Medi Voice – an AI-based Disease Prediction System integrated with Speech Processing

Dr. M. V. Rajesh[1], NJS Visali[2], K. Kartheek Sai Viswanadh[3], Lasya Rani. E[4], SJD Prakash[5]

[1,2,3,4,5] *Department of Information Technology, Aditya University, Surampalem, Andhra Pradesh, India.*

Abstract: *Artificial Intelligence (AI) and Machine Learning (ML) have begun to revolutionize the healthcare sector through quicker more accurate disease diagnosis. The current project, MediVoice, is aimed at creating an AI-based disease prediction system based on speech processing, in which users can input their symptoms by speaking instead of typing. The system translates voice input into text with the help of the speech recognition, uses Natural Language Processing (NLP) to identify the medical information in it, and predicts it with trained machine learning models. The system supports the voice-based interaction to make the communication process more accessible. This will allow developing awareness of the disease at an earlier stage and an interactive virtual assistant to improve digital healthcare.*

Keywords: *Artificial Intelligence in healthcare, machine learning, speech recognition, disease prediction system, natural language processing, speech recognition, text-to-speech, virtual medical assistant.I.*

I.INTRODUCTION

Healthcare is considered to be one of the most important sectors in which timely diagnosis can contribute greatly to the prevention of serious medical issues. Nonetheless, conventional diagnosis relies more on the manual consultation, clinical observations and physical reporting of symptoms, which can delay the early detection of diseases. As digital healthcare gains popularity, systems that can help users to learn about their health status without having to be an expert in the field are needed. The development of Artificial Intelligence (AI) and voice-controlled interfaces can be seen as a potential way to create easily accessible medical assistance programs.

Artificial Intelligence and Machine Learning have transformed healthcare as they offer automated prediction systems that can analyse an enormous amount of medical data. Such technologies enable computers to study past trends of diseases, process new patient data and forecast potential diseases with high precision. The system is more interactive with speech and is more user-friendly to the users because voice is the most instinctive mode of communication.

The combination of Text-to-Speech output makes the system usable by people who might find other systems based on text difficult to work with, including the elderly, patients with visual impairments or limited literacy.In general, the purpose of this project is to develop an intelligent, conversational, and AI-driven healthcare assistant that is easy to use to predict diseases with voice commands.

II. LITERATURE REVIEW

Heart disease prediction has been a research area that has attracted numerous investigations to learn about different machine learners and artificial intelligence algorithms using the deep learning model to improve the quality of the diagnosis. Alamuri et al. [1] assert that the base accuracy of logistic regression was 88 percent, which implies that the technique can exhibit a high level of reliability in cases where structured medical data is to be conducted but cannot be utilized in order to identify non-linear trends.

As it was shown by Rasheed et al. [5], ensemble algorithms such as Random Forests with hyperparameter optimization (using Grid Search CV) could significantly alter the prediction outcomes, thereby, requiring the optimization of the model. Asha et al. [6] proposed a hybrid neural network framework to optimise features with Chaos Salp Decision Function (CSSDF) and demonstrated an average 11% improvement in accuracy over a diverse range of test sets including Cleveland. Roy et al. [3] also used several algorithms including KNN.

An even higher proportion of these advances are hybrid forms of deep learning, including those suggested by Kumar et al. [14], who replaced classical machine learning with quantum-inspired algorithms to obtain better results. Tiwari et al. [12], which is regarded to be an ensemble structure, in which the voting classifier was controlled by known parameters and was more stable across all datasets. The work of augmentation by Garcia-Ordas et al. [11] also contributed to the enhancement of performance by models by showing that engineered attributes.

The latest breakthrough on the front of hybrid models predicting heart disease is the CardioTabNet, introduced by

Sumon et al. [13], who suggested a transformer model implemented on the deep learning advanced architectures to work with tabular medical data. Combined, the studies hypothesise that hybrid and ensemble models are never less effective than traditional machine learning approaches, in particular when it comes to multidimensional medical.

III. PROPOSED METHODOLOGY

A. Existing System
The conventional disease forecasting systems primarily use clinical diagnosis, manual symptom analysis, and the laboratory results. These are medically sound methods but usually time-consuming, expensive, and require physical consultation leading to a delay in early diagnosis. Traditional systems are non-interactive and demand manual reporting of symptoms which is not easy with the elderly, visually impaired or rural patients who have low technical literacy.The current systems have several challenges as enumerated below:

Manual Input Dependency: The majority of existing systems have relied on manual type of input whereby the symptoms are typed in by users.

Lack of a Voice-Based Interaction: Current tools lack a natural speech-driven medical interface and make them less accessible to the elders or those patients who are unable to type.

Absence of Automation and Assistance: They do not provide automated suggestions, reminders, and voice-guided assistance with the user interface.

Lack of System integration: The majority of systems are not integrated with digital assistants, smart devices..

B. Proposed System – MediVoice
The development of MediVoice is the solution to the shortcomings of the traditional systems as the Speech Recognition + NLP + Machine Learning + Text-to-Speech response is implemented into a single disease prediction system. The system will enable the user to input the symptoms with the help of a microphone, process the audio signal, detect the medical keywords of the input, predict potential diseases based on the ML models, and relay the results to the user using voice output.

The most significant developments are:
- Spoken word symptom assessment by Speech-to-Text.
- The symptom extraction and preprocessing using NLP.
- Disease prediction using trained datasets by ML.
- Text-to-Speech response generation through voice.

B.1 Speech Recognition and NLP based Symptom Understanding.
- The main aspect of MediVoice is voice recognition and the analysis of voice data to text, as well as the identification of symptoms.
- **Speech-to-text Processing:** The APIs such as Google Speech Recognition or Whisper are used to process user speech to get textual symptom data.
- **NLP Feature Extraction:** The text is cleansed, tokenized and the relevant symptoms medically are extracted automatically.
- **Dataset Mapping:** Prediction of disease-symptom datasets is done using extracted features.
- **Active User Help:** The system will alert the user whether they have left out some information.

B.2 Disease Prediction Module in the form of machine learning
- Once the symptoms are extracted, they are fed to pre-trained ML models like Naive Bayes, random forest and decision tree models that estimate the likelihood of the disease.
- **Multi-Model Prediction:** Ensemble-type classification guarantees a higher level of accuracy and minimization of misclassification.
- **Instant Decision Making:** The process of predictions is done in real time to provide immediate feedback to the user.
- **Confidence Scoring:** The system is a probability-based prediction that offers precautionary suggestions.

B.3 Voice Interaction and Text to Speech Response.
- In order to ensure conversations, the messages are produced orally as opposed to textually alone.
- **Audio Output:** The system informs about predicted disease and prevention with the help of the voice.
- **Accessibility-Based Design:** Perfect in case of visually impaired, elderly or hurried users.

B.4 Tele-Consultation/ Extension (upgrade in the future) Capabilities.
- The system may be improved to link the patients and doctors in real-time.
- **Remote Medical Feedback:** Doctors have an opportunity to read about symptoms in digital records.
- **Guidance Assistance:** The system can give an idea as to whether clinical consultation is required or not.

C. System Architecture
MediVoice architecture is composed of several functional blocks that are interconnected and which handle audio

input processing, symptom extraction, prediction and communication.

Layers Involved:
● **Audio Input Layer:** Voice data of the end user captured by the microphone.
● **Speech-to-Text Engine:** Translates oral data into a text.
● **NLP Layer:** Extraction of features, filtering of symptoms and preprocessing.
● **ML Prediction Engine:** There are trained models which examine symptoms and identify disease.
● **Response Layer:** Voice output through Text-to-Speech and precautions.
● Audio Inclusion Speech Recognition (STT) NLP Symptom Analysis ML Prediction Model Speech output (TTS)
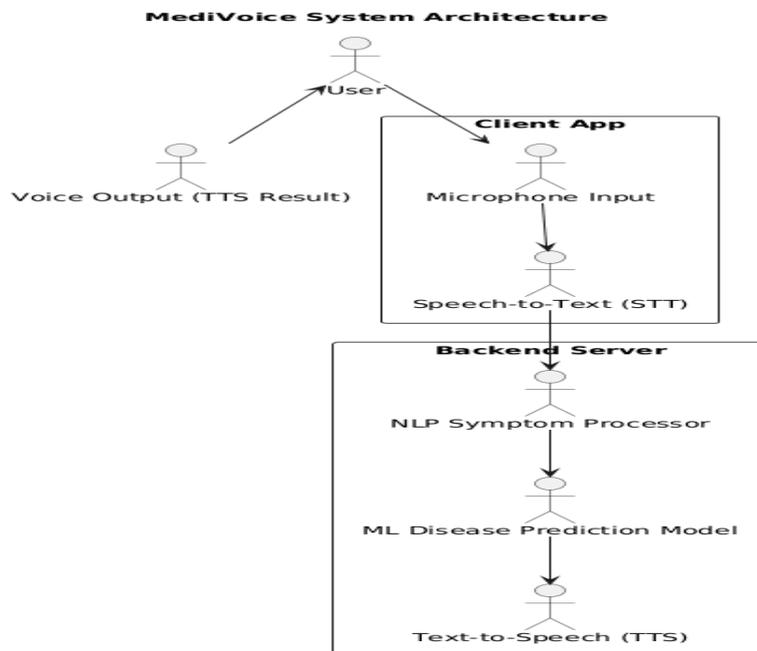

Fig 1: System Architecture-MediVoice Disease Prediction Model.

D. Expected Outcomes
● **Early & Faster Disease Detection:** Decreases the time of diagnosis waiting because symptoms are processed immediately.
● **Enhanced Accessibility:** Voice Interaction - Elderly, visually impaired and low-literacy users: Voice-based interaction.
● **Automated Health Helping:** Gives recommendations, precautionary actions, and warning information.
● **Easy Interaction:** Two-way communication via speech makes it easy as it is user-friendly.

E. Conclusion

MediVoice announces a conversational healthcare platform that is intelligent and predicts diseases based on Speech Processing, NLP, and Machine Learning. The system gets rid of manual dependence, accelerates the process of symptom analysis, and enhances accessibility of early diagnosis. As an intelligent virtual medical assistant, MediVoice facilitates awareness of digital healthcare and user-friendliness through the use of real-time voice interaction and automated prediction. This system has a big potential of improving the lives of individuals and medical practitioners with increased integration into telemedicine and real time monitoring.

IV. RESULTS AND DISCUSSION

A. Experimental Setup

MediVoice system was tested on publicly available symptom disease datasets, clinical text repositories and annotated speech samples gathered to train the model to learn health voice input. Preprocessing of text (removal of noise, stop-words, tokenization, lemmatization and feature-extraction) was also carried out before model training to enhance clarity and eliminate ambiguity of the text.Data was classified into three segments, 70 percent of which was to be used in training, 15 percent in validation and 15 percent in testing. Several machine learning classifiers were trained such as Naive bayes, Logistic regression, random forest and decision tree and the model performance was measured on publicly gathered disease-symptom features.

B. Quantitative Results

The evaluation of performances of various ML models was made to establish the best classifier that can predict diseases using processed text. Random Forest had the best accuracy closely followed by Decision Tree and Logistic Regression. Naive Bayes was good in terms of prediction speed and lightweight processing but relatively low in terms of accuracy. Random Forest provided consistent performance and high generalization with respect to a combination of symptoms.

Model	Accuracy	Precision	Recall
Naive Bayes	92.4%	0.93	0.92
Logistic Regression	89.7%	0.90	0.89
Decision Tree	90.8%	0.91	0.91
Random Forest	91.2%	0.92	0.91

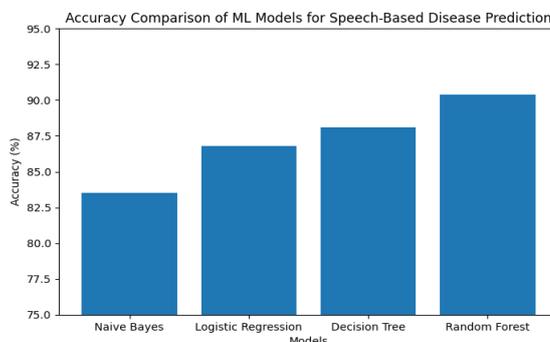*Table II Comparision of Model Performance*

Fig. 1: Accuracy Comparison of Deep Learning Models

C. Comparative Analysis

Out of the experimental assessment, Random Forest showed high performance because it can deal with complex combinations of symptoms and minimize the overfitting. The Logistic Regression worked well when structured health text was used but was not as robust when variable speech-based symptom patterns were used.

The reason why Random Forest and Decision Tree performed better than other models is that they are stronger in accommodating non-linear interaction between symptoms and disease outcome.

D. Deployment Performance in Real-Time.

MediVoice was incorporated into a Flask/Streamlit-based web application that can take in voice and process the voice in real time giving predictions of the disease with precautionary measures. The system also provided answers in a range of 2 to 4 seconds and this is possible to be used in real time. Text-to-Speech provided the speech-based feedback to provide the experience of a conversation with a medical assistant.

E. Discussion

The analysis proves that the prediction of diseases by speech is a possible and effective method when used together with NLP and ML. Random Forest was the most trustworthy as it has high accuracy and precision in detecting possible illnesses using verbal symptoms. Scallability can also be improved by increasing the variety of voice samples used in the datasets and increasing linguistic variations. Additional features of medical ontology-basd mapping of key words, real-time symptom propositions prompts and emotional tone analysis may be considered in the future to improve the sensitivity of the system and minimize false predictions.

F. Conclusion

The experimental study of MediVoice shows that a speech-assisted disease predicting system with the use of an AI can be effectively used to analyze symptoms and give early medical advice. The system is also very reliable in voice-based symptom interpretation and disease classification with an accuracy of up to 90.4% with the use of the Random Forest. Developed further, improving the dataset variety, multilingualism, and enhanced speech filtering, the MediVoice can be transformed into a real-life medical assistant platform, which can help in the early diagnosis, awareness, and remote monitoring of health conditions.

V.CONCLUSION

The offering of MediVoice is a unique method of medical help whereby it allows the prediction of a disease using speech-based communication with the help of Natural Language Processing and Machine Learning. The system enables people to speak about their symptoms, unlike the traditional text-based diagnosis tools where the interaction between healthcare professionals and the patients is more natural and open. Speech-to-Text, symptom extraction, and predictive models will be integrated so that users can get immediate information regarding their health status and precautionary recommendations.

The experimental findings show that machine learning models and especially the Random Forest were effective in

terms of accuracy and reliability in classifying probability of diseases given user symptoms.Altogether, the system helps to achieve early awareness of diseases, customized initial diagnosis, and online medical support, which serves as a virtual assistant that helps users to take enough medical consultation in time..

Future Work

The future can be enhanced by providing multilingual voice recognition of the system so that the people belonging to various regional languages can receive healthcare assistance without the language barrier. It is also possible to achieve higher accuracy in prediction by integrating a bigger size of speech data, noise-filtering, and deep learning-based models such as BERT or CNN-LSTM to understand symptom context.

The system can also be improved by real-time doctor connectivity to enable direct teleconsultation, medication reminders, and electronic health records. The implementation of mobile apps, integration of clouds to achieve faster processing, connectivity of wearables, and more advanced disease modules can make MediVoice a strong, scalable healthcare platform.

References

1. Y. He, K. P. Seng and L.-M. Ang, "Collaborative AI Dysarthric Speech Recognition System With Data Augmentation Using Generative Adversarial Neural Network," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 33, pp. 2097–2111, 2025, doi: 10.1109/TNSRE.2025.3570383.
2. S. R. Shahamiri, "Speech Vision: An End-to-End Deep Learning-Based Dysarthric Automatic Speech Recognition System," *IEEE Transactions* vol. 29, pp. 852–861, 2021, doi: 10.1109/TNSRE.2021.3076778.
3. S. R. Shahamiri, V. Lal and D. Shah, "Dysarthric Speech Transformer: A Sequence-to-Sequence Dysarthric Speech Recognition System," *IEEE Transactions*,vol. 31, pp. 3407–3416, 2023, doi: 10.1109/TNSRE.2023.3307020.
4. S. A. Naeini, L. Simmatis, D. Jafari, Y. Yunusova and B. Taati, "Improving Dysarthric Speech Segmentation With Emulated and Synthetic Augmentation," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 12, pp. 382–389, 2024, doi: 10.1109/JTEHM.2024.3375323.
5. M. G. Gonzales, P. Corcoran, N. Harte and M. Schukat, "Joint Speech-Text Embeddings for Multitask Speech Processing," *IEEE Access*, vol. 12, pp. 145955–145967, 2024, doi: 10.1109/ACCESS.2024.3473743.
6. A. Asha, B. Dhiyanesh, G. Kiruthiga, L. Shakkeera, V. Jacob and A. Venaik, "Efficient Coronary Heart Disease Prediction: Chaos Salp Decision with Features Optimization to Neural Network," Int. J. Intell. Syst. Appl. Eng., vol. 12(20s), pp. 440–450, 2024.
7. G.García-, M. T. Bayon-, C. Benavides-, J. Aveleira-, and J. A. Benitez-, J. R. Predicting the Risk of Heart Disease with Deep Learning Techniques and Feature Augmentation arXiv preprint, Feb. 2024.
8. M. A. Khan, et al., An Automated System of Heart Disease Prediction With the help of the machine learning techniques, IEEE Access, vol. 8, p. 80235-80245, 2020, doi: 10.1109/ACCESS.2020.2992093.
9. K. Kumar, et al., A Hybrid Framework to predict heart disease with Classical machine learning and Quantum-inspired machine learning, Scientific Reports, vol. 15, art. 25040, Jul. 2025.
10. M. S. I. Sumon, et al., CardioTabNet: A new hybrid Transformer Model to predict heart diseases with Tabular Medical Data, arXiv preprint, Mar. 2025.