

Singer and Genre recognition using deep learning

Zainab Iqra Yasrab ¹, Akhtar Hussain Jalbani² and Saima Siraj Soomro ³

¹ Department of Computer Science/Information Technology, Quad-e-Awam University of engineering science and Technology Nawabshah

² Department of Computer Science, Quad-e-Awam University of engineering science and Technology Nawabshah

³ Department of Information Technology, Quad-e-Awam University of engineering science and Technology Nawabshah

*Correspondence: Zainab Iqra Yasrab, yasrabrahu@gmail.com

ABSTRACT: In this paper, we try to solve the two mostly occurred problems which are genre recognition and singer identification from common pieces of music. For that, we use a system that helps us to achieve our target that is based on a music information retrieval system, in which the system can process our files and the system can generate the desired results. The system is divided into different phases and for the music recognition, first, there is a need to separate the vocal features of the singer from the music clip because, with the background music, it's slightly difficult to analyze the vocal of the singer, then extract music features, here we work on some of the music features like tempo, RMS, song duration, frequency, dynamic range, and tonality. The RNN (recurrent neural network) has been used for training the model and testing our data set. After the training of the models, the test classifier with different unknown audio files, and the results are 91% accurate for singer identification and 90% for music genre recognition. The system will identify the singer as well as the genre of the song from the same segment of the file. For the data set, the training data used 10 different Indian and Pakistani male and female singers' audio files in Sindhi, Hindi, and Urdu language for six types of genres like rock, pop, jazz, disco, hip-hop, and classic. The songs were sung by only one singer, the group singing and live concerts music are limitations of our system.

Keywords: Genre, Singer, Classification, CRNN, Deep learning

Introduction:

Music information retrieval (MIR) is an interdisciplinary field that combines elements of signal processing, machine learning, and music theory to analyze musical content. The area of automatically categorizing the genre of music in audio format has seen an increase in recent years.

[1]. MIR provides intelligent musical data understanding and processing in computer algorithms [2]. to make machines understandable real-life problems and generate solutions related to music recognition. Nowadays, music data files are increasing daily, so finding a particular song is a complicated task. It is also becoming impossible for humans to structure and organize such a massive amount of music [3]. So it is necessary to address this type of problem with an intelligent system.

In this paper, we are going to address two types of mostly acquired and complicated problems with the help of deep learning approaches. The problems that are going to solve are our "music genre classification" and "singer identification". In the field of music information retrieval, it might be difficult to classify music files by genre (MIR) [2] because some music genres slightly differ from each other, so classifying music according to the genre is a very complicated task, but deep learning models can help to successfully identify the genre. Almost everyone is quite familiar with the singing voice,

Citation: Rahu, Z. I. Y. Singer and Genre recognition using deep learning: Singer and Genre recognition. *Pakistan Journal of Emerging Science and Technologies (PJEST)*, 4(2).

<https://doi.org/10.58619/pjest.v4i2.102>

Academic Editor: Dr. M. Javaid Afzal

Received date: 28th March, 2023

Revised date: 13th April, 2023

Accepted date: 14th April, 2023

Published date: 15th May, 2023



Pakistan Journal Emerging Sciences and Technologies (PJEST) in collaboration with [Govt. Islamia College Civil Lines Lahore, Pakistan](#) is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#)

which is the oldest musical instrument. It is not unexpected that our auditory physiology and perceptual apparatus have developed to a high level of sensitivity to the human voice, given the significance and utility of vocal communication. Even with very little training, it is rather simple to recognize a particular person's speaking voice once have been exposed to it [4]. So the identification of a singer from a particular song is possible.

Here a recurrent neural network (RNN) is used for training the system, and to recognize singers and genres, recurrent feature-based and trending recurrent neural network (RNN)-based algorithms are taken into consideration and contrasted. For the training dataset, 10 different Indian and Pakistani singers and 6 different genres are used to create our dataset and then select 10 songs from every singer for every genre, which means we select 120 songs from every artist. Also in this paper, we experiment with songs in different languages because the songs also show the different cultures and communities. The main target is to make a system that works for all types of songs, so here we only test with three different languages; these are spoken in areas of India and Pakistan. To achieve the target, we select some Sindhi songs along with Urdu language songs and Hindi language songs after the training system categorizes the genre of these languages too. Somehow, the Urdu and Hindi languages are mostly similar, but when we talk about the Sindhi language, which is different from both of the other languages, that was a challenging task for our system. In our dataset, we select some songs by Sindhi singers and arrange them according to different genres. First, we remove the background music and extra instrumental music because that step helps to easily detect the vocal cords. After that, we extract different features for genres as well as a singer. For singers, we select features like tempo, tonality, sample frequency, and frequency rate, and for the genre, we select RMS, song duration, and dynamic range [5].

After collecting all these features, the normalized data is used to train the neural network model and implement the dataset. For the neural network, Neuroph Studio is used for training and testing. After training the neural network, we test different files and compare their results; in the end, we identified the type of music, genre, and singer's name accurately.

Related Work:

In the past year many researchers have done a lot of work in music information retrieval and deep learning approaches are helpful to successfully generate the results our goal is to solve two mostly accrued problems these are song genre recognition and singer identification, according to past work one of the researchers propose a neural network that combines a two approaches LSTM and SVM and make a hybrid model to categorize the songs [6]. Deep learning models are really helpful in this task a hierarchical model is using LSTM to increase the number of genres categorizing from 6 genres to 10 genres of songs [7]. The researchers also compare the results of different models to improve the accuracy of genre classification and the CRNN has accurately identified the genre with the best accuracy [8, 9]. In conferring the following papers we analyzed that the neural network works on a high level to classify music according to the genre as the paper said that is about genre classification using neural network in this paper the researchers use images of spectrogram generated from time slices of songs and that success to classify music according to the genre [10]. nowadays music become an important part of people's lives and that is necessary to categorize it because it also plays an important role in cultures the researchers test a model for music classification

and recommendation engine by extracting its acoustic features and training the neural network [11], still the researcher trying to make the performance of this system better than before so for that, some more researchers experiment with different deep learning models, they compare the results accuracy of three models CNN, RNN, and CRNN and also proposed the two ways to improve the performance of RNN and CRNN architectures [12]. The multimodal system predicts more accuracy than the unimodal the researchers also proposed the model in which CNN is used for the spectrogram of image and RNN for sequential data of sound the dropout is used to regularize the data and for avoiding over lifting [13] The music genre recognition also works with excellent accuracy by applying deep bidirectional approaches using LSTM and RNN the models can predict the genre with better accuracy than previous [14, 15].

Significant work is also done on the singer identification problem the researchers identify the singer by using bidirectional long and short-term memory (BLSTM) Recurrent neural network RNN this model can take past temporal features of songs and can predict the voice of the singer [16]. The RNN is helpful to the proposed number of different models and projects like the model JDC Joint detection classification that is used to detect melody singing voice and also label the pitch and detect non-voice status[17]. The researchers also search to detect the song regions where only one singer is singing which was a difficult task so to address such task RNN is used and LSTM in different experiments found the best accuracy [18]. The singing voice classification is also introduced on the bases of phoneme transcript in which researchers use a recurrent neural network with CTC connectionist temporal classification that is for training [19]. The singer classification problem also solves to the songs of Chinese singers In which they use different techniques with recurrent neural network RNN like the used LSTM and MFCC and classify the singer as well as in the second approach they use singer voice synthesis SVM to categorize Chinese singer's voices[20,21]. The RNN is used in many types of voice recognition systems to solve a related problem that also helps to find the composer of music in this work LSTM is also helping the system with 70% accuracy can predict the composer [22]. The singer identification field has two most complex and interesting problems these are singer identification and detecting music parts of polyphonic music the researchers try to address both of these problems with the help of deep learning and feed-forward neural network [23], most of the singer identification system used features engineering techniques to detect the music but here they used the music to here and detect it for that they used long-term recurrent convolutional neural network LRCN for vocal detection, convolutional layer LRCN for feature extraction and LSTM for learning features and sequence relationship [24]. An automatic singer age, height, race, and gender assessment system proposed by using Monaural source separation techniques, training with large data, and testing with LSTM helps to improve the accuracy of results [25].

Proposed Framework:

In this work, the Recurrent Neural Network RNN model is used which is a deep learning model and works on sequential data and time series also RNN helps to predict the output along with training the system to predict similar data in the future, for the training the dataset of 10 different singer's songs and 6 different types of genre. For every singer 10 songs for every genre are selected but a few singers sing only some of the common genre songs like "Nazia Hassan" only sing the songs of pop, rock somehow jazz genre, same as one of the famous singer of Pakistan "Abida Parveen" doesn't sing

rock and hip-hop songs so here only select the related genres otherwise we try to collect all 6 genre songs. After collecting the dataset of songs the Python program helps to extract some of the features and one online tool that is helpful to find the tempo and tonality of the song the link will be found at the last of this page¹, at the Python program code, we extract MFCC features to identify the speech because every person's voice features are different from others so with the help of these features system easily identify the vocal of the singer. Now let's discuss that mechanism deeply with the help of the methodology model that is in Figure 1

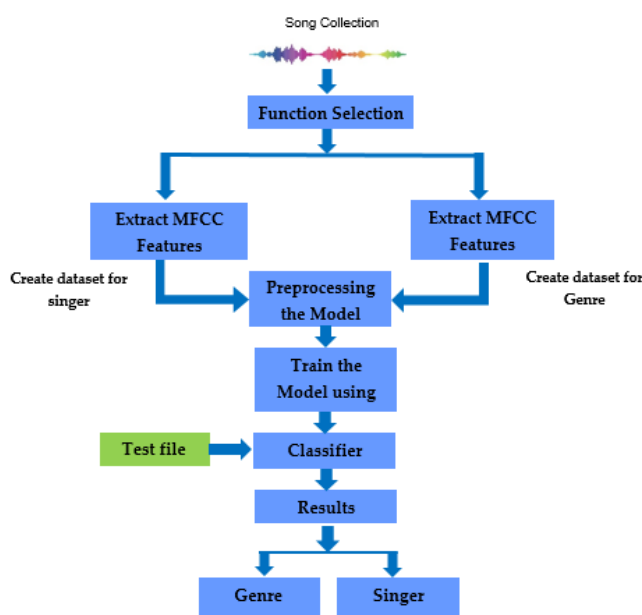


Figure 1 Research Methodology

Model Discussion:

Song Selection:

In song selection, the appropriate song files are selected from different genres and singers to create a collection of audio features and metadata for the training of models.

Feature Extraction:

Feature extraction is a process to transform the raw data into numerical features, audio feature extraction is a necessary step in music recognition it process and manipulates the audio features of an audio signal and removes extra noise from the song, and balances the frequency and other features. Here first extract features genre and singer's vocals.

Preprocessing the models:

In this step, the extracted data of audio signals are arranged and saved in the dataset to train models.

Train the model using RNN:

In this step, the data collected from different songs are used to train our model by using a Recurrent neural network to solve the problems of classification and identification.

Test Classifier:

The classifier is a trained program that has been predicting the genre and singer by recognizing its vocal features.

Results:

In the end, we found two types of results: getting the song genre and the singer of the song.

Feature Selection:

For the identification of singers and genres, the ten different male and female different voices are used, working on different features like the following feature for the identification of singers.

- ✓ Tempo
- ✓ RMS
- ✓ Song Duration
- ✓ Frequency

For Genre Classification:

- ✓ Frequency
- ✓ Dynamic Range
- ✓ Tonality

Results and Discussion:

Data:

For the dataset, we use songs of different Pakistani, and Indian male and female singers, the songs that are selected for the dataset are sung by only one singer either male or female. After that categorize these songs into 6 different genres in which we create a separate file for every single singer and 10 audio files for every genre. After setting the dataset first separate the vocal sound from background noise and instrumental music with the help of an online tool² because extracting only vocal sound will be easier for the system to extract the features of an audio signal. Here it has been used .wave file format extracted some of the features with the help of MFCC (Mel frequency cepstral coefficient). After that made a text file to label all those features and normalized them with the help of a Python program. In the end, by using the normalized data file to train the neural network, for neural network “Neuroph studio” has been used.

Data Description:

Table 1 Genre Pattern

S.No	Genre	Code Pattern
1	Jazz	001
2	Rock	010
3	Pop	100
4	Classis	011

5	Disco	101
6	Hip-hop	110

Table 2 Singer Pattern

S.No	Singer Name	Code Pattern
1	Abida Parveen	0001
2	Shaman Ali	0010
3	Kumar Sanu	0100
4	Manzoor Sakhirani	1000
5	Lata Mangeshkar	0011
6	Mehdi Hassan	0101
7	Atif Aslam	1001
8	Kaifi Khalil	0010
9	Saeed Tunio	0110
10	Nazia Hassan	1100

Results and Discussion:

First of all, load the dataset in the data model and then train it the training graph shows the 0.009 error rate after the completion of training.

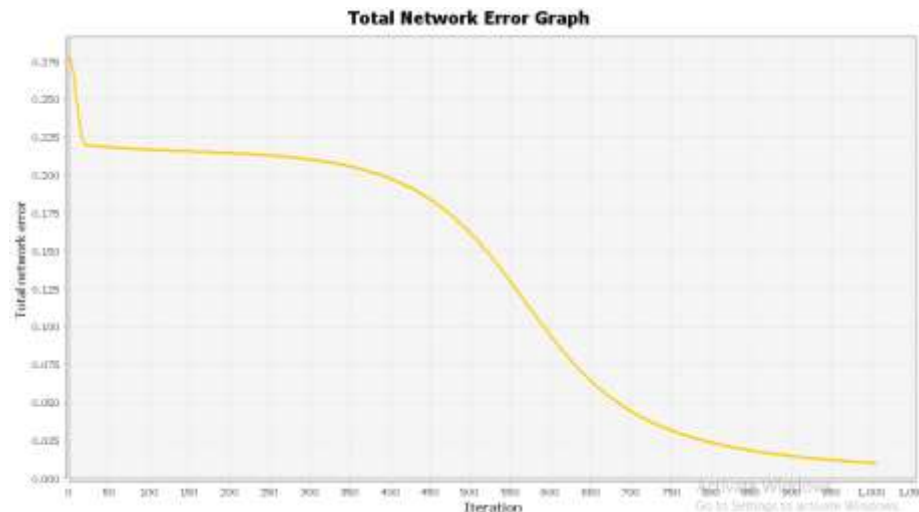


Figure 2 Training Eroor Graph

The Figure 2 shows the results of network graph that are found during the training of the neural network.

Neural Network Explanation:

After training the neural network model, test the system with different unknown audio files that haven't used to train the system, these experimental results are explained below one by one. In Figure 3 the architecture of neural network shown.

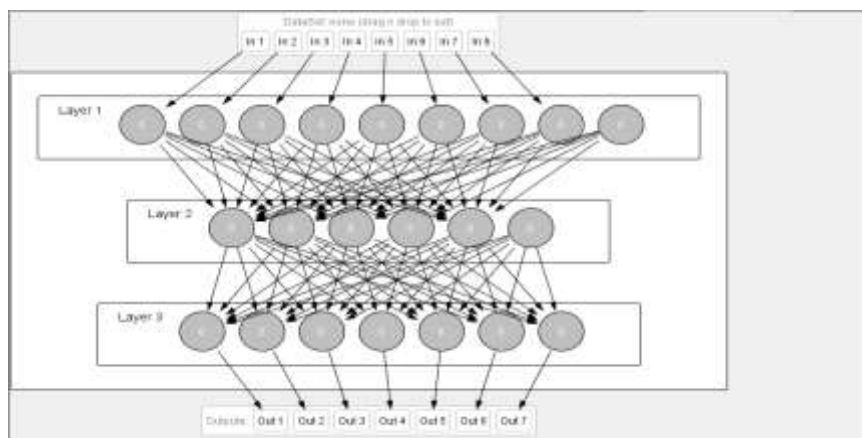


Figure 3 Neural Network

According to Figure 3, the structure of neural network is subject to three layers the first layer is the input layer in which the 8 input neurons values of these neurons are acquired from the features of different audio files, the second layer is hidden layer here five neurons and at the end, we have the output layer that is comprising on 7 neurons. The output neurons are divided into two parts the first three code pattern indicates the genre of the song and the rest of the code that is laying on 4 neurons shows the singer that you can see in Figure 4

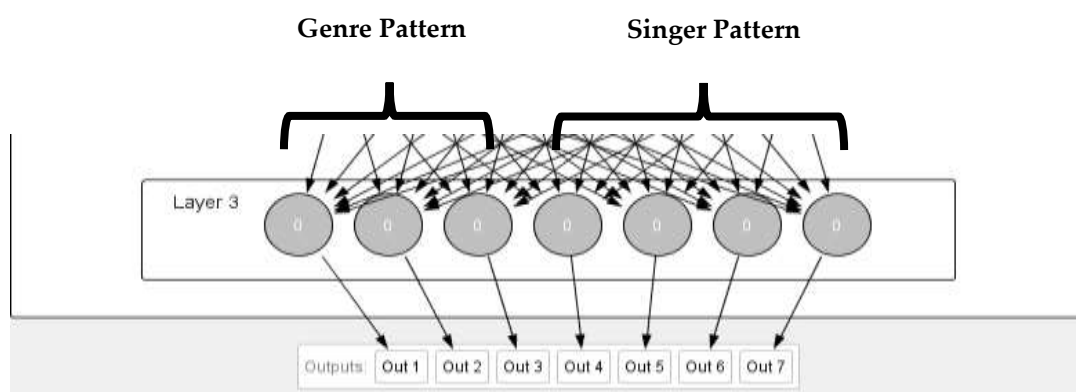


Figure 4 Neural Network

Experimental Results:

We test the different singer's audio files in the first experiment and test the neural network with the famous song of "Abida Parveen" that is "Moula e kul "This song belongs to the classic genre and according to the results in Figure 5 the first session of our code pattern show 011 that is a classic genre you can check in Table 1 for the code patterns and the second section show the code for a singer that is 0001 shows the "Abida Parveen" you can also check in Table 1.

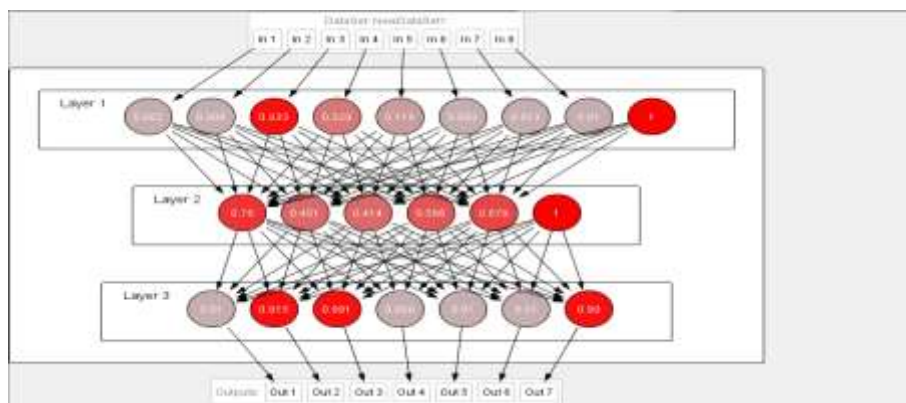


Figure 5 Testing result of 1st experiment

In the second experiment, we test the model with features of "Abida Parveen" song that is " Main Naraye Mastana" neural network predict correct results you can check in Figure 6 the code shows the 011 and classic genre and 0001 the singer is "Abida Parveen" also you can check in Table 1 and Table 2 the code patterns.

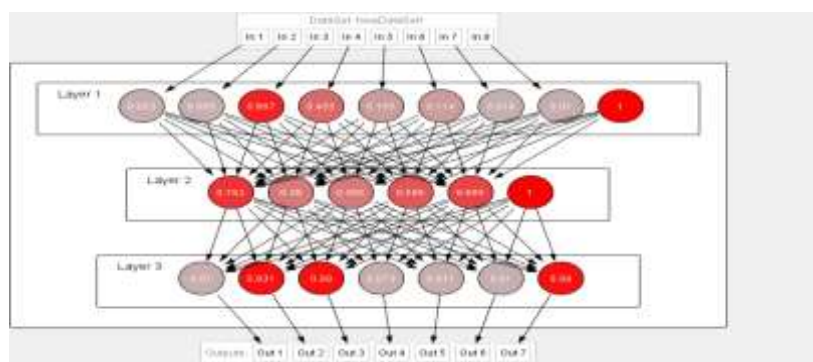


Figure 6 Testing results of 2nd experiment

The third test is takes place on different music file the song that is selected is "Khair manga me teri" the song's genre as well as singer are different and the results you can check in the the Figure 7. According to the output pattern that show the genre of song is jazz the code is 001 and the singer is Atif aslam as 1001.

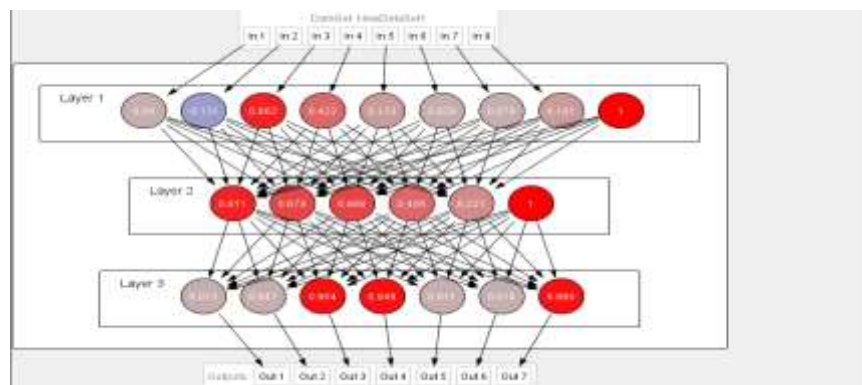


Figure 7 Testing results of 3rd experiment

Next testing file is also one of the famous song of Atif aslam “Tera Huwa” and you can see in Figure 8 the neural network can accurately detect the genre and singer as well.

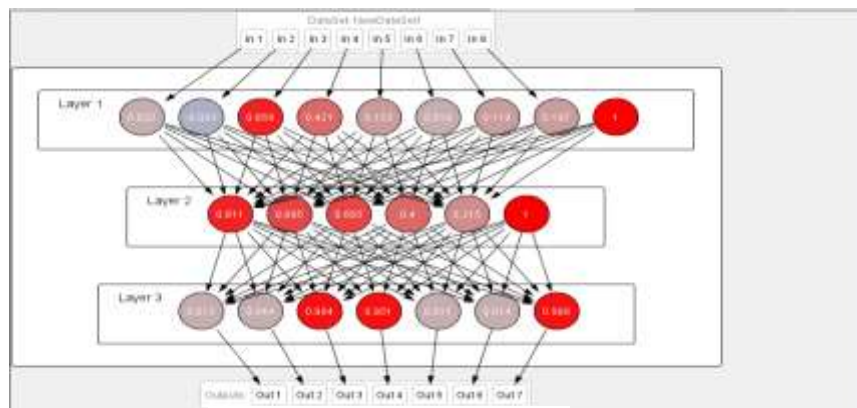


Figure 8 Results of 4th experiment

An deep learning approach RNN recurrent neural network play an important role during training and generating the results when the training phase is completed we test our classifier with different unknown audio files of different and similar singers and genre and the system gives 90% accurate results for the genre and 91% accuracy for singer identification are shown in confusion matrix that is Figure 9 for genre and Figure 10 for singer identification.

TP: 108
Overall Accuracy: 90%

Class	n (truth) ②	n (classified) ③	Accuracy	Precision	Recall	F1 Score
1	20	20	96.67%	0.90	0.90	0.90
2	20	20	90.33%	0.95	0.95	0.95
3	20	20	95%	0.85	0.85	0.85
4	20	20	90.33%	0.95	0.85	0.90
5	20	20	96.67%	0.90	0.90	0.90
6	20	20	95%	0.85	0.85	0.85

Figure 9 Confusion Matrix for genre

TP: 182
Overall Accuracy: 91%

Class	n (truth) ②	n (classified) ③	Accuracy	Precision	Recall	F1 Score
1	20	20	99%	0.95	0.95	0.95
2	20	20	99%	0.95	0.95	0.95
3	20	20	97%	0.85	0.85	0.85
4	20	20	99%	0.95	0.95	0.95
5	20	20	90%	0.90	0.90	0.90
6	20	20	90%	0.80	0.80	0.80
7	20	20	90%	0.90	0.90	0.90
8	20	20	99%	0.95	0.95	0.95
9	20	20	90%	0.90	0.90	0.90
10	20	20	99%	0.95	0.95	0.95

Figure 10 Confusion matrix for singer identification

Conclusion:

The whole paper is discuss the working and significance of system that identify the genre and singer of a song from a common piece of the audio signal, it was not an easy task to develop a system which is capable to identify the singer and genre, because

every single singer have different vocals like every person have different fingerprints, so to identify the singers first extract there vocal features like tonality, tempo, root mean square RMS, same as every different genre have unique features, for classification of genre, we extract the songs duration in minutes, tempo, dynamic range, root mean square, frequency rate and sampling rate, by using the RNN recurrent neural network we train our system with the dataset that creates with the audio songs of different male and female Pakistani Indian singers. To achieve the target we collect the songs of different genre from different singers and extract there features, for extraction of features MFCC (Mel frequency capstral coeficient) helps to extract the frequency, time and amplitude as every person have different vocals so according to their vocals they have different frequency and amplitude values like the female voice have higher pitch that means the amplitude is also high and male voice have low pitch as well. The training with the dataset make system's sense to match the features with related genre and singer's vocal with the deep learning now the system automatically identify the audio file and differentiate between the different songs, the system is work with the accuracy of 90% for genre and 91% for singer identification.

Reference:

- [1] Ghosal, D., & Kolekar, M. H. (2018, September). Music Genre Recognition Using Deep Neural Networks and Transfer Learning. In *Interspeech* (pp. 2087-2091).
- [2] Bahuleyan, H. (2018). Music genre classification using machine learning techniques. *arXiv preprint arXiv:1804.01149*.
- [3] Zhang, W., Lei, W., Xu, X., & Xing, X. (2016, September). Improved music genre classification with convolutional neural networks. In *Interspeech* (pp. 3304-3308).
- [4] Mesaros, A., Virtanen, T., & Klapuri, A. (2007, September). Singer identification in polyphonic music using vocal separation and pattern recognition methods. In *ISMIR* (pp. 375-378).
- [5] Lippens, S., Martens, J. P., & De Mulder, T. (2004, May). A comparison of human and automatic musical genre classification. In *2004 IEEE international conference on acoustics, speech, and signal processing (Vol. 4, pp. iv-iv)*. IEEE.
- [6] Fulzele, P., Singh, R., Kaushik, N., & Pandey, K. (2018, August). A hybrid model for music genre classification using LSTM and SVM. In *2018 Eleventh International Conference on Contemporary Computing (IC3)* (pp. 1-3). IEEE.
- [7] Tang, C. P., Chui, K. L., Yu, Y. K., Zeng, Z., & Wong, K. H. (2018, July). Music genre classification using a hierarchical long short term memory (LSTM) model. In *Third International Workshop on Pattern Recognition (Vol. 10828, pp. 334-340)*. SPIE.
- [8] Bisharad, D., & Laskar, R. H. (2019). Music genre recognition using convolutional recurrent neural network architecture. *Expert Systems*, 36(4), e12429.
- [9] Choi, K., Fazekas, G., Sandler, M., & Cho, K. (2017, March). Convolutional recurrent neural networks for music classification. In *2017 IEEE International conference on acoustics, speech and signal processing (ICASSP)* (pp. 2392-2396). IEEE.

- [10] Pelchat, N., & Gelowitz, C. M. (2020). Neural network music genre classification. *Canadian Journal of Electrical and Computer Engineering*, 43(3), 170-173.
- [11] Elbir, A., & Aydin, N. (2020). Music genre classification and music recommendation by using deep learning. *Electronics Letters*, 56(12), 627-629.
- [12] Rafi, Q. G., Noman, M., Prodhon, S. Z., Alam, S., & Nandi, D. (2021). Comparative analysis of three improved deep learning architectures for music genre classification. *International Journal of Information Technology and Computer Science*, 13(2), 1-14
- [13] Kim, S., Kim, D., & Suh, B. (2016). Music genre classification using multimodal deep learning. In *Proceedings of HCI Korea* (pp. 389-395).
- [14] Yu, Y., Luo, S., Liu, S., Qiao, H., Liu, Y., & Feng, L. (2020). Deep attention based music genre classification. *Neurocomputing*, 372, 84-91.
- [15] Qiu, L., Li, S., & Sung, Y. (2021). DBTMPE: Deep bidirectional transformers-based masked predictive encoder approach for music genre classification. *Mathematics*, 9(5), 530.
- [16] Leglaive, S., Hennequin, R., & Badeau, R. (2015, April). Singing voice detection with deep recurrent neural networks. In *2015 IEEE International conference on acoustics, speech and signal processing (ICASSP)* (pp. 121-125). IEEE.
- [17] Kum, S., & Nam, J. (2019). Joint detection and classification of singing voice melody using convolutional recurrent neural networks. *Applied Sciences*, 9(7), 1324.
- [18] Lehner, B., Widmer, G., & Bock, S. (2015, August). A low-latency, real-time-capable singing voice detection method with LSTM recurrent neural networks. In *2015 23rd European signal processing conference (EUSIPCO)* (pp. 21-25). IEEE.
- [19] Teytaut, Y., & Roebel, A. (2021, August). Phoneme-to-audio alignment with recurrent neural networks for speaking and singing voice. In *Proceedings of Interspeech 2021* (pp. 61-65). International Speech Communication Association; ISCA.
- [20] Shen, Z., Yong, B., Zhang, G., Zhou, R., & Zhou, Q. (2019). A deep learning method for Chinese singer identification. *Tsinghua Science and Technology*, 24(4), 371-378.
- [21] Gu, Y., Yin, X., Rao, Y., Wan, Y., Tang, B., Zhang, Y., ... & Ma, Z. (2021, January). Bytesing: A chinese singing voice synthesis system using duration allocated encoder-decoder acoustic models and waverrnn vocoders. In *2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP)* (pp. 1-5). IEEE.
- [22] Micchi, G. (2018). A neural network for composer classification. In *International Society for Music Information Retrieval Conference (ISMIR 2018)*.
- [23] Kooshan, S., Fard, H., & Toroghi, R. M. (2019, March). Singer identification by vocal parts detection and singer classification using lstm neural networks. In *2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA)* (pp. 246-250). IEEE.

-
- [24] Zhang, X., Yu, Y., Gao, Y., Chen, X., & Li, W. (2020). Research on singing voice detection based on a long-term recurrent convolutional network with vocal separation and temporal smoothing. *Electronics*, 9(9), 1458.
 - [25] Weninger, F., Wöllmer, M., & Schuller, B. (2011). Automatic assessment of singer traits in popular music: Gender, age, height and race.