TelMedAI: A Framework for Patient Speech Recognition and Conversion into Desired Language Towards Telemedicine System

Original Scientific Paper

MrudulaOwk*
GITAM University,
Department of CSE, GITAM School of Technology
Rushikonda, Visakhapatnam -530045
mowk@gitam.edu

Pusarla Sindhu
GITAM University,
Department of CSE, GITAM School of Technology
Rushikonda, Visakhapatnam -530045
spusarla@gitam.edu

Deepthi Godavarthi
VIT-AP University,
School of Computer Science and Engineering(SCOPE)
Amaravati, Andhra Pradesh, India- 522237
deepthi.g@vitap.ac.in

T. Krishna Mohana
Department of ECE Aditya College of Engineering
Krishnamohana_ece@acoe.edu.in

*Corresponding author

Abstract *– Telemedicine is the practice of technology-enabled remote communication between patient and doctor. This phenomenon in healthcare has the potential to make services affordable and save time and money. Besides telemedicine allows care givers and family members to join conversations with doctors. Indian government initiated the National Telemedicine Network (NTN) to serve remote areas in healthcare by integrating existing healthcare facilities.Literature has revealed that existing works lack in an integrated approach for patient speech translation in language-independent fashion and automatic detection of disease and symptoms based on speech.There is a need for an automated system using Artificial Intelligence (AI) to recognize patient's speech and identify symptoms based on given audio description. We proposed a framework known as* ***TelMedAI*** *which is designed to recognize patient speech to comprehend disease symptoms besides converting the speech text into desired language. The framework is useful for realizing a telemedicine system. Speech to Speech (STS) module takes the patient's audio content into English audio. STS module exploits the Bi-LSTM model with an encoder, decoder and attention mechanism for translation. Then Google Speech API is used to convert English audio into English text. Then the framework exploits Natural Language Processing (NLP) to improve the quality of text. Afterwards, the disease and symptoms miner module eventually recognizes a list of diseases and corresponding symptoms. We proposed an algorithm known as Learning based Disease and Symptom Recognition from Patient Speech (LbDSRPS). This algorithm has the functionality to develop* ***TelMedAI*** *which helps doctors in telemedicine. Our empirical study has revealed that* ***TelMedAI*** *takes technology-driven telemedicine research forward significantly. The highest accuracy achieved by the proposed framework is 68.13% which is much better than the baseline LSTM model used for voice translation.*

Keywords: *Telemedicine System, Patient Speech Recognition, Deep Learning, Artificial Intelligence, Multi-Lingual Text Conversion*

Received: November 20, 2023; Received in revised form: April 30, 2024; Accepted: April 30, 2024

1. INTRODUCTION

Research on the realization of telemedicine systems with technology-enhanced approaches is most relevant in the contemporary world. The recent COVID-19 pandemic has emphasized the importance of telemedicine. Telemedicine is the system which enables people of all walks of life to gain access to healthcare services just by making a phone call. This phenomenon has many significant advantages such as saving time, effort, and money besides getting medical advice without time and geographical restrictions. In order to develop a telemedicine system, it is important to have efficient translation services. The rationale behind this is that the patient may speak any language and doctor needs to understand what the patient says while doctor does not know the language of the patient. Therefore, telemedicine has many implementation challenges as explored in [1]. Nevertheless, there has been research endeavours to realize technology-driven telemedicine systems as found in the literature.

Machine learning models are explored in [2] and [3] towards understanding patient diseases and symptoms based on the description given by persons. Speech recognition research carried out as found in [4-8] provide the basis for understanding the importance of patient speech recognition in telemedicine besides how the existing methods could achieve it. Voice recognition in an IoT-integrated home automation system is investigated in [9] while a similar kind of approach is used in [10] as part of remote patient monitoring. Researchers explored telemedicine with novel technologies as discussed in [1, 11-13]. The realization of telemedicine has many challenges as discussed in [11]. One significant challenge is the voice-to-voice conversion from source to target language. Artificial intelligence and its usage in telemedicine is studied in [12]. Though there are innovative technologies, it is observed in [1] that telemedicine implementation is still challenging due to several complexities involved. Patient engagement early and disease diagnosis through the telemedicine system is investigated in [14]. From the literature, it is found that there is need for a system that can translate patient voices into English and mine diseases and symptoms to help doctors realise a telemedicine system. Towards this end, we proposed an AI enabled system for patient speech translation into desired target language and analyse patient's health symptoms and diseases. The novelty of the proposed system is it is designed to be language-independent and helps in analysing and identifying patient's disease and symptoms to assist doctors. The proposed system plays crucial role in realizing a technology-driven telemedicine system. Our contributions to this paper are as follows. We proposed a framework known as **TelMedAI** which is designed to recognize patient speech to comprehend disease symptoms besides converting the speech text into desired language and recognize the patient's disease and symptoms to develop the telemedicine system. We proposed an algorithm known as learning-based Disease and Symptom Recognition from Patient Speech (LbDSRPS) to develop**TelMedAI** which helps doctors in telemedicine. We built an application for evaluating **TelMedAI** and the underlying algorithm. The utility of bi-LSTM model used in the STS module is evaluated and the results are compared with baseline LSTM model.The remainder of the paper is structured as follows. Section 2 reviews the literature on existing research efforts using machine learning in the field of telemedicine. Section 3 presents the proposed framework along with its modus operandi in detail. Section 4 presents experimental results while section 5 concludes our work and provide limitations of the current work and directions for future research.

2. RELATED WORK

This section reviews the literature on existing research endeavours on the usage of techniques towards telemedicine. Zahia et al. [2] opines that pressure injuries burden healthcare systems. Non-invasive imaging, including Deep Learning, aids accurate assessment, but limited data hinders progress. Their investigation has revealed the need for non-invasive healthcare services that could be beneficial to general public. Dargan et al. [3] explored deep learning to show versatility and progress across various fields. Challenges include optimizing hierarchies and maintaining databases. The insights of their research reveal that deep learning models are very useful in computer vision applications. Stanovov et al. [4] proposed a cloud-based speech-controlled wheelchair system, emphasizing the benefits of multiple cloud speech recognition APIs for accuracy. Recognition of speech could help in solving many real world problems such as directing navigation of a wheel chair. Sescleiferet al. [5] focused on online crowdsourcing which aids in efficient and effective perceptual speech assessments, particularly beneficial for cleft palate surgeries.Speech perception plays crucial role in certain applications such as telemedicine where patient's speech needs to be understood accurately. Ivan et al. [9]. An IoT-enabled voice recognition system is explored. Kho et al. [11] observed that telemedicine faces implementation challenges due to inadequate attention to change management. A process-oriented approach is recommended. From their work is understood that there are gaps in existing telemedicine systems. One such gap is lack of usage of AI. Olivia et al. [12] investigated on advancements in digital technologies that offer transformative potential for ophthalmology, notably amid the COVID-19 pandemic, yet challenges persist. In their research, there is inference that improving end-to-end communications in healthcare considering every possibility helps in developing more robust applications. Spachos et al. [15] observed that the pandemic spurred changes in interactions with objects. Voice-activated IoT devices have healthcare potential, yet security concerns persist. Voice based communications in telemedicine system needs technology adoption to leverage seamless telemedicine services.

Lesso et al. [16] opined that telehealth in respiratory care lacks efficient cough detection. Proposed Hu moments-based system offers high sensitivity and specificity. However, it is observed that there is need for enhancing the system to have generalized system for all kinds of diseases. Erdene et al. [17] observed that mobile health technologies aid in early stroke detection through continuous monitoring with diverse sensor-equipped devices. Such systems are useful for alerting people of health issues. However, the availability of telemedicine can expect possibilities in rendering healthcare services. Jamshidiet et al. [18] stated that COVID-19's global impact necessitates AI-driven solutions for diagnosis and treatment acceleration, integrating varied data sources for effective platforms. Covid pandemic also necessitated and reinforced the need for technology driven approach where people can have health services over phone. Panganibanet et al. [19] a deep learning model is evaluated for diagnosis of a patient's disease.Since deep learning models are found good at learning audio signals, using them in development of a telemedicine system could have a positive impact. Wang et al. [20] observed that COVID-19 intensified the role of robotics in healthcare. Despite advancements, challenges like cost and accessibility persist. Such challenges can be overcome with an efficient telemedicine system for some of the health issues. Dharmale et al. [21] exploited phonetic system in healthcare industry for leveraging speech recognition system. Choutri et al. [22] focused on human-drone interaction where speech is recognized of multiple languages. Kerwagen et al. [23] investigated on diagnostic management along with speech recognition and usability in healthcare. Shindel et al. [24] studied healthcare applications integrated with blockchain technology. Javaid et al. [25] explored machine learning and its utility in healthcare domain besides providing valuable insights. From the literature, it is found that there is a need for a system that can translate patient voice into English and mine diseases and symptoms for helping doctors realise a telemedicine system.

3. PROPOSED FRAMEWORK

We proposed a framework known as **TelMedAI** which is designed to recognize patient speech to comprehend disease symptoms besides converting the speech text into desired language. The framework is useful for realizing a telemedicine system. Overview of **TelMedAI** is shown in Fig.1. The source speaker (patient) may speak over the telephone in any language. If that language is not English, the proposed framework helps in converting a voice in different language to a voice in English. This conversion is known as Speech to Speech (STS) conversion. The STS module in the proposed system exploits the deep learning model LSTM for converting patient speech (voice) into English (voice). Fig. 2 illustrates how the STS conversion module does it. Once the patient's voice is converted to voice in English, it is one of the outcomes of the system which can be listened by the doctor directly. Further **TelMedAI** has provision for converting speech in English to English text. This conversion is carried out by speech to text conversion module of the framework.

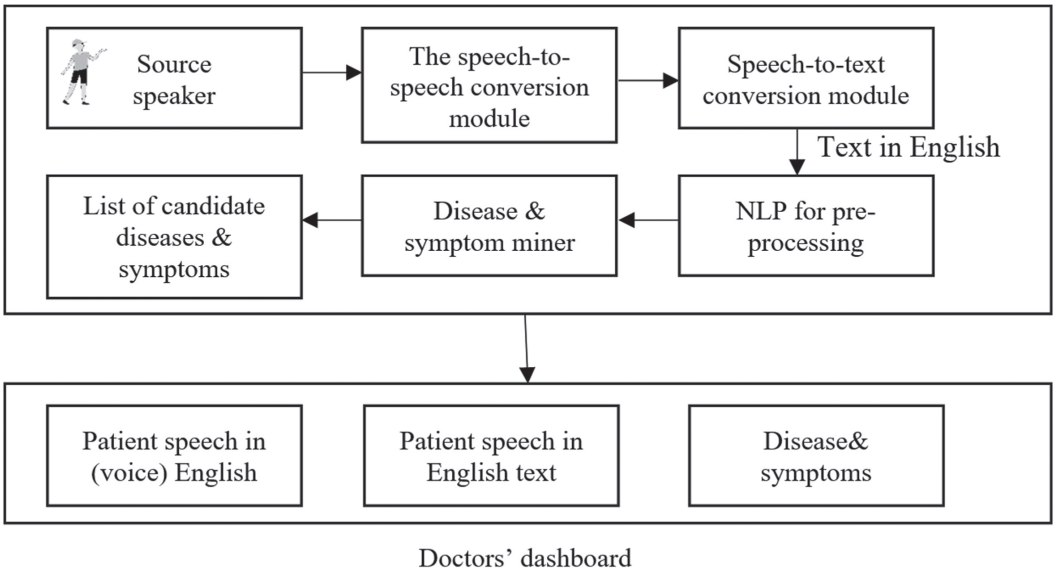


Fig. 1. Overview of the proposed framework known as TelMedAI useful for telemedicine

Text in English is another desired output useful to doctor in the telemedicine system. Afterwards, the text in English is subjected to pre-processing using NLP followed by text mining to discover disease and symptoms found in the patient's speech. The rationale and motivation behind three outcomes in doctor's dashboard is that it will enable doctor to revisit and correlate findings. Therefore, the third outcome of the system useful for doctor is the identified disease and its symptoms. Since voice input is translated into English, only textual data is analysed for identification of disease and symptoms. The STS conversion module makes use of bi-directional LSTM used as an encoder and decoder network. We preferred the LSTM model as order of the inputs and outputs is important. In other words, LSTM is well known for its ability to function in the temporal domain. Using the Fourier transform, the given patient's voice is converted to a Mel spectrogram of speech in the original language. This high level representation of the patient's voice is given to the Bi-LSTM model to convert data into linear spectrograms of speech in English (the target language). The spectrogram of speech of English is subjected to Griffin Lim Vocoder to convert the high-level representation to target voice file.

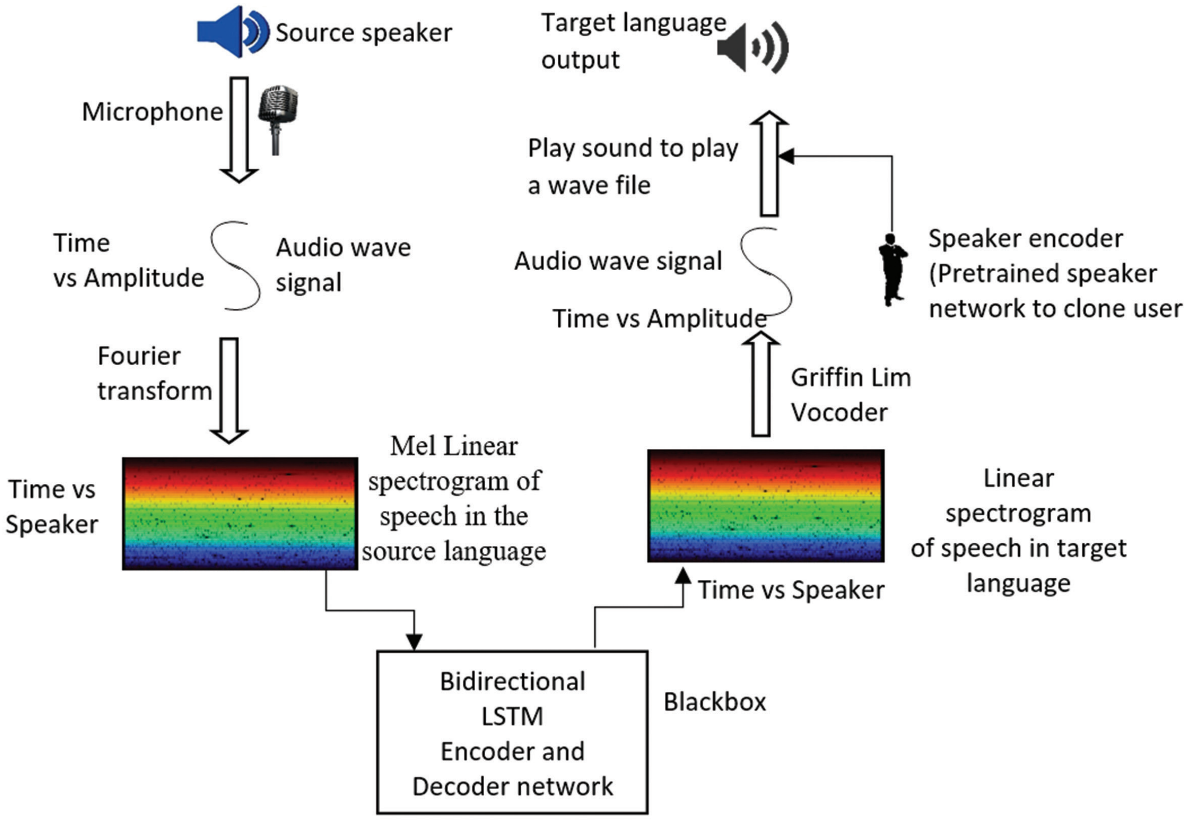


Fig. 2. Functional flow of STS conversion module

Once the source voice is converted to the target voice file, speech to text conversion module comes into picture. This module makes use of Google Speech API to translate voice in English to text in English. Before elaborating further process involved in the TelMedAI, we describe the encoder and decoder functionality involved in the STS conversion module. The basic encoder and decoder functionality of LSTM is shown in Fig. 3.

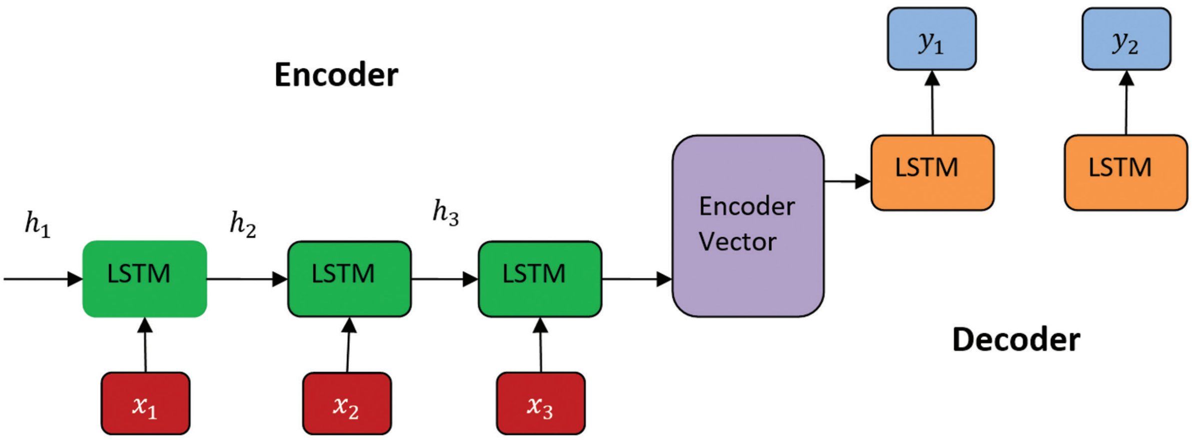


Fig. 3. LSTM based encoder and decoder network

The encoder is made up of three recurrent LSTM units while the decoder is made up of two recurrent LSTM units. Each unit acts on the given input and propagates output to the next LSTM unit. The encoder computes hidden states as expressed in Eq. 1 where h_t denotes hidden state in given time step, W refers to weight matrix.

$$h_t = f(W^{hh}h_{t-1} + W^{hx}x_t) \quad (1)$$

It involves the multiplication of weights associated with the previous hidden state and the input vector. Concerning decoder, computation of the hidden state is done as expressed in Eq. 2 where h_t denotes hidden state in given time step, W refers to weight matrix.

$$h_t = f(W^{hh}h_{t-1}) \quad (2)$$

Computation of final output in the absence of any activation function is carried out as expressed in Eq. 3 where y refers to output, W indicates weight matrix and t refers to given time step.

$$y_t = Softmax(W^s h_t) \quad (3)$$

It reflects output in the form of a multiplication matrix derived from a weight matrix associated with hidden vector in the given time step. An important consideration is the attention mechanism which has the potential to improve the network as it lies between encoder and decoder. The attention mechanism gets rid of any possible misalignment between the encoder and decoder. The Bi-LSTM-based block box shown in Fig. 2 plays an important role in converting patient's voice from one language to another language (STS). This process is further elaborated as illustrated in Fig. 4. Stacking LSTM units is done to develop the encoder which produces outputs form Mel spectrograms of the patient's voice in native language. Then the output is concatenated. Afterwards, an attention mechanism is used to optimize outcomes to be given to the decoder in order to avoid error. Then a stacked LSTM unit is used as decoder followed by a dense layer and Griffin Lim Vocoder to generate Mel spectrograms of target language. Here Griffin Lim Vocoder algorithm is used to generate target spectrograms. The attention mechanism used in the proposed system is multi-head attention as it is run many times instead of computing attention once.

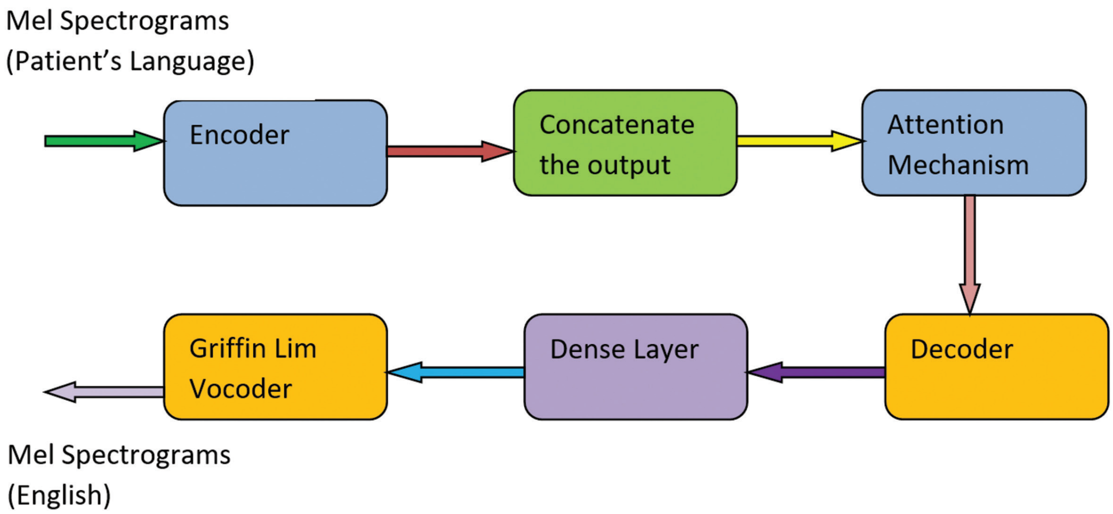


Fig. 4. Technical details of the pipeline involved in the STS module

The dense layer involved in the network is itself a well-connected neural network. It is used to generate outcomes as expressed in Eq. 4.

$$Output= activation\ (d\ ot(input,\ kernel)+bias) \quad (4)$$

It makes use of weight kernel matrix and input tensor to have a dot product which is nothing but second LSTM unit's output in the decoder of the network. Here the bias is optional and set to zero. MSE is used as a loss function to minimize error in computations. We also used Adam optimizer to improve accuracy of the model. Now let us get back to the framework **TelMedAI** presented in Fig. 1. After converting from English speech to text using Google Speech API, the resultant text is subjected to NLP for pre-processing. It is used to get rid of meaningless words in the speech. Afterwards, the disease and symptom miner is responsible for discovering disease (s) and corresponding symptoms. This module provides the desired convenience to doctors as it provides list of candidate diseases and corresponding symptoms. Our implementation of the disease and symptom miner module is influenced by the work of [21] where more technical information about how the diseases and the symptoms are identified from text in English. As presented in Algorithm 1, it takes patient speech as input. It is the voice of patient describing about his/her disease to seek doctor's advice through telemedicine system. After completion of the processing, the proposed algorithm results in three outputs that are useful to doctor in disease diagnosis and treatment. They are known as translated English voice files, English text and recognized disease and symptoms. The given patient voice is used to generate a Mel spectrogram. This is the high-level representation of patient's voice information sent to doctor. Then the algorithm makes use of Bi-LSTM with encoder, decoder along with attention mechanism to convert Mel spectrogram of source language voice into the Mel spectrogram of target language that is English.

Algorithm 1: Learning based disease and symptom recognition from patient speech

Algorithm: Learning-based Disease and Symptom Recognition from Patient Speech

Input

Patient speech (voice file)ov

Outputs

Translated English voice file tv

English text tt

recognized disease and symptoms R

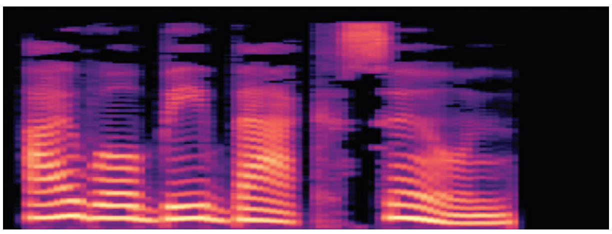
1. Begin
2. S←GenerateMelSpectogram(ov)
3. Build Bi-LSTM model m

4. Compile the model *m*
5. *T*←*GenerateTargetMelSpectogram*(*m, S*)
6. *tv*←*GenerateEnglishVoiceFile* (*T, Griffin Lim Vocoder algorithm*)
7. *tt*←*GoogleSpeechAPI*(*tv*)
8. *tt'*←*NLPTechniques*()
9. *R*←*DiseaseAndSymptomMiner*(*tt'*)
10. Display *tv*
11. Display *tt*
12. Display *R*13.
13. End

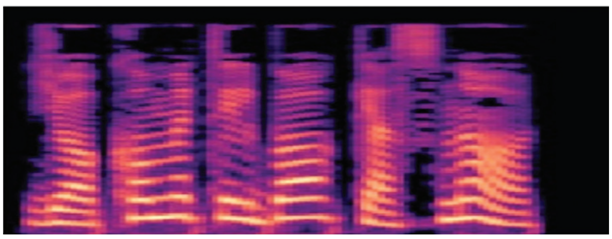
The target Mel spectrogram is given to the Griffin Lim Vocoder algorithm which generates English voice file. This process is known as STS (speech to speech) conversion. Once STS is completed, its resultant English voice file is converted to English text using Google Speech API. The resultant English text contains information provided by the patient to doctor but translated to English. This text is subjected to NLP techniques such as stop word removal, stemming and lemmatization to improve quality of the text. Afterwards, the disease and symptoms miner module is used to discover disease and corresponding symptoms from English text.

4. EXPERIMENTAL RESULTS

We implemented our framework TelMedAI using Python language, machine learning library and Google Speech API. This section presents experimental results in terms of patient voice converting to Mel spectrograms or source and target (English) languages. It also provides the accuracy of deep learning model proposed in this paper and compare the same with baseline LSTM model. Fig. 5 presents Mel spectrogram in source and target languages.



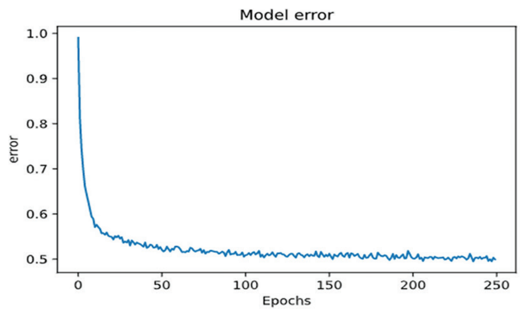
(a)



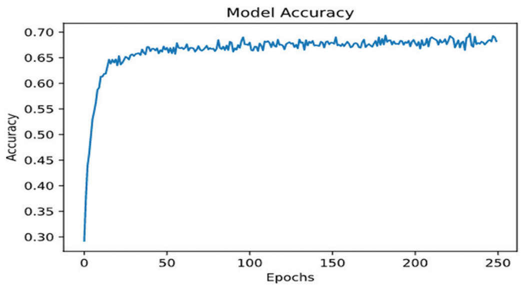
(b)

Fig. 5. Mel Spectrograms (a) for English sentence "I am suffering from fever" and (b) equivalent in Swedish

Patient's speech saying "Jag harbour" in Swedish language is converted to source language Mel Spectrogram and target language (English) Mel Spectrogram. English equivalent of the patient's speech is "I am suffering from fever".



(a)



(b)

Fig. 6. Performance in terms of (a) error and (b) accuracy of the Bi-LSTM with attention in the proposed system

The model performance in terms of error and accuracy is visualized against number of epochs as shown in Fig. 6. It is the result of Bi-LSTM with attention in the proposed system. Table 1 shows performance comparison.

Table 1. Shows performance comparison

Model	Accuracy
Baseline LSTM	62.24
Bi-LSTM with Attention (proposed system)	68.13

Accuracy of the Bi-LSTM model with the attention mechanism used in the proposed framework is compared with that of a baseline LSTM model.

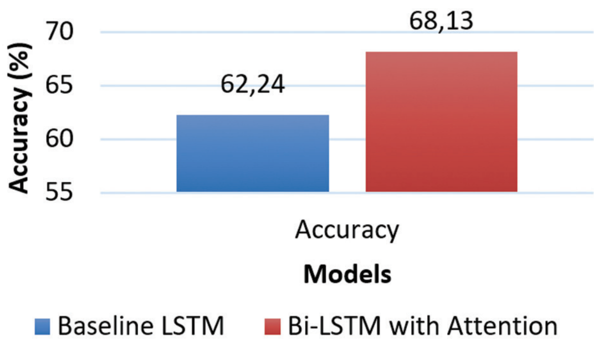


Fig. 7. Performance comparison between the proposed and existing models

Performance of the deep learning model used in the proposed system (Bi-LSTM with attention) is compared with the baseline LSTM model, as shown in Fig. 7, in terms of accuracy in translating the patient's source language voice to English voice. The proposed model in the paper outperformed the baseline model. Highest accuracy achieved by the proposed framework is 68.13%. Considering the difficulty in STS conversion, this accuracy is significant. However, the proposed system has several limitations that need to be overcome in future. First, it is implemented as a preliminary system and it is no way perfect to be used in a telemedicine system without further improvement. Second, the proposed system is yet a laboratory study which takes pre-recorded patient voice as input. However, it needs to be improved to consider live voice of patients. Third, as of now, the system is yet to be integrated with a telephone line to capture patient's voice and evaluate the functionality. Fourth, the proposed system has to be improved further to deploy in a healthcare unit where doctors can take patient calls to understand their health issues, diagnose and give treatment.

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a framework known as **TelMedAI** which is designed to recognize patient speech to comprehend disease symptoms besides converting the speech text into desired language. The framework is useful for realizing a telemedicine system. Speech to Speech (STS) module takes patient's audio content into English audio. STS module exploits the Bi-LSTM model with encoder, decoder and attention mechanism for translation. The deep learning model is used to convert source language Mel Spectrogram into English Mel Spectrogram. The target Mel spectrogram is given to Griffin Lim Vocoder algorithm which generates English voice file. Then Google Speech API is used to convert English audio into English text. Then the framework exploits Natural Language Processing (NLP) to improve quality of text. Afterwards disease and symptoms miner module eventually recognizes a list of diseases and corresponding symptoms. We proposed an algorithm known as Learning based Disease and Symptom Recognition from Patient Speech (LbDSRPS). This algorithm has the functionality to develop **TelMedAI** which helps doctors in telemedicine.

Our empirical study has revealed that TelMedAI takes technology-driven telemedicine research forward significantly. As of now, our system is tested with patients' voice files. However, to develop a complete telemedicine system, there is need for much work to be one. In future, we intend to improve our system in two phases. In the first phase, we evaluate it with live patient's voice and in the second phase we deploy it in a healthcare unit for live testing and further improvement. Improving accuracy is also to be considered in future.

6. REFERENCES

[1] A. Khodadad-Saryazdi, "Exploring the telemedicine implementation challenges through the process innovation approach: A case study research in the French healthcare sector", Technovation, Vol. 107, 2021, p. 102273.

[2] S. Zahia, G. Zapirain, M. Begoña, S. Xavier, A. González, P. J. Kim, A. Elmaghraby, "Pressure injury image analysis with machine learning techniques: A systematic review on previous and possible future methods", Artificial Intelligence in Medicine, Vol. 102, No. C, 2020.

[3] S. Dargan, M. Kumar, M. R. Ayyagari, G. Kumar, "A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning", Archives of Computational Methods in Engineering, Vol. 27, 2019, pp. 1071-1092.

[4] A. Koložvari, R. Stojanović, A. Zupan, E. Semenkin, V. Stanovov, D. Kofjač, A. Škraba, "Speech-Recognition Cloud Harvesting for Improving the Navigation of Cyber-Physical Wheelchairs for Disabled Persons", Microprocessors and Microsystems, Vol. 69, No. C, 2019, pp. 179-187.

[5] A. M. Sescleifer, C. A. Francoisse, A. Y. Lin, "Systematic review: online crowdsourcing to assess perceptual speech outcomes", Journal of Surgical Research, Vol. 232, 2018, pp. 351-364.

[6] B. F. Zaidi, S. A. Selouani, M. Boudraa, M. S. Yakoub, "Deep neural network architectures for dysarthric speech analysis and recognition", Neural Computing and Applications, Vol. 33, 2021, pp. 9089-9108.

[7] P. T. Bradley, J. Patterson, "Attitudes to the Implementation of Speech and Language Therapist Led Low Risk Two Week Wait Clinic in the UK: A Survey Exploration Using Normalization Process Theory", Journal of Voice, Vol. 38, No. 1, 2021, pp. 86-95.

[8] C. Millar, L. B. Carey, A. E. Hill, S. Attrill, M. I. Avgoulas, E. Drakopoulos, C. A. Sutton, "Global citizenship and social justice among speech-language pathologists: A scoping review", Journal of Communication Disorders, Vol. 103, 2023, pp. 1-11.

[9] I. F. Míguez, P. F. Lamas, T. M. F. Caramés, "Design, Implementation, and Practical Evaluation of a Voice Recognition Based IoT Home Automation

System for Low- Resource Languages and Resource-Constrained Edge IoT Devices: A System for Galician and Mobile Opportunistic Scenarios", IEEE Access, Vol. 11, 2023, pp. 63623-63649.

[10] M. Talal et al. "Smart Home-based IoT for Real-time and Secure Remote Health Monitoring of Triage and Priority System using Body Sensors: Multi-driven Systematic Review", Journal of Medical Systems, Vol. 43, No. 3, 2019.

[11] J. Kho, N. Gillespie, M. K. Melinda, "A systematic scoping review of change management practices used for telemedicine service implementations", BMC Health Services Research, Vol. 20, No. 1, 2020.

[12] J.-P. O. Li et al. "Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective", Progress in Retinal and Eye Research, Vol. 82, 2021, p. 100900.

[13] A. H. Talal, U. Jaanimägi, K. Davis, J. Bailey, B. M. Bauer, A. Dharia, S. S. Dickerson, "Facilitating engagement of persons with opioid use disorder in treatment for hepatitis C virus infection via telemedicine: Stories of onsite case managers", Journal of Substance Abuse Treatment, Vol. 127, 2021.

[14] I. Kaur, T. Behl, L. Aleya, H. Rahman, A. Kumar, S. Arora, I. J. Bulbul, "Artificial intelligence as a fundamental tool in management of infectious diseases and its current implementation in COVID-19 pandemic", Environmental Science and Pollution Research, Vol. 28, 2021, pp. 40515-40532.

[15] P. Spachos, S. Gregori, M. J. Deen, "Voice Activated IoT Devices for Healthcare: Design Challenges and Emerging Applications", IEEE Transactions on Circuits and Systems II: Express Briefs, Vol. 69, No. 7, 2022, pp. 3101-3107.

[16] J. M. Alvarez, C. H. Barcelo, P. Lesso, P. C. Higuera, "Robust Detection of Audio-Cough Events using local Hu moments", IEEE Journal of Biomedical and Health Informatics, Vol. 23, No. 1, 2019, pp. 184-196.

[17] B. O. Erdene, J. L. Saver, "Automatic Acute Stroke Symptom Detection and Emergency Medical Systems Alerting by Mobile Health Technologies: A Review", Journal of Stroke and Cerebrovascular Diseases, Vol. 30, No. 7, 2021, p. 105826.

[18] M. B. Jamshidi et al. "Artificial Intelligence and COVID-19: Deep Learning Approaches for Diagnosis and Treatment", IEEE Access, Vol. 8, 2020, pp. 109581-109595.

[19] E. B. Panganiban, A. C. Paglinawan, W. Y. Chung, G. L. S. Paa, "ECG diagnostic support system (EDSS): A deep learning neural network based classification system for detecting ECG abnormal rhythms from a low-powered wearable biosensors", Sensing and Bio-Sensing Research, Vol. 31, 2021, p. 100398.

[20] X. V. Wang, L. Wang, "A literature survey of the robotic technologies during the COVID-19 pandemic", Journal of Manufacturing Systems, Vol. 60, 2021, pp. 823-836.

[21] G. Dharmale, D. D. Patil, T. Ganguly, N. Shekapure, "Effective speech recognition for healthcare industry using the phonetic system", Journal of Autonomous Intelligence, Vol. 7, No. 5, 2024, pp. 1-14.

[22] K. Choutri, M. Lagha, S. Meshoul, M. Batouche, "A Multi-Lingual Speech Recognition-Based Framework to Human-Drone Interaction", Electronics, Vol. 11, No. 12, 2022, p. 1829.

[23] F. Kerwagen, K. F. Fuchs, M. Ullrich, A. Schulze, "Usability of a mHealth Solution using Speech Recognition for Point of care Diagnostic Management", Journal of Medical Systems, Vol. 47, 2023, pp. 1-10.

[24] R. Shinde, S. Patil, K. Kotecha, V. Potdar, G. Selvachandr, "Securing AI-based healthcare systems using blockchain technology: A state-of-the-art systematic literature review and future", Transactions on Emerging Telecommunications Technologies, Vol. 35, No. 1, 2024, p. e4884.

[25] M. Javaid, A. Haleem, R. P. Singh, R. Suman, R. Shanay, "Significance of machine learning in healthcare: Features, pillars and applications", International Journal of Intelligent Networks, Vol. 3, 2022, pp. 58-73.