Model Training

To train the genre classification model, the extracted audio features were first organized into a structured dataset. This dataset included labelled segments for seven genres, each represented by a consistent set of 49 features.

The training workflow included the following steps:

1. **Data Preparation**

   - Non-essential identifiers (like segment numbers) were removed to focus purely on audio features.

   - Genres were converted into numerical labels using a label encoder.

   - The dataset was split into training and testing subsets using **stratified sampling** to maintain balanced genre representation.

2. **Feature Scaling** Standardization was applied to ensure all features contributed equally during training. This helped the model converge more effectively.

3. **Reshaping for Temporal Modeling** Features were reshaped to match the expected format of the CRNN: each sample treated as a sequence input for convolution and LSTM layers.

4. **Class Balancing and Augmentation** To address class imbalance, particularly underrepresented genres like *Afropop* and *Lekompo*, the model employed:

   - **Class weights** to influence the loss function.

   - **Data augmentation** by adding subtle Gaussian noise to existing minority samples, increasing their presence in the training set.

5. **Model Optimization**

   - The model used **CrossEntropyLoss** adjusted by class weights and was trained using the **Adam optimizer**.

   - Training occurred over **30 epochs**, with performance evaluated on the test set every five epochs.

This setup allowed the CRNN to learn both **spectral patterns and temporal dynamics**, capturing genre-specific traits while remaining resilient to data imbalance.