Speech Understanding Assignment 1: Spectogram and Windowing Techniques

Aditi Baheti
Roll No: M23CSA001

February 3, 2025

1 Problem Statement - Task A

This task analyzes the impact of Hann, Hamming, and Rectangular windowing techniques on spectrogram generation using STFT. Extracted spectrogram features train an SVM classifier, evaluated by accuracy, precision, recall, and F1-score.

1.1 Colab Link

For colab file, visit this link.

1.2 Approach

We used the UrbanSound8k dataset (8,732 samples across 10 categories). Audio samples were processed with STFT using Hann, Hamming, and Rectangular windows to generate spectrograms. MFCCs and their derivatives were extracted as feature vectors. An SVM classifier was trained with hyperparameter tuning via grid search. Performance was evaluated using a test set and a confusion matrix.

1.3 Results and Discussion

1.3.1 Spectrogram Analysis

The visual comparison of spectrograms revealed that:

- The **Hann Window** produced smooth transitions in the frequency domain, reducing spectral leakage and ensuring a well-defined spectral representation.
- The **Hamming Window** performed similarly to the Hann Window, with minor improvements in frequency resolution at the cost of slightly higher side lobes.
- The **Rectangular Window** led to significant spectral leakage, distorting the frequency representation and making it less suitable for classification tasks.

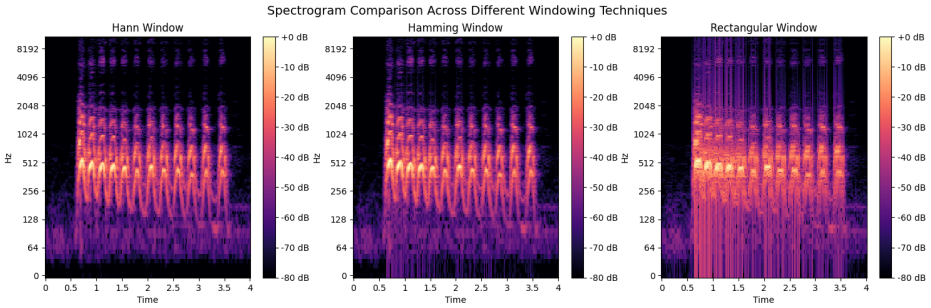


Figure 1: Spectrogram Comparison Across Different Windowing Techniques.

1.3.2 Classification Performance

The classifier achieved an accuracy of **89%**. Table 1 summarizes the evaluation metrics.

Metric	**SVM (Linear)**
Accuracy	89%
Precision (Macro Avg)	0.70
Recall (Macro Avg)	0.74
F1-Score (Macro Avg)	0.70

Table 1: Classification performance metrics.

1.3.3 Confusion Matrix Analysis

Key Observations:

- **Strong Performance for Certain Classes:** The classifier performs well on classes with a higher number of samples, such as class 2 (11 correct predictions), class 5 (8 correct predictions), and class 8 (15 correct predictions).
- **Misclassifications in Similar Classes:** Some classes show minor confusion, such as class 3 being misclassified as class 0 and class 9 being misclassified as class 8. These errors may be due to overlapping frequency content in audio samples.
- **Challenges with Underrepresented Classes:** Classes like 1, 4, and 6 have sparse correct predictions, likely due to low representation in the dataset or overlapping features with other classes. The lower recall for these classes suggests a need for data augmentation or re-sampling strategies to improve classification performance.

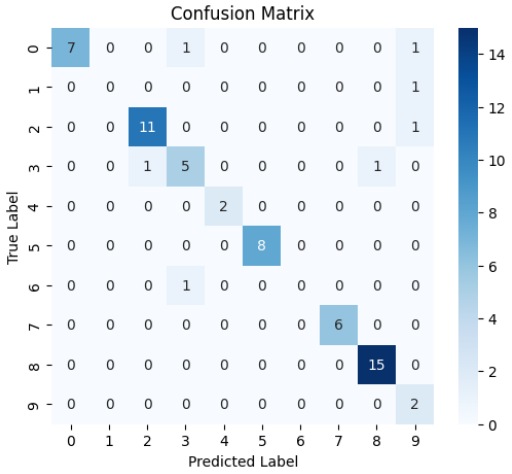


Figure 2: Confusion Matrix for SVM Classifier.

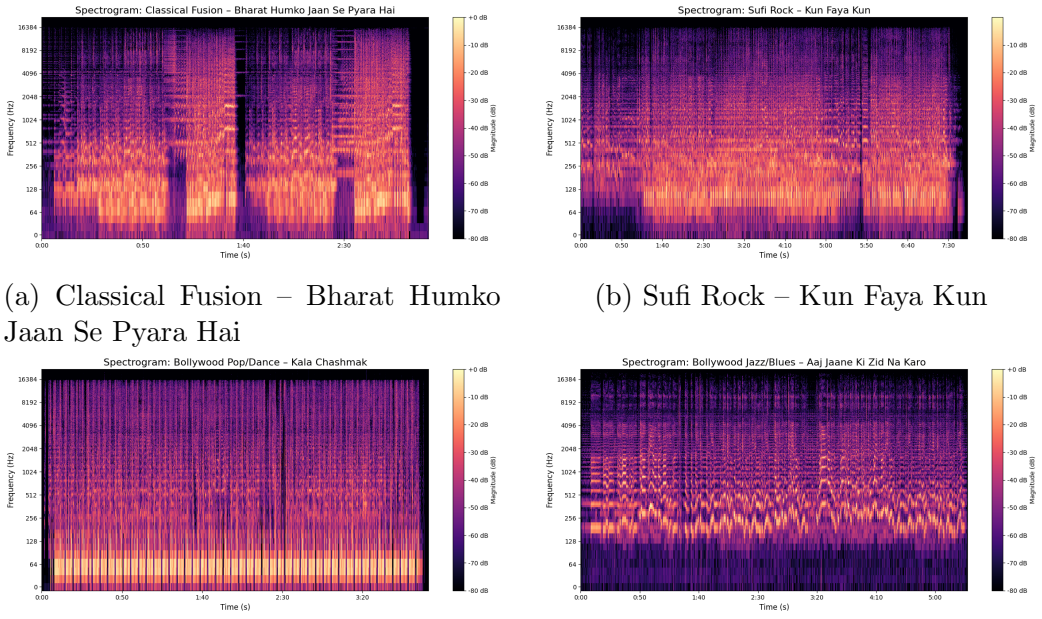
2 Task B: Comparative Analysis of Spectrograms from Different Music Genres

This section compares spectrograms of four songs from different genres—Classical Fusion, Sufi Rock, Bollywood Pop/Dance, and Bollywood Jazz/Blues—based on their spectral characteristics.

2.1 Colab Link

For colab file, visit this link.

2.2 Spectrogram Comparison



(a) Classical Fusion – Bharat Humko Jaan Se Pyara Hai

(b) Sufi Rock – Kun Faya Kun

(c) Bollywood Pop/Dance – Kala Chashmak

(d) Bollywood Jazz/Blues – Aaj Jaane Ki Zid Na Karo

Figure 3: Spectrogram Comparison Across Four Music Genres

2.3 Analysis

Classical Fusion (Bharat Humko Jaan Se Pyara Hai): The spectrogram displays a wide frequency range with distinct harmonic patterns, reflecting the orchestral depth and traditional Indian instruments like flute and veena. Smooth transitions and dynamic shifts emphasize the song's evolving melodic structure, characteristic of classical fusion.

Sufi Rock (Kun Faya Kun): Sustained mid and low frequencies dominate, highlighting deep, meditative qawwali-style vocals. Smooth spectral transitions indicate spiritual fluidity, while subtle rhythmic elements blend traditional Sufi essence with modern rock instrumentation.

Bollywood Pop/Dance (Kala Chashmak): Strong low-frequency beats and periodic bursts of high frequencies signify electronic bass and percussion-driven energy. The structured rhythmic pattern, typical of Bollywood dance tracks, enhances its upbeat and club-like appeal.

Bollywood Jazz/Blues (Aaj Jaane Ki Zid Na Karo): A fluid frequency spectrum with alternating intensity regions showcases expressive vocals, jazz-influenced piano, and wind instruments. The presence of blue notes and syncopation creates a soothing, emotive ambiance, fitting Bollywood's classic romantic style.

3 Conclusion

This study examined windowing techniques for spectrogram generation (Task A) and analyzed spectrograms across music genres (Task B).
In Task A, Hann and Hamming windows produced clearer spectral representations, improving classification accuracy (89%), while the Rectangular window exhibited spectral leakage.
In Task B, genre analysis revealed distinct spectral patterns—Classical Fusion and Jazz/Blues displayed rich harmonic structures, while Pop/Dance and Sufi Rock emphasized rhythmic patterns and sustained frequencies.