A Hybrid Deep Learning Framework for Speech-to-Text Conversion as Part of Telemedicine System Integrated With 5G

Review Paper

Medapati Venkata Manga Naga Sravan*
Andhra University
sravan.medapati@gmail.com

K Venkata Rao
HOD, Dept of Computer Science, Andhra University
professor_venkat@yahoo.com

*Corresponding author

Abstract *– In today's world, aligning healthcare research with the third sustainable development goal of the United Nations (UN) is crucial. This goal focuses on ensuring health and well-being for all. Technological innovations like the Internet of Things (IoT) and Artificial Intelligence (AI) are vital in improving healthcare systems. Developing a technology-driven telemedicine system can have a significant impact on society. While current approaches focus on various methods for developing telemedicine modules, advancing these models with the latest technology is essential. Our paper proposes a deep learning-based framework that allows patients to provide information through voice. The system automatically analyzes this information to provide valuable insights in the doctor's dashboard, making diagnosis and prescriptions easier for the patient. Our proposed hybrid deep learning framework integrates with 5G technology and focuses on speech-to-text conversion. We introduce a hybrid deep learning model to improve performance in speech-to-text conversion. Our proposed algorithm, AI-Enabled Speech-to-Text Conversion (AIE-STTC), has the potential to match and surpass many existing deep learning models. Our empirical study, conducted using a benchmark dataset, demonstrated an impressive accuracy rate of 95.32%. In comparison, the baseline models showed lower accuracy rates: CNN achieved 88%, ResNet50 reached 90%, and VGG16 had 89%. Therefore, our proposed methodology has the potential to realize a technology-driven telemedicine system by integrating it with other necessary modules in the future. It significantly improves remote patient healthcare, making it more accessible and cost-effective, leading to a hopeful paradigm shift in healthcare services.*

Keywords*: Telemedicine System, Artificial Intelligence, 5G Technology, Deep Learning, Patient Voice-to-Text Conversion*

Received: September 9, 2024; Received in revised form: December 23, 2024; Accepted: December 24, 2024

1. INTRODUCTION

Health and well-being for all are critical sustainable development goals the United Nations sets. Researchers have been developing various technologies and approaches to improve healthcare services in alignment with this goal. Traditional healthcare systems have been enhancing service delivery and disease diagnosis due to healthcare equipment and technologies innovations. However, the experience of the COVID-19 pandemic made the world more conscious about people's health regardless of their country or region. Efforts have been made to explore different means of providing healthcare services, including remote patient monitoring with the help of artificial intelligence and Internet of Things technology, as well as creating a personalized medicine system by integrating required technologies such as 5G.

Regarding telemedicine, the Indian government has been working to provide healthcare services through public healthcare units accessible to people from all walks of life with just a phone call. However, there are many challenges in realizing such a telemedicine system. These challenges include technical issues related to infrastructure development, software integration, and security and privacy of existing healthcare-related IT systems. Regulatory challenges include licensing and credentialing, reimbursement policies, legal and compliance issues, and problems associated with user acceptance and experience, accessibility, provider train-

ing, and patient engagement. Clinical and operational challenges include technical support, secure data management, and quality of patient care. Furthermore, ethical challenges include patient privacy and ensuring equitable access. Economic factors associated with the telemedicine system, such as cost-benefit analysis and implementation costs, also pose challenges.

Many ongoing research efforts focus on providing advanced health services using 5G technology. Privacy concerns must be explored in a telemedicine system with 5G technology [1]. Integrating telemedicine with 5G and blockchain technology aims to enhance security, privacy, and non-repudiation [2]. Additionally, exploring a decision support system for telemedicine, incorporating edge computing and 5G technology along with sensors and available devices, is being pursued [3]. Technologies such as artificial intelligence and the Internet of Things are also being investigated as part of developing telemedicine systems to improve healthcare services [4]. Intelligent healthcare systems utilizing IoT technology and wearable devices are also being studied to integrate them with telemedicine and explore the ecosystem's role, including cloud, artificial intelligence, and artifice technologies [5]. Finally, a technical framework with various components is being explored to identify and plan the development of telemedicine systems for the future, leveraging innovative infrastructures, including 5G technology [6]. Based on the literature review, it is evident that developing a telemedicine system is a complex process, requiring investigation of various methods and gradual development to realize the potential of exploiting emerging technologies and advancing healthcare services.

This paper proposes a deep learning-based framework that enables patients to provide information through voice. The system automatically analyses this information to provide valuable insights on the doctor's dashboard, making diagnosis and prescriptions easier for the patient. Our hybrid deep learning framework integrates with 5G technology and focuses on speech-to-text conversion. We propose a hybrid deep learning model to enhance performance in speech-to-text conversion. Our algorithm, AI-Enabled Speech Conversion (AIESC), has the potential to outperform many existing deep learning models, as our empirical study using a benchmark dataset showed an impressive 95.32% accuracy rate. Therefore, our methodology has the potential to realize a technology-driven telemedicine system by integrating it with other necessary modules in the future. It significantly improves remote patient healthcare, making it more accessible and cost-effective, leading to a hopeful paradigm shift in healthcare services. The remaining sections of the paper are structured as follows: Section 2 reviews existing methods available in the literature for developing various modules in the telemedicine system. Section 3 presents the proposed deep learning-based or AI-enabled framework, which facilitates remote patients accessing healthcare services. Section 4 presents the experimental results from our empirical study. Section 5 discusses the significance of the proposed system and its limitations. Section 6 concludes our work and provides directions for the future scope of the research.

2. RELATED WORK

Various existing methods are found in the literature about developing modules required for telemedicine systems. Lin *et al.* [1] improved connectivity between devices and energy efficiency with a 5G network. 5G enables safe, anonymous identity management for privacy in telemedicine, which improves the healthcare industry. Hameed *et al.* [2] proposed an IoHT-based healthcare system that combines blockchain, NN, and 5G for illness severity assessment and prediction. It improves healthcare efficiency and guarantees data confidentiality and privacy with 98.98% accuracy. PSO technology use, patient profiling, optimization, and various algorithms are examples of future developments. Wang *et al.* [3] suggested a 5G MEC-based telemedicine architecture incorporatingOpenEMR with wearables. The multi-layered technique improves efficiency, scalability, and connection for applications other than Afib detection. Yu *et al.* [4] suggested a cloud-converged Internet of Things health architecture that prioritizes emotional engagement and multimodal sensing. A QoS framework for LAN-based health on the Internet of Things has been created. Suleiman *et al.* [5] examined how developments in 5G, IoT, AI, telemedicine, and networked competent healthcare are combined. It talks about difficulties, advantages, and potential futures.

Sadia *et al.* [6] included strategic planning, collaboration with medical professionals, and ongoing eHealth literacy. Public and private sector investments are necessary for affordable telemedicine in rural places with limited infrastructure. Research is essential for sustainable eHealth infrastructure, and 5G's role in connection is critical. Li *et al.* [7] assessed and validated the viability, effectiveness, and improved safety features of a 5G Telemedicine Network Latency Management System for telesurgery. Lin *et al.* [8] suggested a user-controlled single sign-on (SC-UCSSO) for telemedicine systems based on smartcards thatensure increased security, privacy, and performance. Hewa *et al.* [9] promoted using blockchain, 5G, and Multi-access Edge Computing (MEC) in digital healthcare infrastructure to improve patient privacy, data integrity, and scalability. Lu *et al.* [10] established the safety and efficacy of a telemedicine system for managing several diseases in a confined area.

Chettri *et al.* [11] concentrated on developing 5G wireless communication technologies that use Filter Bank Multicarrier (FBMC) for telemedicine effective transmission. The suggested method improves data rates to enable prompt patient monitoring. The simulation findings indicate possibilities for incorporating smartphones in telemedicine systems, increased effi-

ciency, and decreased delays. Diong *et al.* [12] exploited 5G to effectively handle the necessity for telemedicine in high-speed situations. By reducing needless handovers by 80.3%, a suggested changeover algorithm improves the quality of telemedicine services. Sadia *et al.* [13] presented a 5G healthcare system that is more efficient and less prone to latency than 4G using the TRILL protocol for data transport and mobility management. Figueiredo *et al.* [14] achieved fast speeds of 115 ps for an ultrafast electro-optical switch based on a chip-on-carrier semiconductor optical amplifier. Alenoghena *et al.* [15] examined eHealth, wireless technology, communication protocols, and problems in light of the COVID-19 pandemic's spike in telemedicine.

Lin *et al.* [16] presented an ID-based secure communication technique that protects privacy in 5G-IoT telemedicine systems. It integrates telemedicine with emergency medical services (EMS) and ensures the safe transfer of patient information, prompt delivery of emergency signals, and resilience to possible assaults. Liou *et al.* [17] suggested an affordable QoS benchmark system with good performance, simulating 5000 telemedicine devices for 5G uRLLC and mMTC scenarios. Adarsh *et al.* [18] suggested using effective communication technologies, dynamic prioritization, health service prioritization, and a cognitive radio-based telemedicine network for e-health. The performance of the proposed scenario can be improved by implementing WiMAX connectivity for mobility speeds less than 300 kmph and integrating 5G. The network may be further enhanced at the PCC level by utilizing SRD and UWB technology. Colella *et al.* [19] increased forecast precision, lowered expenses, and guaranteed successful BLM production quality. With its excellent parameter optimization performance, the suggested systematic design approach may be used for various industries and light sources. Silva *et al.* [20] presented a Local 5G Operator (L5GO) architecture emphasizing robotic surgery and augmented reality for delay-critical telemedicine. Regarding latency, the suggested L5GO performs better than conventional and Multi-access Edge Computing (MEC) networks, providing unique benefits for telehealth applications sensitive to delays.

Mihuba *et al.* [21] presented a mobile terminal and general packet radio service-based remote medical monitoring system that enables quick and affordable wireless telemedicine. The suggested architecture improves mobility and convenience by integrating sensors, CPUs, and communication. Bailo *et al.* [22] accessed healthcare expanded by telemedicine, which was essential during the epidemic. Lawmakers must support telesurgery since it presents both technological and legal issues. Chettri *et al.* [23] emphasized using Filter Bank Multicarrier (FBMC) in 5G telemedicine to improve remote healthcare in underprivileged regions by transmitting vital signs and imaging data efficiently. Arunsundar *et al.* [24] suggested integrating telemedicine into 5G networks to handle emergencies using massive MIMO and cognitive radio networks. Peralta-Ochoa *et al.* [25] examined how 5G technology may be used in intelligent healthcare applications, focusing on theoretical ideas and small-scale applications. The research indicates that intelligent healthcare is becoming increasingly important, especially in light of the COVID-19 pandemic. A SWOT analysis to evaluate technical support and suggest alternatives may be part of future efforts.

Cabanillas-Carbonell *et al.* [26] examined 66 pertinent articles about how 5G could affect healthcare applications, focusing on cloud, AI, and IoT technology. The evaluation has significance for forthcoming investigations that seek to augment healthcare via 5G technologies, cultivating more intelligent, effective, and enduring healthcare systems. Albahri *et al.* [27] assessed networks, services, and applications related to IoT in telemedicine. By highlighting effective telemedicine for larger populations using IoT technology, it unearths answers from 141 publications. Ahmad *et al.* [28] Despite its importance during COVID-19, telehealth and telemedicine confront obstacles. Blockchain improves data security and privacy in healthcare by providing decentralized, traceable, and secure solutions. Jain *et al.* [29] proposed a 5G Network Slice-based digital system for real-time patient-centric healthcare to meet the post-COVID healthcare demand. Sadia *et al.* [30], Adford *et al.* [31] developed speech models trained on 680,000 hours of diverse data without fine-tuning. However, they acknowledged the need for extensive datasets and aimed to improve model accuracy and robustness in future research.

Based on the literature review, developing a telemedicine system is a complex process requiring investigation of various methods and gradual development to realize the potential of exploiting emerging technologies and advancing healthcare services. Orynbay *et al.* Specifically, [32] take a very newfound look to the integrations/synthesizes of speech, text and vision modalities, and mediates the devoted multi-modal interaction systems as a consequence. AbstractThis paper is a review of recent advances in novel methods and technologies that facilitate joint, bidirectional communication between multiple modalities, through the enhancement of a single modality through behaviour image generation or through a more connected multisensory experience between modalities. In the study by Dhakad and Singh [33] the authors have provided a survey and performance analysis of speech to text technologies implemented by using python and their performance analysis, characteristics domain of usage. The paper evaluates different tools, frameworks focuses on their performances on Accuracy and Usability. Madhusudhana Reddy *et al.* Deep learning-based methods for speech-to-text and text-to-speech recognition are discussed to improve the performance and accuracy [34]. It identifies breakthrough neural models that are leading to the development of speech and text process-

ing systems. An extensive survey is provided by Sethiya and Maurya [35], which introduces advanced neural architectures for end-to-end speech-to-text translation systems. Authors define important challenges, methods, and future directions for work toward continuous, multilingual translation systems. Korchynskyi *et al.* [36] propose a method for enhancing the quality of speech-to-text conversion by noise suppression and language modelling. The true potential of the study is to understand it and make it easy, accurate, and energy-efficient for several applications. telemedicine combines medical aid with technology during catastrophes, providing a tactical method for practical victim assessment, treatment prioritization, and coordination. Dar and Pusharaj [37] proposed a CNN-BLSTM hybrid model with Connectionist Temporal Classification (CTC) for speech recognition. The model achieved a word error rate of 36.97% but noted accuracy and training time challenges. Baevski *et al.* [38] introduced wav2vec 2.0, which performs well in voice recognition with less labelled data while identifying pre-training reliance as a limitation.

3. PROPOSED FRAMEWORK

The telemedicine system is a complex phenomenon that involves various components, protocols, networks, and communication methods. With the emergence of technologies like 5G and Artificial Intelligence, it has become possible to implement complex systems that were not feasible before. However, due to its complexity, we are focused on developing different methods to realize a technology-driven telemedicine system. In other words, we are addressing the challenge of developing various strategies that are part of a technology-driven telemedicine system. This section introduces the proposed methodology, algorithm, and hybrid deep learning model involved in the proposed system.

3.1. PROBLEM DEFINITION

The telemedicine system proposal includes a crucial method for translating spoken English into English text. This module is essential for developing the telemedicine system, as discussed in section 3.2. The challenging problem addressed in this work involves developing a hybrid deep learning model to convert English speech into English text as part of the telemedicine system. Additionally, our proposed telemedicine system will require other modules, the implementation of which will be deferred to our future work.

3.2. OUR FRAMEWORK

A technology-driven telemedicine system is envisioned as a game changer in health service provision for the general public. Many minor ailments may not require an in-person visit to a doctor. Providing affordable healthcare services to accommodate people from various economic backgrounds is crucial, reducing unnecessary expenses and time wastage. Commuting and spending a whole day to consult a doctor for a simple ailment that could be addressed through a phone consultation could be more efficient. Therefore, there is a need to develop a novel healthcare system, such as a telemedicine system, to enable people to seek advice from doctors without spending excessive money or time. Although developing a telemedicine system is complex, we propose a technology-driven one and implement one of the modules discussed in this paper. Implementing other models or methods necessary for actualizing a telemedicine system is deferred for our future endeavours. Fig. 1 displays the technology framework that leverages artificial intelligence and 5G technology to develop a telemedicine system.

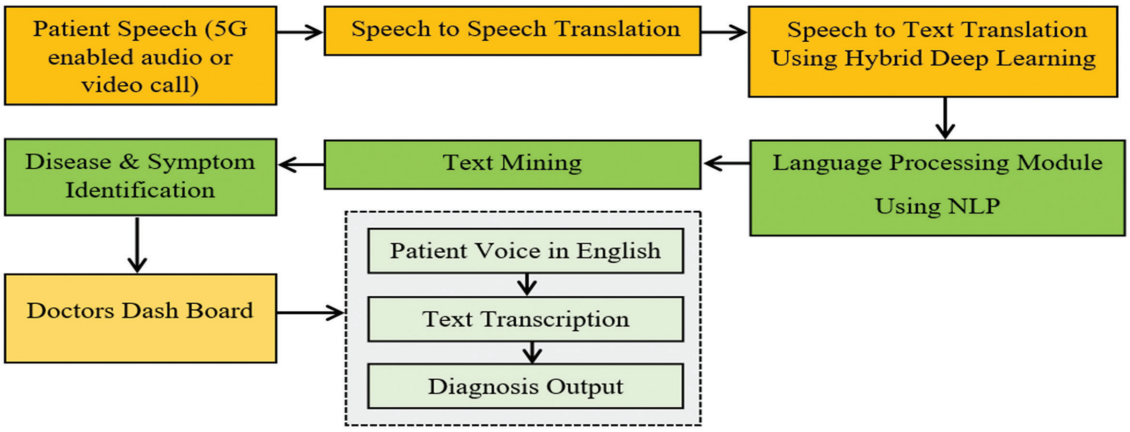

Fig. 1. Overview of the proposed telemedicine system

The proposed system aims to provide inclusive healthcare services accessible to patients of any region, religion, or language. The system is designed to understand spoken language, translate it, and provide information to doctors, making it easier for them to prescribe treatment. Patients can access healthcare services through a simple phone call and receive necessary prescriptions or advice, significantly impacting the lives of people in society. The system utilizes voice-to-voice translation using deep learning models to translate a patient's native language into English accurately. Once translated, the speech is converted

to text, enabling advanced natural language processing and machine learning techniques for data analytics. The system also employs artificial intelligence to identify possible symptoms and diseases, presenting information to the doctor's dashboard. Without a doctor, healthcare professionals can handle patient calls, record the conversation, and make the data analytics results available to the doctor for decision-making. The doctor can then prescribe a solution or medicine, and the patient is informed immediately, either verbally or via messaging in their native language, eliminating the need to visit a healthcare facility physically.

The system ensures that patients can access healthcare services based on a set protocol, regardless of location. Integration with cloud technology allows doctors to access patient information whenever necessary. At the front end, the system receives patient calls in their native language and translates them into English, filtering background noise and using speech signals for accuracy. Deep learning models, including hybrid deep learning, recognize speech patterns and essential features for speech recognition. Language models play a significant role in understanding English text, preprocessing, and aiding machine learning models in data analytics. The proposed system uses artificial intelligence to streamline diagnosis, empowering doctors to make well-informed decisions.

3.3. 5G TECHNOLOGY

Technologies like 5G play a crucial role in developing telemedicine systems. This technology enables data transfer at a much faster rate compared to its predecessors. In other words, this technology allows patients to have video calls to consult with doctors, so the doctors can not only listen to the patients but also see the patient's condition better. The 5G technology enables high-quality data transfer, which is crucial for realizing a telemedicine system. When this technology is spread to remote areas, it helps people communicate seamlessly with doctors through telemedicine. The access to the telemedicine system by people from all walks of life will be improved with 5G technology. Therefore, 5G technology can be adopted to improve healthcare services.

Considering the COVID-19 pandemic, where the world has learned a lesson about the importance of health, it is crucial to understand the importance of consulting doctors without physically moving to hospitals. Therefore, people from different fields need seamless access to the telemedicine system. People in remote areas can quickly access healthcare services through the telemedicine system. The 5G technology can also enable mobile health applications that provide interactivity between people and healthcare professionals. The technology can also be used to develop remote patient monitoring systems with the help of the Internet of Things and artificial intelligence to monitor patient vitals in real time and provide appropriate medical intervention. In the contemporary era, 5G technology is also being used for virtual surgeries from remote areas due to its real-time approach, low latency, and very high speed of data, making it possible to have remote-controlled systems with robotics for disease diagnosis and performing surgical procedures as well.

3.4. ENGLISH SPEECH TO ENGLISH TEXT TRANSLATION

We used an artificial intelligence-enabled approach with a hybrid deep learning model to convert patient speech into English text. The process, shown in Fig. 2, includes training and testing. In the first phase, the hybrid deep learning model is trained with features extracted from the training dataset, a speech recognition challenge dataset. After preprocessing the extracted features as spectrograms, they are used to train the model. The trained model can then translate any patient's speech into English text.

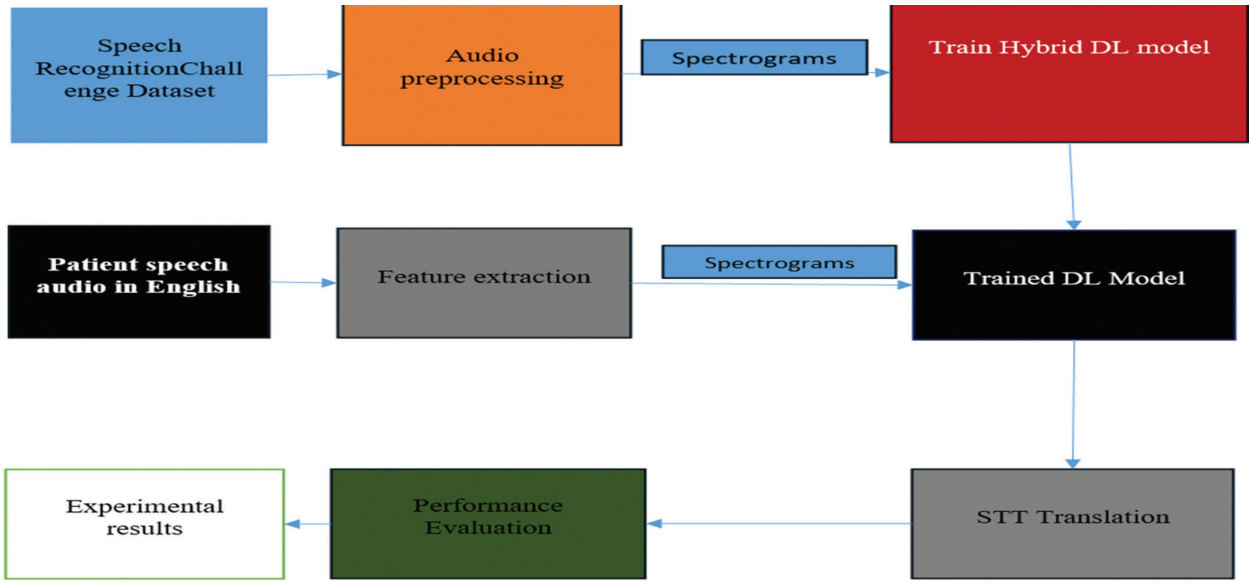

Fig. 2. Proposed methodology for speech-to-text (STT) conversion

Our previous discussion acknowledged that creating an entire telemedicine system is a complex task that demands significant resources.

Therefore, the development and empirical study outlined in this paper is focused solely on one module of the proposed telemedicine system: the conversion of English speech to English text. As previously mentioned, the patient communicates in their native language, which is then translated into English speech. The development of this particular aspect of the framework is deferred to our future work. The primary focus of this paper is the translation of English speech to English text, which is accomplished through an enhanced deep learning model, as illustrated in Fig. 3. The deep learning model we have devised is a hybrid model that effectively utilizes convolutional layers and bidirectional GRU layers to efficiently convert English speech to text.

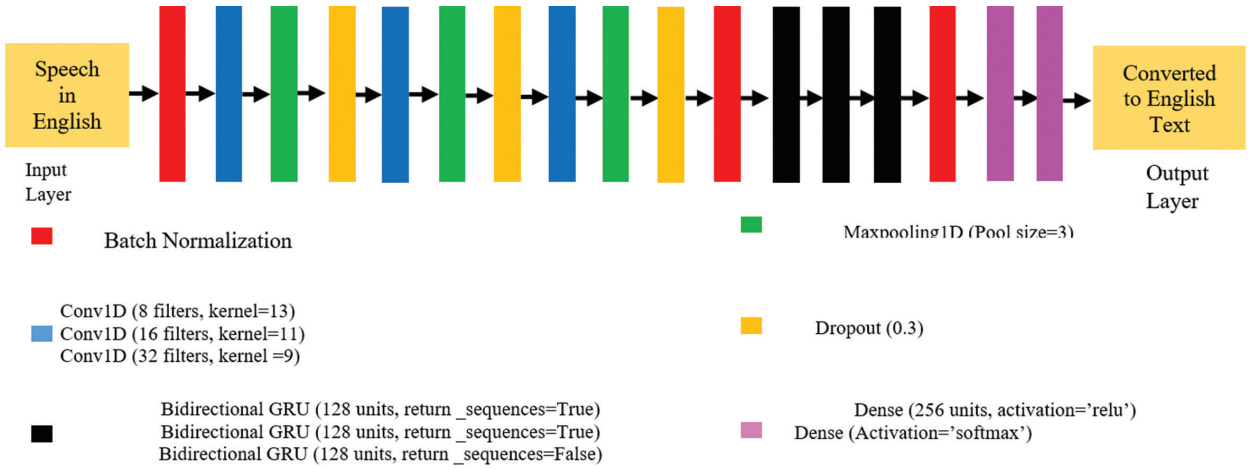

Fig. 3. The proposed hybrid deep learning model, a significant advancement for translating speech to English in a telemedicine system

In Fig. 3, we have a hybrid deep learning model designed for translating spoken English into written English text, specifically for use in a telemedicine system. The English speech input is fed into the model through the input layer, which then normalizes the input data for improved training performance and stability. The model employs convolutional layers, such as Conv1D with different filter sizes and kernels, and Maxpooling1D with a specified pool size to reduce the dimensionality of the feature maps. Dropout rate of 0.3 is used to address overfitting. The data is processed through several Bidirectional GRU layers with different configurations, including returning or not returning sequences. Finally, the model uses Dense layers with ReLU and softmax activations for the final output, which is the translated text in English. In summary, the model employs convolutional layers, max pooling, dropout, bidirectional GRUs, and dense layers to effectively translate spoken English into written English, enhancing the telemedicine system's ability to process and comprehend speech.

When combining CNN and bidirectional GRUs, batch normalization is essential for improving converting speech-to-text efficiency. The normalization process works on inputs of layers to leverage speed in training and achieve model stability. For speech recognition, normalization enhances the capability of the deep learning model. We are learning complex patterns from the data to make more accurate transcription. The normalization process is essential for leveraging the model's performance in the proposed hybrid architecture. In the proposed hybrid deep learning model, convolutional layers extract feature maps from the audio input. The filters used in the convolutional layers are meant to detect particular patterns in the voice and help improve transcription performance. The feature maps developed by convolutional layers help understand the phonetic elements in the given inputs and contextual information. Therefore, convolutional layers are crucial to understanding the difference between similar words and dealing with complexities in spoken language.

The proposed hybrid deep learning model also uses Max pooling layers, which take the feature maps obtained from convolutional layers and reduce spatial dimensions of the outcomes of convolutional layers. These layers use a sliding window on the given feature maps to get a value for which the max pooling is intended. This process involved in Max Pooling enables the model to reduce the feature maps and optimize for further processing. The method of lowering feature maps or optimizing them has its influence on reducing computational complexity, besides helping the hybrid model reduce overfitting problems, as it provides a concept known as translation invariants that helps understand unseen data. As deep learning models are extended neural networks, they are designed to eliminate overfitting problems. This reduction of overfitting is achieved with dropout layers, which can help improve the learning process and reduce noise, besides addressing the issue of overfitting by setting some

inputs to zero in the training process towards making the model more robust in learning and understanding from the data.

A Bidirectional Gated Recurrent Unit (BiGRU) is used in the proposed hybrid deep learning model to enhance the capability of the model to understand dependencies in the sequential nature of data. The BiGRU has two layers that work in the forward and backward directions. This dual approach can help understand the context of the past and the future states while dealing with language data, which is essentially time series data. The proposed deep learning model has a fully connected layer, an essential component in the network. Every neuron is connected to every other neuron in the preceding layer to ensure holistic interaction and data flow among the layers. This layer helps understand the complex patterns related to voice data towards converting data from speech to English text.

3.5. PROPOSED ALGORITHM

The proposed algorithm, AI-Enabled Speech-to-Text Conversion (AIE-STTC), is a critical component of this research. It aims to develop a mechanism to automatically convert patient speech into text using a hybrid deep learning model. The algorithm takes the patient's speech in audio format and the training data to train the deep learning model. Once trained, the model can automatically convert the patient's speech into English text. This algorithm is designed to be a part of a Telemedicine system, offering numerous benefits for healthcare services.

Algorithm: AI-Enabled Speech to Text Conversion (AIE-STTC)

Input: Patient speech audio q, training dataset T

Output: STT conversion results in R, performance statistics P

1. Begin
2. Initialize features map M
3. For each sample t in T
4. features←ExtractFeatures(t)
5. Add t and features to M
6. End For
7. Configure hybrid DL model m (as in Fig. 3)
8. Compile m
9. m'←TrainDLModel(m, M)
10. Persist m'
11. Load m'
12. R←STTConversion(m', q)
13. P←Evaluation(R, ground truth)
14. Print R
15. Print P
16. End

Algorithm 1: AI-Enabled Speech-to-Text Conversion (AIE-STTC)

Algorithm 1 is designed for speech-to-text conversion, an essential component of the telemedicine system. It employs an AI-enabled approach using a hybrid deep learning model illustrated in Figure 3. The algorithm involves training the deep learning model with a provided training dataset. Before teaching the model, there is a preprocessing step where each training sample is converted into spectrograms. These features are then used to train the proposed deep learning model. The training process involves using all the samples in the training dataset, where each sample consists of the patient's speech and the corresponding converted English text.

This helps the deep learning model to learn from the samples and acquire sufficient knowledge. Once the model has gained knowledge, it is saved for future use. When the algorithm receives a new patient's speech as input, it is converted into English text using the trained deep learning model. The algorithm progresses toward achieving patient speech-to-text conversion, ultimately benefitting further modules associated with that element. During testing, the deep learning model demonstrates its capability to translate patient speech into English text, and the algorithm assesses the model's performance. This algorithm enables the realization of the speech-to-text conversion module in the telemedicine system, which plays a crucial role in the data analytics module. Implementing various modules within the telemedicine system can help improve the quality of healthcare services. This novel, technology-driven approach can bring significant benefits to people at large.

3.6. DATASET DETAILS

This paper's empirical study uses the benchmark data set collector from [39].

4. EXPERIMENTAL RESULTS

This section presents the results of our empirical study. The proposed telemedicine system consists of multiple modules, with the STT conversion module being the focus of this paper's empirical study. This module enables the telemedicine system to translate the patient's speech audio into English transcription. To train the deep learning model, we utilized a dataset comprising pairs of samples, each containing speech audio and its corresponding text. The training set shall comprise 80% of the data, while the testing set comprises 20%.The number of epochs used for model training is 10, while the learning rate is 0.001.Given that the input is audio content, the proposed algorithm leverages M Fcc features to train the model. It is then saved and reused for new patients to convert their speech audio into English transcription. The English text obtained is subsequently utilized by other modules, such as the data analytics module, to aid in identifying patient diseases and symptoms. This supports doctors in the process of diagnosis and prescription.

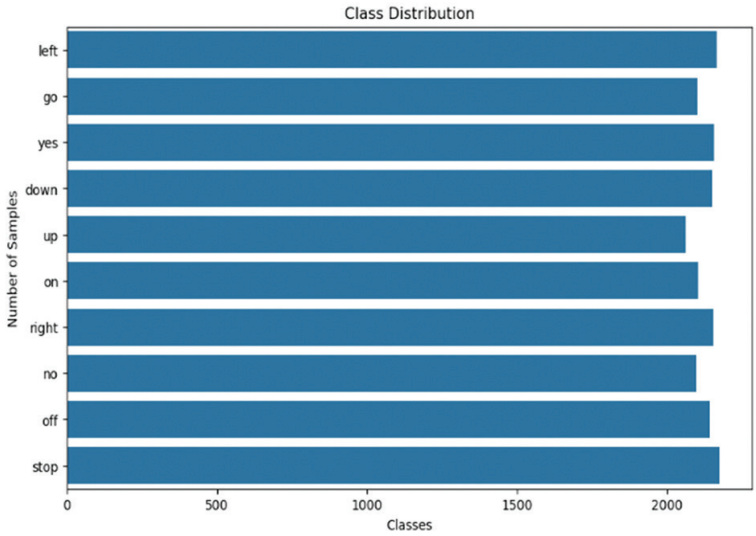

Fig. 4. Class distribution dynamics of the dataset

Fig. 4 illustrates the distribution of classes in the dataset using a bar graph. The horizontal axis represents the class counts, while the vertical axis represents the class samples. Each class in the dataset contains a tentative range of 1800 to 2200 samples. This balanced dataset provides valuable support for artificial intelligence models to learn from the data and effectively perform their tasks.

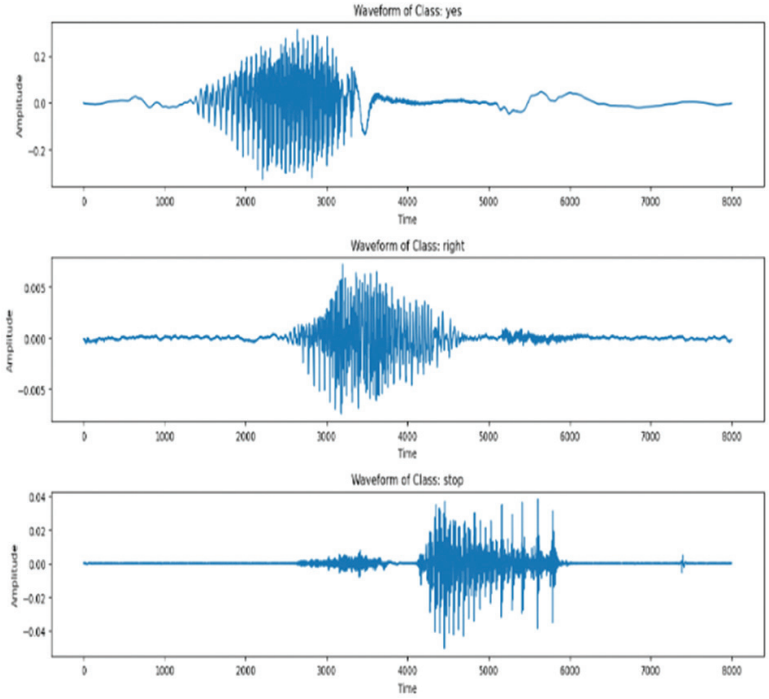

Fig. 5. Waveforms of some audio samples

In Fig. 5, waveforms of various media samples illustrate the amplitude of signals over time. Each sample shows a different amplitude waveform. These waveforms offer essential clues for deep learning models to understand language better and convert it to English text. Additionally, the waveforms provide temporal dynamics related to the audio content, giving valuable insights to artificial intelligence models that aim to capture the essence of audio samples during training and speech-to-text conversion processes.

Fig. 6 shows spectrograms of given audio samples. The proposed telemedicine system uses a hybrid deep learning model for speech-to-text conversion. Before training the model, features are extracted from the audio using spectrograms, visual representations associated with a spectrum of frequencies linked to a given audio signal. Spectrograms aid in understanding time-frequency analysis, enabling the model to convert speech audio into English text. An audio signal is a continuous waveform reflected in sound pressure vari-

ations. Short-Time Fourier Transform (STFT) is the underlying technique used to create spectrograms from given audio samples, facilitating further processing by deep learning models. Fig. 7 shows various distribution dynamics and also skewness distribution dynamics.

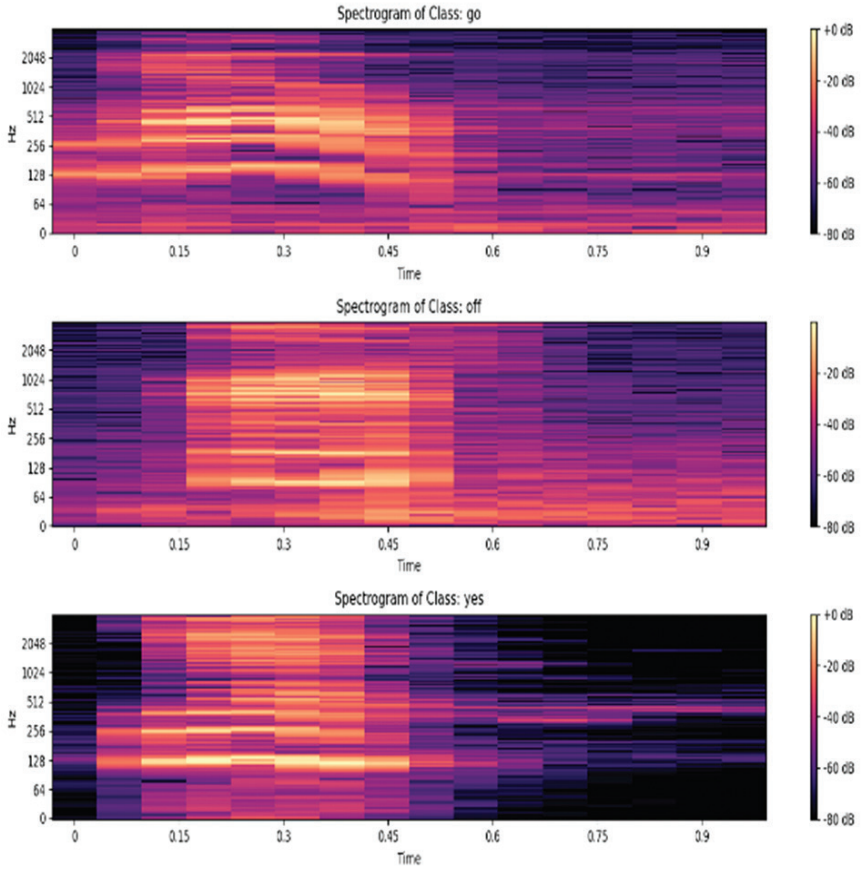

Fig. 6. Spectrogram of audio samples

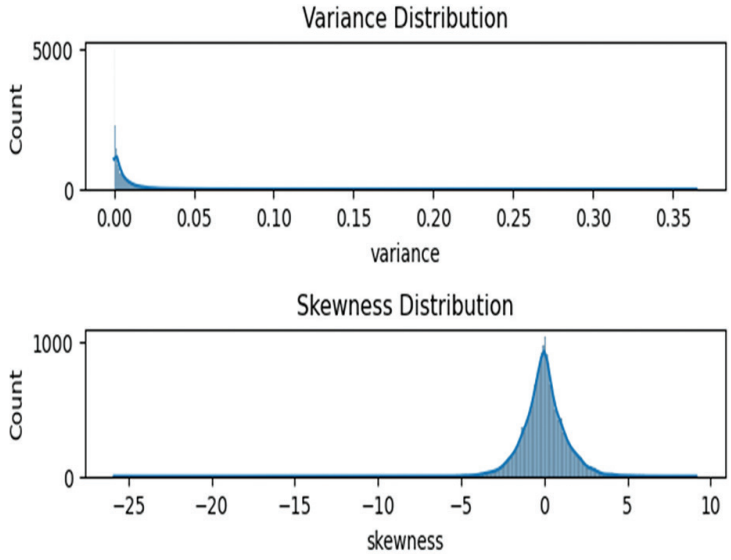

Fig. 7. Variance and skewness distribution dynamics

The variance distribution of a given audio sample visualizes how the features appear across various time windows in the frequency domain, aiding in understanding the differences in emotional tones or phonetic changes. The skewness distribution, visualized in the figure, measures asymmetry in the audio data.

A positive skewness indicates a longer tail on the right side, while a negative skewness indicates the opposite. Fig. 8 shows the proposed hybrid depth learning model's loss dynamics for STT conversion. The model's loss dynamics are shown against the number of epochs. Model loss visualization is provided for training and test

data over multiple learning cycles. The results indicate that the training performance remains consistent as the number of learning cycles increases. This means that the training loss gradually decreases until convergence as the number of epochs increases. For the test data, there are some fluctuations evident in the loss function, unlike the training data. However, there is an overall decrease in the loss as the number of epochs increases, indicating improvement in the model. It's important to note that while overall improvement in the model, there is some instability in the model's performanceconcerning the test data. Fig. 9 shows the model accuracy dynamics against several epochs.

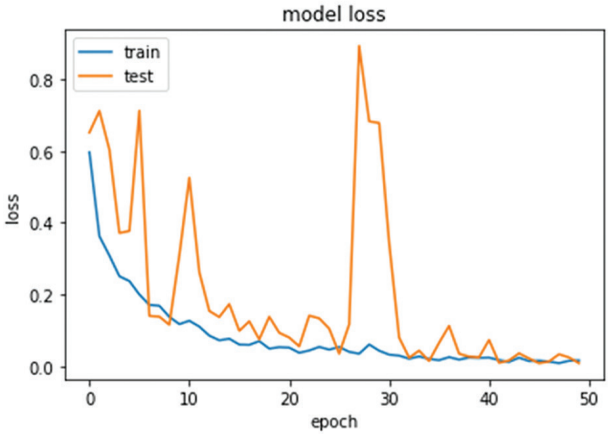

Fig. 8. The loss dynamics of the proposed deep learning model against the number of epochs

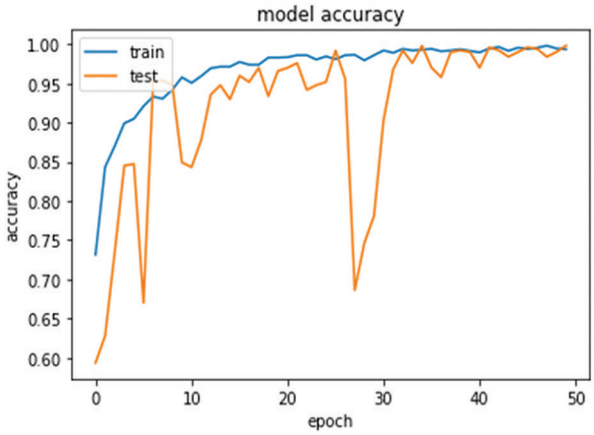

Fig. 9. Accuracy of the proposed hybrid deep learning model against several epochs

The proposed deep learning model's training and test accuracy are provided, showing that as the number of epochs increases, the accuracy of the training and test data gradually increases. The model demonstrated consistency in training accuracy, while there were fluctuations in test accuracy. Overall, the observations indicate that the proposed hybrid deep learning model consistently improves performance in terms of accuracy as the number of epochs increases until convergence. Fig. 10 compares deep learning models' performance in the STT conversion process.

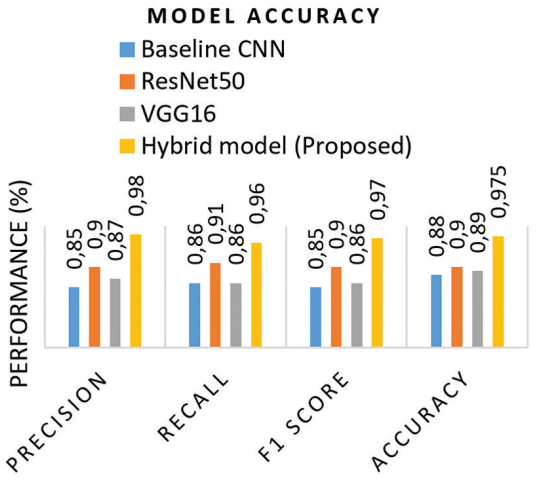

Fig. 10. Performance comparison among deep learning models in STT conversion

The hybrid deep learning model's performance is compared to state-of-the-art models. The existing deep learning models include baseline CNN and pre-trained models such as ResNet-50 and VGG-16. When all the models are used to convert the patient's speech into English, their performance is observed to be different due to their other methods of operation and as they have different layers in the deep learning process. The results show that the baseline CNN model achieved 85% precision, ResNet-50 90%, VGG-16 87%, and the proposed hybrid deep learning model achieved 98% precision. Regarding the recall measure, the baseline CNN model achieved 86%, ResNet-50 91%, VGG-16 86%, and the proposed deep learning model achieved 96% recall. Regarding the F1 score measure, the baseline CNN model achieved 85%, ResNet-50 90%, and VGG-16 86%, while the proposed deep learning model achieved a 97% F1 score. In terms of accuracy measure, the baseline CNN model achieved 88%, ResNet-50 90%, VGG-16 89%, and the proposed hybrid deep learning model achieved 97.50% accuracy. The results show that the proposed deep learning model achieved the highest accuracy, outperforming all the state-of-the-art models with 97.50%. Performance is also evaluated using the metric in Eq. 1.

$$WER = \frac{Substitutions + Insertions + Deletions}{Total\ Words} \quad (1)$$

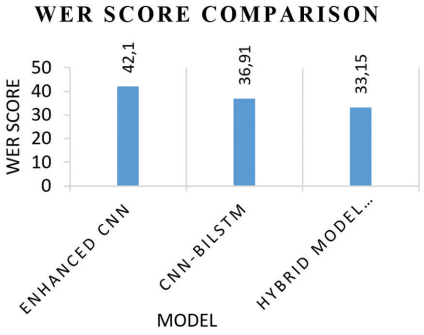

Fig. 11. WER score comparison

As presented in Fig. 11, the proposed model performs better than existing models, with a WER score of 33.15%, compared with the baseline CNN (42.1%) and CNN-BILSTM (36.91%) [38] model.

5. DISCUSSION

With the emergence of innovative technologies, telemedicine in various real-world applications is becoming increasingly important, especially in light of the recent lessons learned from the COVID-19 pandemic. The global understanding of the need for advanced healthcare systems to assist patients without requiring them to visit a hospital has become evident. Therefore, it is crucial to develop telemedicine systems that can save patients time, effort, and money for various reasons. Technological advancements such as the Internet of Things, Artificial Intelligence, and 5G technology have made exploring new avenues in remote patient monitoring and telemedicine systems possible, bringing healthcare services within easy reach with just a simple phone call. 5G technology enables seamless and efficient real-time communication between patients and doctors through wearable and non-wearable devices, including video calls. This allows doctors to listen to the patient's speech and see the patient live, enabling them to provide timely prescriptions. Patients can also have video calls, allowing the doctor to observe the patient for a more accurate diagnosis and treatment. This paper outlines a technology-driven architecture for a telemedicine system with multiple modules. Given the complexity of telemedicine systems, this paper focuses on one module - converting patient speech audio into English text, which other modules will use for data analytics to provide possible disease information and symptoms to the doctor. This will assist the doctor in taking the necessary steps for prescription and guiding the patient in overcoming their ailments. This telemedicine system can help remote patients access healthcare services without disrupting their daily activities, saving them significant time, effort, and money. While this paper proposes a technologically advanced telemedicine system, it is essential to note that itmust be fully implemented. Only the speech-to-text conversion module has been implemented, while the implementation of other modules is deferred to future work. The implemented speech-to-text conversion module has been evaluated, and certain limitations have been identified, as discussed in section 5.1.

5.1. LIMITATIONS

The speech-to-text conversion module described in this paper is based on a hybrid deep learning model. The model has been evaluated, and its results are compared with state-of-the-art models. While the proposed model demonstrates superior accuracy compared to existing models, the speech conversion module has some limitations. A significant limitation is that it has been evaluated with limited samples in the dataset. The findings can only be generalized with a diverse range of real-time patient speech samples. Another significant limitation is that the system has yet to be integrated with any existing healthcare applications used by healthcare units.

6. CONCLUSION AND FUTURE WORK

Our paper presents a framework based on deep learning that enables patients to provide information through voice. The system automatically analyzes this information and provides valuable insights on the doctor's dashboard, making diagnosis and prescriptions easier for the patient. Our proposed hybrid deep learning framework integrates with 5G technology and emphasizes speech-to-text conversion. We introduce a hybrid deep learning model to enhance performance in speech-to-text conversion. We propose an algorithm called AI-Enabled Speech Conversion (AIESC), which utilizes the improved hybrid deep learning model to convert speech to text efficiently. Using a benchmark dataset, our empirical study demonstrated that our proposed model outperforms many existing deep learning models with a 97.50% accuracy rate. In the future, we intend to improve the system by developing a method for converting patient speech to English speech, analyzing patient speech for disease diagnosis, and identifying various symptoms based on the patient's voice information. It is also desirable to integrate multiple methods involved in the telemedicine system to realize a complete and technology-driven telemedicine system that can serve remote patients without the need to visit healthcare facilities and incur significant expenses. This paradigm shift in healthcare services will be possible with an efficient telemedicine system integrated with 5G technology.

DECLARATION

FUNDING

No financial support was received by the authors in this research.

COMPETING INTERESTS

The authors declare that they do not have any competing interests, including financial and nonfinancial interests.

ETHICAL APPROVAL

This research does not involve humans or animals, so no ethical approval is required.

CONSENT FOR PUBLICATION

The authors give consent for their publication.

DATA AVAILABILITY

Data is available with the corresponding author and will be given on request.

AUTHOR CONTRIBUTION

All authors contributed to the study's conception and design. Medapati Venkata Manga Naga Sravan and Prof K Venkata Rao performed material preparation, data collection, and analysis. Medapati Venkata Manga Naga Sravan wrote the first draft of the manuscript. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

7. REFERENCES

[1] T.W. Lin, C. L. Hsu, "FAIDM for Medical Privacy Protection in 5G Telemedicine Systems", Applied Sciences, Vol. 11, No. 3, 2021, p. 1155.

[2] K. Hameed, I. S. Bajwa, N. Sarwar, W. Anwar, Z. Mushtaq, T. Rashid, "Integration of 5G and Block-Chain Technologies in Smart Telemedicine Using IoT", Journal of Healthcare Engineering, Vol. 2021, 2021, pp. 1-18.

[3] Y. Wang, P. Tran, J. Wojtusiak, "From Wearable Device to OpenEMR: 5G Edge Centered Telemedicine and Decision Support System", Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies, Vol. 5, 2022, pp. 491-498.

[4] H. Yu, Z. Zhou, "Optimization of IoT-Based Artificial Intelligence Assisted Telemedicine Health Analysis System", IEEE Access, Vol. 9, 2021, pp. 85034-85048.

[5] T. A. Suleiman, A. Adinoyi, "Telemedicine and Smart Healthcare - The Role of Artificial Intelligence, 5G, Cloud Services, and Other Enabling Technologies", International Journal of Communication, Network and System Sciences, Vol. 1, 2023, pp. 31-51.

[6] A. Sadia, P. Ramjee, "Framework for Future Telemedicine Planning and Infrastructure using 5G Technology", Wireless Personal Communications, Vol. 100, 2018, pp. 193-208.

[7] C. Li et al. "Telemedicine network latency management system in 5G telesurgery: a feasibility and effectiveness study", Surgical Endoscopy, Vol. 38, 2023, pp. 1592-1599.

[8] T. W. Lin, C. L. Hsu, T. V. Le, C. F. Lu, B. Y. Huang, "A Smartcard-Based User-Controlled Single Sign-On for Privacy Preservation in 5G-IoT Telemedicine Systems", Sensors, Vol. 21, No. 8, 2021, p. 2880.

[9] T. Hewa, A. Braeken, M. Ylianttila, M. Liyanage, "Multi-Access Edge Computing and Blockchain-based Secure Telehealth System Connected with 5G and IoT", Proceedings of GLOBECOM 2020 - 2020 IEEE Global Communications Conference, Taipei, Taiwan, 7-11 December 2020.

[10] J. Lu, X. Wang, X. Zeng, W. Zhong, W. Han, "Application of telemedicine system on the management of general patient in quarantine", Scientific Reports, Vol. 13, 2023, p. 12215.

[11] L. Chettri, "Design and Implementation of 5G Wireless Telemedicine Systems", International Journal of Information Studies, Vol. 8, No. 2, 2016, pp. 45-53.

[12] B. W. Diong, M. I. Goh, S. K. Chung, A. Chekima, H. T. Yew, "Vertical Handover Algorithm for Telemedicine Application in 5G Heterogeneous Wireless Networks", International Journal of Advanced Computer Science and Applications, Vol. 12, No. 8, 2021, pp. 611-617.

[13] D. Sadia, P. Anand, A. Awais, R. Seungmin, "Emerging Mobile Communication Technologies for Healthcare System in 5G Network", Proceedings of the IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, 14th International Conference on Pervasive Intelligence and Computing, 2nd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress, Auckland, New Zealand, 8-12 August 2016, pp. 47-54.

[14] R. C. Figueiredo, N. S. Ribeiro, A. M. O. Ribeiro, Cri, "5G-Enabled Health Systems: Solutions, Challenges and Future Research Trends", Journal of Lightwave Technology, Vol. 33, No. 1, 2019, pp. 1-9.

[15] C. O. Alenoghena, H. O. Ohize, A. O. Adejo, "Telemedicine A Survey of Telecommunication Technologies, Developments, and Challenges", Journal of Sensor and Actuator Networks, Vol. 12, No. 2, 2023, p. 20.

[16] T. W. Lin, "A Privacy-Preserved ID-Based Secure Communication Scheme in 5G-IoT Telemedicine Systems", Sensors, Vol. 22, No. 18, 2022, p. 6838.

[17] E. C. Liou, S. C. Cheng, "A QoS Benchmark System for Telemedicine Communication Over 5G uRLLC and mMTC Scenarios", Proceedings of the IEEE 2nd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability, Tainan, Taiwan, 29-31 May 2020, pp. 24-26.

[18] A. Adarsh, S. Pathak, B. Kumar, "Design and Analysis of a Reliable, Prioritized and Cognitive Radio-Controlled Telemedicine Network Architecture for Internet of Healthcare Things", International Journal of Computer Networks and Applications, Vol. 8, No. 1, 2021, pp. 54-66.

[19] R. Colella, L. Catarinucci, "10.525 GHz Backscattering RFID System Based on Doppler Radar Technology for 5G Applications and Telemedicine", Proceedings of the Photonics & Electromagnetics Research Symposium, Rome, Italy, 17-20 June 2019, pp. 3689-3695.

[20] R. D. Silva, Y. Siriwardhana, T. Samarasinghe, M. Ylianttila, M. Liyanage, "Local 5G Operator Architecture for Delay Critical Telehealth Applications", Proceedings of the IEEE 3rd 5G World Forum, Bangalore, India, 10-12 September 2020, pp. 257-262.

[21] G. G. Mihuba, M. J. Shundi, O. K. Bishoge, An, "Design of Telemedicine System Based on Mobile Terminal", International Journal of Emerging Technologies in Engineering Research, Vol. 7, No. 1, 2019, pp. 1-8.

[22] P. Bailo, F. Gibelli, A. Blandino, A. Piccinini, "Telemedicine Applications in the Era of COVID-19 Telesurgery Issues", International Journal of Environmental Research and Public Health, Vol. 19, No.1, 2022, p. 323.

[23] L. Chettri, R. Bera, "Design and Analysis of 5G Telemedicine Systems", Mobile and Forensics, Vol. 3, No. 2, 2021, pp. 48-57.

[24] B. Arunsundar, R. Srinivasan, "Implementation of Obiquitous Telemedicine on 5g Networks", International Journal of Advanced Research Trends in Engineering and Technology, Vol. 2, No. 1, 2015, pp. 1-5.

[25] A. M. Peralta-Ochoa, P. A. Chaca-Asmal, L. F. Guerrero-Vásque, "Smart Healthcare Applications over 5G Networks: A Systematic Review", Applied Sciences, Vol. 13, No. 3, 2023, p. 1469.

[26] M. Cabanillas-Carbonell, J. Pérez-Martínez, J. A. Yáñez, "5G Technology in the Digital Transformation of Healthcare: A Systematic Review", Sustainability, Vol. 15, No. 4, 2023, p. 31278.

[27] A. S. Albahri, J. K. Alwan, Z. K. Taha, S. F. Ismail, R. A. Hamid, A. A. Zaidan, O. S. Albahri, B. B. Zaidan, A. H. Alamoodi, M. A. Alsalem, "IoT-based telemedicine for disease prevention and health promotion: State-of-the-Art", Journal of Network and Computer Applications, Vol. 173, 2021, pp. 1-59.

[28] R. W. Ahmad, K. Salah, R. Jayaraman, I. Yaqoob, S. Ellahham, M. Omar, "The role of blockchain technology in telehealth and telemedicine", International Journal of Medical Informatics, Vol. 148, 2021.

[29] H. Jain, V. Chamola, Y. Jain, Naren, "5G network slice for digital real-time healthcare system powered by network data analytics", Internet of Things and Cyber-Physical Systems, Vol. 1, 2021, pp. 14-21.

[30] A. Sadia, P. Ramjee, B. S. Chowdhary, M. R. Anjum, "A Telemedicine Platform for Disaster Management and Emergency Care", Wireless Personal Communications, Vol. 106, No. 1, 2019, pp. 191-204.

[31] A. Adford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever, "Robust Speech Recognition via Large-Scale Weak Supervision", Proceedings of the 40th International Conference on Machine Learning, Honolulu, HI, USA, 23-29 July 2023, pp. 28492-28518.

[32] L. Orynbay, B. Razakhova, P. Peer, B. Meden, E. Žiga, "Recent Advances in Synthesis and Interaction of Speech, Text, and Vision. Electronics", Electronics, Vol. 13, No. 9, 2024, p. 1726.

[33] A. Dhakad, S. Singh, "Python-Powered Speech-to-Text: A Comprehensive Survey and Performance Analysis", International Journal of Engineering Research & Technology, Vol. 12, No. 9, 2023.

[34] V. M. Reddy, T. Vaishnavi, K. P. Kumar, "Speech-to-Text and Text-to-Speech Recognition Using Deep Learning", Proceedings of the 2nd International Conference on Edge Computing and Applications, Namakkal, India, 19-21 July 2023.

[35] N. Sethiya, C. K. Maurya, "End-to-End Speech-to-Text Translation: A Survey", arXiv:2312.01053, 2024.

[36] V. Korchynskyi, I. Vynogradov, "Methods of Improving the Quality of Speech-to-Text Conversion", Scientific Collection InterConf, Vol. 194, 2024, pp. 426-437.

[37] M. Ahmad Dar, J. Pushparaj, "Hybrid Architecture CNN-BLSTM for Automatic Speech Recognition", Proceedings of the 3rd International Conference on Artificial Intelligence for Internet of Things, Vellore, India, 3-4 May 2024, pp. 1-15.

[38] A. Baevski, Y. Zhou, A. Mohamed, M. Auli, "Wav-2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations", Proceedings of Advances in Neural Information Processing Systems, Vol. 33, 2020, pp. 12449-12460.

[39] TensorFlow Speech Recognition Challenge dataset, https://www.kaggle.com/c/tensorflow-speech-recognition-challenge/data (accessed: 2024)