Synopsis on

Audio Processing using Pattern recognition for Music Genre Classification

Department of Information Technology

By
Avik Bose (Roll No. 15 & University Roll No. 12200218052)
Hrithik Raj Prasad (Roll No. 45 & University Roll No. 12200218040)
Sivangi Chatterjee (Roll No. 19 & University Roll No.12200218016)
Srishti Ganguly (Roll No. 30 & University Roll No. 12200218010)

Under the guidance of
Dr. Arijit Ghosal
&
Dr. Ranjit Ghoshal

St. Thomas’ College of Engineering and Technology

Affiliated to

Maulana Abul Kalam Azad University of Technology, West Bengal

We are submitting the synopsis on Audio Processing using Pattern recognition for Music Genre Classification as a part of our final year seventh semester project under the guidance of Dr. Arijit Ghosal & Dr. Ranjit Ghoshal.

Avik Bose

Hrithik Raj Prasad

Sivangi Chatterjee

Srishti Ganguly

Dr. Arijit Ghosal
Department of Information Technology

Dr. Ranjit Ghoshal
Department of Information Technology

Vision:

To promote the advancement of learning in Information Technology through research-oriented dissemination of knowledge which will lead to innovative applications of information in industry and society.

Mission:

- To incubate students, grow into industry ready professionals, proficient research scholars and enterprising entrepreneurs.
- To create a learner-centric environment that motivates the students in adopting emerging technologies of the rapidly changing information society.
- To promote social, environmental and technological responsiveness among the members of the faculty and students.

PEO:

PEO1: Exhibit the skills and knowledge required to design, develop and implement IT solutions for real life problems.

PEO2: Excel in professional career, higher education and research.

PEO3: Demonstrate professionalism, entrepreneurship, ethical behavior, communication skills and collaborative team work to adapt the emerging trends by engaging in lifelong learning.

Project Mapping with Program Outcomes

PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
3	3	2	3	3	3	-	3	2	3	3	3

Enter correlation levels 1, 2 or 3 as defined below:

1: Slight (Low)   2: Moderate (Medium)   3: Substantial (High)

Justification:

1. Engineering knowledge: Application of the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems is required in this project, hence, it satisfies PO1.
2. Problem analysis: Identifying, formulating, reviewing research literature, and analyzing complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences is highly required in this project, hence, it satisfies PO2.
3. Design/development of solutions: Designing solutions for complex engineering problems and designing system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations is required to be done in this project, thus, PO3 stands applicable.
4. Conduct investigations of complex problems: Using research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions is required to a great extent. PO4 stands applicable.
5. Modern tool usage: Appropriate modern tools and methods are used. Hence PO5 stands applicable.
6. The engineer and society: Application of reasoning informed by the contextual knowledge to assess cultural issues and the consequent responsibilities relevant to the professional engineering practice is recognized in this project, hence, PO6 stands applicable.
7. Environment and sustainability: This project's impact is more cultural and technology based, hence, PO7 is not applicable.
8. Ethics: Applying ethical principles and commitment to professional ethics and responsibilities and norms of the engineering practice is required. PO8 is applicable.
9. Individual and team work: Functioning effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings: this is applicable here.
10. Communication: Communicating effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, making effective presentations, and giving and receiving clear instructions: PO10 is substantially applicable in this project.
11. Project management and finance: Demonstrating knowledge and understanding of the engineering and management principles and applying these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments: this is highly applicable in this project, hence PO11 has high correlation level (3)

12. Life-long learning: Recognizing the need for, and having the preparation and ability to engage in independent and life-long learning in the broadest context of technological change: this is relevant with respect to this project, hence, PO12 has a high correlation level of 3.

PSO: Project Mapping with Program Specific outcomes

PSO1	PSO2	PSO3
3	-	3

Enter correlation levels 1, 2 or 3 are as defined below:

1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)

Justification:

1. This project requires me to apply programming knowledge to build an efficient and effective solution of the problem with an error free, well documented and reusable code, user friendly interface and well-organized database. Hence, the project substantially satisfies PSO1.
2. Creation of multimedia enabled web solutions using information in different forms for business, education and the society at large isn't applicable in this project, hence, PSO2 stands irrelevant/ not applicable.
3. Understanding and analyzing a big complex problem and decomposing it into relatively smaller and independent modules algorithmically is done in this project, hence, PSO3 is satisfied substantially.

Index Page

Contents **Page No.**

1. Introduction:

1.1 Problem Statement:

Music Genre Classification using Machine Learning.

1.2 Problem Definition:

With the different classification techniques , we aim to identify the different genres of music based on their characteristic features.

1.3 Objective:

The aim of this project is to build a model which correctly recognizes the genre of the music based on their characteristic audio features using Machine Learning. The project intends to build an interface which can predict the genre of a given audio file using a given model with acceptable accuracy.

1.4 Literature Survey:

1. Vishnupriya S, and K.Meenakshi, "Automatic Music Genre Classification using Convolution Neural Network", IEEE Conference 2018.

2. Matthew Creme, Charles Burlin, Raphael Lenain, "Music Genre Classification ", Stanford University, December 15, 2016.

3. Arsh Chowdhry, "Music Genre Classification using CNN",May 7, 2021.

4.Gautam Chettiar ,Kalaivani S, " Music Genre Classification Techniques" Conference paper, Vellore Institute of Technology, 11 November, 2021.

5.Parul Pandey, "Music Genre Classification with Python", December 13, 2018.

6. Introduction to Machine Learning by Alex Smola and S.V.N. Vishwanathan.

1.5 Brief Discussion on Problem:

The objective of this project is Audio Processing for Music Genre Classification .The word 'genre' as used in other forms of art, like literature, cinema ,etc combines pieces of work belonging to similar backgrounds , highlighting the distinctive properties among them. A music genre, too identifies and categorizes music as belonging to different styles, grouping the ones with similarities in background or style. In other words, a music genre is the art of performing instrumental or vocal tunes in a manner that gives the music distinctive properties. For example - Rock music is characterized by electric guitars backed by electric bass and percussion from a drum kit while country music is often based around acoustic guitars and bass. Similarly, disco music is largely high temp and dance beat. Artistic compositions belonging to the same music genre share similarities in form or style.

With the start of the 21st century, the pool of music tends to be ever-expanding .The music industry today is a cornucopia of different artists , songs , genres and playlists and people nowadays tend to bid for a total genre of music rather than a single artist or band, since genres have the power to boost our discernment of music and it also enhances our enjoyment of music . As of 2021, there exists 41 primary genres of music and within those primary categories, 337 sub categories of music. Music is an ever growing pool. People tend to listen to the older ones as well as the newer trends in the music industry. Thus sorting music into separate sub-categories of different genres, of different timeframes manually would be an impossible task at the present time.

Companies nowadays use music classification either to be able to place perfect recommendations aligned to their customers' preferences, like Spotify or SoundCloud or simply as a product like Shazam. Therefore determining which music falls under which category or genre is the first step of advancement to this. This project aims to identify and distinguish among music of five different genres - Country, Rock, Reggae, Hip-Hop and Disco. We hope to achieve this with the help of Machine learning techniques and classification algorithms in order to understand the nuances of patterns or trends of similar genres of music, and thus sort them categorically.

2. Concepts and Problem Analysis :

This project is an application of Machine Learning. Machine Learning is the study of algorithms that gives computers the capability to learn and adapt on its own from previous experiences without being explicitly programmed and draw inferences from patterns in data. A general machine learning block diagram is illustrated in Fig 1. A typical machine learning problem starts with data acquisition from various sources followed by data pre-processing where raw data is filtered to improve its quality. The processed data is then used for extracting features and the dataset is split in two groups - one for training and the other for model evaluation. After the selection of a suitable machine learning model depending on the data, the model is trained using the training data set. This is followed by the training model being tested against the evaluation data set for model evaluation which involves calculating the accuracy and precision of the trained model. This evaluation can be further used to fine tune the model until a desired accuracy is achieved that can be deployed.

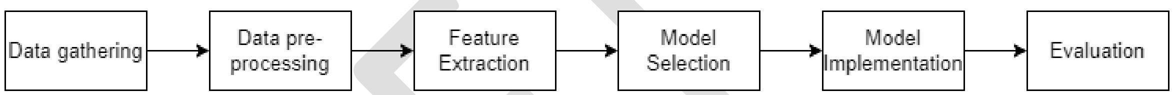

Fig. 1 : Machine Learning Block Diagram

2.1. Dataset :

The dataset used in this project is the GTZAN Genre Collection Dataset. It has been designed and written by George Tzanetaki. The dataset has been taken from the popular software framework MARSYAS[1]. Marsyas (Music Analysis, Retrieval and Synthesis for Audio Signals) is an open source software framework for audio processing with specific emphasis on Music Information Retrieval applications.

Dataset consists of 1000 audio tracks each 30 seconds long. It contains 10 genres (Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae and Rock), each represented by 100 tracks. The tracks are all 22050 Hz Mono 16-bit audio files in .wav format. We will only be working on five genres i.e. Country , Disco , Hip-Hop , Reggae and Rock.

2.2. Feature Extraction :

Before any audio signal can be classified into different classes(in this case different genres), the features in that audio signal should be extracted such that it helps us to classify accordingly. Feature Extraction involves the analysis of the input of the audio signals. A typical sound signal can be expressed as a function of amplitude and time. Audio processing generally involves extracting acoustic features germane to the task followed by classification. Due to complexity of human audio perception, no feature has yet been designed having 100 % distinguishing power between classes. That is why a combination of features can be used to achieve high classification accuracy. We hope to use Python libraries to achieve this .

2.2.1. Features Proposed :

Every audio signal has various features. Therefore it is imperative to select features which will make the classification process efficient. Hence, we have selected features which we think are pertinent to our project and will help us achieve accurate results.

Zero Crossing Rate :

The Zero Crossing Rate(ZCR) depicts the concentration of energy in the spectrum. It is defined as the rate at which the signal crosses zero within the frame i.e. it is the number of times the signal changes value, from positive to negative and vice versa within the frame. The ZCR values are in the range of 0-1. Low ZCR value corresponds to low frequency and vice versa . ZCR can be used to distinguish between voiced and unvoiced speech since unvoiced speech has a higher frequency than voiced speech. ZCR has higher values for rock music and that is why we are considering this feature for our project.

Spectral Centroid :

The Spectral Centroid(SC) is the centre of gravity of the spectrum. It is used to classify speech and noise and the same feature is also used to distinguish among various music genres. It is the weighted mean of frequencies present in the sound. A higher value of SC corresponds to brighter sound. So we hope that it would be useful in distinguishing between genres with high energy such as rock and other genres with relatively low energy.

Spectral Roll-off :

It measures the skewness of the spectral shape. It represents the frequency below which 85% of the spectral distribution is concentrated. This can help distinguish between rock music where the specified percentage of energy will correspond to a higher frequency to that of country or reggae music where the spectral energy is concentrated at lower frequencies.

Mel Frequency Cepstral Coefficients :

The envelope of the time power spectrum of the speech signal is representative of the vocal tract, which determines any sound that comes out of the human mouth based on its physical structure and MFCCs accurately represents this envelope. Now, MFCCs are the Mel frequency cepstral coefficients of a signal that are a small set of features (usually about 10–20). The MFCCs encode the timbral properties of the music signal by encoding the rough shape of the log-power spectrum on the Mel frequency scale. This will help us in distinguishing between the genres.

Chroma Shift :

Chroma features are closely related to 12 pitch classes. They can identify pitches that vary by an octave and hence are very much useful in establishing variations in timbre and correlate to the musical aspect of harmony. They are a very good tool in processing and analyzing music data. Hence, this feature is helpful in genre classification.

2.3. Classification :

In machine learning, classification refers to a predictive modeling problem where a class label is predicted for a given example of input data. A model will use the training dataset and will calculate how to best map examples of input data to specific class labels. As such the training dataset must sufficiently be representative of the problem and have as many examples of each class label.

Types of classification techniques:

1. K Neighbors Classification
2. Logistic Regression
3. Random Forest
4. Decision Tree
5. Support Vector Machines

2.3.1. K Neighbors Classifier

The k-nearest neighbors (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. It is easy to implement and understand. It is a versatile algorithm which is also used for imputing missing values and resampling datasets.

2.3.1.1 Algorithm

Step-1: Load the data

Step-2: Initialize 'K' to your chosen number of neighbors

Step-3: For each example in the data,

a. Calculate the distance between the query example and the current example from the data.

b. Add the distance and the index of the example to an ordered collection.

Step-4: Sort the ordered collection of distances and indices in ascending order by the distances.

Step-5: Pick the first K entries from the sorted collection.

Step-6: Get the labels of the selected K entries.

Step-7: Return the mode of the K labels.

2.3.2. Logistic Regression

Logistic Regression is a 'Statistical Learning' technique categorized in 'Supervised' Machine Learning (ML) methods dedicated to 'Classification' tasks. It has gained a tremendous reputation for the last two decades especially in the financial sector due to its prominent ability of detecting defaulters.

Logistic regression is named for the function used at the core of the method, the 'logistic function'. In order to map predicted values to probabilities, we use the Sigmoid function. The function maps any real value into another value between 0 and 1. In machine learning, we use sigmoids to map predictions to probabilities. The sigmoid curve can be represented with the help of the following graph. We can see the values of the y-axis lie between 0 and 1 which crosses the x-axis at 0.5.

2.3.2.1 Brief Discussion

1. The following equation is used to presents Sigmoid function:

   $g(z)= 1/1+e^{-z}$.

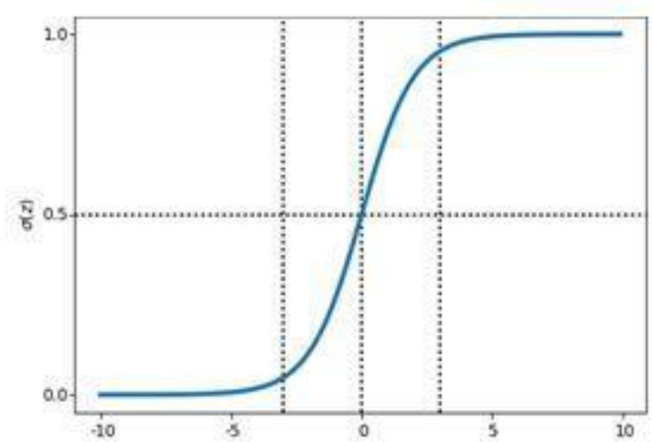

Fig. 2 : Graph for Sigmoid Function

2. h(xi) represents the predicted i^{th} observation, i.e. xi .The formula is as follows :

   $h(xi) = g(\beta^T x_i) = 1/1+e^{-\beta^T x_i}$

3. The conditional probabilities for y = 0 and 1 can be generalized as:

   $P(y_i|x_i;\beta) = (h(x_i))^{yi}(1-h(x_i))^{1-yi}$

4. Likelihood of the parameters i.e. Probability of data for a given model and parameter values:

   $L(\beta) = \prod_{i=1 \text{ to } n} P(y_i|x_i;\beta)$ or $L(\beta) = \prod_{i=1 \text{ to } n}(h(x_i))^{yi}(1-h(x_i))^{1-yi}$

5. A cost function is a measure of how wrong the model is in terms of its ability to estimate the relationship between X and Y. This is typically expressed as a

difference or distance between the predicted value and the actual value. The cost function for logistic regression is proportional to inverse of log-likelihood of parameters :

$$J(\beta) = \sum_{i=1 \text{ to } n} -y_i \log(h(x_i)) - (1-y_i)\log(1-h(x_i))$$

6. After finding the cost function for Logistic Regression, our job should be to minimize it,i.e., min J(θ). Gradient Descent is an optimization algorithm that helps machine learning models to find out paths to a minimum value using repeated steps. Gradient descent is used to minimize a function so that it gives the lowest output of that function.

$$\partial J(\beta) / \partial \beta_j = (h(x)-y)x_j$$

2.3.3. Random Forest

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. Random Forest can also be defined as the cluster of decision trees. Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

2.3.3.1. Algorithm

The "forest" it builds is an ensemble of decision trees, usually trained with the “bagging” method. The general idea of the bagging method is that a combination of learning models increases the overall result. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

The below diagram explains the working of the Random Forest algorithm:

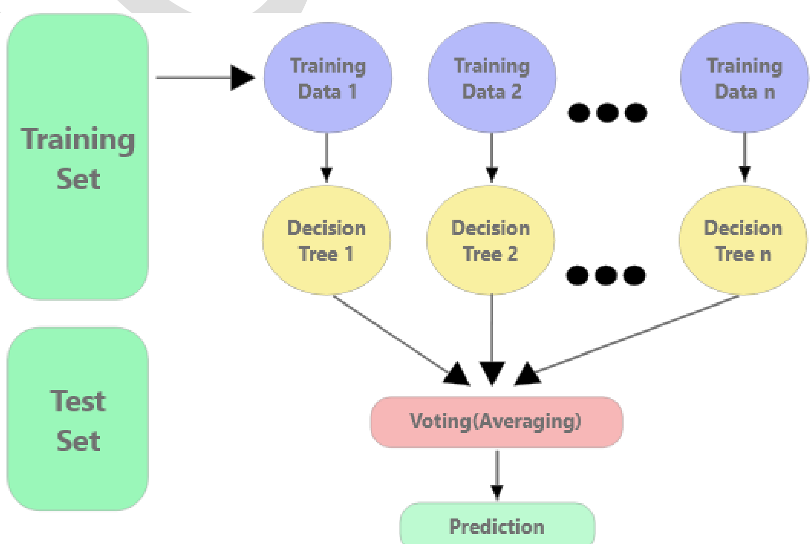

Fig. 3 : Flowchart for Random Forest Algorithm

Random Forest works in two-phases-

First is to create the random forest by combining N decision trees, and

Second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (subsets).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

3. Conclusion:

This project is an application of Machine Learning. For distinguishing among various genres, we hope to use Chroma-based feature, Time-Domain based feature, Spectral-domain and Cepstral-domain based feature. The proposed classification techniques are KNN and Logistic Regression. We plan to use a combination of above-mentioned features to achieve high classification accuracy.

4. Future Planning:

In the upcoming months we plan on completing feature selection and classification models and completing the entire design of the project. Our aim is to have a high classification accuracy, so selection of features and classifiers will be important.

Also, we hope to create an user interface where our model could be deployed and the user could input music files to yield proper results.

Gantt Chart :

Name of Module	Sept 2021	Oct 2021	Nov 2021	Dec 2021	Jan 2022	Feb 2022	Mar 2022	Apr 2022	May 2022
Requirement Analysis									
Feasibility									
Literature Survey									
Design									
Training									
Testing									
Report Writing									

References :

1. Bishop C. M., “Pattern Recognition and Machine Learning”, Springer, 2006
2. Duda, R. O.,Hart P. E., and Stork D. G., “Pattern Classification”, John Wiley & Sons, 2004
3. Bin Yu and Baozong Yuan (1993) A more efficient branch and bound algorithm for feature selection, Pattern Recognition, Vol. 26, No. 6, pp. 883-889.
4. Lie Lu, Hong-Jiang Zhang, and Hao Jiang. "Content analysis for audio classification and segmentation". Speech and Audio Processing, IEEE Transactions on, 10(7):504-516, 2002.
5. George Tzanetakis and Perry Cook. "Musical genre classification of audio signals". Speech and Audio Processing, IEEE transactions on, 10(5):293-302, 2002.
6. Kristopher West and Stephen Cox. “Features and classifiers for the automatic classification of music audio signals”. In *ISMIR*, 2004.
7. Benjamin Kedem. "Spectral analysis and discrimination by zero-crossings". Proceedings of the IEEE, 74(11):1477-1493, 1986.