

# Assignment-2 Q2 Report

Aryan Kumar M23CSA510

March 30, 2025

## GitHub Repository:

<https://github.com/Aryank47/Speech-Understanding-Assignment2>

## Abstract

This report presents a comprehensive study of audio signal analysis using Mel-Frequency Cepstral Coefficients (MFCCs) for 10 Indian languages. The study is divided into two major tasks: visual and statistical comparative analysis of MFCC spectrograms (Task A), and the development of a language classifier using MFCC features (Task B). The visual inspection of MFCCs helped uncover similarities and differences in acoustic patterns between languages, while the statistical analysis quantified these variations. A simple neural network classifier achieved a high accuracy of 87.85% on the test set, demonstrating the effectiveness of MFCCs in language identification tasks.

## Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>                                     | <b>2</b>  |
| <b>2</b> | <b>Task A: MFCC Extraction and Comparative Analysis</b> | <b>2</b>  |
| 2.1      | MFCC Spectrogram Visualization . . . . .                | 2         |
| 2.2      | Statistical MFCC Analysis . . . . .                     | 8         |
| <b>3</b> | <b>Task B: Language Classification using MFCCs</b>      | <b>10</b> |
| 3.1      | Preprocessing and Feature Engineering . . . . .         | 10        |
| 3.2      | Model Architecture . . . . .                            | 10        |
| 3.3      | Training Dynamics . . . . .                             | 10        |
| 3.4      | Confusion Matrix and Accuracy . . . . .                 | 11        |
| <b>4</b> | <b>Discussion</b>                                       | <b>12</b> |
| 4.1      | Acoustic Reflections in MFCCs . . . . .                 | 12        |
| 4.2      | Challenges with MFCC-based Classification . . . . .     | 12        |
| 4.3      | Recommendations for Improvement . . . . .               | 12        |
| <b>5</b> | <b>References</b>                                       | <b>13</b> |

# 1 Introduction

Indian languages exhibit a rich diversity in phonetic structure and acoustic features. Mel-Frequency Cepstral Coefficients (MFCCs), which model human auditory perception, offer a robust approach to capture and analyze such characteristics. This report explores MFCC-based representation for visual, statistical, and classification tasks on audio samples from 10 Indian languages.

## 2 Task A: MFCC Extraction and Comparative Analysis

### 2.1 MFCC Spectrogram Visualization

MFCCs were extracted for multiple audio samples from four selected languages: Bengali, Gujarati, Marathi, and Urdu. Two samples from each language were visualized to observe temporal variation and spectral energy distribution.

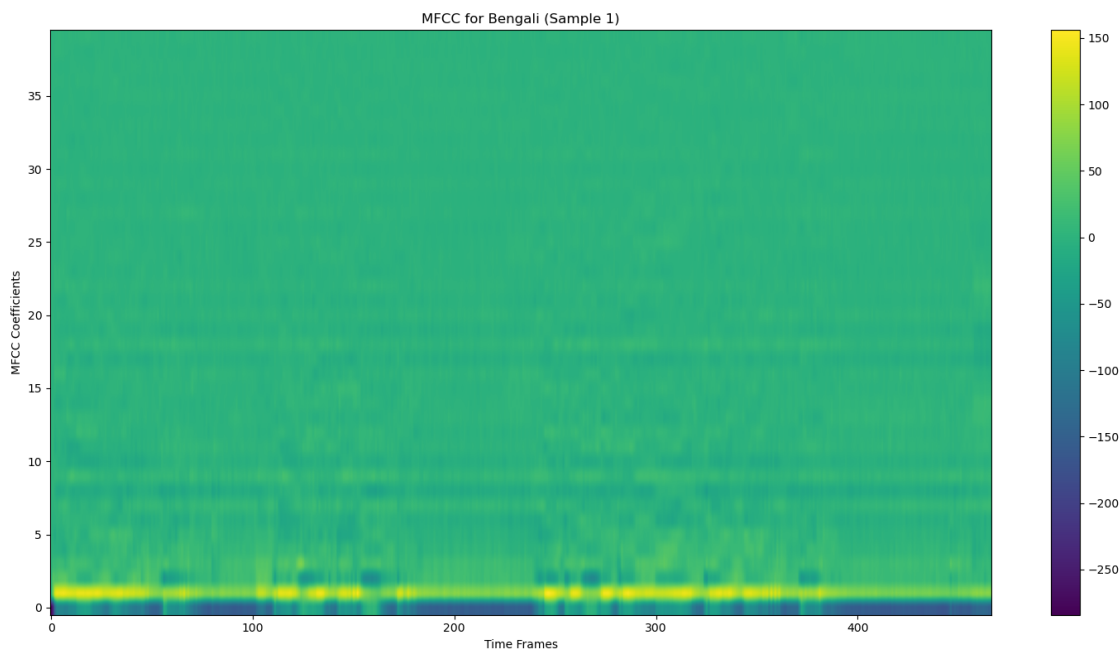


Figure 1: MFCC Spectrogram for Bengali (Sample 1)

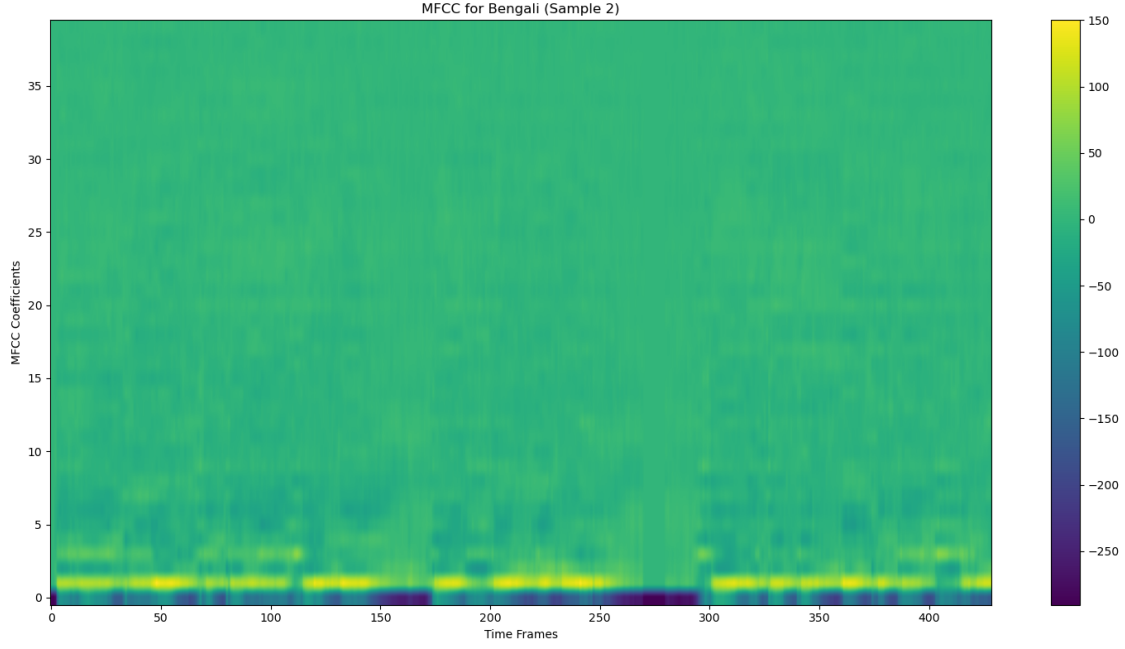


Figure 2: MFCC Spectrogram for Bengali (Sample 2)

**Observation:** Bengali shows sharp low-frequency energy peaks in early MFCC coefficients with moderate variation over time.

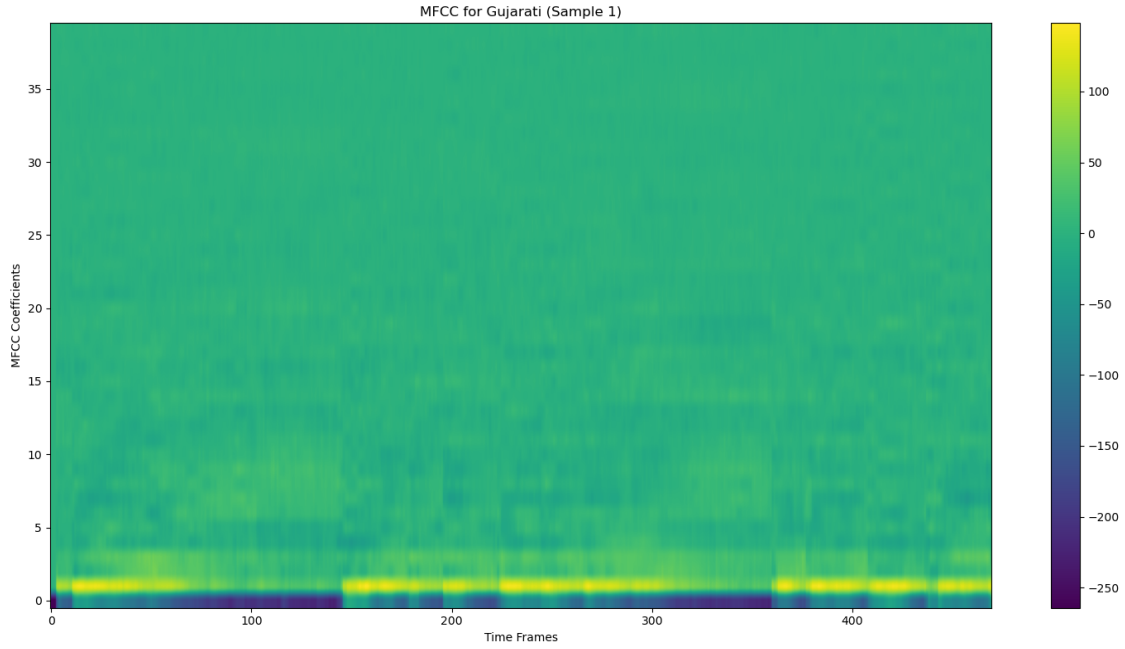


Figure 3: MFCC Spectrogram for Gujarati (Sample 1)

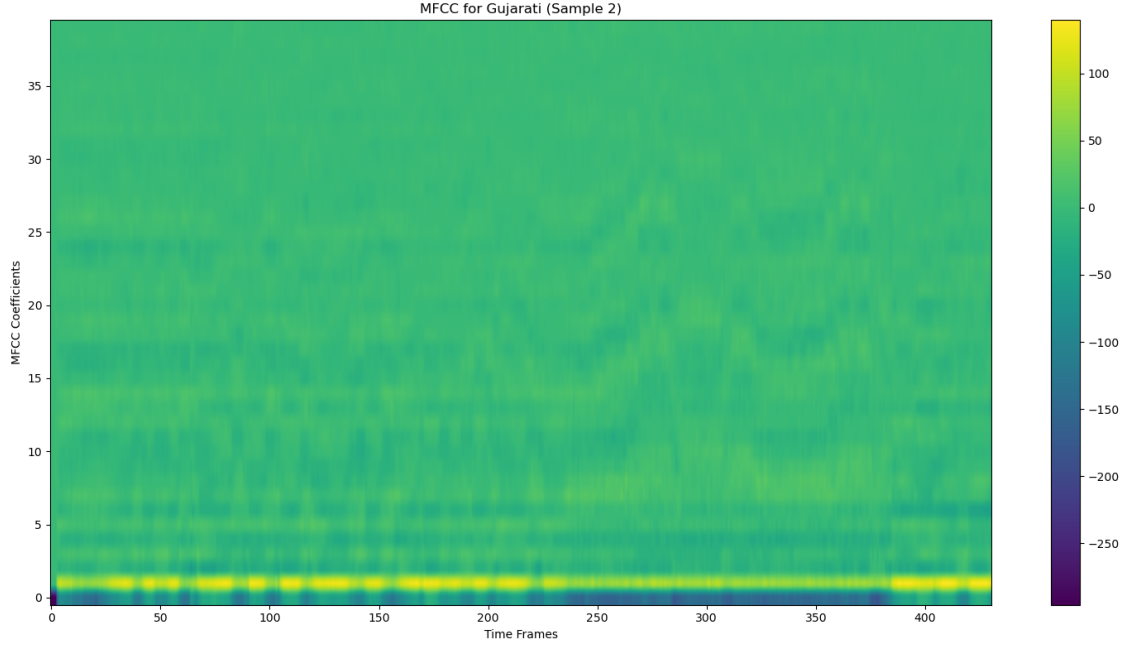


Figure 4: MFCC Spectrogram for Gujarati (Sample 2)

**Observation:** Gujarati shows similar energy in low and mid bands with visible phoneme-rich transitions.

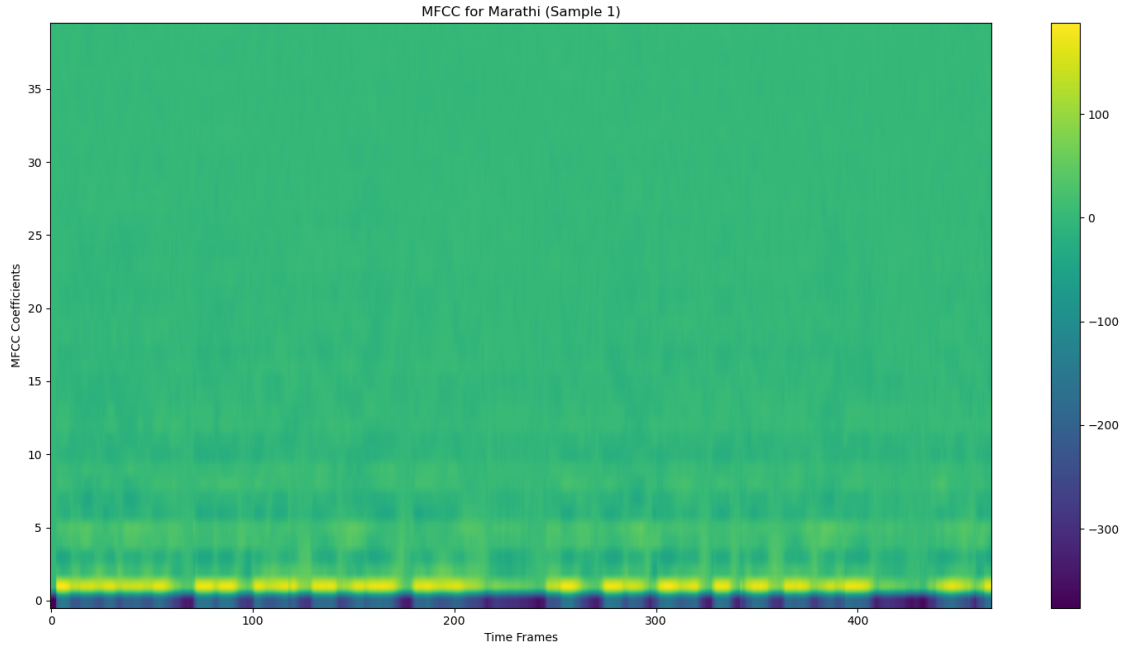


Figure 5: MFCC Spectrogram for Marathi (Sample 1)

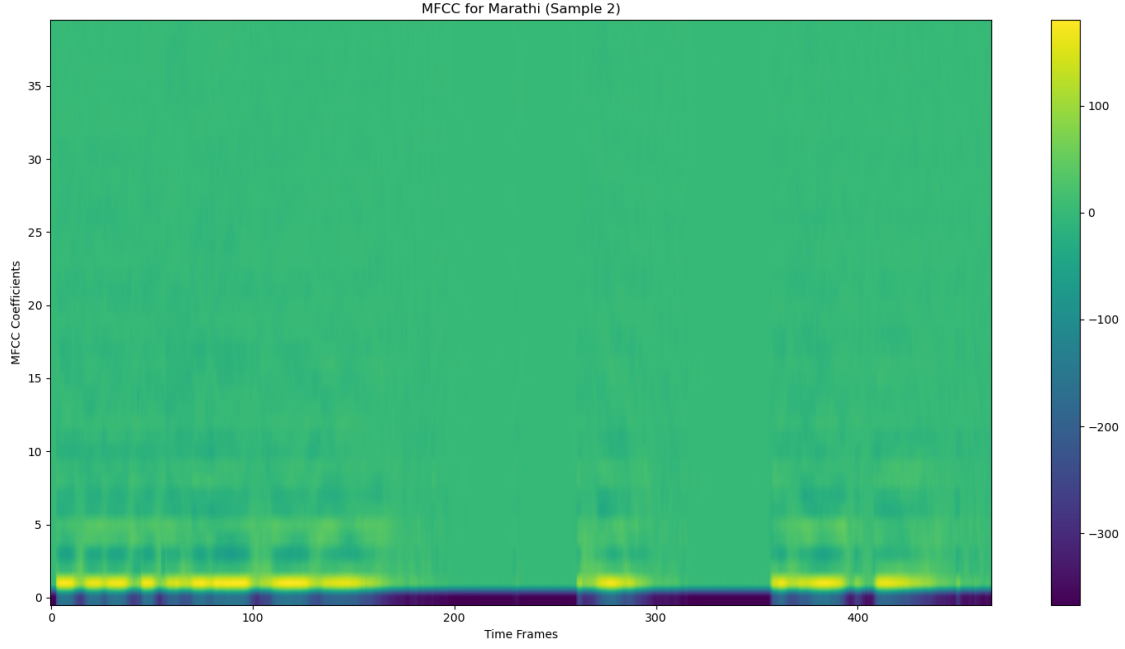


Figure 6: MFCC Spectrogram for Marathi (Sample 2)

**Observation:** Marathi maintains consistent spectral energy with structured vowel distribution across time.

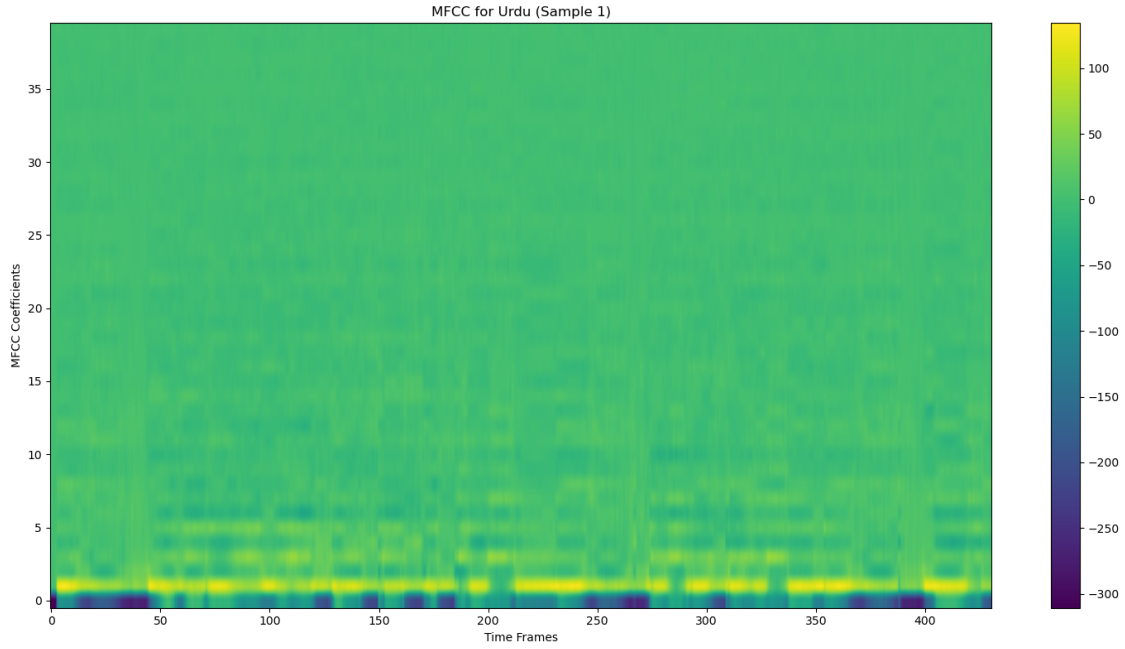


Figure 7: MFCC Spectrogram for Urdu (Sample 1)

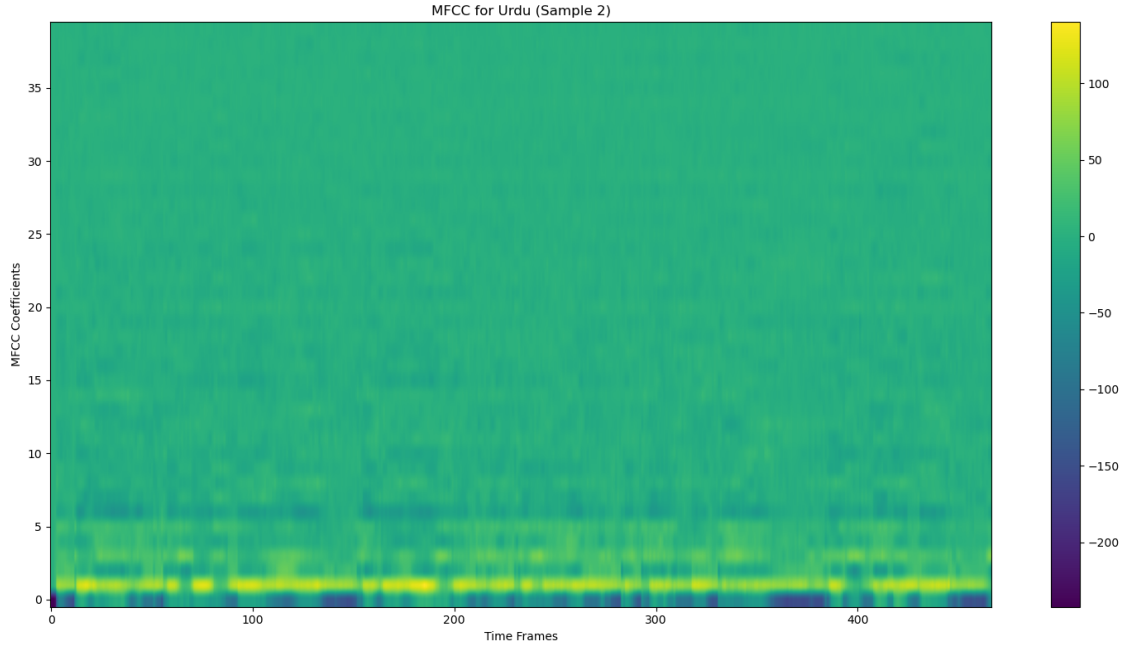


Figure 8: MFCC Spectrogram for Urdu (Sample 2)

**Observation:** Urdu shows high clarity and distinction in phoneme structure with strong energy in lower coefficients.

### MFCC Image Intensity Distributions

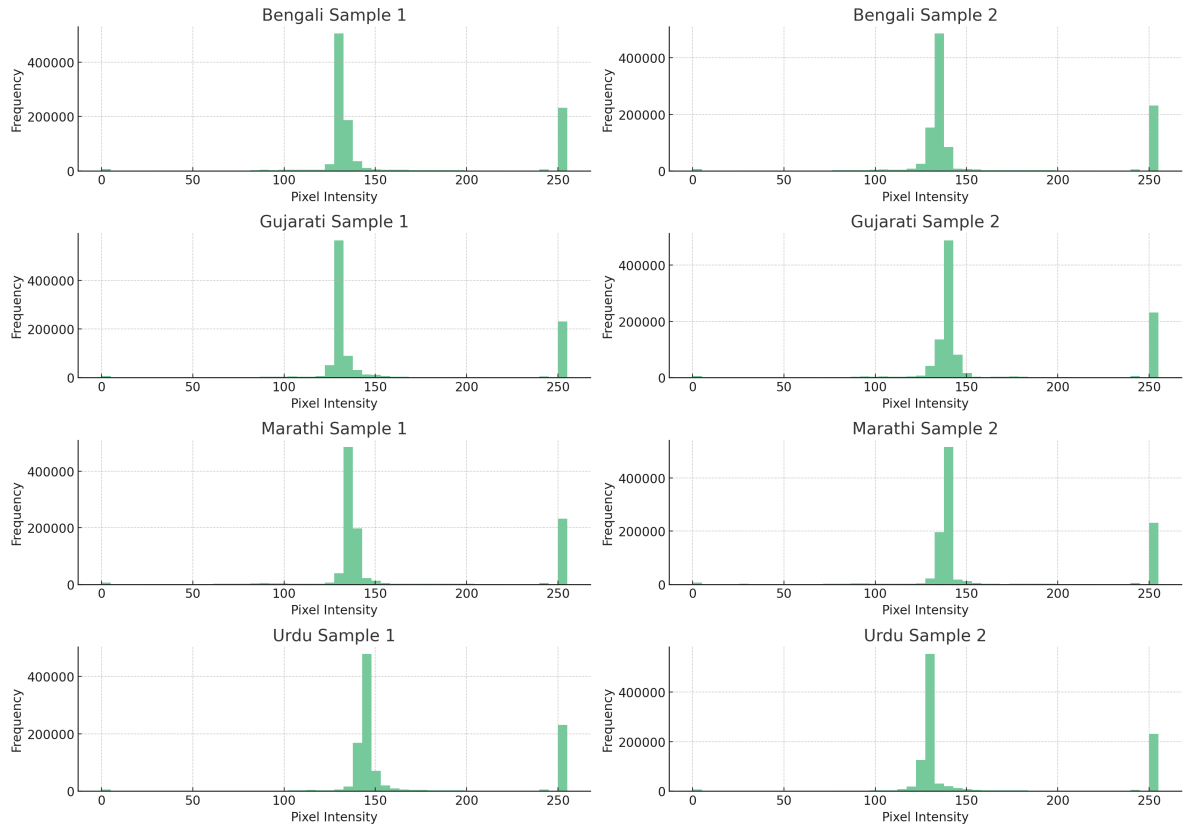


Figure 9: MFCC Image Intensity Distribution

## 2.2 Statistical MFCC Analysis

The mean and standard deviation were computed for each coefficient across all samples from the 10 languages. This statistical representation enables quantitative comparisons.

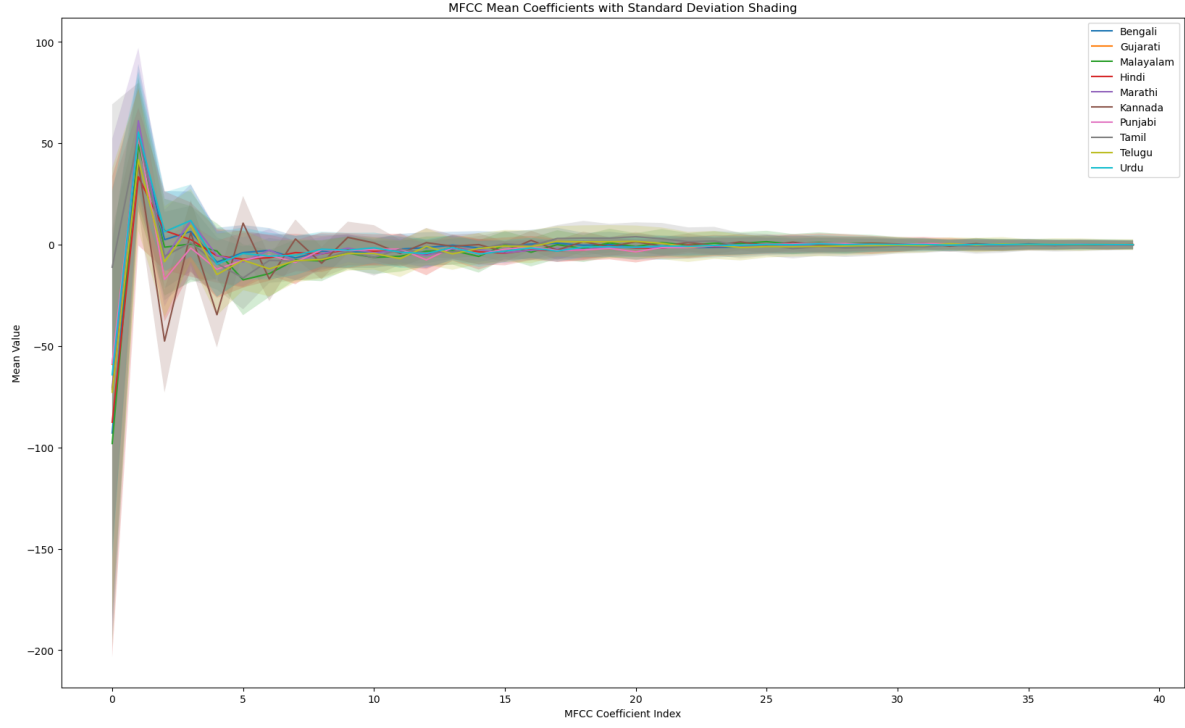


Figure 10: MFCC Mean Coefficients with Standard Deviation Shading for 10 Indian Languages



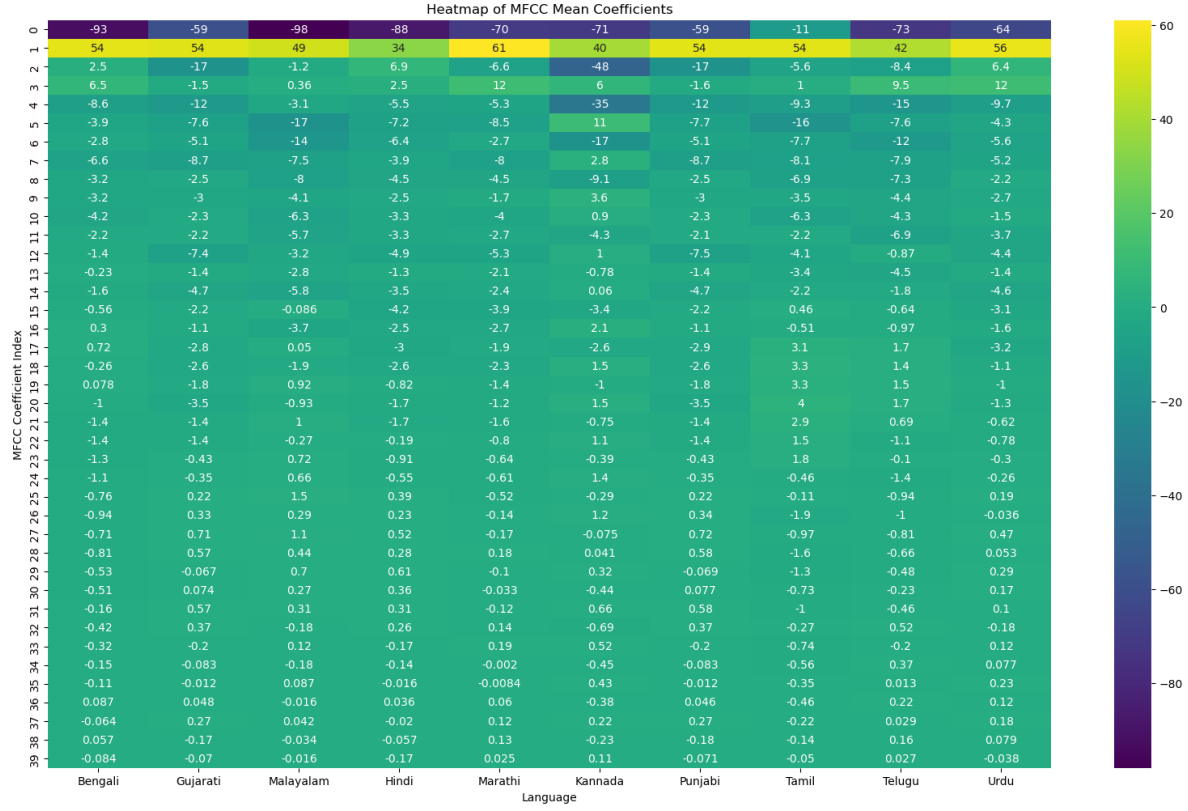


Figure 11: Mean Heatmap for 10 Indian Languages

### Insights:

- **Telugu** and **Marathi** have high variance, indicating rich phoneme diversity.
- **Tamil** and **Kannada** show mid-frequency emphasis.
- **Bengali**, **Hindi**, and **Urdu** have stable profiles across MFCCs.
- **Gujarati** and **Punjabi** exhibit nearly overlapping mean curves, later reflected in classification confusion.

## 3 Task B: Language Classification using MFCCs

### 3.1 Preprocessing and Feature Engineering

- All audio was resampled to 16kHz.
- MFCCs (40 coefficients) were extracted and time-averaged.
- Features were normalized, and a train-validation-test split was performed.

### 3.2 Model Architecture

- Input: 40 MFCC features
- 1 Hidden Layer: 225 neurons, ReLU activation
- Output: 10-class Softmax
- Loss: CrossEntropyLoss, Optimizer: Adam
- EarlyStopping used to prevent overfitting

### 3.3 Training Dynamics

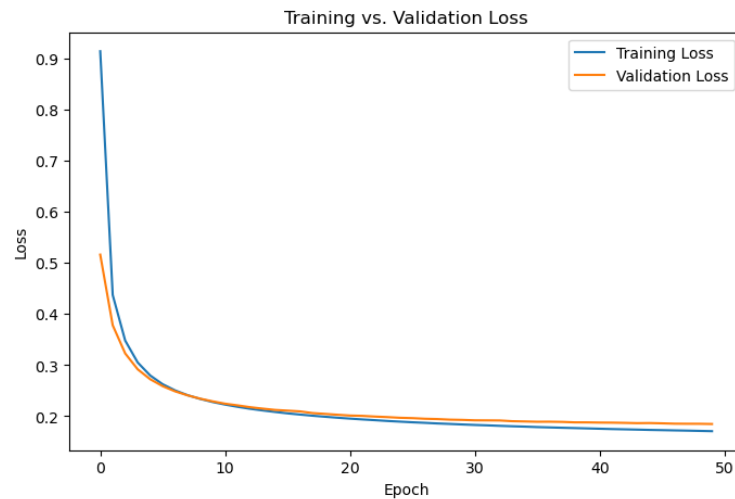


Figure 12: Training vs Validation Loss over Epochs

**Observation:** Loss steadily decreases. The gap between train and validation is minimal, confirming generalization.

### 3.4 Confusion Matrix and Accuracy

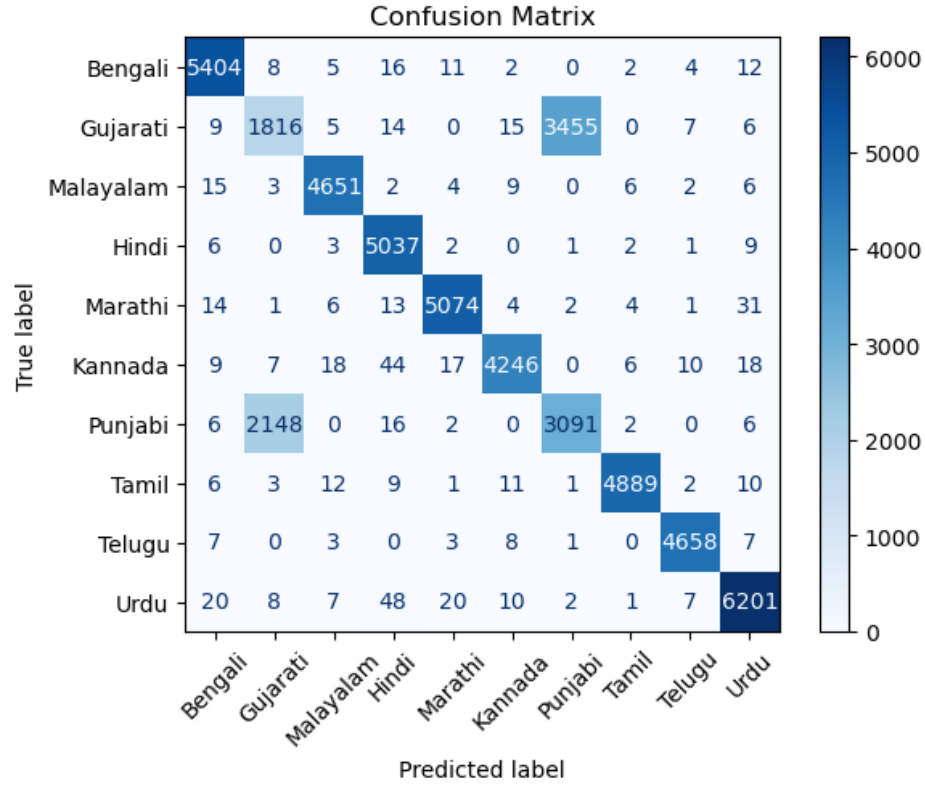


Figure 13: Confusion Matrix of Predicted vs Actual Labels

**Final Test Accuracy: 87.85%**

**Class-wise Insights:**

- Urdu, Bengali, Hindi showed highest classification accuracy.
- Gujarati-Punjabi showed major confusion due to statistically similar MFCCs.
- Tamil, Telugu, Marathi maintained good separation.

## 4 Discussion

### 4.1 Acoustic Reflections in MFCCs

MFCCs effectively capture formants and phoneme energy. Languages like Tamil show strong mid-frequency energy, while Urdu exhibits consistent low-band strength. Variance trends correlate well with phonetic richness.

### 4.2 Challenges with MFCC-based Classification

- **Speaker Variability:** Differences in pitch, accent, and speaking speed can significantly affect MFCC values, even for the same language.
- **Background Noise:** Noisy environments distort spectral features, reducing classification accuracy.
- **Regional Accents and Dialects:** Pronunciation variation within the same language can cause MFCC shifts and intra-class variability.
- **Temporal Averaging of MFCCs:** While useful for simplicity, averaging removes time-based phonetic transitions that are important for language identity.
- **Recording Quality and Devices:** Different microphones and encoding formats can impact the consistency of extracted MFCC features.

### 4.3 Recommendations for Improvement

- Use CNNs or RNNs to model MFCC sequences.
- Augment data with noise and time-warping.
- Balance datasets to avoid class bias.

## 5 References

1. <https://www.kaggle.com/code/islamic/audio-language-classification>.
2. <https://www.kaggle.com/code/kripa5661/audio-data-preprocessing>.