INDIAN INSTITUTE OF TECHNOLOGY
COMPUTER SCIENCE AND ENGINEERING

CS 354N Computational Intelligence

Project Report

Music Genre Classification

Advisor: Aruna Tiwari

IIT INDORE, MAY 2023

Contents

Member list

No.	Full name	Student ID
1	Mir Razee Mohideen	200001045
2	Nishit Singh	200001056

1 Problem definition

Music genre classification is the task of automatically categorizing a piece of music into one or more predefined genres, such as rock, pop, jazz, or classical. It involves using machine learning algorithms to analyze the acoustic features of a song, such as its rhythm, melody, harmony, and timbre, and using those features to make predictions about which genres the song belongs to.

To build a music genre classification system, one typically needs a large dataset of labeled music samples, where each sample is labeled with its corresponding genre(s). The dataset is then divided into training, validation, and test sets, and machine learning models are trained on the training set using various algorithms,

During training, the model learns to recognize the patterns in the acoustic features that are most indicative of each genre, and adjusts its internal parameters to improve its predictions on the validation set. Once the model is trained, it can be tested on the test set to evaluate its accuracy and generalization performance.

Some of the challenges of music genre classification include dealing with the subjective and culturally dependent nature of genre labels, handling the high-dimensional and noisy nature of audio data, and addressing the issue of multi-label classification, where a song may belong to multiple genres at the same time.

Overall, music genre classification is a challenging and active research area that has many practical applications in music recommendation, playlist generation, music search, and music analysis.

2 Study of Algorithm

Our approach to solve this problem is by using convolutional neural networks (CNN) to extract features from audio data, and then feed those features into a fully connected neural network (FCNN) to make the final classification. The audio data is usually preprocessed by converting it into a spectrogram, which is a visual representation of the frequency components of the audio signal. The CNN is trained on a dataset of audio samples labeled with their respective genres, and the weights of the network are adjusted during training so that it can accurately classify new audio samples based on their features. Once the model is trained, it can be used to predict the genre of an unseen audio sample.

3 Data Visualisation and Preprocessing

We did the following in Data Preprocessing and Visualization:

- GTZAN dataset was first checked for corrupted audio files Only one corrupted audio file was found in the dataset and it was removed.
- The features dataframe was checked for NA values but none were found.
- Plotted Spectral Roll-off, Amplitude, Linear-frequency power spectrogram to get brief idea of the dataset.

4 Performance measurement and evaluation of model

We use the Sparse Categorical CrossEntropy function as the loss function for our CNN, since it is commonly used in multi-class classification problems. Here, the labels are integers rather than one-hot encoded vectors. In other words, it is used when the target labels are not binary, but instead, there are multiple classes, and each instance belongs to one of these classes. The formula for the sparse categorical crossentropy loss function is:

$$L = -\sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij} \log(p_{ij})$$

There are various other methods to test and evaluate the algorithm. Some of them are as follows;

- Accuracy
- Receiver Operating Characteristic (ROC)
- Precision
- Recall
- F1 score

5 Our algorithm

Our approach to solve this problem is by using convolutional neural networks (CNN) to extract features from audio data, and then feed those features into a fully connected neural network (FCNN) to make the final classification. The audio data is usually preprocessed by converting it into a spectrogram, which is a visual representation of the frequency components of the audio signal. The CNN is trained on a dataset of audio samples labeled with their respective genres, and the weights of the network are adjusted during training so that it can accurately classify new audio samples based on their features.

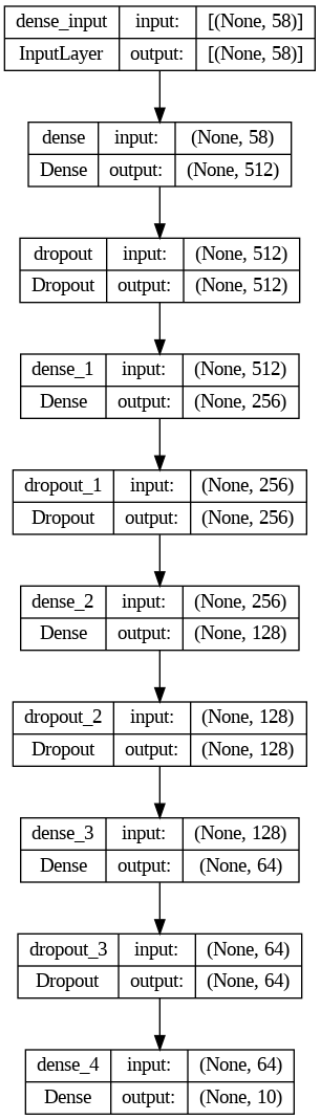
6 Comparison with other models

We have compared our model with three baseline models:

- Linear Regression: It is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. The goal of linear regression is to find the best fit line that minimizes the distance between the predicted values and the actual values of the dependent variable.
- DecisionTreeClassifier: It is a classification algorithm that builds a tree-like model by recursively splitting the data based on the values of input features. The algorithm selects the best feature to split the data based on some criterion, such as information gain or Gini impurity. The tree is built until a stopping criterion is met, and each leaf node is assigned a class label based on the majority class of the samples in that node.

- KNeighborsClassifier: It is a classification algorithm that assigns a label to a data point based on the class labels of its nearest neighbors in the training dataset. The algorithm works by calculating the distance between the new data point and each point in the training dataset, and then selecting the k-nearest neighbors. The class label assigned to the new data point is then determined by a majority vote of the labels of the k-nearest neighbors.

For the proposed CNN model we have used 5 dense layers with decreasing number of neurons with each layer having ReLU activation. We have added dropout layers to prevent overfitting. The final layer is the softmax activation layer for final prediction.



7 Conclusion

Our CNN model outperformed the baselines models by a significant amount

Model results	
Model	Accuracy
Linear Regression	26%
DecisionTreeClassifier	64%
KMeansClassifier	85.5%
CNN	91.3%

Below is the confusion Matrix obtained for our Model

References

[1] GTZAN Dataset. [Online]. Available: https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification

[2] Cross Entropy Loss Function. [Online]. Available: https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e

[3] S. Prince, J. J. Thomas, S. J. J, K. P. Priya, and J. J. Daniel, "Music genre classification using deep learning - a review," in *2022 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, 2022, pp. 1–5.

[4] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.