Raag Sense

A DL outlook to Raga-Recognition in Indian Classical Music

Bhattacharya Brothers

May 2, 2025

Outline

Thaat

What is **Mela/Thaat** system?

A set of 7 swaras arranged in ascending order. The most prominent raga in each group gives its name to the thaat.

Counting of Thaats:

- Total number of thaats in NICM = 1^{2}* 2^{5} = **32**
- Total number of thaats in Carnatic Music = 1^{2} * 2^{3} * 3^{2} = **72**

In practice, we only use 10 Thaats: *Asavari*, *Bhairav*, *Bhairavi*, *Bilawal*, *Kafi*, *Kalyan*, *Khamaj*, *Marva*, *Poorvi*, *Todi*

Raga

- **Definition**: Raga structurally includes the arrangement of swaras, shrutis, Aroha/Avaroha (ascending/descending order of the selected notes) and performatively (raag-bhava) includes ornamentations, phrasing and expression.
- **"Raga"** comes from the word **"Ranj"** meaning to colour or to delight
- Dr. Adrian McNeil's (Australian Sarod-player) analogy to Raga:
  - Ingredients $\rightarrow$ Swaras, phrases
  - Proportions $\rightarrow$ Note emphasis, ornamentation
  - Cooking Method $\rightarrow$ improvisation, presentation
  - Outcome $\rightarrow$ The performance at a glance

Design parameters of a raga

1. **Exhaustive usage**: All swaras selected for the raga must be used.
2. **Vadi system**:
    1. **Vadi**: The most important swara [King of the raga kingdom]
    2. **Samvadi**: 2^{nd} most important swara [Minister]
    3. **Anuvadi**: Other supporting swaras [Subjects of the kingdom]
    4. **Vivadi**: Forbidden notes [Enemies of the kingdom]
3. **Alpatva** & **Bahutva**
4. **Graha** & **Nyasa**
5. **Gamakas** (ornamentation):
    1. **Meend**: Glides
    2. **Kampana**: Vibrato
    3. **Jhatka**: sudden jump
    4. **Khatka**: fast note cluster
    5. **Zamzama**: fast ornamentation with repeated notes

Indian Classical Music (ICM)

- ICM is a rich tradition based on **ragas** — melodic frameworks encoding:
  - Note sequences
  - Ornamentation (gamakas)
  - Emotional context and time of performance
- **Challenge**: Automatic raga classification is complex due to:
  - Scale variance
  - Microtonal nuances
  - Data inconsistencies
- **RaagSense** objective: A deep learning system to identify ragas present in any music audio, using Convolutional Neural Networks (CNNs).

Literature Review

- **Tonic-Independent Classification** (Tejaswi & Chowdhary, 2021):
  - Proposed tonic-independent framework using data augmentation.
  - Limitation: Assumes structured dataset, ignores data quality.
- **Explainable Deep Learning Analysis for Raga Identification in Indian Art Music** (Singh & Arora, 2022):
  - Introduced PIM-v1 dataset (191 hours of Hindustani music).
  - Limitation: Does not address tonic variation or non-ICM content.
- **Our Contribution**:
  - Emphasis on dataset curation and authenticity.
  - Manual scrutiny of audios containing alap, jor and bandish sections.
  - Avoidance of pitch normalization to preserve microtonal features.

Dataset fetched

- *Data Source*: Thaat and Raga Forest (TRF) Dataset [4]
- Count of audio samples: 1180
- Total duration: 241.35 hours
- **Observations**: Major portions of the dataset contained non-ICM audios, noisy audio. There was a lack of quality in the dataset. Post expert scrutiny and filtering out the non-ICM audios, the findings were:
  1. # ICM audio = 559
  2. # non-ICM audio = 621

Dataset Curation

1. Eliminated non-ICM audios based on expert's scrutiny
2. Populated the dataset with ICM audios downloaded from public audio libraries.
3. Resultant volume of dataset: 654 files
4. Total duration: 197.50 hours

- **Challenges**: Despite such a large dataset, for every raga class, different audio files are sung in different scales. Misleads the extracted feature information.

Final Working Dataset

1. *Verified raga elaboration*: Presence of alap, jor and bandish.
2. Ensured *scale consistency* per raga class.
3. Only those classes were considered that have significant number of audio samples.
4. Resultant volume of dataset: **315** files (each of 3min 30seconds)
5. Total duration: **17.15 hours**
6. Number of classes (ragas): **41**

EDA

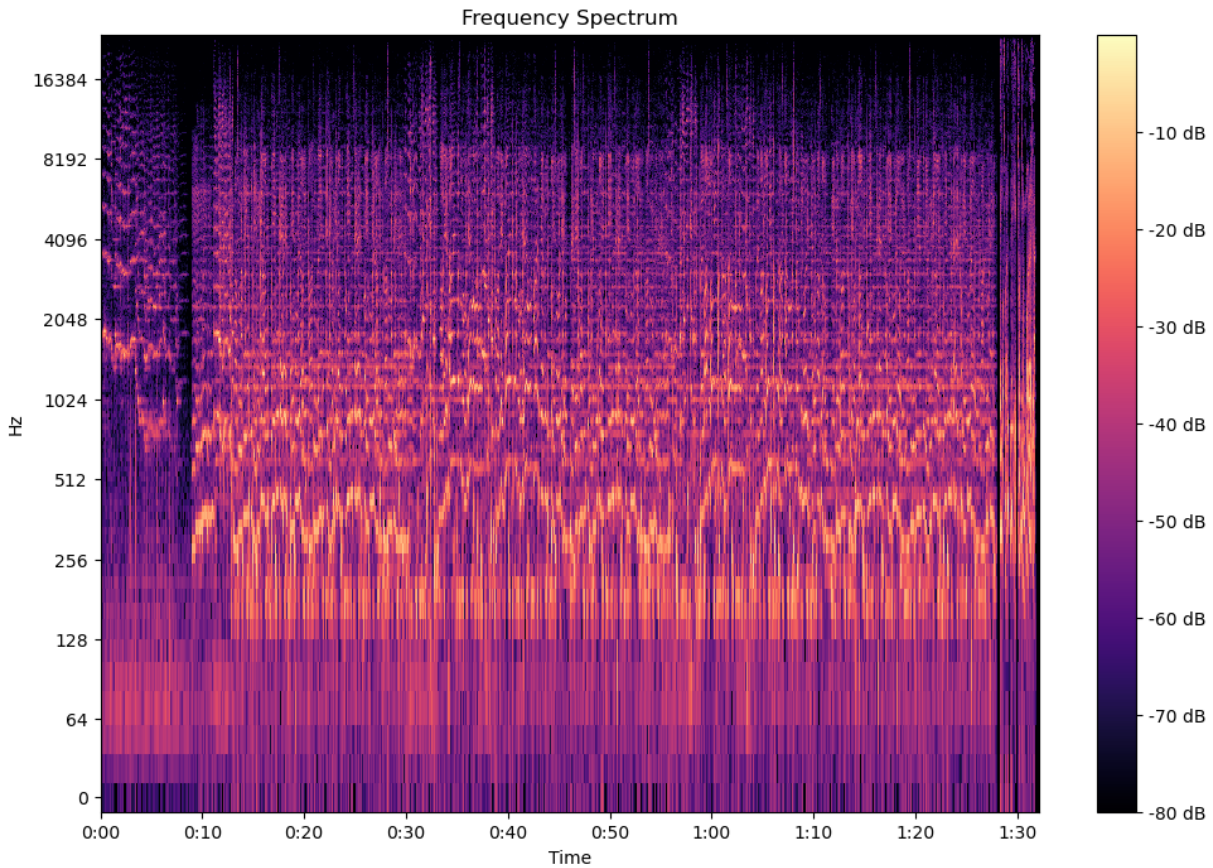
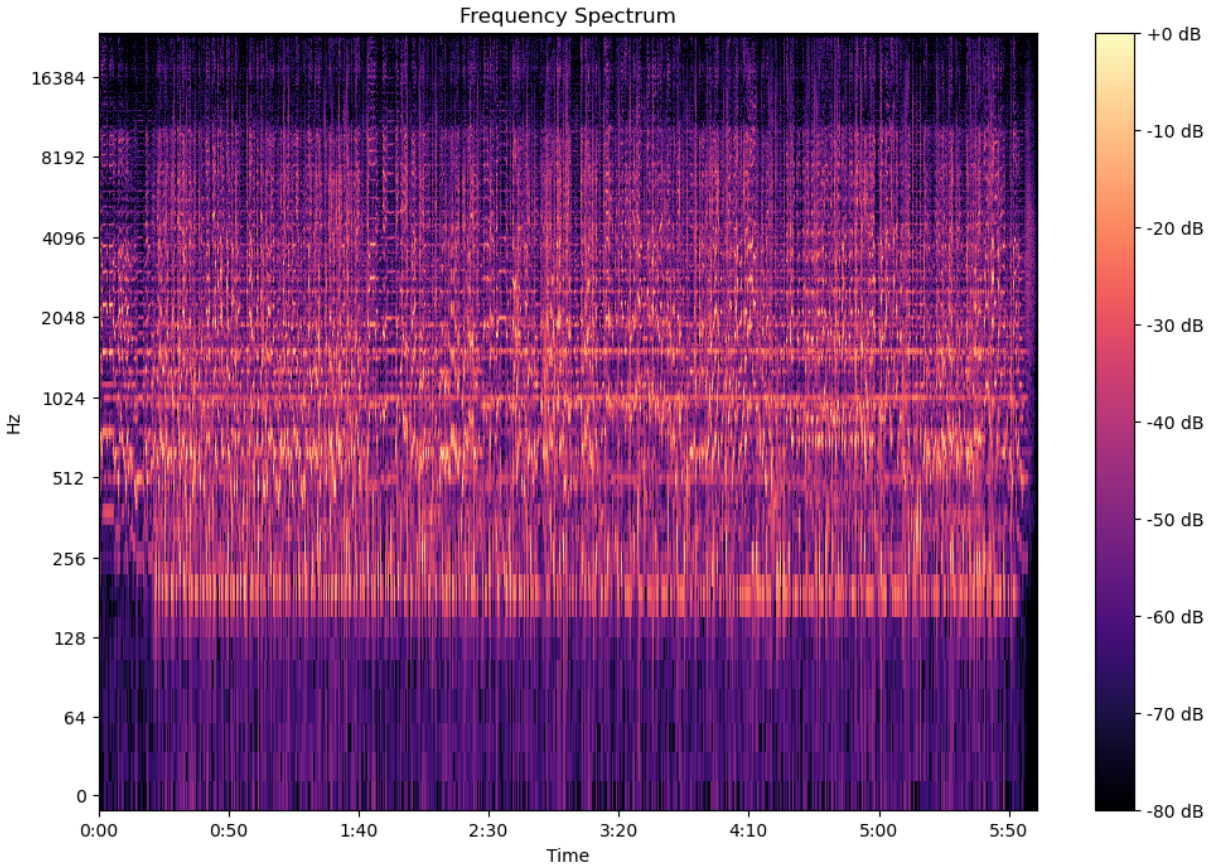


Figure 1: Non-ICM audio

EDA



ICM audio

EDA I

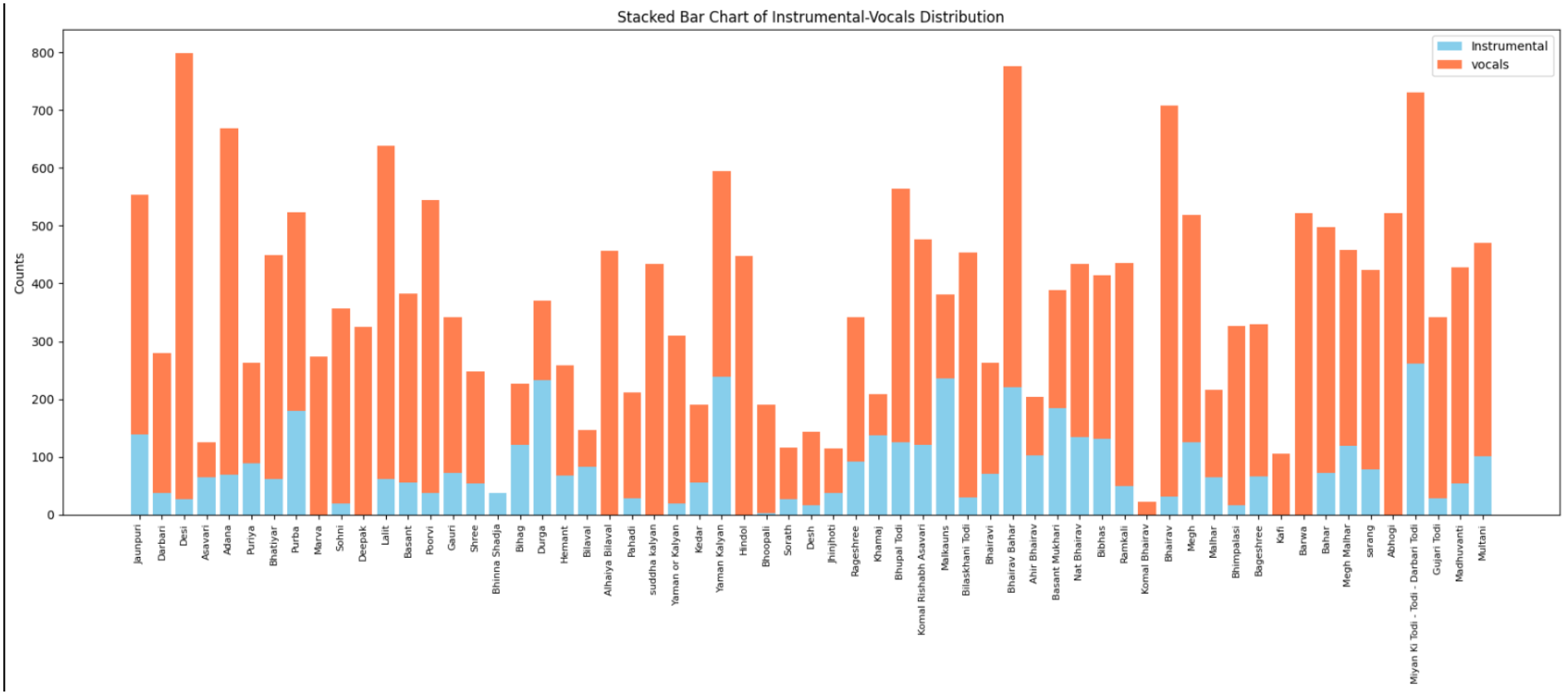


Figure 2: Instrument-Vocals Distribution of Curated Dataset

EDA II

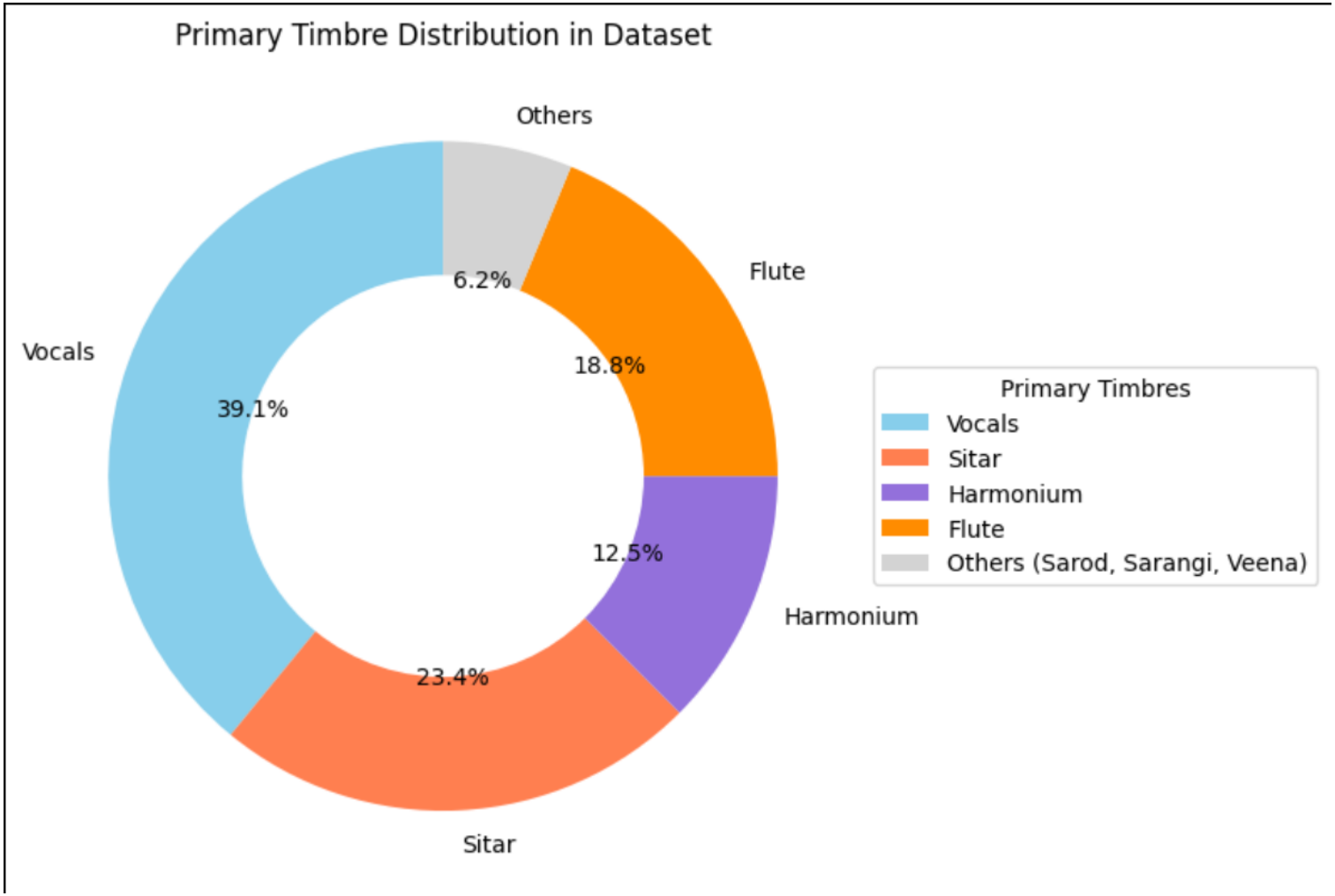


Figure 3: Primary Timbre Distribution of Curated Dataset

EDA III

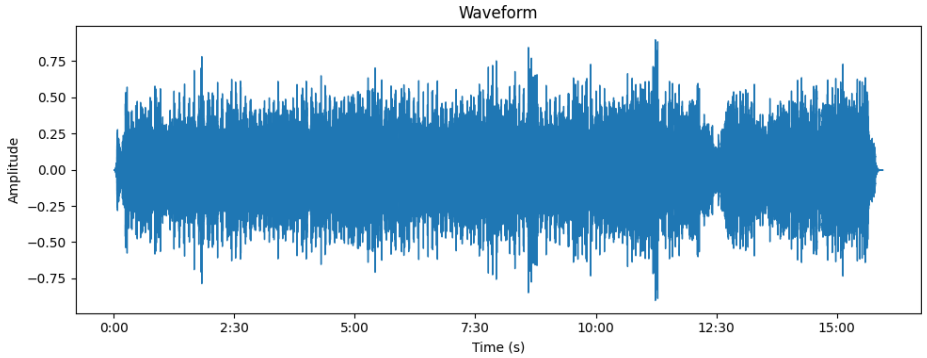


Figure 4: Raag Ahir Bhairav (Thaat Bhairav) waveform

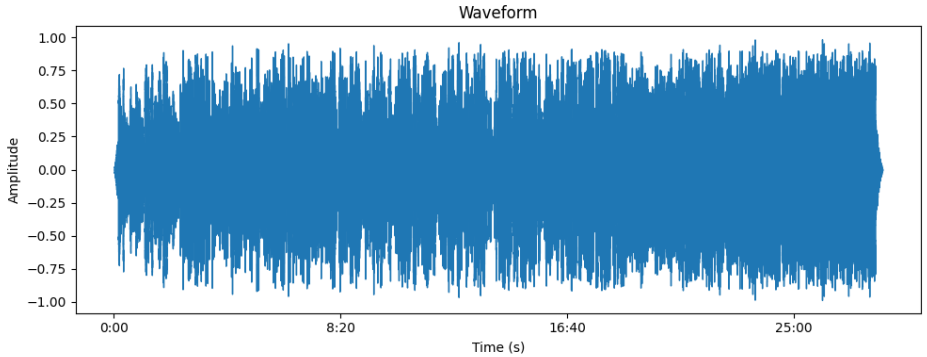


Figure 5: Raag Durga (Thaat Bilaval) waveform

EDA IV

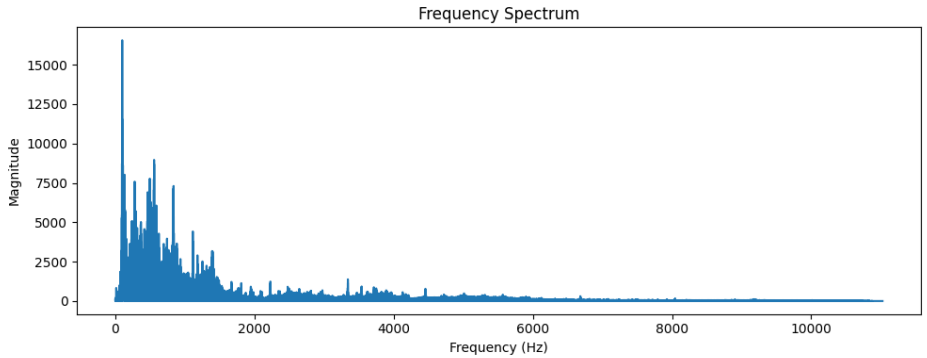


Figure 6: Raag Ahir Bhairav (Thaat Bhairav) frequency spectrum

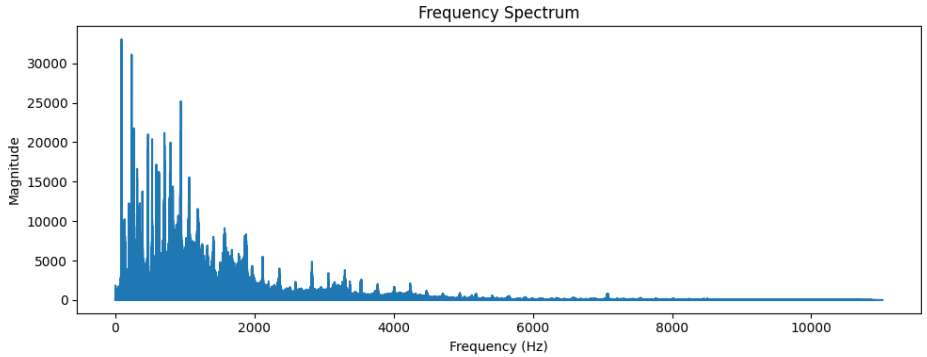


Figure 7: Raag Durga (Thaat Bilaval) frequency spectrum

EDA IV

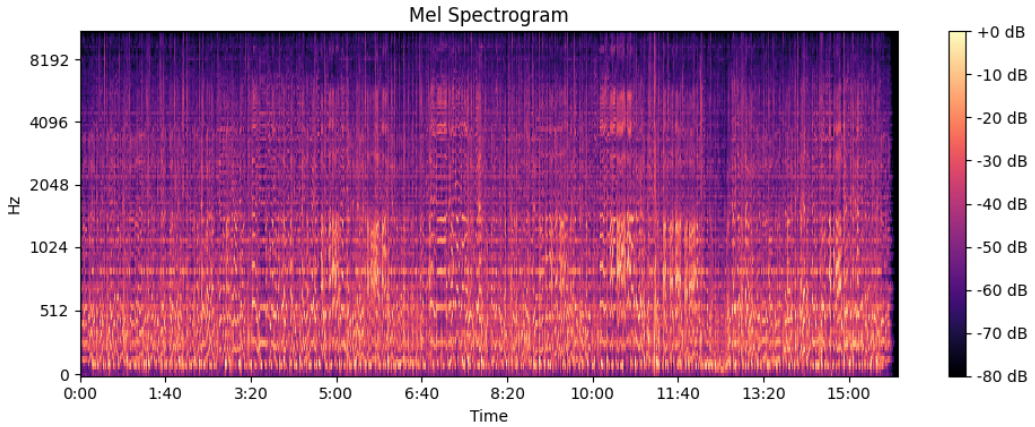


Figure 8: Raag Ahir Bhairav (Thaat Bhairav) Mel-spectrogram

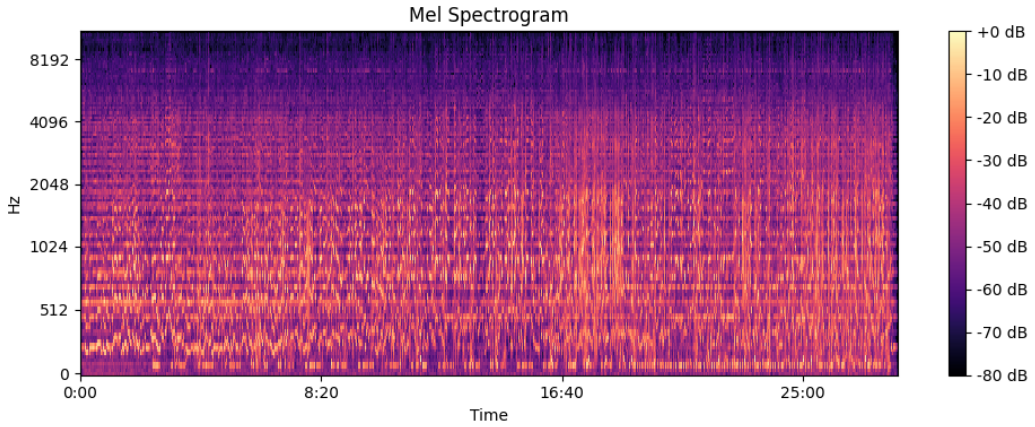


Figure 9: Raag Durga (Thaat Bilaval) Mel-spectrogram

Preprocessing

- Considered the best ICM-explainable audio for each of the 41 ragas (or classes)
- Standardized to 3:30 duration
- 22,050 Hz sampling rate.
- Raga labels extracted from file names (e.g., `Bageshree_vocals_16_8.wav` $\rightarrow$ `Bageshree`).

Feature Extraction

- **MFCC** (Mel-Frequency Cepstral Coefficients): It transforms an audio signal into a set of coefficients that capture the essential characteristics of the sound.
  - $PreEmphasis \rightarrow Framing \rightarrow Windowing \rightarrow FFT \rightarrow Mel - FilterBank \rightarrow DiscreteCosineTransform$
- **Captures General Tonal Texture**: Useful for distinguishing broader categories of ragas or identifying stylistic patterns.
- **Train/Test Split**: 80-20.

CNN Architecture

- Implemented a 1D CNN using `TensorFlow Keras` for raga classification.
  - Used two `Conv1D` layers with 32 and 64 filters (kernel size = 5), each followed by `MaxPooling1D`.
  - Flattened output passed to a dense layer with 128 units and ReLU activation.
  - Applied 0.5 `Dropout` to reduce overfitting.
  - Final `Dense` layer with 41 units and `softmax` activation for multiclass output.
  - Compiled with `Adam` optimizer and `categorical_crossentropy` loss.

Results I

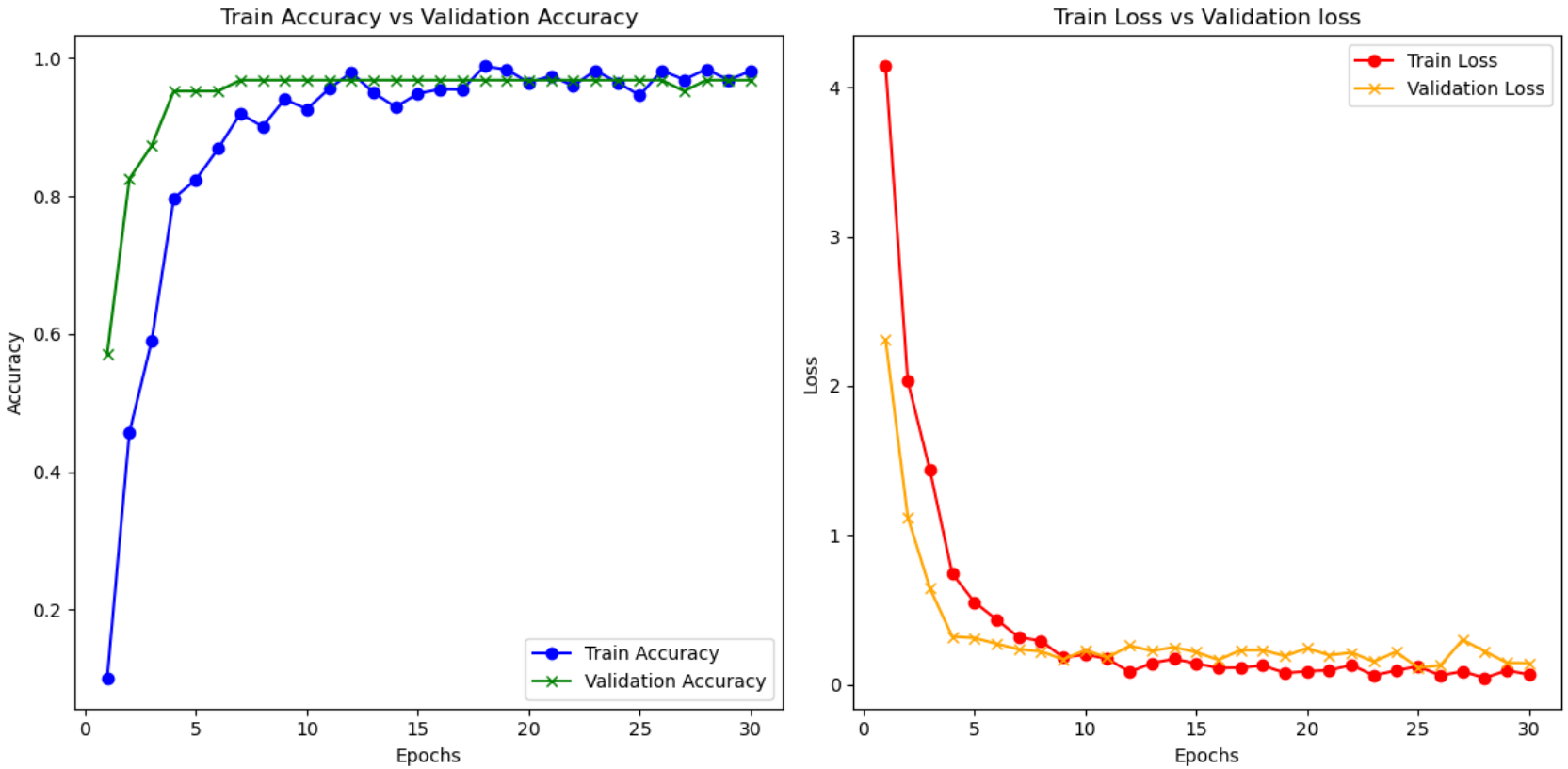


Figure 10: Accuracy & Loss Plot

Results II

- **Dataset Details**:
  - Total audios: 653
  - Classes: 41 ragas
  - Duration: 3:30, Sampling Rate: 22,050 Hz
- **Performance**:
  - Test Accuracy: **96.82%**
  - Validation Loss: Stabilized at 0.1452
  - Training Time: 6 seconds (GPU)

Future Work

- **Feature-Specific Models**: Train ensemble models on individual features (MFCC, Chroma, Pitch) and combine via soft voting.
- **Frontend UI**: Develop a real-time raga recognition application.
- **Pitch Normalization**: Explore non-destructive methods for scale alignment.
- **CRNN Models**: Implement recurrent architectures to capture temporal dynamics of ragas.

Summary

- RaagSense achieves **96.82% accuracy** in classifying 41 ragas using a CNN-based approach.
- Key contributions:
  - Meticulous dataset curation for ICM authenticity.
  - Focus on MFCC features for efficient classification.
  - Comprehensive EDA to ensure data quality.
- Sets a high benchmark for ICM-aware raga classification, with potential for real-world applications.

Thank You!

References

[1] *Appreciating Hindustani Music - Course — onlinecourses.nptel.ac.in*. https://onlinecourses.nptel.ac.in/noc22_hs57/preview. [Accessed 04-04-2025].

[2] Sathwik Tejaswi Madhusudhan and Girish Chowdhary. *Tonic Independent Raag Classification in Indian Classical Music*. https://openreview.net/pdf?id=HJz9K7kJcX. 2023.

[3] Parampreet Singh and Vipul Arora. *Explainable Deep Learning Analysis for Raga Identification in Indian Art Music*. Accessed: 2025-05-02. 2024. DOI: 10.48550/arXiv.2406.02443. arXiv: 2406.02443 [eess.AS]. URL: https://arxiv.org/pdf/2406.02443.

[4] *Thaat and Raga Forest (TRF) Dataset — kaggle.com*. https://www.kaggle.com/datasets/suryamajumder/thaat-and-raga-forest-trf-dataset. [Accessed 04-04-2025].