ÇANKAYA UNIVERSITY

Software Design Document

Speech Emotion Recognition

Abdullah Özder - 202011410, Furkan Duran - 201811027, Elif Aybüke Coşkun - 201811018, Şima Kayısı - 201811043, İhsan Bardakcı - 201717007

26/12/2022

Table of Contents

List of Figures

1. INTRODUCTION

1.1 Purpose

The purpose of this Software Design Document (SDD) is explaining the system which is called audio-visual emotional recognition. This system's goal to provide recognition of emotions from speech and text.

We define the SER system as a collection of methodologies that process and classify speech signals to detect embedded emotions. Such systems can be used in speech analytics with interactive visitor agents, as well as in a variety of operational areas. In this study, we try to determine the underlying emotions(happy, sad, angry..) of recorded speech by examining the acoustic properties of the auditory data of the recordings.

1.2 Scope

Emotions are an integral part of human behavior and an inherited characteristic of all modes of communication.

We thought of your experience reading detecting different emotions which makes us more rational and understanding. However, while machines can easily understand text, audio, or video information, they are still far behind in accessing the depth of content.

Additionally, voice sentiment analysis has many applications in various fields such as healthcare, banking, defense, call center and IT.

In this project, user will make the upload as a voice or text file and the uploaded audio file and the recording will be the mood in the voice, and the uploaded text will be the mood analysis from the words used.

1.3 Glossary

Term	Definition
SER	The main project is an abbreviation of the name Speech Emotion Recognition
User	People who want to use application

1.4 Overview of Document

The remaining chapters and their contents are listed below.

Section 2 is an architectural design that describes the development stages of the project. Also included is a project system and architectural design class diagram that describes actors, exceptions, basic sequences, priorities, preconditions, and postconditions. Additionally, this section contains an activity diagram for the Scenario Generator.

Section 3 is realization of the use case. This section presents and describes a block diagram of a system designed according to the use cases in the SRS document.

Section 4 is about the environment. This section showed a sample frame of the environment from the prototype to illustrate the scenario.

1.5 Motivation

We are a group of final-year students of computer engineering department. As a group, we chose the project of making mood analysis from sound and text. In this project, we aimed to include technologies in the fields of education, social and humanity. We are interested in speech processing, image processing and artificial intelligence fields.

2. DESIGN OVERVIEW

a 2.1 Description of Problem

One problem with speech emotion recognition is that it can be difficult to accurately perceive and identify emotions in speech because of the complexity and variability of human emotions. Different people can express the same emotion in different ways, and the same person can express different emotions in similar ways.

Another problem is that speech can be affected by a variety of factors that can distort expressed emotions. For example, background noise, accent, and speaking style can affect perceived emotion in speech.

b 2.2 Technologies Used

This software will communicate with Python to handle image, audio and text processing. Machine learning algorithms, artificial neural networks will be used to train the model. Librosa, a Python package for music and audio analysis, will also be used. Java Script and React will be used for the design. This software can run on Microsoft Windows, Linux and macOS.

c 2.3 Design Summary

To create a speech emotion recognition system, the following steps are followed:

Collect and preprocess data: The first step in building a speech emotion recognition system is to collect a large dataset of spoken language samples associated with the emotions they are expressing.

This may involve labeling audio recordings with appropriate emotion categories (such as happy, sad, angry, neutral). Data may also need to be cleaned and preprocessed to eliminate noise or other defects.

Extract features: After the data has been collected and preprocessed, the next step is to extract relevant features from the audio recordings or transcriptions. This includes continuous speech features (e.g. pitch and energy), voice quality features (e.g. signal amplitude, energy, duration, phrase, phoneme, word, feature boundaries, temporal structures), spectral-based speech features (e.g. expressions containing happiness have high energy in the high frequency range, and expressions containing sadness have low energy in the same range) or nonlinear TEO-Based features (e.g., to find stress in speech).

Train a classifier: With the extracted features, the next step is to train a machine learning or deep neural network classifier to identify the emotions expressed in spoken language samples. In the studies, using the deep neural network classifier, 71.61% success was achieved in the RAVDESS dataset, 86.1% in the EMO-DB dataset, and 64.3% in the IEMOCAP dataset. In another study, using machine learning, 83.0% success was achieved in the Berlin Emo DB dataset and 70.37% in the RML dataset.

Evaluate the system: Once the classifier has been trained, it is important to evaluate its performance on a separate test dataset to ensure it can correctly classify emotions in unseen data.

3. ARCHITECTURE DESIGN

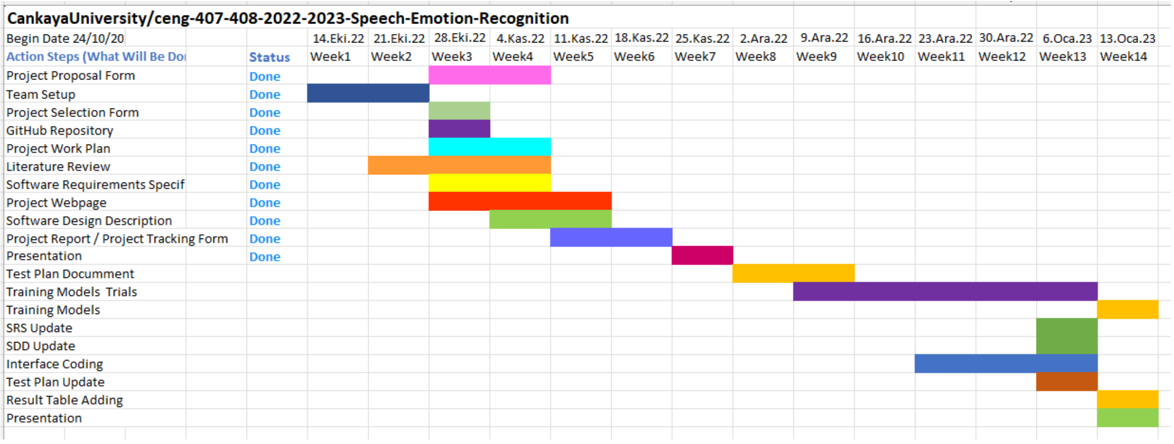
3.1 Simulation Design Approach

We preferred Agile Methodology, one of the software development methodologies, to develop the project. The Agile methodology is a way to manage a project by dividing it into several phases. It includes improvement at every stage. The agile methodology encourages ongoing testing and development throughout the project's software development lifecycle. One of the simplest and most efficient methods for translating a vision for a business need into software solutions is the agile software development methodology. The term "agile" is used to describe methods for developing software that involves ongoing planning, learning, and improvement, teamwork, evolutionary development, and early delivery. It promotes adaptable reactions to change.

We chose the Scrum method, one of the Agile methodologies, for our project. Scrum is iterative and incremental. In Scrum, main work is divided into sprints which should be completed within a certain period of time. Every Sprint includes tasks.

The development team and the project's mentor should hold a meeting of at least 40 minutes each week. Scrum has three main roles which are product owner, scrum master and development team. Product owner delivers the requirements, scrum master manages the development team. Development team is the group of developers who work on the project according to schedule. There are some benefits to using Scrum. The first benefit is that short sprints and constant feedback make it simpler to deal with changes. Another benefit is problems can be handled quickly due to weekly meetings. Also, it makes it possible to create quality products in scheduled time.

We also kept the work done and the progress made at the scheduled times on a Project Work Plan table (Figure 1).



CankayaUniversity/ceng-407-408-2022-2023-Speech-Emotion-Recognition

Begin Date 24/10/20		14.Eki.22	21.Eki.22	28.Eki.22	4.Kas.22	11.Kas.22	18.Kas.22	25.Kas.22	2.Ara.22	9.Ara.22	16.Ara.22	23.Ara.22	30.Ara.22	6.Oca.23	13.Oca.23
Action Steps (What Will Be Do	Status	Week1	Week2	Week3	Week4	Week5	Week6	Week7	Week8	Week9	Week10	Week11	Week12	Week13	Week14
Project Proposal Form	Done			■	■										
Team Setup	Done	■	■												
Project Selection Form	Done			■											
GitHub Repository	Done			■											
Project Work Plan	Done			■	■										
Literature Review	Done		■	■	■										
Software Requirements Specif	Done			■	■										
Project Webpage	Done			■	■	■									
Software Design Description	Done				■	■									
Project Report / Project Tracking Form	Done					■	■								
Presentation	Done							■							
Test Plan Documment									■	■					
Training Models Trials										■	■	■	■	■	
Training Models															■
SRS Update														■	
SDD Update														■	
Interface Coding												■	■	■	
Test Plan Update														■	
Result Table Adding															■
Presentation															■

Figure 1 Project Work Plan

3.1.1 Class Diagram

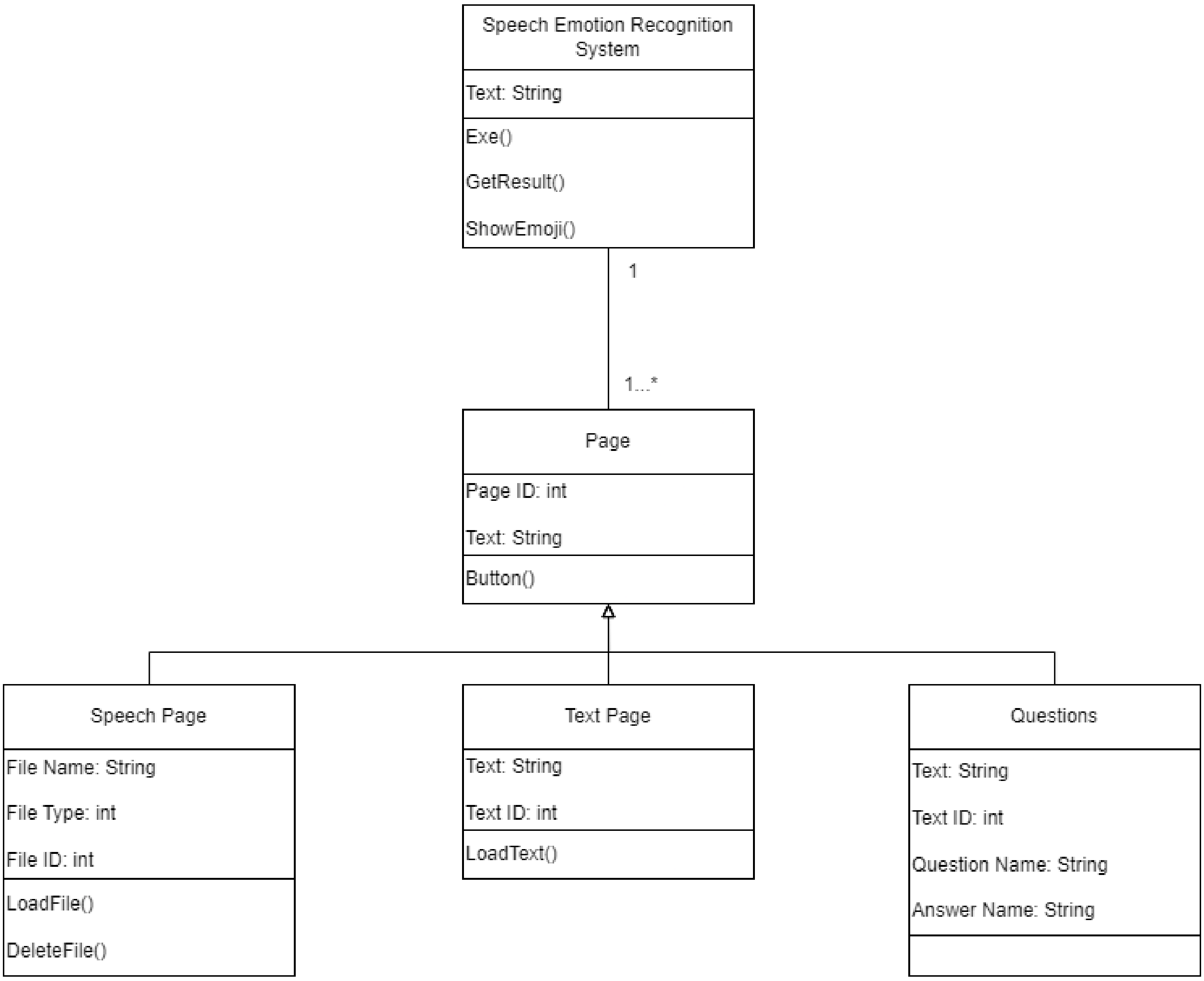


Figure 2 Class Diagram of Speech Emotion Recognition project

The figure (figure 2) above gives information about the connections and interactions within the system. SER (Speech Emotion Recognition) includes other classes, that is, all systems included in the project. The profile class represents all users using the system. The Text class expresses and contains the text written inside itself. In the Speech section, there is the audio file uploaded by the user. On the Questions page, there are brief information about the use of the system and questions about the use of this system. In the Speech emotion Recognition class, there are expressions containing the result of the text or audio file uploaded to the Speech Emotion Recognition system.

In addition to these, the exact measurement or calculation given by the system is directly indicated with a percentile expression. However, there is also information about the content of the uploaded file.

3.2 Architecture Design of Speech Emotion Recognition

3.2.1 Main Page

Summary: On the main page, the user can choose between the audio file loading screen or the page to enter text.

Actor: User

Precondition: User must be open the system.

Basic Sequence:

2. The user can choose what he wants from the "Speech" or "Text" section on the main page.
3. The user is directed to the relevant page according to the selection user has chosen.
4. User can exit from the system by selecting the exit button.

Exception: None

Post Conditions: None

Priority: Medium

3.2.2 Speech Page

Summary: User should upload the audio file that he wants to analyze to the system.

Actor: User

Precondition: User must be select "Speech" from the main page.

Basic Sequence:

1. The user must upload the desired file to the system by dragging it from the "File Upload" section or selecting it from his own computer.
2. After the installation to the system is finished, user should press the "Submit" button.
3. If user has uploaded a wrong file, user should delete the uploaded file by pressing the trash icon.
4. After pressing the "Submit" button, the system will redirect to the "Result" page.

Exception: None.

Post Conditions: After the file is uploaded and submitted, the system should process this file.

Priority: High

3.2.3 Text Page

Summary: The user should write the text that user wants to analyze into the system.

Actor: User

Precondition: User must be select "Text" from the main page.

Basic Sequence:

1. The user should write the text that user wants to write in the relevant part of the "Text" page.
2. After the writing process is finished, user should press the "Submit" button.
3. After pressing the "Submit" button, the system will redirect to the "Result" page

Exception: None.

Post Conditions: After the text is written, the system must process it.

Priority: High

3.2.4 Result Page

Summary: The user will see the result of the text user has written on this page or the result of the audio file he has uploaded.

Actor: User

Precondition: The user has written the text that user wants to be processed from the "Text" section or uploaded the file that he wants to be processed from the "Speech" section.

Basic Sequence:

1. The user should make a choice according to which type user wants to progress on this page.
2. After completing his selection, user should press the "Evaluate" button.
3. After pressing the "Evaluate" button, the system will process this file and generate a result.
4. This result will be displayed to the user on the page with emojis.

Exception: None.

Post Conditions: None

Priority: High

3.3 Activity Diagram

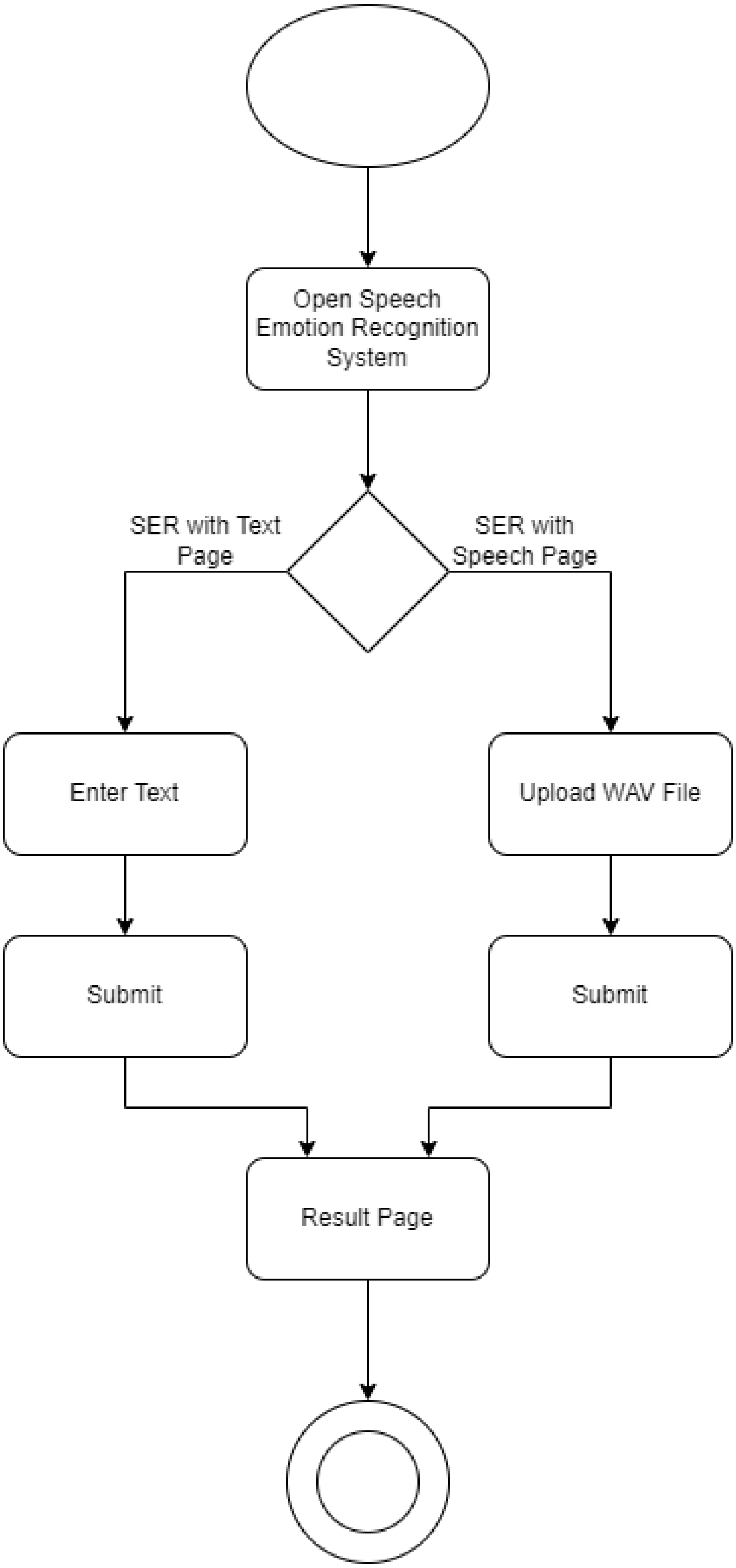


Figure 3 Activity Diagram of SER

Figure 3 shows how the SER works as an activity diagram. The user will be able to use the SER system by following a simple way. The user must decide whether he wants to do sentiment analysis with Text or Voice. After making this decision, the user should either upload an audio file or enter text by following the relevant section. After completing this process, he should go to the result page and evaluate the part he chose. After this process, the sound analysis will be completed and the result will be displayed.

4. USE CASE REALIZATIONS

5.

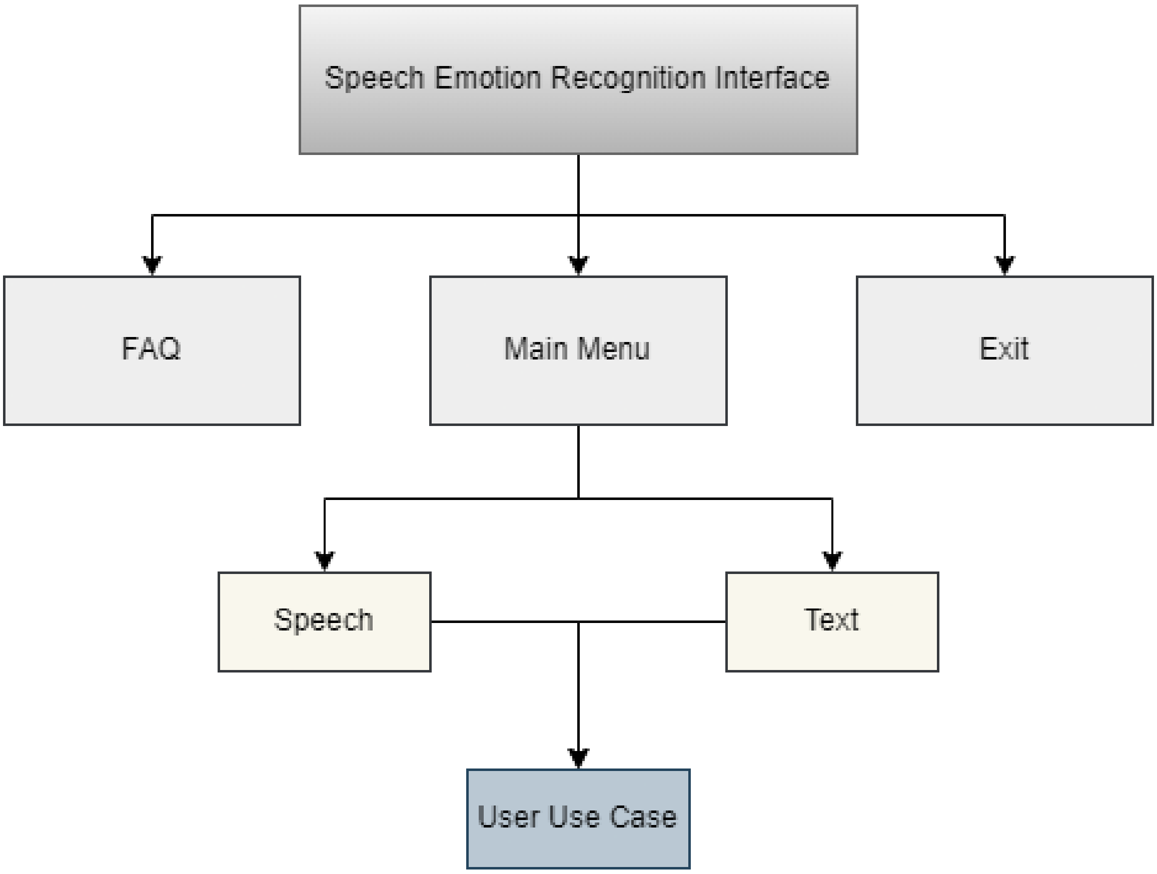


Figure 4 Project Components of Speech Emotion Recognition

4.1 Brief Description of *Figure 4*

Speech Emotion Recognition Project Components are shown in Figure 4. All designed systems of the speech Emotion Recognition project are shown in the block diagram in the figure. The system has 3 subcomponents, 1 of which has its subcomponent.

4.1.1 Frequently Asked Questions Design

Frequently Asked Questions (FAQ) are designed to contain information about the software and the Main Menu. Here are tips on how to use the software. In addition, information such as on which metrics it was resolved and approximately what accuracy value was found are also included.

4.1.2 Main Menu Design

The Main Menu is designed to allow users to easily use the developed Speech Emotion Recognition software. The GUI consists of three main heads. These headings are Main Menu, Frequently Asked Questions and Exit. There are 2 subtitles in the Main Menu. These subheadings are “Speech” and “Text”. Users can upload an audio file and learn emotions from this file. The user can also write an article and learn the emotion from that article.

4.1.3 Exit Design

The exit header is used to exit of the software.

5. DETECTION

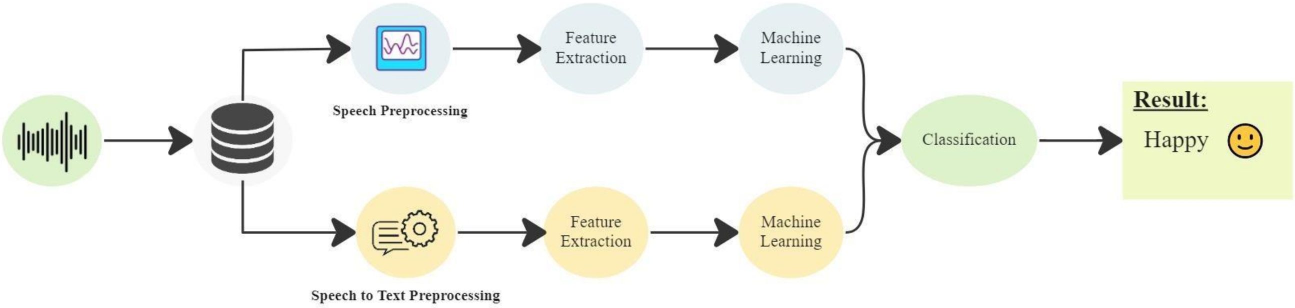


Figure 5 Project System Overview

In this project, the technique of receiving and processing voice data as input and emotion recognition with voice + text was used to create Speech Emotion Recognition.

The data is taken as input by the algorithm. At this stage, sound and text are processed with separate features and extractions. In the next step, the data obtained from both speech and text are subjected to feature extraction. The features obtained as a result of this process are given to the Machine Learning algorithm. In the last stage, the data coming to the classification algorithm shows the emotion according to the "anger", "excited", "sadness", "frustration", "happiness", "neutral" emotions.

6. REFERENCES

[1] https://deliverypdf.ssrn.com/delivery.php?ID=505074082114096020101003119003107029049047056084030089124066087009123068127099064002052103055052104116023097116122003064082097028045036041065026018101019006026003025077028077026080123094096093068095091117071118025067012027074116117097106115010120083126&EXT=pdf&INDEX=TRUE