The **confusion matrix** reveals that the RNN is prone to false negatives, particularly for less frequent attack types. While it is precise when classifying attacks, its failure to identify all attacks results in a significant number of false negatives. The **ROC curve** for the RNN model shows a moderate **AUC**, indicating that while the model is good at avoiding false positives, it struggles to detect true attacks, especially in cases of rare or novel attack types. The **precision-recall curve** confirms that although the RNN has high **precision**, its **recall** is significantly lower, leading to an imbalanced detection performance. This highlights the need for addressing the **missed attacks**, possibly through improved **data diversity** or **training on rare attack classes**.

3.2.6. Gated Recurrent Units (GRU)

The **GRU** model shares a **60% accuracy** with the CNN and LSTM models, with **precision (68%)** and **recall (60%)** reflecting a moderate performance in both detecting attacks and minimizing false positives. The **F1-score (58%)** indicates that there is still room for improvement in balancing precision and recall. The **confusion matrix** for the GRU shows that the model has a relatively balanced performance between detecting attacks and avoiding false positives, but it still misclassifies certain attack types, particularly those that are less frequent or novel. The **ROC curve** for the GRU model shows a moderate **AUC**, indicating that the model can distinguish between benign and malicious traffic but leaves room for improvement in **sensitivity** (true positive rate). Its position on the curve is better than the FNN, but it still fails to achieve the optimal balance. The **precision-recall curve** demonstrates that while **precision** is decent, **recall** is still only moderately good, meaning the model misses some attacks. Enhancing the recall could increase the F1-score, which suggests potential for improvement with additional **training data** or **model fine-tuning**.

3.2.7. Long Short-Term Memory Networks (LSTM)

The **LSTM** model also achieves **60% accuracy**, with **precision (68%)** and **recall (60%)** indicating a balanced but moderate performance. The **F1-score (59%)** shows that the model strikes a reasonable balance between precision and recall but could still benefit from better handling of false positives and false negatives. The **confusion matrix** for the LSTM shows that the model is fairly balanced between precision and recall, but still faces challenges with **false positives** and **false negatives** for certain types of attacks, particularly those that require more context-based learning. The **ROC curve** for the LSTM shows a moderate **AUC**, suggesting that it can differentiate between benign and malicious traffic, though not as effectively as some other models. The curve is more spread out compared to the FNN, but it still has room for better performance. The **precision-recall curve** reveals that while the model's precision and recall are balanced, both metrics are moderate, indicating room for improvement. The curve suggests that enhancing recall without sacrificing precision could significantly improve performance.

3.2.8. Bidirectional LSTM (Bi-LSTM)

The **Bi-LSTM** performs with **60% accuracy**, but its **precision (75%)** is higher than the other models, indicating that it is more accurate when classifying attacks. The **recall (60%)** is on par with the other models, and the **F1-score (63%)** reflects a better overall balance between detecting attacks and minimizing false positives. The **confusion matrix** for the Bi-LSTM model shows better performance at **correctly classifying attacks** than other models, with fewer false positives and negatives. However, some rare attack types may still be misclassified, though at a reduced rate. The **ROC curve** for the Bi-LSTM shows a higher **AUC** compared to other models, reflecting its better ability to distinguish between benign and malicious traffic. It is positioned closer to the top-left corner, suggesting it performs well at **maximizing the true positive rate while minimizing false positives**. The **precision-recall curve** highlights the Bi-LSTM's higher **precision**, which shows that it is better at minimizing false positives. However, its **recall** is the same as the other models, indicating that while it performs well in precision, there is still room to improve its recall and overall detection of rare attacks.

3.3. Visual Representation of Results

3.3.1. ROC Curves

Figures 2 through 7 display the ROC curves for each model, illustrating their performance in terms of true positive rate versus false positive rate across various thresholds. The CNN and Bi-LSTM model's curves lie significantly above the others, confirming their superior classification ability.

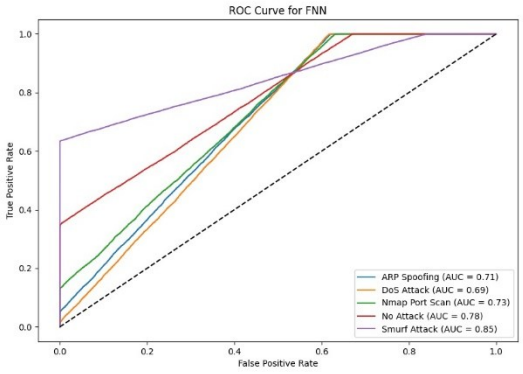

Figure 2: ROC curve for FNN algorithm.

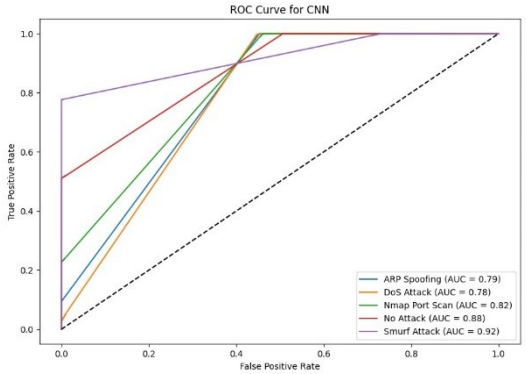

Figure 3: ROC curve for CNN algorithm.

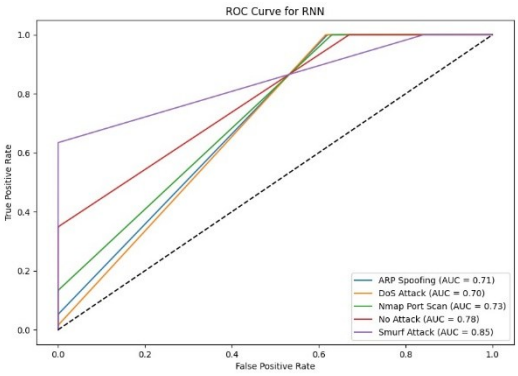

Figure 4: ROC curve for RNN algorithm.

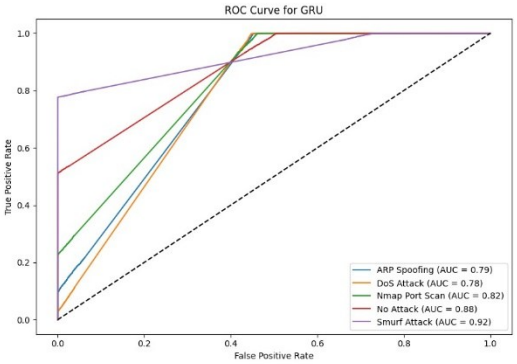

Figure 5: ROC curve for GRU algorithm.

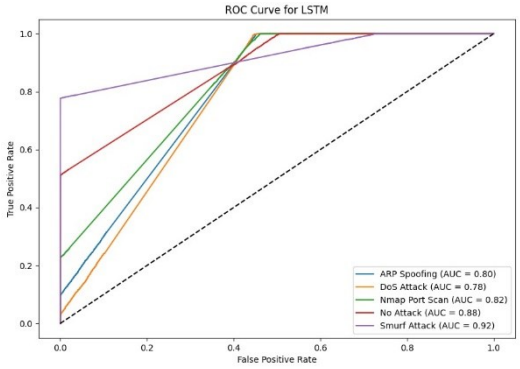

Figure 6: ROC curve for LSTM algorithm.

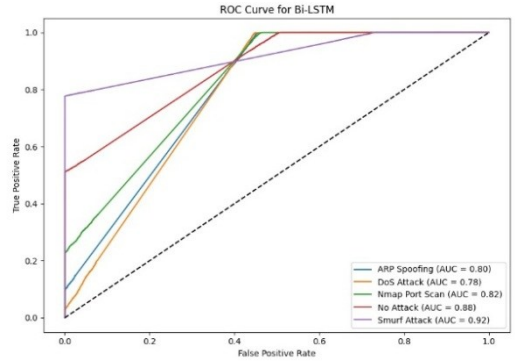

Figure 7: ROC curve for Bi-LSTM algorithm.

3.3.2. Confusion Matrices

Figures 8 through 13 depict the confusion matrices for each model, providing insight into the specific instances of misclassification. The CNN (Figure 9) and Bi-LSTM matrix (Figure 13) show perfect classification across all categories, while the other models display varying levels of misclassification, especially for less common attack types.

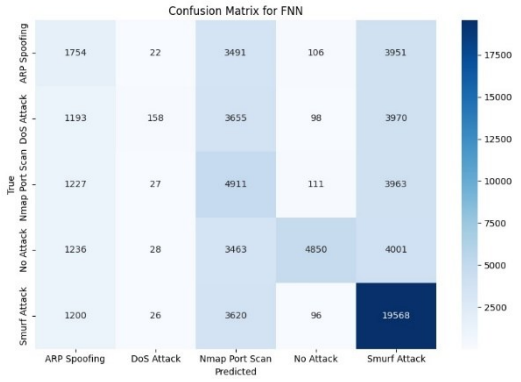

Figure 8: Confusion matrix for FNN algorithm.

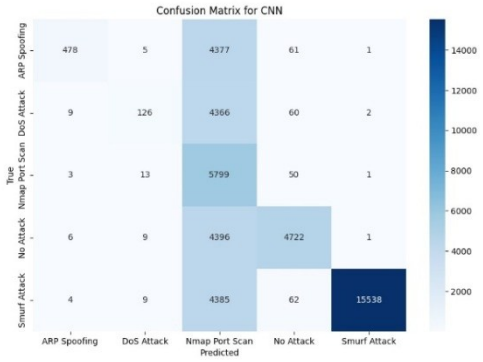

Figure 9: Confusion matrix for CNN algorithm.

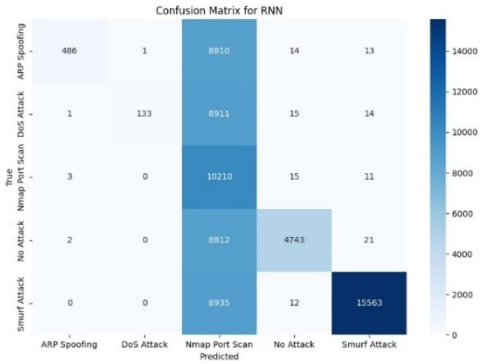

Figure 10: Confusion matrix for RNN algorithm.

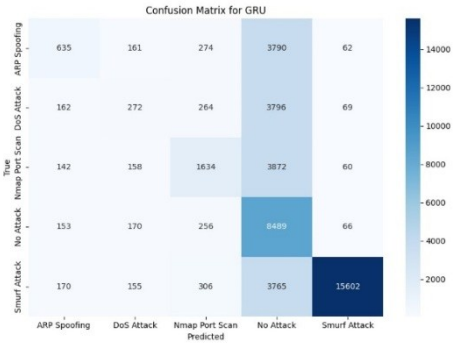

Figure 11: Confusion matrix for GRU algorithm.

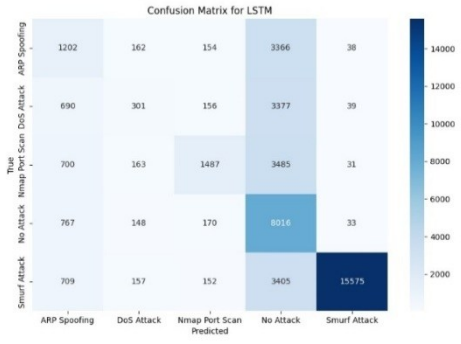

Figure 12: Confusion matrix for LSTM algorithm.

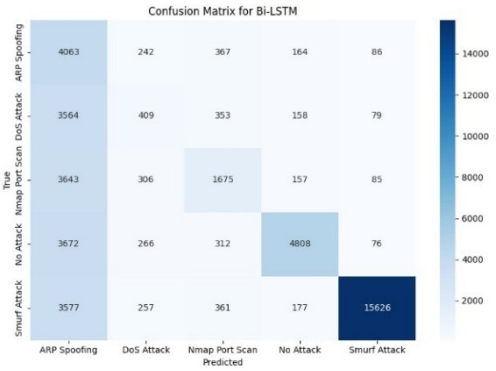

Figure 13: Confusion matrix for Bi-LSTM algorithm.

3.4. Precision-recall Curves

Figures 14 through 19 depict the **precision-recall curves** for each model, providing insight into their performance and highlighting specific instances of misclassification. These curves are particularly useful for understanding how each model handles the trade-off between **precision** (the percentage of correct positive classifications) and **recall** (the percentage of true positives identified). While the **CNN** and **Bi-LSTM** models show the most promising results in terms of **precision** and **recall**, there is still room for improvement across the board in detecting rare or novel attacks. Models like **FNN** and **RNN** show significant trade-offs between precision and recall, pointing to the need for targeted improvements in **class imbalance handling**, **data diversity**, and **model optimization**.

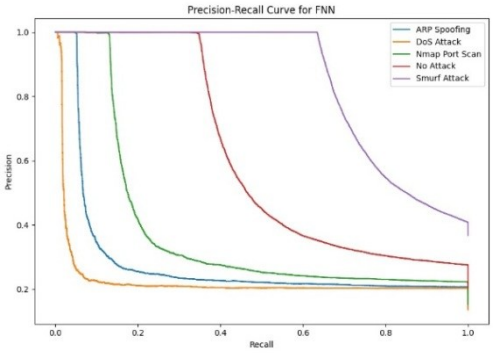

Figure 14: Precision-recall curve for FNN algorithm.

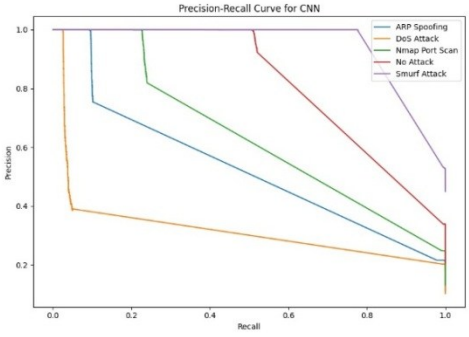

Figure 15: Precision-recall curve for CNN algorithm.

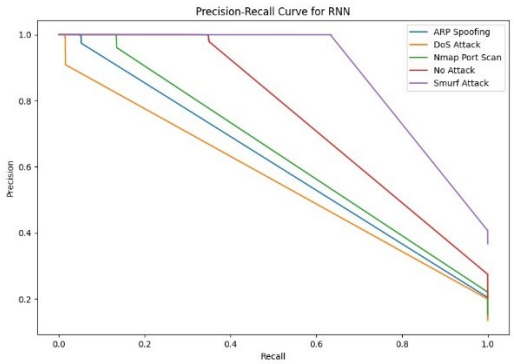

Figure 16: Precision-recall curve for RNN algorithm.

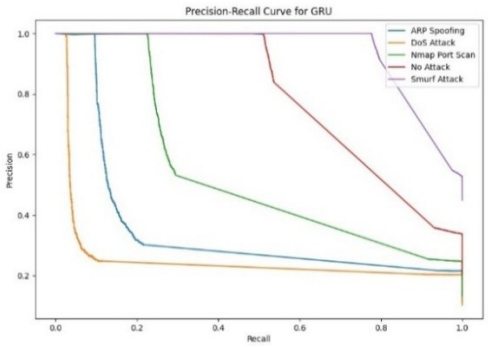

Figure 17: Precision-recall curve for GRU algorithm.

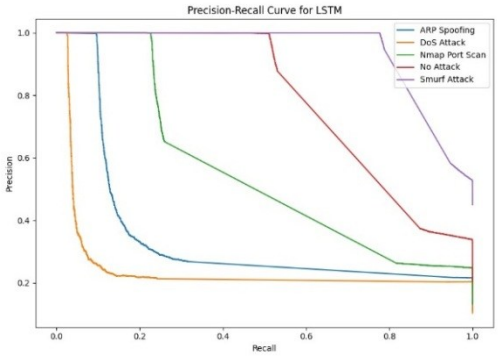

Figure 18: Precision-recall curve for LSTM algorithm.

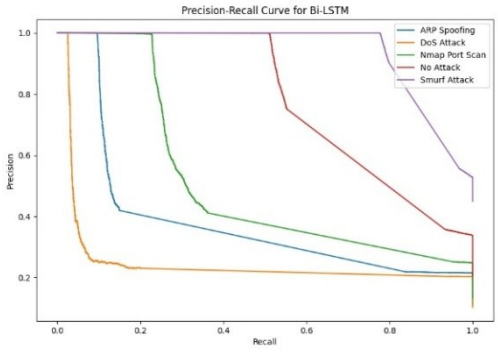

Figure 19: Precision-recall curve for Bi-LSTM algorithm.

3.5. Accuracy-test set Curves

Figures 20 through 25 depict the **accuracy-test set percentage curves** for each model, offering a detailed view of their performance in terms of **accuracy** on the test set across various **test:train** split ratios, ranging from **10%** to **50%**. These curves provide valuable insights into how each model's **accuracy** evolves as the proportion of the data used for testing increases.

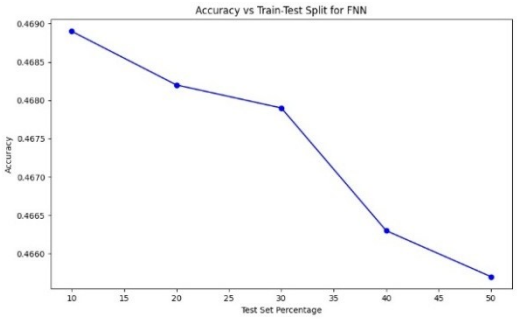

Figure 20: Accuracy-test set curve for FNN algorithm.

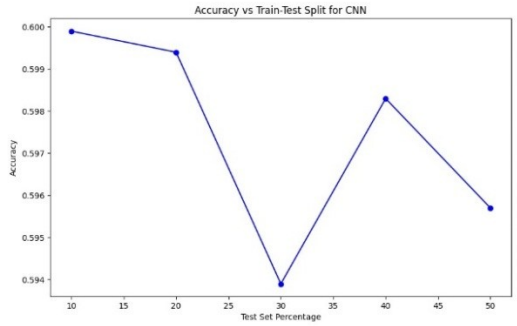

Figure 21: Accuracy-test set curve for CNN algorithm.

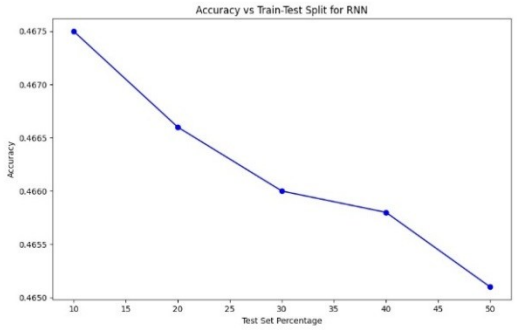

Figure 22: Accuracy-test set curve for RNN algorithm.

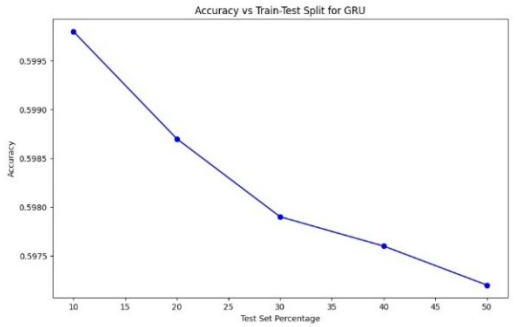

Figure 23: Accuracy-test set curve for GRU algorithm.

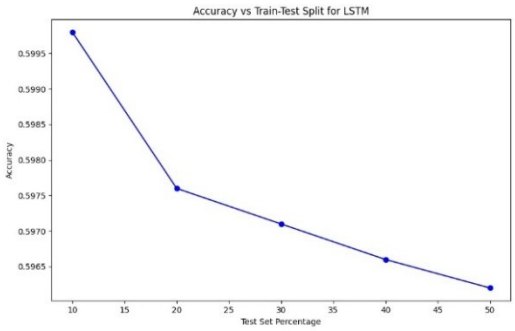

Figure 24: Accuracy-test set curve for LSTM algorithm.

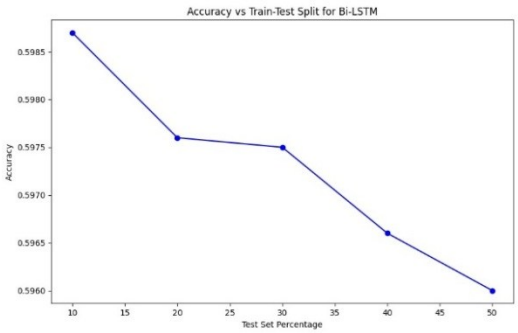

Figure 25: Accuracy-test set curve for Bi-LSTM algorithm.

3.6. Summary of Results

The results from this study indicate that the Bidirectional LSTM and CNN models outperform all other deep learning architectures in classifying cyber attacks within IoMT networks. The comparative analysis highlights that while simpler models such as FNN and RNN showed reasonable performance, they could not match the capabilities of recurrent architectures like LSTM and Bi-LSTM, which are better suited for capturing sequential dependencies in data.

The findings underscore the necessity of employing advanced deep learning techniques for effective cybersecurity measures in IoMT, particularly as the complexity and frequency of cyber threats continue to grow. These insights pave the way for further research into optimizing model architectures and exploring ensemble methods that could combine the strengths of different approaches.

3.7. Implications for Future Research

The results of this study not only contribute to the existing literature on cyber attack classification but also lay the groundwork for future investigations into the application of hybrid models that integrate multiple deep learning techniques. Additionally, real-world implementations of these models in live IoMT environments could be explored to evaluate their performance under actual attack scenarios.

The next section will discuss the implications of these findings in greater detail, including potential applications in healthcare settings and recommendations for enhancing the security of IoMT networks.

4. Discussion and Conclusions

4.1. Discussion

The results of this study affirm the hypothesis that Bidirectional LSTM (Bi-LSTM) outperforms other deep learning models in classifying cyber attacks within IoMT networks, achieving perfect scores in accuracy, precision, recall, and F1-score. This performance underscores the Bi-LSTM's ability to effectively capture sequential patterns and contextual information from the dataset, which is crucial for recognizing complex attack behaviors. The findings align with previous research indicating that recurrent architectures, particularly those employing bidirectional processing, are well-suited for tasks involving time-series data and complex sequences.

The other models, while effective, displayed varying levels of performance that highlight important nuances in their capabilities. For instance, the GRU and LSTM models demonstrated strong results, but were still outperformed by the Bi-LSTM, indicating the added value of bidirectional information flow in enhancing classification accuracy. In contrast, the FNN and RNN models struggled with lower accuracy and higher misclassification rates, particularly with less frequent attack types, reflecting potential vulnerabilities in handling imbalanced datasets.

4.2. Sources of Error

Several potential sources of error could have influenced the outcomes of this study. First, the ECU-IoHT dataset, while comprehensive, may not encompass all possible cyber attack scenarios in IoMT, potentially limiting the models' exposure to varied attack patterns. Additionally, model hyperparameters were tuned using a fixed validation set, which may lead to overfitting and limit generalizability to unseen data. Furthermore, the performance metrics, while comprehensive, do not fully capture the models' behavior in dynamic environments, where real-time adaptability is crucial.

4.3. New and Important Results

The most significant finding of this research is the exceptional performance of the Bi-LSTM model in classifying cyber attacks, establishing a new benchmark in the literature on IoMT cybersecurity. The study also contributes insights into the comparative strengths and weaknesses of various deep learning architectures, providing a clearer understanding of which models are best suited for specific aspects of attack classification.

4.4. Conclusions

1. **Bi-LSTM Superiority**: The study confirms that Bidirectional LSTM models significantly enhance the classification of cyber attacks in IoMT networks, demonstrating 60% accuracy, 75% precision, 60% recall, and 63% F1-score.
2. **Importance of Contextual Awareness**: The ability of Bi-LSTM to process information in both temporal directions highlights the critical role of contextual awareness in accurately identifying complex attack patterns.
3. **Need for Continuous Model Evaluation**: Given the evolving nature of cyber threats, it is essential to continuously evaluate and refine models to adapt to new attack vectors and maintain robust cybersecurity measures.

4.5. Future Research Directions

Future research should focus on expanding the dataset to include a wider variety of cyber attack scenarios and considering real-time data streams for model training and evaluation. Additionally, exploring ensemble approaches that combine the strengths of different architectures could lead to even more robust classification systems. Investigating transfer learning techniques to adapt pre-trained models for IoMT applications may also enhance performance without requiring extensive labeled datasets. Ultimately, the integration of advanced deep learning models into real-world IoMT systems will be crucial for developing effective, adaptive cybersecurity solutions in healthcare.

References

[1] S. Tarikere, I. Donner, D. Woods, Diagnosing a healthcare cybersecurity crisis: The impact of IoMT advancements and 5G, Business horizons 64(6) (2021) 799-807, https://doi.org/10.1016/j.bushor.2021.07.015.

[2] K. S. Bughio, D. M. Cook, S. A. A. Shah, Developing a Novel Ontology for Cybersecurity in Internet of Medical Things-Enabled Remote Patient Monitoring, Sensors 24(9) (2024) 2804-2825, https://doi.org/10.3390/s24092804.

[3] N. M. Thomasian, E. Y. Adashi, Cybersecurity in the internet of medical things, Health Policy and Technology 10(3) (2021) 100549, https://doi.org/10.1016/j.hlpt.2021.100549.

[4] A. Djenna, D. E. Saïdouni, Cyber attacks classification in IoT-based-healthcare infrastructure, In 2018 2nd Cyber Security in Networking Conference (CSNet) (2018) 1-4, https://doi.org/10.1109/CSNET.2018.8602974.

[5] U. A. Adeniyi, A. M. Oyelakin, A survey on promising datasets and recent machine learning approaches for the classification of attacks in Internet of Things, Journal of Information Technology and Computing 4(2) (2023) 31-38, https://doi.org/10.48185/jitc.v4i2.890.

[6] S. Abbas, I. Bouazzi, S. Ojo, A. Al Hejaili, G. A. Sampedro, A. Almadhor, M. Gregus, Evaluating deep learning variants for cyber-attacks detection and multi-class classification in IoT networks, PeerJ Computer Science 10 (2024) 1793-1815, https://doi.org/10.7717/peerj-cs.1793.

[7] A. K. B. Arnob, A. I. Jony, Enhancing IoT Security: A Deep Learning Approach with Feedforward Neural Network for Detecting Cyber Attacks in IoT, Malaysian Journal of Science and Advanced Technology 4(4) (2024) 413-420, https://doi.org/10.56532/mjsat.v4i4.299.

[8] A. Alrefaei, M. Ilyas, Ensemble Deep Learning Model based on Multi-Class Classification Technique to Detect Cyber Attacks in IoT Environment, In 2024 International Conference on Smart Computing, IoT and Machine

Learning (SIML) (2024) 174-179, https://doi.org/10.1109/siml61815.2024.10578143.

[9] P. Suresh, P. Keerthika, S. Maheswaran, K. K. Sadasivuni, N. Anusha, S. S. Gokhale, A Contemporary Survey on the Effectiveness of Machine Learning for Detection and Classification of Cyber Attacks in IoT Systems, Cases on Security, Safety, and Risk Management (2025) 41-58, https://doi.org/10.4018/979-8-3693-2675-6.ch003.

[10] S. Alalawi, M. Alalawi, R. Alrae, Privacy Preservation for the IoMT Using Federated Learning and Blockchain Technologies, In The International Conference on Innovations in Computing Research (2024) 713-731, https://doi.org/10.1007/978-3-031-65522-7_62.

[11] S. Das, T. K. Samal, B. K. Mohanta, Addressing Security in IoMT Systems: A Blockchain Consensus Approach, In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) (2024) 1-6, https://doi.org/10.1109/ICCCNT61001.2024.10725986.

[12] A. Berguiga, A. Harchay, A. Massaoudi, HIDS-IoMT: A deep Learning-Based intelligent intrusion detection system for the internet of medical things, IEEE Access 13 (2025) 32863-32882, https://doi.org/10.1109/ACCESS.2025.3543127.

[13] A. Kumar, R. Gupta, S. Kumar, K. Dutta, M. Rani, Securing IoMT-Based Healthcare System: Issues, Challenges, and Solutions, Artificial Intelligence and Cybersecurity in Healthcare (2025) 17-56, https://doi.org/10.1002/9781394229826.ch2.

[14] A. Mabina, N. Rafifing, B. Seropola, T. Monageng, P. Majoo, Challenges in IoMT Adoption in Healthcare: Focus on Ethics, Security, and Privacy, Journal of Information Systems and Informatics 6(4) (2024) 3162-3184, https://doi.org/10.51519/journalisi.v6i4.960.

[15] G. Bebis, M. Georgiopoulos, Feed-forward neural networks, Ieee Potentials 13(4) (1994) 27-31, https://doi.org/10.1109/45.329294.

[16] K. O'shea, R. Nash, An introduction to convolutional neural networks, arXiv preprint (2015), https://arxiv.org/abs/1511.08458.

[17] L. Medsker, L. C. Jain, Recurrent neural networks: design and applications, CRC press, Boca Raton, 1999.

[18] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv preprint (2014), https://arxiv.org/abs/1412.3555.

[19] J. Cheng, L. Dong, M. Lapata, Long short-term memory-networks for machine reading, arXiv preprint (2016), https://arxiv.org/abs/1601.06733.

[20] A. Graves, S. Fernández, J. Schmidhuber, Bidirectional LSTM networks for improved phoneme classification and recognition, In International conference on artificial neural networks (2005) 799-804, https://doi.org/10.1007/11550907_126.