

Feature Extraction

To prepare the segmented audio clips for classification, each one underwent a series of transformation steps to ensure consistency and maximize the amount of meaningful information captured:

1. **Audio Standardization** Each clip was converted from stereo to mono to simplify the signal and reduce redundancy. If the segment's sampling rate differed from 22,050 Hz, it was resampled to match this standard, which balances audio quality with computational efficiency.
2. **Feature Encoding** A total of 49 audio features were extracted from each segment, capturing diverse aspects of sound such as energy, texture, rhythm, and pitch. These were saved in a labelled CSV file for training the CRNN model.

Extracted Features Overview

Category	Features Extracted
Energy & Time Domain	Zero Crossing Rate Energy Samples RMSE Samples Temporal Centroid
Spectral Features	Spectral Centroid Roll-off Spectrum Spectral Flux
Spectral Contrast	Spectral Contrast Spectrum [6]
Cepstral Features	MFCCs [13]
Chroma Features	Chroma Features [12]
Statistical Measures	Statistics Samples [5] Statistics Spectrum [5]