Classification of music genre using data augmentation in neural network based on Sports universities data

Sejong[1]

Abstracts

The field of artificial intelligence is among the most promising in AI. To train Artificial Intelligence algorithms, a significant quantity of data is required. We employ Sports colleges data augmentation in a neural network based on the real categorization of music genres by utilizing Sound Smith's algorithm. It can categorize music genres with a 60 percent accuracy and employs millions upon millions of audio samples. In numerous real-world applications, the categorization of genres is a crucial undertaking. The need for accurate meta-data for database management and search/storage purposes increases as the amount of music released daily continues to rise, particularly on internet platforms like YouTube music and spottily (a 2016 estimate suggests that tens of millions of songs were released every month on Playlists). The classification of musical genres is based on characteristics taken from a database of football players from sports universities. The classification was accomplished using Deep Learning. Specific musical features are derived from the acoustic waves generated by football players during games. Convolutional Neural Network is utilized for classification (CNN). The classification of musical genres is based on the technology used by football players. The characteristics are taken from the auditory waves produced by football players during games. A comparison is made between the actual categorization procedure and the classification created by CNN. The comparison demonstrates that the conventional approach, which uses classical features, provides a more accurate classification than the CNN-generated classification based on music features.

Keywords: Real Classification, Music Genre, using Data Augmentation, Neural Networks. Sports universities data

1. Introduction

Following the harrowing catastrophe that followed a soccer match in 1994, the sport has gained worldwide popularity. From the late 19th century to the 1930s, soccer encountered challenges in the United States. Nevertheless, according to FIFA's official website, it is the most popular sport today:
"More than 265 million people play soccer globally, and it is played in almost every country in the world".
According to statistics from a website called "Soccer Statistics", an average of more than 37,227 watched every match. It is also estimated that football will reach about 1.6 billion fans by 2018 and around 2 billion by 2026. Nowadays, it is essential to analyze and predict soccer data to help us exploit better soccer trading opportunities. We selected the Football dataset as our primary source of analysis due to its rich characteristics of different football players' attributes, such as their skills, goals scored, assists provided, and passes made.
This paper analyzes how people classify different types of music genres based on their favorite football players. We use Tensor flow as our machine learning platform and have collected data from 46,064 users of the Sound Cloud website who listened different types of music and mapped their favorite football players for every country.
We reported 46,064 users based on the following number from the sound cloud website. This was utilized for training our neural network, and as a consequence, we discovered that people in Brazil appear to enjoy rock music, which explains why their favorite football players are either Milan or Ronaldo or someone else. We demonstrate a correlation between football players and music genres for each country in the world to dissect the issue further. In the context of a music retrieval system, structuring a vast quantity of music-related data is crucial. Music data can be archived systematically based on metadata such as genre and artist. Manually annotating a piece of music by a domain expert is one technique for extracting such metadata from players. For a long time, computational approaches have been introduced to classify music genres. Tzanetakis and Cook (2002) used the mixture of Minutia-based model and k-nearest neighborhood and three sets of meticulously hand-extracted features spanning tumbrel texture, rhythmic content, and pitch content in a study published in 2002. They reached a 61 percent accuracy rate (Mohamed, 2017). Music is the most famous art performed and listened to by billions of people every day. There are numerous musical genres, including metal, pop, jazz, classical, reggae, blues, disco, and hip hop. To classify a music sample or song manually, one must listen to the song and select the genre. This time-consuming activity requires knowledge of diverse musical genres, which is extremely tough given the millions of songs that exist today (Yanchenko, 2017). The

[1] University the Graduate School Department of Music Applied Music, Seoul, Korea, 05006
lixiaoyi@sju.ac.kr

automation of music classification expedites the discovery of vital information such as trends, popular genres, and performers. The primary advantage of classifying music by genre is the ability to build a play list of comparable genres. Gaussian Mixture Models (GMM), k-Nearest Neighbors (k-NN), Neural Networks (NN), Support Vector Machines (SVM), and Hidden Markov Models are the most often employed classification techniques for this audio class recognition (HMM). Texture, rhythm, and pitch are proposed to be represented by three distinct feature sets. Feature vectors were generated using Short-time Fourier Transform (STFT), Mel frequency cepstral coefficients (MFCCs), Wavelet Transform (WT), and a few more parameters (Mohamed, 2017).

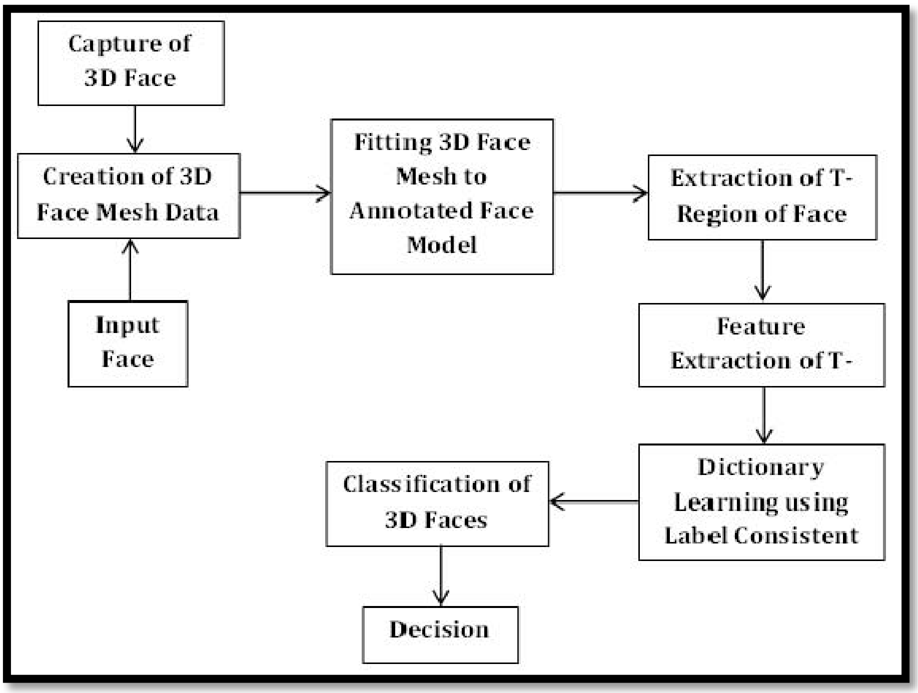

Figure 1. Block diagram of the proposed method

The principal objective of this paper is to characterize music sorts utilizing profound learning methods for students of sports universities, particularly CNNs and RCNNs. Since CNNs are sed for picture characterization, the sound information is changed over to spectrogram pictures and sent as aontribution to the model (Sethares & Toussaint, 2015). We performed information expansion on the crude waveform to expand the exhibition of the model. As development expands the dataset size, which is based on sports universities data of Korean universities, the model is prepared and tried on many examples, so the speculation capacity of the model increments with increase Figure 1 shows the general square graph of our methodology. This work utilizes Recurrent Convolutional Neural Network (RCNN) to display both fleeting and recurrence conditions of the music's spectrogram proficiently (Tzanetakis & Cook, 2002).

The primary commitments of this work are the following:

1. Utilizing the recurrence and transient characteristics of the spectrograms, this study focuses on how CNNs and RNNs are helpful for music classification based on data from sports universities, utilizing the recurrence and transient conditions of the spectrograms.
2. Observing how CNNs and RNNs classified spectrograms into distinct categories with high precision using different frameworks (Elbir et al., 2018).
3. Performed data enhancement on the raw data to improve precision and demonstrated how accuracy boosted the growth of each data enhancement approach.

1. Genre Based Classification

Sports and music are two of the few things that most people enjoy. If you will, the two hobbies (or passions) share similar traits and characteristics, which often go hand-in-hand. Both require a competitive nature, and both contain a sense of community among the viewers and participants. Both have been proven to bring hope to those that are feeling down on their luck--the list goes on. Here we will discuss how these aspects may contribute to why sports love music so much and vice versa.

One of the most common ways for musicians to gain recognition is by providing entertainment at sporting events. This occurs most commonly in the United States, where DJ AM and Kanye West frequently perform at football games, baseball stadiums, and other athletic events. Jay-Z and Rihanna's performance during the NBA Finals victory parade in 2005, following the team's triumph over the Dallas Mavericks, is one of the most memorable occurrences. In 2004, they also performed during a soccer match between Manchester United and Arsenal F.C. at the Vodafone Arena. In contrast, pop artists such as Madonna and Britney Spears often do not participate in sporting events. Their performances on the field have been limited to music concerts and other non-sporting activities. Another common way that rock stars get involved in sports is by visiting locker rooms and meeting with athletes. Many musicians are avid fans of a sport, and this often leads them to come out to watch their favorite team play and even take part in the game. This was a common sight during Bill Clinton's presidency when baseball players would frequent First Lady Hillary Clinton's political events and vice versa. However, this was also prevalent in the social milieu of the 1960s, when sportsmen and coaches engaged several rock stars in the strongly supported anti-war movement. Such examples include Jimi Hendrix playing for the American League Champion New York Yankees in a pre-game ceremony and John Lennon guest-starring on Monday Night Football for a few weeks during his time with The Beatles. More recently, Oasis would famously attend Manchester United games and make appearances at their events.

Historically, musicians have had a soft spot for sports teams because they frequently identify with the players. In England, for instance, it is more inexpensive for musicians to support lower-tier football clubs than teams like Manchester United or Arsenal F.C. Genre is a common classification for identifying a piece of music that belongs to a shared tradition. It relates to numerous aspects of music. There are multiple ways to classify musical genres. Such as the period and musical style in which musical composition was composed. The instruments utilized in the music and their respective treatments. The geographical origin of the music and its cultural and ethnic background play a significant effect in determining a music clip's genre (T. Li et al., 2003).

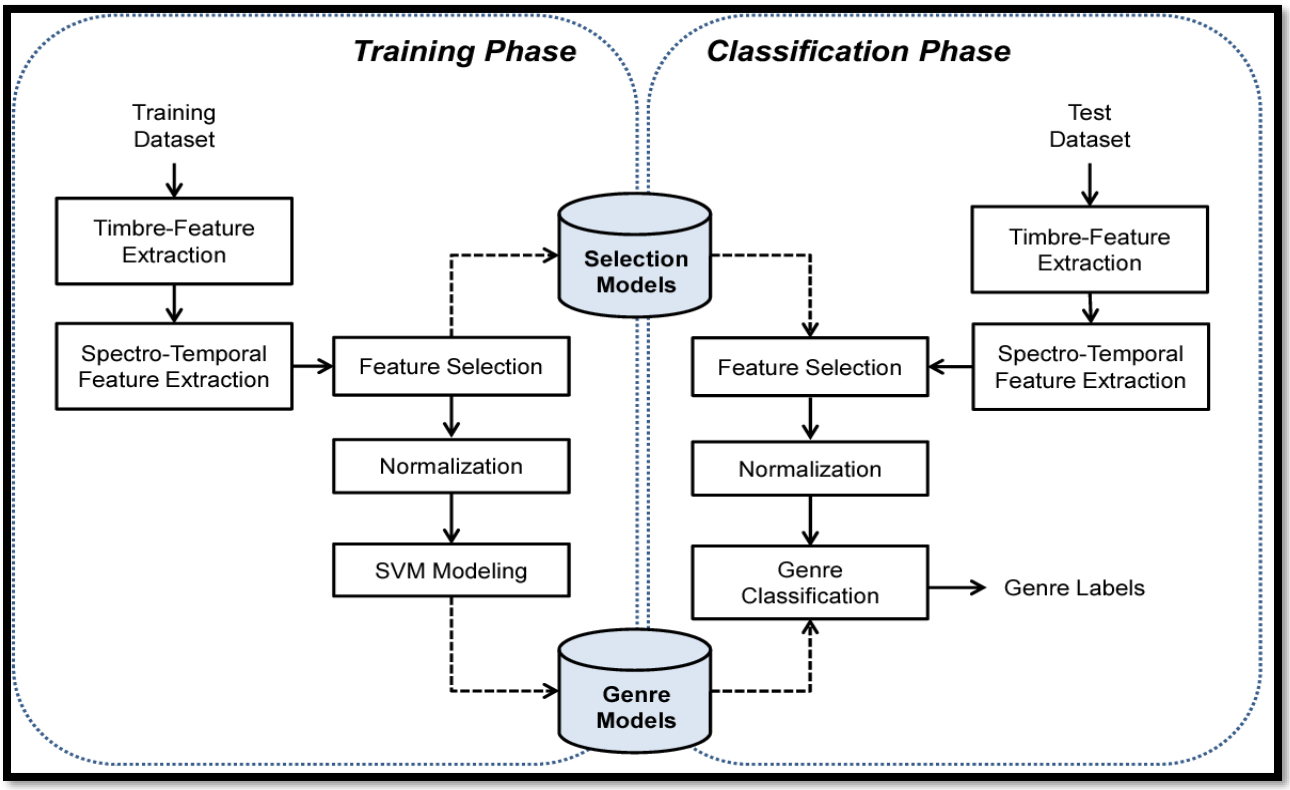

Figure 2. Genre Based Classification

1.1 Singer Based Classification

A song is a composition of a singing voice and accompanying instrumental music. To capture the singer's characteristics, one major issue is to extract the segments which contain only the singing voice. But it is not easy to have such elements as it is usually accompanied by instruments, and the existence of only voice is quite rare. Observers note that certain measures are made to eliminate the impact of instruments. Separating the singing voice from the accompanying instrumental music is a difficult process that is still the subject of continuous research. However, studies have suggested focusing on voice dominant parts in order to identify singers (Panagakis et al., 2009).

1.2 Emotion Based Classification

Music is associated with emotion, and accordingly, it generates an intuitive feeling to the listener. A music clip's ability to convey emotion depends on its structural qualities. Tempo, melody, mode, volume, and rhythm are examples. Tempo indicates the speed or pace of a musical excerpt. A music excerpt with a fast tempo normally conveys emotions like happiness, excitement, and anger, whereas a slow tempo conveys sadness and serenity. Mode indicates the type of scale or tonality of the music excerpt. A music excerpt with major tonality conveys emotions like happiness, joy whereas minor tonality conveys sadness. A music excerpt's loudness is defined as its physical force and amplitude (Pedregosa et al., 2011).

1.3 Raga Based classification

Indian Classical Music is regarded as one of the most prestigious and highest music classes. It is broadly divided into North Indian or Hindustani classical music and South Indian or Carnatic classical music. Hindustani classical music is mainly found in Northern part of India, Bangladesh and Pakistan. Raga (the composition) and tala (rhythmic cycle) remain the central notion in both the systems. Indian Classical Music follows a particular musical style which is known as Gahanna. Guarani's have their basis in the traditional musical training and education mode. Each Gahanna has its style (Tzanetakis & Cook, 2002).

1.4 Sports-based classification of songs.

These are some of the fastest-growing musical genres, so you have heard your favorite song on the radio and landed on a similar one. The very influential genre has now generated sufficient demand for a comparison classification that assists listeners in keeping track of their favorite tracks. The website Music Genre Guru provides easy-to-read descriptions of the 37 most popular genres and connections to each area's top albums and musicians.

The website also includes instructions on making personalized playlists based on specific emotions, frequencies (e.g., high energy, mellow, complex structure), and locations (i.e., party, car, work). Classification based on sports can be utilized in a variety of ways. The first is for radio stations to use the classification to determine which songs are played during game time and when. The music then assists the listeners in staying informed about the game. Sports teams can also utilize sports-based classification to cultivate stronger victory attitudes for their team and inspire excellence through times of defeat or failure. The music could perhaps increase the players' mood and create an environment conducive to fostering stronger ties amongst players.

Sports enthusiasts can also use the classification of songs based on sports to enlighten themselves about the music played during games and as a tool for making their playlists. Certain songs and their lyrics, or even a newfound interest in the teams they support, could drive fans to learn more about their teams.

There are also different genres that these songs could fall under, such as:

* Alternative Rock * Hip Hop * Metal * Rock n Roll * Soul

2. Related Work

Classifications of the musical genre have always been based on either the song's sound or lyrical content. While many different genres have been identified, one type of classification, which has not yet received much attention, is sports-based classification. This initiative will identify and categorize songs about sports to provide musicians with a new means of identifying and targeting their audience. This project will also serve as a repository for scholars interested in finding potential themes within song lyrics and how they relate to particular audiences. This data may establish the groundwork for future research on the effect these themes have on listening habits and the desire for more music with comparable content. The most influential taxonomy of musical genres was devised in.

The authors explored the automatic classification of audio signals into a hierarchy of musical genres. They proposed three feature sets representing timbre texture, rhythmic content, and pitch content. The performance and relative importance of the proposed features are investigated by training statistical pattern recognition classifiers using real world audio collections. Using the 23 proposed feature sets, a classification accuracy of 61 percent for ten musical genres is achieved. A complete survey of both features and classification (Xu et al., 2003).

3. Data Augmentation and Spectrograms

We used the GTZAN dataset (marsyas.info/downloads/datasets.html) in this work, which is most widely used in music genre classification tasks. The dataset contains ten different genres: pop, metal, jazz, classical, rock, hip-hop, reggae, country, disco, and blues. Each genre has 100 audio clips, and each clip is of 30 seconds long. We have taken a subset of these datasets comprising only eight different genres. However, the unlimited data is insufficient to learn the parameters of CNN, so data augmentation is performed on the given data (Elbir et al., 2018).

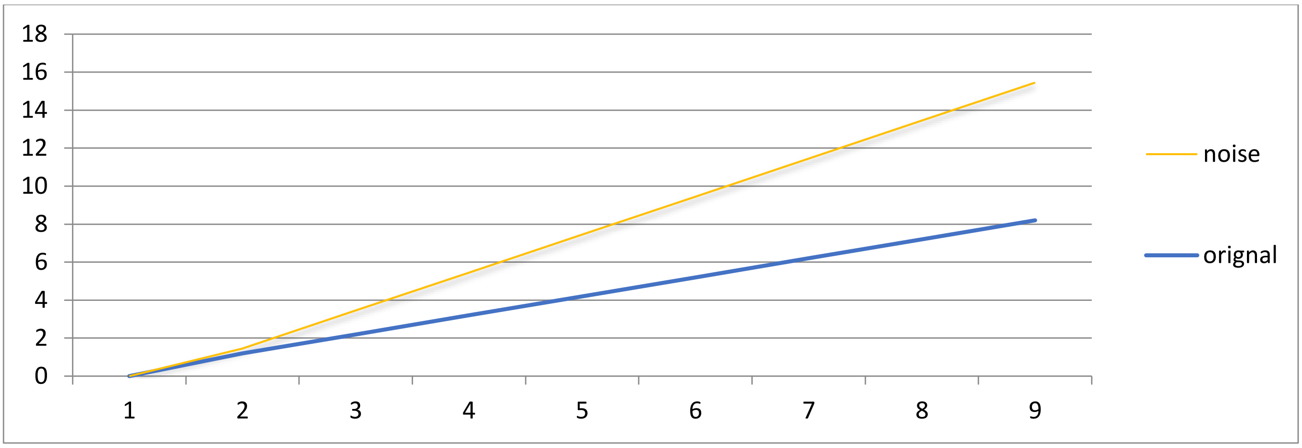

Figure 3. Data Augmentation and Spectrograms

3.1 Data Augmentation

Data augmentation is a process that can be described as developing new data by adding additional data. It can have many different forms, including the addition of new items (e.g., points in golf), the addition of new attributes (e.g., wickets, runs in cricket), or simply amending existing data to keep it relevant (for example, changing the number of games played by each team). In recent years, there has been a growth in the number of sports enthusiasts, with more and more individuals getting interested and involved in their preferred sport. This has resulted in a rise in the number of new sports games. A football manager, for instance, allows users to control their squad and test out different strategies. This also covers cricket, badminton, and basketball, among many others.

Nonetheless, as the popularity of each game grows, people desire to play more and more games. People desire to play these games independently, not merely via computer or television. The electronic version of athletics is a popular pastime that people participate in for various reasons. Some individuals may be interested in the mathematical aspects of sports, while others may believe it is preferable to spectate rather than participate. This forces them to choose whatever sport they wish to play and how much time they want to devote to mastering it. For instance, if a person is passionate about cricket, they may want to study everything they know about the sport, meaning it will be a while before they play their first game. Alternatively, if a person merely wants to play without learning much about the sport, it will take less time, but they will not pick up advanced strategies as quickly (Tzanetakis & Cook, 2002).

Due to the complexity and depth of the different sports, one would need considerable data collection to study sportsstudy sports thoroughly. One way of collecting quality data is by using surveys. There are many benefits of using a study; however, it is essential to consider its drawbacks. Surveys can be an effective and trustworthy method of data collection. For instance, it is feasible for organizations to use polls while hiring new personnel (e.g., through an email). However, this can lead to erroneous assumptions about the individuals selected for recruiting (for example, assuming that they would have applied if they were asked directly). Sport research can also make use of questionnaires. As a researcher, you must ensure that the questionnaires used to collect data are trustworthy. It is also vital to evaluate the types of questions asked and the variables being measured. This can facilitate a researcher's comprehension of the facts and statistics. However, it is also essential to remember that surveys have limitations because they do not represent the entire population (Tzanetakis & Cook, 2002).

On the other hand, data augmentation can be used to improve sports games. This can give players a more realistic experience, as they will have to use more advanced tactics and strategies to win matches. It also gives people who want to play a sport other than their favorite one an opportunity to play their favorite sport using a computer or their hand. Examples of data augmentation can be found in many sports, such as cricket and football. An example of cricket data augmentation is the addition of wickets to a game, meaning that a user would have to score more runs than their opponent to win. An example of football data enhancement is the addition of players and points, resulting in a lengthier and more time-consuming game (and therefore more realism). This helps users to practice more complicated strategies, as they will be need to adapt in order to win. This also means that users can create their scenarios and play against opponents with different strengths and weaknesses.

Noise Injection It simply adds some random value to the data. Here, we added the data sampled from Gaussian distribution at the same position by 0.005 times as much as the length of the data, normalized it to a smaller number, and then applied element-wise add to the data. Equation 1 shows the mathematical representation for the same (Tzanetakis & Cook, 2002).

$$\boldsymbol{Noise} = \boldsymbol{rand}\,()\ \%\boldsymbol{Len}\,(\boldsymbol{data})$$
$$\boldsymbol{Data\ noise} = \boldsymbol{data} + \mathbf{0.005} * \boldsymbol{noise}$$

4. Spectrograms

The features extracted from the audio are Mel spectrograms, Mel frequency cepstral coefficients (MFCC), Spectral Centroid, Spectral Roll-off, Zero Crossing rate. This study extracted Mel spectrograms from the audio waveforms since it is regarded as detailed and accurate information for audio signals. The low low-order MFCCs describe the slow variations of the spectral envelopehereas the fast variations of the envelope are defined by the higher-order MFCCs (Scaringella et al., 2006). Typically, the top thirteen coefficients of MFCCs are used to classify musical genres. In this study, we applied techniques of deep learning that circumvent the selection of audio signal features. Moreover, deep learning permits hierarchical architecture that is congruent with the layering structure of music in both the temporal and frequency domains (Elbir et al., 2018).

$$\boldsymbol{Spectrogram}\,\{\boldsymbol{x}\,(\boldsymbol{t})\}\,(\boldsymbol{\tau},\boldsymbol{\omega}) \equiv |\boldsymbol{X}\,(\boldsymbol{\tau},\boldsymbol{\omega})|\,\mathbf{2}$$

5. Extraction of Vocal Component

A song is the composition of the singing voice and instrumental music. Some segments may contain voice with or without accompanying background music, and some may have only background music. We refer to such segments as vocal and non-vocal segments, respectively. Extracting the segments, which contain only the singing voice, will be the best scenario for subsequent use in characterizing the singer's voice. However, it is not easy to attain, as the existence of only a voice is quite rare. There are primary segments with singing voices and background music and segments with only the music. Figure 3 shows a sample song clip where the deep blue colored segments are vocal (contains voice) parts and the light blue colored segments are non-vocal parts without any singing voice (Zhang & Kuo, 1998).

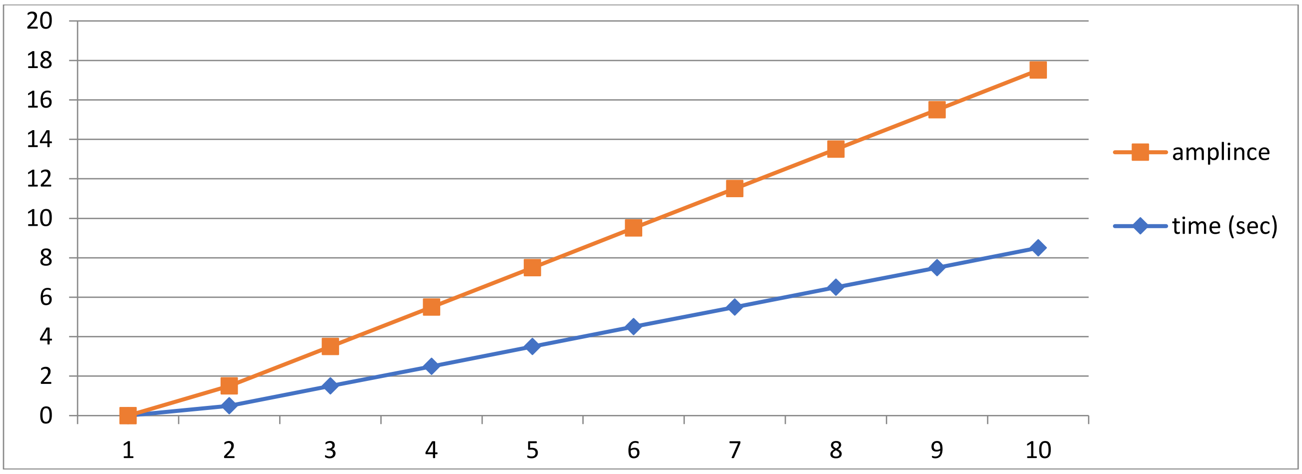

Figure 4. Extraction of Vocal Component

6. Review of literature

Fujihara et al. (2010) used the hidden Markov model (HMM) to detect vocal and non-vocal segments at the first stage. Sub-band-based log frequency power spectrum is considered the feature for this purpose. Identified vocal segments are further verified using hypothesis test. The portions of a song are then represented by perceptual characteristics such as harmonic, vibrato, and timbre. The verification method is utilized for singer identification, and the hypothesis test is used to validate the input feature vectors against the singer models.

Shen et al. (2012) proposed hybrid singer identification (HSI) model. Using SVM, audio clips are separated into vocal and non-vocal parts as input vectors to SVM, MFCC, energy, zero-crossing rate (ZCR), and spectral characteristics such as spectral centroid and spectral flux used. For singer identification, characteristics are computed from both the vocal and non-vocal portions of the audio. The linear predictive cepstral coefficients (LPCC) based on vocal timbre feature and pitch histogram characterize the vocal segments. It is assumed that an artist performs within a limited set of genres. Hence moments are computed from Daubechies wavelet coefficient histogram (DWCHs) to capture the genre information. It is further assumed that a singer sings with a more or less fixed set of instruments. Hence, from the non-vocal segment, MFCC is computed to represent instrument information.

L. Li et al. (2015) have considered sparse representation-based classification to detect the vocal segments. Different auditory features based on MFCC, linear predictive Mel-frequency cepstral coefficients (LPMCC), and Gamma tone cepstral coefficient (GTCC) are extracted from the vocal segments. Finally, GMM is used to model the singers.

Saha et al. (2018) have applied a pre-processing technique to extract the vocal component. For singer-based classification, characteristics based on the variation pattern of zero-crossing rates and short-term energy are applied. Ratanpara and Patel (2015) have utilized timbre, Chroma gram, loudness, MFCCs, LPCCs, and Ad boost as classifiers for popular Indian video music. Tsai et al. (2015) addressed compressed (mp3) files. These files are decompressed before MFCCs are extracted. GMM is used to determine the coefficient distribution.

Rauber and Schindler (2015) used both audio and visual-based features to classify music genres. The acoustic features they have extracted are statistical spectrum descriptors (SSD), rhythm patterns, rhythm histograms, MFCCs, and Chroma gram. Color statistics and emotion-related features are taken from music videos as visual information.

Additionally, the music incorporates highly structured characteristics such as pitch, rhythm, and tempo. For categorization purposes, the values of the characteristics and their spatial-temporal structure are employed to represent the music. Researchers have explored a variety of strategies for categorization, including k-nearest neighbors, Gaussian mixture model (GMM), Support Vector Machines (SVM), and artificial neural network similarity functions (Elbir et al., 2018).

7. Research Methodology

7.1. Descriptor for Genre: two methodologies have been presented. In one case, empirical mode decomposition (EMD) was deployed, and subsequently, features were computed. In the other approach, low-level features have been calculated from the signal without any pre-processing in decomposition. EMD is a prolonged process (Silla et al., 2008).

7.2. Descriptor for Singer: A song is the composition of the singing voice and instrumental music. Some segments may contain voice with or without accompanying background music, and some may have only background music. The music excerpts are segmented, and non-vocal segments are removed based on the time-domain energy distribution. A simple frequency domain filtering is applied to the vocal segments to minimize the impact of accompanying instruments. Then MFCCs features and spectrogram-based vocal-print features are extracted from the filtered vocal segments to represent the singer characteristics (Pelchat & Gelowitz, 2019).

7.3. Descriptor for Emotion: two methodologies have been proposed. One is based on deep learning, and the other is based on low-level features. We have used low-level feature-based methods to keep costs down and the dataset manageable (Meng & Shawe-Taylor, 2005).

8. Data Analysis

To experiment, we have prepared a dataset consisting of 202 music clips. The dataset has been annotated with three distinct types of metadata: Genre, Singer, and Player Emotion. The collection includes the recordings of six singers: Abbasuddin Ahmed, Asha Bhosle, Anita Saha, Pandit Jasraj, Kishore Kumar, and Anup Jalote. Folk, Rabindra sangeet (a subset reflecting a specific genre type), Devotional, and Classical are the four distinct genres. Songs are associated with four distinct emotions: Joy, Peace, Romance, and Sadness. To experiment, we have considered the mono channel music excerpts of 50 seconds duration, sampled at 23650 Hz (Tzanetakis & Cook, 2002).

Table 1

Accuracy (in %) for genre based classification

Genre	Accuracy (in %)
Folk	90.25
Tagore	98.00
Devotional	96.36
Classical	100.00
Overall	96.32

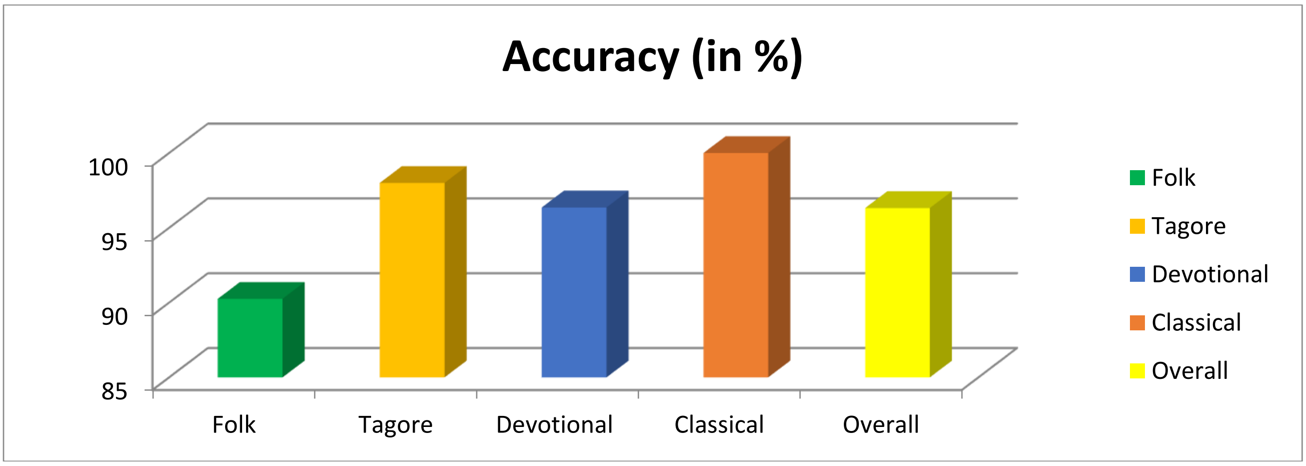

Figure 5. Accuracy (in %) for genre

Part of the data has been used to train the network, and the rest is used for testing. Fivefold cross-validation has been done, and average results are being reported. Table 1 displays the classification accuracy based on genre. 96.32 percent accuracy in classification is recorded overall. The identity categorization accuracy for Singer is 98.36%. The respective accuracy of several singers is given in Table 2. We have also attempted to identify the singer with an overall accuracy of 91.27 percent on our Raga dataset. Table 3 shows the classification accuracy for emotion detection. The most task, success is limited,, and overall accuracy of 75.33% has been obtained.

Table 2

Accuracy (in %) for singer based classification

Singer	Accuracy (in %)
Leela	96.67
Mansi	98.45
Harshita	100.00
Surbhi	98.36
Overall	98.36

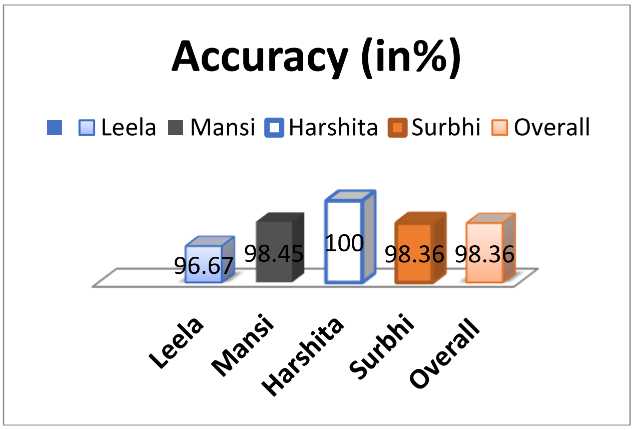

Figure 6. Accuracy (in %) for singer

Table 3

Accuracy (in %) for emotion based classification

Singer	Accuracy (in%)
Joy	85.36
Peaceful	75.33
Romantic	80.00
Sadness	75.23
Overall.	75.33

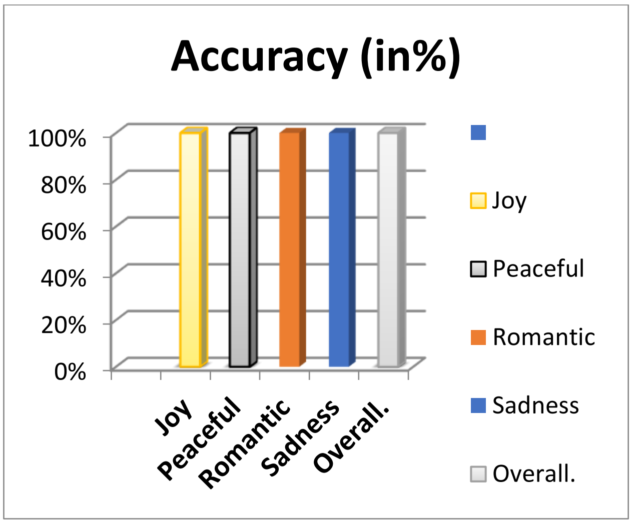

Figure 7. Accuracy (in %) for emotion

Table 4

Overall classification accuracy (in %) for combined aspects.

Combination	Accuracy (in %)
Genre	97.26
Singer	99.02
Emotion	75.20
Genre+ Singer	94.36
Emotion+ singer	75.22
Emotion + genre +Singer	75.22

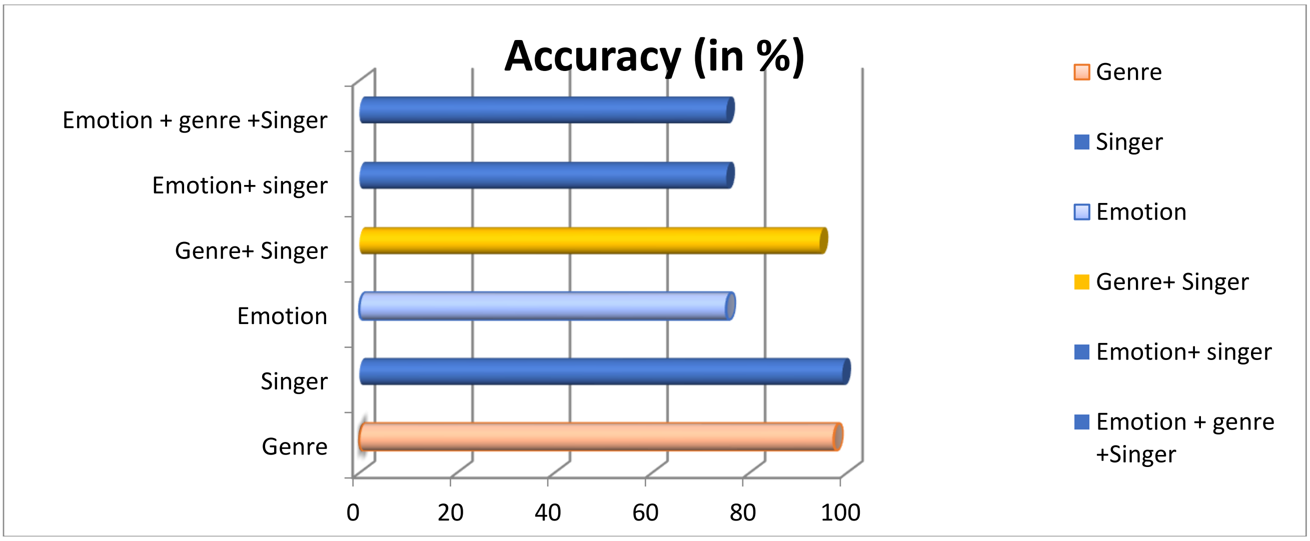

Figure 8. classification accuracy (in %) for combined aspects

Accuracy for classification on multiple aspects is reported in Table 4. Different combinations of the aspects (i.e. emotion, genre, and singer) have been considered. It is observed as the accuracy for detecting emotion limits the performance. A detailed result for all the aspects taken together and an overall accuracy of 75.22% have been obtained.

9. Result and Discussion

This work combines low-level descriptors to express music metadata such as style, genre, performer, and mood. A dataset consisting of music clips of different genres, singers, and emotional categories has been prepared and annotated. The methodology for the multi-aspect classification of music data has been tested. Even without costly descriptors based on EMD (for genre) and a deep learning-based approach, the proposed method is satisfactory. It is observed that genre identification performance is quite high for singers, and it suffers in identifying the emotion. It also affects the multi-aspect classification. In the future, efforts may be directed to improve in this respect.

Table 5

Accuracy (in %) for classification based on genre- singer-emotion taken together.

Emotion + genre +Singer	Accuracy (in %)
Folk – Leela- Joy	50.00
Folk – Leela- Peaceful	45.00
Folk – Leela-Romantic	45.00
Folk – Leela-sadness	40.00
Tagore –Mansi- Joy	60.00
Tagore – Mansi- Peaceful	45.80
Tagore – Mansi- Romantic	47.80
Tagore – Mansi-sadness	75.00
Classical – Harshita- Joy	95.00
Classical – Harshita- Peaceful	60.00
Classical – Harshita- Romantic	45.30
Classical – Harshita- sadness	85.00
Tagore –surbhi –joy	75.00
Tagore- surbhi-peaceful	85.00
Tagore- surbhi-romantic	75.00
Tagore- surbhi-sadness	50.00

10. Conclusion

For the majority of this project, our team has collaborated closely. We worked together to generate a project concept, establish the project's direction, and determine data processing and implementation details. We never felt the need to delegate tasks clearly when composing the project proposal, milestone reports, and this final report. Arianna handled most of the design work for the poster, but all three members of the group contributed to the content.

We observed that each music class's performance at a sports university is affected differently by each growth approach, suggesting that the exhibition of the model could be improved by employing class-restrictive information expansion. Our future projects include song characterization based on the vocalist and artist using productive brain network models and artists seeing a movie using computer vision and artificial consciousness techniques. As music is frequently multi-instrumental or vocal with instrumentals, a robust filtering mechanism to separate the components can significantly assist when building the features for apps of this nature. Time-consuming is the proposed EMD-based method for genre identification. It is primarily because of the breakdown process. In the future, it may be a goal to develop a different, speedier strategy with equivalent capacity. A CNN-based approach for emotion recognition has been proposed. CNN-LSTM network can be considered in the future. In general, it is recognized that emotion recognition is still a work in progress due to our limited results.

Regarding the raga-based classification of Indian classical music, there is room for further comprehension and application of domain knowledge. Only in the instance of emotion-based categorization did we rely on deep learning to overcome the challenge of designing relevant features. However, there is a tremendous possibility to investigate the design and application of deep learning networks for classifying music data on each considered element.

11. References

Elbir, A., İlhan, H. O., Serbes, G., & Aydın, N. (2018). Short Time Fourier Transform based music genre classification. *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)* (pp. 1-4). IEEE. https://doi.org/10.1109/EBBT.2018.8391437

Fujihara, H., Goto, M., Kitahara, T., & Okuno, H. G. (2010). A modeling of singing voice robust to accompaniment sounds and its application to singer identification and vocal-timbre-similarity-based music information retrieval. *IEEE Transactions on Audio, Speech, and Language Processing, 18*(3), 638-648. https://doi.org/10.1109/TASL.2010.2041386

Li, L., Cai, H., Zhang, Y., & Qian, J. (2015). Sparse representation based image quality assessment with adaptive sub-dictionary selection. *Pacific Rim Conference on Multimedia* (pp. 53-62). Springer. https://doi.org/10.1007/978-3-319-24078-7_6

Li, T., Ogihara, M., & Li, Q. (2003). A comparative study on content-based music genre classification. *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval* (pp. 282-289). https://doi.org/10.1145/860435.860487

Meng, A., & Shawe-Taylor, J. (2005). An Investigation of Feature Models for Music Genre Classification Using the Support Vector Classifier. *ISMIR. 2005* (pp. 604-609). Citeseer. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.106.3531&rep=rep1&type=pdf

Mohamed, A. E. (2017). Comparative study of four supervised machine learning techniques for classification. *International Journal of Applied, 7*(2), 5-18. https://www.academia.edu/download/54482697/2.pdf

Panagakis, Y., Kotropoulos, C., & Arce, G. R. (2009). Music genre classification via sparse representations of auditory temporal modulations. *2009 17th European Signal Processing Conference* (pp. 1-5). IEEE. https://ieeexplore.ieee.org/abstract/document/7077429

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research, 12*, 2825-2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https://githubhelp.com

Pelchat, N., & Gelowitz, C. M. (2019). Neural network music genre classification. *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)* (pp. 1-4). IEEE. https://doi.org/10.1109/CCECE.2019.8861555

Ratanpara, T., & Patel, N. (2015). Singer identification using MFCC and LPC coefficients from Indian video songs. *Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1* (pp. 275-282). Springer. https://doi.org/10.1007/978-3-319-13728-5_31

Rauber, A., & Schindler, A. (2015). An audio-visual approach to music genre classification through affective color features. *European conference on information retrieval* (pp. 61-67). Springer. https://doi.org/10.1007/978-3-319-16354-3_8

Saha, S. K., Sarkar, R., Singh, A., & Mondal, A. (2018). Automatic Extraction and Identification of Bol from Tabla Signal. In R. Chaki, A. Cortesi, K. Saeed, & N. Chaki (Eds.), *Advanced Computing and Systems for Security: Volume Five* (pp. 139-151). Springer Singapore. https://doi.org/10.1007/978-981-10-8180-4_9

Scaringella, N., Zoia, G., & Mlynek, D. (2006). Automatic genre classification of music content: a survey. *IEEE Signal Processing Magazine, 23*(2), 133-141. https://doi.org/10.1109/MSP.2006.1598089

Sethares, W. A., & Toussaint, G. T. (2015). Expressive timbre and timing in rhythmic performance: analysis of Steve Reich's Clapping Music. *Journal of New Music Research, 44*(1), 11-24. https://doi.org/10.1080/09298215.2014.935736

Shen, G., Nguyen, Q., & Choi, J. (2012). An environmental sound source classification system based on mel-frequency cepstral coefficients and gaussian mixture models. *IFAC Proceedings Volumes, 45*(6), 1802-1807. https://doi.org/10.3182/20120523-3-RO-2023.00251

Silla, C. N., Koerich, A. L., & Kaestner, C. A. (2008). A machine learning approach to automatic music genre classification. *Journal of the Brazilian Computer Society, 14*(3), 7-18. https://doi.org/10.1007/BF03192561

Tsai, T.-H., Huang, Y.-S., Liu, P.-Y., & Chen, D.-M. (2015). Content-based singer classification on compressed domain audio data. *Multimedia Tools and Applications, 74*(4), 1489-1509. https://doi.org/10.1007/s11042-014-2189-6

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing, 10*(5), 293-302. https://doi.org/10.1109/TSA.2002.800560

Xu, C., Maddage, N. C., Shao, X., Cao, F., & Tian, Q. (2003). Musical genre classification using support vector machines. *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 5* (pp. V-429). IEEE. https://doi.org/10.1109/ICASSP.2003.1199998

Yanchenko, A. (2017). *Classical music composition using hidden Markov models.* (Doctoral dissertation). Duke University. https://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/15245/Yanchenko_duke_0066N_13967.pdf

Zhang, T., & Kuo, C.-C. J. (1998). Hierarchical system for content-based audio classification and retrieval. *Multimedia Storage and Archiving Systems III. 3527* (pp. 398-409). SPIE. https://doi.org/10.1117/12.325832