Project Report: Music Genre Classification Using Deep Learning

1. Project Title

   Music Genre Classification Using CNN and Audio Feature Extraction

2. Abstract

This project explores the use of deep learning for classifying musical genres based on audio input. By converting audio signals into structured, multi-channel image-like data, a convolutional neural network (CNN) is trained to distinguish between 10 music genres from the GTZAN dataset. The model is further deployed through a Flask-based web application for real-time predictions on uploaded audio files.

3. Objectives

   - To classify audio files into genres using CNN
   - To extract meaningful audio features such as MFCC, tempo, and chroma
   - To visualize misclassification and improve model performance
   - To deploy a web application for live genre prediction

4. Dataset

   GTZAN Dataset: 10 genres, each with 100 audio files (30s)
   Genres: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, rock

5. Methodology

   Audio Feature Extraction: Each audio file is converted to a 3D array using MFCC, Delta and Delta-Delta MFCC, Tempo-synced MFCC (beat-aware), and Chroma. This results in a shape of (40, 130, 5) per audio chunk.
   CNN Architecture:
   - Conv Block 1: Conv2D(5→32) → ReLU → BatchNorm → MaxPool → Dropout
   - Conv Block 2: Conv2D(32→64) → ReLU → BatchNorm → MaxPool → Dropout
   - Global Pool: AdaptiveAvgPool2D(1×1)
   - FC Layers: Linear(64→128) → ReLU → Dropout → Linear(128→10)
   - Output: LogSoftmax

6. Training Pipeline

   Optimizer: Adam with weight decay (1e-4)
   Loss Function: CrossEntropyLoss
   Metrics: Accuracy, Confusion Matrix
   Tools: torchmetrics.Accuracy, GradScaler (AMP training)

7. Deployment

   A Flask web application allows users to upload .wav files. Audio is sliced into 3-second chunks, each chunk is classified independently, and majority voting determines the final genre.

8. Result

Achieved high test accuracy on GTZAN.
Confusion matrix revealed common genre misclassifications:
- hiphop ↔ reggae
- metal ↔ rock
- disco ↔ pop

9. System Requirement

Python 3.10+
Dependencies: PyTorch, TensorFlow (optional), Librosa, Flask, Jupyter
Compatible with both CPU and CUDA-enabled GPUs

10. Project Structure

mainthing.py - Flask App
model/ - CNN model
ckpt/ - Trained model weights
templates/index.html - Upload UI
static/ - CSS/JS if any
requirements.txt
README.md
notebook.ipynb - Training and analysis

11. Final Accuracy Metrics

Train Accuracy: 86.42%
Validation Accuracy: 85.62%
Test Accuracy: 85.90%