MUSIC GENRE CLASSIFICATION

J Krishna Sai Roshith[1] , P. Siddharth[2] , P. Hitesh Sai[3], D. Harshith Naidu[4]

R.V.S Nagar, Moinabad, Road, near AP Police Academy, Aziznagar, Hyderabad, Telangana 500075

Abstract:

The Python program for music genre classification utilizes machine learning techniques to categorize music tracks into distinct genres. Leveraging libraries like TensorFlow and scikit-learn, it first preprocesses audio data, extracting features such as spectral characteristics, tempo, and rhythm. These features serve as inputs to a classification model, typically a neural network or a traditional machine learning algorithm. The program begins by collecting a diverse dataset comprising audio samples from various genres like rock, jazz, pop, classical, etc. Each audio sample undergoes feature extraction, transforming the raw audio signals into a format conducive to machine learning analysis. Feature selection is crucial, as it determines the model's ability to discern genre-specific patterns effectively. After preprocessing, the program trains the classification model using labeled data, employing techniques like cross-validation to ensure robustness. Once trained, the model is tested on unseen data to evaluate its accuracy and generalization capabilities. Hyperparameter tuning and model optimization may be performed to enhance classification performance further. The final program allows users to input audio files, which are then processed and classified into their respective genres. With its modular structure and extensive documentation, this Python program serves as a versatile tool for music genre classification, aiding researchers, music enthusiasts, and industry professionals in analyzing and organizing vast collections of music.

Keywords:

Feature Extraction, Data Preprocessing, Model Selection, Training and Evaluation, Hyperparameter Tuning, Performance Metrics.

1. INTRODUCTION

The art of music is the ability to convey emotions through vocal, instrumental, or even a combination of the two. There are many different types of music, from jazz and classical to pop, rock, and hip-hop. With the widespread usage of digital music in modern technology, the demand for automatically categorizing music depending on its genre has also increased because doing so enables more specialized music suggestions, supports music search engines, and creates playlists for users.

The goal of a music genre classification application is to automatically assign a certain piece of music to the appropriate genre. A typical person would listen to a music file the entire time if they attempted to categorize it according to its specific genre. This program would be incredibly helpful because it decreases the amount of time required and human error a person would encounter when classifying independently. Therefore, machine learning and deep learning algorithms are required for implementation in order to automatically classify. Since audio data contains information in both the time and frequency domains, as opposed to image data, which typically consists of two- dimensional arrays of pixels with information encoded in the intensity and color values, audio processing is more complex than image processing and other classifications.

Deep learning methods for classifying music genres, like convolutional neural networks (CNNs) and recurrent neural networks, have produced encouraging results. CNNs are a particular class of neural network that excel at processing image data, while they may also be used to handle audio data by treating audio waves like images. and RNNs can be useful for tasks like classifying the musical genre since they can identify temporal connections and patterns in audio data.

The issue of music genre classification using ML and DL algorithms will be explored in this term paper. The paper will begin with an introduction to the fundamentals of ML and DL and how they apply to processing audio data. The following section will go over the many methods used to categorize music genres, including feature extraction techniques, model structures, and evaluation metrics.

Following that, the paper will present a review of recent research in music genre classification using CNNs and RNNs, emphasizing the performance of various models and the factors that influence their performance. Furthermore, the paper will compare the performance of CNN-based models to RNN-based models in music genre classification tasks, as well as ML algorithms such as SVM and XGBoost.

The paper will wrap up with a summary of the main findings and how they may affect future attempts to classify music genres using ML and DL algorithms.

2. DATASET DESCRIPTION

The most popular publicly available dataset for evaluation in machine listening research for music genre identification (MGR) is the GTZAN dataset. In order to represent a range of recording circumstances, the files were gathered in 2000 and 2001 from a number of sources, including personal CDs, radio, and microphone recordings.

- Context

For a very long time, experts have been attempting to comprehend sound and what makes one song different from another. ways to picture sound. what distinguishes one tone from another.

- Content

An illustration of every audio file. Data classification using neural networks is one method. The audio files were changed into Mel Spectrograms in order to make this possible because NNs (like CNN, which is what we will be utilizing today) typically take in some form of picture representation.

CSV files: These files include the audio files' features. A mean and variance computed over numerous features that can be recovered from an audio stream are contained in one file, one for each song (30 seconds long). The other file is identical structurally, but the songs were previously divided into 3 second audio files, increasing the amount of data we feed into our classification algorithms by a factor of 10. More data is always better in this case.

3. Methodologies

To categorize music into its various genres, genre classification is utilized. We employ identical approaches to implement the project in both deep learning and machine learning. We must extract the audio features from the provided data sets (GTAZAN Dataset) in order to classify the musical genres.

3.1 ML Methodologies

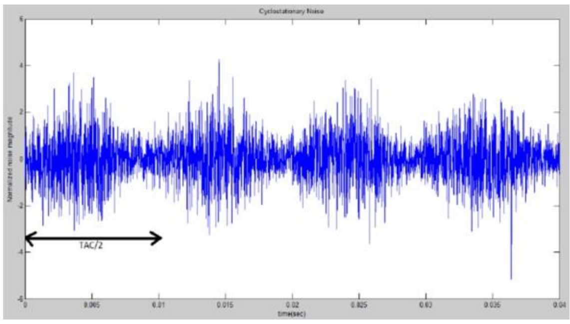
3.1.1 Feature Extraction

We noticed that the data in the GTAZAN dataset is presented as wav files. Therefore, in order to use machine learning and deep learning models on it, we must transform that raw data into numerical features.

3.1.2 Wave Form

In order to increase the frequency of the wavefronts over time, each musical genre is turned into a waveform. An accessible interface to the audio WAV format is the wave module found in the Python standard library. The functions in this module can read the properties from a WAV file and write audio data in raw format to a file-like object.

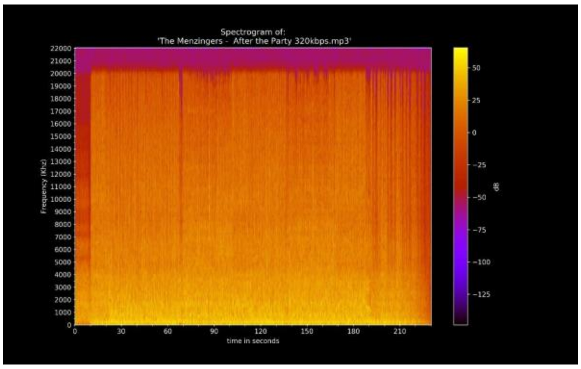
A spectrogram is a visual depiction of the spectrum of frequencies of a signal as it evolves with time in audio signal processing. The signal's amplitude or power at a specific time and frequency is depicted in a two- dimensional plot with time on the x-axis and frequency on the y-axis.



3.1.3 Spectogram

Spectrograms can be made and modified in Python using the matplotlib and NumPy packages. An audio source can be utilized to create a spectrogram using the matplotlib.pyplot.specgram() function. The audio signal, sampling rate, and additional optional parameters like window size, hop length, and frequency range are all inputs for this function.

For a visual representation of signal intensity or loudness over time at different frequencies contained in a specific waveform, spectrogram conversion is required.

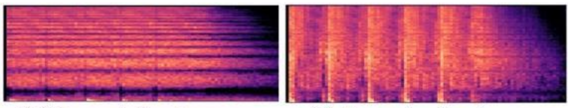


3.1.4 Mel Spectrogram

A Mel spectrogram is a spectrogram in which the frequency axis is converted to the mel scale, a scale of pitches based on perceptual principles. Since it more accurately captures how people hear pitch differences, this scale is better suited for human hearing perception. We can see audio files and the pressure that sound waves produce using mel spectrograms, and this enables us to see the shape and form of the recorded music.

3.1.5 Harmonic and percussive

To separate an audio signal into two components, one harmonic and the other percussive, is the aim of harmonic and percussive separation. Percussive sound is what we hear as a clash, knock, clap, or click, whereas harmonic sound is what we hear as a pitched sound.



(b) Harmonic Spectrogram (c) Percussive Spectrogram

3.1.6 Chromogram

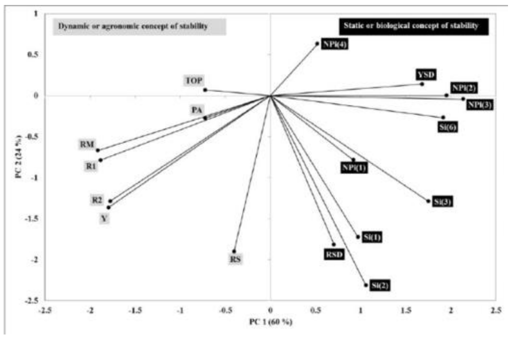
A chromogram is used to deduce characteristics about how a signal's energy, chroma, and time are distributed. A Python object for mass spectrometry data is provided by the spectrum class. The spectrum object provides ways to query the spectrum's attributes as well as basic information about the spectrum. Mass over charge (m/z) and intensity decoding are examples of data that are performed on demand and accessible through their attributes. The chromatogram class permits inquiry with profile data (time, intensity) in a total ion chromatogram, much like the spectrum class does.

3.1.7 Principal Component Analysis

The creation of principal components follows the degree of variance in the covered genera. Each of the principal component analyses listed below provides us with a value for the principal component analysis and some information about the data.

3.1.8 PCA 1 and PCA2nd PCA2

Principal component analysis 1 is used to capture the most variation. Principal component



analysis 2 is used to capture the second most variation and so on.

3.1.9 Models used for Machine Learning

- Support Vector Machine
- Naïve Bayes Classifier
- Logistic Regression
- K-Nearest Neighbors
- MLP
- Extreme Gradient Boost

1. The supervised learning algorithm known as Support Vector Machines (SVMs) can be applied to classification or regression applications. Finding a hyperplane that maximally separates the various classes in the training data is the basic goal of SVMs.

2. The Nave Bayes algorithm is a supervised learning method for classification issues that is based on the Bayes theorem. It is mostly employed in text categorization with a large training set.

3. Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

4. K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

5. Multilayer perceptron (MLP) is a technique of feed-forward artificial neural networks using a back propagation learning method to classify the target variable used for supervised learning. MLP's can be applied to complex non- linear problems, and it also works well with large input data with a relatively faster performance. The algorithm tends to achieve the same accuracy ratio even with smaller data.

6. Extreme Gradient Boost or XGBoost is an implementation of Gradient Boosted decision trees. This library was written in C++. It is a type of Software library that was designed basically to improve speed and model performance. It has recently been dominating in applied machine learning. XGBoost models majorly dominate in many Kaggle Competitions. In this algorithm, decision trees are created in sequential form. Weights play an important role in XGBoost.

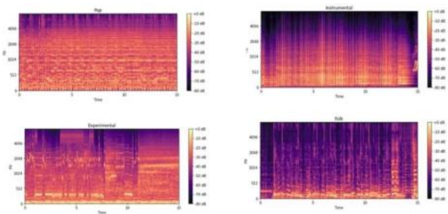
3.2 Deep Learning Methodologies

3.2.1 Feature extraction in DL

Data input for deep learning models must be in a set format. Extraction of pertinent information from the audio files is therefore crucial. Due to their ability to capture the spectrum properties of the audio signal, MFCCs are frequently utilised in tasks involving the classification of music genres. To extract these features, utilise libraries like librosa or Essentia.

3.2.2 Pre-processing

To apply deep learning algorithms to an audio file, we need features like Mel- Frequency Cepstral Coefficients (MFCCs), Spectral Centroid, Spectral Flux etc.



3.2.3 Models used in DL

- VGG 5
- VGG 16
- VGG 19
- LSTM
- ResNet
- LeNet

I. Visual Geometry Group (VGG) is a very popular deep learning technique that helps in image processing. The VGG contains small convolutional filters (3x3) with a deep network structure (16-19 layers) architecture. To use VGG for music genre classification, the input data should be pre- processed into a format suitable for use in the network. Typically, this involves representing the audio signal as a spectrogram, which is a visual representation of the frequency content of the signal over time. The spectrogram can then be transformed into a 2D image-like format that can be fed into the VGG network.

II. Recurrent Neural Networks (RNNs) are a class of neural networks that are well-suited for modelling sequential data, such as audio signals. RNNs have been used for music genre classification tasks by modelling the temporal dependencies in the input data. To use RNNs for music genre classification, the input data should be preprocessed into a sequence of feature vectors. Typically, this in- volves representing the audio signal as a spectrogram, and then segmenting the spectrogram into overlapping frames. Each frame is represented as a feature vector, which can be fed into the RNN as a sequence. RNNs like Long Short Term Memory(LSTM), ResNet, LeNet are used.

4. CONCLUSION:

Music genre classification is a difficult task that can be addressed with machine learning and deep learning techniques. On this task, both algorithms have advantages and disadvantages. SVM, Logistic Regression, and XGBoost are examples of machine learning algorithms that have achieved high levels of accuracy. Deep learning algorithms such as VGG16 and LSTM have demonstrated great accuracy. As a result, deep learning and machine learning both have solid algorithms, and their performance is entirely dependent on feature selection and other factors in the dataset.In essence, the expense tracking system serves as a practical tool for users to monitor their financial activities and improve budgeting strategies. Its intuitive interface and database-driven backend enable seamless recording and manipulation of expense data. While the system may not incorporate sophisticated machine learning algorithms, its reliance on basic data structures and algorithms ensures efficient management and analysis of expenses, empowering users to gain insights into their spending habits and make informed financial decisions. With further enhancements, such as integration with predictive analytics or advanced reporting capabilities, the system holds the potential to provide even greater value in helping users achieve their financial goals.

5. References

5.1 Sami Abu-El-Haija, Rohan Kothari, Joonseok Lee, Apostol Natsev, George Toderici, Balakrishnan Varadarajan, and Sudheendra Vijayanarasimhan. The youtube-8m dataset. arXiv preprint arXiv:1909.08676, 2019.

5.2 Shashank Anand and Shashidhar G Chakravarthy. A comparative study of music genre classification using machine learning algorithms. Interna- tional Journal of Innovative Technology and Exploring Engineering (IJI- TEE), 7:36–40, 2018.

5.3 Nicolas Boulanger-Lewandowski, Yoshua Bengio, and Pascal Vincent. Au- dio chord recognition with convolutional neural networks. Neural Networks,44:72–82, 2013.

5.4 Yuan Cao and Ronghua Chen. Music genre classification based on deep learning. Journal of Physics: Conference Series, 1634(1):012057, 2020.

5.5 Keunwoo Choi, Gyorgy Fazekas, and Kyunghyun Cho. Convolutional re- current neural networks for music classification. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(6):1216–1227, 2017.

5.6 Fran¸cois Chollet. Deep learning with Python. Manning Publications, 2017.

5.7 Yiheng Deng, Qiong Wu, Xiaomin Lian, and Guojun Zhang. Music genre classification based on hybrid features and convolutional neural network. In 2020 2nd International

Conference on Computer Communication and Informatics (ICCCI), pages 333–337. IEEE, 2020.