

---

**ARABIC MUSIC ANALYSIS USING ARTIFICIAL INTELLIGENT TEQNIQUES**

*By*

*Eman. A. Esmaeil*

*Faculty of Specific Education.  
Mansoura University*

*S.M.Khater*

*Faculty of Specific Education.  
Mansoura University*

*S.M.K.tobar*

*Faculty of Specific Education.  
Mansoura University*

*M. E. ElAlami*

*Faculty of Specific Education.  
Mansoura University*

**Research Journal Specific Education**

**Faculty of Specific Education  
Mansoura University**

**ISSUE NO. 67, MAY , 2022**

مجلة بحوث التربية النوعية - جامعة المنصورة  
العدد السابع والستون - مايو ٢٠٢٢

---



## **ARABIC MUSIC ANALYSIS USING ARTIFICIAL INTELLIGENT TECHNIQUES**

*Eman. A. Esmail\**

*S. M. Khater\*\**

*S. M. K.tobar\*\*\**

*M. E. ElAlami\*\*\*\**

### **Abstract:**

Singing is the use of the human voice to make musically meaningful sounds, and it is used in most cultures for enjoyment or self-expression. Songs are audio signal and musical instrument representations. Speech, background noise, and music able to be identified by an audio signal analysis and separation system. The singing voice in a song provides useful information on pitch range, music content, music pace, and rhythm. Nowadays, with multimedia technology, there are many audio editing software available as well as audio merging software by mixing singing voice and music together, but most of these applications are in the field of Western music and there is a less application in the field of arabic music as well as its not free . One of the primary qualities that identify music based on a specific set of patterns is genre. However, the genres of Arabic music on the web are loosely defined, making automatic classification of Arabic audio genres difficult. In this paper, a system has been proposed that consider a form of arabic music analysis and classification which is a two-stage based system : The first stage is separation between arabic singing voice and melody using CRPCA which is an extension of RPCA in our previous work. Then extracting the arabic musical genres which forms musical melody that had extracted from the previous process . Mel Frequency Cepstral Coefficients and pitch used to extract feature from musical signal, then Supports Vector Machine algorithm used for classification process. The experimental results show that CRPCA can achieve greater separation performance than earlier approaches, particularly

---

\* Faculty of Specific Education. Mansoura University

\*\* Faculty of Specific Education. Mansoura University

\*\*\* Faculty of Specific Education. Mansoura University

\*\*\*\* Faculty of Specific Education. Mansoura University

when using temporal frequency masking. Furthermore, the duration of operation on CRPCA under the same conditions, is shorter than others, as well as the most important benefit of the separation improvement that is infact of improving the classification and analysis process in the next stages.

**Index Terms**— Source separation, Musical genres classification, Mel frequency cepstral coefficients (MFCC) , Music information retrieval (MIR) , pitch feature, Discrete wavelet transform (DWT) , genres of arabic music, Support Vector Machine , spectrogram.

### ***1- Introduction:***

Music separation and classification is a critical task in signal processing, particularly in the field of music. In general, singing sound separation systems have applications in a variety of domains, including automatic word recognition and alignment, identification of the singer, retrieval of information about music, karaoke, identifying musical genre, extraction of melody, sound signal categorization, and so on [1].

Currently, genre hierarchies, which are typically built manually by human specialists, are one way used to arrange music information on the Web [2]. This method, which may be automated via automatic musical genre categorization, is an important component of a comprehensive music information retrieval system for audio data. There is also a framework for developing and analysing musical content features [3]. The bulk of audio analysis algorithms for music that have been proposed employ feature extraction for similarity retrieval, classification, segmentation, and audio thumbnailing. As we can see, our research is the first step in identifying and evaluating various types of Arabic music, as well as isolating the Arabic singing voice from the background music. The following is the paper's outline. Section 3 proposes a system. Section 4 is about the automatic classification and evaluation of the proposed features, and Section 5 is about the conclusions and future work.

## **RELATED WORK**

One of the most common task for getting Musical Information Retrieval (MIR) is analysis process of musical jins. There were many works in this process, but it was limited to Western music thus works in Arabic music consider non-existent . This section presented some studies related to separating singing voice from music background as a first step in the analysis process, then determining music genres as a second step.

**In [4 ]** The bowing movements are utilized to create additional limitations on audio dictionary activation entries using non-negative matrix factorization (NMF). However, this method is only tested on randomly produced video sequences of string instruments where each player's different bowing actions are clearly captured.

**In [5]** Similar work has been proposed in which audio and visual correspondences are learned unsupervised to aid source separation. This research topic produces positive results separation of audiovisual music for musical instrument performances, but not separation of vocal voices.

**In [6]** The author articulates a generic formalism for source model adaptation in a Bayesian model framework. The suggested approach is then tested on the challenge of differentiating between voice and music in popular songs. According to the data, an adaptation technique can greatly improve separation performance on a continual basis

**In [7]** After focusing on the fluctuation of a singing voice, A technique for improving the singing voice in monaural music signals is based on "two-stage harmonic/percussive sound separation on various resolution spectrograms. This technique relies heavily on the pitch estimation parameter, and the pitch estimation parameter is critical to its effectiveness.

**In [8]** The author suggested an unique unsupervised technique for singing voice separation that uses rank-1 constraint to improve RPCA. The mixer and DSD100 datasets were employed in the experiment, and the findings demonstrated that CRPCA outperforms regular RPCA and WRPCA, especially when time frequency masking is used.

**In [9]** Huang et al. began with the notion that music accompaniments are unimportant and difficult to change, whereas singers are more versatile but scarce in the auditory mix. Based on these assumptions, it is advised that the Robust Principal Component Analysis (RPCA) technique, for solving low-rank and sparse matrices, be utilized.

**In [10]** the work was used to identify three types of folk music: German, Irish, and Austrian. To investigate statistical differences across the various folk, the dataset was evaluated and compared using several Hidden Markov Models (HMM) architectures. The categorization performance is now averaging 77%.

**In [11]** Support Vector Machines (SVM) as a statistical ML approach performed better in MGC than HMM .

**In [12]** They chose frequency-domain properties and low-level traits using a genetic approach. For comparative comparison, different classification approaches such as Nave Bayes (NB), K-nearest neighbor (KNN), and SVM are applied.

**In [13]** The researchers used samples from the GTZAN dataset to distinguish between ten genres. With, we were able to get an ideal classification accuracy of 80.1 %.

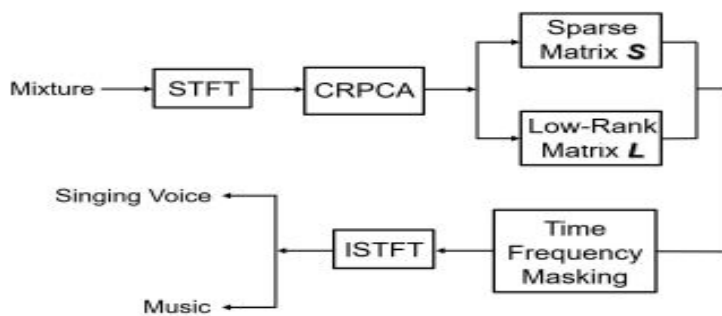
Many different classifiers and feature extraction algorithms are utilised for automatic MGC. Gaussian mixture model (GMM) [14], radial basis function (RBF) [15], AdaBoost [16], and semi-supervised method [17] are the classifiers.

### ***The PROPOSED SYSTEM***

This study discusses an arabic music analysis system that employs artificial intelligence techniques and plays an essential role in retrieving music information (MIR). There are two primary steps in our analytical system: extracting the melody by isolating the singing voice from the music background Then there's melodic analysis to determine the genres that make it up.

**A. Singing Voice (A) Robust Principal Component Analysis is required for separation. Rank-1 Constraint Exploitation**

The suggested CRPCA as a first portion of our analysis system is discussed in this section. Despite the fact that RPCA has been effectively used to the problem of separation of singing voices, in order to lower the nuclear norm, it overlooks the changing SVD characteristic values and computing complexity [18]. To address these two problems, CRPCA, a partial sum minimization of singular value based on rank-1 constraint, is applied. The purpose of RPCA is to maximize the partial sum of singular values by applying a prior rank-1 restriction fully [19]. Due to the rank-1 limitation, In many songs, background music has a greater range of richness than the singer vocal . Figure 1 depicts the processes of the CRPCA separation technique.



**Figure 1. Block diagram of CRPCA technique for separation process**

**• Arabic songs dataset for separation process**

The proposed Arabic music database is the first database that separates Arabic songs. It contains 100 Arabic songs ranging in length from

90 to 120 seconds. It includes Arab singers such as Umm Kulthum, Abdel Halim, Najah and others. These clips are found on the Internet.

• **CRPCA Principle**

CRPCA is a variant of RPCA that separates singing voices using the rank-1 constraint. This is how the model is defined [20] :

$$\begin{aligned} &\text{Minimize } \|S_x\|_* + \lambda_x \|S_y\|_1 \quad (1) \\ &\text{Subject to } S_x + S_y = M_x \end{aligned}$$

**Where :**  $M_x \in \mathbb{R}_{n1 \times n2}$ ,  $S_x \in \mathbb{R}_{n1 \times n2}$ ,  $L \in \mathbb{R}_{n1 \times n2}$ . and  $\|\bullet\|_*$  and  $\|\bullet\|_1$  consider The nuclear norm (single value summation) is denoted. L1 norm (Summation of absolute matrix entry values).  $\lambda_x > 0$  is a trade-off parameter between range L and sparsity S,  $\lambda_k = k/\max(n1, n2)$ .

CRPCA **Algorithm 1** is used to separate singing voices. The value of M is a music signal derived from the audio data observed. After separating them with CRPCA, We can obtain a sparse matrix S and a low-rank matrix L (music accompaniment) (singing voice) [21].

---

**Algorithm 1** CRPCA for singing voice separation

---

**Input:** Mixture signal  $M \in \mathbb{R}_{m \times n}$ .  
 1: **Initialize:**  $\rho > 1, \mu_0 > 0, k = 0, L_0 = S_0 = 0$ .  
 2: **While** not convergence,  
 3: **do** :  
 4:  $L_{k+1} = P_{L, \mu_k^{-1}}(M - S_k + \mu_k^{-1} J_k)$ .  
 5:  $S_{k+1} = Q_{\lambda \mu_k^{-1}}(M - L_{k+1} + \mu_k^{-1} J_k)$ .  
 6:  $J_{k+1} = J_k + \mu_k(M - L_{k+1} - S_{k+1})$ .  
 7:  $\mu_{k+1} = \rho * \mu_k$ .  
 8:  $k = k + 1$ .  
 9: **end while**.  
**Output:**  $L_{m \times n}$ ,  $S_{m \times n}$ .

---

• **CRPCA for separating singing voices**

To improve separation performance, we use ideal binary time frequency masking (IBM) estimate following CRPCA separation to further improve separation outcomes.  $M_{ibm}$  is defined as follows [22]:



$$M_b(y1, y2) = \begin{cases} 1 & |S(y1, y2)| > gain * |L(y1, y2)| \\ 0 & otherwise \end{cases} \quad (2)$$

for all  $y1 = 1 \dots n_1$  and  $y2 = 1 \dots n_2$



**Figure 2. A comparison of RPCA, WRPCA, CRPCA, and CRPCA on SDR, SIR, and NSDR on our dataset with IBM.**

We partition the test dataset using CRPCA and conduct a short-time Fourier transform (STFT) and an inverse STFT (ISTFT). In addition, we employ the IBM technique to improve separation outcomes. Finally, the low-rank matrix L (accompanying music) and the sparse matrix S can be obtained (singing voice).

### **B. Arabic Music Genre Classification Using support vector classification**

One of the primary features that distinguishes music based on a certain set of patterns is genre. The internet's arabic music genres are poorly defined, making computerised categorization of Arabic audio genres difficult. The first stage in this project is to build a well-annotated dataset that includes thirteen of the most well-known Arabic music genres.

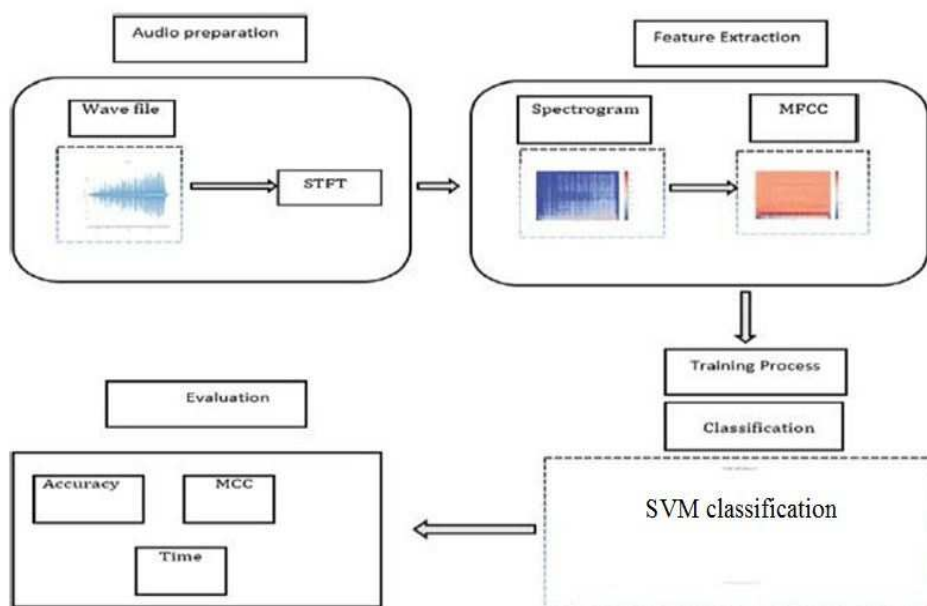


Figure 3. Show the framework of genres classification step

- **Building a Dataset**

The Arabic Music Genre (AMG) dataset's vast corpus is developed by recording various audio snippets. The dataset contains thirteen distinct genre classes. Each music composition is 60 to 120 seconds long and is saved as a 1200 MB wav audio file.

- **WAV Audio File**

Figures 4 and 5 illustrate the waveforms of two AMGs, rast and nahwand, utilizing the x-axis and the y-axis for amplitude. In this case, Both genres' visual representations vividly demonstrate the difficulties in differentiating one from the other.

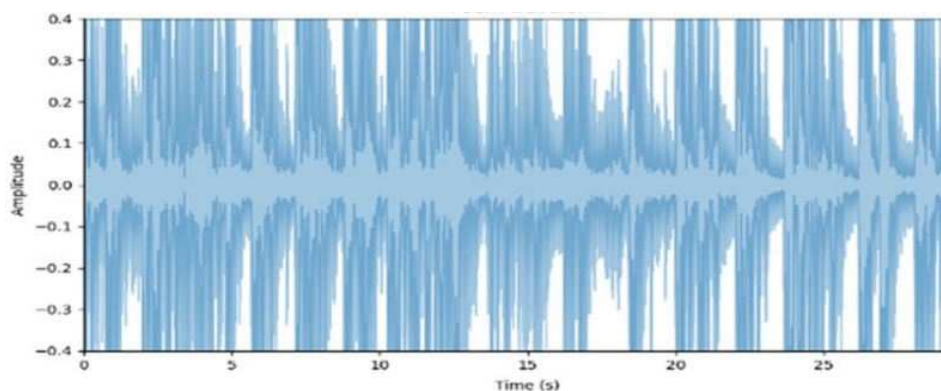


Figure 4. The Nahwand genre's spectrogram

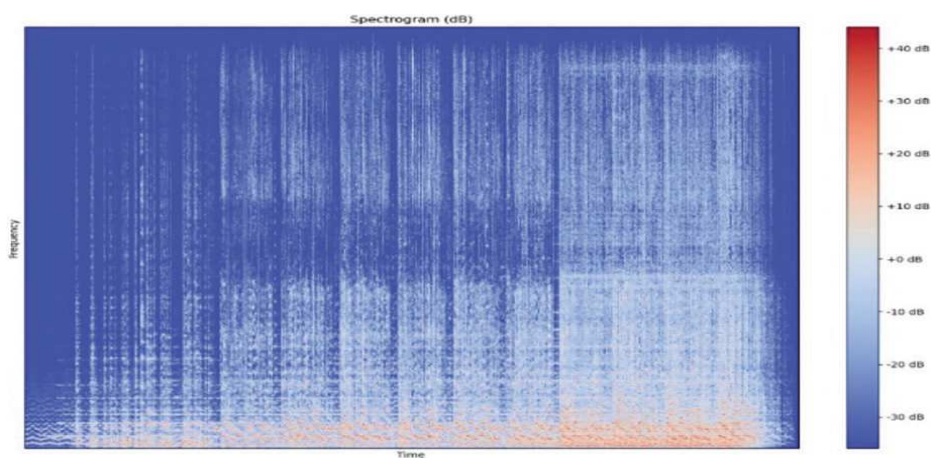


Figure 5. Spectrogram of the Rast genre

The power spectrogram was created using STFT with the following parameters: The sample rate (sr) is 44100.

- The window size (n FFT) is set to 1024.
- 512 is the time interval between frames (hop size).
- Window Function of Hann Window.

• **MFCC feature**

Davis et al. [24] proposed Mel-Frequency Cepstral Coefficient (MFCC) in the 1990s. The MFCC summarizes the frequency distribution over a window size, allowing analysis of the sound's time and frequency

features. It is one of the most widely used cutting-edge feature extraction methods.

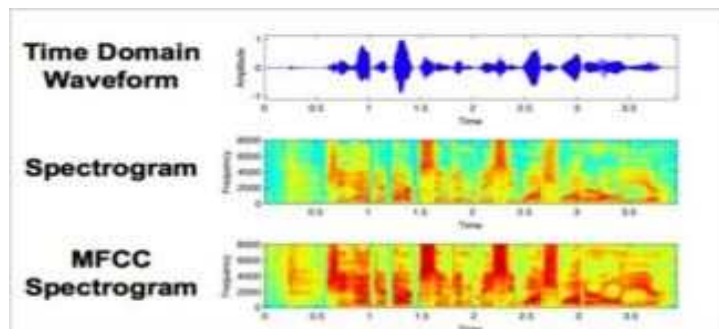


Figure 6 . MFCC Spectrogram

## A. SVM CLASSIFICATION

SVM is a supervised learning methodology that aims to develop a decision-making strategy to better characterize training data by using a specific training data set. The method attempts to use a hyperplane to classify decision boundaries that distinguish data points of different groups and are commonly employed in pattern identification and classification. The hyper plane equation is defined as [26]:

$$N_1 \cdot x_1 + b_1 \quad (3)$$

**Where:**

$N_1$  is weight vector and  $b_1$  is bias.

In this work, multi-class SVM divides the research sample into 35 pieces. There are several techniques available to deal with various groups. In this work, "one vs. other" techniques have been employed to isolate items from one class from those from another [27]. The appropriate hyper level that distinguishes each class from the other elements is calculated. The  $N$  classes are distinguished by a set of binary classifiers, with each training identifying one class from the other. SVM seeks the best separation hyper plane (i.e., a "decision boundary" that separates instances of one class from those of another). Data from two classes can always be separated by a hyper

plane with a suitable nonlinear mapping. The SVM method is precise because it can model complex, nonlinear decision boundaries [28].

### Experimental results

Our analysis system has been applied to two datasets: first one is arabic songs database for singing voice separation which contains male and female vocalists with a sampling rate of 16 kHz and audio clip lengths ranging from 90 to 120 seconds. Data were taken from arabic music tracks that included more than two background music instruments. Second , arabic music genres database for genre identification that comprises 35 classes of arabic music genres.

#### • Singing Voice Separation using Arabic songs dataset

CRPCA generates better separation results than SDR, SIR, and NSDR. Table I displays the execution time of each technique on two identical datasets. According to CRPCA, a previous target rank can be utilized to separate a singing voice from a mixed music signal.

GNSDR is calculated by averaging the NSDRs of each collection over all mixtures, weighted by time, and evaluating the terms of source to artefact ratio (SAR), source to distortion ratio (SDR), and source to interference ratio (SIR) using BSS-EVAL. [29] GNSDR is defined as follows:

$$GNSDR_{v,v,x} = \frac{\sum_{n=1}^N w_n NSDR(\tilde{V}_n, V_n, X_n)}{\sum_{n=1}^N w_n}$$

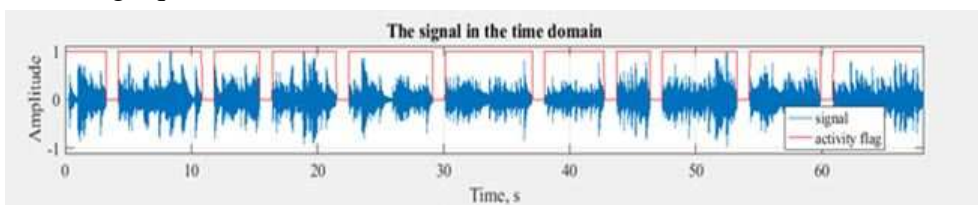
As : N is the total number of tracks,  $w_n$  is the track time, and n is a song index.

Table1. Evaluation table for 4 song files with 90 s duration using CRPCA for separation

Song name	SDR	SIR	SAR	GNSDR
Ya msafer wahdak	1.7360	5.0676	30.8097	2.6610
shaghuly f alhubi badr	2.2495	7.0665	22.1440	1.2282
ana f intizarak	2.0895	6.7545	16.408	3.2028
bent el balad	1.0568	4.9963	20.4936	2.4452

## • Arabic Music Genre Classification

According to the experiment results, pitch with (MFCC) has the best effect for extracting music signal features, and MFCC is a good feature to employ for individual recordings. The accuracy classification achieved by SVM reaches 99 % when the recognition rate is used, as demonstrated in the following equation.



**Figure. 6.4 experimental results to identify music genres at first 40 second of shaghuly f alhubi badr Arabic song**

On our dataset, we tested the recommended approach. The trial findings using SDR, SIR, and NSDR clearly show that CRPCA achieves higher separation outcomes. The process of analysis and classification of races was evaluated from the following equation [30]:

$$\text{Recognition Rate} = \frac{\text{Number of correctly recognized items}}{\text{Total number of items}} * 100\%$$

The results of the evaluation of the experiment indicated the effectiveness of the proposed system, as the accuracy of the results reached to 99%.

## 2. CONCLUSION

The classification of musical genres is difficult in light of the large amount of fresh data produced in an unstructured manner, which is a topic of interest in Music Information Retrieval. Furthermore, certain musical genres are formed of similar sorts of instruments with similar rhythmic patterns, adding to the difficulty. In the future, we hope to create an automatic music genre using deep learning classifier model that might be

included in automatic music information retrieval systems or applications that benefit from music genre information.

### **REFERENCES**

- [1] Sagun, M. ,A., and Bolat , B., (2019) . Classification of Classic Turkish Music Makams by Using Deep Belief Networks , IEEE , 978-1-4673-9910-4/16/31.
- [2] Y.M.D. Chathuranga and K.L. Jayaratne , July (2017) . Automatic Music Genre Classification of Audio Signals with Machine Learning Approaches , GSTF International Journal on Computing , Vol. 3 No.2.
- [3] ] Bahoura, M., 2019. Efficient FPGA-Based Architecture of the Overlap-Add Method for Short-Time Fourier Analysis/Synthesis. Electronics, 8(12), p.1533.
- [4] Bahoura, M. and Ezzaidi, H., (2017) . FPGA implementation of a feature extraction technique based on Fourier transform. 2012 24th International Conference on Microelectronics (ICM).
- [5] Imen Trabelsi and Dorra Ben Ayed , (2016) , “On the Use of Different Feature Extraction Methods for Linear and Non Linear kernels”, 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications , Tunisia , PP.1-8 .
- [6] Mehta , L . , . Mahajan S. , and Dabhade , A. , (2015) , Comparative Study Of MFCC And LPC For Marathi Isolated Word Recognition System, IJSER International Journal of Scientific & Engineering Research, Volume 6, Issue 5, 147 ISSN 2229-5518.
- [7] Wu , J.D., and Lin, B.F. (2009). Speaker identification based on the frame linear predictive coding spectrum technique. Expert Systems with Applications, 36, 8056–8063.
- [8] Childers, D. G. (2000). Speech processing and synthesis toolboxes. New York: John Wiley & Sons.
- [9] Gupta , H. , Gupta , D. , (2016) , LPC AND LPCC METHOD OF FEATURE EXTRACTION IN SPEECH RECOGNITION SYSTEM , IEEE , 978-1-4673-8203-8/16/\$31.00.
- [10] Signal Processing Methods for Identification of Induction Motor Bearing Fault. (2019). International Journal of Recent Technology and Engineering, 8(3), pp.143-151.

- [11] Ali, H., et al , (2014) : DWT features performance analysis for automatic speech recognition of Urdu , Springer Plus , 3:204 [11] Ali, H., et al , (20148) : DWT features performance analysis for automatic speech recognition of Urdu , Springer Plus , 3:204.
- [12] Zahid, S., et al , (2019) , Optimized Audio Classification and Segmentation Algorithm by Using Ensemble Methods , Hindawi Publishing Corporation Mathematical Problems in Engineering , Volume 2015, Article ID 209814, 11 pages.
- [13] D. J. Ghosh B. J. Love, J. Vining, X. Sun, "Automatic Speaker Recognition Using Neural Network", EE371D Intro. To Neural Networks Electrical and Computer Engineering Department, The University of Texas at Austin, Spring 2004, PP. 1-25, 2004.
- [14] S. Haykin, "Neural Networks: A comprehensive Foundation", International Journal of Neural Systems, Vol. 5, No. 4, PP. 363-364, Dec. 1994.
- [15] Bhuvanewari , M . ,(2016) , Gaussian mixture model: An application to parameter estimation and medical image classification Journal of Scientific and Innovative Research , 5(3): 100-105.
- [16] Bahoura, M., and Pelletier, C. 2004. Respiratory sounds classification using cepstral analysis and Gaussian mixture models. In 26th Annual Conference of the IEEE EMBS, San Francisco, CA, September 1–5, pp. 9–12.
- [17] S Goienetxea, I., Martínez-Otzeta, J., Sierra, B. and Mendiáldua, I. (2018). Towards the use of similarity distances to music genre classification: A comparative study. PLOS ONE, 13(2), p.e0191417.
- [18] D. J. Ghosh B. J. Love, J. Vining, X. Sun, "Automatic Speaker Recognition Using Neural Network", EE371D Intro. To Neural Networks Electrical and Computer Engineering Department, The University of Texas at Austin, Spring 2004, PP. 1-25, 2004.
- [19] Chandna, P., Blaauw, M., Bonada, J. and Gomez, E. (2019). A Vocoder Based Method for Singing Voice Extraction. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
- [20] Li, F. and Akagi, M. (2018). Unsupervised Singing Voice Separation Using Gammatone Auditory Filterbank and Constraint Robust Principal Component



- Analysis. 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC).
- [21] Rafii, Z. and Pardo, B. (2011). A simple music/voice separation method based on the extraction of the repeating musical structure. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
- [22] Liutkus, A., Rafii, Z., Badeau, R., Pardo, B. and Richard, G. (2012). Adaptive filtering for music/voice separation exploiting the repeating musical structure. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
- [23] Jansson , A., Humphrey, E. (2017). singing voice separation with deep u-net convolutional networks. Proceedings of the 18th ISMIR Conference, Suzhou, China, October 23-27, 2017.
- [24] Simpson, A., Roma, G. and Plumbley, M. (2015). Deep Karaoke: Extracting Vocals from Musical Mixtures Using a Convolutional Deep Neural Network. Latent Variable Analysis and Signal Separation, pp.429-436.
- [25] Hsu, C., Wang, D., Jang, J. and Hu, K. (2012). A Tandem Algorithm for Singing Pitch Extraction and Voice Separation From Music Accompaniment. IEEE Transactions on Audio, Speech, and Language Processing, 20(5), pp.1482-1491.
- [26] Abouzid, H. and Chakkor, O. (2017). Blind Source Separation Approach for Audio Signals based on Support on Audio, Speech, and Language Processing, 20(5), pp.1482-1491.
- [27] Hu, Y. and Liu, G. (2015). Separation of Singing Voice Using Nonnegative Matrix Partial Co-Factorization for Singer Identification. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23(4), pp.643-653.
- [28] Nasrullah, Z. and Zhao, Y. (2019). Music Artist Classification with Convolutional Recurrent Neural Networks. 2019 International Joint Conference on Neural Networks (IJCNN).
- [29] Liu ,c. and Feng ,L (2019) . Bottom-up Broadcast Neural Network For Music Genre Classification. Elsevier 24 Jan (2019).
- [30] Chillara ,S. and Haldia ,S (2019). Music Genre Classification using Machine Learning Algorithms: A comparison. International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 05 | May 2019.