CSL7770: Speech Understanding

Assignment 1 - Question 2

Report

Aditya Dhaduk
B21AI014

1. Introduction

The **UrbanSound8K** dataset is a collection of 8,732 labeled sound excerpts (<= 4s) from ten urban sound classes such as air conditioner, car horn, and children playing. The dataset is used for machine learning tasks related to environmental sound classification.

This report explores:

- Windowing techniques (Hann, Hamming, Rectangular)
- Spectrogram generation using Short-Time Fourier Transform (STFT)
- Training a Support Vector Classifier and comparing results with different windowing methods

---

2. Windowing Techniques

Windowing is essential for spectral analysis and STFT. I implemented the following:

a) Hann Window

- Smooth window reducing spectral leakage.

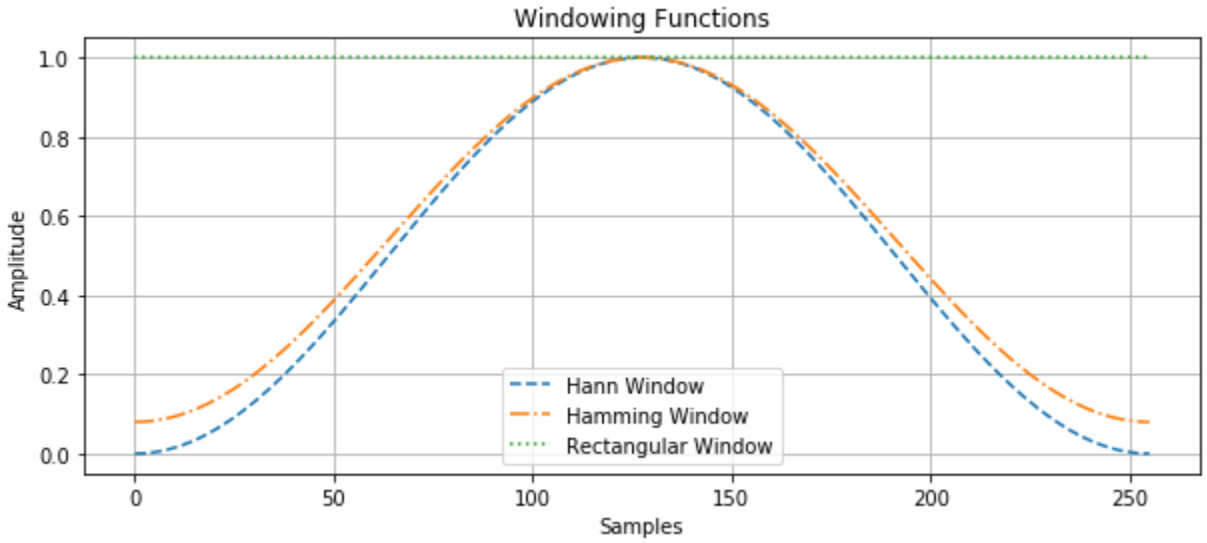
b) Hamming Window

- Similar to Hann but with a slightly different shape, reducing side lobes.

c) Rectangular Window

- Simple window without tapering, leading to higher spectral leakage.

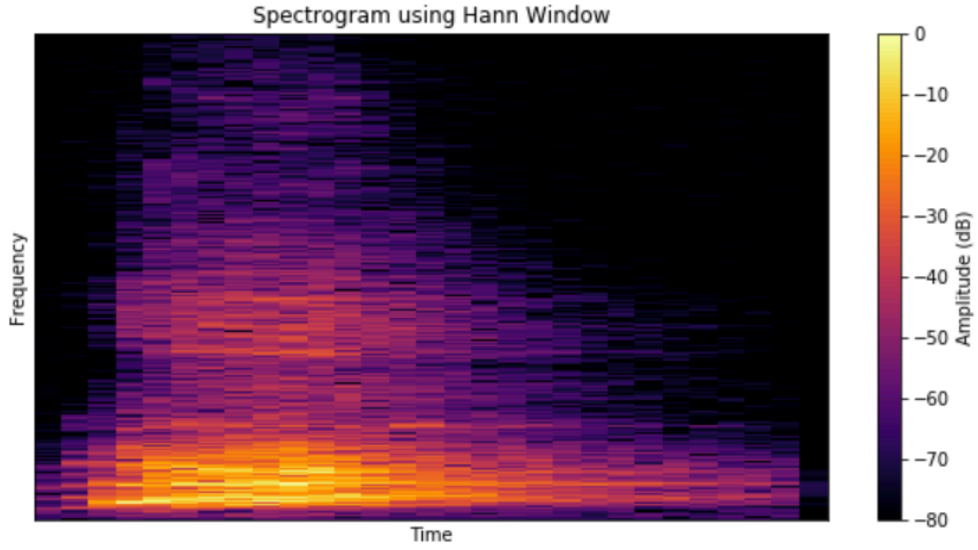
Below is the visualisation for all 3 window functions.

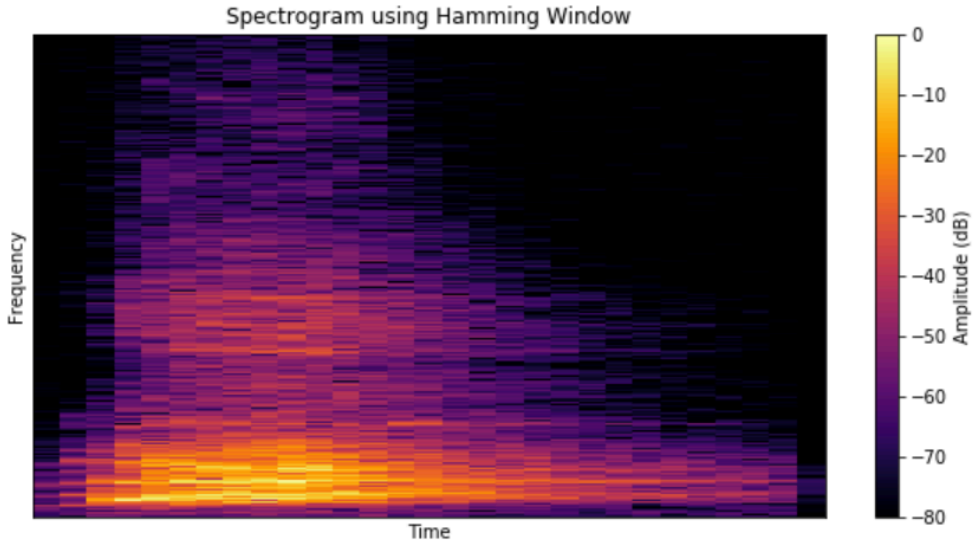


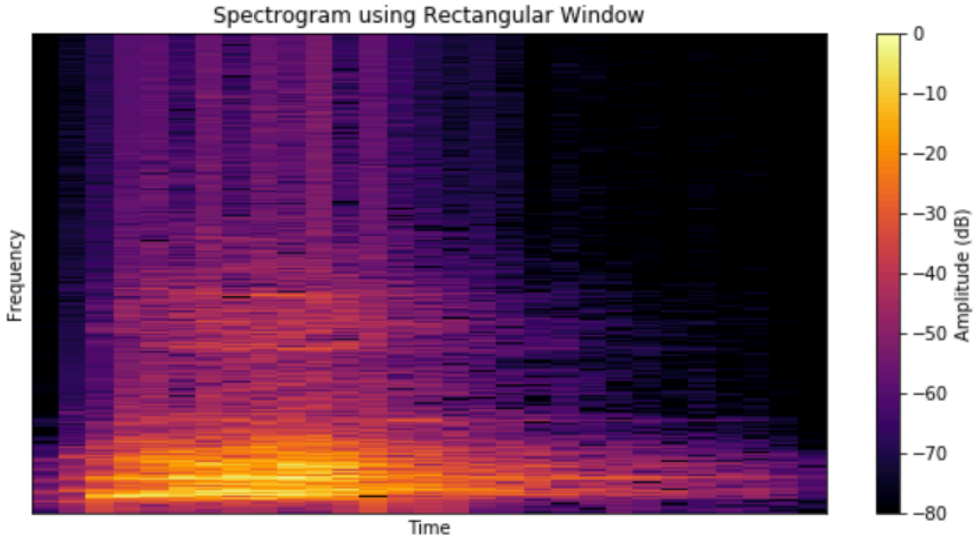
---

3. Spectrogram Generation using STFT

Using **Librosa** and **SciPy**, I generated spectrograms from audio samples using different window functions. STFT converts signals from the time domain to the frequency domain, allowing visualization of spectral characteristics.







Observations:

- **Hann and Hamming Windows** provide clearer frequency separation.
- **Rectangular Window** introduces more artifacts due to higher spectral leakage.

---

4. Classification Model

I trained a **Support Vector Machine (SVM)** classifier using features extracted from the spectrograms.

Feature Extraction:

- **Log-Mel Spectrograms** were computed for each audio file.
- Each spectrogram was flattened into a feature vector.

Training:

- **Train-Test Split**: 80% training, 20% testing.
- **Three separate models** trained for different window types.

Accuracy Comparison:

Window Type	Accuracy
Hann Window	27.88%
Hamming Window	28.11%
Rectangular Window	25.30%

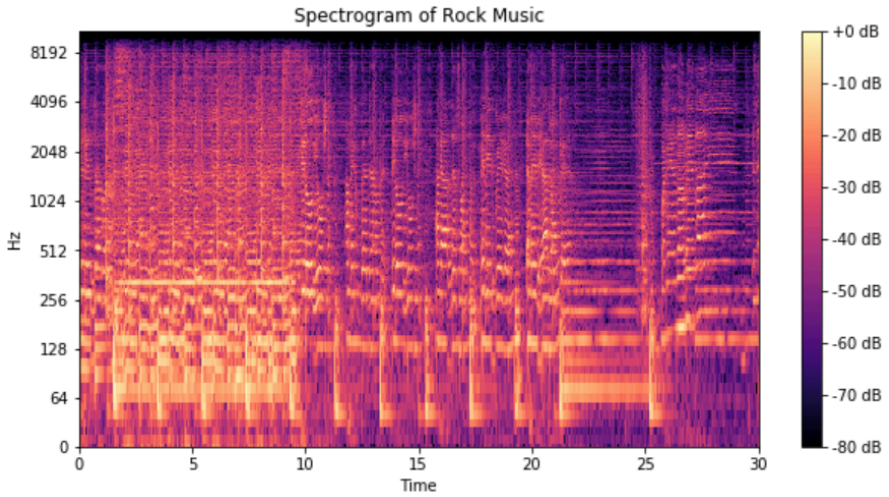
The Hann and Hamming windows performed better due to their ability to reduce spectral leakage.

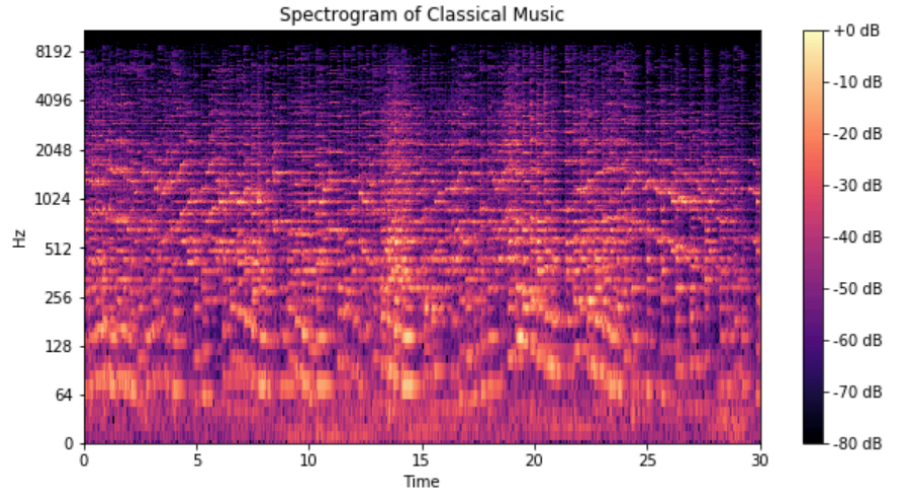
---

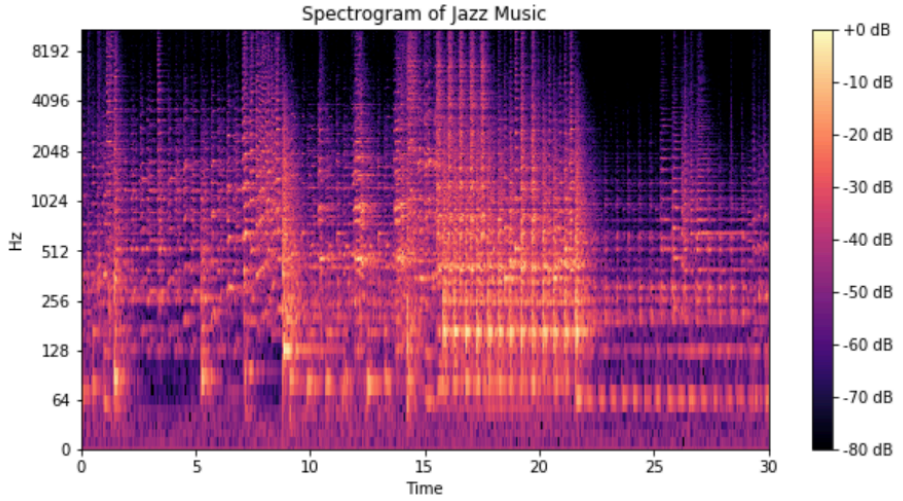
5. Genre-Based Spectrogram Analysis

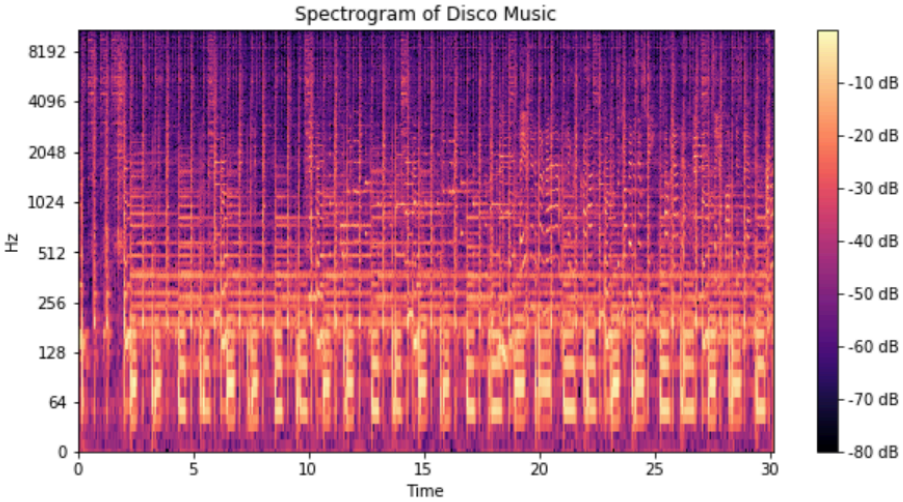
I selected four genres (Disco, Classical, Rock, Jazz) from the GTZAN dataset and compared their spectrograms.

Observations:









- **Disco & Rock**: High-energy, rapid transients, dense spectral components.
- **Classical**: Smooth transitions, harmonic structures.

- **Jazz**: Moderate complexity, structured energy distribution.

---

6. Conclusion

This assignment demonstrated the impact of **windowing techniques** on spectrogram quality and **classification performance**. The Hann and Hamming windows provided better results for urban sound classification, while the rectangular window introduced spectral artifacts.

---

7. Github Repository Link

https://github.com/Aditya2814/Speech-Understanding---Assignment-1