Grenze International Journal of Engineering and Technology, June Issue

An Insight to AI-Powered Cardiovascular Health Monitoring System using Vocal Biomarkers

Sai Pavan G[1], Prabhavathi C N[2] and Suguna G C[3]
[1-2]Dept of ECE, RNS Institute of Technology, Bengaluru, India
Email: 1rn22ec102.saipavang@rnsit.ac.in, prabhavathi.cn@rnsit.ac.in
[3]Dept of ECE, JSS Academy of Technical Education, Bengaluru, India
Email: sugunagc@jssateb.ac.in

***Abstract*— Cardiovascular Diseases (CVDs) remain high on the list of causes of death globally, so early, convenient, and non-invasive diagnostic possibilities are required. Conventional cardiovascular monitoring relies on clinical- grade equipment and physical checks, which limit scalability and real-time measurement. The current paper presents a design for an AI-driven framework for cardiovascular health monitoring via speech biomarkers measurable voice parameters representing physiological change due to cardiac disease. The proposed methodology integrates acoustic preprocessing; feature extraction techniques such as Mel-Frequency Cepstral Coefficients (MFCCs), jitter, and breath-to-speech ratio; and classification using machine learning and deep learning architectures such as support vector machines, convolutional neural networks, and recurrent neural architectures. Since proper public datasets with annotated cardiovascular speech data are lacking, speech recordings will be collected manually via pilot trials in healthy volunteers and subsequently via clinical collaborations for extensive validation. The project is currently in the design phase, and the paper introduces the research aims, system structure, data collection plan, and implementation plan for the development of a non-invasive, scalable, real-time cardiovascular monitoring system.**

***Index Terms*— Speech biomarkers, Cardiovascular monitoring, Non-invasive diagnostics, Machine learning, Deep learning, AI in healthcare, Telemedicine.**

I. INTRODUCTION

Cardiovascular diseases (CVDs) are the world's leading cause of death, with over 17 million fatalities per year. Despite improved diagnostics development, most cardiovascular assessments need to be performed with specific equipment and in clinical facilities by echocardiography, electrocardiography, or biochemical tests. While these conventional methods are effective, they are likewise limited in terms of cost, accessibility, and scalability, particularly for continuous or population- level health monitoring. More recent developments in digital health have underscored the need for non-invasive, real-time methodologies that can be performed outside the clinic. Among these possibilities, speech-based biomarkers have proved to be of particular promise as a modality for physiological health monitoring. These biomarkers are measurable changes in speech characteristics like prosody, vocal tremor, amplitude variation, and spectral energy distribution that are dependent on underlying physical states, like respiratory and cardiovascular status.

Grenze ID: 01.GIJET.12.2.60_21

Studies have shown that cardiac dysfunctions, heart failure or hypertension, can alter vocal fold and breathing behaviour and therefore cause subtle yet perceptible changes in voice. With improving artificial intelligence (AI) technologies in the form of machine learning (ML) and deep learning (DL), speech analysis is increasingly proving to be a useful diagnostic tool. AI has previously been employed to diagnose neurological and psychological illnesses through voice analysis. However, the exploration of speech biomarkers for cardiovascular monitoring has not been widespread. This project seeks to fill that gap by proposing an AI-powered cardiovascular health monitoring framework that takes speech signals as input. Unlike earlier works which utilized pre- curated public data for training and testing purposes, this project acknowledges there are no publicly available annotated speech recording datasets that correspond to cardiovascular diseases. Therefore, the information utilized in this project will be collected directly through pilot studies on healthy volunteers. Subsequent phases will involve partnerships with medical facilities to access medically certified speech recordings that correspond to diagnosed cardiovascular ailments. This approach facilitates more situation-dependent data collection and ensures real-world deployment setting correspondence. The system, at the design phase, will consist of modules for speech preprocessing, feature extraction, and predictive modeling, with integration of both traditional ML classifiers as well as deep learning models. Acoustic features like Mel-Frequency Cepstral Coefficients (MFCCs), speech-to-breath ratios, and vocal tremor features will be extracted and utilized to train models like support vector machines (SVMs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs). In the end, the long-term goal is to develop a deployable tool that can be integrated into telemedicine platforms and wearables to facilitate real-time and non-invasive cardiovascular monitoring in clinical and remote settings.

II. Related Work

The identification of vocal biomarkers as a method for tracking cardiovascular status has been of increasing interest since they are non-invasive, readily obtainable, and suitable for real-time analysis protocols. Speech signals, modulated by respiratory, neuromuscular, and cardiovascular mechanisms, have the potential to reflect underlying physiological changes with accompanying cardiac dysfunction. As digital health continues to progress, biomedical engineers, signal processors, and clinical cardiologists have investigated the use of voice as a surrogate measure of cardiovascular status. This body of work spans from early clinical trials through signal-level feature extraction to deep learning-based modeling and is applied within mobile health (mHealth) and telemedicine platforms. This research direction intersects with existing multidisciplinary research, driven primarily by clinical basis and signal processing methods, in addition to AI-driven systems underpinning voice use in cardiology diagnostics.

A. Clinical Foundations of Vocal Biomarkers in Cardiology

One of the first of this series of milestone studies was by Maor et al. [1], who demonstrated that voice characteristics extracted from brief telephonic voice recordings were powerfully linked with death and hospitalization in heart failure patients. Of a population of over 10,000 patients, those in the highest quartile of the vocal biomarker score had a roughly doubled risk of death or admission when compared with those in the lowest quartile. The AHF-Voice study conducted by Kerwagen et al. [2] also confirmed this clinical use. Using the day-to-day speech samples of acute heart failure patients, the study linked vocal parameters to the traditional measures such as N-terminal pro-B-type natriuretic peptide (NT- proBNP) level and fluid congestion scores. This longitudinal study established the feasibility of integrating speech analysis into the common care pathway for chronic heart failure. At the ESC 2024 congress, Sharma [3] also showed AI-based speech tools, which detected abnormalities in voice, e.g., shimmer and creakiness, in decompensated heart failure. The tools were advocated for early triage and integration into telemedicine platforms.

B. Signal Processing and Machine Learning Techniques

The physiological justification for cardiovascular diagnosis with speech has also been investigated using acoustic signal analysis. Rao et al. [4] demonstrated zero-crossing rate, energy entropy, and spectral centroid to be dependable features for heart rate estimation and supported that vocal tract function indexes cardiovascular and respiratory states. Building on this research, Cohn et al. [5] developed models incorporating linguistic features and acoustic features (e.g: MFCCs, jitter, and shimmer) for heart rate variability prediction. Their conclusion was that even neutral talk could yield considerable insight into autonomic nervous system function. Affective computing studies by van der Werff et al. [6] showed that emotional stress affected vocal parameters in a way that was predictive of physiological arousal, for example with increased heart rate and blood pressure. Using

Gaussian Mixture Models (GMMs) and Support Vector Machines(SVMs), the group classified emotional states following cardiovascular changes. In the same way, Levanon and Lerman [7] presented speech features in patients with coronary artery disease (CAD) and noted stress-induced changes in harmonic-to-noise ratio (HNR) and spectral entropy, providing evidence of the relationship between hemodynamic stress and acoustic modulation.

C. Deep Learning, Explainable AI, and System-Level Integration

Recent advances in representation learning have taken speech-based cardiac diagnosis to new heights. Dhamala et al. [8] employed Wav2Vec 2.0, a self-supervised model, in order to learn the latent speech features required for NT-proBNP and troponin levels without the need for manual feature engineering. Their results were strongly correlated with clinical biomarker data. Based on this, Ahmadli et al. [9] proposed a hybrid model that combined speech embeddings with laboratory-based measures to forecast short-term mortality among heart failure patients. Using Shapley Additive Explanations (SHAP), the model provided clinical decision-making clarity by determining relevant speech segments that had an impact on forecasts. From the systems perspective, Ismael and Fagherazzi [10] illustrated a digital health architecture to incorporate vocal biomarkers into mHealth environments. Universality of linguistic media, ethical management of data, and potential for edge-device deployment were underscored by their work. In line with this, Gouda et al. [11] evaluated voice assistants like Amazon Alexa in the management of cardiovascular diseases and showed improved patient adherence and activation, but with speech recognition challenges among the older population. A broader overview by Fagherazzi et al. [12] captured the entire pipeline from raw audio acquisition and cleaning to machine-learning-inferred interpretation and clinical deployment. Their summary emphasized the need for multilingual datasets, real-time capabilities, and clinical utility, specifically for cardiovascular applications where vocal and cardiac systems interact.

III. PROBLEM STATEMENT AND RESEARCH OBJECTIVES

A. Problem Statement

CVDs are a global health issue, causing more than 17 million deaths per year. Despite the efficacy of early diagnosis and repeated monitoring, existing diagnostic techniques like electrocardiography (ECG), echocardiography, and biochemical tests are limited by invasiveness, expense, and reliance on clinical facilities. These limitations prevent repeated assessment, mass screening, and monitoring from a distance particularly in low-resource settings. Recent developments in AI and speech signal processing indicate that non-invasive voice-guided health markers, or speech biomarkers, have the potential to offer clinically relevant information regarding a person's cardiovascular condition. Alterations in vocal measures like jitter, shimmer, rate of speech, prosody, and phonation threshold pressure can be indicators of underlying cardiac impairment, such as fluid overload, compromised cardiac output, and neuromuscular exhaustion. Yet, the research currently available in this field is con strained by several key limitations:

- Lack of publicly accessible, annotated voice datasets associated with cardiovascular diseases.
- Inconsistency in model performance between different populations and languages.
- Lack of deployable real time systems ready for integration with telemedicine or mobile platforms.

This work is designed to overcome these limitations through the creation of an AI-based framework that processes speech signals to forecast cardiovascular risk. In contrast to previous work based on publicly curated data, the suggested framework will have access to speech data gathered with pilot testing in healthy volunteers, which would be followed by extended clinical partnerships for real-world validation.

B. Research Objectives

To rectify the gaps identified, the following research objectives are formulated:

- Develop and design a speech-based cardiovascular health monitoring system with the ability to process raw voice recordings and derive clinically significant acoustic features for risk estimation.
- Employ preprocessing and feature extraction pipelines with standardized noise-reduction procedures (e.g: spectral subtraction, Wiener filtering) and derive features like MFCCs, jitter, shimmer, spectral energy distribution, and breath-to-speech ratio.
- Build and test machine learning and deep learning models, such as standard classifiers (e.g: SVM, Random Forest) and neural networks (e.g: CNN, RNN, LSTM), with incorporation of self-supervised learning models such as Wav2Vec 2.0 and HuBERT as necessary.

- Acquire and label a personalized speech dataset by conducting pilot recordings of healthy speakers, followed by procuring clinically tagged speech samples with the help of healthcare.
- Develop a deployable system for telehealth and wearable platforms for real-time continuous cardiovascular health monitoring in non-clinical and resource-poor environments.
- Consider ethical, privacy, and deployment issues by guaranteeing adherence to data protection rules and ethical principles for human-subject research.

IV. Proposed Methodology

This section presents the methodology intended for the creation of a speech-based cardiovascular health monitoring system. The method combines voice signal acquisition, acoustic feature extraction, machine-learning-based classification, and real time, non-invasive deployment planning for diagnostic support.

A. Data Acquisition and Collection Strategy

Since publicly annotated speech datasets related to cardiovascular diseases are not available, a dedicated dataset is being gathered for this research. During the pilot phase, recordings of speech are acquired from healthy volunteers with the help of the Audacity software program such that standardized recording is conducted at a mono-channel setup with a sampling frequency of 16 kHz. Patients are asked to read predefined phrases in a quiet room to minimize ambient noise and add stability to phonation patterns. Audio files are stored in .wav format to maintain fidelity and compatibility with subsequent processing equipment. Future stages will involve clinical partnerships to record voice data from patients with heart failure, hypertension, or arrhythmia diagnoses. All samples will be identified with analogous clinical measurements like blood pressure, NT-proBNP levels, ejection fraction, or diagnostic results. All data gathering will adhere to relevant privacy and ethics practices, including informed consent.

B. Signal Preprocessing

In order to maintain consistency and reduce variability due to recording conditions, all audio samples will be preprocessed before feature extraction. The processes involve:

- Resampling to a consistent 16 kHz mono-channel format.
- Denoising with Spectral Subtraction and Wiener filtering to eliminate background noise.
- Amplitude and duration normalization to minimize speaker variation.
- Segmentation to extract useful phonation segments from silences and pauses.

This preprocessing pipeline guarantees downstream feature computation to be carried out on acoustically clean and normalized input.

C. Feature Extraction

After preprocessing, speech signals will be converted to multidimensional acoustic representations. Two parallel approaches will be employed:

Conventional Acoustic Features

- Mel-Frequency Cepstral Coefficients (MFCCs): Characterize the perceptual and spectral attributes of speech.
- Spectral Energy Distribution: Recognizes atypical phonatory habits affected by pulmonary congestion.
- Jitter and Shimmer: Measure microvariations in pitch and amplitude, respectively, that could indicate fatigue or fluid overload.
- Breath-to-Speech Ratio (BSR): Monitors respiratory effort and coordination of respiration during phonation.

Learned Representations

- Mel-Spectrograms: Employed as image-like input to CNN models for spatial feature extraction.
- Self-Supervised Speech Embeddings: Wav2Vec 2.0 and HuBERT models will be used to produce task-specific latent representations.

D. Model Architecture and Classification

Both feature-based and deep-learning-based classification approaches will be tested by the system:

Feature-Based Models:

- Support Vector Machines (SVM): Best suited for linearly separable feature spaces.

- Random Forest and XGBoost: Ensemble classifiers with the ability to address feature non-linearity and interaction.
- Gradient Boosting Classifiers: Suitable for high-variance datasets with small training size.

Deep Learning Models:

- Convolutional Neural Networks (CNNs): Process spectrogram inputs to learn frequency-time patterns indicative of CVD.
- Recurrent Neural Networks (RNNs): Process sequential MFCC features to capture temporal dependencies in speech.
- Transformers: Sequence-to-sequence models for encoder-decoder formulations that capture long-range dependencies in inputs.
- Long Short-Term Memory (LSTM): Augmented RNNs that can model longer dependencies and memory effects.
- Transfer Learning: Pretrained models will be fine-tuned with gathered cardiovascular speech samples to enhance generalization.

Model training will use stratified cross-validation, and performance will be measured in terms of accuracy, F1-score, sensitivity, specificity, and AUC-ROC.

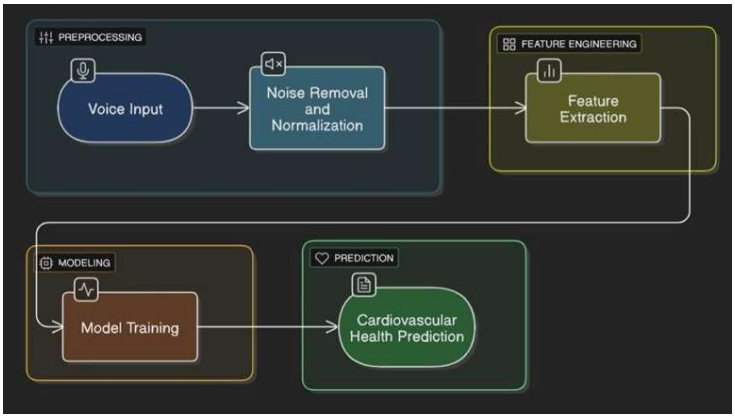

Fig. 1. Overview of the proposed AI pipeline for speech-based cardiovascular health monitoring

TABLE I: CARDIOVASCULAR HEALTH MONITORING EVALUATION MATRIX

Metric	Category	Description
Accuracy	Model Performance	Correct prediction proportion.
Precision	Model Performance	True positives over predicted positives. Minimizes false alarms.
Recall (Sensitivity)	Model Performance	True positives over actual positives. Essential for disease detection.
Specificity	Model Performance	Proper identification of healthy people.
F1-Score	Model Performance	Harmonic mean of recall and precision.
AUC-ROC	Model Performance	Area under ROC curve. Quantifies discrimination ability.
Inference Latency	System Deployment	Time taken by the model to make predictions. Critical for real-time applications.
Model Size	System Deployment	Memory usage of the trained model.

E. Deployment and Integration Plan.

The trained models, once validated, will be integrated into modular system architecture for deployment onto telemedicine or wearable platforms. The system will include:

- Voice input module: Smartphone or microphone compatible.
- Preprocessing engine: Acoustic cleaning on device or in the cloud.
- Inference engine: Containerized ML/DL model optimized (e.g., TensorFlow Lite or ONNX).
- Feedback interface: Easy-to-use output for visualization and alerts of health status. Feasibility of real-time processing and system latency will be assessed through prototyping.

F. Challenges and Risk Mitigation.

Anticipated challenges are data skewness (fewer clinical cases), inter-speaker variation, language variability, and model overfitting. To counter these:

- Synthetic minority oversampling (SMOE) and data augmentation will address class imbalance.
- Demographic bias will be minimized by speaker normalization and domain adaptation methods.
- Regularization, dropout, and early stopping will be utilized to avoid overfitting while training models.

V. Expected Outcomes and Future Scope

The anticipated speech-based cardiovascular health monitoring platform is expected to provide a non-invasive, scalable, and clinically useful solution for the early detection of risk in patients with possible cardiac dysfunction. Through the use of acoustic biomarkers extracted from human speech, the system will facilitate real-time monitoring over existing limitations with classical diagnostic modalities that necessitate invasive tests or hospital-based equipment. Once the pilot data collection and model prototyping phases are completed, the system is expected to show high recall and specificity in distinguishing healthy individuals from those with vocal correlates of cardiovascular abnormalities. These enhancements are anticipated to come from the inclusion of domain-specific acoustic features like jitter, shimmer, harmonic-to-noise ratio (HNR), and breath-to-speech ratio, and latent speech embeddings learned through self-supervised learning models like Wav2Vec 2.0 and HuBERT. Previous work has demonstrated that these features can capture underlying physiological conditions affected by heart failure, congestion, or fatigue [2], [4]. The near-term future research includes growing the dataset with clinical cooperation so that speech samples can be labelled using ground-truth cardiovascular metrics such as levels of NT-proBNP, ejection fraction, and official diagnostic results. This shift from pilot recordings to clinically annotated datasets is intended to enhance the model's robustness, generalizability, and diagnostic accuracy. Previous research on hospitalized patients with heart failure has established strong correlations between certain voice patterns and clinical worsening, justifying the potential of speech as a diagnostic indicator [1], [2]. In terms of deployment, the system will be optimized for real-time inference on edge and mobile devices. This will involve the application of light model structures, quantization methods, and memory-constrained deployment formats like ONNX or TensorFlow Lite. Such endeavours resonate with previous literature in digital health highlighting the significance of latency, portability, and energy efficiency in mobile diagnostics [6]. Minimizing inference time and model size will become critical for embedding the system into telemedicine platforms or smartphone-based screening devices. Future work will also tackle the known issues in vocal biomarker modeling, which are speaker variability, language-dependency, and external noise. To counter these, the system will integrate speaker normalization, data augmentation, and adversarial training methods. Domain-adaptation techniques will also be investigated to enable multilingual and demographically diverse usage. This process is important to make the system work uniformly across populations and dialects and make it more globally applicable [4]. Ethical and regulatory aspects will be an integral component of the deployment plan. Voice data will be gathered and processed in line with privacy-preserving protocols, upholding adherence to informed consent principles and data governance policies.

Federated learning or encrypted model-training methods could be investigated in consultation with clinical partners to improve privacy across data sharing and model development [3]. In summary, the system is not only a research prototype, but also a usable tool for real-world cardiovascular screening. By ongoing clinical validation, algorithmic improvement, and deployment testing, the framework will be an effective addition to standard screening techniques, allowing early glimpses into cardiac risk using nothing more than a person's voice.

References

[1] E. Maor et al., "A vocal biomarker is associated with hospitalization and mortality among heart failure patients," J. Am. Heart Assoc., vol. 9, no. 7, p. e013359, 2020.

[2] F. Kerwagen et al., "Vocal biomarkers in heart failure-design, rationale and baseline characteristics of the AHF-Voice study," Front. Digit. Health, vol. 3, p. 8600, 2021.

[3] A. Sharma, "Can speech analysis detect heart failure?" EMJ Cardiol., vol. 12, 2024.

[4] A. Rao et al., "Speech signal analysis for the estimation of heart rates," Interspeech, 2020.

[5] J. Cohn, A. Rao, Y. Xu, "Predicting heart activity from speech using data-driven and linguistic features," Interspeech, 2019.

[6] M. van der Werff et al., "Speech emotion recognition for estimating heart rate from speech," Interspeech, 2015.

[7] Y. Levanon, A. Lerman, "Detecting voice biomarkers for coronary artery disease using telephonic speech," Mayo Clinic Proc., 2019.

[8] R. Dhamala et al., "Self-supervised speech representation learning for predicting cardiac biomarker values," arXiv:2406.06341, 2024.

[9] N. Ahmadli et al., "Voice-driven mortality prediction in hospitalized heart failure patients," arXiv:2402.13812, 2024.

[10] M. Ismael, G. Fagherazzi, "Artificial intelligence and digital biomarkers," Digit. Biomarkers, vol. 4, pp. 101–114, 2022.
[11] H. Gouda et al., "Feasibility of incorporating voice technology and virtual assistants in cardiovascular care," J. Card. Fail., 2023.
[12] G. Fagherazzi et al., "Voice for health: the use of vocal biomarkers from research to clinical practice," Digit. Biomarkers, vol. 5, pp. 78–88, 2021.
[13] J. Lee, G. Kim, I. Ham, K. Ko, S. Park, Y.-J. Choi, D. O. Kang, J. Y. Choi, E. J. Park, S. Lee, et al., "Voice as a Biomarker to Detect Acute Decompensated Heart Failure: Pilot Study for the Analysis of Voice Using Deep Learning Models," medRxiv, Sep. 12, 2023, doi: 10.1101/2023.09.11.23295393.
[14] M. A. Sarsilă, N. Ahmadli, B. Mizrak, K. Karauzum, A. Shaker, E. Tulumen, D. Mirzamidinov, D. Ural, and O. Ergen, "Voice-Driven Mortality Prediction in Hospitalized Heart Failure Patients: A Machine Learning Approach Enhanced with Diagnostic Biomarkers," arXiv:2402.13812, version 2, 2024.
[15] K. Okada, D. Mizuguchi, Y. Omiya, K. Endo, Y. Kobayashi, N. Iwahashi, M. Kosuge, T. Ebina, K. Tamura, T. Sugano, T. Ishigami, K. Kimura, and K. Hibi, "Clinical Utility of Machine Learning-Derived Vocal Biomarkers in the Management of Heart Failure," Circulation Reports, vol. 6, pp. 303–312, Aug. 2024, doi: 10.1253/circrep.CR-24-0064.
[16] F. Kerwagen, M. Bauser, M. Baur, F. Kraus, C. Morbach, R. Pryss, K. Rak, S. Frantz, M. Weber, J. Hoxha, and S. Störk, "Vocal biomarkers in heart failure—design, rationale and baseline characteristics of the AHF-Voice study," Frontiers in Digital Health, vol. 7, May 2, 2025, Art. 1548600, doi: 10.3389/fdgth.2025.1548600.
[17] M. Bauser, F. Kraus, F. Koehler, K. Rak, R. Pryss, C. Weiß, A. Hotho, G. Fagherazzi, S. Frantz, S. Störk, and F. Kerwagen, "Voice Assessment and Vocal Biomarkers in Heart Failure: A Systematic Review," Circulation: Heart Failure, vol. 18, Aug. 2025, Art. e012303, doi: 10.1161/CIRCHEARTFAILURE.124.012303.