International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 05, May-2021

Musical Instrument Recognition System with Dynamic Feature Selection Method

Sujith. S, Jithin Kumar I J, Appu.V
GEC Palakkad, Palakkad

***Abstract*— The automatic recognition of musical instruments is a challenging task in the field of music information retrieval (MIR). This article describes an Automatic Musical Instrument Recognition (AMIR) System. Selection of minimum set of features from a pool of features for each instrument is done. In music (polyphonic audio signals) several sound sources are active at the same time. The identification of the instruments present in the audio track provides important information about the composition of music.**

***Index Terms*—Musical instrument recognition, music information retrieval (MIR), acoustics, instrument classification, feature extraction.**

I. INTRODUCTION

One of the main functions lacking in current computer technology is real world awareness. People use a variety of information obtained from the real world through their eyes and ears to assess situations and appropriate behavior in everyday life. However, the computer's ability to recognize acoustic and visual scenes is severely limited.

As human beings, we have a clever ability to recognize sounds acoustically. Although the human ear can hear a sound, that sound is perceived by the human brain or mind. Sounds are presented to the ear and collected by the ear, but it is the brain that organizes the acoustic inputs in such a way that they can be categorized and thus recognized by the consciousness. The recognition of a sound thus depends on a number of processes: the presentation of auditory stimuli, the collection of these stimuli by the ear, the transmission of this stimulus data to the brain, the organization of the data transmitted by the brain, the conscious recognition of the organized data as one certain sound of the mind.

In particular, there have been relatively few attempts to study sound recognition, with the exception of speech recognition studies. Techniques for recognizing a wide variety of sounds, not limited to speech, are important to realizing sophisticated computers that make extensive use of real information. A main reason why it is difficult for computers to recognize audio scenes is that in the real world the audio scenes usually contain several simultaneous sound sources.

Music is an art form whose medium is sound and silence. Its common elements are *pitch* (which governs melody and harmony), *rhythm* (and its associated concepts tempo, meter, and articulation), *dynamics*, and the sonic qualities of *timbre* and *texture*. And music is a good domain for studying computer-aided recognition of audio scenes, since several instruments are usually played at the same time. The difficulty in dealing with music lies in the fact that signals (sources to be recognized) and sounds (sources to be ignored) are not clearly defined. Therefore, several sources should be modeled at the same time and recognized in parallel.

Selecting the functions is arguably the most important step in creating an automatic instrument classifier. Regardless of what type of classifier is used, he does not know in advance the instruments that he needs to recognize - it can only learn from the features it is given. It follows that no classifier can be trained to its full potential if an appropriate set of features from a sufficient set of samples is not used as training data.

In music, the timbre, which in psychoacoustics is also referred to as timbre or tone quality, differs from other notes with a similar pitch and volume. Two notes of the same pitch played on different instruments can be very different from each other, which means that the recognition of an instrument depends on its timbre. Therefore, in order to correctly identify an instrument, we must find a way to accurately measure its timbre. The physical properties of sound that determine how timbre are perceived include the spectrum and envelope.

Since the mid-1990s, machine learning studies have been conducted to address the problem of recognizing musical instruments. These studies did musical instrument classifiers using methods such as multi-layered perceptrons (Nielson et al., 2007), support vector machines (Essid et al., 2006), k-nearest neighbors (Livshin and Rodet, 2004) and self-organizing maps (Cosi et al., 1994) among others. They use a number of characteristics calculated from a selection of sound samples to train and test these classifiers.

From these studies it appears that there is no consensus on which classification method is best. The studies cite different results from different classifiers. More importantly, their experiments vary in terms of the instruments studied and the characteristics used. Some studies cite a large number of instruments, but they were only classified between instrument families, while others only tried to classify between instruments of the same family.

The rest of the paper is organized as follows. In Section II, we present the current state of the art related to the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 05, May-2021

automatic instrument recognition in music. In Section III, we discuss the methods of classification and Section IV proposed method of instrument recognition. In Section V we discuss the results and Section VI discuss the limitations and future work. Section VII, applications and finally Section VIII summarizes this paper with some concluding remarks.

II. Current State of the Art

There are many literature found on automatic instrument recognition systems. And found that the sound is characterized by pitch, loudness and timbre. Timbre is considerate to be distinctive between two instruments playing the same note with the same pitch and loudness. Therefore, there is a direct relationship between timbre and musical instrument identification. The challenge is to determine which attributes characterize best the multidimensional perceptual timbre.

The current ability of computers to recognize auditory events is severely limited when compared to human capabilities. Although computers can accurately detect noises sufficiently close to those trained in advance. Despite the importance of recognizing musical instruments, until recently studies have mainly focused on monophonic sounds. Although the number of studies dealing with polyphonic music has increased, their techniques have not yet reached a sufficient level to be applied to MIR or other real world applications. Our current implementation handles the isolated tone condition well, and we are hoping that it will generalize to still more realistic contexts. The commonly used classifiers used for instrument recognition were the Gaussian [1], gaussian mixture model(GMM) [1] [2] [3], genetic algorithms (GA) [4], multi-layered perceptrons(MLP) [2] [4], artificial neural network(ANN) [5], k-nearest neighbours(k-NN) [6] [2] [7], and support vector machines(SVM) [8] [6] [9] classifiers. While almost all recent speech recognition studies used Hidden Markov Models (HMMs), only a few studies used them to recognize musical instruments [9]. Some introduced hierarchical schemes. A number of studies achieved detection rates of 70-80% for more than 10 target instruments and some achieved around 90%; However, these studies cannot be compared directly because different data and different assessment methods were used.

Although until recently the goals of studies of musical instrument recognition have been monophonic sounds, the number of studies that are now dealing with polyphonic music is increasing.

A. Various Methods of Classification

1) Artificial Neural Network (ANN): In computer science and related fields, artificial neural networks are computational models inspired by animal central nervous systems (in particular the brain) [10] that are capable of machine learning and pattern recognition. They are usually presented as systems of interconnected ”neurons” that can compute values from inputs by feeding information through the network.

The inspiration for neural networks came from examination of central nervous systems. In an artificial neural network,

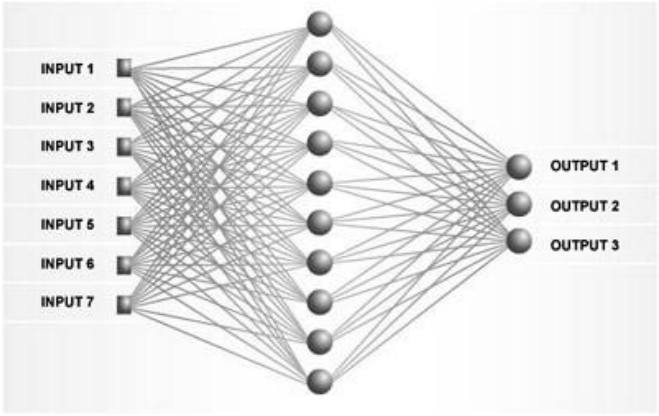

Fig. 1: Artificial neural network.

simple artificial nodes, called *neurons*, *neurodes*, ”processing elements” or *units*, are connected together to form a network which mimics a biological neural network. There is no single formal definition of what an artificial neural network is. Commonly, though, a class of statistical models will be called *neural* if they

1) consist of sets of adaptive weights, i.e. numerical parameters that are tuned by a learning algorithm, and
2) are capable of approximating non-linear functions of their inputs.

The adaptive weights are conceptually connection strengths between neurons, which are activated during training and prediction. Fig. 1 shows a simple neural network with one hidden layer.

2) Support Vector Machine (SVM): In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns that are used for classification and regression analysis. Support Vector Machine creates a hyperplane or set of hyperplanes in high or infinite dimensional space that can be used for classification, regression, or other tasks. A good separation is achieved intuitively by the hyperplane that has the greatest distance to the next training data point of a class (so-called function edge), since in general the generalization error of the classifier is smaller, the larger the edge is. Fig. 2 shows a maximum-margin hyperplane and margins for an SVM trained with samples from two classes [11]. The hyperplanes in the higher-dimensional space are defined as the set of points whose dot product with a vector in that space is constant.

3) k-Nearest Neighbors Algorithm (k-NN): The k-Nearest Neighbors (k-NN) [12] [2] classifier is a typical example of a distance-based classifier. It saves all training examples and then calculates a distance between the test observation and all training observations. The k-NN classifier is easy to implement and can form arbitrarily complex decision boundaries. Hence it was used in many of our simulations. The problem with the k-NN classifier is that it is sensitive to irrelevant features that may dominate the distance metric. In addition, the calculation requires a significant computational load if a large number of training instances is stored.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 05, May-2021

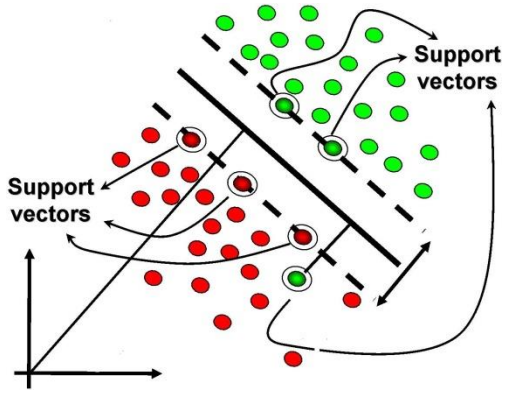

Fig. 2: Support vector Machine.

4) Multi-Layer Perceptron (MLP): Multilayer perceptron (MLP) [2] is a feed forward neural network. Layered perceptrons are networks of multiple layers of interconnected neurons made up of input neurons, output neurons, and hidden neurons, as shown in Figure 3. A forward neural network is one in which the neurons do not form a directed cycle. That is, a neuron in layer i 1 is connected to every neuron in layer i, but not to any other neuron in layer i+1. Learning in multilayer neural networks is done using a technique called error back propagation. Back propagation is a search algorithm for gradient descent and can suffer from both a slow convergence time and from being trapped in local minima. Extensive research has been carried out to improve the back-propagation algorithms. One of these optimization algorithms is the scaled conjugate gradient descent. In our proposed system, the MLP network was back propagation using a scaled conjugate gradient optimization.

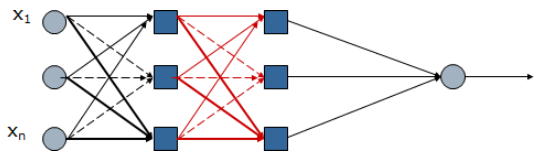

Fig. 3: Multi-Layered Perceptron

5) Gaussian Mixture Model (GMM): A Gaussian mixture model [2] presents each class of data as a linear combination of several Gaussian densities in the feature space. The parameters of the component densities can be iteratively estimated with the well-known Expectation Maximization (EM) algorithm. The EM algorithm is guaranteed to find a local maximum likelihood model regardless of the initialization, but different initializations can lead to different local maxima. It consists of a combined Matlab and C implementations of the basic structure of the model and the EM-algorithm.

III. Proposed Method

The musical instrument recognition system for solo instruments is shown in Figure 4 and can be applied to mixed sounds by recognizing either ensembles instead of isolated instruments.

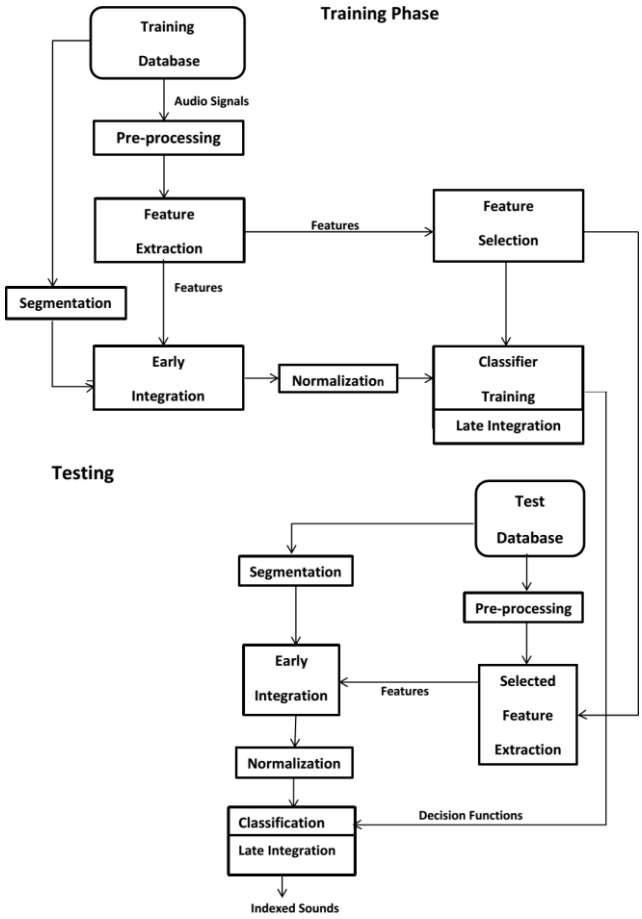

Fig. 4: Architecture of the musical instrument recognition system.

The method used in this thesis for recognizing musical instruments is based on the core system for musical phrases from solo instruments. The system is based on traditional signal analysis over successive overlapping windows. A reduced number of features are then obtained by a feature selection algorithm and it is used to train an ANN classifier. The segmentation module used to break the signal into segments that correspond to musical notes.

A. Principal of Musical Instrument Recognition :

The musical instrument recognition consists of preprocessing, feature extraction, classification and test part. Various operations are performed on the input signal, e.g. B. Removal of the silence part, pre-emphasis, segmentation, framing, windowing, feature analysis and recognition (matching) of the isolated musical phrases. The Musical instrument recognition algorithms consist of two parts i.e. testing and training

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 05, May-2021

phases as mentioned earlier. The block schematic is given in the Fig. 5.

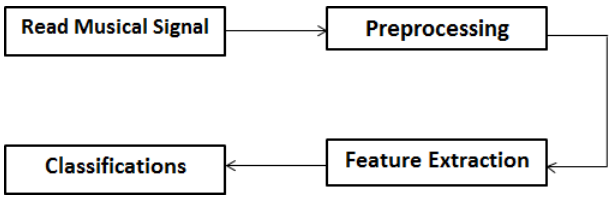

Fig. 5: Block schematic of MI Recognition.

B. Removal of silence part:

Removal of silence part of musical notes is used to reduce the dimension of feature vector and to improve the recognition rate. We have used maximum amplitude feature to remove the silence part. Maximum amplitude of each frame is calculated and based on threshold value the silence part is removed. Silence part removal algorithm is given below.

1) Divide the signal into number of frames.
2)Calculate maximum amplitude for each frame.
3)Define threshold.
4)Compare amplitude of each frame with threshold.

If maximum amplitude of frame is greater than threshold then consider it otherwise it is silence part of signal and eliminates it.

C. Segmentation :

The aim of this segmentation is to obtain segments that contain a single musical note. In such semantically rich segments, the time integration should be more efficient as the features above them should be more pronounced for each instrument.

D. Feature Extraction:

There is no consensual set of features for the instrument recognition problem. Numerous proposals have been made in the past and the strategy is to extract a wide range of potentially useful features in order to select the most relevant to our task, using a feature selection algorithm (FSA) as shown in Fig. 6. Instrument recognition is the Extraction of important features for a better parametric representation of the acoustic model of a musical instrument. The accuracy of this phase is important for the next phase as it affects the detection accuracy. There are many features listed in Table I that can be used to characterize audio signals. Generally they can be grouped into five categories:

1) Temporal
2)Spectral
3)Perceptual
4)Harmonic
5)Statistical and Chroma features .

Mel-frequency cepstral coefficient(MFCC) is based on human hearing perceptions which are linear below 1 KHz and logarithmic above 1 KHz. MFCCs are the way of representing the spectral information of a sound in compact form. There is no standard number of MFCC coefficients for recognizing the tone in any literature. Here we performed the experiments with MFCC to identify the No. of coefficients, 8-14 No. of MFCC coefficients are sufficient to recognize the instruments. MFCC functions are very popular in many language and speaker recognition literatures. The algorithm for getting the MFCC function is given below

1) Step 1: Preemphasis: This step processes the passing of signal through a first order FIR filter which emphasizes higher frequencies. This process will increase the energy of signal at higher frequency.

2) Step 2: Framing: The note of a music signal is divided into frame of 36 ms as their most of spectral characteristics remain same in this duration.

3) Step 3: Windowing: After framing each frames are shaped with hamming window to remove edge effects. Hamming window works better than other window.

4) Step 4: Fast Fourier Transform(FFT): FFT of 1024 point is used to determine spectrum, which is used get log magnitude spectrum to determine MFCC. We have used 1024 point to get better frequency resolution.

5) Step 5: Mel Filter Bank Processing: The 20 mel triangular filters are designed with 50% overlapping .From each filter the spectrum are added to get one coefficient each.

6) Step 6: Discrete Cosine Transform(DCT): DCT of Each mel-frequency cepstum are taken for de-correlation and energy compaction which is called as MFCC. The set of coefficient is called MFCC acoustic vectors. Therefore, each input note is transformed into a sequence of MFCC acoustic vector from which reference templates are generated. One feature extraction process [7] is shown here.

Chroma features are an interesting and powerful representation for music audio in which the entire spectrum is projected on to 12 bins representing the 12 distinct semi tones (or chroma) of the musical octave. Since, in music, notes exactly one octave apart are perceived as particularly similar, knowing the distribution of chroma even without the absolute frequency (i.e. the original octave) can give useful musical information about the audio- -and may even reveal perceived musical similarity that is not apparent in the original spectra.

E. Feature Selection:

The large set of features can be redundant, and some features can be noisy or simply not relevant to class distinction. The feature selection is then essential to reduce the complexity of the problem (by reducing the dimensionality) as well as to eliminate the non-discriminatory features. Instrument recognition for solo sounds has been treated comparatively with many types of instruments. Various acoustic features were used; some were designed based on the knowledge of musical acoustics (e.g., spectral centroid and odd/even energy ratio) and some were used in speech recognition (e.g.,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 05, May-2021

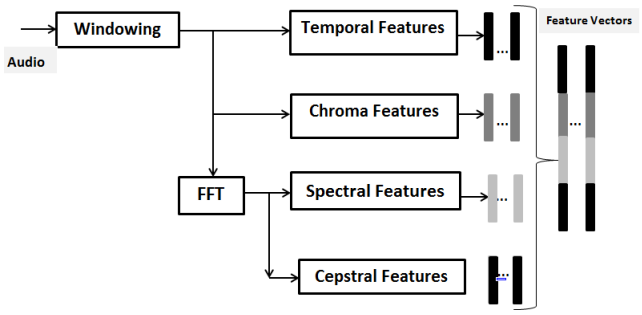

Fig. 6: Feature Extraction Process.

S.No.	Feature
1	Fundamental frequency – f0
2	MFCCs (13 values)
3	Chroma features (12 values)
4	Spectral skewness
5	Spectral spread
6	Temporal centroid
7	Zero crossing rate
8	Energy of signal
9	Spectral roll off
10	Crest factor
11	Spectral flux
12	Spectral centroid

TABLE I: Features selected for this work.

MFCCs). Some studies used techniques to reduce dimensionality or to select features in order to avoid the redundancy of high-dimensional feature spaces. The selected features are shown in Table I and the classifier used is Artificial Neural Networks (ANN). The achieved recognition rate is above 95% for 6 classes of instruments.

IV. Results and Discussion

Experiments were conducted to test the above system. The system is designed in MATLAB R2013a for the recognition of musical instruments. Six classes of musical instruments have been used for studies, and the accuracy of the system is about 95-98%. Some particular set of features selected provides high accuracy rate for one instrument, and some another set of features selected provides high accuracy rate for another instrument. If we provide different set of features for different instruments increases the efficiency of the system considerably. The University of Iowa (MIS) musical instrument samples used for study. The main disadvantage is that many factors affect the characteristics calculated from real sounds. This includes the different playing styles and dynamics that vary the sound spectrum. Very few features are constant across the pitch range of an instrument. In addition, the recording environment affects, samples recorded in an anechoic chamber are well recognized, whereas more realistic environments, or synthetic samples pose much extra difficulty for the task. The result is shown in Table II.

V. Limitations and Future Work

The system recognition rate is increased by selecting different set of features for different instruments. Single set of features is not enough to get same high accuracy for all instruments. So if we need to add more instruments for recognition, we should add more features that we omitted in this study. In such case the complexity of the system will also increase. Selection of minimum set of features for entire musical instrument recognition system is not relevant as we are selecting different features for different instruments; the accuracy of the system is increased. Maximum number of features should be added to the pool of features in this instrument recognition system, and then selection of minimum number of features from the pool of features for each instrument to be recognized is to be done. It will provide most reliable and accurate instrument recognition system. As we are selecting different features for different instruments, the accuracy is increased.

The main challenge for the construction of musical instrument recognition systems is increasing their robustness. Many factors influence the features calculated from real sounds. These include the different playing styles and dynamics that vary the sound spectrum. Very few features are constant across the pitch range of an instrument. Instruments radiate sound unevenly at different directions. In addition, the recording environment affects, samples recorded in an anechoic chamber are well recognized, whereas more realistic environments, or synthetic samples pose much extra difficulty for the task. The problem of generalizing is by no means a trivial one: the system must recognize different pieces of violin as belonging to the same class and different members of the string family as a part of the string class.

Although until recently the goals of studies on the recognition of musical instruments have been monophonic sounds, the number of studies now dealing with polyphonic music is increasing. The first major difficulty with instrument recognition in polyphonic music (a mixture of several instrument sounds) is the fact that the sounds contained in the mixture interfere with each other, and this interference makes it difficult to extract acoustic features from the sounds accurately and robustly. If a clean sound could be achieved for each instrument using sound separation technology, instrument detection for polyphonic music would be the same as detection of monophonic sound for each instrument. In practice, however, it is difficult to separate a mixture of sounds without distortion. When some partials (harmonic components) of some sounds in the mix overlap in frequency, the separation is very difficult and therefore the acoustic features extracted from the mix differ significantly from those extracted from monophonic sounds. The second main difficulty with instrument recognition in polyphonic music is that the preliminary recognition processes (e.g. beginner recognition and F0 estimation) are not

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 05, May-2021

sufficiently reliable for polyphonic music. In some frameworks, instrument recognition was performed for each note. They had to estimate the start time and F0 of each note in order to extract the segment corresponding to the note before identifying the instrument for the note. However, onset detection and F0 estimation for polyphonic music are still challenging problems and their errors can adversely affect instrument recognition. The detection of musical instruments in polyphonic music takes place with the help of the Missing Feature Approach [1], the dynamic model of the spectral envelope [13] and the joint modeling of continuous and attack sounds [14] etc..

VI. APPLICATIONS

This section refers to some of the main uses of automatic musical instrument recognition systems. From the MIR (Musical Information Retrieval) point of view, such a system can be implemented in any music indexing context or using general musical likeness. Tag propagation, recommendation, or playlist generation systems, to name a few, conceptually use the information to instrument a piece of music. In addition to the pure administrative capabilities of large archives, music indexing also opens up opportunities for educational aspects. Music students can search sound archives for compositions that contain a specific solo instrument; or search for the appearance of certain instruments or instrumental combinations in a musical recording.

VII. CONCLUSION

We have described a system that can hear and recognize a musical instrument. The work began by examining human perception: how well people can recognize different instruments and what underlying phenomena occur in the auditory system. Then we studied the characteristics of musical sounds that distinguish them from one another, as well as the acoustics of musical instruments. The knowledge of the perceptibly prominent acoustic cues that may be used by human test persons for recognition was the basis for the development of feature extraction algorithms.

Selection of minimum set of features for entire musical instrument recognition system is not relevant as we are selecting different features for different instruments; the accuracy of the system is increased. Maximum number of features should be added to the pool of features in this instrument recognition system, and then selection of minimum number of features from the pool of features for each instrument to be recognized is to be done. It will provide most reliable and accurate instrument recognition system. As we are selecting different features for different instruments, the accuracy is increased.

REFERENCES

[1]D. Giannoulis and A. Klapuri, "Musical instrument recognition in polyphonic audio using missing feature approach," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 9, pp. 1805–1817, 2013.

[2]S. Gunasekaran and K. Revathy, "Recognition of indian musical instruments with multi-classifier fusion," in *International Conference on Computer and Electrical Engineering*, 2008, pp. 847–851.

[3]G. Hall, H. Hassan, and M. Bahoura, "Hierarchical parametrisation And classification for musical instrument recognition ," in *International Conference on Information Science, Signal Processing and their Applications*, 2012, pp. 1039–1044.

[4] R. Loughran, J. Walker, and M. O'Neill, "An exploration of genetic algorithms for efficient musical instrument identification," in *International Conference on Signals and Systems*, 2009, pp. 1–6

[5] C.-J. Lin, C.-L. Lee, and C.-C. Peng "Chord recognition using neural networks based on particle swarm optimization" in *The 2011 International Joint Conference on Neural Networks (IJCNN)*, 2011, pp 821-827.

[6] J. Yu, X. Chen, and D. Yang, "Chinese folk musical instruments recognition in polyphonic music," in *International Conference on Audio, Language and Image Processing*, 2008, pp.1145-1152.

[7]A. Azarloo and F. Farokhi, "Automatic musical instrument recognition using k-nn and mlp neural networks," in *International Conference on Computational Intelligence, Communication Systems and Networks*, 2012, pp. 289–294.

[8]S. Essid, G. Richard, and B. David, "Instrument recognition in polyphonic music based on automatic taxonomies," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 68–80, 2006.

[9]C. Joder, S. Essid, and G. Richard, "Temporal integration for audio classification with application to musical instrument classification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 1, pp. 174–186, 2009.

[10][Online]. Available: https://www.coursera.org/course/neuralnets

[11][Online]. Available: http://nlp.stanford.edu/IR-book/html/htmledition/support-vector-machines-and-machine-learning-on-documents-1.html

[12][Online]. Available: http://www.statsoft.com/textbook/k-nearest-neighbors

[13]J. Burred, A. Robel, and T. Sikora, "Polyphonic musical instrument recognition based on a dynamic model of the spectral envelope," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 173–176.

[14]J. Wu, E. Vincent, S. Raczynski, T. Nishimoto, N. Ono, and S. Sagayama, "Polyphonic pitch estimation and instrument identification by joint modeling of sustained and attack sounds," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1124–1132, 2011.