African Leadership University

COMPUTER SCIENCE CAPSTONE

Project Title:

"Speech Emotion Recognition for Danger Detection: A Tool to Enhance Protection of Women from Gender-Based Violence."

Authors' full names:

Catherine Muthoni

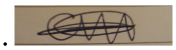
Adviser's names:

Dr Tatenda Duncan Kavu

Year of Graduation: 2023

Declaration

I certify that the work presented in this capstone is my own and that any work that has been performed by others is appropriately cited.

Date: 04/05/2023 Name & Sign: ……Catherine Muthoni…….. ……

Approval (Supervisor)

This is to acknowledge that this capstone project has been submitted with my approval.

Name & Sign:----Tatenda Kavu-----:--Date----08/05/2023-

Acknowledgement

I would like to express my sincere gratitude to all those who have contributed to completing this paper. First and foremost, I would like to thank **God Almighty**. All this would not have been possible without him.

I would also like to thank my capstone supervisor **Dr Tatenda Duncan Kavu**, for his invaluable guidance, support, and encouragement throughout the entire research process. His insightful feedback and suggestions have been instrumental in shaping the direction of this paper.

I would also like to thank the **participants from ALU and beyond** who volunteered their time and filled out my numerous surveys, and participated in my interviews. Without their contributions, this research would not have been possible.

Additionally, I am grateful to the **ALU administration** for the opportunity to study here and for providing the necessary resources to achieve my highest potential.

Finally, I would like to acknowledge the support of **my family and friends**, who provided me with the emotional support and encouragement I needed to complete this research.

Thank you all for your contributions and support.

Abstract

This paper presents a project aimed at developing a tool that utilizes speech emotion recognition to enhance protection for women against Gender-Based Violence (GBV). The paper identifies the significance of speech emotion recognition in detecting dangerous situations and preventing GBV, particularly in Africa, where the rates of GBV are high, and practical solutions are limited. The literature review discusses the evolution of speech emotion recognition, its limitations, and the proposed solution, which includes a close contact information system triggered upon danger detection by sound. The paper then outlines the system analysis, design, and prototype of the proposed solution, which includes a deep learning model called Temporal-aware bI-direction Multi-scale Network (TIM-Net), tested on the RAVDESS test data, achieving an accuracy score of 93%. The proposed system is a speech-activated safety mobile application that detects feelings of fear or danger from human voices, providing women with a means to faster protection. The paper concludes with recommendations for further research and improvements, such as expanding the dataset, adding features, and testing on a larger sample size to ensure effectiveness. Overall, the project demonstrates the potential of speech emotion recognition in preventing GBV and highlights the importance of further research and improvements to make the tool more effective in different regions and cultures.

Keywords: *Speech Emotion Recognition, Gender-Based Violence, Protection, Deep Learning, Danger Detection, Temporal-aware bI-direction Multi-scale Network (TIM-Net)*

Table of Contents

Table of Figures

List Of Images

List of Tables

List of Acronyms, and Abbreviations

GBV - Gender Based Violence

SER - Speech Emotion Recognition

RAVDESS - Ryerson Audio-Visual Database of Emotional Speech and Song

TIM-Net - Temporal-aware bI-direction Multi-scale Network

APK - Android Application Package

GPU - Graphics Processing Unit

UI - User Interface

GMM - Gaussian Mixture Models

HMM - Hidden Markov Models

kNN - k-nearest neighbourhood

SVM - Support Vector Machines

MFCC - Mel Frequency Cepstral Coefficient

Chapter 1: Project Introduction

1.1 Introduction and Motivation

Speech Emotion Recognition (SER) has recently emerged as an integral component within Human-computer Interaction (HCI) and other high-end speech processing systems (Wani, Gunawan, Qadri, Kartiwi, & Ambikairajah, 2021). Its usefulness has been demonstrated in a variety of areas, including smart speakers and virtual assistants, language conversion, treatment, online trading and the entertainment industry (Abbaschian, Sierra-Sosa, & Elmaghraby, 2021). The increased study and exploration of deep learning techniques have also been of great advantage to SER as it has increased the accuracy of its models over time. Given the numerous applications and the great developments in the fields, there is reason to believe that SER could have a major impact in helping in the fight against Gender-Based Violence. This can be achieved by using Speech Emotion Recognition to identify dangerous situations from women's voices. It has been established through studies that gender-specific SER models perform better than generalized models(Shaqra, Duwairi, & Al-Ayyoub, 2019). There are therefore significant inferences that can be made from emotion detection from women's voices. There is also evidence to prove that speech emotion recognition is significant in detecting human stress in dangerous situations (Pavol Partila, Jaromir Tovarek, Rozhon, & Jakub Jalowiczor, 2019).

Given the above context, this project aims to utilize speech-emotion recognition to develop a solution to rapidly protect women against Gender-Based Violence scenarios. Given the current high rates of GBV in Africa and the slow response from authorities and justice systems within the region, the tool derived from this project will benefit all women by empowering them with a tool to protect themselves against assault. Díaz Gorfinkiel, Díaz Gandasegui, & Gómez García (2021), after interviews with experts in matters Gender Based Violence, also uphold the effectiveness of such a tool by saying that it will help women by providing knowledge of their reactions and opening a path to self-discovery in at-risk situations.

1.2 Problem Statement

Gender-based violence can be described as any type of malicious behavior directed at an individual that is based on gender. (Dugbazah , Glover, Mbuli, Kungade, & Shikwambane, 2022). It has been a threat to the safety of girls and women in Africa since time immemorial. It is estimated that almost 1 in 3 women globally has experienced some form of violence in her lifetime(Tackling gender-based violence, once and for all, 2021), and women in low- and lower-middle-income countries are disproportionately impacted with 30% of African Women having experienced GBV at least once (Dugbazah, Glover, Mbuli, Kungade, & Shikwambane, 2022). Efforts have been made worldwide to prevent Gender Based Violence and protect victims against it. However, there is a significant lack of active and practical solutions to deal with Gender-Based Violence, particularly in Africa. Technological solutions like Safetipin App, GovChat and Nokaneng, among others, have been adopted to reduce the risk of violence among women by providing platforms for women to report GBV cases, informing women of safe spaces around their area and educating them on GBV (Binder & Poulton, 2022). While all these solutions are helpful, they are more informative than actionable which means that they only work after the damage has occurred and not during the incident. Díaz Gorfinkiel, Díaz Gandasegui, & Gómez García (2021) noted that despite the approval of a law to recognize and deal with Intimate Partner violence, a form of GBV, there has been minimal progress in eradicating GBV and achieving universal social rejection of the same. This shows that this is seen as a problem requiring urgent attention. Hence, it is precisely in this context that this project aims to develop a tool that implements speech-emotion recognition using machine learning to detect dangerous situations and protect the victim from further assault in addition to other features.

1.3 Project's main objective

The main objective of this project is to develop a tool that implements speech emotion recognition using machine learning to detect dangerous situations and protect the victim from further assault in addition to other features.

1.4 List of the project's specific objectives

The following describes the objectives aimed to be achieved during the implementation of this project:

- Data collection, analysis and preparation: Data will be collected from the open source sources. The data used for this project with be collected from the Ryerson Audio-Visual

Database of Emotional Speech and Song (RAVDESS) dataset. This data will be prepared and made ready for machine learning processing.

- Building of the machine learning model: The project will use a Temporal-aware bI-direction Multi-scale Network (TIM-Net) model.
- Building of the application's interface and API: The UI of the application will be built using Flutter. Django will be used for the API and integration of the model while PostgreSQL will be used for persistence.
- Deployment of the system: The mobile app will be deployed using Firebase and saved as an APK on AppTester while the API will be deployed on Render.
- User testing, collection and implementation of feedback: The testing will be conducted on the project's case study which will be African Leadership University students.

1.5 Project scope

This project will be centered around African women, particularly those based in Rwanda. The case study for this phase of the project will be female students and staff at The African Leadership University. This is because of the diversity in nationality and in that population. The tool that will be developed from this project will only be a mobile application. Other user-friendly approaches like wearables will be considered beyond the scope of this project. Due to the time constraints involved in the project, the databases for training the speech emotion recognition model will be outsourced from open-source resources. In addition, the features that will be included in the mobile application will be danger detection and close contact information. Recording of surroundings will not be included in the scope of this project but will be considered in later phases. Finally, the project will be undertaken for a duration of four months, that is between January and April 2023.

1.6 Technical requirements

The following describes the functional requirements of the application:

- Google Colaboratory and Jupyter Notebook with GPU will be used to program the model
- TensorFlow and Librosa will be used as libraries for creating the deep learning model.
- Flutter will be used to build the UI for the application
- PostgreSQL will be used for the data persistence of the application

- Firebase and Render will be used to deploy the application

1.7 Project time frame (Gantt chart)

Link to Gantt Chart

1.8 Project Budget

Plagiarism Checker and Citation Aid - 30$(3 month subscription)

Chapter 2: Literature Review

2.1 Introduction

Abbaschian, Sierra-Sosa, & Elmaghraby (2021) define Speech Emotion Recognition as the task of recognizing emotions from speech signals. A Speech Emotion Recognition system targets the speaker's existence of varied emotions by extracting and classifying the prominent features from a preprocessed speech signal (Wani, Gunawan, Qadri, Kartiwi, & Ambikairajah, 2021). Given the immense developments in gender-specific speech emotion recognition and the use of speech emotion recognition in danger detection, there is potential for speech emotion recognition to aid in the protection of women against Gender Based Violence. Various scholars like Díaz Gorfinkiel, Díaz Gandasegui, & Gómez García (2021) have studied the use of technology particularly SER in the fight against Gender Based Violence.

2.2 The History and Evolution of Speech Emotion Recognition

There is a uniform voice from scholars that the study of Speech Emotion Recognition has greatly evolved over time. Wani, Gunawan, Qadri, Kartiwi, & Ambikairajah (2021) highlighted the very beginning of this journey during the era of classical automatic speech recognition systems which focused less on some of the essential paralinguistic information passed on by speech like emotions. They then proceeded to highlight that upon the discovery of emotion recognition, simple models like Linear discriminant classifiers, Gaussian Mixture Models (GMM), Hidden Markov Models (HMM), k-nearest neighbourhood (kNN) classifiers, Support Vector Machines (SVM), decision tree were used to classify emotions dependent on their acoustic features of intrigue. In recent times, deep learning classifiers such as Deep Belief Networks, Deep Neural Networks, Deep Boltzmann Machines, Convolution Neural Networks, Recurrent Neural Networks, and Long Short-Term Memory have become more common. With scholars looking to improve feature extraction abilities in their models, stronger classifiers are being developed.

2.3 Why Speech Emotion Recognition for Rapid Protection against Gender Based Violence?

There exists numerous solutions to prevent Gender Based Violence majority of which have from the global north. Cardoso, Sorenson, Webb, & Landers (2019) highlighted Athena, a wearable that sends information to family and friends through the click of a button. It has a silent and loud alarm mode for different scenarios. They also highlighted Safelet, which is a wrist wearable which informs family and friends of the situation at hand through the tap of a button. While all these solutions are good and applicable, they may not be effective because before the victim reaches for the button, the assailant may have already gained access to it. There are also cases where people freeze during such scenarios making them unable to or forget to press a button. The advantage of Speech Emotion Recognition(SER) systems is that one just needs to let out an alarm and the device will automatically act accordingly. Partila, Tovarek, Rozhon, & Jalowiczor (2019) affirmed this by saying that a reason for dealing with speech emotion recognition is its broad applicability and utility. Even people with disabilities can easily use these devices.

2.4 Existing solutions against GBV applying SER

Upon research, it was noted that there are indeed very few technological solutions for rapid protection against Gender-Based Violence that implement Speech Emotion Recognition. In fact, in Africa there is none. This already affirms the need for such a solution given the current high rates of GBV in the continent. Díaz Gorfinkiel, Díaz Gandasegui, & Gómez García (2021) highlighted one such solution which is still under development in Spain. The solution is a wearable that uses physiological sensory data, audio analysis and machine-learning algorithms to detect intense emotional states (such as panic, fear or stress) caused by Intimate Partner Violence situations. The device also facilitates the geo-localization of the victim, gathering evidence of the alleged offence and transferring the warning to a 'guard circle', who can act immediately to protect the user (Díaz Gorfinkiel, Díaz Gandasegui, & Gómez García, 2021). While this is a good solution, there is still a need for something similar within the African context because of differences in speech accent and accessibility factors. In addition to the features in the mentioned device, the product proposed in this proposal will also include an alarm system will be triggered upon danger detection by sound. In later phases, there will also be a fresh data collection of data from African voices so that the product is better suited for the African demographic.

2.5 Limitations of SER Technologies

While Speech Emotion Recognition(SER) is a good solution to implement in the rapid protection against GBV, it is important to acknowledge its limitations. Díaz Gorfinkiel, Díaz Gandasegui, & Gómez García (2021) pointed out that SER systems may be prone to insufficient signal coverage or false alarms when the physiological parameters incorrectly indicate a situation of violence. The process of social acceptance of the new technology especially in Africa which has the greatest digital divide, may also be very slow thus hindering achievement of the goal. Cueto (2015) also critiqued the technology saying that sexual assault might be too big of a problem, one that is too finely enmeshed in our culture, and whose roots are too complex and widespread for there to be any technological quick fixes. This is a very important point to consider. In the proposed solution, alongside the product, part of its proceeds will be used in organising forums where men are taught about respecting women and exercising healthy masculinity.

2.6 Conclusion

It is clear from the previous dissertations that SER has the potential in promoting the rapid protection of women against Gender-Based Violence. It is however also clear that the scholars highlighted solutions and works from developed countries which show that little effort has been made in the exploration of the incorporation of technology in at-risk protection against Gender-Based Violence within Africa. In conclusion, there is still a need for a solution that rapidly protects women against GBV in Africa because there is none within and also, the existing inefficient systems put in place emphasize the need for self-empowerment in this fight. Given the literary analysis above, the proposed solution in this project is an application that detects emotions of fear or danger through sound and sends an alert to the listed contacts indicating that their kin is in danger.

Chapter 3: System Analysis, Design and Prototype

3.1 Introduction

In the previous chapter, the need for a solution to rapidly protect women against Gender Based Violence was highlighted. It was also identified that Speech Emotion Recognition technology has the potential to offer such a solution. In this chapter, audio data is collected, prepared and trained with a machine learning model to predict its emotion. The results from the model are analyzed and recommendations on a solution are made based on the analysis. The architecture of the recommended solution is discussed in depth as well as its functional and non-functional requirements. Lastly, the standard technologies involved in the development of the proposed system are highlighted.

3.2 Data analysis and presentation

3.2.1 Data collection

The main data source for this project was The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) by Livingstone SR, Russo FA (2018). It is an emotional-based database, containing speech emotional sounds collected from actors' voices. There are 8 emotions assessed in this database including calm, happiness, sadness, anger, fear, disgust, surprise. The database consists of 7356 audio files from 24 different male and female actors of different ages but in the context of this project, 1440 were used. This was the preferred database because compared to other publicly available emotional databases, RAVDESS is more realistic and non-scripted with no language restriction.

3.2.2 Data Preparation(Feature Extraction)

Before any kind of analysis was performed on the data, it had to be transformed to fit the needs of the project. A data frame was generated containing the path of each file and emotions displayed through that file. The figure below shows the first 5 rows of the resulting database:

	Emotions	Path
0	neutral	/content/drive/My Drive/MachineLearningStuff/s...
1	angry	/content/drive/My Drive/MachineLearningStuff/s...
2	calm	/content/drive/My Drive/MachineLearningStuff/s...
3	disgust	/content/drive/My Drive/MachineLearningStuff/s...
4	neutral	/content/drive/My Drive/MachineLearningStuff/s...

Figure 1. The first 5 rows of Audio files and corresponding emotions in the RAVDESS dataset

Signal processing was then conducted to convert the audio files into a numerical form that can be used by the deep learning models later. In addition, significant features were extracted from the processed files to distinguish the embedded emotions easier. The features extracted were the Mel Frequency Cepstral Coefficient(MFCC) which represents the short-term power spectrum of a sound. The code snippet below shows the whole signal processing and feature extraction process:

```
get_feature(file_path: str, mfcc_len: int = 39, mean_signal_length: int = 100000):

    signal, fs = librosa.load(file_path)
    s_len = len(signal)

    if s_len < mean_signal_length:
        pad_len = mean_signal_length - s_len
        pad_rem = pad_len % 2
        pad_len //= 2
        signal = np.pad(signal, (pad_len, pad_len + pad_rem), 'constant', constant_values
    else:
        pad_len = s_len - mean_signal_length
        pad_len //= 2
        signal = signal[pad_len:pad_len + mean_signal_length]
    mfcc = librosa.feature.mfcc(y=signal, sr=fs, n_mfcc=39)
    mfcc = mfcc.T
    feature = mfcc
    return feature
```

Figure 2: MFCC Feature Extraction

3.2.3 Model Creation

Based on the pre-processed data from the previous section a deep learning model was developed. The model was a Temporal-aware bI-direction Multi-scale Network, termed TIM-Net, which is a novel temporal emotional modelling approach to learning multi-scale contextual affective representations from various time scales. This model was adapted from a study by Jiaxin et al. (2022). It provides a new approach to Speech Emotion Recognition by exploring how to model the temporal patterns of speech emotions from dynamic temporal scales other than the usual way of mining spatiotemporal information from hand-crafted features (Jiaxin et al., 2022).

The model was structured to have 2 main components, the feature extraction part and the temporal-aware block. The feature extraction part comprises a bi-directional module and fusion module while the temporal aware block consists of a series of 1D convolutional and dense layers. The figure below shows a detailed structure of the model:

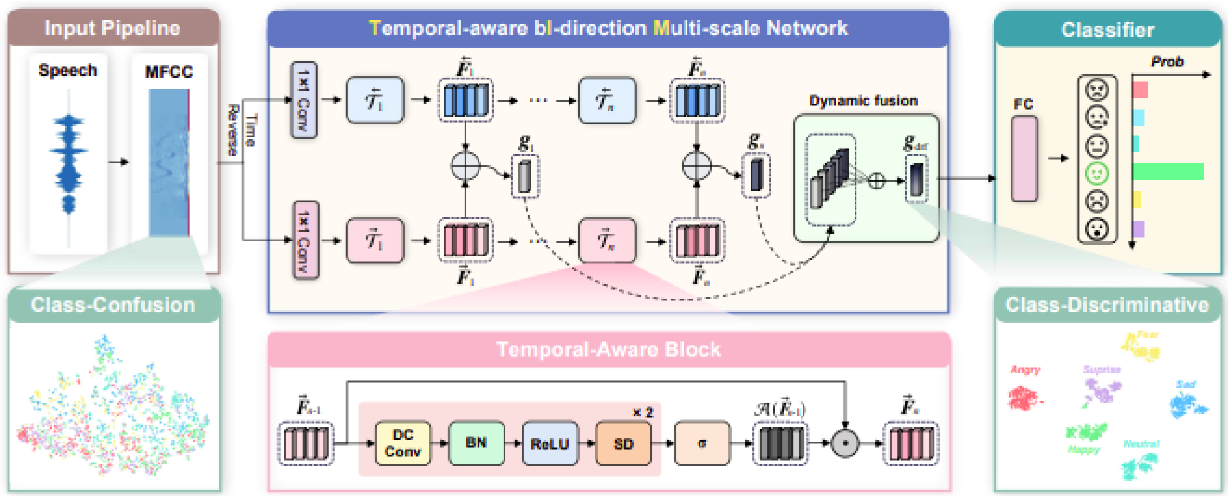


Figure 3: TIM-Net Model Structure (Ye, 2022)

The model was fit with a batch size of 64 and 300 epochs. The model was tested on the test data, and the predicted outputs for the first ten rows of the test data were as follows:

	Predicted Labels	Actual Labels
0	happy	happy
1	sad	sad
2	fear	fear
3	fear	pleasure
4	fear	fear
5	happy	surprise
6	sad	sad
7	neutral	pleasure
8	sad	sad
9	pain	pain

Figure 4: Model predictions vs Actual labels as tested on test data

.

3.2.4 Model Evaluation

A classification report was generated for the model. It presented the precision, recall, f1 score and support for each predicted emotion in the model. The main focus of this project was on the fear emotion, which has an f1 score of 97% which was impressively the highest recorded score. The accuracy score of the entire model was recorded as 93%. A figure of the full classification report is shown below:

	precision	recall	f1-score	support
angry	1.00	0.96	0.98	24
calm	0.97	0.97	0.97	29
disgust	1.00	0.92	0.96	12
fear	0.95	1.00	0.97	18
happy	1.00	0.75	0.86	12
neutral	0.82	0.90	0.86	10
sad	0.84	0.84	0.84	19
surprise	0.87	1.00	0.93	20
accuracy			0.93	144
macro avg	0.93	0.92	0.92	144
weighted avg	0.93	0.93	0.93	144

Figure 5: Classification matrix

3.3 Interpretation of findings/results

From the findings gathered, it can be concluded that Speech Emotion Recognition has a high potential to detect feelings of fear or danger from human voices. A speech-activated safety mobile application was proposed as a suitable solution.

The next section highlights the details of the proposed system.

3.4 Software Development Life Cycle of the new system

As mentioned in the previous section, the proposed system was a speech-activated safety mobile application. The application would be developed using the Big Bang Software Development Life Cycle(SDLC) methodology (GeeksforGeeks, 2021). This methodology was decided upon because of its flexible nature. It would combine time, effort and resources when building the application. In addition, due to the experimental nature of this project, the product's requirements would be built as they arrived. In using this model, there was a possibility that the end product would differ from the initial requirements. The figure below shows an illustration of the Big Bang methodology.

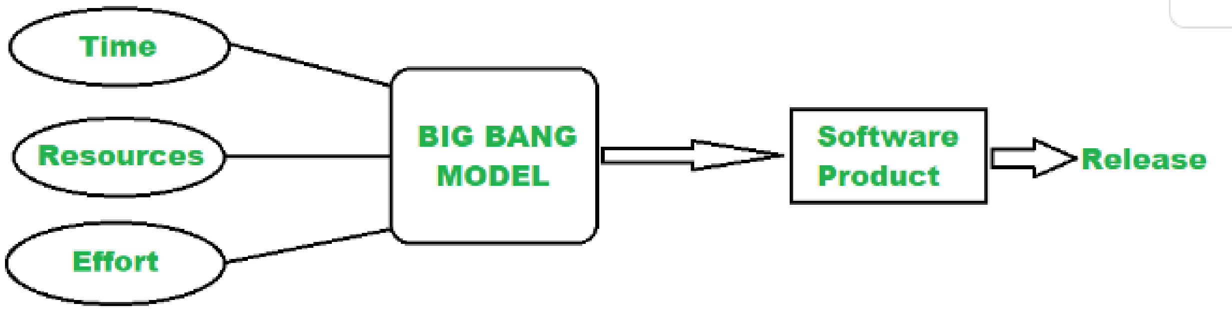


Figure 7. Big Bang SDLC Methodology

3.5 Architecture Design of the new system

The proposed system architecture consists of five main components: The mobile user interface(UI), The device microphone, The embedded machine learning model, the application programming interface(API) and the database. The user interface will be the component that the user will be interacting with. The user will create an account, log in, add contact people and confirm danger alerts through the application. The microphone will be used for recording the surroundings while the embedded ML model will be used to detect emotions of danger from the sound recorded by the microphone. Users' and Contact people's details will be saved on the database. The API will be used to connect the database to the interface. The figure below gives an illustration of the system's architectural design:

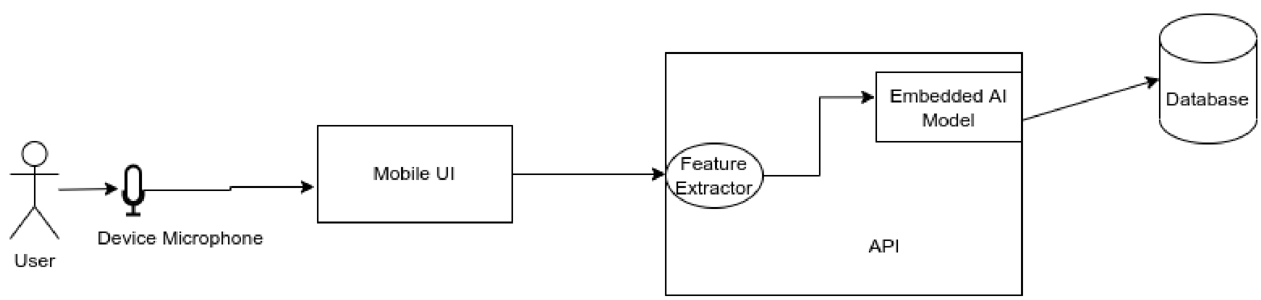


Figure 8. Proposed System Architecture

3.6 Description of the new system

3.6.1 Functional Requirements

The following describes the functional requirements of the application:

User authentication: The user shall be able to create an account and log into the system.
Contact people addition: The user shall be able to add the details of one or more contact people to the system. The contact people shall be contacted when the application detects danger.
Danger detection: The system shall be able to detect danger from sound recorded through the device's microphone.
Messaging contact people: The system shall send a text message to the user's contact people in the event of danger detection.

3.6.2 Non-Functional Requirements

The following describes the non-functional requirements of the application:

Portability and compatibility: The system shall be a cross-platform mobile application compatible with Android and iOS devices.
User Experience: The system's user interface shall be uniform across all pages with visually appealing colors and typography. The flow of pages shall also be clearly structured
Security: Both the system and the database will be strongly password protected to prevent external users from accessing the data. In addition, additional code will be added to the forms in the application to prevent SQL injection from malicious users.

3.6.3 Development Tools

The software and hardware tools used in developing the proposed system are described in this section.

Programming Languages

Python: Python has numerous libraries useful in data science. For this project, it was used to develop the Machine Learning model. Django, a Python framework, was also used to develop the API.

Dart: Dart is the main language used in Flutter, a framework used for cross-platform mobile development. It was chosen for this project to allow for the mobile application to run on both Android and iOS.

SQL: SQL is the main language used in database management systems. For this project's scope, it was used to write queries for PostgreSQL, the database used in this application.

Software Tools

Google Colab Notebook: This tool was used as the main environment for developing the speech emotion recognition model.

Tensorflow: This software was used for the creation of the model. Tensorflow-Lite was also used to create a mobile friendly model.

Scikit learn: Was used to provide libraries for data processing, model training and evaluation.

Librosa: Was used for handling audio files in Python.

Android Studio: It was used as the IDE to develop the mobile application.

Visual Studio Code: It was used as the IDE for the development of the Django REST API

Github: It was used for creating the project's repository and version control.

Render: Was used for the deployment of the database and the API

Hardware Tools

Operating System: Ubuntu 20.04, Windows 10

Laptop: HP 15 Intel Core i7 processor 16GB RAM

3.7 Illustration of New system

3.7.1 Data Flow Diagram

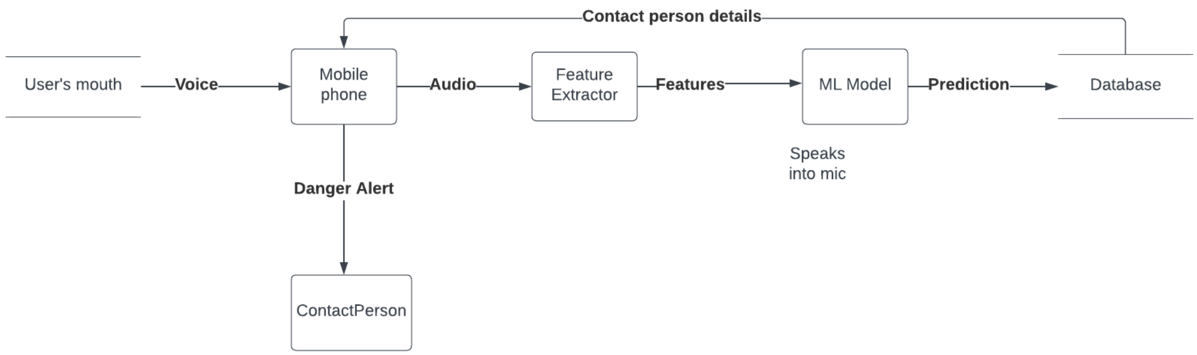


Figure 9. Data Flow Diagram

3.7.2 Use Case

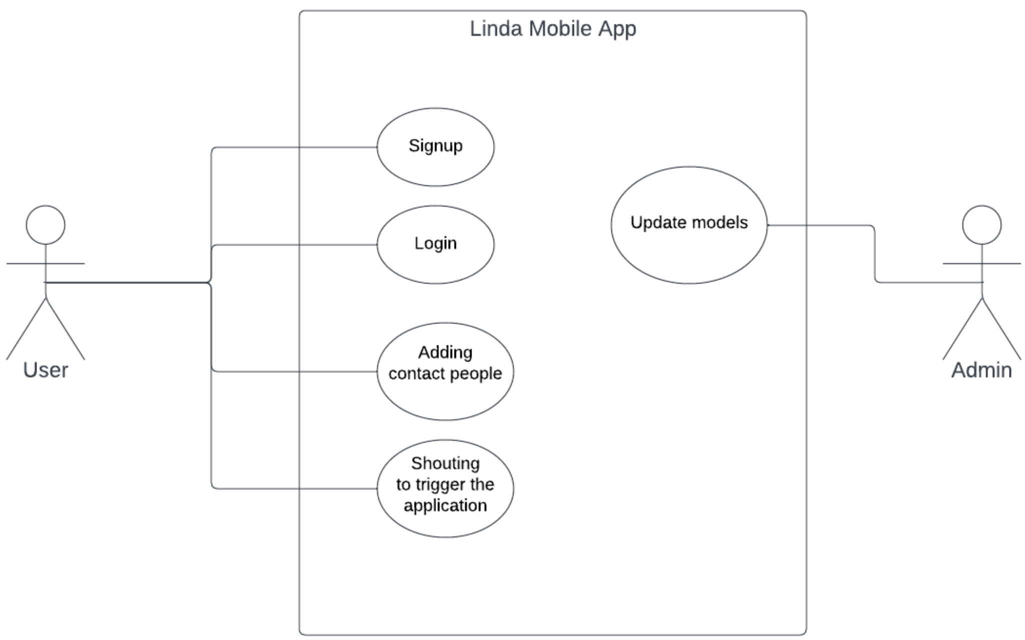


Figure 10. Use Case Diagram

3.7.3 Data Dictionary

Entity Name	Entity Description	Column Name	Column Description	Data Type	Length	Primary Key	Nullable
User	These are the users of the application. They create accounts on the application	ID	The unique number that identifies the user	INT	10	TRUE	FALSE
		Email	The email address of the user	VARCHAR	255	FALSE	FALSE
		Name	The name of the user	VARCHAR	255	FALSE	FALSE
		Password	The password the user set for their account	VARCHAR	255	FALSE	FALSE
		Phone	The user's phone number	INT	10	FALSE	FALSE
		Gender	The user's gender	VARCHAR	255	FALSE	FALSE
ContactPerson	A person that the application user lists as a person to contact in case they are in danger	ID	The unique number that identifies the user	INT	10	TRUE	FALSE
		Email	The email address of the user	VARCHAR	255	FALSE	FALSE
		Name	The name of the user	VARCHAR	255	FALSE	FALSE
		Phone1	The user's phone number	INT	10	FALSE	FALSE
		Phone2	The user's phone number	INT	10	FALSE	FALSE
Contact	The entity that contains the relationship between one user and one contact person. Avoids a many-many relationship between the contact and user entities	ID	The unique identifier of the contact	INT	10	TRUE	FALSE
		UserID	The foreign key representing the user's ID	INT	10	FALSE	FALSE
		ContactPersonID	The foreign key representing the contact person's ID	INT	10	FALSE	FALSE

Figure 11: Data Dictionary

3.7.4 Entity Relationship Diagram

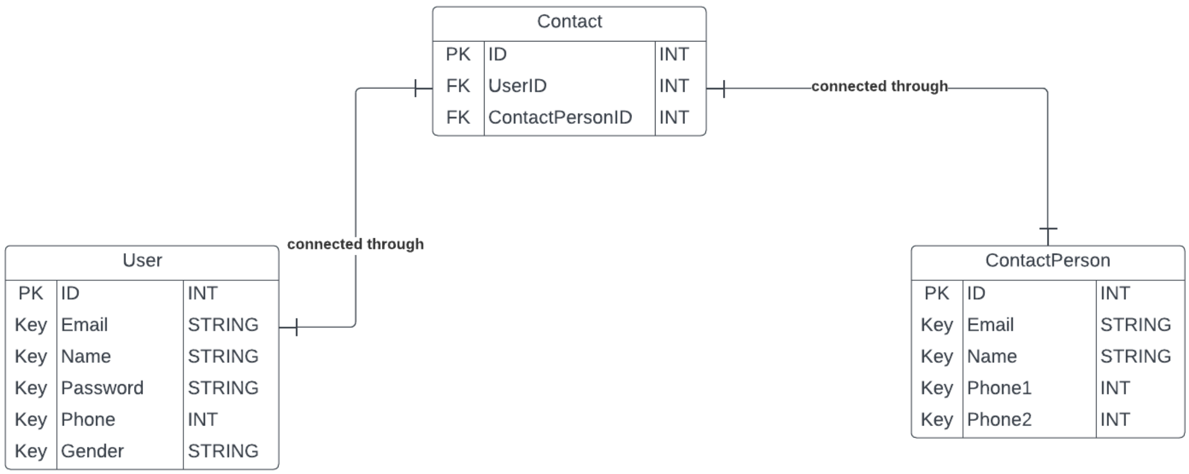


Figure 12. Entity Relationship Diagram(ERD)

Chapter 4: Implementation and Testing

4.1 Implementation and coding

4.1.1 Introduction

In this chapter, the practical aspects of implementing the proposed solution to the problem at hand are delved into. In addition, a description of the proposed solution and an outline of the tools and technologies used in developing the solution are given.

4.1.2 Description of implementation tools and technology

The proposed solution involved developing a mobile application that detects emotions of fear through sound and alerts the user’s registered contacts of possible danger upon their kin. To implement the solution, the following tools and technologies were used:

Flutter - Flutter was used for creating the mobile application's user interface and client-side logic. We also used the Flutter plugin Dio to communicate with our Django-based back-end.

Django - Django was used for creating the application's server-side logic and API endpoints. Several Django models were created to represent the database schema and Django's built-in ORM was used to interact with the Postgresql database. Django's authentication system was also used to secure the API endpoints and implement custom middleware for handling cross-origin requests.

PostgreSQL - Postgresql was used as the primary database for storing user data and other application-related data. Several tables and relationships were created to represent the application's data model and leveraged Postgresql's advanced indexing and query optimization features to ensure fast and efficient data retrieval.

4.1.3 Application Screenshots

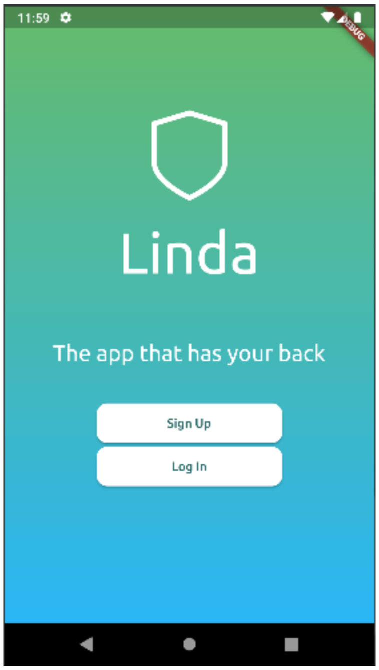


Figure 13. Landing Page/Splash

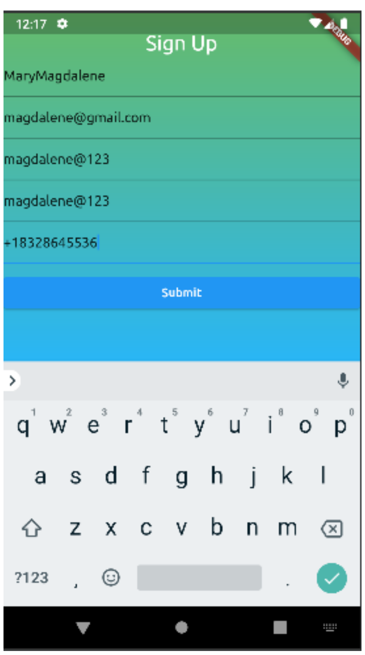


Figure 14. Sign Up

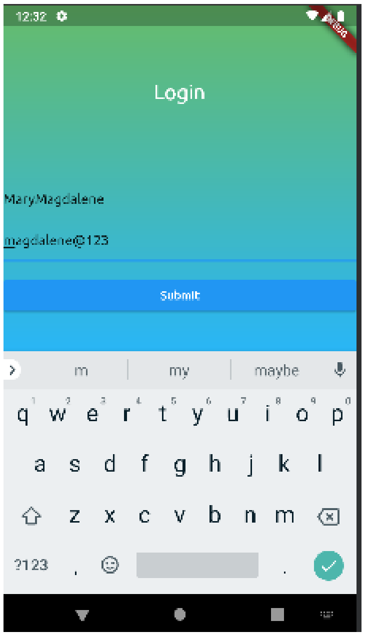


Figure 15. Login

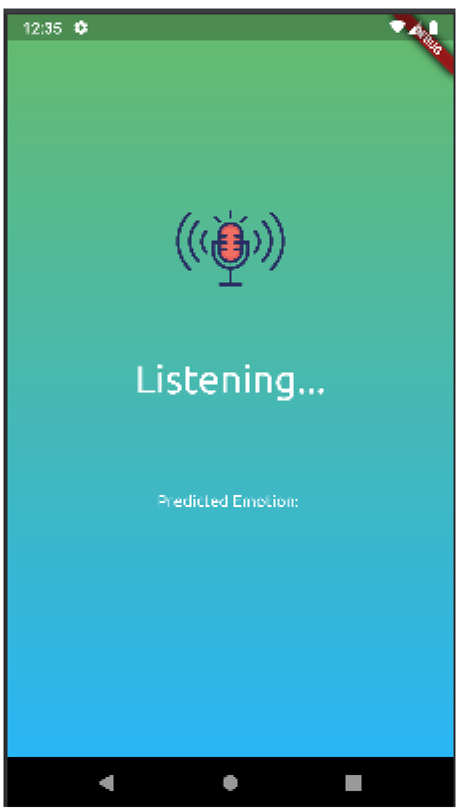


Figure 16. Recording sound

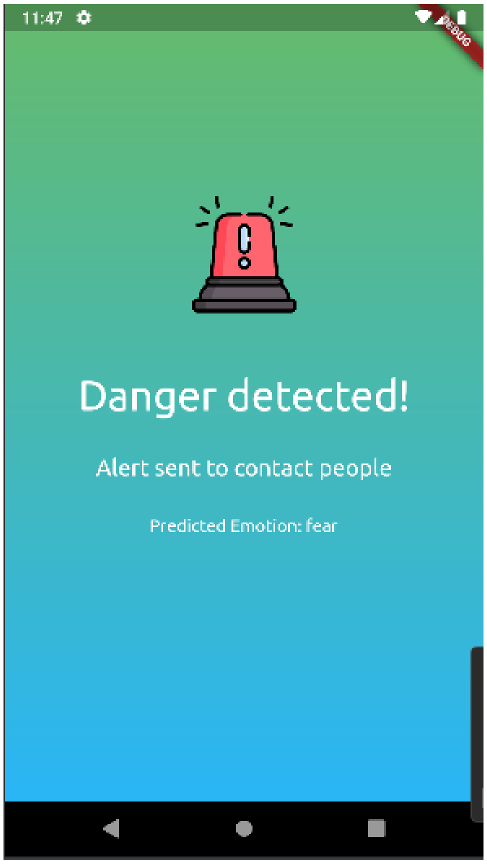


Figure 17. Danger detection

4.2 Testing Results

4.2.1 Introduction and the objective of testing

This section highlights the different tests made on the application after building. The main objective of testing this application is to to find maximum defects the application while validating whether the program is working as per the earlier stated requirements or not.

4.2.2 Functional, Unit and Validation Testing outputs

4.2.2.1 Testing User Creation with different inputs

```
def test_create_valid_user(self):
    response = self.client.post(
        self.register_url,
        data=self.valid_payload,
        format='json'
    )
    self.assertEqual(response.status_code, status.HTTP_201_CREATED)

def test_create_invalid_user(self):
    response = self.client.post(
        self.register_url,
        data=self.invalid_payload,
        format='json'
    )
    self.assertEqual(response.status_code, status.HTTP_400_BAD_REQUEST)
```

Figure 18. Test 1

4.2.2.2 Test User Login with different inputs

```
def test_login_valid_user(self):
    response = self.client.post(
        self.login_url,
        data=self.valid_payload,
        format='json'
    )
    self.assertEqual(response.status_code, status.HTTP_200_OK)
    self.assertIn('token', response.data)

def test_login_invalid_user(self):
    response = self.client.post(
        self.login_url,
        data=self.invalid_payload,
        format='json'
    )
    self.assertEqual(response.status_code, status.HTTP_400_BAD_REQUEST)
```

Figure 19. Test 2

4.2.2.3 Test Contact person creation

```
self.user = get_user_model().objects.create_user(
    email='test@example.com',
    password='test_password'
)
self.token = Token.objects.create(user=self.user)
self.client.credentials(HTTP_AUTHORIZATION='Token ' + self.token.key)
self.valid_payload = {
    'name': 'Test Contact',
    'email': 'test_contact@example.com',
    'phone': '1234567890'
}
self.invalid_payload = {
    'name': 'Test Contact',
    'email': 'invalid_email',
    'phone': '1234567890'
}
```

```
def test_create_valid_contact(self):
    response = self.client.post(
        self.url,
        data=self.valid_payload,
        format='json'
    )
    self.assertEqual(response.status_code, status.HTTP_201_CREATED)
```

Figure 21. Test 3

```
def test_create_invalid_contact(self):
    response = self.client.post(
        self.url,
        data=self.invalid_payload,
        format='json'
    )
    self.assertEqual(response.status_code, status.HTTP_400_BAD_REQUEST)
```

Figure 22. Test 4

4.2.2.4 Test Model Prediction

```
@patch('app.views.get_feature')
def test_predict_emotion_not_fear(self, mock_get_feature):
    mock_get_feature.return_value = np.array([0.5, 0.6, 0.7])
    with patch('app.views.SermodelConfig') as mock_config:
        mock_model = MagicMock()
        mock_model.model.predict.return_value = np.array([[0, 0, 0, 0, 0, 1, 0, 0]])
        mock_config.ser_model = mock_model
        response = self.client.post(
            self.url,
            data=self.payload,
            format='json'
        )
    self.assertEqual(response.status_code, status.HTTP_200_OK)
    self.assertEqual(response.data['data'], 'neutral')
    self.assertEqual(mock_get_feature.call_args[0][0], 'test_audio_file')
    self.assertEqual(mock_model.model.predict.call_args[0][0].shape, (1, 196, 39))
```

```
@patch('app.views.get_feature')
def test_predict_emotion_fear(self, mock_get_feature):
    mock_get_feature.return_value = np.array([0.5, 0.6, 0.7])
    with patch('app.views.SermodelConfig') as mock_config:
        mock_model = MagicMock()
        mock_model.model.predict.return_value = np.array([[0, 0, 0, 1, 0, 0, 0, 0]])
        mock_config.ser_model = mock_model
        response = self.client.post(
            self.url,
            data=self.payload,
            format='json'
        )
    self.assertEqual(response.status_code, status.HTTP_200_OK)
    self.assertEqual(response.data['data'], 'fear')
    self.assertEqual(mock_get_feature.call_args[0][0], 'test_audio_file')
    self.assertEqual(mock_model.model.predict.call_args[0][0].shape, (1, 196, 39))
```

Figure 24. Test 5

4.2.2.5 Test Authorisation

```
def test_predict_unauthorized(self):
    self.client.credentials()
    response = self.client.post(
        self.url,
        data=self.payload,
        format='json'
    )
    self.assertEqual(response.status_code, status.HTTP_401_UNAUTHORIZED)
```

Figure 25. Test 6

Chapter 5: Conclusion & Recommendations

5.1 Introduction

Speech Emotion Recognition (SER) has been widely explored in recent times as a key component in human-computer interaction and speech processing systems. It has been applied in various fields such as language conversion, smart speakers, online trading, and entertainment. The study and advancement of deep learning techniques have benefited SER as it has improved the accuracy of its models. In light of the numerous applications of SER, this project aimed to use it to protect women against Gender-Based Violence by detecting dangerous situations from women's voices. This chapter presents the conclusion, areas for further research, and recommendations for the study.

5.2 Conclusion (s)

The project successfully achieved its main objective, which was to develop a tool that implements speech emotion recognition using machine learning to detect dangerous situations and rapidly protect the victim from further assault. A Temporal-aware bI-direction Multi-scale Network(TIM-Net) model was used to build the machine learning model. The application's interface and API were built using Flutter and Django. The app was saved as an APK on AppTester and deployed on Render. The testing was conducted on the project's case study, African Leadership University. The tool derived from this project empowers women by providing a means to protect themselves against Gender-Based Violence and provides them with evidence to seek legal assistance.

5.3 Area(s) for further research

Further research can be done on the effectiveness of speech emotion recognition in detecting other forms of violence, such as child abuse or domestic violence. Research can also be done on the tool's effectiveness in different regions and cultures to determine if there are any cultural or regional differences in its effectiveness. In addition, research can be conducted on the various ways in which machine learning can be applied in building more accessible software solutions for the prevention of Gender Based Violence.

5.4 Recommendations for improvement for the study

The project can be improved by expanding the dataset used for training the machine learning model to include more diverse voices and scenarios, particularly from Africa. Additionally, extra features like geo-location and recording of the surroundings can be added to increase the effectiveness of the application. The machine learning algorithm can also be applied on edge devices and microcontrollers so as to make more user-friendly tools and to cater for the limitations that come with a mobile application which include high latency and high battery consumption. Finally, further testing can be done on a larger sample size to ensure the tool's effectiveness in protecting women against Gender-Based Violence.

In conclusion, this project has demonstrated the potential of speech emotion recognition in helping to fight Gender-Based Violence. Using machine learning to detect dangerous situations from women's voices, women can be empowered to protect themselves against assault and seek legal assistance. Further research and improvements can be made to make the tool more effective in different regions and cultures.

REFERENCES

Abbaschian, B. J., Sierra-Sosa, D., & Elmaghraby, A. (2021). Deep learning techniques for speech emotion recognition, from databases to models. Sensors, 21(4), 1249. doi:10.3390/s21041249

Binder, G., & Poulton, C. (2022, September 15). UNICEF blog. Retrieved November 1, 2022, from https://blogs.unicef.org/east-asia-pacific/category/child-protection/

Cardoso, L. F., Sorenson, S. B., Webb, O., & Landers, S. (2019). Recent and emerging technologies: Implications for women's safety. Technology in Society, 58, 101108. doi:10.1016/j.techsoc.2019.01.001

Cueto, E. (2015, September 12). Could A "high tech rape whistle" prevent sexual assaults? some are hopeful, but there are also a lot of questions. Retrieved February 8, 2023, from https://www.bustle.com/articles/110079-could-a-high-tech-rape-whistle-prevent-sexual-assaults-some-are-hopeful-but-there-are-also

Díaz Gorfinkiel, M., Díaz Gandasegui, V., & Gómez García, M. V. (2021). New technology proposals for tackling intimate partner violence: Challenges and opportunities. Technology in Society, 67, 101714. doi:10.1016/j.techsoc.2021.101714

Dugbazah, J., Glover, B., Mbuli, B., Kungade, C., & Shikwambane, N. (2022). Leveraging Smart Technologies to tackle gender-based violence in Africa: Auda-NEPAD. Retrieved November 1, 2022, from https://www.nepad.org/blog/leveraging-smart-technologies-tackle-gender-based-violence-africa

Jiaxin, Y., Xin-cheng, W., Yujie, W., Yong, X., Kunhong, L., & Hongming, S. (2022). Temporal Modeling Matters: A Novel Temporal Emotional Modeling Approach for Speech Emotion Recognition. *Temporal Modeling Matters: A Novel Temporal Emotional Modeling Approach for Speech Emotion Recognition*. doi:2211.08233

Landry, D. T., He, Q., Yan, H., & Li, Y. (2020). ASVP-ESD: A dataset and its benchmark for emotion recognition using both speech and non-speech utterances. ASVP-ESD: A Dataset and Its Benchmark for Emotion Recognition Using Both Speech and Non-speech Utterances.

Livingstone SR, Russo FA (2018) The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. PLoS ONE 13(5): e0196391. https://doi.org/10.1371/journal.pone.0196391.

Overview of Big Bang Model - GeeksforGeeks. (2021). Retrieved 21 March 2023, from https://www.geeksforgeeks.org/overview-of-big-bang-model/

Partila, P., Tovarek, J., Rozhon, J., & Jalowiczor, J. (2019). Human stress detection from the speech in danger situation. Mobile Multimedia/Image Processing, Security, and Applications 2019. doi:10.1117/12.2521405

Shaqra, F. A., Duwairi, R., & Al-Ayyoub, M. (2019). Recognizing emotion from speech based on age and gender using hierarchical models. Procedia Computer Science, 151, 37-44. doi:10.1016/j.procs.2019.04.009

Wani, T. M., Gunawan, T. S., Qadri, S. A., Kartiwi, M., & Ambikairajah, E. (2021). A comprehensive review of Speech Emotion Recognition Systems. IEEE Access, 9, 47795-47814. doi:10.1109/access.2021.3068045