Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume7-Issue1, January 2024
Paper: 64

Enhancing Music Genre Classification Accuracy using Machine Learning Models

U Venkata Pavan Kumar Reddy, *Student*, *Department of Computational Intelligence SRM Institute of Science and Technology, Kattankulathur, Chennai, TN, India*
up6089@srmist.edu.in

Thontepu Datta Naga Sri Manikanta, *Student*, *Department of Computational Intelligence SRM Institute of Science and Technology, Kattankulathur, Chennai, TN, India*
td8950@srmist.edu.in

Dr. M Vimaladevi
Assistant Professor, Department of Computational Intelligence SRM Institute of Science and Technology, Kattankulathur, Chennai, TN, India
vimaladm@srmist.edu.in

Abstract—Music is extremely beneficial to one's mental health. Communities are held together by music, which has the power to bring people from all walks of life together. We will tell a lot about a community by the kind of music they produce or like. Our study's and research's objective are to create a model which has the capability to predict songs in a certain genre more accurately than current models. musical style and form must be separated Many several classification schemes that divide music into its many subgenres.

Humans developed the classified designations known as musical genres to describe various types of music. To define a musical genre, we look at the shared traits of its members. These features are usually connected to rhythm and harmony. It is standard practice to employ genre hierarchies to organize the vast music libraries accessible on the Internet. At now, annotation of musical genres is done manually. There is a need for automated musical genre categorization in music systems since it may either supplement or replace the human user.

Automatic musical genre categorization offers a framework for generating and assessing musical signal analysis characteristics. The study is conducted to classify the music into its respective genres and recommends similar music. As a preliminary step, we looked for compelling traits that neatly distinguished across different genres. From the GTZAN genre dataset, feature vectors for the classifiers were constructed using the following five features: MFCC, chroma frequency, spectral roll-off, spectral centroid, and ZCR. Numerous classifiers were developed and put to use for classification, with variable degrees of success in terms of prediction accuracy.

I. INTRODUCTION

The characteristics of sound is encoded in an audio signal's frequency, decibel level, bandwidth, and other measurable quantities. The characteristics like Amplitude are considered to express an audio signal. These indicators are utilized with many advanced characteristics that facilitate and simplify computer analysis.

A song's virtual signatures for acoustics, danceability, tempo, strength, and other criteria are completed as part of music analysis in order to instruct listeners on the types of songs they should pay attention to.

Genres are broad categories used to categorize different types of music. These types of music are developed with the help of the participation of humans. The features which are commonly shared between the members of the genre are taken into account while categorizing them into their own distinct categories. It's common

knowledge that a song's rhythm, instrumentation, and harmony all play a role in defining its distinctive

qualities. Classifying track documents into their many genres is definitely challenging work in the subject of retrieving required features of audio, which is concerned with accessing, organizing, and looking at big-track collections.

As a result expansion of the virtual display industry, the concept of the computerized track category has emerged as extremely elegant in recent years. Putting songs into genres is bigoted, but the instruments, rhythm shape, and texture of a song can help define a style. The classification of musical styles for digitally distributed music has been done manually up until now. Consequently, methods for automated style classification could be a important complement to the audio record retrieval structures for tracks.

II. RELATED WORK

A.Papers that utilize CNN

Nikki Pelchat and Craig M. gelowitz[1].The research involved converting songs into short-time segments and representing them as spectrogram images. These images were labeled by genre and used as inputs for a CNN. The results were 85% accurate. Future research includes modifications to the algorithms and feature engineering, such as changes to weight initialization and experimenting with different filters. Additionally, future work may include additional features embedded into a colored spectrum.

Che-Nan Kuo [4] Music genre classification is a useful tool to help users find music that interests them, especially for beginners and those looking for music by specific musicians. The proposed convolutional neural network achieved an accuracy of 83.3%, which can be improved in future work. The goal is to integrate streaming media to complete the CNN architecture and increase efficiency

Md Sabbir Ahmed [8] proposed using Mel Frequency Cepstral Coefficients (MFCC) spectrograms to identify the Marsyas dataset. CNN was preferred for extracting features from

Volume7-Issue1, January 2024
Paper: 64

images and the genre waveform is converted into an MFCC spectrogram to minimize the time taken for training the model. The accuracy of classification using CNN and MFCC spectrogram is 92%, which proves the correctness of the hypothesis. The paper proposes to explore RNN or LSTM to improve the classification accuracy for larger data sets and different sound features.

Nirmal M R and Dr. Shajee Mohan B [9] proposed using a CNN classifier for music genre categorization on three genres from the GTZAN dataset. A user-defined sequential CNN model and a pre-trained CNN known as MobileNet were used, with classification accuracies of 40% and 67%, respectively. The study offers further experiments with alternative model settings and data augmentation using a Generative Adversarial Network (GAN).

Keunwoo Choi, Gyorgy Fazekas, and Mark Sandler [10] proposed CRNN for tagging the music and conducted a study for understanding how to control the size of the networks. The results show that the 2D convolution with 2d kernels (k2c2) and the CRNN perform similarly for a moderate usage of parameters. However, for a fewer or huge amount of parameters, there is a tradeoff between the speed of computation and memory of the neural networks. The k2c2 is faster than the CRNN, but CRNN tends to perform better for the same number of parameters.

Dr. Levi Ford, Dr. Sylvia Bhattacharya, Dr. Red Hayes, and Dr. Wesley Inman [11] offer a machine-learning approach for categorizing multilingual music into multiple genres. Although the model's accuracy can be improved, it is important to note that it is capable of detecting low-confidence misclassifications. If a considerable amount of songs are automatically categorized, this capability may be beneficial for manual verification.

Nikki Pelchat's [17] Neural Network Music Genre Classification proposed a study that suggests research in music and genre classification using machine learning and offers preliminary work on a neural network model for genre classification using spectrogram pictures of short-time segments of songs. The early findings revealed a 67% accuracy on testing data. Future studies will concentrate on algorithm adjustments, testing with different dataset sizes, assessing the effect of categorizing different frequency ranges and investigating possibilities such as binary classification for user preferences and employing a recurrent neural network with MIDI files as input.

Yunming Liang, Yi Zhou, Tongtang Wan and Xiaofeng Shu [19] suggest using CNNs and CRNNs with (DSC) for categorizing the given audio file into respective genres. Previous research utilizing Xception with DSC for picture classification and DSC for audio recordings inspired this technique. The authors also cite earlier work on utilizing neural networks to improve feature learning. By considering music characteristics as pictures, the suggested models attempt to employ model parameters more effectively and achieve greater accuracy.

Partha Protim Das [20] described the use of VGG and ResNet convolutional networks with a double layer of additional neural networks for genre classification using spectrogram images of 10 different genres. The VGG16-based model performed best with an accuracy of 84%, which is a significant improvement over previous studies. The authors suggest that more data and longer audio files could lead to even better results in the future.

Ahmet Elbir1, and Hilmi Bilal am2 [23] propose to categorize and suggest music using auditory information derived from digital signal processing technologies and convolutional neural networks. SVM outperformed other approaches in feature extraction and classification by utilizing digital signal processing technologies and CNNs. MFCC outperformed other algorithms, although deep learning methods did not exhibit significant performance changes in music genre categorization. Music recommendations based on the genre can be utilized to solve the problem. The research will further progress by concentrating on advanced feature extraction approaches for getting a better understanding and analysis.

Rafael L. Aguiar [27] studied ways to enhance music genre classification efficiency by employing data augmentation techniques with the help of Convolutional networks by Exploring various Data Augmentation techniques for a Better understanding of Music Genre Classification Convolutional networks. The Latin Music Database was used for utilizing four alternative data augmentation strategies, with the Tone Shifting technique of the data augmentation approach yielding the greatest results (89.45%). Future studies will look into other audio data augmentation approaches as well as merging other sorts of humanly created features with non-manually made features. The article also suggests that data augmentation be used to solve other audio challenges, such as bird species categorization and auditory event detection.

Shun-Hao Chang[28] made a customized music recommendation system using CNN as well as a collaborative filtering technique. The CNN method is used to label music based on audio signals, and the CF algorithm generates music suggestions based on the labeled data and the user log file. Further work will involve exploring metadata utilization, extracting user information for improved recommendations, and considering different DNN algorithms for music categorization, such as GRU and LSTM.

Fady Medhat and David Chesmore [30] studied the use of Conditional Neural Networks (CLNN) for detecting multivariate temporal signals. It is a CLNN variant with a band-like structure, allowing the network to learn frequency bands for automating feature combination discovery. Without the inclusion of extensions or musical perceptual data, the MCLNN surpasses earlier attempts at music genre recognition. Further research will look at several MCLNN designs as well

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume7-Issue1, January 2024
Paper: 64

as the MCLNN's applicability to other multichannel temporal signal formats.

B. papers that utilize SVM

Naofumi AOKI and Yoshinori DOBASHI [2] have conducted research on the automatic classification of musical genres and suggested an alternative method for extracting summaries from music data that collect summaries from many parts rather than just one. The findings of the experiment demonstrate that a proper summary of musical data may result in a more accurate classification of musical genres. The proposed method will be tested by adding more songs from each genre to the list of music genres to be classified in order to find the ideal conditions for the procedure. Further research will examine feature vectors besides MFCC.

Manuel Theodore Leleuly and P. H. Gunawan[3] explained the feature selection by using correlation in the feature correlation of music genre classification. The PNN approach and supervised learning with entropy as a feature are both used in this study. The proposed method uses four of the eight retrieved features to produce the best accuracy score of 90%. Further work would be dedicated to the extraction of feature sets that are highly efficient, utilizing various extraction techniques, and investigating various classifiers like SVM and ANN.

Moses Setiadi[5] has suggested classifying the genres comparing different algorithms on metadata and used to analyze the Spotify music dataset. The performance of classifiers used is compared in the study, with SVM-RBF proving to be the most accurate. The study comes to the conclusion that categorizing music genres based on metadata features can have accuracy comparable to that of audio feature extraction, with the benefit of being quicker as no conversion process is needed. The following study is advised to assess the accuracy and computational efficiency of the identification and categorization of music into its respective genres according to features described in metadata and audio features side by side.

Deepak S[13] has explained the role of feature extraction techniques in his paper. The model performs well on the GTZAN dataset and uses LSTM and LSTM+SVM to detect patterns specific to each class. Future research can concentrate on examining current changes in mixed-genre music, where the LSTM+SVM model can be used to train the SVM model for fresh fusion genres, and music from different genres can be divided and identified according to its genre.

Dipjyoti Bisharad [18] has proposed that identification of music genres can be done using RNN by describing a residual deep learning model for genre classification that, for finding the most probable genre classifications, obtains accuracy levels of 82%, 91%, and 94.5%, respectively. The performance of the model closely resembles how people perceive genre, with some genres being very distinct and others overlapping. These subtleties—such as the variations in rhythm and beat between jazz and classical music and the parallels between rock, metal, and country music—have all been successfully incorporated into the proposed work.

Jin, and Zhiyuan Cheng[21]: a supervised SVM multiclassification model and an unsupervised learning approach known as Multilevel Wasserstein Means. When the proper parameters are chosen, both models can produce acceptable results, but only the unsupervised model requires data training. When working with data that is equally accurate, the article suggests utilizing the unsupervised model.

Ahmet Elbi's [24] research was based on the Classification of genres using STFT and examined the changes that occurred in accuracy by using various window sizes, and different kinds of overlap ratios in the classification of musical genres. Three window sizes and overlap ratios were used to test six different window types, and the Parzen window has given the best accuracy with a size of 512 and an overlap ratio of 256. The study also discovered that the window with a size of 512-point is the most effective and that the best overlap ratio for classifying music genres is when analysis windows overlap by half their size. It is thought that the Parzen window's performance is enhanced by the well-localized frequency characteristics.

Nimagna Biswas [26] and Rajib Sarkar suggested a frequency domain feature-based approach for categorizing music genres using SVM and random forest classifiers. Characteristics are chosen to accurately represent the music's timbre, tone, and pitch. In three benchmark datasets, the suggested feature set beats other accepted published work, and the performance is constant regardless of the classifier employed. This implies that the suggested feature set is useful for classifying music genres.

C.Papers that utilize other Algorithms

Leisi Shi[6] established a paradigm for categorizing musical genres according to the chroma features of the audio files and utilized deep learning techniques that emphasizes the critical role of harmony and disregards unimportant elements like human timbre, volume, and absolute pitch. To represent fundamental data, it depends on chroma feature and an architecture of an upgraded VGG16 deep learning network. The model outperforms all earlier models in terms of classification accuracy, reaching a maximum of 92.12%. The subsequent work will examine how network structure affects the classification of music genres and test it using various datasets to determine the best network structure.

Meimei Wu and Xingli Liu[7] proposed a new methodology using DW-KNN for classifying music into genres, which outperforms the conventional KNN algorithm in two ways: by taking into account the correlation between features and their classes, as well as by taking into account the similarity between the nearest neighbors and the audio files to be classified. The DW-KNN algorithm is better suited for cross-class field classification and has a greater classification accuracy. It is also effective at classifying large amounts of musical data because of its straightforward and symmetrical calculation.

Vatolkin, Igor [12] purposed the project was to improve feature processing for music genre categorization using structural complexity and several semantic audio feature groupings. To determine the ideal feature combination and calculate the effect of grouping structural complexity on feature sets, they used evolutionary multi-objective feature selection. The findings demonstrated that the structural complexity method performs better for window sizes which are small, and a novel feature known as the chord vector was established. This feature outperformed comparable statistics from earlier research. Other structural complexity features and processing parameter combinations will be tested in future research for various classification tasks, including emotion identification and music segmentation.

Deepak S[13] introduced a music genre classification model employing processing methods on audio data to extract features, mainly MFCC. The model employs LSTM to classify audio based on genre. The model successfully competes on the GTZAN dataset using LSTM and LSTM+SVM to identify patterns specific to each class. The next work can concentrate on investigating the current evolution in mixed genre music, where the combination of LSTM and SVM model can be employed to train the SVM model for new fusion genres, and audio belonging to multiple genres can be divided and identified as belonging to a specific genre.

Prateek Verma [14] proposed a method by using melodic features to distinguish between Hindustani and Turkish music and found that features based on energy helped distinguish Turkish music from Indian Classical Music when classifying cultural music using features with melody proposed by Amruta Vidwans. The chosen model was able to discern genres based on listener signals, and listening experiments indicated that style distinction was audible in melody. It was discovered that timbre-based features enhanced the fundamental features.

Abdul Aziz Md, Ganga Raju Ravula [15] created a classification model using Naive Bayes to categorize songs based on their lyrics and discovered that the model was completely effective. The suggested approach can be used to automatically categorize music in paid music download services and offer suggestions depending on user moods. The study also made the case for further investigation into the model's behavior with more classes and the application of the morphological generator to produce more precise root words for Telugu terms.

Classification of Music Genres Leisi Shi [16] developed an upgraded VGG16 deep learning network and suggested a system for categorizing music genres based on Chroma Features and Deep Learning. The framework overlooks unimportant elements like timbre and volume in favor of the harmony of the music. It can achieve excellent classification accuracy—the best accuracy being 92.12%—and is resistant to background noise. The use of Resnet50, which had reduced accuracy, is also mentioned in the report, and more research examining the effect of network on music genre classification is recommended.

Zhen Wang[22] proposed that both the temporal dimension and the audio frequency dimension dependencies are extracted using RNN. CRNN-TF outperforms CRNN and in two music classification tasks utilising three real-world music data sets, employing a range of assessment measures.

Wenli Wu's [25] proposed a technique uses a 5-layer IndRNN with preprocessing technique of scattering coefficient, is a method for categorizing music genres. The method is demonstrated to have competitive performance with other models in comparision of classification accuracy and computation time. On the GTZAN dataset, the scattering transform and IndRNN combination successfully categorize different musical genres.

Sharaj Panwar [29] proposed that by categorizing audio signals based on elements like genre, instrument, and mood, music information retrieval can be used to make it simpler for listeners to choose their favorite music. For precise tag retrieval, deep feature extraction and learning techniques, particularly the union of CNN and RNN networks, are useful. The MagnaTagATune database's proposed CRNN network for tag recognition produced a high AUCROC of 0.893, demonstrating the effectiveness of the straightforward structure with fewer parameters over conventional approaches.

III.METHODOLOGY

1)Proposed Architecture Diagram

These are the steps to be followed

1. Data preparation: The GTZAN dataset is downloaded and the audio files are preprocessed to extract a set of features from each file. This involves splitting each audio file into 1-second windows and computing the set of features for each window using libraries like Librosa or Essentia.
2. Feature selection: Once the features are computed, a feature selection step may be performed to identify the most relevant features for classification. This can be done using techniques like mutual information, correlation analysis, or PCA.
3. Model selection and training: With the features selected, a classification model is chosen and trained

on a subset of the GTZAN dataset. Popular models for music genre classification include decision trees,Logistic Regression, support vector machines (SVM),XGBoost and neural networks.

4. Model evaluation: Once our model is trained, it is tested on a held-out test set of the GTZAN dataset. Metrics like accuracy, precision, recall, and F1 score can be used to measure the performance of the model.

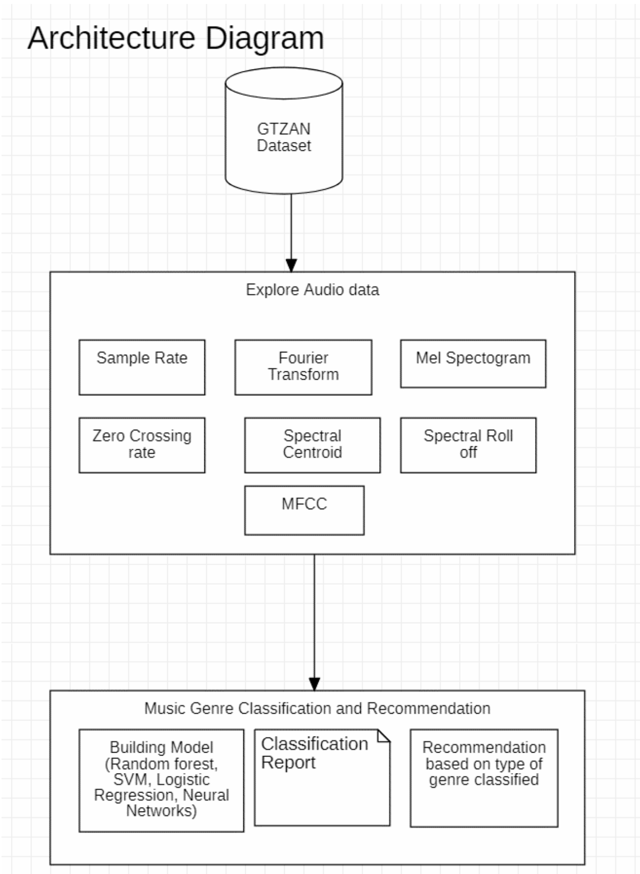


Figure 1: Architecture Diagram

2) Dataset

There are 1,000 audio files included throughout 10 different genres. There is a 30-second clip of music in each audio file. The.wav records are monophonic and 16 bits in size, with a sampling rate(sr) of 22050.0 Hz. The GTZAN dataset is a popular for music genre classification.The features used in this dataset are typically extracted using various signal processing techniques and then used as input to the machine learning models for genre classification.The dataset is being used in research on music genre classification.

3) DATA EXPLORATION

Data exploration can be called as the first step in data analysis and is an important step in understanding characteristics of the dataset,viewing and presenting data in order to get preliminary insights or suggest regions and patterns that should be further investigated. Identifying potential issues or biases that may affect the efficiency of machine learning models trained on the dataset.Here are the some common data exploration steps for the GTZAN dataset.

Import the necessary libraries

we can use libraries like numpy ,pandas for loading the dataset, matplotlib for visualing the data,librosa for feature extraction.

Loading the dataset

The dataset is available in wav format and the data can be loaded using using various libraries such as librosa.The dataset can be split into training and training sets.

Visualize the data

Visualizing the audio signals can help identify any issues with the dataset such as noise recorings or inconsistent sampling rates.One common visualization technique is to use matplotlib to visualize the audio waveform,spectograms and mel spectrograms of the audio clips to get an idea of the different genres and their characteristics

Computing summary statistics

Computing summary statistics such as mean,standard deviation and range for each feature can help identify any outliers in the data.To visualize the distribution of each feature histogram can be used.

Analyze the data

Pandas to create a dataframe to store information about the audio files such as genre,file name, and duration.We can use this dataframe to perform statistical analysize on the data such as finding the mean and standard deviation of the features of the audio files for each genre.
Preprocess the data
can use the librosa library to preprocess the audio files by extracting the features such as the mel spectrogram,chroma features and MFCCs.

Feature correlations

Examing the correlation between features can help identify potential redundancies or collinearites that may affect the performance of machine learning models.Correlation matrices or scatter plots can be used to visualize the relationships between pair of features.

Class distribution

Examing the distribution of classes can help identify any class imbalances that may affect the efficiency of machine learning models. The distribution of classes can be visualized using a bar chart.

Exploring genre characteristics

Examing the characteristics of each genre can help identify any unique features or patterns that specific to each genre.For example certain genres such as metal music may have high

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume7-Issue1, January 2024
Paper: 64

emphasis on certain spectral or rhythmic features.

Further analysis

Can perform analysis on the dataset such as using dimensionality reduction techniques like PCA to visualize the data in a lowe-dimensional space

4) EXPLORE AUDIO DATA

Sample rate

The number of audio signals per second taken from an analog signal to be converted into a digital signal, and it can be expressed in hertz (Hz) or kilobits per second (kHz).The higher the sample rate, the more accurately the digital signal can represent the the original analog signal.The sample rate is an important factor in determining the quality of the audio recording as a higher sample rate can capture more detail in the audio signal.However, higher rates also require more storage space and processing power.
2D Representation: Sound Waves

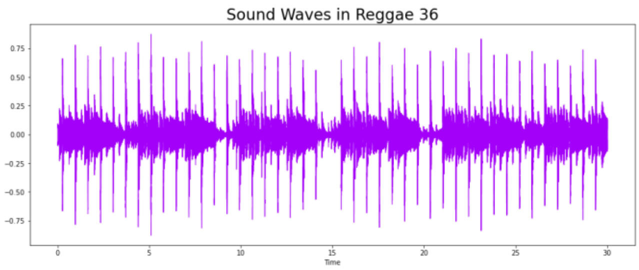


Figure 2:Sound waves in Raggae

Fourier Transform

Returns the result of decomposing a time-domain signal itself into the component frequencies. It is possible to examine and break down complicated signals into simpler parts using the Fourier transform mathematical approach. Since it makes it possible to comprehend the fundamental frequencies and harmonics of music, it is frequently utilised in signal processing and music analysis.

An analysis of a sound signal's frequency content in music may be done using the Fourier transform.

This may be achieved by describing the sound wave as the sum of sine and cosine waves with various amplitudes and frequencies.

This is referred to as the Fourier series.

A time domain signal can also be converted to a frequency domain signal using the Fourier Transform.

Spectrogram

A spectrogram displays, graphically, the signal's frequency spectrum as it evolves over time. When spectrograms are applied to an audio stream, they are often referred to as sonograms.
In music, spectograms are used to analyze and visualize the different frequencies and harmonics present in a music.Spectograms are generated using Fourier Transform.The Fourier Transform is applied to small time segments of the sound signal and the resulting frequency components are displayed as a color-coded graph,with frequency is pictured on the vertical axis,time is pictured on the horizontal axis, and amplitude is represented by color as shown in the figure [No].Here, we do the opposite and make the frequency axis logarithmic.

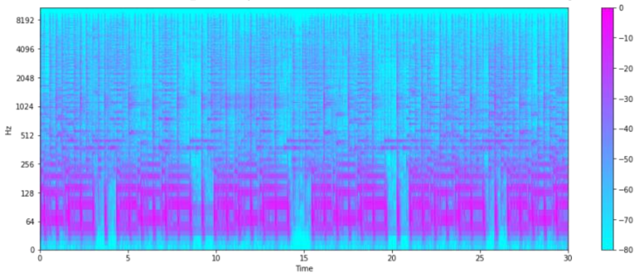


Figure 3:Fourier Transform

Mel Spectrogram

From a mathematical perspective, the Mel Scale is the product of non-linear transformation applied on frequency scale. These are commonly used features in audio signal processing and are derived by taking the Fourier transform of a short term window of audio signal and then mapping the resulting spectrum onto the mel scale.The resulting coefficients are then transformed using the discrete cosine transform to obtain a set of MFCC s typically between 12 to 20 coefficients per frame.
"Sound waves that are equally spaced on the Mel Scale sound similarly to our ears."

Mel Spectrogram for metal genre is as shown in the figure

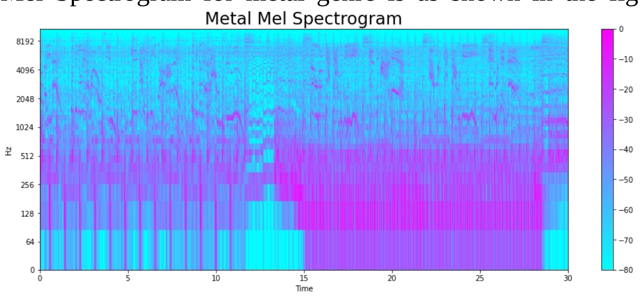


Figure 4: Metal Mel Spectogram

Audio Features

Zero Crossing Rate

ZCR can be described as how quick the signal flips from positive to negative and vice versa.This is a statistical feature that describes the number of times the signal crosses the zero axis within a certain time window.The zero crossing rate can be used to analyze and identify the rhythmic,tempo characteristics of music.The zero crossing rate can also be used as a feature in music analysis processing algorithms such as automatic rhythm detection or beat tracking .By analyzing the zero crossing rate over time these algorithms can identify the underlying rhythmic structure of a music and align it witha metronome other reference tempo.Overall the zero crossing rate is a useful measure for analyzing the rhythmic and dynamic characteristics of music, and it can be used in a variety of music analysis.

Harmonics and Perceptual

A human's hearing isn't sensitive enough to pick up on harmonics In terms of the brain's ability to process information, a shock wave symbolizes the music's tempo and the feelings in it. Tempo and rhythm features are related to the rhythmic content of the audio signal including tempo,beat histogram,rhythm patterns..

Harmonics are created when a musical instrument or voice produces a sound wave that contains multiple frequencies that are related to each other by simple mathematical ratios.These related frequencies create harmonic overtones that are heard as part of the overall sound.

The perception of harmonics is important in music, as it contributes to the characteristics of sound and timbre of different instruments and voices.

Harmonice play a fundamental role in the perception of music and they are an important consideration in music analysis,production, engineering.The harmonices and perception are represented in the figure.

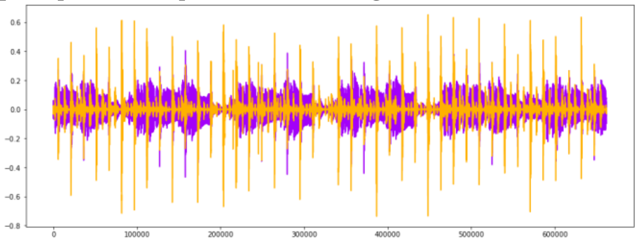


Figure 5 : Harmonics

Spectral Centroid

Spectral Centroid is used to display the location of the sound's "center of mass," which is determined by taking a weighted average of the sound's frequency.Spectral Centroid is a commonly used feature in music analysis, particularly in the field of music information retrieval. It refers to the center of gravity of the spectrum of a sound signal and is calculated as the weighted mean of the frequencies present in the signal.

In music, spectral centroid can be used to identify the brightness or darkness of a sound. For example, a high spectral centroid value indicates that a sound has a lot of high-frequency content and is therefore likely to be perceived as bright or sharp, while a low spectral centroid value indicates that a sound has more low-frequency content and is likely to be perceived as dull or mellow.

Spectral centroid is often used in combination with other features to classify music into different genres or to identify specific types of instruments.

Overall, spectral centroid is a useful feature for music analysis and can provide valuable information about the spectral content of a sound signal.

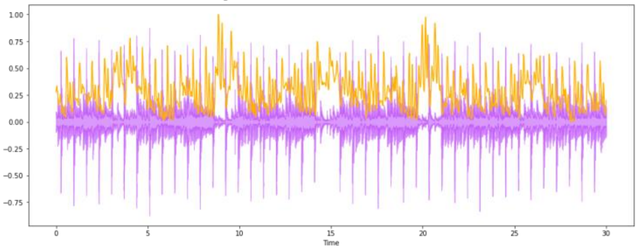


Figure 6 : Spectral Centroid

Spectral Roll off

Spectral Roll off defines the signal's profile in numerical form. It's a measure of how low in frequency a certain amount of the spectrum's energy may be found.Spectral rolloff is another commonly used feature in music analysis, particularly in the field of music information retrieval. It is a measure of the frequency at which a specified percentage of the total spectral energy of a sound is contained below that frequency.

In music, spectral rolloff can be used to identify the high-frequency content of a sound. A high spectral rolloff value indicates that a sound has a lot of high-frequency content, while a low spectral rolloff value indicates that a sound has more low-frequency content.

Spectral rolloff is often used in combination with other features for classifying music into its genres or to identify specific types of instruments. For example, in a study on music genre classification, spectral rolloff was found to be one of the important features for distinguishing between different genres of music, along with features such as spectral centroid, zero-crossing rate, and MFCCs.

Overall, spectral rolloff is a useful feature for music analysis and can provide valuable data regarding the spectral content of a sound signal, particularly in the high-frequency range.

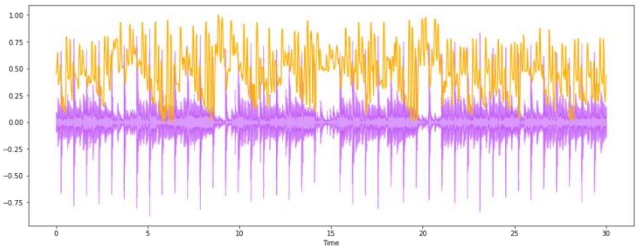


Figure 7 : Spectral Rolloff

Mel-Frequency Cepstral Coefficients

A signal's MFCC can be considered as a small group of characteristics that can be used to represent the complete pattern of a spectral envelope. It resembles the sounds of a human voice.

Mel-Frequency Cepstral Coefficients (MFCCs) are one of the

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume7-Issue1, January 2024
Paper: 64

most commonly used features in music analysis and other audio signal processing tasks. MFCCs are derived from the Mel frequency scale, which is a perceptually relevant frequency scale based on the human auditory system.

To compute MFCCs for a given audio signal, the signal is first divided into overlapping frames of equal duration. Then, a Fourier transform is applied to each frame to obtain the power spectral density, which represents the distribution of signal power over different frequencies. The power spectral density is then transformed into the Mel scale using a set of Mel filter banks, which are triangular filters that are uniformly spaced on the Mel frequency scale. The logarithm of the Mel-filtered spectral energies is then computed and transformed using the discrete cosine transform (DCT) to obtain a set of coefficients, which are the MFCCs.

In music, MFCCs can be used to represent the spectral characteristics of a sound signal in a compact and efficient manner. They capture both the spectral shape and the temporal variations of the sound signal, making them suitable for a wide range of music analysis tasks such as genre classification, instrument recognition, and emotion recognition.

MFCCs are often used in combination with other features such as spectral centroid, spectral flux, and zero-crossing rate to improve the accuracy of music analysis algorithms. Overall, MFCCs are a powerful tool for analyzing the acoustic characteristics of music and have become a standard feature in many music analysis systems.

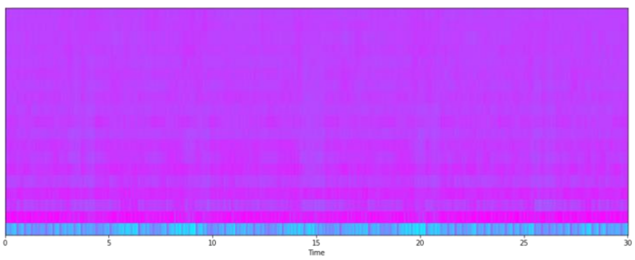


Figure 8: MFCC

Chroma Frequencies

Chroma features are based on the idea of representing the pitch,content of an audio signal by projecting classes.Chroma features are derived by then, divide the audio signal into tiny time intervals.,then computing the short term fourier transform of each frame and finally mapping the results powerful spectrum.Audio with chroma characteristics, here we will make the complete spectrum projected onto 10 bins which reflects on the 10 unique semitones of the music octave, is an unusual and powerful representation.

Chroma Frequencies are a set of features used in music analysis that represent the distribution of energy over twelve pitch classes in an audio signal.To compute the Chroma features for a given audio signal, the signal is first divided into frames of fixed duration.The frequency spectrum is then generated for each frame using a short-time Fourier transform (STFT).Then, each frequency band's energy is mapped to the relevant pitch class using a formula that takes into consideration the human auditory system's frequency sensitivity.. Finally, the energy in each pitch class is summed over all the frequency bands to obtain the Chroma feature vector for the frame.

Chroma features can be used to capture the harmonic content of a music signal, as they represent the energy in each pitch class regardless of its octave. They are particularly useful for tasks such as chord recognition, key detection, and melody extraction. For example, in chord recognition, Chroma features are often used to compute the similarity between the chord sequence of a given song and a set of known chord progressions.

Chroma features can also be combined with other features such as MFCCs and spectral features to improve the accuracy of music analysis algorithms. Overall, Chroma frequencies are a powerful tool for analyzing the harmonic content of music and have become a standard feature in many music analysis systems.

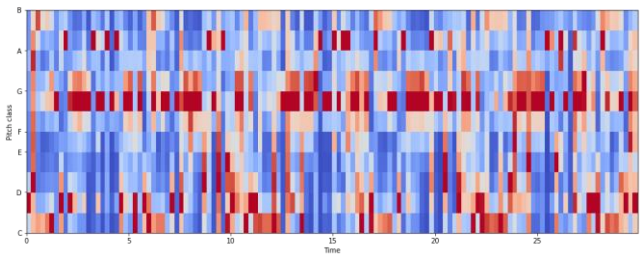


Figure 9 : Chroma Frequencies

5)Exploratory Data Analysis

Correlation for features used

Can compute the matrix of correlation coefficients to investigate the relationship between these features. A statistical metric that expresses the direction and intensity of a relationship between two variables is the correlation coefficient. It has a value range of -1 to 1, with a value of -1 denoting a perfect negative correlation, a value of 0 denoting no correlation, and a value of 1 denoting a perfect positive correlation.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume7-Issue1, January 2024
Paper: 64

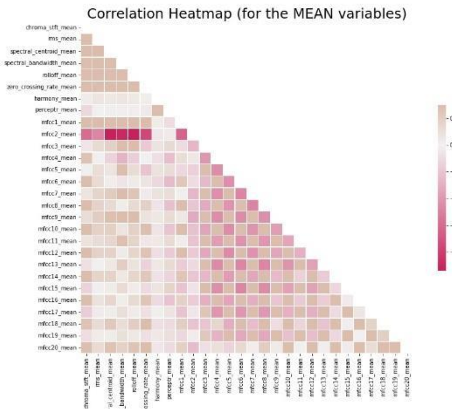


Figure 10 : Correlation

Beat per minute for genres

The beats per minute (BPM) of a song is a measure of the tempo or speed of the song. The BPM of a song can vary widely between different genres of music.here can be significant variations within each genre, and some sub-genres may have significantly different BPM ranges than the ranges listed above.

It's worth noting that BPM is just one aspect of the tempo of a song, and there are other factors such as rhythm, groove, and syncopation that can also affect the perceived tempo. Additionally, many modern songs use electronic production techniques to manipulate the tempo and create rhythmic effects, so the BPM of a song may not always be a straightforward measure of its tempo.

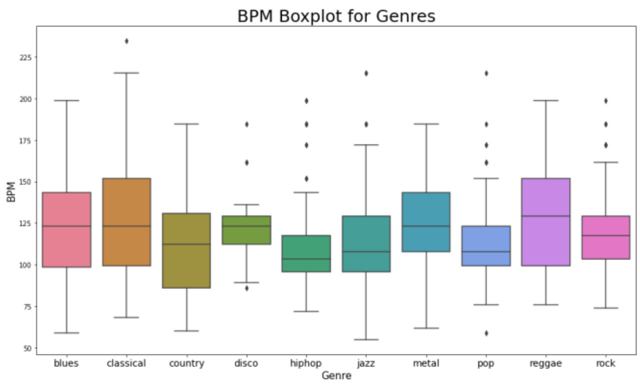


Figure 11 : Beats per minute for genre

Principal Component Analysis on Genres

PCA is a popular approach for dimensionality reduction and feature extraction in. music analysis. It can be applied to the GTZAN dataset to identify the underlying factors that contribute to the variability in the music features across different genres.

To perform PCA on the GTZAN dataset,the first step is to extract the appropriate audio elements from the files. Popular features for music genre classification include spectral features such as Mel-Frequency Cepstral Coefficients (MFCCs), spectral centroid, spectral rolloff, and spectral flux, as well as temporal features such as zero-crossing rate and tempo.

Once the features have been extracted, they can be standardized by centering and scaling them to have a mean of zero and a variation of one unit. The standardized features can then be used as input to the PCA algorithm, which computes the principal components of the dataset.

The principal components are linear combinations of the original attributes that encapsulate the most significant causes of data variability. The first principal component explains the greatest amount of variation in the data, the second principle component explains the next greatest amount of variation, and so on. To understand how the different genres are dispersed in the reduced feature space, the principal components can be represented in a scatter plot or a heatmap.

PCA can be used to identify the features that are most important for distinguishing between different genres of music. For example, if the first principal component is dominated by features related to tempo and rhythm, then it may be a good indicator of dance music genres such as EDM and hip-hop. On the other hand, if the first principal component is dominated by features related to harmonic content and melody, then it may be a good indicator of genres such as classical and jazz.

In summary, PCA is a powerful tool for analyzing the underlying structure of music features in the GTZAN dataset and identifying the factors that contribute to the variability across different genres.

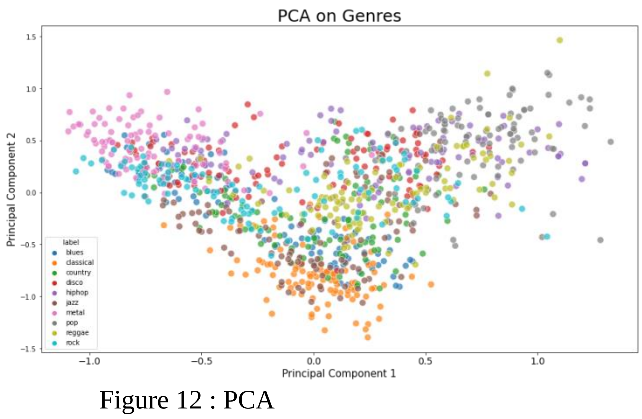


Figure 12 : PCA

6)Feature Data Summary

In terms of statistical functions, the dataset includes the Zero Crossing Rate (ZCR) feature, which measures the rate at which the audio waveform crosses the zero axis. For ZCR, the mean and median values are calculated, resulting in two features.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume7-Issue1, January 2024
Paper: 64

The dataset also includes features related to spectral characteristics of the audio signal. Specifically, it includes Spectral Bandwidth, Spectral Centroid, and Spectral Roll-off. For each of these features, the mean and median values are calculated, resulting in a total of six features.

Finally, the dataset includes features derived from Mel-Frequency Cepstral Coefficients (MFCCs). MFCCs are commonly used in speech and music analysis and provide a compact representation of the spectral characteristics of the audio signal. In the GTZAN dataset, 39 MFCC coefficients are derived, resulting in 39 features.
Overall, the GTZAN dataset includes a total of 47 features.

Feature	Statistical Functions	# of Features
ZCR	Mean, median	2
Spectral Bandwidth		2
Spectral Centroid		2
Spectral Roll off		2
MFCC Derivation		39
Total		**47**

Table 1 : Feature Data Summary

7) Algorithms used

Different machine learning algorithms have been trained and evaluated on the GTZAN dataset and their accuracies have been reported. The choice of the algorithms used for this task is likely based on their previous success in similar classification tasks,as well as their ability to handle the specific characteristics of the dataset. Here's a brief explanation of why each algorithm may be suitable for our dataset GTZAN.

Naive bayes because this is a simple and efficient probabilistic algorithm that can operate well with huge datasets with a large number of features such as the audio features extracted from the GTZAN dataset.

Decision Tree because this can handle both categorical and continuous data and can be useful for identifying patterns in the audio features that are most important for distinguishing between different music genres.

Multi layer perceptron because this a oneof neural network that can learn complex non-linear relationships between the input features get the output labels,making it suitable for music genre classification tasks.

Logistic regression because this is a simple efficient algorithm that can work well on dataset with a large number of features.

Support Vector Machine because this algorithm work well on datasets with complex and non linear relationships between the input features and output labels, and has been shown to perform well on music genre classification tasks.

K-Nearest Neighbors because this algorithm can be useful for identifying similar patterns in the audio features of different songs which can be used to categorize them into different music genres.

Random Forest because this is a ensemble learning method that combine multiple decision tress help to increase the accuracy of the classification findings.

XG Boost because it is a gradient boosting algorithm that can improve the accuracy of the classification results by minimizing the prediction error of the previous model.

Ensemble Model because this model combines the predictions of the multiple learning models to help increase the accuracy of the classification findings.

CNN because is a type of neural network that is specifically designed to work well with image and audio data making it suitable for music genre classification tasks.

Logistic Regression

Logistic regression is a supervised learning, is a popular component of Machine Learning. Logistic regression is a classification algorithm that can be used to predict the probability of an observation belonging to a certain class, in this case, the probability of an audio file belonging to a certain genre. The logistic regression algorithm works by fitting a logistic curve to the input features and using this curve to estimate the probability of a particular outcome.

Therefore, the final outcome has to be either categorical or discrete. The output offers a variety of possibilities between Zero and one, rather than the binary values Zero and one.

Random Forest (RF)

As an ensemble learner, Random Forest could utilize the vital knowledge from many decision trees to generate a more reliable prediction. It accomplishes this by mixing two areas of potential, such as: 1) During the training phase, bootstrap aggregation only employs a portion of the whole data set to generate each individual decision tree (or bagging) 2) Each tree in the decision-making mechanism must predict the Random Forest's class based on a fully independent set of attributes. The class is determined by a majority vote among some of the classifiers.

K-Nearest Neighbor

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume7-Issue1, January 2024
Paper: 64

One of the simplest machine learning methods is the K-Nearest Neighbor method, which offers analysis by means of the Supervised Learning paradigm. The K-Nearest Neighbor approach focuses on the grouping of content into instances that are most resemble those that are grouped. The K-Nearest Neighbors (K-NN) technique takes into consideration every piece of data and analyzes the data's relationships to identify how to label a new data instance. Thus, the K- NN technique could be used to efficiently categorize newly available data into a respective category. The K-NN method is frequently utilized for classification problems, at the same time it can be utilized for Regression problems.

Gradient Boosting (XGB)

An ensemble classifier, or "boost," is the result of integrating several "weak learners" (such as decision trees). Not like RFs, however, which can be learned in parallel, boosting algorithms must go through a lengthy and tedious process of training in stages.
Early rounds in a neural network's training process result are very straightforward decision trees. As training continues, classifier improves because it is taught to focus on the situations in which it previously failed. At the end of the training process, the model performance may be a linear weighted average of the results from the various students. Extreme Gradient Boosting, or XGB for short, is a version of boosting that facilitates rapid, parallelized model training.

Naive Bayes

Naive Bayes algorithm is an example of supervised learning paradigm which utilizes Bayes theorem to solve classification problems.
In order to construct a machine learning model that can generate predictions rapidly and efficiently. Naive Bayes Classifier is often used due to its simple and effective properties. It makes its predictions based on the object's likelihood since it is a probabilistic classifier. Because Naive Bayes believes that all variables are unrelated, it is incapable of learning the link between features.

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

Figure 13 : Conditional Probability

Convolutional Neural Network(CNN)

Convolutional neural networks are built from a feed-forward neural network with several hidden layers stacked in a certain sequence. The capability of CNN to learn hierarchical properties is a byproduct of the sequential architecture. In a standard CNN architecture, the first set of layers includes of convolutional layers, subsequently activation layers, then grouping layers, and finally the hidden layers. Input is provided via the picture. Image pixels are read in as arrays by the input layer. It is possible for CNNs to have numerous hidden layers that each do their own set of computations to extract features from the input picture. Convolution, pooling, and fully - connected layers are all part of CNN architecture

Architecture of CNN

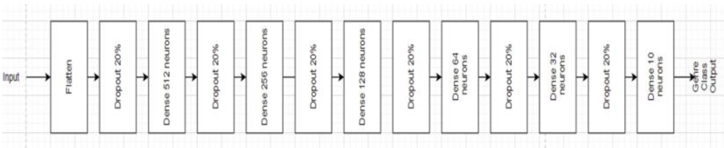


Figure 14 : Architecture of CNN

Cosine Similarity

For music recommendations, we turned to cosine similarity.
Cosine similarity is used as a statistic to get the degree to which documents of varying sizes are similar to one another. Cosine angle is a mathematical measure of the projection of two vectors into a higher dimensional space. Cosine similarity is preferable because it increases the likelihood that two documents with comparable content will be orientated more similarly, even if the documents' Euclidean distances are large. In general, the cosine similarity increases as the angle decreases.

Cosine similarity is a common similarity metric used in recommendation systems. In the case of music recommendation, it can be used to recommend songs that are similar in terms of their audio features. The GTZAN dataset, which contains audio files of different music genres, can be used to demonstrate how cosine similarity can be used to recommend similar songs.

To use cosine similarity for song recommendation, we first need to extract audio features from the audio files. As mentioned in the previous answer, the GTZAN dataset includes various audio features such as ZCR, Spectral Bandwidth, Spectral Centroid, Spectral Roll-off, and MFCCs.

Once the features are extracted from the music files, we can calculate the cosine similarity between each pair of songs. The cosine similarity between two songs is calculated as the cosine of the angle between the two feature vectors representing the songs. The cosine similarity value ranges from -1 to 1, where 1 indicates that the two songs are very similar, and -1 indicates that they are very dissimilar.

To recommend similar songs, we can take a user's preferred song as input and calculate the cosine similarity between this song and all the other songs in the dataset. We can then rank the songs based on their cosine similarity values and

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume7-Issue1, January 2024
Paper: 64

recommend the top M songs that are very much similar to the user choice.

In summary, cosine similarity can be used to recommend similar songs based on their audio features in the GTZAN dataset. The process involves extracting audio features, calculating cosine similarity between songs, and ranking and recommending similar songs based on their cosine similarity values.

This technique estimates the cosine similarity in between two music pairs in the dataset. This yields a 1000*1000 matrix (with some information duplication, since similarity between items A and B equals similarity between items B and A).

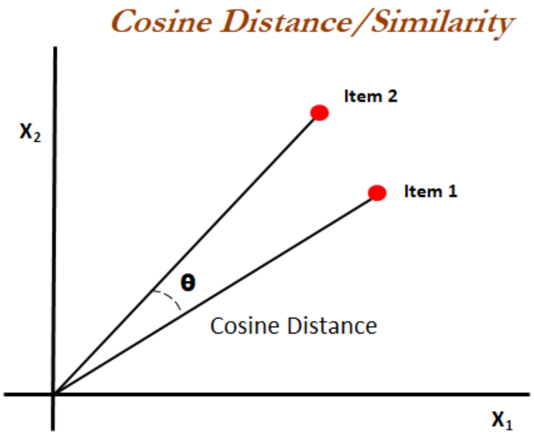


Figure 15 Cosine Similarity

8) Evaluation metrics

Accuracy: Accuracy refers to how well our model predicts the correct class or labels. If our dataset is somewhat balanced and all classes are of equal importance, this should be our baseline metric for measuring the performance of our version.Accuracy is a commonly used metric for evaluating classification models, including those used for our dataset GTZAN. It measures the proportion of correctly classified instances out of all instances in the dataset.

$$\text{ACCURACY} = \frac{\text{TRUE POSITIVES} + \text{TRUE NEGATIVES}}{\text{ALL SAMPLES}}$$

IV Result

As discussed in the methodology we have chosen the models mentioned above.The table shows the accuracies of different machine learning models.Each model has been trained on GTZAN dataset.Each row represent a different machine learning model and the columns show the name of the model and its accuracy score.

In the table 2 the accuracy scores ranges. The table shows a range of accuracies scores for different machine learning models from from 0.51 to 0.94 with the highest accuracy being achieved by CNN model with a score of 0.94.This suggest that the CNN model is the most accurate and best performing model and naive bayes model has the lowest accuracy score of 0.51 while.The other models fall in these two extremes.

Other models such as XG Boost and the ensemble model also achieved high accuracy scores indicating that they are strong performers for the GTZAN dataset.

MODEL	Accuracy
Naïve Bayes	0.51
Decision Tree	0.64
Multi-layer Perceptron	0.67
Logistic Regression	0.69
Support Vector Machine	0.75
KNN	0.80
Random Forest	0.81
XG Boost	0.90
Ensembled Model	0.92
CNN	0.94

Table 2: Results

If the same information is represented in a bar graph, it would look something like this as shown in the figure 16**.**In this bar graph,the accuracy scores are shown on the vertical axis,and each model is represented by a orange colour bar.The height of each bar corresponds to the accuracy score for that model.

The bars are arranged of increasing accuracy with smallest accuracy model(Naive Bayes) on the left and the highest accuracy model on the right.The models are labeled on the horizontal axis.

The bar graph provides a visual representation of the same information provided in the table 2,but it allows us to more easily compare the accuracy scores of different models.One can see that there is a large difference between the lowest-scoring model(Naive Bayes) and the highest scoring model(CNN) and can also see there is a clear progression in accuracy from left to right.The bar graph also makes it easy to quickly identify the highest and lowest-scoring models.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume7-Issue1, January 2024
Paper: 64

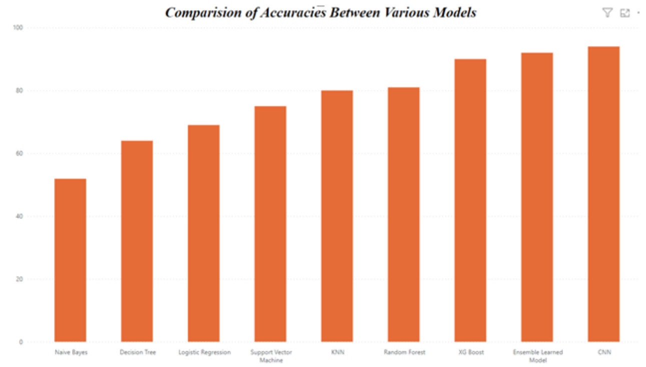
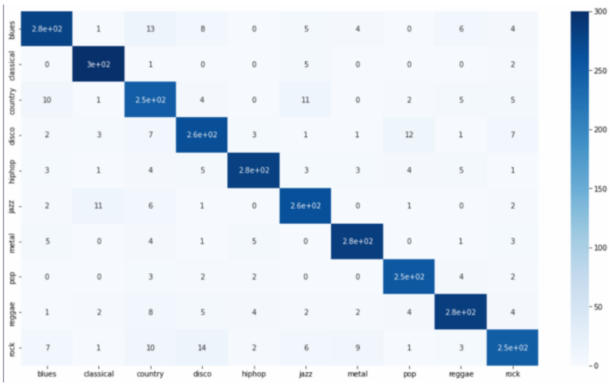


Figure 16 : Comparison of Accuracies between various models

Confusion matrix for xg boost predictions



Conclusion

Based on our thorough evaluation of the accuracy of each model, we conclude that CNN is the best option for the GTZAN dataset with 94% accuracy. Other models such as XG Boost and the ensemble model also achieved high accuracy scores indicating that they are strong performers for the GTZAN dataset. In this study, we provide the specifics of a program that uses Machine Learning strategies to categorize musical genres. The software's categorization is accomplished by use of a model built using a number of machine-learning methods. Each song in the GTZAN dataset is converted into a Mel Spectrum. This is accomplished with the help of the librosa python package. The process of sorting a massive music library into its many genres is automated by a piece of software.

Overall, our findings demonstrate the potential of machine learning techniques for music genre classification. However, further research is needed to optimize and fine-tune these models for even higher accuracy. Additionally, it would be interesting to explore the effectiveness of these models on different datasets and for different types of music.

we may expand our focus to include larger data sets and/or include more track formats (mp3). Furthermore, as time passes, the aesthetics that each genre embodies will evolve. As such, moving forward, we want to keep up with the ever-evolving aesthetics of various genres and adapt our software accordingly.

REFERENCES

[1] N. Pelchat and C. M. Gelowitz, "Neural Network Music Genre Classification," in Canadian Journal of Electrical and Computer Engineering, vol. 43, no. 3, pp. 170-173, Summer 2020,
doi: 10.1109/CJECE.2020.2970144.

[2] Y. KIKUCHI, N. AOKI and Y. DOBASHI, "A Study on Automatic Music Genre Classification Based on the Summarization of Music Data," 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan, 2020, pp. 705-708,
doi: 10.1109/ICAIIC48513.2020.9065046.

[3] M. T. Leleuly and P. H. Gunawan, "Analysis of Feature Correlation for Music Genre Classification," 2020 8th International Conference on Information and Communication Technology (ICoICT), Yogyakarta, Indonesia, 2020, pp. 1-4,
doi: 10.1109/ICoICT49345.2020.9166333.

[4] Y. -H. Cheng, P. -C. Chang and C. -N. Kuo, "Convolutional Neural Networks Approach for Music Genre Classification," 2020 International Symposium on Computer, Consumer and Control (IS3C), Taichung City, Taiwan, 2020, pp. 399-403,
doi: 10.1109/IS3C50286.2020.00109.

[5] D. R. Ignatius Moses Setiadi et al., "Comparison of SVM, KNN, and NB Classifier for Genre Music Classification based on Metadata," 2020 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 2020, pp. 12-16, doi: 10.1109/iSemantic50169.2020.9234199.

[6] L. Shi, C. Li and L. Tian, "Music Genre Classification Based on Chroma Features and Deep Learning," 2019 Tenth International Conference on Intelligent Control and Information Processing (ICICIP), Marrakesh, Morocco, 2019, pp. 81-86, doi: 10.1109/ICICIP47338.2019.9012215.

[7] M. Wu and X. Liu, "A Double Weighted KNN Algorithm and Its Application in the Music Genre Classification," 2019 6th International Conference on Dependable Systems and Their Applications (DSA), Harbin, China, 2020, pp. 335-340, doi: 10.1109/DSA.2019.00051.

[8] M. S. Ahmed, M. Z. Mahmud and S. Akhter, "Musical Genre Classification on the Marsyas Audio Data Using Convolution NN," 2020 23rd International Conference on Computer and Information Technology (ICCIT), DHAKA, Bangladesh, 2020, pp. 1-6, doi: 10.1109/ICCIT51783.2020.9392737.

[9] N. M R and S. Mohan B S, "Music Genre Classification using Spectrograms," 2020 International Conference on Power, Instrumentation, Control and Computing (PICC), Thrissur, India, 2020, pp. 1-5, doi: 10.1109/PICC51425.2020.9362364

[10] K. Choi, G. Fazekas, M. Sandler, and K. Cho, "Convolutional recurrent neural networks for music classification," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017, pp. 2392-2396, doi: 10.1109/ICASSP.2017.7952585.

[11] L. Ford, S. Bhattacharya, R. Hayes and W. Inman, "Using Deep Learning to Identify Multilingual Music Genres," 2020 SoutheastCon, Raleigh, NC, USA, 2020, pp. 1-5, doi: 10.1109/SoutheastCon44009.2020.9368270.

[12] P. Ginsel, I. Vatolkin and G. Rudolph, "Analysis of Structural Complexity Features for Music Genre Recognition," 2020 IEEE Congress on Evolutionary Computation (CEC), Glasgow, UK, 2020, pp. 1-8, doi: 10.1109/CEC48606.2020.9185540.

[13] S. Deepak and B. G. Prasad, "Music Classification based on Genre using LSTM," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 985-991, doi: 10.1109/ICIRCA48905.2020.9182850.

[14] A. Vidwans, P. Verma and P. Rao, "Classifying Cultural Music using Melodic Features," 2020 International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, 2020, pp. 1-5, doi: 10.1109/SPCOM50965.2020.9179597.

[15] A. Aziz Md., G. Raju Ravula, M. T. Shaik and S. Potharaju, "Telugu Lyrics Based Classification By Using Naive Bayes," 2020 International Conference on Artificial Intelligence and Signal Processing (AISP), Amaravati, India, 2020, pp. 1-6, doi: 10.1109/AISP48273.2020.9073597.

[16] L. Shi, C. Li and L. Tian, "Music Genre Classification Based on Chroma Features and Deep Learning," 2019 Tenth International Conference on Intelligent Control and Information Processing (ICICIP), Marrakesh, Morocco, 2019, pp. 81-86, doi: 10.1109/ICICIP47338.2019.9012215.

[17] N. Pelchat and C. M. Gelowitz, "Neural

Network Music Genre Classification," 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), Edmonton, AB, Canada, 2019, pp. 1-4, doi: 10.1109/CCECE.2019.8861555.

*[18]*D. Bisharad and R. H. Laskar, "Music Genre Recognition Using Residual Neural Networks," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, pp. 2063-2068, doi: 10.1109/TENCON.2019.8929406.

*[19]*Y. Liang, Y. Zhou, T. Wan and X. Shu, "Deep Neural Networks with Depthwise Separable Convolution for Music Genre Classification," 2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP), Weihai, China, 2019, pp. 267-270, doi: 10.1109/ICICSP48821.2019.8958603.

*[20]*P. P. Das, A. Acharjee and Marium-E-Jannat, "Double Coated VGG16 Architecture: An Enhanced Approach for Genre Classification of Spectrographic Representation of Musical Pieces," 2019 22nd International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2019, pp. 1-5, doi: 10.1109/ICCIT48885.2019.9038339.

*[21]*J. Wei, C. Jin, Z. Cheng, X. Lv and L. Song, "Music Classification using Multiclass Support Vector Machine and Multilevel Wasserstein Means," 2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS), Beijing, China, 2019, pp. 408-412, doi: 10.1109/ICIS46139.2019.8940172.

*[22]*Z. Wang, S. Muknahallipatna, M. Fan, A. Okray and C. Lan, "Music Classification using an Improved CRNN with Multi-Directional Spatial Dependencies in Both Time and Frequency Dimensions," 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 2019, pp. 1-8, doi: 10.1109/IJCNN.2019.8852128.

*[23]*A. Elbir, H. Bilal Çam, M. Emre Iyican, B. Öztürk and N. Aydin, "Music Genre Classification and Recommendation by Using Machine Learning Techniques," 2018 Innovations in Intelligent Systems and Applications Conference (ASYU), Adana, Turkey, 2018, pp. 1-5, doi: 10.1109/ASYU.2018.8554016.

*[24]*A. Elbir, H. O. İlhan, G. Serbes and N. Aydın, "Short Time Fourier Transform based music genre classification," 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), Istanbul, Turkey, 2018, pp. 1-4, doi: 10.1109/EBBT.2018.8391437.

*[25]*W. Wu, F. Han, G. Song and Z. Wang, "Music Genre Classification Using Independent Recurrent Neural Network," 2018 Chinese Automation Congress (CAC), Xi'an, China, 2018, pp. 192-195, doi: 10.1109/CAC.2018.8623623.

*[26]*R. Sarkar, N. Biswas and S. Chakraborty, "Music Genre Classification Using Frequency Domain Features," 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT), Kolkata, India, 2018, pp. 1-4, doi: 10.1109/EAIT.2018.8470441.

*[27]*R. L. Aguiar, Y. M. G. Costa and C. N. Silla, "Exploring Data Augmentation to Improve Music Genre Classification with ConvNets," 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 2018, pp. 1-8, doi: 10.1109/IJCNN.2018.8489166.

*[28]*S. -H. Chang, A. Abdul, J. Chen and H. -Y. Liao, "A personalized music recommendation system using convolutional neural networks approach," 2018 IEEE International Conference on Applied System Invention (ICASI), Chiba, Japan, 2018, pp. 47-49, doi: 10.1109/ICASI.2018.8394293.

*[29]*S. Panwar, A. Das, M. Roopaei and P. Rad, "A deep learning approach for mapping music genres," 2017 12th System of Systems Engineering Conference (SoSE), Waikoloa,

HI, USA, 2017, pp. 1-5, doi: 10.1109/SYSOSE.2017.7994970.

[30] F. Medhat, D. Chesmore and J. Robinson, "Automatic Classification of Music Genre Using Masked Conditional Neural Networks," 2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, 2017, pp. 979-984, doi: 10.1109/ICDM.2017.125.