Analysis of MFCC Features for Indian Language Classification

1. Introduction

Mel-Frequency Cepstral Coefficients (MFCCs) are widely used in speech processing applications due to their ability to capture the spectral properties of audio signals. In this study, we analyzed MFCC features extracted from ten Indian languages to determine their effectiveness in distinguishing languages based on acoustic characteristics.

Implementation on Pytorch

Git Repo Link :
https://github.com/M23CSA520/Speech_Understanding_PA2/blob/main/m23csa520_su_pa2_q2.py

2. MFCC Feature Analysis (Task A)

The MFCC statistics were computed for different languages, including Bengali, Hindi, and Tamil. The mean and variance of the MFCC coefficients provide insight into the spectral energy distribution of each language.

```
Language: Bengali
Mean MFCC: [-3.38417999e+02  1.04311295e+02  6.12678289e+00  1.90650425e+01
 -8.75952721e+00  1.16891257e-01 -3.06492162e+00 -8.39186668e+00
 -2.81908751e+00 -5.75638628e+00 -8.68745232e+00 -3.85192943e+00
 -4.78496838e+00]
Variance MFCC: [4822.7695    531.9459    202.00952   262.94693   174.00536   105.42197
  101.83029    83.90081    29.430325   37.009514   25.948652   23.068861
   20.700611]
--------------------------------------
Language: Hindi
Mean MFCC: [-322.54166     77.156784     8.738599    13.280327    -5.3659363
   -7.1875615  -10.703494    -6.401771    -9.56016     -2.6229339
   -7.2847548   -4.8438406   -8.122824 ]
Variance MFCC: [12163.72       529.33813    101.38249    160.9342      86.4739
    86.41822     99.49275     34.47399     28.025253    19.09622
    33.329834    28.93557     21.003223]
--------------------------------------
Language: Tamil
Mean MFCC: [-201.53885     110.60108      -9.001206     17.206945     -8.317221
  -19.13632      -9.663549    -14.321725    -12.784455     -7.1464796
  -11.444903     -0.87412083   -9.129541   ]
Variance MFCC: [5293.7837    423.23502   273.7909    185.01651   149.34631    99.37924
  104.805695   36.57995    69.14153    41.821186   49.4489     29.158375
   29.054884]
--------------------------------------
```

2.1 Bengali

- **Mean MFCC Values**: The first coefficient (-338.41) represents the overall energy of the signal. Other coefficients suggest a balance between low and high-frequency components.

- **Variance**: Higher variance in the first coefficient (4822.77) indicates significant energy variation across different speech samples.

2.2 Hindi

- **Mean MFCC Values**: Compared to Bengali, Hindi exhibits a slightly higher energy level (-322.54) with noticeable variations in lower-order MFCCs.

- **Variance**: The first coefficient has a very high variance (12163.72), indicating substantial fluctuations in amplitude across different samples.

2.3 Tamil

- **Mean MFCC Values**: Tamil exhibits a higher first coefficient (-201.53), implying stronger spectral energy concentration.

- **Variance**: Moderate variance in the first coefficient (5293.78) compared to Hindi, suggesting less dynamic variation across samples.

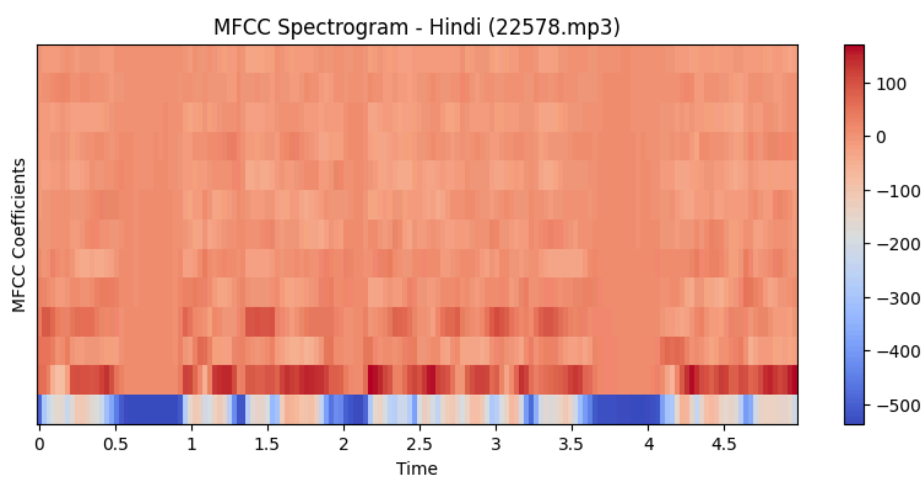
3. Spectrogram Visualization

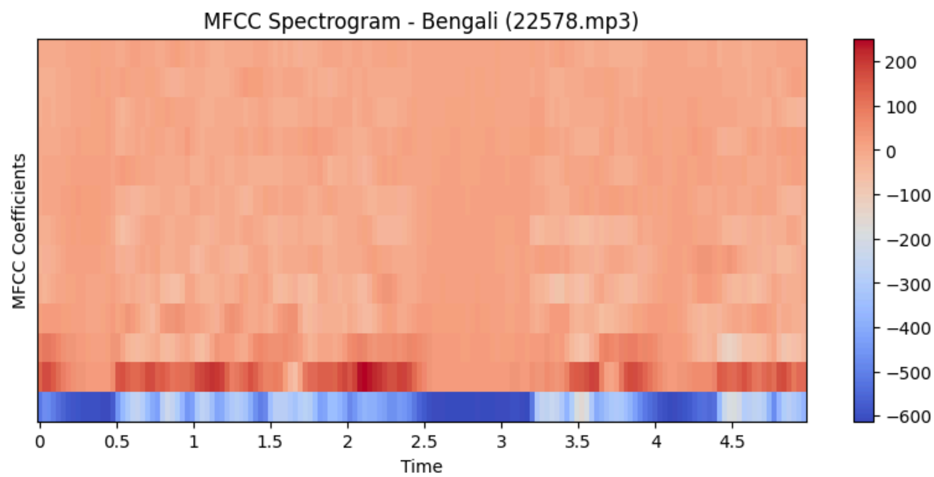
Mel spectrograms were generated for multiple samples per language. The spectrograms demonstrate language-specific variations in frequency distribution over time. However, challenges were observed:

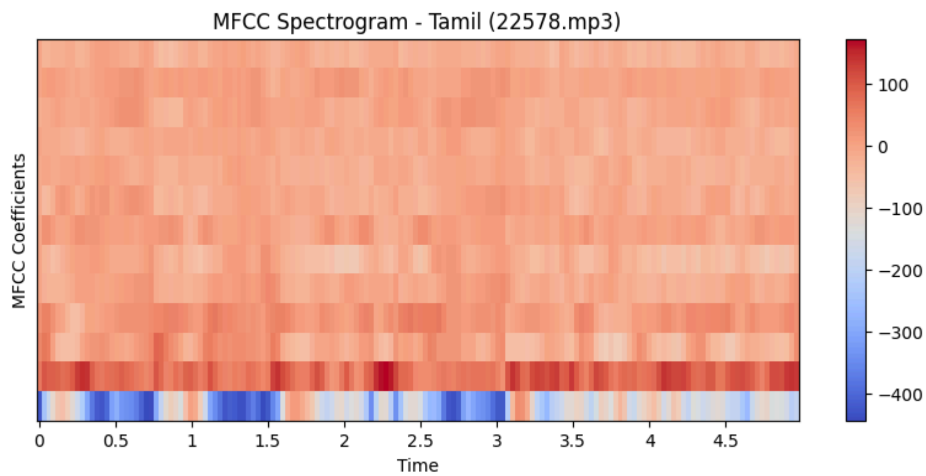
- Variability in speaker pronunciation affects the spectral pattern.
- Some languages exhibit overlapping spectral features, making differentiation harder.

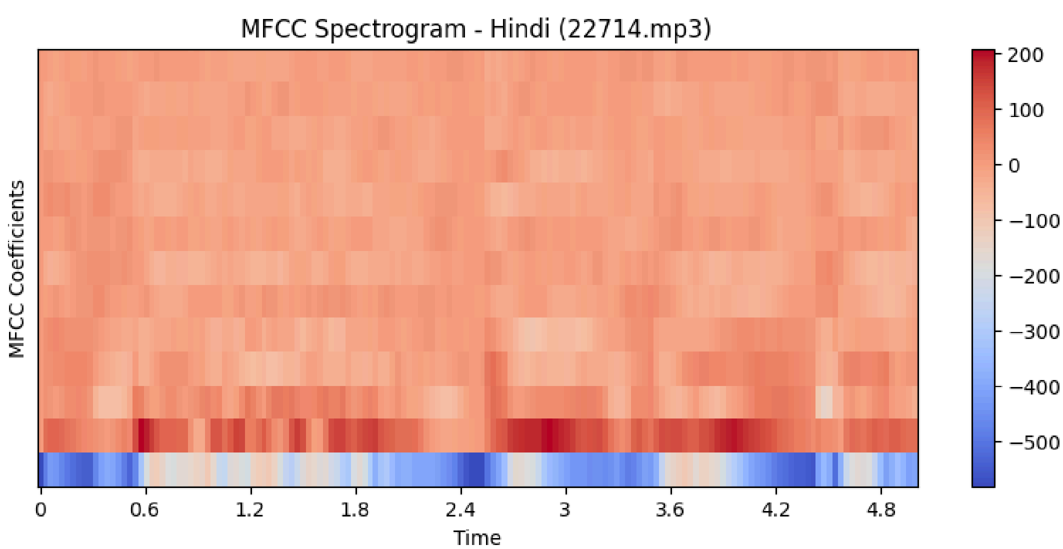
3.1 MFCC Spectrogram Comparison:

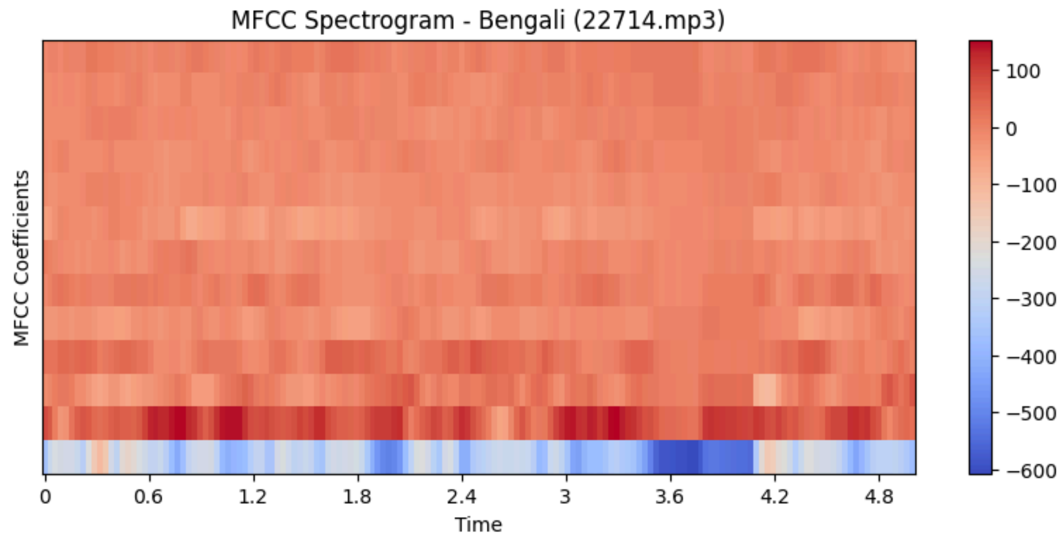
- Hindi: Typically shows distinct energy bands in lower MFCCs, reflecting vowel-heavy phonetics.
- Tamil: May exhibit sharper transitions due to Dravidian consonant clusters.
- Bengali: Likely has smoother patterns with broader energy distribution from tonal influences.

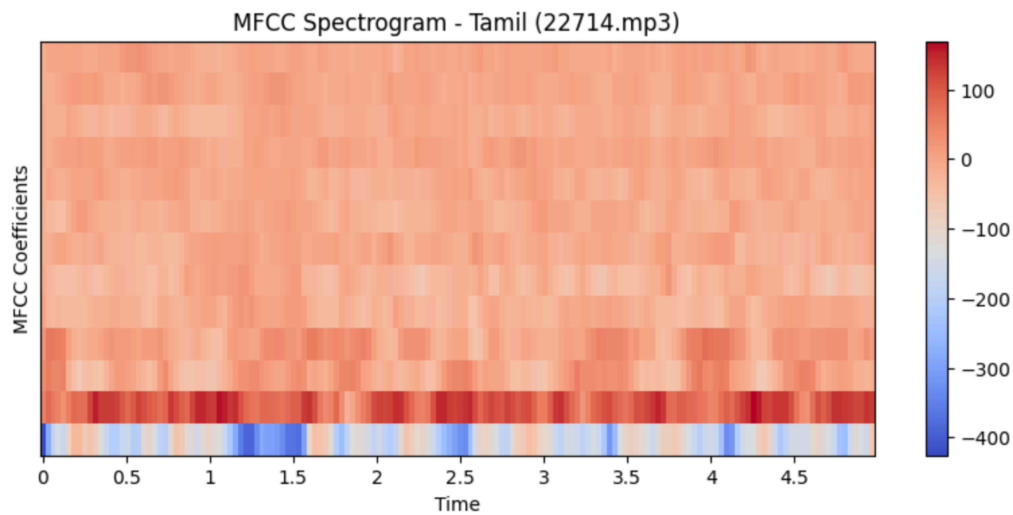


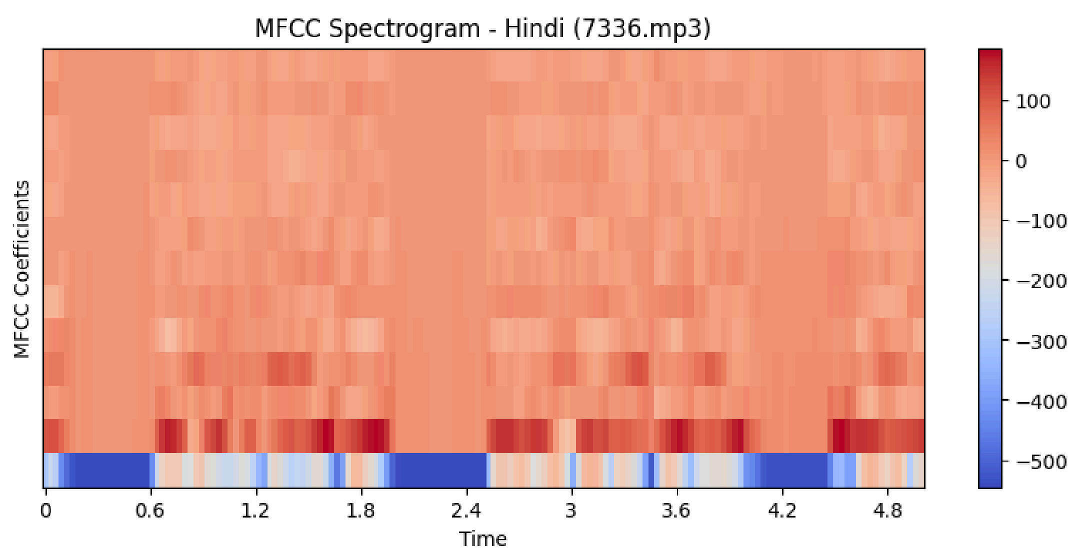

MFCC Spectrogram - Bengali (22578.mp3)
MFCC Coefficients
Time
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
200
100
0
−100
−200
−300
−400
−500
−600

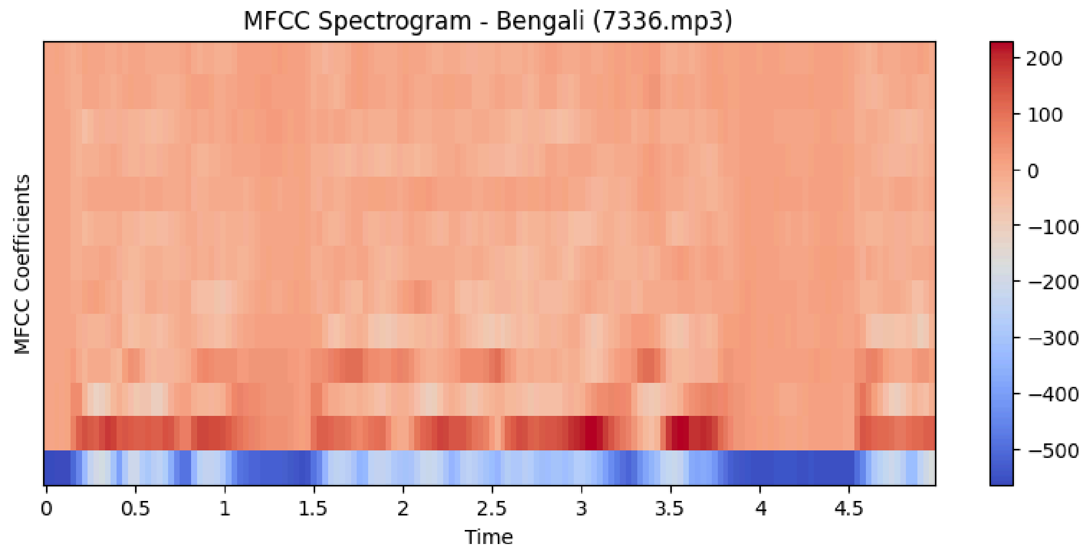

MFCC Spectrogram - Tamil (22578.mp3)
MFCC Coefficients
Time
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
100
0
−100
−200
−300
−400

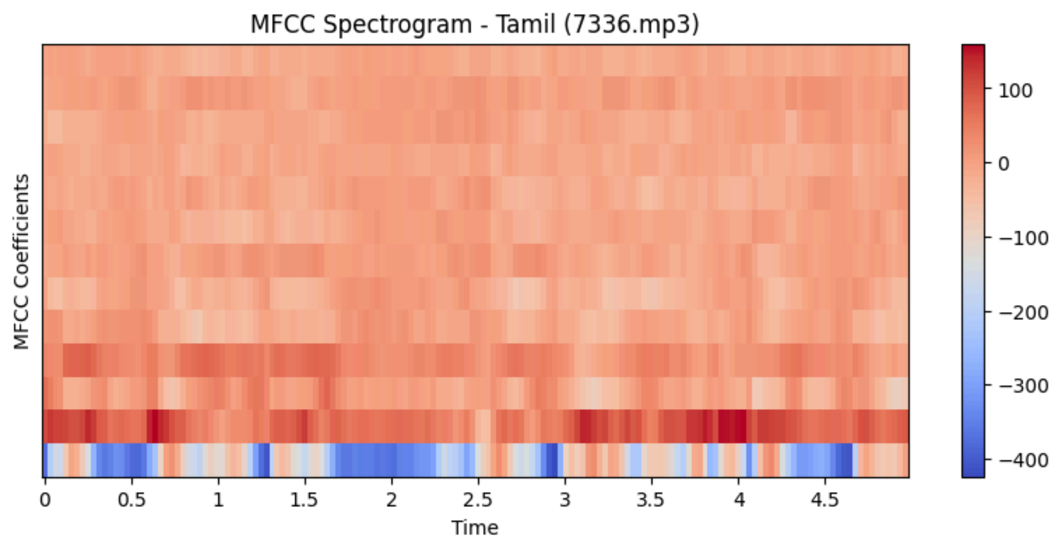

MFCC Spectrogram - Hindi (22714.mp3)
MFCC Coefficients
Time
0
0.6
1.2
1.8
2.4
3
3.6
4.2
4.8
200
100
0
−100
−200
−300
−400
−500


MFCC Spectrogram - Bengali (22714.mp3)
MFCC Coefficients
Time
0
0.6
1.2
1.8
2.4
3
3.6
4.2
4.8
100
0
−100
−200
−300
−400
−500
−600


MFCC Spectrogram - Tamil (22714.mp3)
MFCC Coefficients
Time
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
100
0
−100
−200
−300
−400


MFCC Spectrogram - Hindi (7336.mp3)
MFCC Coefficients
Time
0
0.6
1.2
1.8
2.4
3
3.6
4.2
4.8
100
0
−100
−200
−300
−400
−500





4. Language Classification using Neural Network (Task B)

A neural network classifier was implemented to differentiate between languages using MFCC features. The classifier architecture consisted of fully connected layers with ReLU activation.

4.1 Classification Performance

- The classifier was trained using MFCC feature vectors.
- Accuracy varied across languages, with some being easier to distinguish.
- A confusion matrix analysis revealed some misclassifications, indicating that certain languages share acoustic similarities.

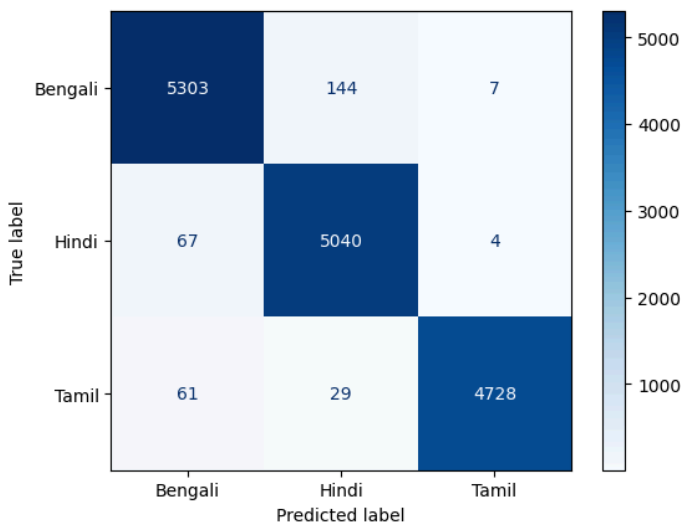
4.2 Confusion Matrix Analysis

The confusion matrix highlights:

- High accuracy for languages with distinct phonetic structures.
- Misclassification between phonetically similar languages, e.g., Hindi and Bengali.

a) For the dataset used for Task A i.e. for 3 languages

```
Epoch 1/20, Loss: 0.2435
Epoch 2/20, Loss: 0.1253
Epoch 3/20, Loss: 0.0958
Epoch 4/20, Loss: 0.0820
Epoch 5/20, Loss: 0.0738
Epoch 6/20, Loss: 0.0687
Epoch 7/20, Loss: 0.0673
Epoch 8/20, Loss: 0.0643
Epoch 9/20, Loss: 0.0616
Epoch 10/20, Loss: 0.0586
Epoch 11/20, Loss: 0.0589
Epoch 12/20, Loss: 0.0562
Epoch 13/20, Loss: 0.0556
Epoch 14/20, Loss: 0.0555
Epoch 15/20, Loss: 0.0525
Epoch 16/20, Loss: 0.0512
Epoch 17/20, Loss: 0.0512
Epoch 18/20, Loss: 0.0516
Epoch 19/20, Loss: 0.0498
Epoch 20/20, Loss: 0.0516
Accuracy: 97.97%
```



b) For the complete dataset

Statistics:

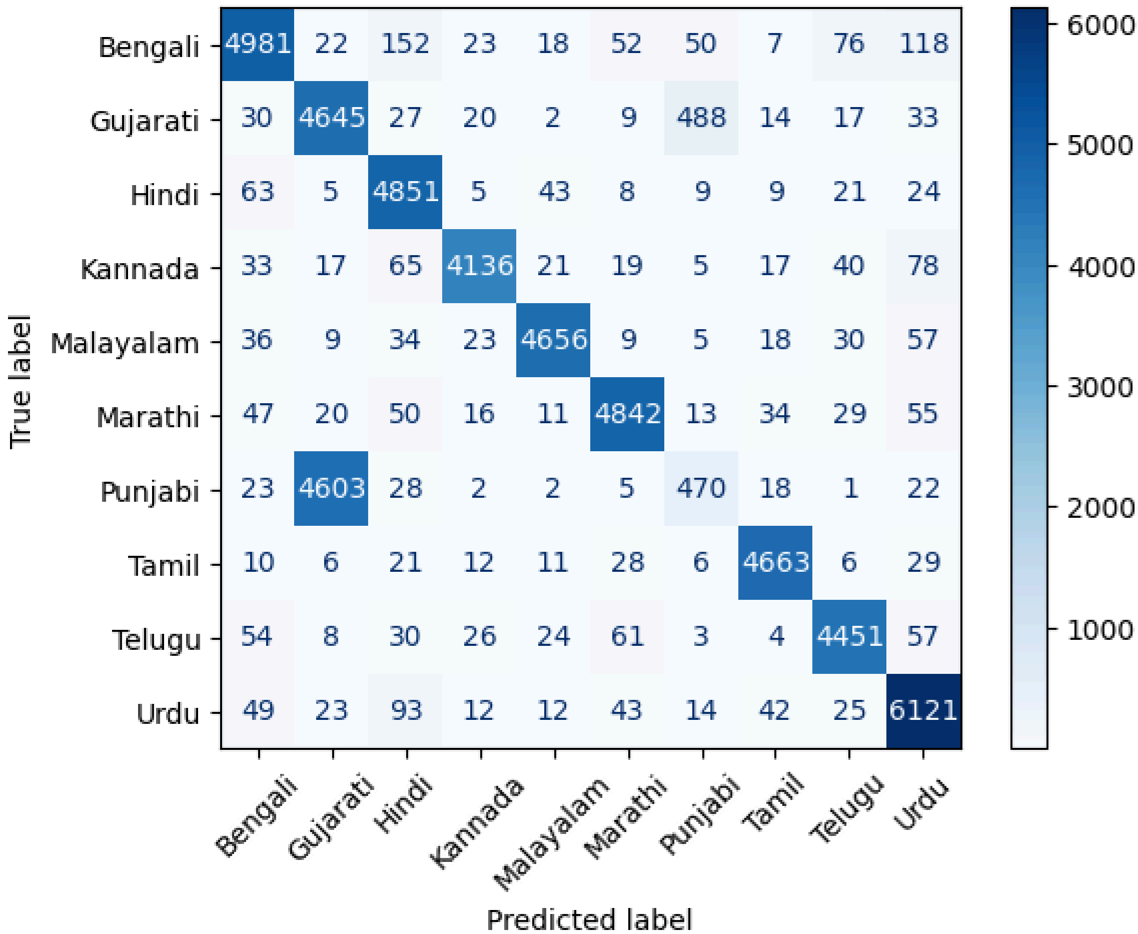
```
Language: Bengali
Mean MFCC: [-3.38417999e+02  1.04311295e+02  6.12678289e+00
1.90650425e+01
 -8.75952721e+00  1.16891257e-01 -3.06492162e+00 -8.39186668e+00
 -2.81908751e+00 -5.75638628e+00 -8.68745232e+00 -3.85192943e+00
 -4.78496838e+00]
Variance MFCC: [4822.7695    531.9459    202.00952   262.94693   174.00536
105.42197
  101.83029    83.90081    29.430325   37.009514   25.948652   23.068861
   20.700611]
--------------------------------------
Language: Gujarati
Mean MFCC: [-286.8727     112.98775    -27.514338     9.51765
-15.783783
   -9.623793    -9.3122      -16.51111     -5.348225    -7.774254
```

```
   -2.893667    -1.4293011   -8.06461  ]
Variance MFCC: [10229.165      259.12497    274.30582     76.586914
86.93688
    69.59949     50.11437     61.301323    30.509954    38.976803
    22.266329    21.345827    18.013695]
--------------------------------------
Language: Hindi
Mean MFCC: [-322.54166     77.156784     8.738599    13.280327
-5.3659363
   -7.1875615  -10.703494    -6.401771    -9.56016     -2.6229339
   -7.2847548   -4.8438406   -8.122824 ]
Variance MFCC: [12163.72       529.33813    101.38249    160.9342
86.4739
    86.41822     99.49275     34.47399     28.025253    19.09622
    33.329834    28.93557     21.003223]
--------------------------------------
Language: Kannada
Mean MFCC: [-296.14578     94.37632    -71.94011     14.752592
-50.843357
   16.682959   -28.943312    -1.5799763  -16.109812     3.1591384
   -1.6090814  -10.110656     6.6712923]
Variance MFCC: [10225.506      284.40094    886.6057     144.52438
405.72934
   134.52827    125.75556     57.434338    38.24019     25.505165
    25.288774    25.312723    51.514538]
--------------------------------------
Language: Malayalam
Mean MFCC: [-347.327       92.03243     -1.0266384   16.441767
5.747666
  -20.340578   -17.158215   -12.322354   -14.710349    -7.12855
  -12.880997    -8.068458    -5.092512 ]
Variance MFCC: [4063.9092    445.55112   197.51689   152.1459    138.17029
119.89247
   77.47965    44.508682   64.60879    56.667324   36.47627    37.399513
   22.821539]
--------------------------------------
Language: Marathi
Mean MFCC: [-300.15338    117.56788    -10.1451645   27.42893
-2.726875
   -6.1159477   -2.6499534  -12.996384    -7.884989    -2.6959121
   -7.0260324   -4.378044    -8.144388 ]
Variance MFCC: [2.62681328e+04 5.25981018e+02 3.74437531e+02
2.11887253e+02
 1.27972755e+02 7.50418777e+01 1.38298569e+02 7.21823425e+01
 5.16352844e+01 3.34368973e+01 2.89190750e+01 1.44503670e+01
 1.89874935e+01]
--------------------------------------
```

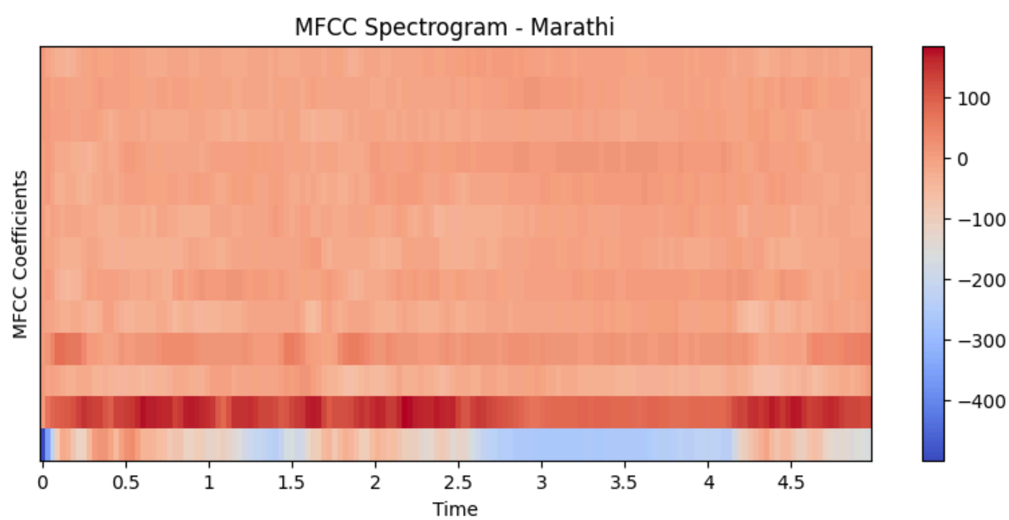
```
Language: Punjabi
Mean MFCC: [-286.2408     113.23248    -27.683508     9.403069
-15.821948
   -9.682325    -9.279035   -16.582664    -5.3308005   -7.8047147
   -2.8850527   -1.4054718   -8.096261 ]
Variance MFCC: [10210.154      250.82321    271.4497      74.62104
85.975365
    69.03857     49.75701     60.635674    30.43391     38.771152
    22.22069     21.073456    17.82212 ]
--------------------------------------
Language: Tamil
Mean MFCC: [-201.53885     110.60108      -9.001206     17.206945
-8.317221
  -19.13632      -9.663549    -14.321725    -12.784455     -7.1464796
  -11.444903     -0.87412083   -9.129541  ]
Variance MFCC: [5293.7837    423.23502   273.7909    185.01651   149.34631
99.37924
  104.805695   36.57995    69.14153    41.821186   49.4489     29.158375
   29.054884]
--------------------------------------
Language: Telugu
Mean MFCC: [-294.44473     84.93009    -11.536548    29.923927   -17.69179
   -2.9616585  -17.753153   -10.901045   -12.484311    -5.9997544
   -9.897438    -8.719697    -2.4484417]
Variance MFCC: [2494.851     745.38745   152.48001   141.80713   296.16037
117.52863
  136.96965    87.016624   56.003662   17.546677   21.52593    25.754898
   52.6728  ]
--------------------------------------
Language: Urdu
Mean MFCC: [-2.9343314e+02  1.0928956e+02  1.0868942e+01  3.0289423e+01
 -9.6097364e+00 -1.2042217e-01 -9.1702032e+00 -9.5645905e+00
 -5.2986135e+00 -4.3679676e+00 -2.6153760e+00 -4.7743087e+00
 -3.9715550e+00]
Variance MFCC: [8198.008     438.79507   248.8397    220.79015   272.50558
106.84272
  144.81123    54.225697   52.723003   38.841175   53.592358   46.72241
   41.81308 ]
--------------------------------------

Epoch 1/20, Loss: 0.6033
Epoch 2/20, Loss: 0.4010
Epoch 3/20, Loss: 0.3639
.
.
.
Epoch 19/20, Loss: 0.2958
```

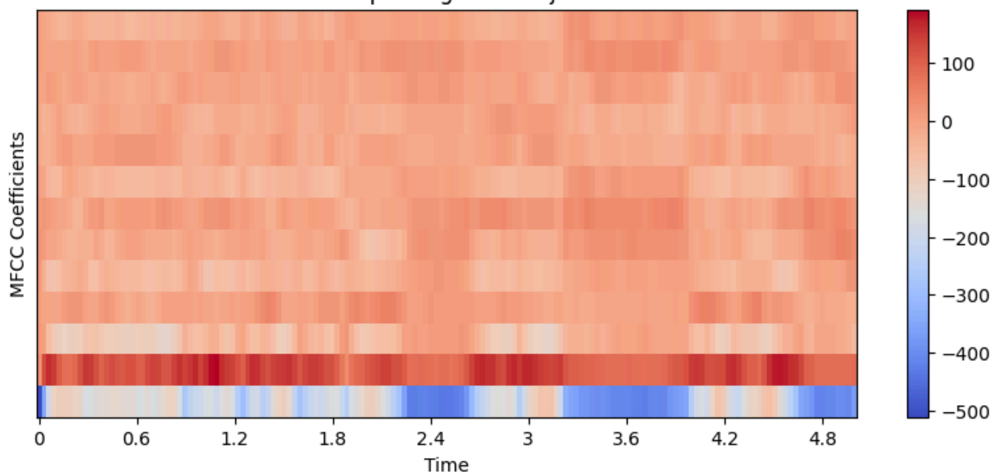
```
Epoch 20/20, Loss: 0.2949
Accuracy: 84.68%
```



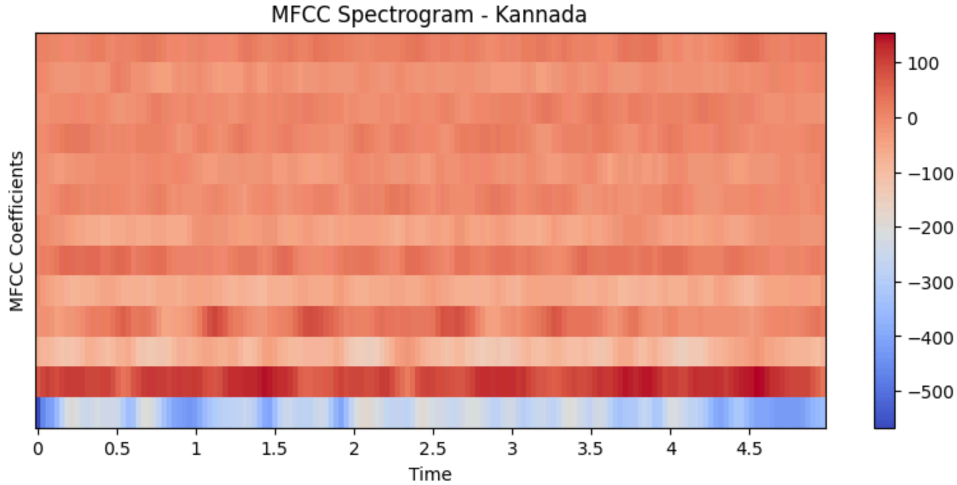

```
Few Other Spectrograms
```



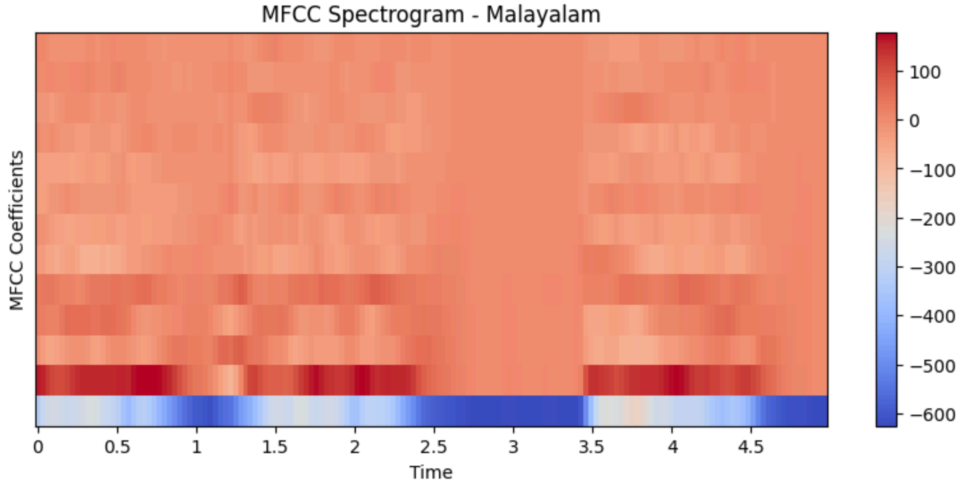
MFCC Spectrogram - Gujarati



MFCC Spectrogram - Kannada



MFCC Spectrogram - Malayalam



5. Challenges in MFCC-Based Language Differentiation

While MFCCs provide valuable spectral information, several challenges arise when using them for language classification:

5.1 Speaker Variability

Different speakers may have varying pitch, articulation, and pronunciation styles, which can influence MFCC values, making classification less reliable.

5.2 Background Noise

Environmental noise and recording conditions can distort MFCC features, leading to misclassification.

5.3 Regional Accents

Languages exhibit significant regional variations in pronunciation, which may overlap with other languages, making it difficult to distinguish them purely using MFCCs.

6. Conclusion and Future Work

MFCCs serve as a useful feature set for distinguishing languages, but they are sensitive to speaker and environmental variations. Future work may incorporate deep learning models, additional acoustic features (e.g., spectrograms, pitch contours), and noise-robust preprocessing techniques to improve classification accuracy.