**MEDI-VOICE**

**1. Tech Stack**

* **Programming Language:**
* Python 3.12
* **Audio Processing & Feature Extraction:**
* **Librosa** — for audio loading and feature extraction (MFCC, pitch, rms, mel frequency)
* **SciPy** — for signal processing (peak detection)
* **NumPy** — numerical operations
* **Machine Learning / Reinforcement Learning:**
* **Stable Baselines3 (SB3)** — for Deep Deterministic Policy Gradient (DDPG) implementation
* **Gymnasium** — custom environment creation for RL training
* **PyTorch** — backend framework used by SB3 for neural network models
* **Model Training:**
* Custom heart rate simulation environment built in Gymnasium
* DDPG RL algorithm for continuous action space prediction of heart rate

**2. Literature Survey**

|  |  |  |  |
| --- | --- | --- | --- |
| **Paper Name** | **Implemented** | **Drawbacks** | **How to solve** |
| **Heart Rate Detection and Classification from Speech Spectral Features Using Machine Learning (2020)** | Regression based algorithms on mel MEL frequency spectrum constants | Only used MEL spectrum constant feature from voice and has only been trained on English dataset | Include pitch and mfcc and other spectral constants from speech as features |
| **Heart rate monitoring using human speech spectral features (2015)** | 20 classifiers used on speech feature MEL spectrum constant | Low classification accuracy  Only one feature used  Results are in a range of heart rates not exact | Use more complex models  And use more features |
| **Analysis and prediction of heart rate using speech features from natural speech (2017)** | Used random forest regression models on SRI BioFrustration Corpus to classify emotional state and heart rate through live continuous speech | Heart rate predictive accuracy is not good | Integrating more features and using more complex models |
| **Heart Rate Extraction from Vowel Speech Signals (2012)** | Estimates heart rate from vowel speech signals mapped on a short-term Fourier transform (STFT) | Only focuses on vowel speech. | Integrating this with machine learning and expanding to more features |
| **Speech signal analysis for the estimation of heart rates under different emotional states (2016)** | Used a empirical linear predictor model to estimate heart rate. Trained on 4000 audio samples with ECG data as labels | Used Feature distances as metric to classify heart rate  And small dataset size | Improve dataset and feature extraction methods |
| **Extraction of Heart Rate Parameters Using Speech Analysis** | Entropy energy mean frequency standard deviation of the speech signal is being used to estimate the heart rate | Only showed correlation between speech and heart rate parameters | Integrating machine learning |
| **How speech processing can help with beat-to-beat heart rate estimation in ballistocardiograms** | Uses speech signlas to estimate heart rate from BCG’s | Limited generalizability | Expanded dataset |
| **Determining heart rate using speech signal** | Uses Fast fourier transforms to map the frequency differences in voice then a regression model to show correlation | Doesn’t estimate the heart rate only shows relation between speech and heart rate | Implementing complex machine learning techniques on the concept |
| **Heart rate from read speech influenced by physical exercise** | Used pre-trained SBreathNet deep learning model to extract breathing patterns on which Independent component analysis was applied | Limited sample size  Only 7 participants data | Use other speech features along with breathing patterns to get a more accurate estimation |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**3. Abstract**

This project explores the estimation of heart rate from voice recordings by leveraging acoustic features extracted from speech signals, including MFCCs, pitch, energy, and speaking rate. Due to the scarcity of labelled heart rate data paired with voice, a heart rate simulation environment was developed to generate synthetic heart rate signals conditioned on speech features, enabling the training of a reinforcement learning model. Using the Deep Deterministic Policy Gradient (DDPG) algorithm within a custom Gymnasium environment, the system learns to predict heart rate values from the extracted speech features. This novel integration of speech signal processing with reinforcement learning provides a promising non-invasive method for estimating physiological parameters from audio data.

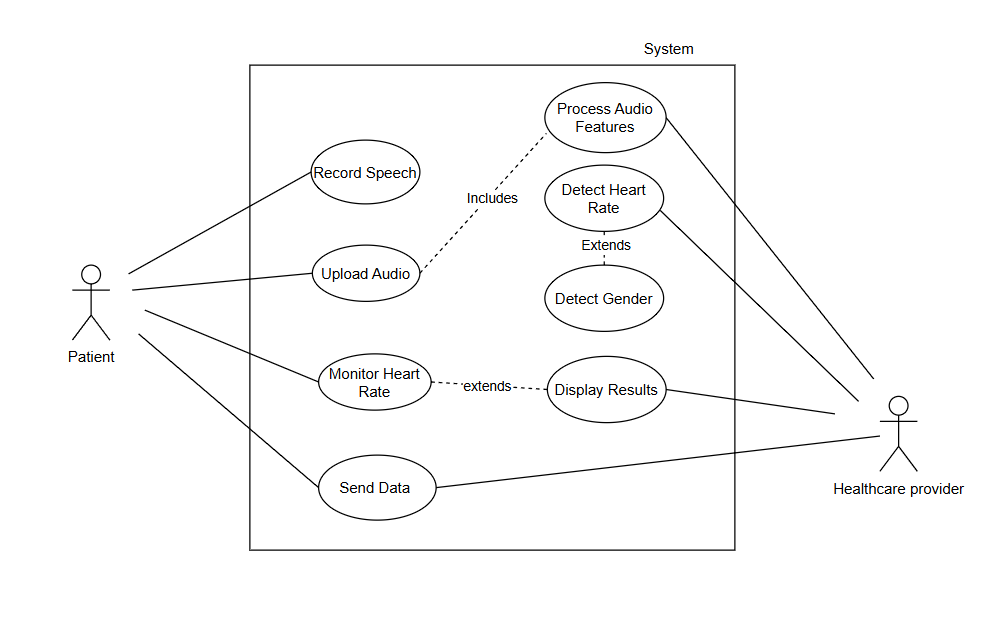
**4. Introduction**

Heart rate estimation through speech analysis exploits the physiological links between vocal production and cardiovascular activity. Speech signals encode multiple features such as pitch, energy, and spectral coefficients which can indirectly reflect heart rate variability. However, acquiring large, labelled datasets pairing speech with accurate heart rate measurements remains challenging.

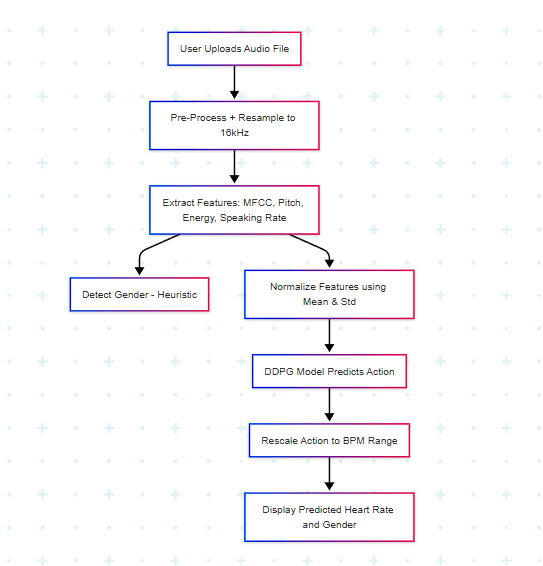
To address data scarcity, synthetic heart rate values are generated based on statistical models of speech features, enabling the creation of a robust training dataset. A reinforcement learning approach using Deep Deterministic Policy Gradient (DDPG) is employed to predict continuous heart rate values from high-dimensional speech features. This method leverages simulated data and advances beyond traditional regression techniques, aiming for improved prediction fidelity and adaptability.

**5. System Design**

**5.1) Use Case Diagram**

****

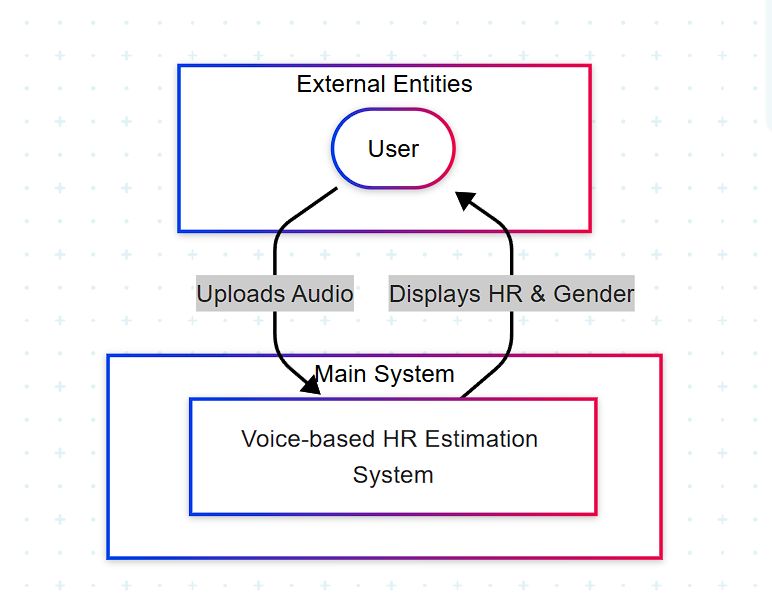
**5.2) Workflow Diagram**

****

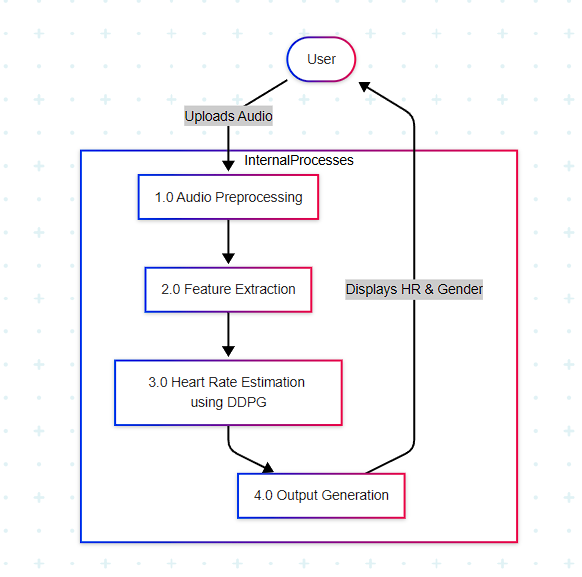
**6. Methodology**

**7. Detailed System Design**

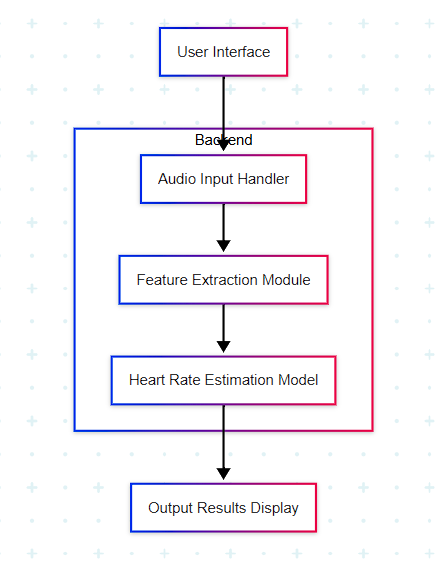
**7.1) DFD level 0 diagram**

****

**7.2) DFD level 1 diagram**

****

**7.3) Birds eye view diagram**

****

**8. Implementation**

**9. Testing**

**10. Results and conclusion**