Journal of Theoretical and Applied Information Technology
31st December 2014. Vol.70 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# AUTOMATIC CLASSIFICATION OF AUDIO DATA USING GRADIENT DESCENT NEURAL NETWORK BASED ALGORITHM

[1]LAKSHMI DEVASENA C, [2]M HEMA LATHA

[1]Assistant Professor, Operations & Systems, IBS Hyderabad, IFHE University, INDIA

[2]Department of Software Systems, Karpagam University, INDIA

E-mail: [1]devaradhe2007@gmail.com, [2]hema.bioinf@gmail.com

**ABSTRACT**

Audio mining is a technique by which the core of an audio signal can be automatically searched and analyzed. This research work addresses feature extraction from audio and audio similarity measures and proposes an algorithm to mine any type of audio data. This Hybrid Algorithm for Audio Mining (HAAM) consists of two different phases. The first phase named Training Phase, takes Training Audio data as input and then Sonogram, Spectrum Histogram, Periodicity Histogram and Fluctuation Pattern for the audio data are calculated. Then the features are extracted using Mel Frequency Cepstral Coefficient (MFCC) and the essential features are reduced from these for further processing. Then, classification is done by using a supervised classification technique. Then the features of the audio files are trained using Gradient Descent Adaptive Learning Back propagation Network. The second phase is the Testing Phase, which takes Testing Audio data as input and then preprocessing is done and the features are extracted. Then classification is done using the findings from the trained data by matching the features of the test audio with the equivalent or closest feature audio classes. The working prototype of this algorithm has been implemented and tested. During the experiments, Instrumental music data have been used as input and the system performed well and classified the input music in an accurate manner and presented the results obtained. The system is suitable for classifying the different types of audio data and can be used in many applications including speech recognition, audio classification in scientific research and engineering, audio data comparison and detection in audio surveillance applications.

**Keywords:** *Audio Mining, Gradient Descent Neural Network, K-Means Classifier, Mel Frequency Cepstrum Coefficient, Principal Component Analysis*

## 1. INTRODUCTION

In digital information storage and acquisition technology, the swift evolution has led to faster and remarkable data stored in data stores and databases. Even valuable information may be hiding behind the data; the awesome data volume makes it complicated for human beings to extort them without powerful tools. Multimedia mining systems, extracting semantically meaningful information (knowledge) from multimedia files without human intervention is increasingly in demand. This research work addresses a neural network based audio clustering and classification model which can be used in any multimedia mining system. Audio mining is a practice by which the content of an audio signal can be automatically searched and analyzed. It is most regularly used in automatic speech recognition and automatic audio signal recognition area, where the analysis tries to recognize any speech within the audio.

There are four important issues concerned with designing similarity measures related to audio. First it is obligatory to decide which features to be extracted from the audio signal. This depends on the targeted concept (e.g. Rhythm, timbre, melody, harmony etc.) And involves psycho-acoustics, signal processing and music perception. The second issue is to determine how to summarize and describe a piece of music based on the extracted features. Often pieces are very inhomogeneous. Simply calculating the mean of the extracted features is generally not useful. The third issue is to evaluate the representation of one piece with another. The fourth issue is how much computational efficiency is achieved. It is not possible to model every nerve cell in the human auditory system when processing music archives

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

with terabytes of data. This research will address these issues and will propose a neural network based model for audio classification.

In this research work, a paradigm for similarity based audio clustering and classification is addressed using statistical and machine learning techniques. The final algorithm can be used for audio mining related applications. For implementing the idea the most advanced audio pattern extraction techniques and artificial neural networks will be used. The figure 1 shows the general idea of the typical audio classification system. Audio signals can be seen as a subset of the larger topic of digital signals in general. While sounds manifest themselves as waves in the physical medium of the atmosphere, they are represented digitally as sample points in the cartesian coordinate plane, with the y-axis representing amplitude and the x-axis representing time. An example can be seen in figure 2.

Although some features can be extracted directly from the raw time-domain signal, most auditory pattern recognition and context inference has made use of either the spectral data or features extracted from the spectrum. Studies have made use of either the standard spectrum [22], [16] or one adjusted to match the sensitivity profile of human hearing [13and 20].

## 2. LITERATURE REVIEW

Changsheng Xiang and ZiYing Zhou [1] proposed a Music Classification Method based on Back Propagation Neural Network which is providing Classification accuracy of up to 88.1% for classifying different types of music genres (like rock, pop, Chinese zither and folk song). Chungsoo Lim and Joon-Hyuk Chang [2] suggested the Adaptive Kernel Function of Support Vector Machine Algorithm that adaptively tunes classification of Speech/Music frames from 3GPP2 Selectable Mode Vocoder (SMV). Shruti Aggarwal and Naveen Aggarwal [3] compared different algorithms (ID3, SVM, Linear SVM, J48, Functional Tree Classifier and Naïve Bayes Classifier) for the classification of Audio Data and concluded that Linear Support Vector Machine Algorithm and Naïve Bayes Classification algorithm have given better results by taking the audio data of five different sets including languages like Hindi, English and Punjabi. In [4], the second order Multi-scale Scattering coefficients can be used to bring great improvement in the Mel Frequency Cepstral Coefficients of the Audio data for the best Classification results is shown. A pragmatic model for Clustering and Classification of Instrumental Music using Principal Component Analysis with Multilayer Perceptron Neural Network is proposed in [5] and classified eight different types of Instrumental music with the classification accuracy of 79%. In [6], two different voice recognition algorithms to improve the voice recognition performance, namely the Mel Frequency Cepstral Coefficient (MFCC) which is a non-parametric method and non-linear sequence alignment known as Dynamic Time Warping (DTW) Techniques are discussed. Voice Recognition System implemented based on MFCC and Dynamic Time Warping (DTW) is discussed in [7].

A proficient speech recognition system with the assimilation of MFCC feature with frequency sub-band decomposition passed to the HMM network is proposed and the results are compared to the existing MFCC method in [8]. Wei Chu [9] discussed about the Auditory-Based Noise-Robust algorithms for Classification of audio data. A mixed type audio (music along with speech) classification system based on Support Vector Machine (SVM) using the different measures like Variance of Zero crossing rate, Silence Ratio, Harmonic Ration and Sub-band Energy is presented in [10] and comparison of results made with other classifiers like k Nearest Neighbor(k-NN), Neural Network (NN), and Naive Bayes (NB). The state-of-art algorithms of classification of music based on audio data is described in [11]. A supervised learning algorithm for classifying recorded music using the AdaBoost algorithm for Music Classification is presented in [12]. The classification performance of techniques used in recent works is summarized in Table 1. The proposed Hybrid Algorithm for Audio Mining (HAAM) is aimed to classify any types of audio data with the improvement in classification accuracy.

## 3. METHODOLOGY AND DESIGN

### 3.1. Sound Similarity Metrics

#### 3.1.1. Fluctuation patterns

Fluctuation Patterns (FPs) describe characteristics of the audio signal which are not described by the spectral similarity measure are described by FPs. Figure 3 shows some examples of FPs. The vertical lines signify reoccurring periodic beats.

Journal of Theoretical and Applied Information Technology
31st December 2014. Vol.70 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

#### 3.1.2. Sone

Each 6-second sequence is cut into overlapping frames with a length of 46ms. FFT is computed for each frame. According to a model of the outer and middle-ear, the frequency bins are weighted to emphasize frequencies around 3-4 kHz and smother very low or high frequencies and they are grouped according to the critical-band rate scale with the unit Bark into frequency bands. A spectral masking model is applied to smooth the spectrum. At last, the loudness is computed with a non-linear function.

#### 3.1.3. Spectrum histograms

Spectrum Histograms (SHs) [23] are the simple approach to summarize the spectral shape and based on a bark/sone representation of the audio signal. By counting how many times each loudness level of a piece of audio exceeded in each frequency band the Spectrum Histogram is described.

#### 3.1.4. Periodicity histograms

Periodicity Histograms (PHs) were originally presented in the context of beat tracking [24]. The PHs are used to describe periodically reoccurring beats. Further processing of the bark/sone representation of audio is used to extract the features.

### 3.2. Mel Frequency Cepstrum Coefficient

In this work, to compare music similarity and to form the music similarity matrix Spectrum Histograms, Periodicity Histograms and Fluctuation Pattern and MFCC Coefficient are used. The extraction of the best parametric representation of acoustic signals is an important task to produce a better recognition performance. MFCC is based on human hearing perceptions which cannot recognize frequencies above 1 KHz. That is, MFCC is based on the known disparity of the human ear's decisive bandwidth with frequency. MFCC has two different types of filter spaced linearly at low frequency below 1000Hz and logarithmic spacing above 1000Hz. Subjective pitch present on Mel Frequency Scale is used to capture important characteristics of phonetic in speech. The entire process of the MFCC is shown in Figure 4 [6].

### 3.3. Principal Component Analysis

In this research, a statistical technique called Principal Component Analysis (PCA) is used for plotting the multidimensional music feature matrix data into a three dimensional space. The dimensionality reduction is used for visualizing the data in the three dimensional space. A data set xi, (i = 1,…,n) is summarized as a linear combination of orthonormal vectors (called principal components):

$f(\mathbf{x},\mathbf{V}) = \mathbf{u} + (\mathbf{x}\mathbf{V})\mathbf{V}^T$ , where $f(\mathbf{x},\mathbf{V})$ is a vector valued function, u is the mean of the data { xi } , and V is a matrix with orthonormal columns. The mapping provides a low-dimensional projection of the vectors xi if . The PCA estimates the projection matrix V minimizing

$$R_{emp}(\mathbf{x},\mathbf{V}) = \frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{x}_i - \mathrm{f}(\mathbf{x}_i,\mathbf{V})\right\|^2$$

Reducing dimensionality combines variables that have a linear relationship, therefore reducing two variables to one or few.

### 3.4. K-Means Classifier

K-means is one of the effortless unsupervised learning algorithms. This algorithm classifies the given data set using a predefined number of clusters using the feature vectors obtained after feature reduction by PCA.

Steps:

1. Identify a centroid for k different clusters. These centroids should be placed in an intelligent way to minimize the number of iterations.

2. Take each point belonging to the given data set and associate it to the nearest centroid.

3. When no point is pending, the first step is completed and a first set of clusters has been formed. At this point, new centroid for all the 'k' clusters is found as barycenters of the clusters, resulting from the previous step.

4. After finding these k new centroids, a new binding has to be made between the same data set points and the nearest new centroid.

5. An iterative process has been generated. As a result of this iteration process, the centroid of each cluster changes its location step by step until no more changes are needed.

### 3.5. Gradient Descent Adaptive Learning Back propagation Network

Neural networks are made up of simple elements operating in parallel. A neural network can be trained to carry out a particular function by fine-tuning the values of the connections (weights) between elements. The network is adjusted based on the comparison of the target and the actual output, until the network's actual output matches

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the given target. In supervised learning, typically many such input/target pairs are used to train a network. Incremental training of the network changes the weights and biases of the network as needed after the presentation of each individual input vector. Incremental training is referred to as "adaptive" training. After initialization of weights and biases, the network is ready for training. The network can be trained with training set for function approximation. The training process involves a set of examples of proper network behavior - network inputs p and target outputs t.

### 3.6. Model of the Proposed Training Scenario

The model for the training scenario of the proposed HAAM algorithm is depicted in Figure 5. The Training process in the proposed algorithm consists of the steps below.

1. Read all the audio files from the training set.

2. Extract the features from the audio file and prepare the feature matrix.

3. Classify the audio files using the supervised classification algorithm.

4. Construct a Multi-layer neural Network and train it with the multi-dimensional feature matrix data using Gradient Descent Adaptive Learning Back propagation algorithm.

### 3.7. Model of the Proposed Testing Scenario

As The model for the testing scenario of the proposed HAAM algorithm is depicted in Figure 6. The Testing process in the proposed algorithm consists of the steps listed below.

1. Open a new audio file and preprocess it.

2. Extract the features and create the feature matrix for the input audio file.

3. Present the feature data to the trained neural network.

4. Find the class label of the input audio file.

## 4. PROPOSED HYBRID ALGORITHM FOR AUDIO MINING (HAAM)

The proposed HAAM Algorithm has the two phases as described below.

**Training Phase:**

Input: Training Audio data.

Stage (1): Preprocessing.

Find Sonogram, Periodicity Histogram, Spectrum Histogram and Fluctuation Pattern for the training audio data.

Stage (2): Feature Extraction

Window the audio data with a Hamming window.

Amplitude values of the DFT of the data are found.

Amplitude values are converted to filter bank outputs.

Log base 10 for the output is calculated.

Find the cosine transform.

Feature vectors are stored in matrix X.

Stage (3): Features reduction

Calculate the empirical mean of X.

The deviations are calculated from the mean and the data are stored in the matrix B [M N].

Covariance matrix C is found.

Eigenvectors and eigenvalues of the covariance matrix C are found.

The eigenvectors and eigenvalues are rearranged to form the feature vector.

The new data set is derived and the eigenvectors with the highest eigenvalues are projected into space.

Put the new dataset in a matrix Y

Stage (4): Classification

Select 'm' initial "means" randomly from the data set Y.

'm' clusters are created by associating every observation with the nearest mean.

The centroid of each of the 'm' clusters becomes the new means.

The above steps are repeated until convergence is reached.

Stage (5): Training

Train the features of the audio files using Gradient Descent Adaptive Learning Back propagation Network.

Output: Trained Audio dataset.

**Testing Phase:**

Input: Test Audio data.

Journal of Theoretical and Applied Information Technology
31st December 2014. Vol.70 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Stage (1): Preprocessing.

Find Sonogram, Spectrum Histogram, Periodicity Histogram and Fluctuation Pattern for the audio data

Stage (2): Feature Extraction

Window the audio data with a Hamming window.

Amplitude values of the DFT of the data are found.

Amplitude values are converted to filter bank outputs.

Log base 10 for the output is calculated.

Find the cosine transform.

Feature vectors are stored in matrix X.

Stage (3): Features reduction

Calculate the empirical mean of X.

The deviations are calculated from the mean and the data are stored in the matrix B[M N].

Covariance matrix C is found.

Eigenvectors and eigenvalues of the covariance matrix C is found.

The eigenvectors and eigenvalues are rearranged to form the feature vector.

The new data set is derived and the eigenvectors with the highest eigenvalues are projected into space.

Put the new dataset in a matrix Y.

Stage (4): Classification

Unknown samples are classified into closest class based on it's feature value using the trained neural network.

Output: Classified audio data.

## 5. IMPLEMENTATION AND RESULTS DISCUSSION

### 5.1. Interface Design

The proposed Hybrid algorithm for Audio Mining has been implemented using Matlab and the interface has been designed as shown in Figure 7.

### 5.2. Results and Discussion

The algorithm has been designed to extract the features of the training audio data given and train the system using Gradient Descent Adaptive Learning Back propagation Network with the extracted features of the given audio data. When the test audio is given as input, the trained network classifies the test audio data into different predefined groups of audio. For testing the performance of the algorithm, the instrumental music data have been taken as sample input and experimented with 14 different classes of it. When the 'Train Neural Network' button is pressed, the Training Phase of the HAAM algorithm is executed. This phase takes the audio data, extracts its features and reduces the features and then classifies the data according to the features extracted and then train the neural network using Gradient Descent Adaptive Learning Back propagation Network is shown in Figure 8.

When the 'Classify Files' button is pressed the test audio is given as input, its features are extracted and based on the features extracted, the trained network classifies the test audio into different classes (Figure 9). Here, 14 different classes of Instrumental music are taken for classification. It has been classified with an accuracy of 84.38% by the trained network after extracting the feature. When the 'Plat Audio in 3D space' button is pressed, the test audio data are plotted on 3D space (Figure 10). In a similar manner, a single unknown audio file can also be classified according to its feature value. Sample result of single audio file classification is shown in Figure 11 and Figure 12.

## 6. PERFORMANCE COMPARISON OF PROPOSED HAAM ALGORITHM WITH EXISTING ALGORITHMS

Totally 14 different classes of Instrumental music data (Accordion, Bassoon, Cello, Clarinet, Drums, flute, Guitar, Harmonium, Mandolin, Oboe, Piano, Quena, Sitar and Banjo) have been taken as input audio and a classification accuracy of 84.38% is achieved using HAAM Algorithm. For one to six different classes of audio data, 100% classification accuracy is achieved. For seven different classes of audio data, 98.61% classification accuracy is arrived. For eight different classes, HAAM algorithm gives 97.56% of classification accuracy. For nine different classes, 93.55% of classification accuracy is arrived. For ten different classes, 91.18% of classification accuracy is reached. For eleven and twelve different classes, HAAM algorithm produces 93.04% & 89.84% of classification accuracy. For thirteen different classes of audio, 85.92% of classification accuracy is produced. The performance analysis of proposed HAAM algorithm is shown in Table 2. The Number

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

of Audio data used in Different number of Audio data, Number of correctly classified Audio data for different number of audio classes, Rand Index Arrived at each time, Time taken for finding the Feature Matrix and the Classification Accuracy Produced by the HAAM algorithm for different number of Audio classes are summarized in Table 2. Performance comparison of the HAAM algorithm for different number of Audio classes based on Accuracy is depicted in Figure 13.

The confusion matrix obtained by the proposed HAAM Algorithm is shown in Table 3.

In Table 3, 1 to 14 represents the instrumental music classes of Accordion, Bassoon, Cello, Clarinet, Harmonium, Mandolin, Oboe, Quena, Sitar, Banjo, Piano, Guitar, Drums and Flute respectively.

The existing audio mining methods (AdaBoost.Tree and AdaBoost.Stump [Igor Vatolkin and Wolfgang Theimer, 2007], Support Vector Machine, J48 and FT Classifier [Shruti Aggarwal and Naveen Aggarwal, 2011], PCA+Multilayer Perceptron [Hemalatha. M et al, 2010]), are compared with the proposed HAAM algorithm in terms of classification accuracy and the proposed HAAM algorithm attained more classification accuracy than those methods. It is summarized in Table 4. From Table 4, it is clear that the proposed HAAM algorithm earns the highest classification accuracy.

## 7. CONCLUSION

This research work addresses techniques for feature extraction from audio and audio similarity measures and proposed an algorithm for music clustering and classification. The working prototype of an Audio classification system has been implemented and tested in MATLAB. During the experiments, the proposed system performed well and classified the input music accurately and the results arrived were remarkable and comparable. The accuracy of the proposed algorithm has been compared with the earlier works, in terms of Correctly Classified Instances, Rand Index and Classification Accuracy. The accuracy of the proposed work is high when compared to the earlier works. The proposed work is suitable for classifying the different types of audio data and can be used in many applications including speech recognition, audio classification in scientific research and engineering, audio data comparison and detection in audio surveillance applications. Limitations here are only 14 classes of Instrumental music are taken with 160 total files. Even the proposed algorithm provides more accuracy, when the number of classes increases, the accuracy decreases. Future work is planned to increase the number of classes and the number of files for testing as well as fine tuning the algorithm to get more accuracy.

**REFRENCES:**

[1] Changsheng Xiang and ZiYing Zhou, “A New Music Classification Method based on BP Neural Network”, *Int. J. of Digital Content Technology and its Applications*, 5(6), 2011, pp. 85-94.

[2] Chungsoo Lim and Joon-Hyuk Chang, “Adaptive Kernel Function of SVM for Improving Speech/Music Classification of 3GPP2 SMV”, *ETRI Journal*, 33(6), 2011, pp. 871-879.

[3] Shruti Aggarwal and Naveen Aggarwal, “Classification of Audio Data using Support Vector Machine”, *Int. J. of Computer Science and Technology*, Vol. 2, Issue 3, 2011, pp. 348-405.

[4] Joakim And´en and St´ephane Mallat, “Multiscale Scattering for Audio Classification”, *12th Int. Society for Music Information Retrieval Conf. (ISMIR 2011)*, 2011, pp. 657-662.

[5] M.Hemalatha, N.Sasirekha, S.Easwari and N.Nagasaranya, “An Empirical model for Clustering and Classification of Instrumental Music using Machine Learning Technique”, *IEEE Explore*, 2010, pp. 1-5.

[6] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, “Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques”, *J. of Computing*, vol. 2, Issue 3, 2010, pp. 138-143.

[7] Anjali Bala, Abhijeet Kumar and Nidhika Birla, “Voice Recognition System based on MFCC and DTW”, *International Journal of Engineering Science and Technology*, Vol. 2(12), 2010, pp. 7335-7342.

[8] Ibrahim Patel and Y. Srinivas Rao, “Speech recognition using hmm with MFCC – An Analysis using Frequency spectral Decomposion technique”, *Signal & Image Processing : An International Journal (SIPIJ)*, 1(2), 2010, pp. 101-110.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[9] Wei Chu, "Auditory-Based Noise-Robust Audio Classification Algorithms", *Ph.D Thesis, McGill University*, September 2008.

[10] Lei Chen, Sule Gunduz and M. Tamer O¨zsu, "Mixed Type Audio Classification with Support Vector Machine", *ICME 2006*, *IEEE*, 2006.

[11] Igor Vatolkin and Wolfgang Theimer, "Introduction to Methods for Music Classification Based on Audio Data", *Research Center NRC-TR-2007-012*, 2007.

[12] James Bergstra, Norman Casagrande, Dumitru Erhan, Douglas Eck and BalazsKegl, "Aggregate Features and AdaBoost for Music Classification".

[13] Brian Clarkson, Nitin Sawhney, and Alex Pentland, "Auditory context awareness via wearable computing", *In Proceedings of the Workshop on Perceptual User Interfaces (PUI'98)*, San Francisco, CA, November 1998.

[14] Cooley. J.W and J.W. Tukey, "An algorithm for the machine calculation of complex Fourier series", *Mathematics Computation*, 19, 1965, pp.297–301.

[15] Ben Gold and Nelson Morgan, "Speech and Audio Signal Processing: Processing and Perception of Speech and Music", *John Wiley & Sons*, 2000.

[16] Marcel Levy, "Ringer mute: Automated phone detection using an artificial neural net", *Unpublished paper produced as part of coursework for CS793k at the University of Nevada*, Reno 2004.

[17] Lie Lu and Hong-Jiang Zhang, "Content analysis for audio classification and segmentation", *IEEE Transactions on Speech and Audio Processing*, 10, October 2002, pp.504–516.

[18] Steven W. Smith, "The Scientist and Engineer's Guide to Digital Signal Processing", *California Technical Publishing*, 1997.

[19] Steven W. Smit,. "Digital Signal Processing: A Practical Guide for Engineers and Scientists", *Elsevier Science*, 2003.

[20] Savitha Srinivasan, Dragutin Petkovic and Dulce Ponceleon, "Towards robust features for classifying audio in the cuevideo system", *In Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, ACM Press New York, NY, USA, 1999, pp. 393–400.

[21] Haleh Vafaie and Kenneth De Jong, "Feature space transformation using genetic algorithms", *IEEE Intelligent Systems*, 13(2), March/April 1998, pp.57–65.

[22] Brian Westphal and Jim King, "A genetic algorithms based automatic phone-ring detection system", *Unpublished paper produced as part of coursework for CS790k at the University of Nevada*, Reno, 2003.

[23] Pampalk. E,S. Dixon, and G. Widmer, "Exploring music collections by browsing different views", *In Proceedings of ISMIR*, 2003.

[24] Scheirer. E. D., "Tempo and beat analysis of acoustic musical signals", *JASA*, 1998.

Journal of Theoretical and Applied Information Technology
31st December 2014. Vol.70 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

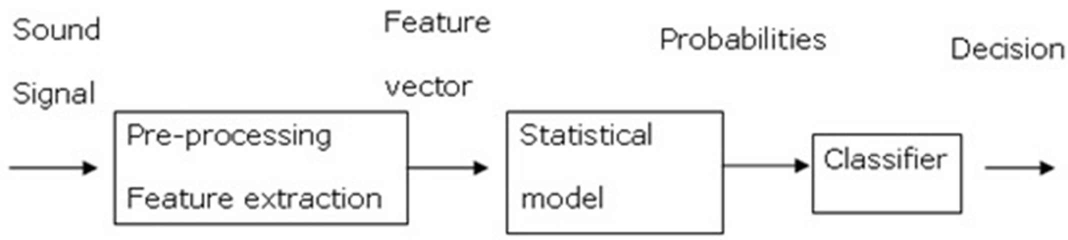


*Figure 1: Typical Audio Classification System*

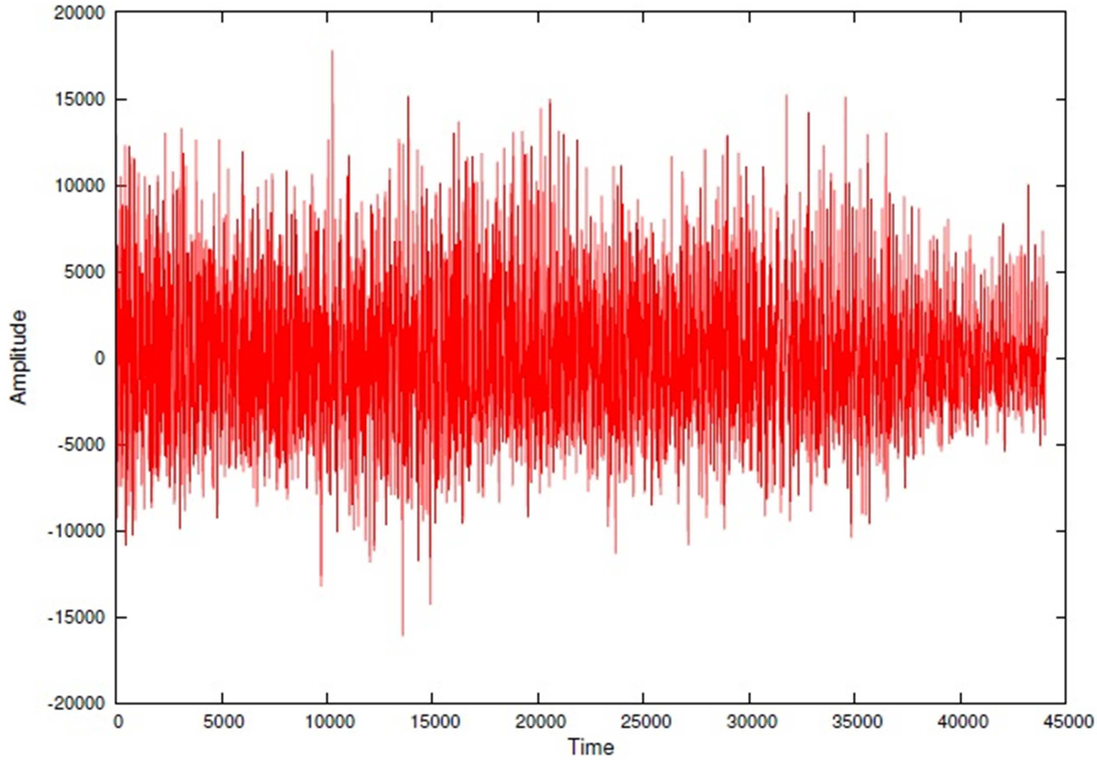


*Figure 2: An Example Of The Digital Representation Of A Raw Audio Signal*

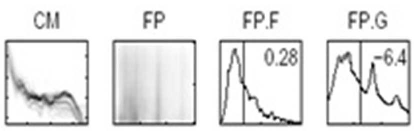


*Figure 3: Fluctuation Patterns*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

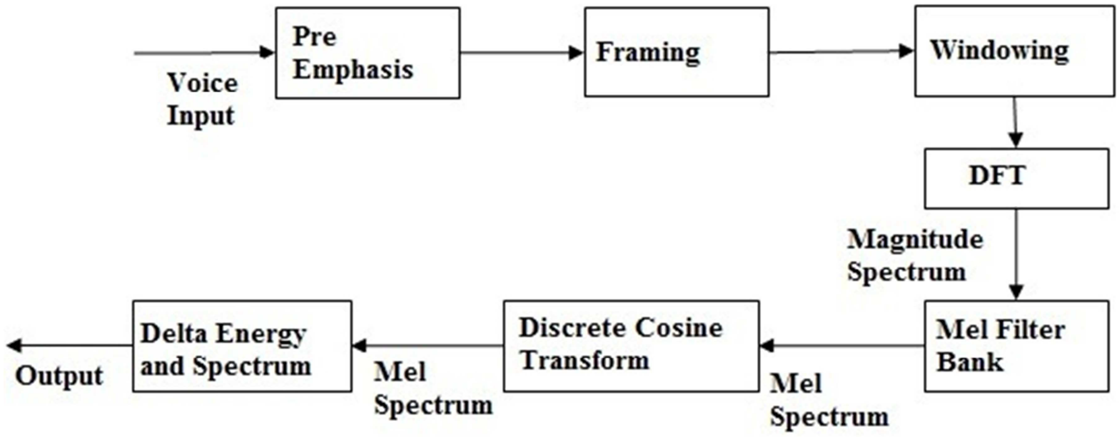


*Figure 4: MFCC Block Diagram*

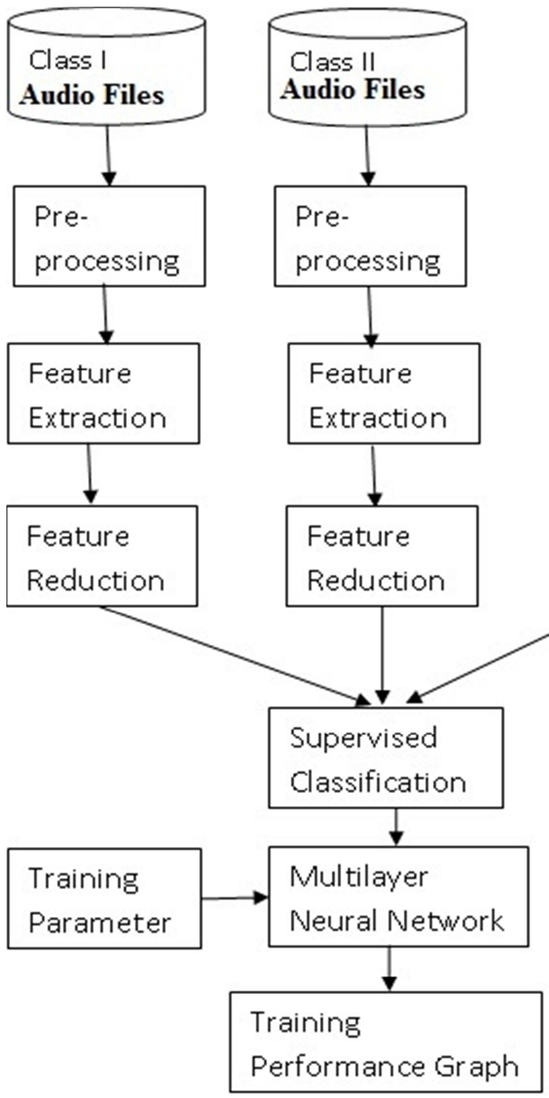


*Figure 5: Schematic Diagram For Training Scenario In HAAM*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

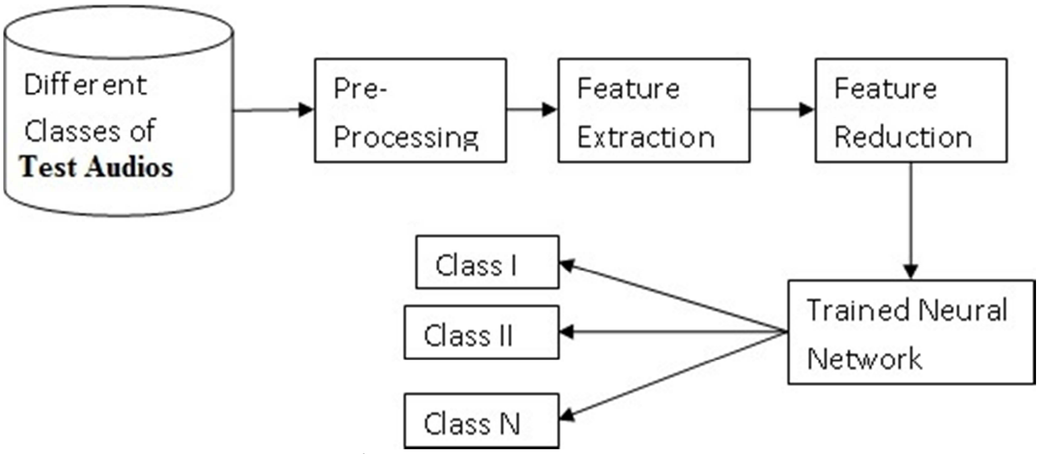


*Figure 6: Schematic Diagram For Testing Scenario In HAAM*

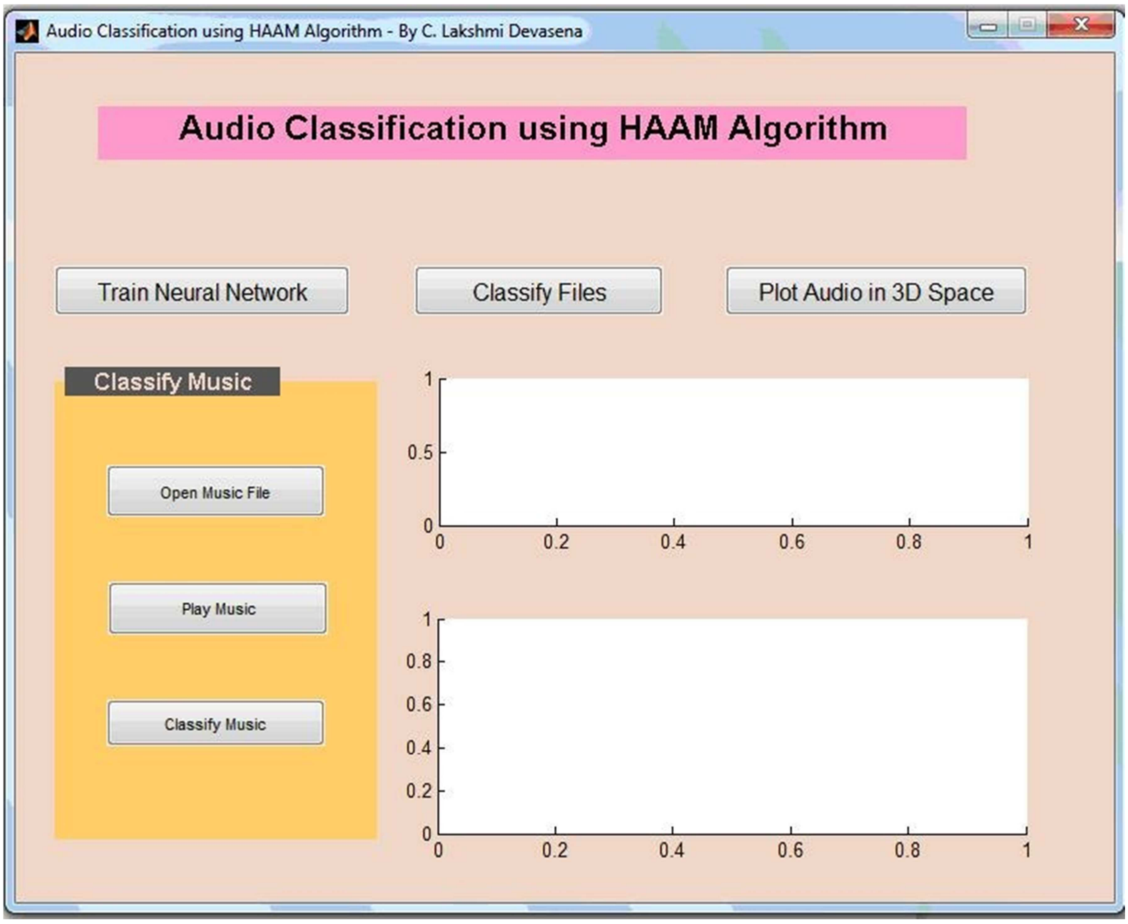


*Figure 7: Interface For Audio Mining System Using HAAM*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

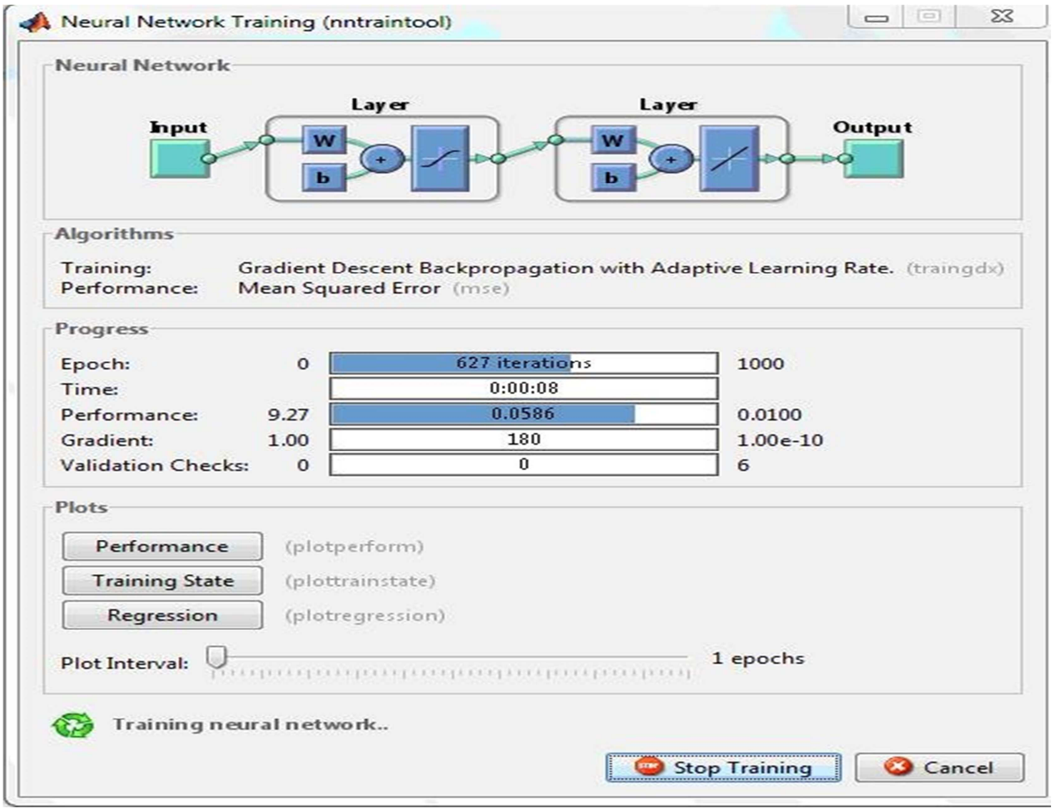

Figure 8: Training The Neural Network Using Gradient Descent Adaptive Learning Back propagation

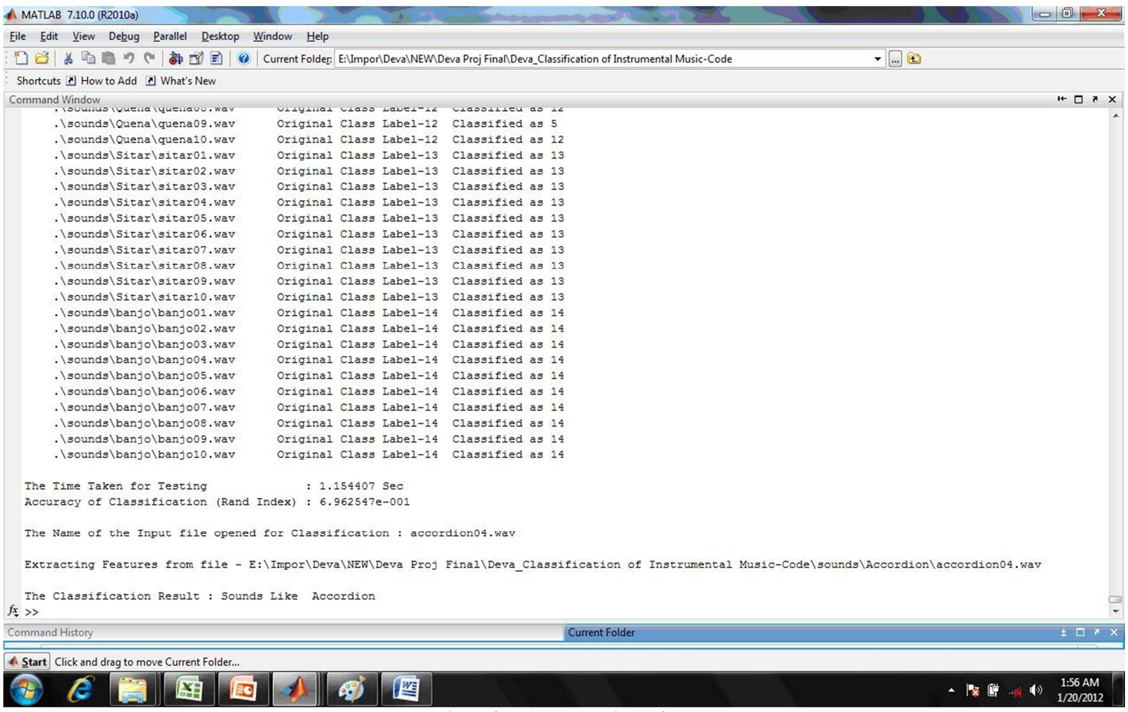

Figure 9: Classification Results Of Testing Process

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

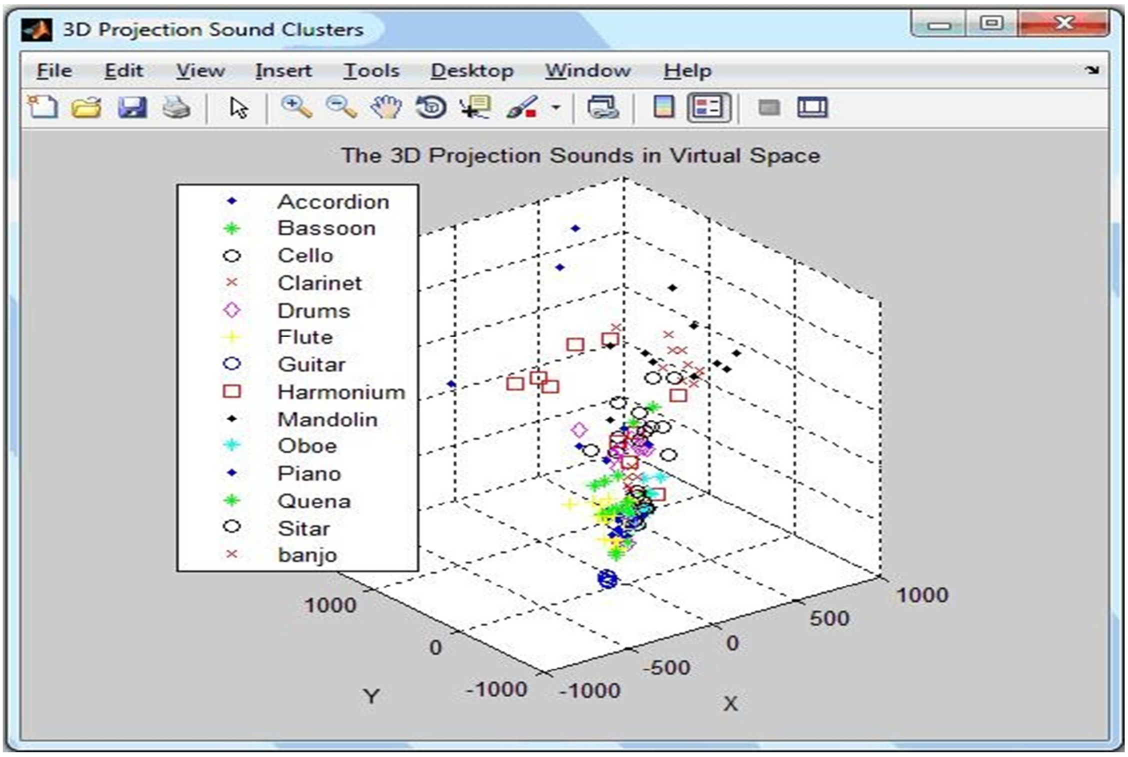


Figure 10: 3D Projection Of Audio Files In Virtual Space

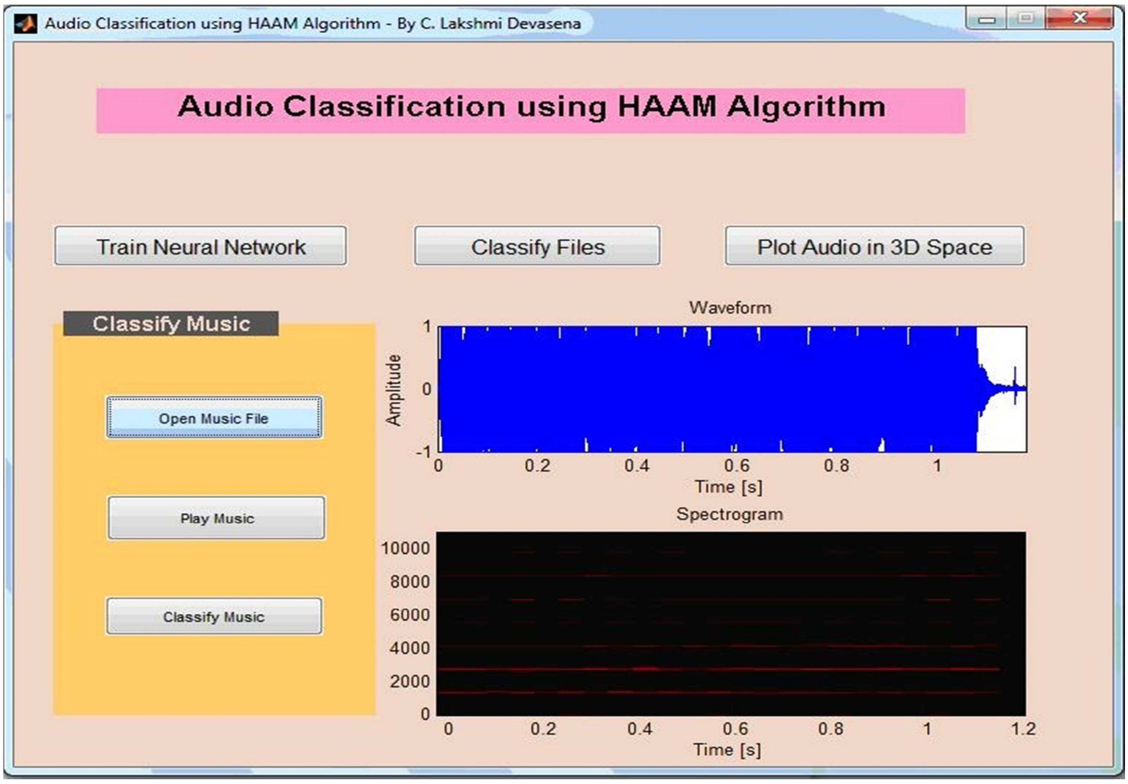


Figure 11: Unknown Audio File Selected For Testing

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

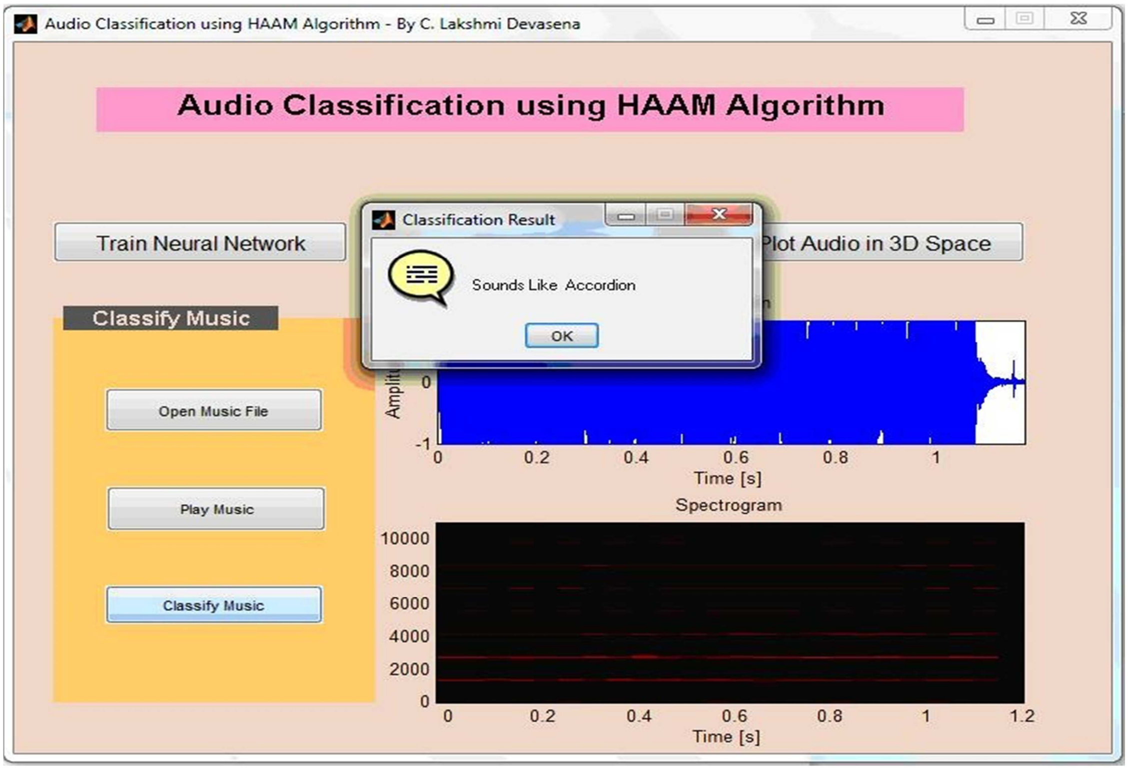


Figure 12: Unknown Audio File Classified Correctly During Testing

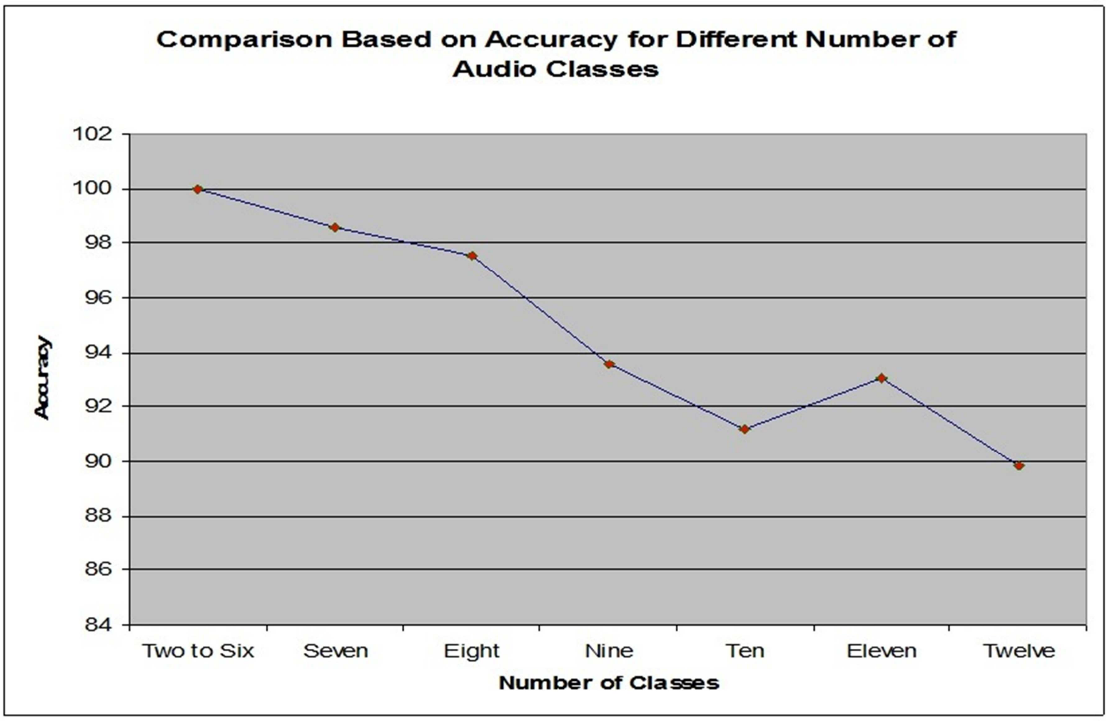


Figure 13: Classification Performance Of Proposed System Based On Accuracy For Different Number Of Audio Classes

Journal of Theoretical and Applied Information Technology
31st December 2014. Vol.70 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Table 1: Classification Performance Of Techniques Used In Recent Works For Audio Classification*

| Technique Used | Data Set Used | No. of Classes | Accuracy |
|---|---|---|---|
| AdaBoost.Tree | Genre and Artist | 6 | 86.92 |
| AdaBoost.Tree | Genre and Artist | 10 | 75.1 |
| AdaBoost.Stump | Genre and Artist | 6 | 86.29 |
| AdaBoost.Stump | Genre and Artist | 10 | 74.71 |
| PCA+Multilayer Perceptron NN | Instrumental Music | 8 | 79 |
| Linear SVM | Environment sound, Music | 10 | 73.17 |
| SVM | Audio type of eng, Hindi | 5 | 88.3 |
| Naïve Bayes | Audio type of eng, Hindi | 5 | 83.18 |
| J48 | Audio type of eng, Hindi | 5 | 86.65 |
| ID3 | Audio type of eng, Hindi | 5 | 81.54 |
| FT (Functional Trees Classifier) | Audio type of eng, Hindi | 5 | 85.37 |
| BP Network | Genre | 4 | 88.1 |

*Table 2: Classification Performance Of Proposed HAAM Algorithm*

| S.No | Number of Classes | Number of Instances | Correctly Classified | Rand Index | Accuracy (%) | Time Taken for finding Feature Matrix |
|---|---|---|---|---|---|---|
| 1 | 2 | 22 | 22 | 1 | 100 | 6.47 |
| 2 | 3 | 32 | 32 | 1 | 100 | 8.58 |
| 3 | 4 | 42 | 42 | 1 | 100 | 12.48 |
| 4 | 5 | 52 | 52 | 1 | 100 | 15.10 |
| 5 | 6 | 63 | 63 | 1 | 100 | 18.02 |
| 6 | 7 | 72 | 71 | 0.9675 | 98.61 | 21.07 |
| 7 | 8 | 82 | 80 | 0.9469 | 97.56 | 23.62 |
| 8 | 9 | 92 | 87 | 0.8859 | 93.55 | 27.79 |
| 9 | 10 | 102 | 93 | 0.863 | 91.18 | 31.06 |
| 10 | 11 | 115 | 107 | 0.854 | 93.04 | 35.68 |
| 11 | 12 | 128 | 115 | 0.821 | 89.84 | 38.15 |
| 12 | 13 | 142 | 122 | 0.7134 | 85.92 | 41.41 |
| 13 | 14 | 160 | 135 | 0.6963 | 84.38 | 45.35 |

*Table 3: Confusion Matrix Obtained By The Proposed HAAM Algorithm*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 11 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| **2** | 0 | 6 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 10 |
| **3** | 0 | 0 | 5 | 1 | 1 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 10 |
| **4** | 0 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 10 |
| **5** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |
| **6** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 10 |
| **7** | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 10 |
| **8** | 0 | 1 | 1 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 10 |
| **9** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 10 |
| **10** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 10 |
| **11** | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 10 | 0 | 0 | 0 | 13 |
| **12** | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 |
| **13** | 0 | 0 | 0 | 0 | 13 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 |
| **14** | 0 | 2 | 0 | 0 | 0 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 18 |
| **Total** | 11 | 9 | 7 | 10 | 17 | 19 | 16 | 11 | 10 | 12 | 12 | 6 | 10 | 10 | 160 |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Table 4: Classification Performance Comparison Of Proposed Method With Other Recent Works*

| Technique Used | Data Set Used | No. of Classes | Classification Accuracy |
|---|---|---|---|
| AdaBoost.Tree | Genre and Artist | 6 | 86.92 |
| AdaBoost.Tree | Genre and Artist | 10 | 75.1 |
| AdaBoost.Stump | Genre and Artist | 6 | 86.29 |
| AdaBoost.Stump | Genre and Artist | 10 | 74.71 |
| PCA + Multilayer Perceptron NN | Instrumental Music | 8 | 79 |
| Linear SVM | Environment sound, Music | 10 | 73.17 |
| SVM | Audio type of eng, Hindi | 5 | 88.3 |
| Naïve Bayes | Audio type of eng, Hindi | 5 | 83.18 |
| J48 | Audio type of eng, Hindi | 5 | 86.65 |
| ID3 | Audio type of eng, Hindi | 5 | 81.54 |
| FT (Functional Trees Classifier) | Audio type of eng, Hindi | 5 | 85.37 |
| BP Network | Genre | 4 | 88.1 |
| Proposed HAAM | Instrumental Music | 4 | 100 |
| Proposed HAAM | Instrumental Music | 5 | 100 |
| Proposed HAAM | Instrumental Music | 6 | 100 |
| Proposed HAAM | Instrumental Music | 8 | 97.56 |
| Proposed HAAM | Instrumental Music | 10 | 91.18 |