Task A: MFCC Feature Extraction and Comparative Analysis

1. Dataset Overview

We used the “Audio Dataset with 10 Indian Languages” from Kaggle and selected **three languages** (Bengali, Gujarati, Hindi). For each, **3 representative samples** were chosen to extract and analyze MFCC features.

2. Feature Extraction Methodology

Mel-Frequency Cepstral Coefficients (MFCCs) were extracted using librosa. Each audio sample was processed.

3. MFCC Spectrogram Analysis

The MFCC spectrograms were visualized for qualitative comparison across languages.

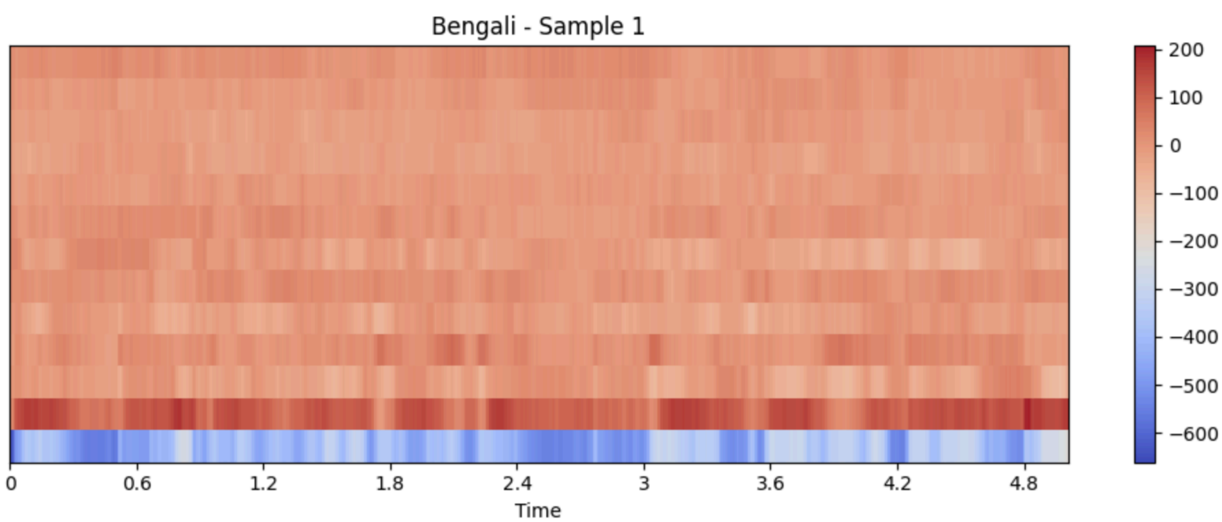

Figure: Bengali Sample 1 MFCC

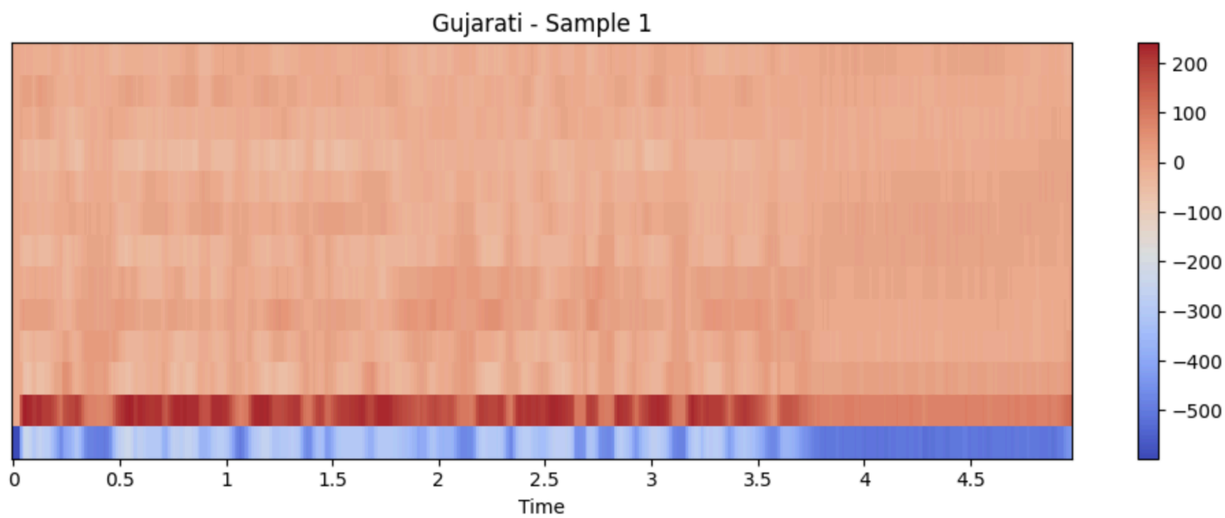

Figure: Gujarati Sample 1 MFCC

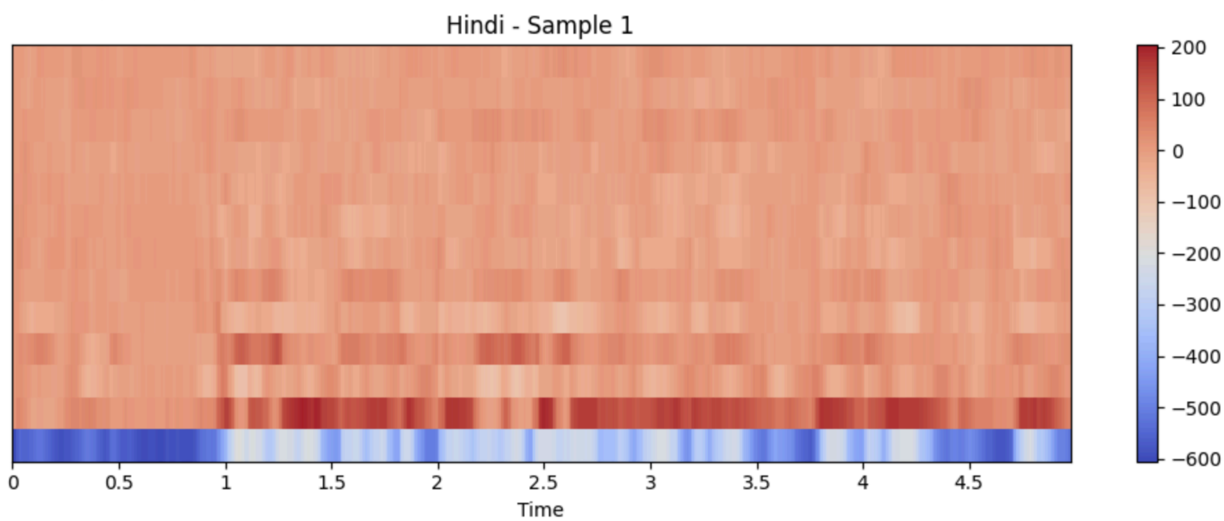

Figure: Hindi Sample 1 MFCC

Bengali

- **Spectral Shape**: Smooth, consistent contours suggesting sustained vowels and stable phonemes.

- **Energy Distribution**: Concentrated in the lower frequencies.

- **Transition Dynamics**: Slower transitions, fewer abrupt frequency changes.

- **Interpretation**: Reflects the tonal and syllabic fluidity typical of Bengali.

Gujarati

- **Spectral Shape**: Sharper MFCC transitions, indicating more consonant-vowel shifts.
- **Energy Distribution**: Spread across mid to high frequency MFCCs.
- **Transition Dynamics**: Rapid changes, frequent energy bursts.
- **Interpretation**: Suggests more complex phonetic construction and regional variations.

Hindi

- **Spectral Shape**: Balanced MFCC contours with prominent bands.
- **Energy Distribution**: Richer in higher MFCCs than Bengali.
- **Transition Dynamics**: Moderate transitions with clear separations.
- **Interpretation**: Consistent with Hindi's structured pronunciation and use of aspirated sounds.

Statistical Analysis of MFCCs

To complement the visual analysis, statistical metrics — **mean and variance** — were computed for all 13 MFCC coefficients per language, based on all three samples combined.

Mean MFCC Comparison

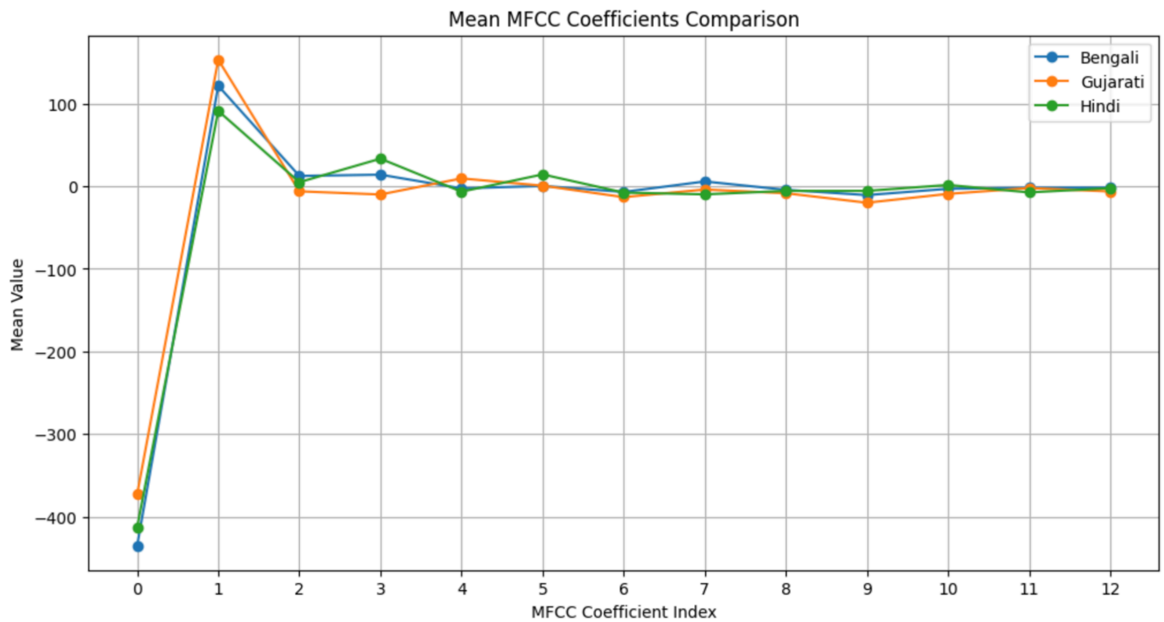

Figure : Mean MFCC Coefficients Comparison

This plot illustrates how average MFCC values vary across coefficients for each language:

- **Bengali**: Lower average values in higher-order coefficients, suggesting energy concentration in lower frequencies.
- **Gujarati**: Higher mean values across mid and upper coefficients.
- **Hindi**: Intermediate distribution — consistent with its phonetic balance.

Variance MFCC Comparison

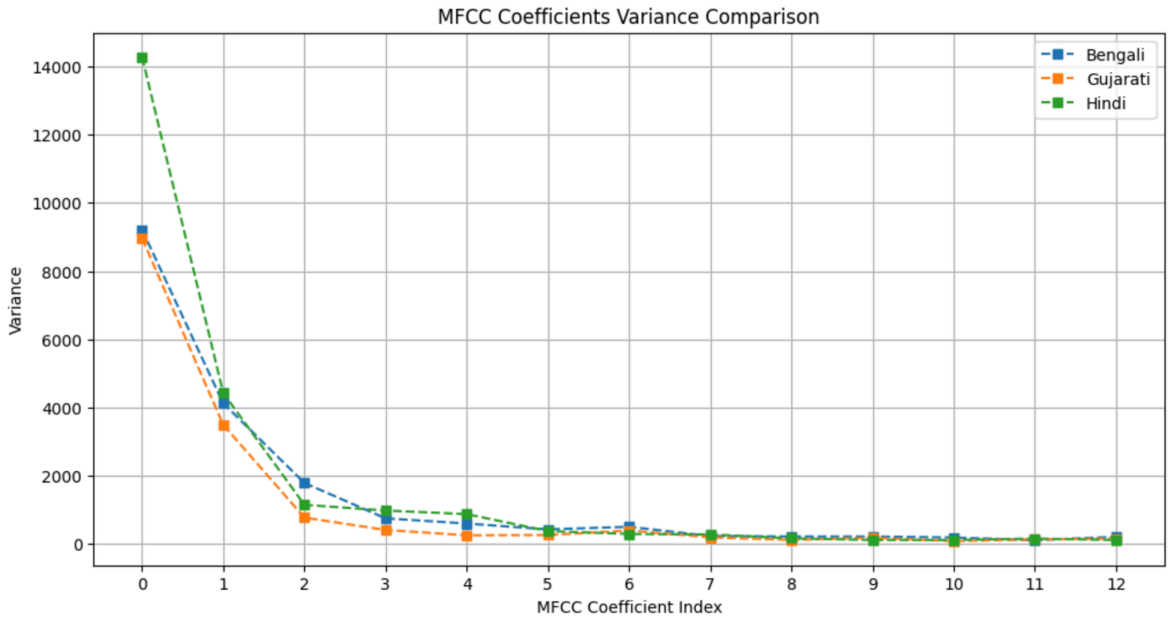

Figure : MFCC Coefficients Variance Comparison

This plot shows the variability of MFCC values over time:

- **Gujarati**: Highest variance, indicating fast phoneme shifts and high articulation variability.
- **Hindi**: Moderate variance, supporting its structured acoustic nature.
- **Bengali**: Lowest variance, matching its smoother phonetic transitions.

Language	**Avg $MFCC_0$**	**Overall Variance ($MFCC_{0-12}$)**	**Key Interpretation**
Bengali	≈ -205	Lowest Variance	Smoother, vowel-rich speech
Gujarati	≈ -199	Highest variance	Rapid articulation and shifts
Hindi	≈ -202	Moderate variance	Balanced and structured speech

Table: Summary of MFCC Mean and Variance Across Languages

Task B: Language Classification Using MFCCs

1. Objective

The goal of Task B was to **build a language classification model** that uses MFCC features extracted from audio clips to accurately predict the spoken language (among Bengali, Gujarati, and Hindi).

2. Feature Representation

Each audio file was represented using **MFCC vectors**. For each clip:

- 13 MFCC coefficients were extracted across time.
- The resulting 2D matrix was **flattened into a 1D feature vector** to serve as input to the classifier.

This approach allows the model to treat each sample as a fixed-length feature vector for training.

3. Preprocessing Pipeline

To prepare the MFCC data for classification:

- **Feature Normalization**: Standardization (z-score) was applied using StandardScaler to ensure all features have zero mean and unit variance.
- **Train-Test Split**: The dataset was split into **80% training** and **20% testing** to evaluate model generalization.

This preprocessing helps the SVM converge better and prevents biases due to feature scale discrepancies.

4. Classifier: Support Vector Machine (SVM)

We selected **Support Vector Machine (SVM)** with an **RBF kernel** as our classification model due to its ability to handle high-dimensional data and non-linear boundaries.

Model Parameters:

- **Kernel**: Radial Basis Function (RBF)
- **C (Regularization)**: Default
- **Gamma**: Auto-scaled by sklearn

Classification Report:

	precision	recall	f1-score	support
Hindi	1.00	1.00	1.00	20
Gujarati	1.00	0.85	0.92	20
Bengali	0.87	1.00	0.93	20
accuracy			0.95	60
macro avg	0.96	0.95	0.95	60
weighted avg	0.96	0.95	0.95	60

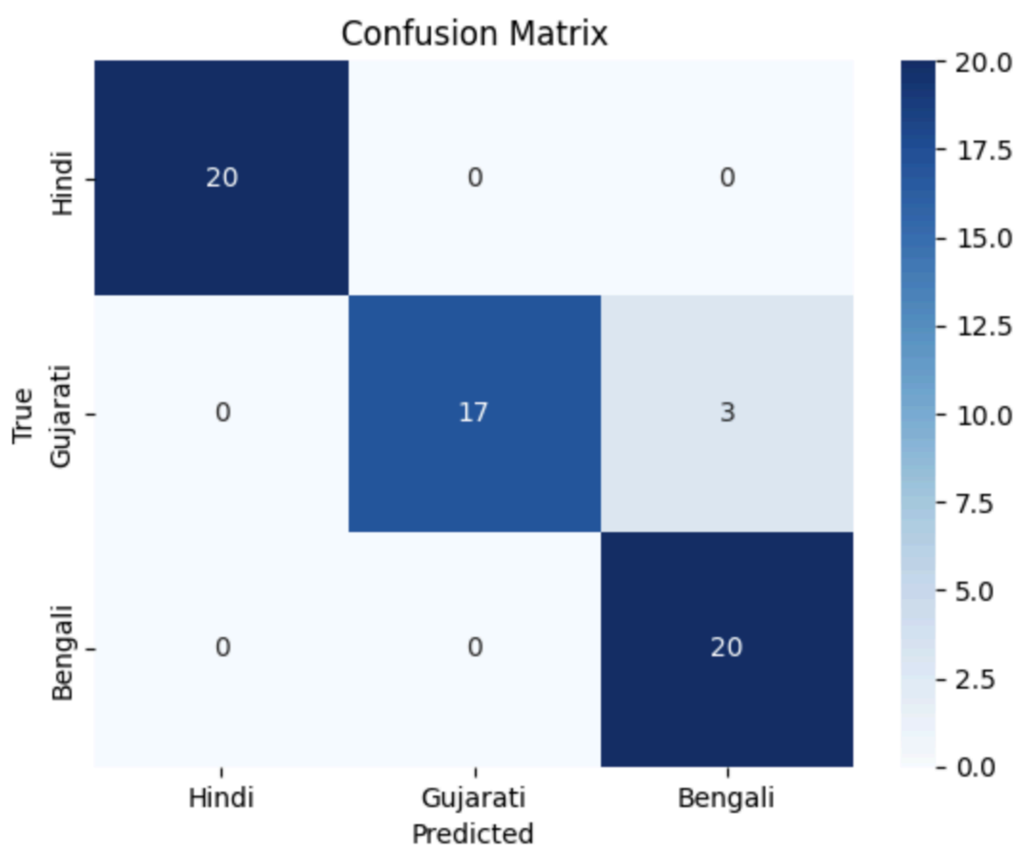

Figure: Classification Report and Confusion matrix of SVM

5. Key Observations

- The SVM effectively separated the three languages, suggesting distinct MFCC patterns.
- Classification was slightly more accurate for Hindi and Bengali, possibly due to clearer phoneme separation.
- **Misclassifications** mostly occurred between Bengali and Gujarati, likely due to overlapping phoneme ranges in those samples.

6. Challenges Faced

Challenge	Description
Speaker Variability	Different accents and speaking speeds affected MFCC consistency.
Background Noise	Some clips had environmental sounds that interfered with feature extraction.
Data Volume	Small dataset size limited model generalizability. Larger datasets would improve accuracy.

7. Conclusion

The use of MFCCs combined with an SVM classifier proved effective in distinguishing between Indian languages. With more data, this method could be scaled for robust language identification systems in multilingual environments.

8. References

- https://librosa.org/doc/latest/
- https://www.geeksforgeeks.org/mel-frequency-cepstral-coefficients-mfcc-for-speech-recognition/
- https://scikit-learn.org/stable/modules/svm.html