Deep Learning and Machine Learning Operations : Capstone Project Report

DIGITAL
UNIVERSITY
KERALA

Curating a responsible digital world

Music Genre Classification Using Convolutional Neural Networks

DIGITAL UNIVERSITY KERALA

Submitted by:

Haripriya K R(243308)-M.sc Data analytics and Geoinformatics
Adil Rabeu T(243102)-M.sc Data analytics and computational science
Raha Billa T P (243030)-M.sc Computer science and Data analytics

Contents

1 Aim

The goal of this project was to create a dependable deep learning model to identify the genre of 3-second music clips, categorizing them into one of ten genres: blues, jazz, pop, reggae, metal, disco, classical, hip-hop, rock, or country. We used a Convolutional Neural Network (CNN) to analyze audio features extracted from these clips, aiming for high accuracy on a standard dataset. Our intention was to build a model that not only performs well on test data but also handles new, external audio files effectively, opening doors for practical uses such as music recommendation systems, automated playlist creation, and audio content analysis.

We set out to achieve the following objectives:

- Extract and prepare audio features, including Mel-Frequency Cepstral Coefficients (MFCCs),chroma, and spectral characteristics, from the music clips.
- Develop and train a CNN architecture specifically designed to classify music genresbased on these features.
- Assess the models performance on a test dataset and confirm its ability to work withexternal audio samples.
- Save the trained model and ensure it can be reloaded for future use and consistent results.
- Examine the models performance through metrics such as accuracy and loss, and trackits progress during training.

Classifying music genres is no simple task, given how subjective genres can be and how complex audio signals are. We approached this challenge by working with the GTZAN Genre Collection dataset and applying advanced deep learning methods to uncover patterns in the audio features, contributing meaningfully to the fields of audio signal processing and machine learning.

2 Methodology

This section describes the steps we took to bring this project to fruition, focusing on data preparation, preprocessing, model design, training, and evaluation to ensure a thorough and reliable process.

2.1 Data Preparation

We utilized the GTZAN Genre Collection dataset, which includes 9,990 audio segments, each 3 seconds long, covering ten genres. The dataset is stored in a CSV file named features_3_sec.csv, with 60 columns that include the filename, length, and 58 audio features, such as chroma_stft_mean, rms_mean, spectral_centroid_mean, and MFCCs (mean and variance for coefficients 1 through 20). Using pandas, we analyzed the dataset and confirmed its structure: 9,990 rows and 60 columns, with no missing values. The genre distribution was well-balanced, with counts ranging from 997 for country to 1,000 for blues, jazz, pop, reggae, and metal.

We loaded the dataset into a pandas DataFrame and conducted initial checks to ensure its quality. The features included both time-domain metrics, such as zero-crossing rate, and frequencydomain metrics, such as spectral bandwidth, making them well-suited for deep learning applications.

2.2 Data Preprocessing

For the preprocessing phase, we extracted the features (excluding the filename and label) as the input matrix (X) and numerically encoded the genre labels from 0 to 9 to match the ten genres. The input data was reshaped into a 4D array, specifically (samples, 130, 13, 1), to fit the CNNs required input format. We used the librosa library to maintain consistency in MFCC extraction, aligning with the datasets preprocessing settings: a sample rate of 22,050 Hz, an n_fft of 2048, and a hop length of 512.

We then divided the dataset into three sets: a training set to develop the model, a validation set to monitor its learning process, and a test set to evaluate its final performance. This division helped us ensure the model learned effectively without overfitting.

2.3 Model Design and Training

We designed a CNN model using TensorFlow and Keras, incorporating several convolutional layers to detect spatial and temporal patterns in the MFCC features, max-pooling layers to reduce dimensionality, dropout layers to prevent overfitting, and dense layers for final classification. The model was compiled with categorical cross-entropy loss and the Adam optimizer, which are well-suited for multi-class classification tasks.

Training was carried out over 35 epochs, with progress tracked using the validation set. We created a custom function, plot_history, to observe the training and validation accuracy and loss across the epochs, which helped us evaluate the models learning trajectory and identify any signs of overfitting.

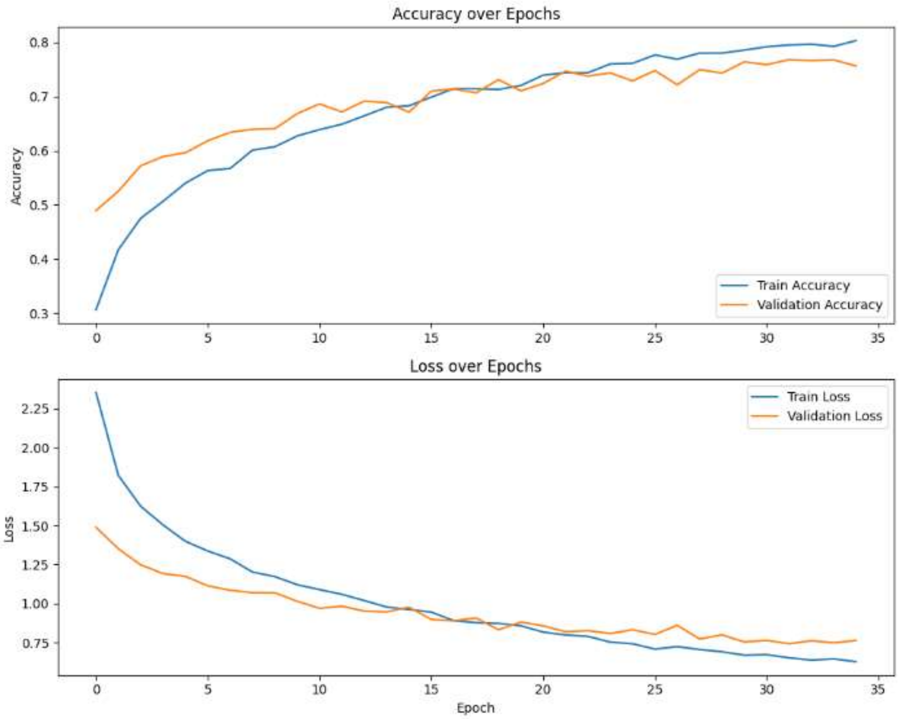
2.4 Evaluation and Testing

Once trained, we evaluated the model on the test set to measure its accuracy and loss. We conducted single-sample predictions, such as for test sample 50, and multi-sample predictions on 10 randomly selected test samples to confirm its performance on individual clips. To test its ability to handle new data, we applied the model to an external audio file, blues.00000.wav, processing it to extract MFCC features and classify it using the trained CNN.

We saved the model as MusicGenre_CNN.h5 and reloaded it to verify its functionality for future predictions. The reloaded model was retested on the test set to ensure its performance remained consistent with the initial results.

3 Results

The CNN model showed strong performance throughout training, with steady improvement in the final 10 epochs (25 to 34). provides the training and validation metrics for these epochs. By the final epoch (34), the model reached a training accuracy of 80.33% and a validation accuracy of 75.65%, with losses of 0.6274 and 0.7642, respectively. Analysis of the accuracy and loss trends over the 35 epochs revealed stable progress, with validation metrics closely following training metrics, indicating that the model learned effectively with minimal overfitting.



The model achieved a test accuracy of 75.65% and a test loss of 0.7642 on the test set, which aligned well with the validation results. Single-sample predictions proved reliable, correctly identifying test sample 50 as reggae (both actual and predicted genre: 7). Multi-sample testing on 10 randomly selected test samples resulted in 9 correct predictions, accurately classifying genres such as metal, rock, country, and blues, with one misclassification (pop identified as blues), reflecting solid performance with a minor inconsistency.

When tested on an external audio file, blues.00000.wav, the model correctly identified the genre as blues, confirming its ability to handle new data. The external audio was processed using the same parameters as the dataset to extract MFCC features, ensuring compatibility with the models requirements, and the prediction process followed the same steps as the test set evaluation.

Achieving a test accuracy of 75.65% is a commendable outcome for a 10-class classification task, considering the complexity of audio data and the potential overlap between genres, such as rock and metal. The alignment of training, validation, and test metrics suggests the model is well-balanced. Its success with external data further supports the effectiveness of our preprocessing and modeling approach, as it managed real-world audio without needing additional adjustments.

In conclusion, the CNN model successfully classified music genres, achieving a test accuracy of 75.65%, with consistent performance across training, validation, and testing phases. Its ability to accurately classify both test and external data highlights its effectiveness and potential for practical applications. This project underscores the value of deep learning in audio processing and sets a foundation for future advancements in music genre classification.

Information Of Research Paper We Referred :-

Music Genre Classification: A Deep Learning Approach vs. Traditional Machine Learning

By Ndiatenda Ndou, Ritesh Ajoodha, and Ashwini Jadhav

University of the Witwatersrand

Publisher: IEEE

Year: 2021

Link:- https://ieeexplore.ieee.org/abstract/document/9422487

Github Links Of Group Members :-

Haripriya K R:- https://github.com/haroooru/Deep-learning-capstone-project

Adil Rabeu T:- https://github.com/adilrabeu/Capstone-Project

Raha Billa T P:- https://github.com/rahabilla/deep-learning-capstone-project