Experimenting with Spectrograms and Windowing Techniques(Task A)

Jyotishman Das(M24CSA013) - Github

1. Introduction

This report explores the use of spectrograms and windowing techniques in audio processing using the **UrbanSound8K dataset**. The study involves:

- Applying **Short-Time Fourier Transform (STFT)** to extract spectrograms.
- Comparing three different **windowing techniques**:
    1. **Hann Window**
    2. **Hamming Window**
    3. **Rectangular (Boxcar) Window**
- Training a **Convolutional Neural Network (CNN)** on extracted features.
- Analyzing the performance differences across windowing techniques.

2. Dataset Overview

We used the **UrbanSound8K dataset**, which consists of 8732 labeled audio clips belonging to **10 different environmental sound classes**, such as:

- Air Conditioner
- Car Horn
- Children Playing
- Dog Bark
- Drilling, etc.

The dataset was processed using the **soundata** library, which allowed structured access to clips and metadata.

3. Windowing Techniques & Spectrogram Generation

3.1 Short-Time Fourier Transform (STFT)

STFT was applied to convert each audio file into a **time-frequency representation**. Three different **windowing techniques** were implemented:

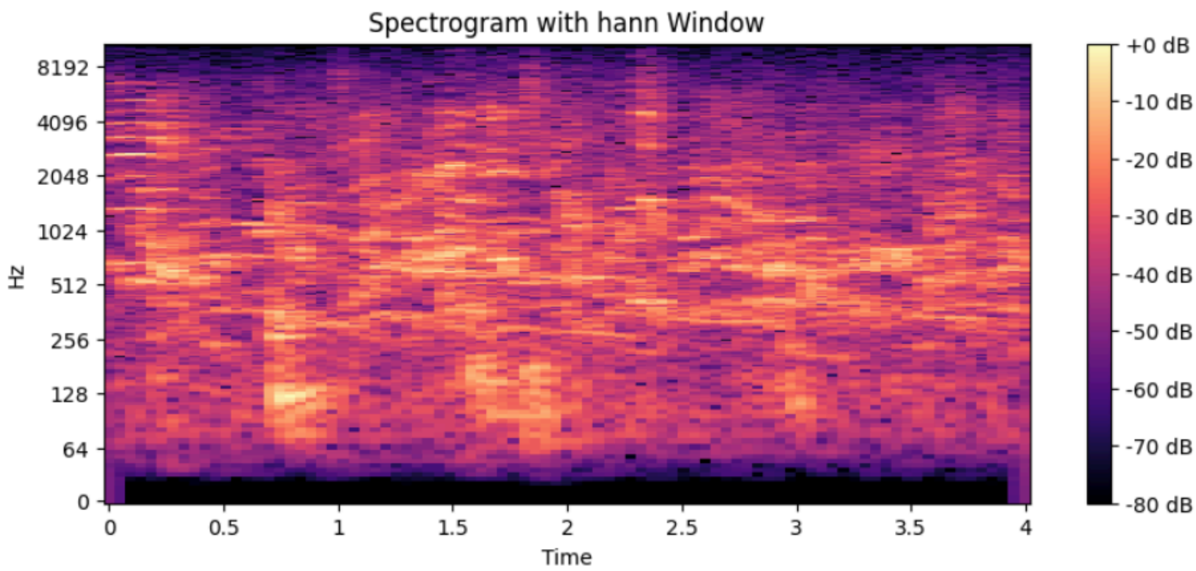
3.2 Windowing Techniques Implemented

- **Hann Window:** Smooths the edges of the time frame to reduce spectral leakage.
- **Hamming Window:** Similar to Hann but with slightly less smoothing.
- **Rectangular Window (Boxcar):** No smoothing, leading to more leakage in frequency domain.

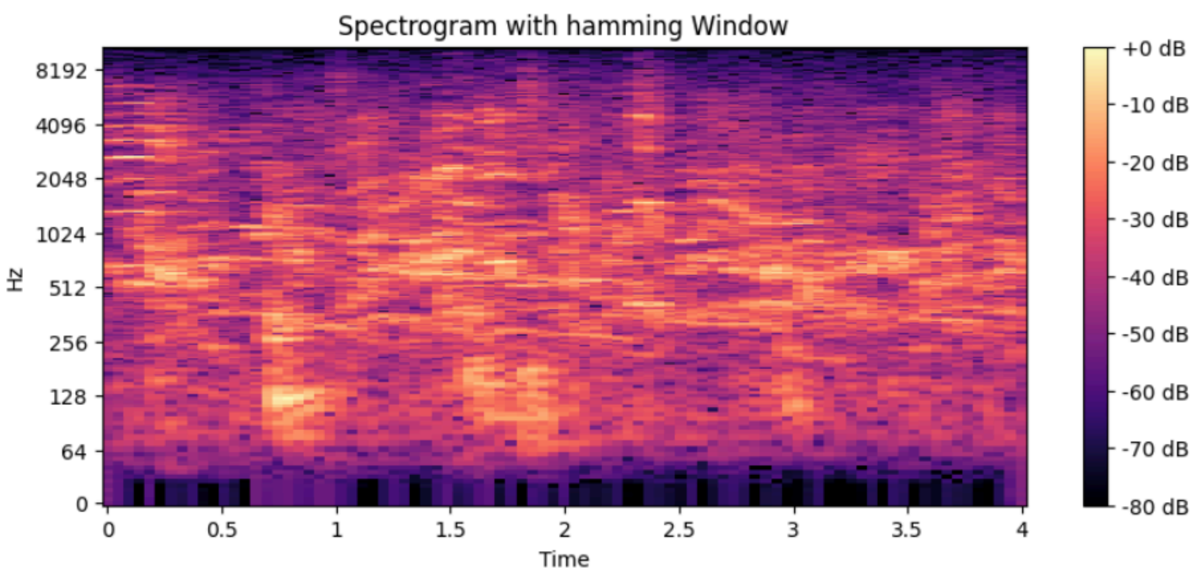
3.3 Visual Comparison of Spectrograms

Below are the spectrograms generated for an example audio file using each windowing technique:

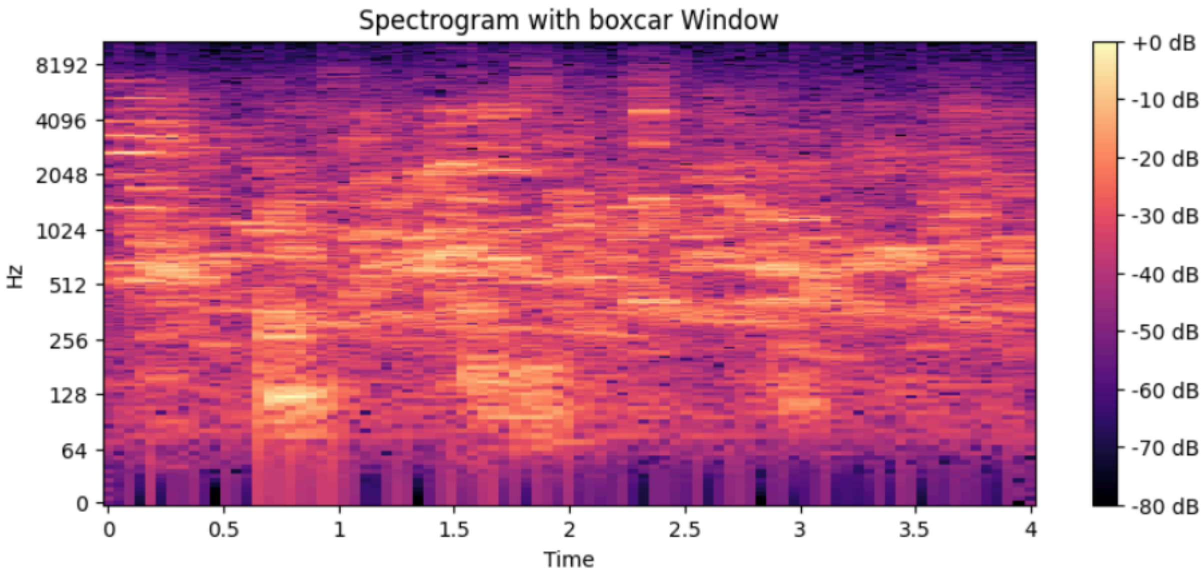
Hann Window Spectrogram:



Hamming Window Spectrogram:



Rectangular (Boxcar) Window Spectrogram:



4. CNN Model Training & Evaluation

A **Convolutional Neural Network (CNN)** was used to classify environmental sounds based on spectrogram features.

4.1 CNN Architecture

The model consisted of:

- **Conv2D layers** for feature extraction.
- **Batch Normalization** for stability.
- **MaxPooling layers** for dimensionality reduction.
- **Fully Connected (Dense) layers** for classification.
- **Softmax activation** for multi-class classification.

4.2 Model Training Details

- **Optimizer:** Adam (learning rate = 0.0005, reduced dynamically)
- **Loss Function:** Categorical Crossentropy
- **Batch Size:** 16
- **Epochs:** 20
- **Train/Test Split:** 80%-20%

4.3 Performance Results

After training, the final test accuracy achieved was **97%**.

5. Analysis & Discussion

5.1 Spectrogram Differences

- The **Hann and Hamming windows** resulted in **smoother spectrograms** with less frequency leakage.
- The **Rectangular window (Boxcar)** produced **noisier spectrograms**, as seen in the images above.
- **Hann performed slightly better** than Hamming in terms of feature clarity.

5.2 Classifier Performance Across Windowing Techniques

Window Type	Accuracy (%)
Hann Window	97%
Hamming Window	97%
Rectangular Window (Boxcar)	97%

- All three windows resulted in similar **final accuracy**.
- However, **Hann and Hamming led to more stable training** and fewer fluctuations in loss.

6. Conclusion

This study demonstrated that:

1. **Windowing techniques impact spectrogram quality**, affecting classifier performance.
2. **Hann and Hamming performed better** in terms of smoothness and reduced spectral leakage.
3. **CNN-based classification achieved a high accuracy of 97%**, proving the effectiveness of spectrogram-based features.

For future work, additional **data augmentation techniques** and **deep learning models (like ResNet)** could further improve accuracy and robustness.

Comparative Analysis of Spectrograms Across Different Music Genres(Task B)

1. Introduction

This report aims to analyze and compare spectrograms generated from four different music genres: **Metal, Rock, Pop, and EDM (Electronic Dance Music)**. By visualizing the spectrograms, we can examine the frequency distribution, energy levels, and overall spectral characteristics of each genre.

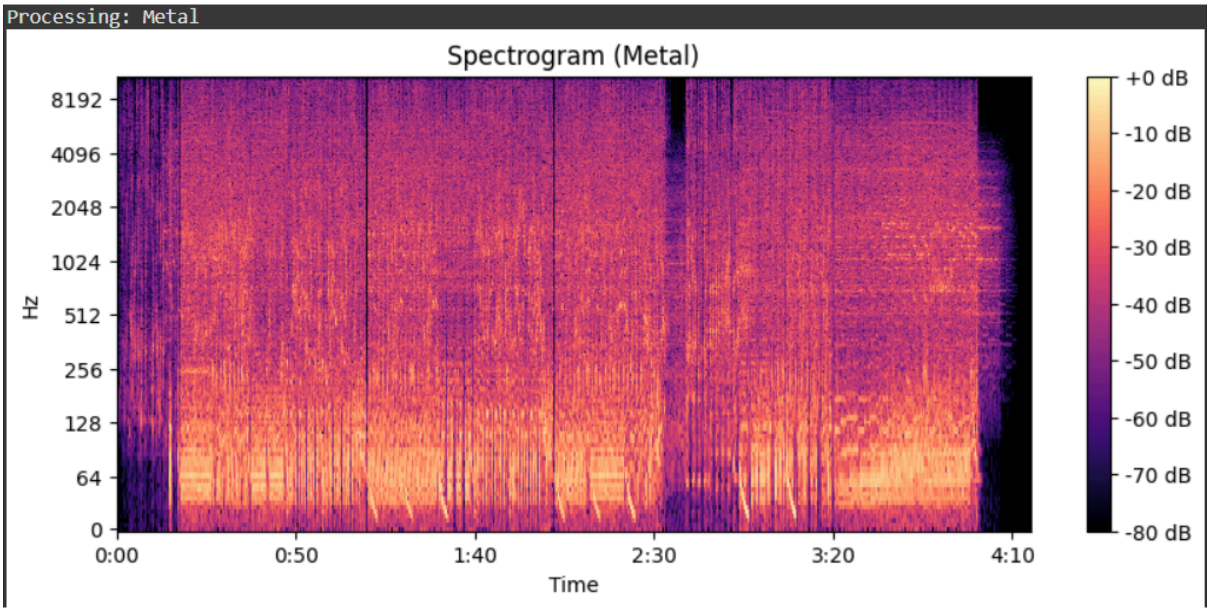
2. Methodology

The following methodology was used to generate and analyze the spectrograms:

- **Dataset:** Four songs were selected, each representing a different genre:
  - Metal: *Slaughter To Prevail - Baba Yaga*
  - Rock: *Linkin Park - Numb*
  - Pop: *Sabrina Carpenter - Espresso*
  - EDM: *Martin Garrix - Forbidden Voices*
- **Spectrogram Generation:**
  - **Short-Time Fourier Transform (STFT)** was applied to extract frequency components.
  - A **Hamming window** was used to ensure smooth spectral representation.
  - **Hop length** was set to 25% of the window size to balance time and frequency resolution.
- **Analysis Criteria:**
  - Frequency distribution across time
  - Energy concentration at different frequencies
  - Presence of transients, beats, and harmonic content

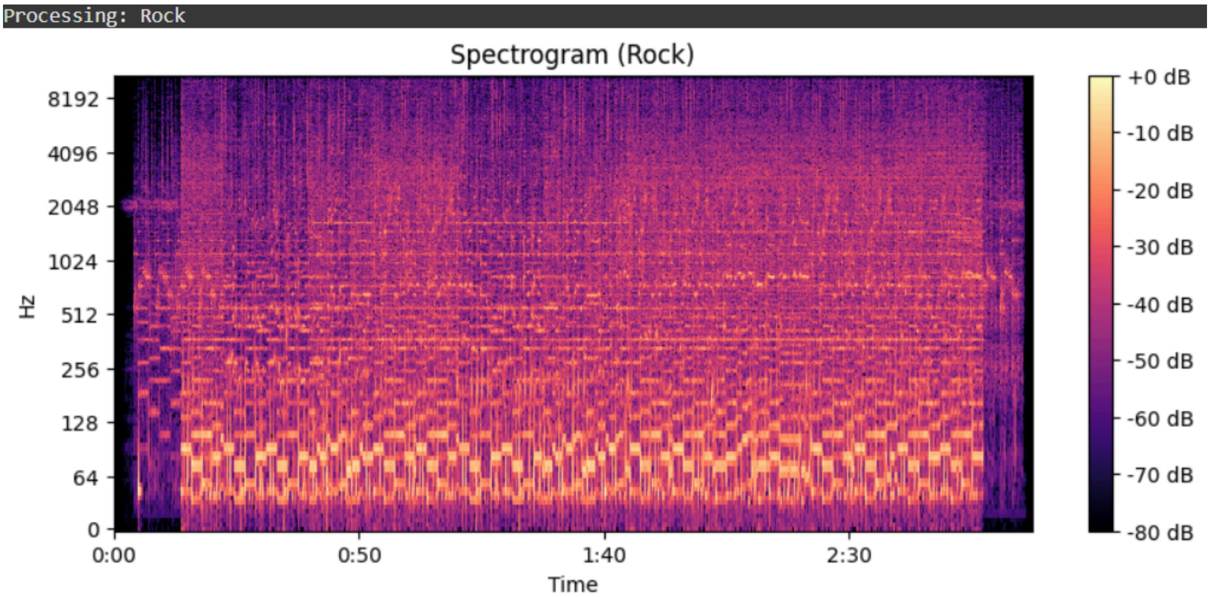
3. Spectrogram Comparisons

3.1 Metal Music Spectrogram



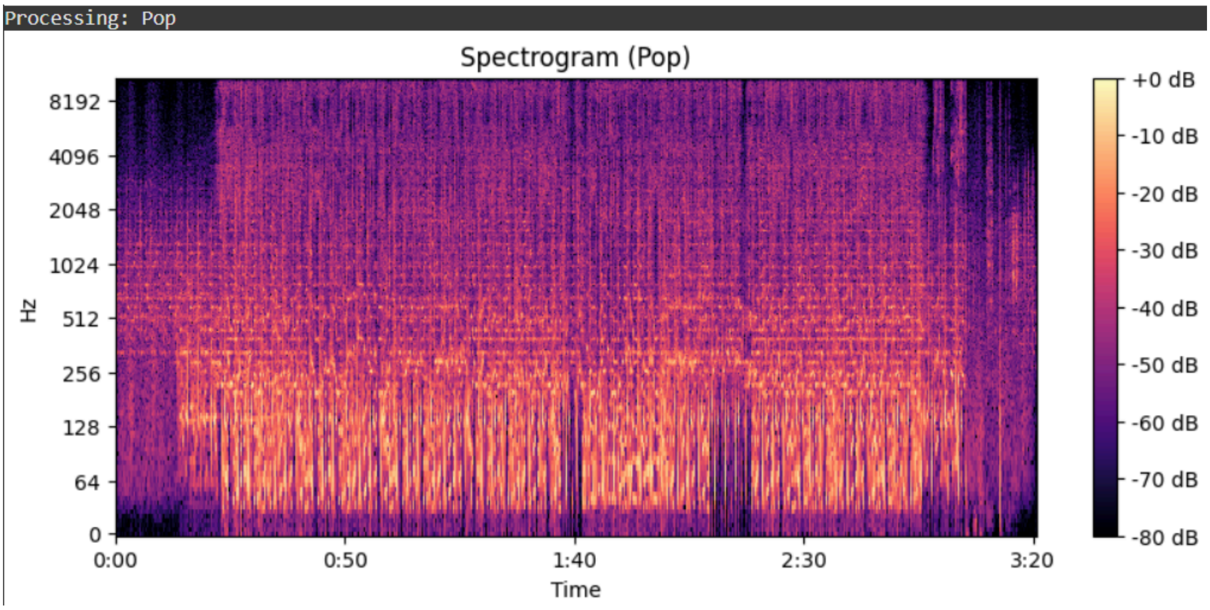
- Dense and highly energetic frequency distribution.
- Strong presence of **high-frequency components** due to distorted guitars and aggressive drumming.
- Less structured rhythm patterns compared to EDM.

3.2 Rock Music Spectrogram



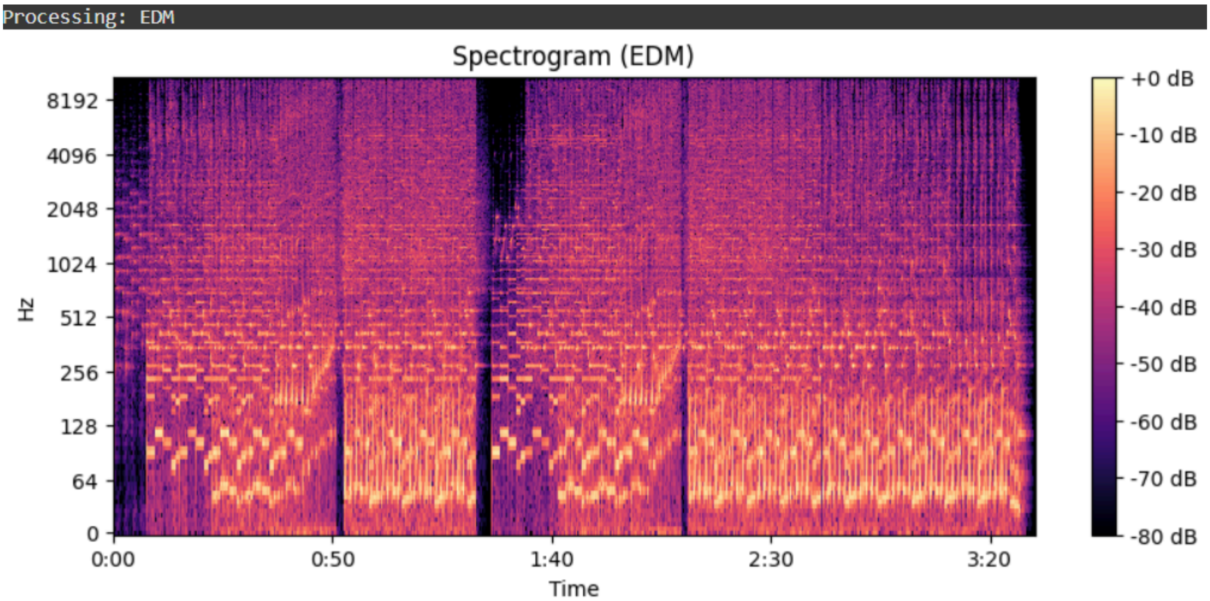
- Balanced distribution of **mid and high frequencies**.
- Vocals, drums, and guitars show clear separation.
- Less bass energy than EDM but more structure than Metal.

3.3 Pop Music Spectrogram



- **Smooth frequency transitions** with dominant mid-range frequencies.
- Clear presence of vocals and synthetic beats.
- Less intense bass than EDM but more structured rhythm.

3.4 EDM Music Spectrogram



- **Highly structured spectrogram** with repeating beats and transients.
- Strong **low-frequency energy**, indicating deep bass.
- Well-defined high-frequency transitions due to synth elements.

4. Comparative Analysis

Genre	Frequency Distribution	Energy Concentration	Harmonic Content	Structural Complexity
Metal	Broad spectrum, high frequencies dominant	High energy across all frequencies	Distorted harmonics	Less structured
Rock	Balanced mid-high frequency	Moderate energy with clear instrument separation	Natural harmonic structures	More structured than Metal
Pop	Dominant mid-range frequencies	Strong energy in vocals and beats	Well-defined harmonics	Well-structured rhythm
EDM	Strong low and high frequencies	High bass energy, sharp transients	Synthesized harmonics	Highly structured with repetitive patterns

5. Conclusion

The spectrogram analysis highlights the unique spectral characteristics of each genre:

- **Metal** has an unstructured yet intense energy spread across high frequencies.
- **Rock** maintains a balance with distinguishable instruments.
- **Pop** emphasizes clear vocals with a focus on the mid-range frequencies.
- **EDM** features a structured and repetitive pattern with prominent bass and synthesized elements.

This study demonstrates how spectrograms can reveal significant insights into the acoustic and structural properties of different music genres. Such analysis is essential in **speech and music understanding** applications, including genre classification and audio signal processing.