Research Article GU J Sci, Part A, 11(1): 123-136 (2024) 10.54287/gujsa.1374878

Gazi University
Journal of Science
PART A: ENGINEERING AND INNOVATION
http://dergipark.org.tr/gujsa

Genres Classification of Popular Songs Listening by Using Keras

İlhan TARIMER[1*] Buse Cennet KARADAĞ[1]

[1] *Muğla Sıtkı Koçman University Faculty of Technology, Muğla, Türkiye*

Keywords

Musical Genre

MFCC

Keras

Classification

Abstract

Listening to the music affects the brain in ways which might help to promote the human health and arrange various diseases symptoms. Music is a phenomenon that is intertwined at every stage of human life. In the modern era music is shaped by the combination of an incredible number of genres, some of which are contemporary, and some come from the previous times. The music genre represents a collection of musical works that develop according to a certain shape, expression and technique. The music genre of interest varies from person to person in society. Most listeners today do not know what kind of music they listen to. In this study, sound features were extracted from music data and the Keras model was trained using these attributes. The correct classification rate of a music genre of the trained model was determined as 71.66%. Mel Frequency Cepstral Coefficients (MFCC), Mel Spectrogram, Chroma Vector and Tonnetz methods in the Librosa library were used to extract sound properties from music data. Using the features probed by means of the library, the most listened songs with Shazam in Türkiye were categorized in with TensorFlow/Keras. Many methods can be used in classification. It is uncertain which method the researchers should opt. It has been emphasized that classification of the genres of newly released songs by using Keras in this study. At result, it is said that the study has presented a sound processing are Keras classification of musical parts.

Cite

Tarimer, İ., & Karadağ, B. C. (2024). Genres Classification of Popular Songs Listening by Using Keras. *GU J Sci, Part A, 11*(1), 123-136. doi:10.54287/gujsa.1374878

Author ID (ORCID Number)

0000-0002-7274-5680 İlhan TARIMER
0000-0002-2488-1047 Buse Cennet KARADAĞ

Article Process

Submission Date	12.09.2023
Revision Date	10.11.2023
Accepted Date	08.02.2024
Published Date	15.02.2024

1. INTRODUCTION

Music is one kind of therapy that boosts our mood and happiness and might help the patients during treatments for some health disorders. Music is one of the indispensable elements of life, which is thought to exist with the existence of people in the world (İmik & Haşhaş, 2020). It is usually defined as the art of expressing certain feelings and thoughts with harmonious sounds within the framework of certain regulations. Music can also be defined as the art of bringing together measured and regulated sounds within artistic thoughts with or without rhythm to express feelings, thoughts or events.

It is known that music has an inevitable locus in human life. It is possible to come across music on television, in cinema films, in advertorials, in cafeterias, and in shopping malls. Some of these musical artifacts can be ones that we listen to and some that we have never heard. In the era in which we live, different types of digital platforms offering musical arranged or covered songs have been developed for people to listen to music and discover new music. The most well-known of these digital platforms are given as Spotify, Youtube Music and iTunes & Apple. By means of these platforms, people can easily access to their sideburn genres of music, and songs.

Shazam application was developed in 1999 by Chris Barton, and his team members, it was only released to use in the UK. It is Shazam which can listen to music currently playing in the environment within 5-10 seconds

*Corresponding Author, e-mail: itarimer@mu.edu.tr

and that find out who is the singer and which song has now been playing through the database, providing access to the lyrics of the song in the Lyrics section. The application makes these as to be listing of the searched songs in My Shazams section. The application which provides the opportunity to monitor the clip of the song on the video side, also permits to share the song on many platforms by using sharing button (Hussain et al., 2019).

Shazam finds music frequencies in fingerprint logic and executes them in the database. Then, the determined frequencies are searched and matched. The sound frequencies are firstly converted into a recognizable format by an algorithm. This process done is called as hashing. The data that they are received through a series of mathematical operations, produce complex outputs by using hash algorithms. This is considered as the most matched song frequencies among the frequencies coded via the hash system and its information is shared.

Acoustic fingerprint, as the name suggests, is a kind of identity that belongs solely to that sound. Each sound type contains sound and bandwidth values together in a frequency band. While the sounds we hear sometimes sound alike, on the other hand, the formulae reveal differences. In order to detect songs, acoustic fingerprint is constituted when identifying a sound, providing voice searching operations by searching among all sounds, and then, Shazam can detect songs searched (Web Source 1). Figure 1 displays operation of a Shazam application.

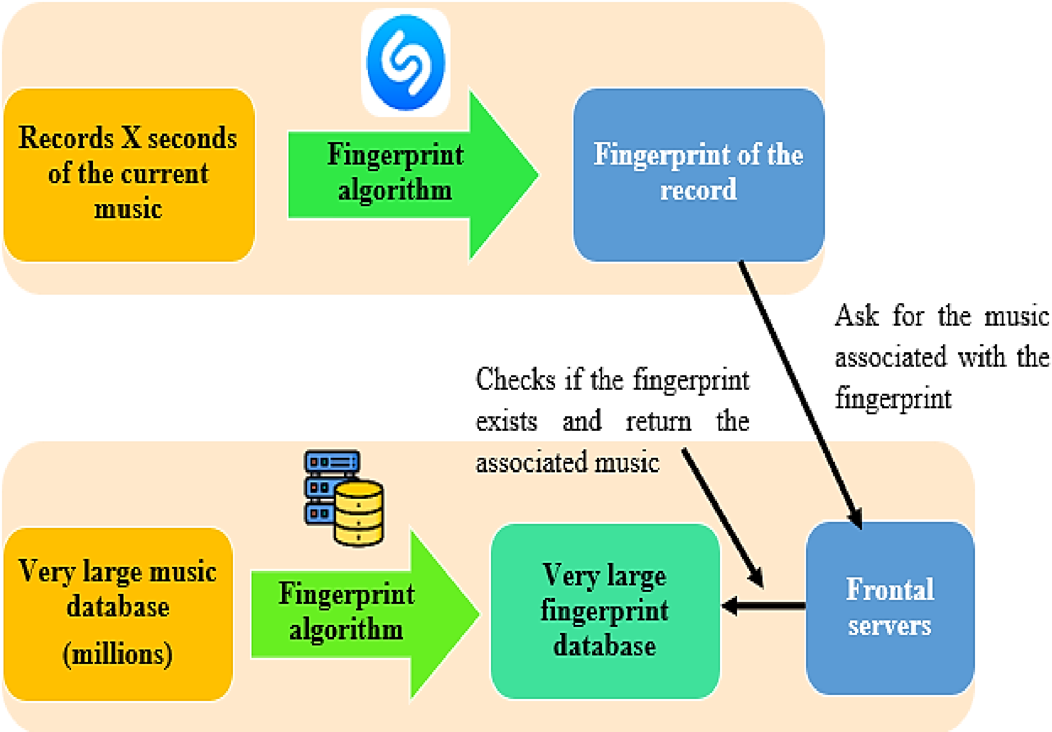


Figure 1. *Working Diagram of Shazam Application*

1.1. Related Works

The recently published papers point out that researches on the sound processing have increased and it indicates that this is an active research field that emerges new music genres. The authors in Tzanetakis & Cook (2020), pioneered the works on music genre classification using a machine learning algorithm. These two researchers created the GTZAN dataset and to date it has been considered that it is a standard for species classification. According to Bahuleyan (2018), an approach to automatic music classification was presented by providing tags to songs in the user’s library. In the study, it is understood that two separate approaches were used. The first is Convolutional Neural Network, the second is various Machine Learning algorithms. Both two approaches have been compared separately, and it has resulted to a conclusion that the VGG-16 CNN model gave the highest accuracy with an accuracy of 0.894. Pelchat & Gelowitz (2020), have studied the classification of musical genres. They have used neural networks to classify songs based upon their musical genres in the research. The images of the created spectrograms were used as input in the neural network from the time intervals of the songs. Vishnupriya & Meenakshi (2018) have conducted a study on the classification of musical genres. In their study, CNN was used training and classifying the genres; feature extraction and sound analysis were performed on the dataset. The proposed system separates the music into various genres by extracting the feature vector. The results show that the accuracy level of the system is around 76% and will widely thrive and facilitate automatically classification of music genres.

Nirmal & Mohan (2020), have proposed a method for classifying music by using spectrograms. In their papers, they approved that the music signals were first converted to their corresponding spectrograms. These spectrograms are then given as input to the classifier. It is seen that a CNN was used as the classifier. They evaluated performance of the classifier via performance measuring such as confusion matrix and classification accuracy. To do this, the GTZAN dataset was used as the dataset. Chillara et al. (2019), put forward multiple classification models. These models were trained using the Free Music Archive (FMA) dataset. A few of the classification models were trained on mel-spectrograms with the sound features of songs, while a few were trained solely on spectrograms of songs. It was found that the Convolutional Neural Network, which is one of the models in which only spectrograms are given as the dataset, gives the highest accuracy with 88.5% among all other models. Ghildiyal et al. (2020) have studied several classification models which have been established for the classification of musical genres. Most of the patterns were trained using the GTZAN dataset, and a few of the models were trained on the spectrogram. CNN model produced the best performance.

Zhang et al. (2019) have compared various classification algorithms and proposed a new near real-time classification model using RNN with a low accuracy of 64%. The authors used the mean and common variance of the MFCC to train their patterns. Yang & Zhang (2019), have extracted the mel-spectrogram of the fragments in the GTZAN dataset and used it as an input. The authors used the double convolution layer, where the output is passed through different pooling layers and a statistical analysis is done. Another paper determined MFCC and compared the CNN model with the Long ShortTerm Memory model (Gessle & Åkesson, 2019). The results showed that CNN had 56.0% prediction accuracy in GTZAN dataset and 50.5% prediction accuracy in FMA dataset, while LSTM model had 42.0% prediction accuracy in GTZAN dataset and 33.5% prediction accuracy in FMA dataset. From this paper, it is known that the CNN pattern has given better accuracy. Chen et al. (2019), have aimed to classify environmental sounds using the UrbanSound8K dataset. The researchers proposed a CNN-based model to classify environmental sounds in the dataset. Moreover, the researchers stated that they also examined the effect of the number of layers on the performance in this study. In this proposed model, an accuracy value of 78% was obtained.

Demir et al. (2020). have converted audio files to images by using the Short Time Fourier Transform to classify the environmental sounds in the UrbanSound8K dataset in their study. The researchers used pre-trained Convolutional Neural Networks architectures in the model proposed, and VGG16, VGG19 and Densenet201 architectures were used for feature extraction. The obtained feature maps were classified in the Support Vector Machines classifier; this pattern gave an accuracy value of 78.14%. Davish & Suresh (2018), have proposed a CNN-based model to classify environmental sounds in the UrbanSound8K dataset in the study. In their research, data augmentation techniques have been used to increase the model's performance. In addition, different augmentation methods were used in the study for determining the best augmentation technique for making environmental sound analysis. Researchers stated that LPCC is the most successful method among data augmentation methods, and an accuracy rate of 67.8% was seemed from the original dataset. Salamon et. al (2017), have used the UrbanSound8K dataset for classification of environmental sounds. In their study, they have stated that they made data augmentation due to the low environmental sounds. In the proposed CNN-based model, an accuracy value of 73% was obtained in the classification of environmental sounds.

1.2. Datasets

In our study, the most popular dataset in music genre classification, GTZAN was presented to utilize. The GTZAN dataset collection is used to classify music into different types. It has been used by Tzanetakis and Cook contained a series of records reflecting different situations, and they were 10 classes, each containing 100 different 30-second audio files with .wav extension (Tzanetakis & Cook, 2020). These classes consist of blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. Some of the sample audio files are seen in Figure 2.

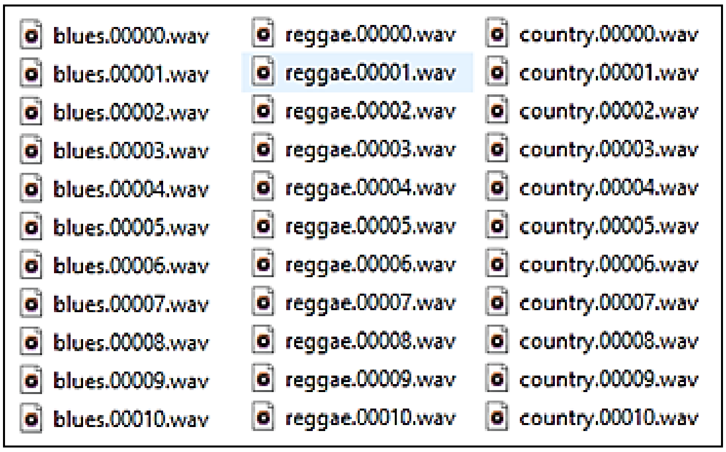


Figure 2. *Audio Files in the Dataset*

The remaining of the paper is organized as follows: Section 2 presents audio analysis methods used to extract new features from existing audio files. Section 3 presents and discusses the extraction of new features from various audio files in the existing dataset, the creation of a new dataset, and the experimental results of the proposed classification method. Chapter 4 concludes the findings, summarizes the paper and makes some prepositions and directions for possible researches in the future.

2. MATERIAL AND METHOD

The method used for classifying music genres in this chapter. Figure 3 displays the generalized block diagram of the proposed method.

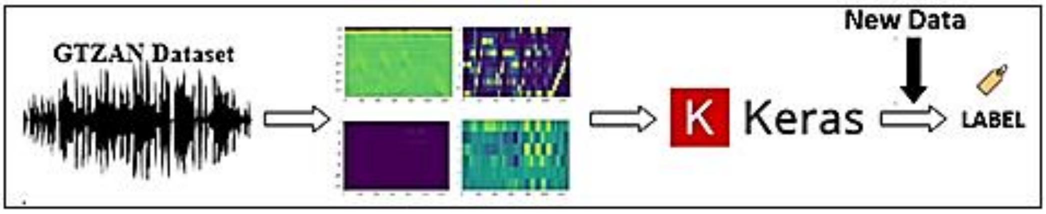


Figure 3. *Block Diagram of the Method*

Each type of music is described by own characteristics, pitch, melody, chord sequences and type of instrumentation. In order to have a reliable classification, a set of features ought to be used to capture essence of these elements. To divulge the features, the Librosa Package of Phyton has been used for analyzing the music and its sound. The library evaluates audio signal recordings. Mel frequency cepstral coefficients, mel spectogram, chroma vector and tonal centroid features are the features that will use to create a single feature vector for each file on which the model will be trained. These are explained in the below lines.

2.1. Audiotic Features

Mel Frequency Cepstrum Coefficients that is the most well-known, robust, accurate and established methods for feature extraction from audio signals are Cepstral coefficients are calculated by a discrete cosine transform applied to the power spectrum of a signal. The frequency bands of this spectrum are logarithmically spaced according to the Mel scale (Web Source 2). The procedures of the MFCC method are given in Figure 4. The training data set includes GTZAN music data. The MFCC algorithm splits an audio stream into frames by reforming it into smaller windows using the Hamming window. The Fourier transforms stage is used to transform the limited area of the audio signal into a frequency spectrum. The spectrum is generated for each frame using the fast fourier transform and each spectrum is weighted using a filter bank. At the last step, the MFCC vector is calculated by using the Logarithm and the Discrete Cosine Transform. In the cepstrum stage, the mel spectrum is converted to the time domain using DCT and produces the result called MFCC (Yıldırım, 2022).

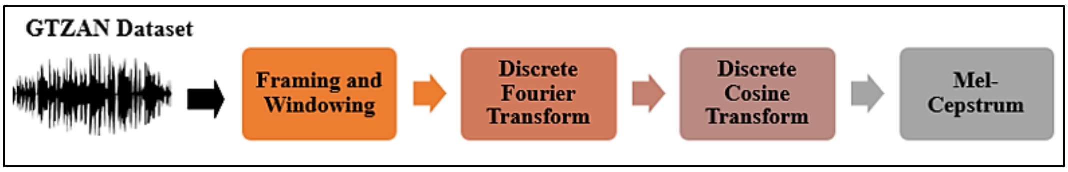


Figure 4. *Steps Involved in the MFCC Extraction*

The MFCC image of the 25th hip-hop audio file in the GTZAN dataset used is shown in Figure 5.

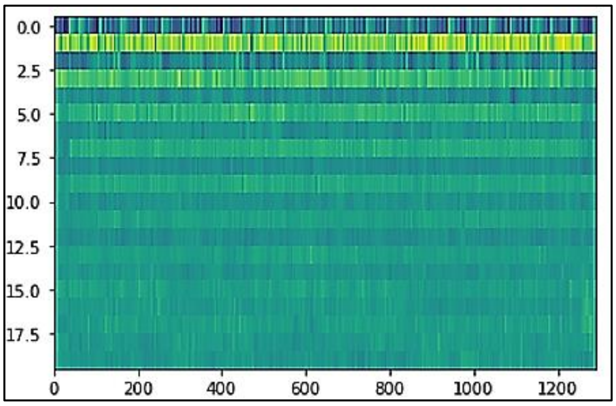


Figure 5. *MFCC of the File hiphop.00025.wav*

As human ear cannot perceive frequencies linearly, a perceptual scaling created according to the hearing feature of human ear, like Mel Spectogram is used for this purpose (Eray, 2008). MFCC is the expression of the short-time power spectrum of the audio signal on the Mel scale. While the unit of true frequency is Hertz, on the other hand Mel scale frequency is stated by the unit of Mel. Equation 1 defines the converting to Mel scaling from frequency (Karhan et al., 2016). 1 KHz is chosen as the reference and is accepted that it corresponds to 1000 mel (Eray, 2008).

$$mel(f) = 2595 * \log_{10}\left(1+\left(\frac{f}{700}\right)\right) \tag{1}$$

The view of Mel spectrogram of the 25th hip-hop sound file in the GTZAN dataset used is given in Figure 6.

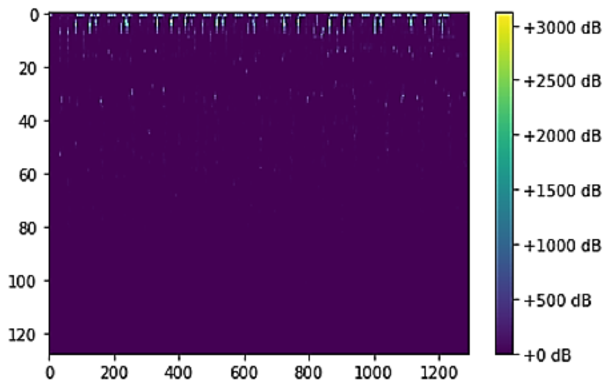


Figure 6. *Mel Spectrogram of the Audio File hiphop.00025.wav*

Chroma Vector (Chromagram) is a qualitative value that it determines a pitch grade quality referring to "colour" of a musical pitch. It can be dissociated into an octave-invariant value called chroma and pitch height presenting the octave the pitch is in the references (Kattel et al., 2019; Patil et al., 2017). It is a strong representation for sound in which 12 parts representing the 12 different halftones (chroma) of the spectrum musical octave are specified (Web source 3). These parts are C, C#, D, D#, E, F, F#, G, G#, A, A#, B. The chroma vector is a perceptually motivated feature vector, and uses the concept of chroma in the cyclic spiral representation of the musical pitch perception. Thus, the Chroma vector represents magnitudes in twelve pitch classes on a standard chromatic scale (Kattel et al., 2019). Chroma vector for the same sound sample is given in Figure 7.

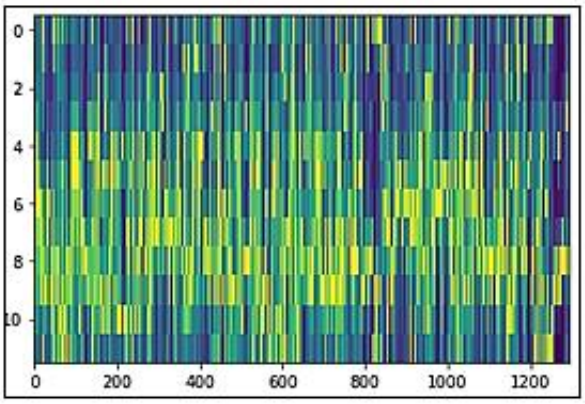


Figure 7. *Chroma Vector of the Audio File hiphop.00015.wav*

Tonal Centroid Features computes the tonal centroid features (tonnetz). This representation uses the method of 1 to project chroma features onto a 6-dimensional basis representing the perfect 5th, minor 3rd, and major 3rd each as two-dimensional coordinates (Web Source 4) (Harte et al., 2006). Figure 8 displays the tonnetz network.

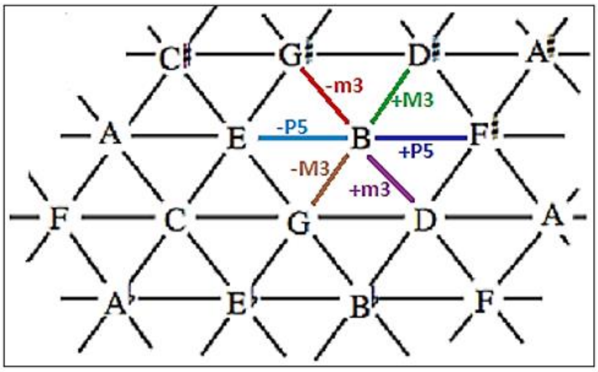


Figure 8. *The Harmonic Vector or the Tonnetz*

In the provided Tonnetz network, the connection shown in green and +M3 indicates minor 3rd, the connection labeled with dark blue and +P5 indicates perfect 5th, purple and the connection labeled with +m3 indicates major 3rd. The created Tonnetz network for Chroma vector of the same sound file is given in Figure 9.

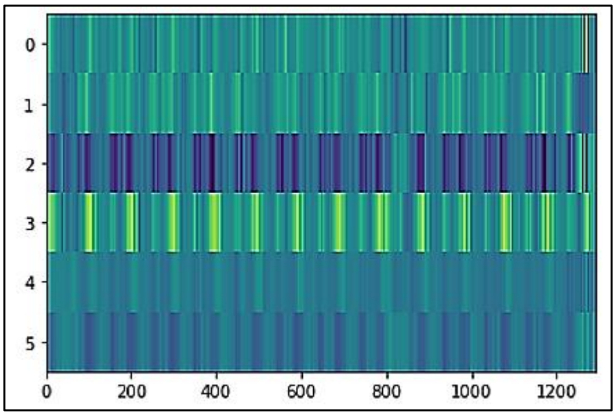


Figure 9. *Tonnetz Feature of the Audio File hiphop.00025.wav*

GU J Sci, Part A 11(1) 123-136 (2024) 10.54287/gujsa.1374878

2.2. Keras

As TensorFlow is an open-source math library for Deep Learning tasks, Deep Learning, which is a subset of the Machine Learning cluster, processes new data from the existed ones. Keras is a Deep Learning library and runs on TensorFlow. While it is written library in Phyton, it does not define almost any Deep Learning model. It is a high-level neural networks API that can work not only on TensorFlow but also on other Theano and CNTK. Creating a deep learning model with Keras is done quietly easy. The procedural steps can be summarized: Defining the trained data, defining layers–model, hyper parameters such as epoch, loss function and optimizer.

3. RESULTS AND DISCUSSION

In this chapter, the sounds in the current dataset are analyzed with the functions provided by the Librosa library. A new dataset is created from the data obtained as a result of the analysis. The new dataset created is used for the training Keras. After training the algorithm, the most searched songs in Türkiye are given as input to the algorithm by using Shazam and it is provided to estimate the type of music.

3.1. Creating a New Dataset with Audiotic Analysis

To extract MFCC, Mel Spectrogram, Chroma Vector and Tonnetz features from the audio files withdrawn from the GTZAN dataset, four functions are created as same name as the above. These functions hold the extracted features of each audio files in separate sequences, are combined using the ***concatenate()*** function available in the Numpy library. In the last generated dataset array, each element represents an audio file. The dataset includes the features of 1000 audio files in total. The feature set of the extracted and combined features from the audio file is given in Figure 10, and the labeled version of each audio file is given in Figure 11.

```
features

array([[0.41249344, 0.35866404, 0.470512  , ..., 0.36730251, 0.14873096,
        0.15761952],
       [0.47382188, 0.38117659, 0.41266283, ..., 0.41842549, 0.12306178,
        0.18224726],
       [0.54743522, 0.40536609, 0.48203942, ..., 0.28925934, 0.04000752,
        0.07743137],
       ...,
```

Figure 10. *View of Analyzed Audio Files in Dataset*

```
array([8, 8, 8, 3, 8, 3, 3, 7, 4, 9, 5, 3, 2, 3, 6, 9, 9, 3, 5, 2, 5, 0,
       2, 1, 3, 1, 6, 4, 7, 4, 4, 9, 0, 6, 8, 6, 8, 9, 1, 7, 7, 6, 1, 3,
       5, 7, 8, 3, 3, 3, 5, 2, 3, 4, 7, 1, 8, 9, 3, 3, 0, 2, 6, 5, 1, 7,
       4, 0, 5, 4, 4, 3, 3, 5, 6, 9, 7, 4, 6, 0, 7, 9, 3, 0, 2, 8, 5, 8,
       8, 9, 4, 8, 7, 8, 1, 5, 4, 5, 6, 7, 2, 8, 2, 5, 8, 1, 5, 7, 1, 8,
       4, 7, 6, 5, 2, 2, 5, 6, 1, 8, 2, 8, 2, 6, 2, 0, 2, 5, 0, 6, 5, 0,
```

Figure 11. *View of Labeled Audio Files in Dataset*

The numbers in the sequence given in Figure 11 represent the music genres included in the GTZAN dataset. 0 blues, 1 classical, 2 country, 3 disco, 4 hip-hop, 5 jazz, 6 metal, 7 pop, 8 reggae and 9 rock.

3.2. Splitting the Dataset

Data splitting is to divide data into two or more subsets. In a two-part partition, one part is used to evaluate or test the data and the other to train the model. The random**.**permutation function of the Numpy library is used to mix the recordings after the property set of the audio files is created so that the model had information about each music genre. This function is used to randomly change an array. During training the model, the dataset is splatted as 60% for training, 20% for validation and 20% for testing. As the result, 600 audio files for training, 200 for validation and 200 for testing are used.

3.3. Training and Evaluation of the Model

The steps how to train and evaluate Keras data are given in Figure 12. First, the model is created While creating the model, some metrics ought to determine. These metrics are a function used to evaluate performance of the model, and are given as arguments to the ***compile()*** function. They are optimizer, accuracy metrics, probability metrics, regression metrics, based on positive and negative classification metrics. The mentioned above metrics configure training data. The trained model obtains the test data by making inferences.

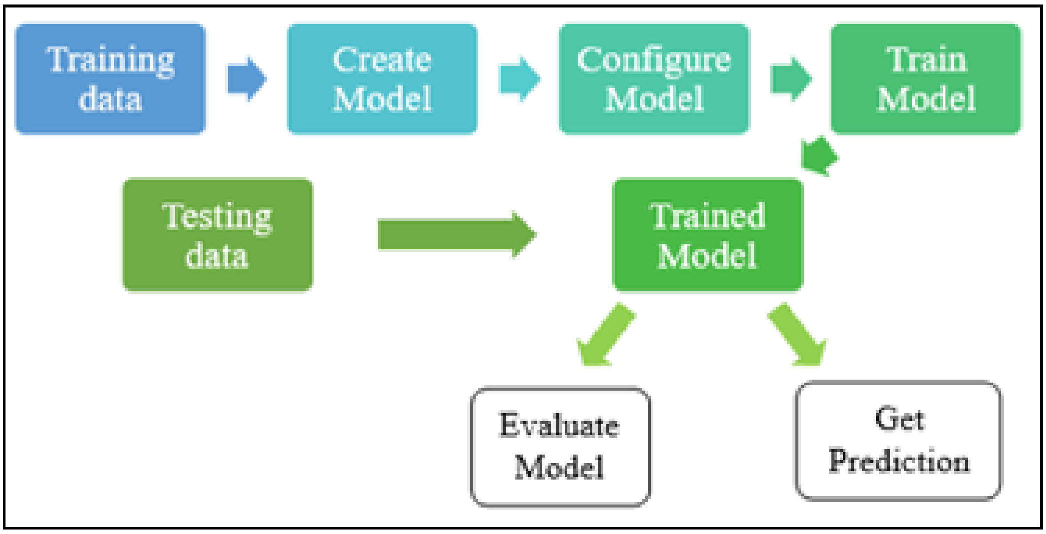


Figure 12. *Keras Algorithm Training Steps*

Two regular densely connected neural network layers which are with a rectified linear unit activation function "relu", and 300 hidden units (the first layer) and 200 (the second layer) are implemented by using Keras for this model. Then, heavily correlated layer with the probability distribution activation function "softmax" is also applied for the output layer. Why Nadam is chosen as the optimizer is chosen is that this give the highest accuracy value in the trials. Next, the model is trained using 64 epochs. The trained Keras is evaluated after the values are obtained. The resulted set of features achieves around 71,66% classification rate.

3.4. Finding the Genre of a New Audio File

The new audio file is withdrawn from the most wanted songs in Türkiye list offering by Shazam, and given in Table 1.

Table 1. *Top 10 shazamed songs in Türkiye*

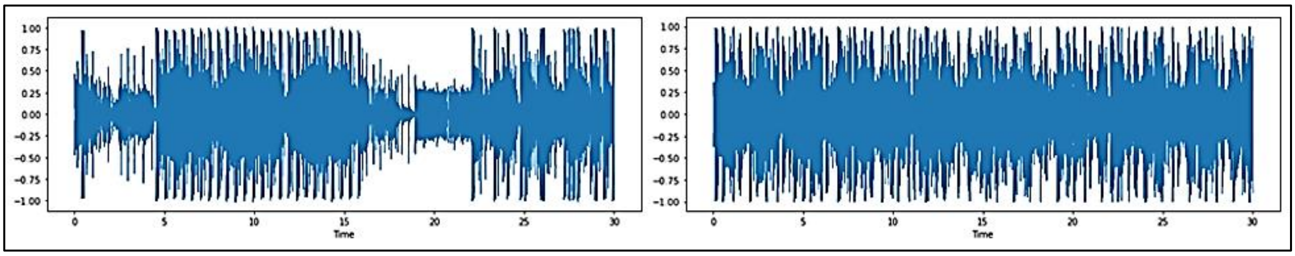
Rank	Artist	Title	Genre
1	Lvbel C5	Dacia	Reggae
2	DEHA INC. & Reckol	Talibana	Hip-hop
3	Heijan & Muti	İmparator	Hip-hop
4	BLOK3	Affetmem	Hip-hop
5	Metro Boomin, The Weeknd & 21 Savage	Creepin'	Hip-hop
6	Güneş	NKBİ	Hip-hop
7	Alok & Alan Walker	Headlights (feat. KIDDO)	Pop
8	Lady Gaga	Bloody Mary	Pop
9	Mert Demir & Mabel Matiz	Antidepresan	Pop
10	Dj Belite	All Eyez On Me	Hip-hop

The audio file of the songs given in Table 1 ought to be created for determining the type of songs. The Youtube link of this song can be used for this aim. The library is used to generate the audio file over the youtube link by determining the metrics required to use. The metrics consist of information of audio file's name in which will be recorded with and what its extension will be. By means of these metrics, the model is established and the audio file is downloaded using the YouTube link of the song. It ought to be noticed that not all the size of the downloaded audio file is used in estimating the genre of the song. Therefore, the pydub library is used to trim the file size. A 30-second portion of the new audio files from 01:00:00 to 01:30:00 is taken since the audio files in the GTZAN dataset are 30 seconds long. The resulting audio files are recorded with a different name. The created sound files are given as input to the Keras model. Genres predicted are given in Table 2.

Table 2. *Top 10 shazamed songs in Türkiye*

Rank	Artist	Title	Genre	Estimated
1	Lvbel C5	Dacia	Reggae	Pop
2	DEHA INC. & Reckol	Talibana	Hip-hop	Hip-hop
3	Heijan & Muti	İmparator	Hip-hop	Hip-hop
4	BLOK3	Affetmem	Hip-hop	Hip-hop
5	Metro Boomin, The Weeknd & 21 Savage	Creepin'	Hip-hop	Pop
6	Güneş	NKBİ	Hip-hop	Hip-hop
7	Alok & Alan Walker	Headlights (feat. KIDDO)	Pop	Pop
8	Lady Gaga	Bloody Mary	Pop	Hip-hop
9	Mert Demir & Mabel Matiz	Antidepresan	Pop	Reggae
10	Dj Belite	All Eyez On Me	Hip-hop	Hip-hop

The graphics of the Talibana song estimated correctly in hip-hop by the established model, and the graphics of 50th sample sound file in the hip-hop genre in the GTZAN dataset are given below. The signal graph of the song named "Taliban", which the model predicts correctly, and the signal graph of the 50th sound file in the hip-hop genre in the sample dataset are shown in Figure 13.

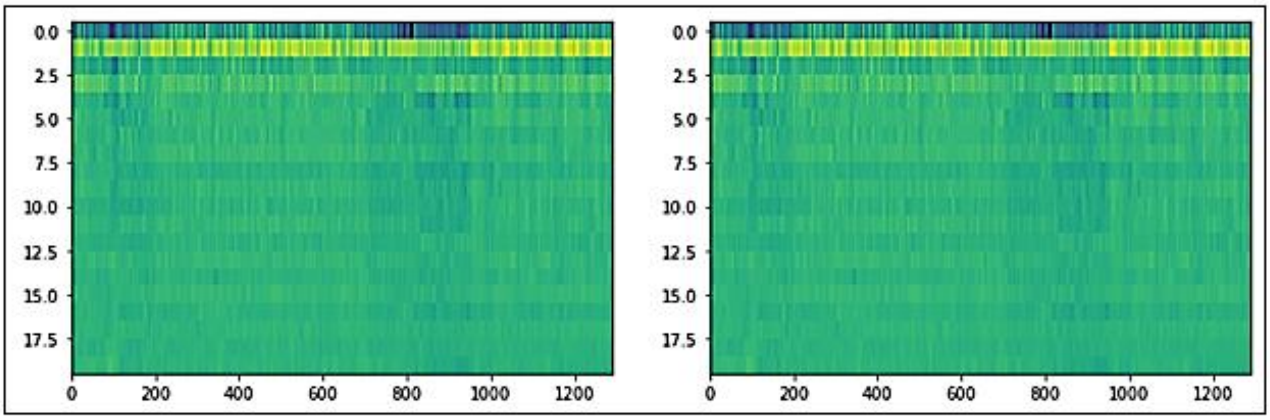


a. *Talibana Song* ***b.*** *Hip-hop 50. example*

Figure 13. *Signal Graph,* ***a)*** *Talibana Song,* ***b)*** *Hip-hop 50. example*

The MFCC graph of the song called Talibana, which the model predicts correctly, and the MFCC graph of the 50th sound file in the hip-hop genre in the sample dataset are seen in Figure 14. The Chromagram graph of the song called Talibana, which the model predicts correctly, and the Chromagram graph of the 50th sound file in the hip-hop genre in the sample dataset are seen in Figure 15. The signal frequencies of the song titled Headlights, which the established model estimated correctly in the pop genre, and the signal frequencies of the 35th sample sound file in the pop genre in the GTZAN dataset are seen in Figure 16. The MFCC signal frequencies of the song titled Headlights, and the MFCC graph of the 35th pop sound file are seen in Figure 17.

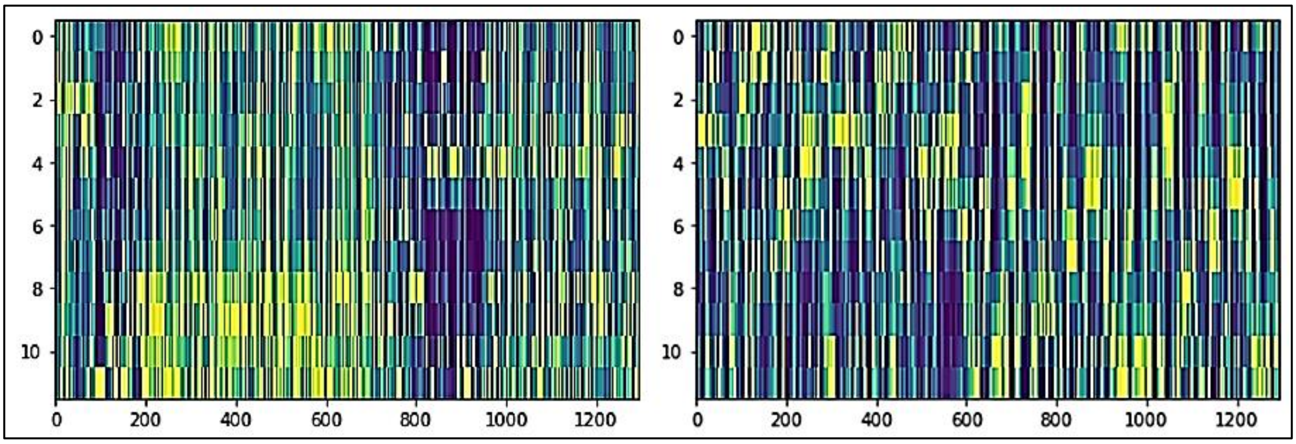
İlhan TARIMER, Buse Cennet KARADAĞ
GU J Sci, Part A 11(1) 123-136 (2024) 10.54287/gujsa.1374878



a. *Talibana Song* ***b.*** *Hip-hop 50. example*

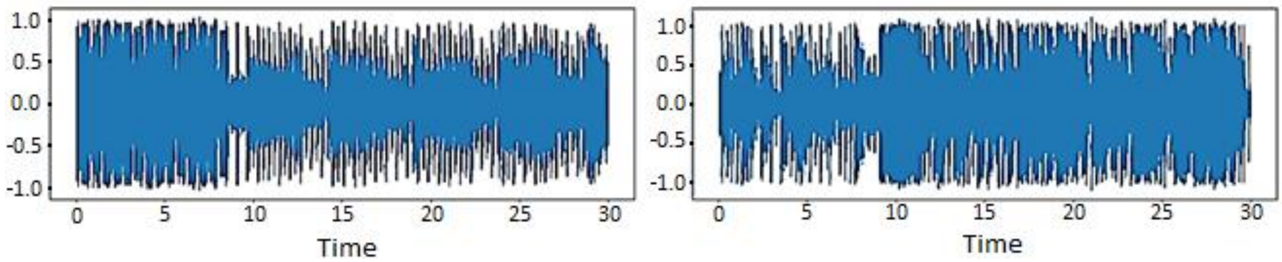
Figure 14. *MFCC Graph,* ***a)*** *Talibana Song,* ***b)*** *Hip-hop 50. example*

In Figure 15, the Chromagram graph of the song called Talibana, which the model predicts correctly, and the Chromagram graph of the 50th sound file in the hip-hop genre in the sample dataset are given. Figure 16 and Figure 17 show signal graphics and MFCC graphics.



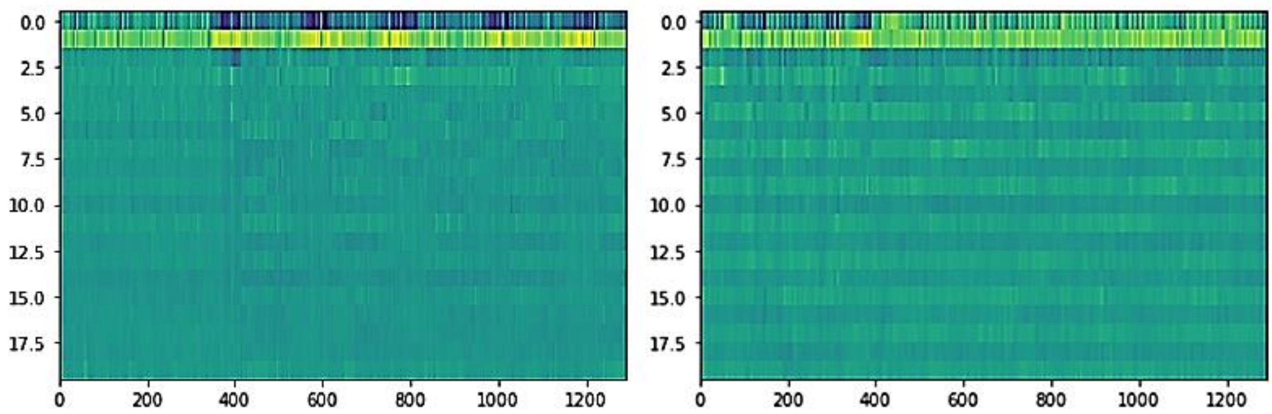
a. *Talibana Song* ***b.*** *Hip-hop 50. example*

Figure 15. *Chromagram Graph,* ***a)*** *Talibana Song,* ***b)*** *Hip-hop 50. example*



a. *Headlights Song* ***b.*** *Pop 35. example*

Figure 16. *Signal Graph,* ***a)*** *Headlights Song,* ***b)*** *Pop 35. example*



a. *Headlights Song* ***b.*** *Pop 35. example*

Figure 17. *MFCC Graph,* ***a)*** *Headlights Song,* ***b)*** *Pop 35. example*

In this step, we benchmark our proposed system with other existing systems. Comparison amid previous studies and our proposed system is presented in Table 3.

Table 3 indicates that all studies which divide the existing dataset into two parts as training and testing data. A comparison related to the accuracy values among the established models is made. New sound files are not given as input to the established models and no genre prediction is made. The study deals with the established model that it is trained with the existing dataset, and is calculated how accurately it could predict new audio files with the data reserved for validation in the dataset. It has been calculated that the estimated model can predict a new audio file with an accuracy rate of 71.66%.

4. CONCLUSION

The paper examines several scientific modelling techniques presenting music genre classification. In order to extract new attributes of the audio files in the used dataset, the Librosa library has been used. A new dataset has been created by combining the new features. This model was created and trained with Tensorflow / Keras to find the music genre that audio files are entered externally. The dataset is being mixed for training the model. Hence, it is assured that the model had information about music genres. The 60% of the dataset is used for training the model, 20% for testing and 20% for validation.

For realizing this model, the corrected linear unit activation function called as relu is used. two normative connected neural network layers are obtained with 300 hidden units for the first and 200 hidden units for the second layer. followed by, a layer closely related to the probability dispersion activation function softmax is also applied for the output layer**.** As Nadam gives the highest accuracy in try-outs, it has been chosen for optimizing element. Next, the model is trained by using 64 epochs. The success rate of the trained model in predicting a new audio file is determined as 71,66%. In order to find out whether the trained model correctly predicted the music genre, the top 10 most searched songs in Türkiye by shazam are used. On the other hand, to estimate the genre of the popular song, the audio file of the song is created first by youtube-dl library is used. The resulting sound file is transmitted to the Keras model for the prediction of the music genre. The Keras model correctly estimates 60% of the new audio files given.

To realize this study, machine learning algorithms are used. Nevertheless, different machine learning algorithms can be using to classify musical genres in future studies. We propose that both the dataset is increased and the research is furthered by adding sample audio files values in the current one.

GU J Sci, Part A 11(1) 123-136 (2024) 10.54287/gujsa.1374878

Table 3. *Comparison of the Proposed System with Similar Studies*

Research, Year	Datasets	Algorithms Used	Results
(Vishnupriy et al., 2018)	GTZAN dataset. Spectrogram images of musics.	Convolutional Neural Network	The system separates the music into various genres by extracting the feature vector. The results show that the accuracy level of the system is around 76%.
(Chillara et al., 2019)	Free Music Archive dataset. Spectrogram images of musics.	CNN–RNN Logistic Regression Simple Artificial Neural Network	The classification accuracies of the CNN model and CRNN are 88.54% and 53.5% respectively. The classification accuracies of the feature-based models LR and ANN are 60.89% and 64.06% respectively.
(Zhang et al., 2019)	GTZAN dataset MFCC features	Support Vector Classification Recurrent Neural Network	They concluded that the short-term MFCC features of the RNN can achieve sufficient classification accuracy. By listening to a piece of music for 0.5 seconds, they determined its genre with 64% accuracy.
(Yang et al., 2019)	GTZAN dataset Mel-Spectrogram features	Convolutional Neural Network	Two different networks, net1 and net2, were created. net1 network achieved 90.7% accuracy, net2 88.3% accuracy.
(Gessle et al., 2019)	GTZAN dataset FMA dataset MFCC features	(CNN) LSTM-Keras	CNN achieved 56% accuracy, and in the FMA dataset, 50.5% accuracy. LSTM achieved 42% accuracy, and in the FMA dataset, 33.5% accuracy.
(Nirmal et al., 2020)	GTZAN dataset. Spectrogram images of musics.	Convolutional Neural Network Pre-trained convnet	Two CNN models are discussed: A user-defined CNN model and a pre-trained convnet. Three music genres (blues, classical and rock) from the GTZAN dataset are selected for testing. The classification accuracies of user-defined CNN model and MobileNet are 40% and 67% respectively.
(Nirmal et al., 2020)	Two music libraries. 1880 songs categorized into 10 genres.	Convolutional Neural Network	Each of these spectrograms was labeled by music genre and then used as inputs in a CNN. The results were obtained 85% accurately during testing data.
(Ghildiyal et al., 2020)	GTZAN dataset	Simple Artificial Neural Network Support Vector Machines Multilayer Perceptron Desicion Tree CNN	The proposed research work has compared few classification models. Among the models created, CNN showed the highest accuracy. The classification accuracy of the model is 91%.
(Proposed Research, 2024)	GTZAN dataset Spectrogram images of musics.	Keras	Keras model is trained by using GTZAN dataset. Therefore, the genres of 10 popular songs in Türkiye are estimated by Shazam application. The Keras model estimates types of the songs with 71.66% accuracy.

ACKNOWLEDGEMENT

As the authors of this paper, we would like to thank to Muğla Sıtkı Koçman University for supporting this research to do.

AUTHOR CONTRIBUTIONS

Conceptualization, methodology, manuscript-review, editing supervision, Tarımer İ.; field and laboratory works, sources, data curation and visualization Karadağ B.C.; research, software, validation, formal analysis, manuscript-original draft and funding, both authors. All authors have read and legally accepted the final version of the article published in the journal.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Bahuleyan, H. (2018). Music Genre Classification Using Machine Learning Techniques, *ArXiv:1804.01149v1*, *1*(1). https://doi.org/10.48550/arXiv.1804.01149

Chen, Y., Guo, Q., Liang, X., Wang, J., Qian, Y. (2019). Environmental Sound Classification with Dilated Convolutions, *Elsevier Applied Acoustics*, *148*(1). 123-132, https://dig.sxu.edu.cn/docs/2019-03/2d147cee7cfb4f3eb25ee73f6e6dd3de.pdf

Chillara, S., Kavitha, A. S., Neginhal, S. A., Haldia, S., Vidyullatha, K. S. (2019). Music Genre Classification Using Machine Learning Algorithms: A Comparison. *International Research Journal of Engineering and Technology*, *6*(5), 851-858, https://www.irjet.net/archives/V6/i5/IRJET-V6I5174.pdf

Davis, N., Suresh, K. (2018). Environmental sound classification using deep convolutional neural networks and data augmentation, *IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, (pp. 41-45). https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8635051&tag=1

Demir, F., Türkoğlu, M., Aslan, M., Sengur, A. (2020). A New Pyramidal Concatenated CNN Approach For Environmental Sound Classification, *Elsevier*, *170*(2020), 7. https://www.sciencedirect.com/science/article/pii/S0003682X20306241

Eray, O. (2008). The speech recognition application with support vector machines, *Pamukkale University*, (pp. 1-90). https://gcris.pau.edu.tr/handle/11499/1501

Gessle, G., & Åkesson, S. (2019). A comparative analysis of CNN and LSTM for music genre classification, Degree Project In Technology 2022 https://www.diva-ortal.org/smash/get/diva2:1354738/FULLTEXT01.pdf

Ghildiyal, A., Singh, K., Sharma, S. (2020). Music genre classification using machine learning, *International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, (pp. 1368-1372). https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9297444

Harte, C., Sandler, M., Gasser, M. (2006). Detecting harmonic change in musical audio, *1st ACM Workshop on Audio and Music Computing Multimedia (AMCMM)*, (pp. 21-26). https://dl.acm.org/doi/10.1145/1178723.1178727

Hussain, A., Mkpojiogu, E., Almazini, H., Almazini, H. (2019). Assessing the usability of Shazam mobile app, *The 2nd Internatİonal Conference on Applied Science and Technology (ICAST'17)*, (pp. 0200571– 0200575). https://pubs.aip.org/aip/acp/article/1891/1/020057/886329/Assessing-the-usability-of-Shazam-mobile-app

İmik, U., Haşhaş, S. (2020). What is Music and Where Is It In Our Lives?, *İnönü University Journal of Culture and Art*, *6*(2), 196-202, https://dergipark.org.tr/en/download/article-file/1491772

Karhan, M., Çakır, M., Uğur, M. (2016). Analysis of Electrical Discharge Sound Data Using MEL Frequency Cepstral Coefficients (MFCC), *In Proceeding of the 2016 cigrė Türkiye*, (pp. 4). http://gsk.cigreturkiye.org.tr/bildiriler2016/5.2.pdf

Kattel, M., Nepal, A., Shah, A. K., & Shrestha, D. (2019). Chroma feature extraction, *Conference: Chroma Feature Extraction Using Fourier Transform*, (pp 1-4). https://www.researchgate.net/publication/330796993_Chroma_Feature_Extraction

Nirmal M. R., & Shajee Mohan B. S. (2020). Music Genre Classification using spectrograms, *International Conference on Power, Instrumentation, Control and Computing (PICC)*, (pp. 1-5). https://ieeexplore.ieee.org/document/9362364

Pelchat, N., & Gelowitz, C. M. (2020). Neural Network Music Genre Classification, *Canadian Journal of Electrical and Computer Engineering*, *43*(3), 170-173. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9165253

Patil, N. M., Nemade, & M. U. (2017). Music Genre Classification Using MFCC, K-NN and SVM Classifier, *International Journal of Computer Engineering In Research Trends*, *4*(2), 43-47. https://www.ijcert.org/index.php/ijcert/article/view/371/313

Salamon, J., & Bello, J. P. (2017). Deep convolutional neural networks and data augmentation for environmental sound classification, *IEEE Signal Processing Letters*, (pp. 279-283). https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7829341

Yang, H., & Zhang, W. Q. (2019). Music genre classification using duplicated convolutional layers in neural networks, *INTERSPEECH*, (pp. 3382-3386). https://www.isca-archive.org/interspeech_2019/yang19f_interspeech.html

Yıldırım, M. (2022). Automatic Classification of Environmental Sounds with the MFCC Method and the Proposed Deep Model, *Fırat University Journal of Engineering Sciences*, *34*(1), 449-457. https://dergipark.org.tr/en/download/article-file/2186156

Tzanetakis, G., Cook, P. (2020). Musical Genre Classification of Audio Signals, *IEEE Transactions on Speech and Audio Processing*, *10*(5), 293-302. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1021072

Vishnupriya, S., & Meenakshi, K. (2018). Automatic music genre classification using convolution neural network, *International Conference on Computer Communication and Informatics (ICCCI)*, (pp. 1-4.). https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8441340

Web Source 1: (Shazam Nasıl Çalışıyor?). (2024). (Accessesd: 12.02.2024) https://musiconline.co/tr/blog/shazam-nasil-calisiyor

Web Source 2: (Music genre classification using Librosa and Tensorflow/Keras). (2024). (Accessesd: 12.02.2024) https://musiconline.co/tr/blog/shazam-nasil-calisiyor

Web Source 3: (Librosa). (2024). (Accessesd: 12.02.2024) https://medium.com/datarunner/librosa-9729c09ecf7a

Web Source 4: (librosa.feature.tonnetz). (2024). (Accessesd: 12.02.2024) https://librosa.org/doc/main/generated/librosa.feature.tonnetz.html

Zhang, S., Gu, H., & Li, R. (2019). Music Genre Classification: Near-Realtime vs Sequential Approach, *PrePrint*. (pp. 6). https://cs229.stanford.edu/proj2019spr/report/3.pdf