



DEPRESSION DETECTION





OVERVIEW

01 Problem Statement

Goal, introduction and dataset

02 Literature Review

Research work already done in relevant field

03 Methodology

Approach used

04 Results & Discussions

Interpretation of the work



01

Problem statement





Problem statement

- ❑ Depression is a serious mental health problem that affects people worldwide. Its complexity and diverse symptom presentation make it challenging for doctors and researchers to understand, diagnose, and treat effectively.
- ❑ The goal is to develop an effective classification system capable of distinguishing between depressed and non-depressed speech based on Low Frequency spectrograms of speech signals. It can be performed by converting audio streams into spectrograms, which provide visual representations of spectrums of frequencies as they vary over time.
- ❑ Comparison and analysis of LF spectrograms of both depressed and non-depressed classes. Use of different machine learning classifiers for performing the classification task. The research utilizes the EATD-corpus dataset which is specifically designed for studying depression in speech.



02

Literature Review

Research work already done in
relevant field





Literature review

- **Rhythm Formant Analysis for Automatic Depression Classification based of Amplitude Modulation and Frequency Modulation :** The classification system is built using a Decision Tree (DT) classifier and its results are compared with logistic regression and random forest. The model's performance is evaluated using the accuracy, F1 scores for each class and their macro and weighted averages.
- **Quantifying and Correlating Rhythm Formants in Speech :** Introduction of a new theory of speech rhythm zones or rhythm formants. Study of three domains Amplitude Modulated Signal (AM) , the amplitude envelope modulation (AEM) and frequency modulation (FM) of the signals.
- **The rhythms of rhythm :** The study explores speech rhythms and their relation to neural patterns using modulation-theoretic methods. The study views speech rhythms as regular oscillations below 10Hz. It also introduces Rhythm Formant Theory (RFT) and Rhythm Formant Analysis (RFA) as methodologies for rhythm analysis.



03

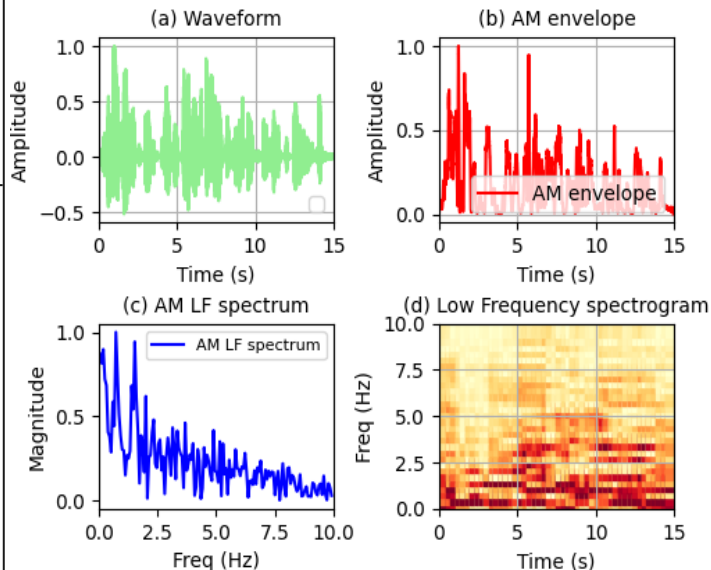
Methodology





Low Frequency Spectrograms

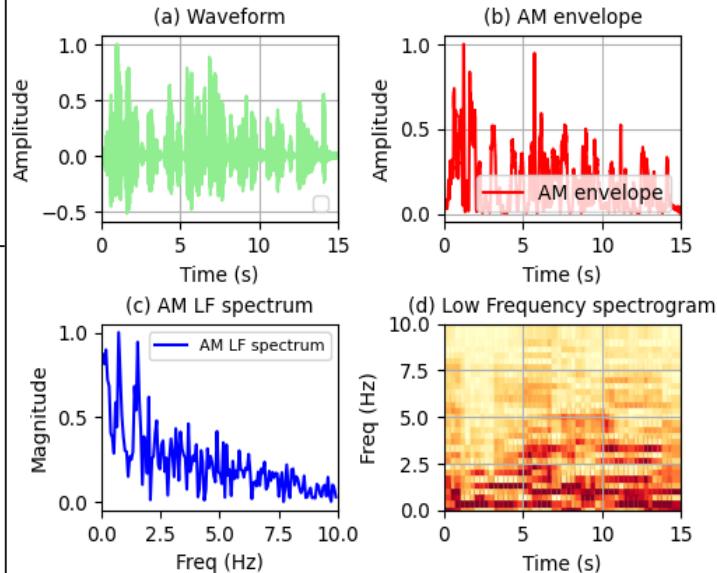
Low-frequency spectrograms typically refer to the visualization of the spectral content of a signal in the lower frequency range.



- A) Waveform – Amplitude vs. Time : In the first plot, we have a waveform after applying Normalization between -1 to 1.
- B) Amplitude Modulation Envelope – Amplitude vs. Time : This involves several steps to transform the original signal into a new representation.
1. Hilbert Transform : Hilbert transform to obtain the analytic signal. It gives amplitude of signal.
 2. Median Filter : It helps to remove any potential noise or fluctuations in the signal, resulting in a smoother representation.



Low Frequency Spectrograms



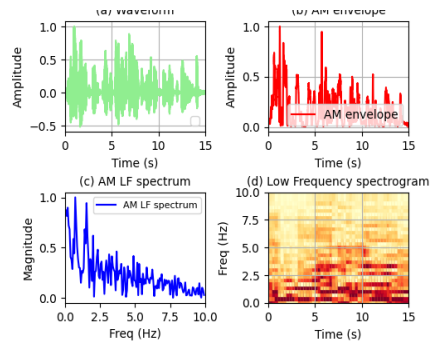
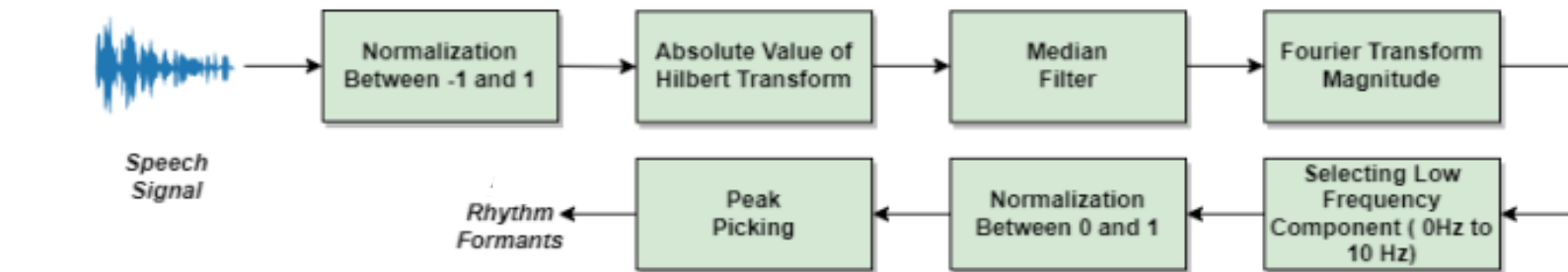
3. Fast Fourier Transform : It transforms a time-domain signal into its frequency representation, revealing the magnitudes of different frequency components.
4. Extract a low-frequency segment of the spectrum based on specified frequency limits. (0 Hz to 10 Hz).
5. Normalization between 0 to 1.

C) Amplitude Modulation Low-Frequency Spectrum - Amplitude vs. Frequency

D) Low Frequency Spectrogram - Frequency vs. Time: The color intensity at each point in the graph indicates the amplitude of the signal at that specific time and frequency. Brighter areas correspond to higher amplitudes and darker areas represent lower amplitudes.



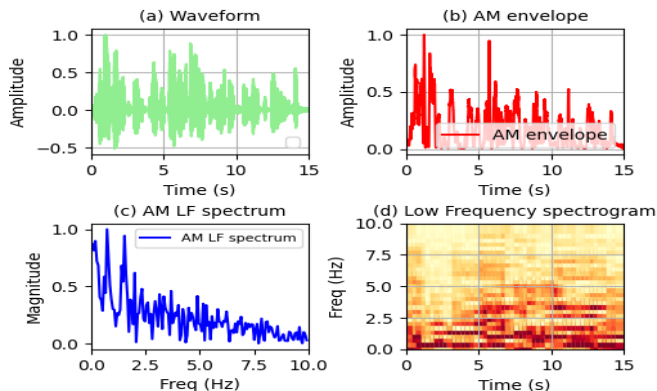
Low Frequency Spectrograms



Block diagram showing signal processing steps for extraction of rhythm formants



Depressed vs Non-Depressed

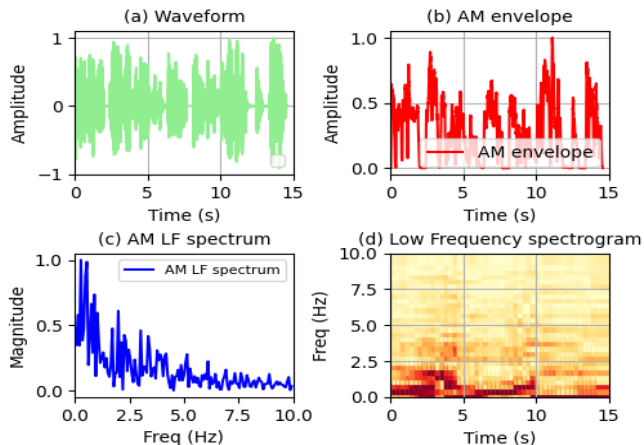


Observation





Frequent Zero Crossing and High Amplitude Variations : More in depressed person, it could indicate hesitations, pause in person's speech or lack of fluency in speech.

Brighter areas : In the low frequency spectrogram of a depressed person, brighter areas appears less compared to a non-depressed person. These brighter areas corresponds to moments of higher amplitude or more expressive speech which depressed person could lack due to overall emotional state.

Darker Areas : In a non-depressed person's spectrogram, darker areas are less. This indicate that moments of lower amplitude or emotional expression are less frequent.





Methods		
	Preprocessing	Preprocessing the data and optimising it for our model. Models implemented are resnet50 and mlp.
	Model Architecture	Creating different layers of neural network and setting its initial parameters .
	Training	Training and fine tuning parameters .
	Testing	Testing on data and deriving conclusions.

Dataset EATD-CORPUS : Test_ND : 201 files Test_D : 33 files
 Train_ND: 192 files Train_D: 57 files



04

Results and Discussions





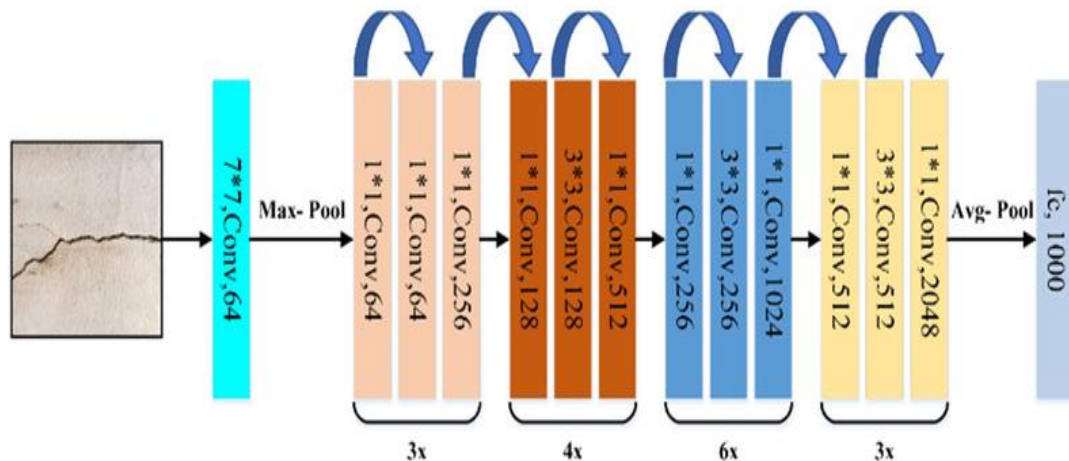
Resnet 50

ResNet50 is a specific type of convolutional neural network (CNN). It is a 50-layer convolutional neural network (48 convolutional layers, one MaxPool layer, and one average pool layer).

The 50-layer ResNet uses 1×1 convolutions, which reduces the number of parameters and matrix multiplications.

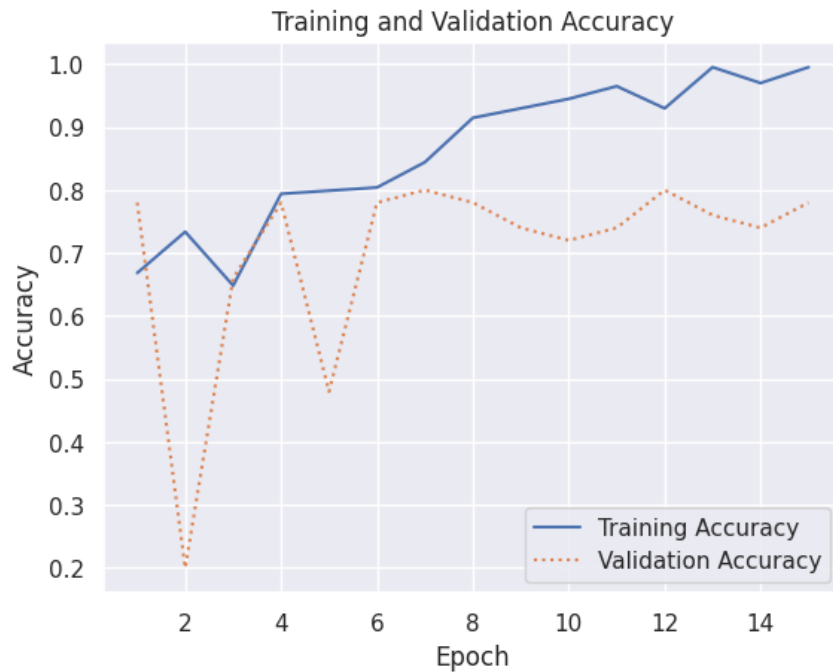
This enables much faster training of each layer. It uses a stack of three layers rather than two layers.

Skip connections, allow for the preservation of information from earlier layers by adding the output of an earlier layer to the output of a later layer.





Resnet 50

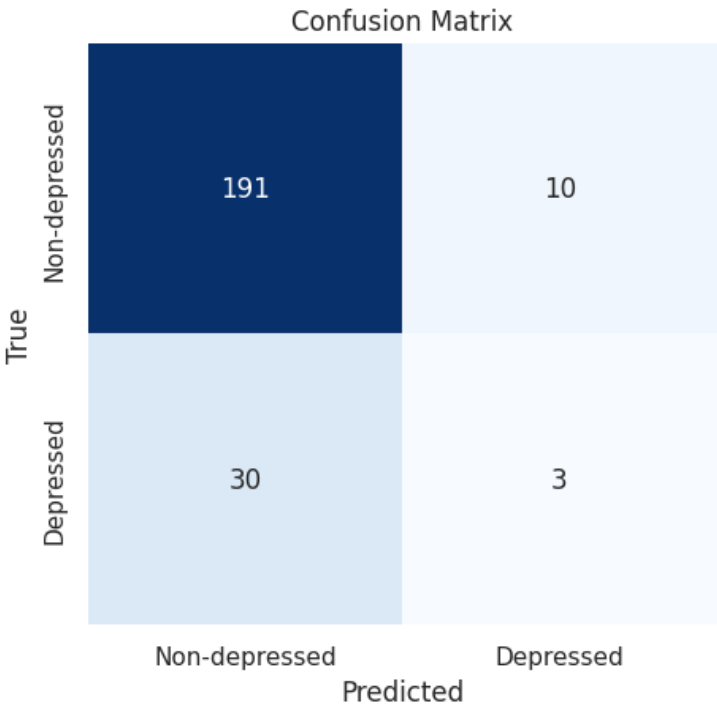




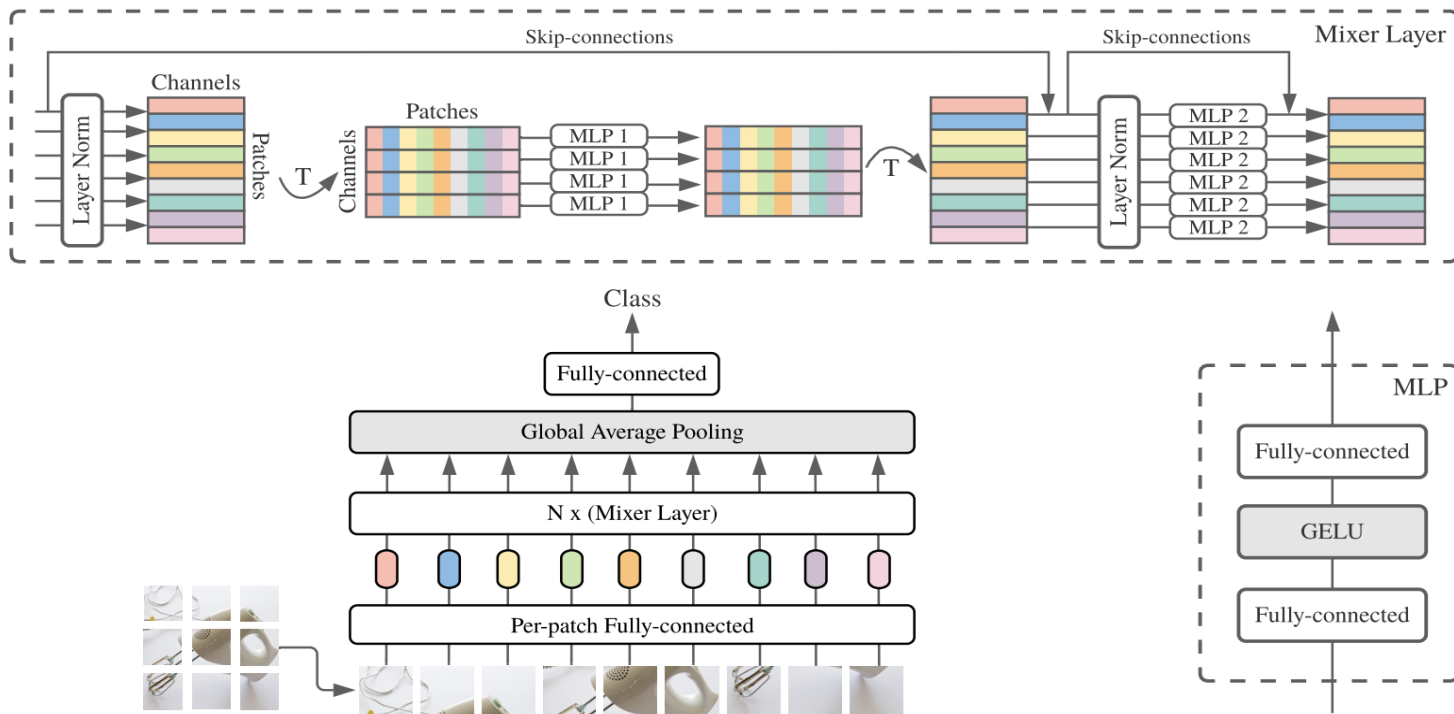
Resnet 50

Test Accuracy: 0.8291

	precision	recall	f1-score	support
non_depressed	0.86	0.95	0.91	201
depressed	0.23	0.09	0.13	33
accuracy			0.83	234
macro avg	0.55	0.52	0.52	234
weighted avg	0.77	0.83	0.80	234



Multi-layer Perceptrons Mixer (MLP-Mixer)





MLP- Mixer

Actual test loader Accuracy: 0.7307692307

Classification Report:

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

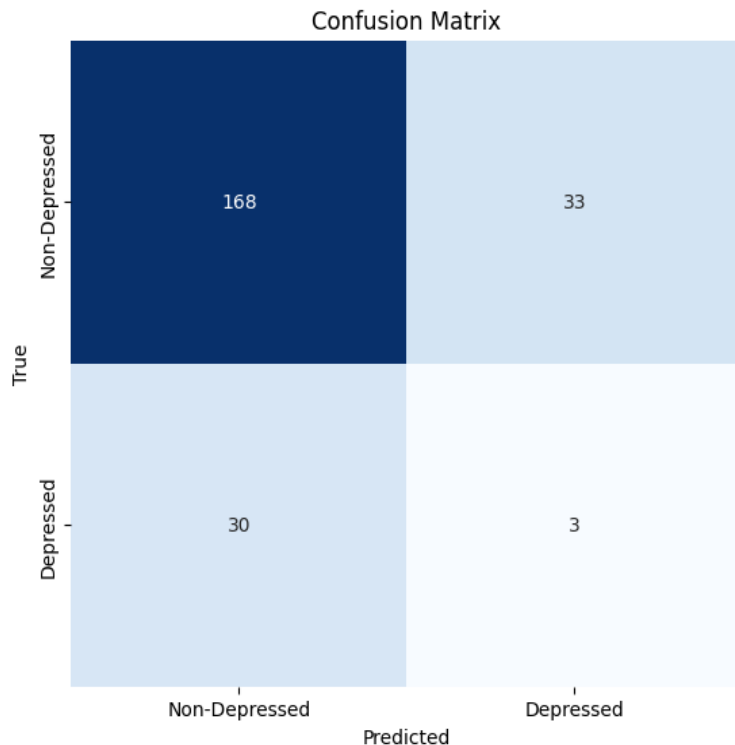
0	0.85	0.84	0.84	201
---	------	------	------	-----

1	0.08	0.09	0.09	33
---	------	------	------	----

accuracy			0.73	234
----------	--	--	------	-----

macro avg	0.47	0.46	0.46	234
-----------	------	------	------	-----

weighted avg	0.74	0.73	0.74	234
--------------	------	------	------	-----





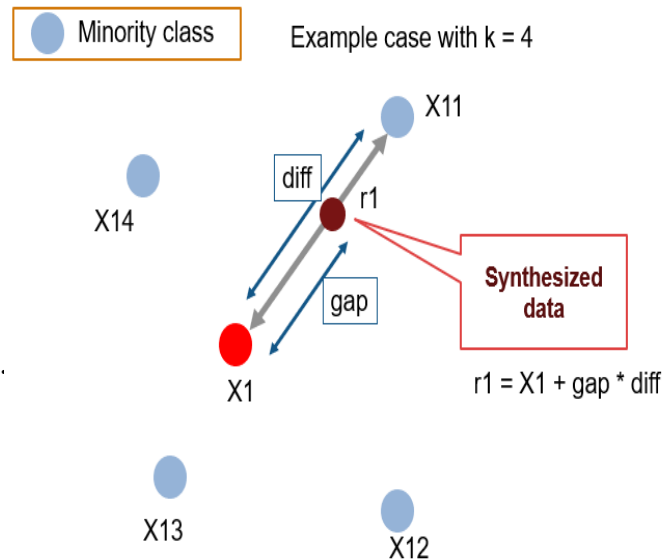
SMOTE Synthetic Minority Over-sampling Technique

SMOTE is an oversampling technique where the synthetic samples are generated for the minority class.

The goal is to balance the binary class distribution, to achieve a 1:1 ratio between the minority and majority classes.

At first the total no. of oversampling observations, N is set up. Then the iteration starts by first selecting a positive class instance at random. Next, the KNN's for that instance is obtained. At last, N of these K instances is chosen to interpolate new synthetic instances.

To do that, using any distance metric the difference in distance between the feature vector and its neighbors is calculated. Now, this difference is multiplied by any random value in $(0,1]$ and is added to the previous feature vector.



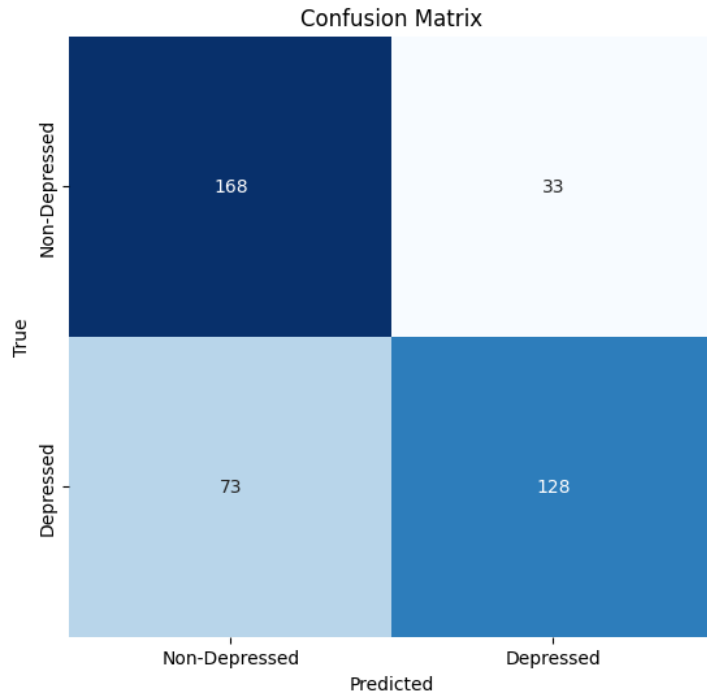


MLP After Oversampling

Accuracy: 0.736318407960199

Classification Report:

	precision	recall	f1-score	support
0	0.70	0.84	0.76	201
1	0.80	0.64	0.71	201
accuracy			0.74	402
macro avg	0.75	0.74	0.73	402
weighted avg	0.75	0.74	0.73	402





Conclusion

- **Extended the RFA work as LF spectrogram based deep learning methods for automation depression classification.**
- **Utilizes EATD - corpus for depression analysis using multiple pretrained models/ deep learning models.**
- **Effectively build classification systems for depression detection using speech.**
- **A collection of more diverse dataset needed to generalize the model**



Future Work

- **Integrate Amplitude Modulation (AM) and Frequency Modulation (FM) rhythm formants along with Low Frequency Spectrograms as additional features to capture variations in the audio signals.**
- **Using fusion of multiple deep learning models (score level) based on LF spectrogram images.**
- **Training on bigger datasets in other languages as well.**
- **Further optimization and fine tuning of model.**



Thanks !