DOI: 10.21917/ijsc.2019.0273

ALGORITHM BASED CLASSIFICATION OF SONGS

Deepti Chaudhary[1], Niraj Pratap Singh[2] and Sachin Singh[3]

[1,2]Department of Electronics and Communication Engineering, National Institute of Technology Kurukshetra, India
[3]Department of Electrical and Electronics Engineering, National Institute of Technology Delhi, India

Abstract

The emotional content perceived from music has great impact on human beings. Research related to music is attaining more and more recognition not only in the field of musicology and psychology but also getting attention of engineers and doctors. The categorization of music can be carried out by considering various attributes such as genres, emotional content, mood, instrumental etc. In this work Hindi music signals belonging to different genres are categorized in four quadrants belonging to Arousal-Valence (AV) plane of emotion categorization i.e. Positive Valence-Positive Arousal (PV-PA), Positive Valence- Negative Arousal (PV-NA), Negative Valence-Positive Arousal (NV-PA) and Negative Valence-Negative Arousal (NV-NA). Features related to music are calculated by using MIR toolbox and the classification techniques used are K-Nearest Neighbor (K-NN), Naive Bayes (NB) and Support Vector Machine (SVM). In this work authors make use of two types of annotation techniques i.e. Subject Based Annotation (SBA) and Algorithm Based Annotation (ABA). The accuracy, precision and recall are considered as evaluation parameter in this work. The evaluation parameters for all the classification techniques using both the annotation methods are compared in the proposed work. Results reveal that SVM classifier outperforms other two classifiers in terms of the parameters considered for both SBA and ABA and it has also been proved that algorithmic annotation outperforms subjective annotation.

Keywords:

Music Emotion Recognition, Human Computer Interaction, MIR Toolbox, Arousal, Valence

1. INTRODUCTION

Most of the people enjoy music in their day to day life. It is an entertainment and pleasure for human's activity [1]. Now a day, huge collections of music is available online. Depending upon the user's interest the user can download and hear music from various internet sources [2]. Musical component has five elements of music such as melody, interval, rhythm, harmony and pitch, which establish the human's emotion and psychological changes [3] [4]. The music can be expressed by considering human's facial expression, pose, speech, which expresses their present state of the user's emotion. There are various types of emotion such as aggressiveness, angry, happiness, sadness, relaxation that can be evoked or felt from music [5]. Music can also be distinguished from images. In the system the images can be captured through the webcam application. According to the task the user ideal playlist is obtained [6]. Emotion is the sensation provoked by concentrating the music. Emotion can be categorized as expressed perceived and feel. The expressed emotion is induced by the singer whereas perceiving and feel are associated with the listener's predictions and feeling while listening the song [7]. Music emotion recognition (MER) is the classification technique to sense the emotion from the music. The emotion considered for MER can be divided into two categories: dimensional approach and categorical/discrete approach [7]. Dimensional approach expresses their emotions in a Cartesian space with Arousal-Valence (AV) model. The arousal model describes high or low value of emotion. The valence model describes positive or negative value of emotion [8]. The dimensional approach are described by Robert Plutchik's, Russel's and Thayer's emotion model illustrated by AV plane. Discrete approach expresses various emotions by using adjectives and trained classifier is used to forecast the overall emotion of the song [9]. Emotion classification can be classified into two steps. They are feature extraction and classifier learning. The aim of feature extraction of music signal is characterized by detection of minimum set of distinguished parameters through different algorithms. Classifier learning is used to relate the features of various classes to emotions under the effect of minimizing prediction error and improved accuracy. The basic automatic music emotion recognition (AMER) system consists of five steps described below.

1.1 MUSIC DATABASE COLLECTION

A large database consisting of all the genres related to different languages is used for AMER. The database should not be from the same album and of same singer and artist. The database should be collected widely from various albums and websites for research. The music information retrieval evaluation exchange (MIREX) is the evaluation campaign for music information techniques and systems coordinated by International Symposium on music information retrieval (ISMIR) annually. The aim of this symposium is to provide the exploration to various techniques and algorithms for research in MIR. MIREX introduces the automatic music classification (AMC) task in 2007 [10] [11]. The database available with MIREX can be used by the researchers by signing the agreement for not sharing the database commercially.

1.2 PREPROCESSING

As the emotion perceived from the song is not constant throughout the entire song and it varies from segment to segment, a short time segment of the song is considered for the research. The song length considered by researchers to avoid emotion variation is 25-45 seconds. If the length of the clip is less than this range then for such short duration clips the emotion cannot be judged correctly and for longer clips the emotion of within the song is not stable. It has been noted from the review that a 30 segment clip is common choice [7].

The music clips are also not available in standard format. It is the prime requirement to convert the music database in standard format for their comparative analysis. The standard format that is normally considered by researchers is 22050Hz maximum frequency and 44100Hz sampling frequency by keeping in view the frequency range of audio signals i.e. 20Hz to 20KHz and 16 bits precision and mono-channel. Music clips also undergo the normalization process. In this process the windowing and framing techniques are used.

1.3 ANNOTATION

Annotation is subjective analysis of the database to categorize it in different classes. Emotion related to music varies from person to person, Thus it is considered as a subjective concept. The database is collected from various sources and annotated by a group of subjects. . The subjective analysis of the music clips can be carried out either by a group of experts or untrained group of people based or subject based [12]. The expert group consists of less number of people generally less than five who have in-depth knowledge of music and are employed for the task of annotation [13]. In untrained group the annotation task is given to more than ten people and each song is annotated by the whole group. The average opinion of the subjects is considered as final category of the music clip. The annotation process can be carried out by considering either categorical approach or dimensional approaches of emotion classification as described in sections 1.3.1 and 1.3.2.

Huron described the four parameters style, genre, emotion and similarity on the basis of which classification of music can be carried out [14] [15]. In the research field of music the keywords used for emotions are well defined by psychologists and they use the words that are used by human beings to express their emotion [16] [17]. From the literature study two main types of approaches: categorical and dimensional are identified to define emotion models. Categorical approach is defined discretely and makes use of clusters using adjective terms to define the emotion by various researchers and dimensional approach is defined dimensionally and represents the emotions on the basis of their positions on the emotion planes.

1.3.1 Categorical Approach:

The relationship between emotion and music is explored by Hevner in 1936 and described the discrete cluster of emotions using adjectives to represent emotions in eight different categorical clusters [18] [19]. Cluster 1 consists of the terms such as spiritual, lofty, inspiring, dignified, sacred, solemn sober and serious. Cluster 2 consists of the terms including pathetic, doleful, sad, mournful, tragic, melancholy, frustrated, depressing, gloomy, heavy and dark. Cluster 3 consists of the terms such as dreamy, yielding, tender, sentimental, longing, yearning, pleading and plaintive. Cluster 4 consists of the terms such as lyrical, leisurely, satisfying, serene, tranquil, quit and soothing. Custer 5 consists of humorous, playful, whimsical, fanciful, quaint, sprightly, delicate, light and graceful. Cluster 6 consists of the terms such as merry, joyous, happy, cheerful and bright. Cluster 7 consists of the terms like exhilarated, soaring, triumphant, dramatic, passionate, sensational, agitated, exciting, impetuous and restless. Cluster 8 consists of the terms such as vigorous, robust, emphatic, martial, ponderous, majestic and exalting. The emotional clusters formed by Hevner were re explored by Farnsworth by using ten groups of emotional in 1958. Ekman broadly categorize the basic emotions universally in six basic classes happy, sad, anger, fear and disgust from which all the other emotions can be obtained [20]. Fransworth regrouped the Hevner's adjectives and proposed ten groups to describe emotion.

1.3.2 Dimensional Approach:

Dimensional approach used for emotion categorization is based on their positions on emotion plane. The dimensions on the plane are given by considering the relationship between basic factors that are used to differentiate the emotions. The placement of the emotion on dimensional graph depends on their correlation between the axes scales and the large number of terms is used to describe the varying emotions on the bases of their variability on axes of emotion plane. In 1980 Robert Plutchik proposed first 2-dimensional wheel model and 3-dimensional model in cone-shape to represent relationship between different types of emotions [21]. Authors considered eight basic types of emotions: anger, disgust, fear, joy, sadness, surprise, anticipation and trust by arranging them circularly as shown in Fig.1. The emotions are represented by different colors in this model. As shown in Fig.1 similar colors are used to represent the similar type of emotions with variable strengths and terms representing opposite emotions are placed against each other. The emotions shown in table in different colors can be mixed up to obtain different emotion. By using this basic differentiation of emotion along the axes various dimensional models are proposed by authors and these models represents the emotions in continuous plane by considering two or three dimensions. These dimensions are related with valence, arousal and dominance. Valence term deals with the positive and negative types of emotional terms, arousal term deals with the energy or stimulation level of song and dominance deals with the level of measuring strength of influencing power.

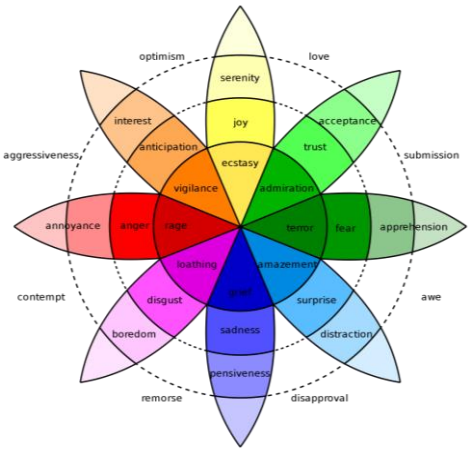

Fig.1. Plutchik's model of emotion [21]

The three dimensions pleasure, arousal and dominance related to emotion were described by A. Mehrabian and J.A. Russell in 1974 [22]. A two dimensional circumplex model of emotion had been proposed by Russel in 1980 [23] [24]. In this model valence and arousal are considered as dimensions. The horizontal dimension of the model is related with positive and negative emotions and vertical dimension of the model is related with positive arousal and negative arousal as shown in Fig.2. The same types of emotions are placed in the same quadrant and opposite emotions are place in the opposite quadrant. For example the first quadrant of the model deals with positive arousal-positive valence emotions covering the emotions such as happy, glad, delighted excited etc., second quadrant deals with positive arousal-negative valence types of emotions covering the emotions such as angry, tense, frustrated etc., third quadrant is related with negative arousal-negative valence type emotions such as sad, bored, tired etc. and fourth quadrant consists of negative arousal and positive valence type of emotions such as calm, relax, satisfied etc.

Authors make use of 28 adjective terms related to emotion in four different ways by making use of Ross technique to obtain the model. This technique is used for ordering the variables in circular pattern, implementation of a multidimensional scaling technique on similar emotional terms and one-dimensional dimensional scaling on presumed degree of valence and arousal dimensions.

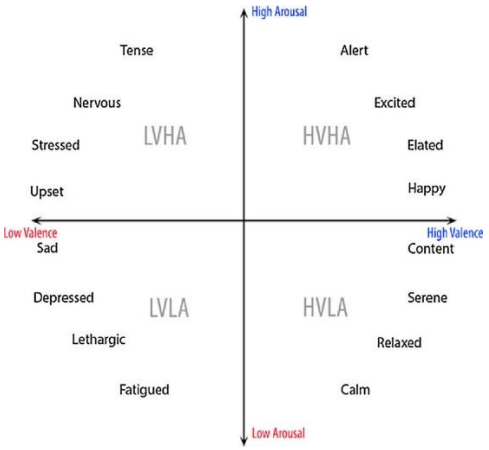

Fig.2. Russel's 2'd model of emotion [23]

Normal mood states are defined in 1990 by Dr. Thayer. Author described the mood states in two dimensions and represented the dimensions of approach as tension and energy [25]. The author described the mood categories in relation with energy and tension and represented them in 2D plane. Further the mood are also analyzed related to different activities such as exercise, sleep, depression, cognition and other daily routine activities related to person. Bigand et al. [26] suggested the third dimension in terms of a continuity-discontinuity or melodic–harmonic contrast instead of dominance and Fontaine et al. [27] proposed fourth dimension to represent emotion. The fourth dimension suggested by author is related to resemblance and contrast among the terms used to describe emotion [27]. The general consent of research community is based on two dimensions till now.

1.4 FEATURE EXTRACTION

A large number of features are related to music to represent the clips in different dimensions. Various toolboxes are available for feature extraction. Music Information Retrieval (MIR) toolbox [28], Psysound [29], Marsyas [30] etc. are used to extract dynamics, rhythm, timbre, pitch and tonality related to sound signals. Dynamics of a song clip is the alteration in loudness of notes. Rhythm represents the pattern of notes of different strength in a song. Timbre defines the quality of notes of a song. Pitch represents the perceived frequency of song clips. Tonality represents the arrangement of pitches in a proper order.

The feature normalization techniques are used after feature extraction for fair comparison of value of each feature. Feature normalization can be carried out by two methods.

- *Linear Normalization* - In this method, the range of each feature is set between zero and one [0,1].
- *Z-Score Normalization* - In this method, each feature is normalized to zero mean value and the standard deviation is set to unity.

Before classification process the feature selection techniques are applied to minimize the number of random variables by selecting the principal variables from the features that are extracted. The various techniques such as Principal Component Analysis, Linear Discriminant Analysis, Relief etc can be used to select the appropriate features from all the extracted features [7].

1.5 CLASSIFICATION

Further classification process is considered as vital part for the AMER system. The performance of the system depends strongly on how the classification process is carried out for the system. Various types of classification processes that are used by the researchers are explained in this section.

1) Regression models are also used as classifiers to determine relationship between dependent and independent variables. The performance parameter for regression models is R2 statistics that is used to fit the data to the regression line [31].
2) SVM classifiers are based on supervised learning techniques and algorithms. The dataset is divided into training and testing part [32] [33]. The training data is used to train the SVM by marking the particular category. Further the test data is analyzed to check their category by determining the hyper plane that maximize the distance between the classes.
3) KNN classifies is used to store the data for all the categories and new classes are classified based on distance functions. The test data is classified based on the majority vote of its neighbors [34] [35].
4) Gaussian Mixture models (GMM) are basically used to detect the likeliness for ordinarily scattered data within overall dataset. The presumption of scattered data belongingness to particular class is not required in this case resulting in unsupervised learning [36].
5) Neural network is a system basically consisting of strongly interrelated elements used for processing the information by their active response to input dataset [37] [38]. The classifiers indicated above are used by various authors across the world to detect the emotion automatically.
6) NB technique is used as classification algorithm. The class labels are assigned to problem instances and described by using feature values in vector form [39]. Class labels can be chosen by any one of the method described above.

2. RELATED WORK

Automatic emotion recognition has been growing interest among researchers since last decade. As the issues related to annotation process are of great concern in this proposed work, so the review related to annotation process is presented in this section. In 2003 Feng et al. analyzes the concept of detecting the mood by retrieving information from music [40]. Authors classified the mood as happiness, sadness, anger and fear respectively. Authors use randomly collected dataset from CD's and online sources and divided them in training and test data. Authors divided training data in various categories on the basis of probability distribution of mean of average silence ratio and standard deviation of average silence ratio. Muyuan et al. makes

use of adaptive scheme to detect the emotion of music in 2004 [41]. Authors make use of online collected database in this approach and annotation process is carried out by 20 subjects. Tao Li and Mitsunori Ogihara analyzes the acoustic features for music emotion detection [42]. Authors collected dataset of different genres and the labeling of emotion class was conducted by two subjects. The detection accuracy reached is up to 80% by this approach. Thereafter in 2006 Lu et al. presented hierarchical framework for automatic detection and tracking of mood of songs [43]. In this approach annotation process is carried out with the help of three experts and Thayer's model is used to represent the moods in 2-d plane. Value of recall and precision achieved experimentally in this work are 84% and 81%. Yang et al. focused on the issues related to detection of emotion in music signals and proposed a regression approach for predicting AV values on the basis of which the songs are categorizes in different classes [7]. In this work 195 music clips are considered and human defined subjective test is carried out for annotation. Psysound and Marsyas is used by authors for feature extraction, Support vector regressor is used for classification and RRelief is used for feature reduction. Experimental results of this work show that R2 statistics achieved 58.3% for arousal and 28.1% for valence. In 2011 Saari et al. proposed wrapper selection feature reduction approach for improving classification process [44]. Yang and Chen [7] proposed an approach to represent the emotion as probability distribution in AV plane [7]. Authors deal with issues due to ambiguity in language of people. The annotation process is carried out by subjects but their opinion is marked on a graphical user interface by a circle on AV-plane. The R2 statistics achieved in this work is 54.3%. Brinker et al. [45] conducted a listening experiment for rating various moods belonging to songs [45]. The analysis deals with the issues of the number of basic dimensions used for emotion classification and VA relationship of emotion and variation in labeling the class to particular emotion. In 2013, Wang et al. proposed maximum a posteriori approach to analyze personalized music emotion by using acoustic emotion Gaussian's model [46]. The database used in this work is MER 60 and annotation is subjective and carried out by 40 annotators. In 2014, Markov and Matsui analyses the applicability of Gaussian processes for music emotion recognition [47]. The authors make use of MediaEval 2013 database for research and annotation is done by annotators. Authors proved that Gaussian model outperforms SVM's. In 2015, Ahsan et al. analyses the annotation as multilabel classification [48]. Nearest neighbor and maximum margin techniques were employed for classification. Wang et al. proposed an automatic system for music emotion detection that is based on mixture model using hierarchical Dirichlet process [49]. Hu and Yang proposed regression based models for mood detection using fifteen audio features and classified songs in five mood categories by focusing on cross dataset generalizability [50]. Experimental results of this work prove that the performance of regression system is affected by size of training database and accuracy of annotation.

In 2017, Mo, and Niu proposed a technique for analyzing music signal for detecting the emotions [51]. The technique makes use of orthogonal matching pursuit, Wigner function for distribution and Gabor functions. Annotation process is carried with the help of five subjects. Experiments were carried out by using four dataset and mean accuracy of 69.5% was achieved by this technique. In 2018, Patra et al. [52] makes use of LibSVM and neural networks to develop music emotion detection system based on lyrics of Hindi and western music songs [52]. The annotation process is subjective in this work. Authors achieved F-measure of 0.751 and 0.83 for above two techniques.

In 2018, authors categorizes the Hindi music signals according to four genres: Classical, Folk, Ghazal and Sufi [52]. Music signals belonging to these genres are divided into positive arousal, negative arousal, positive valence and negative valence by considering arousal and valence as parameters. Spectral features are calculated for the music clips using MIR toolbox. The classification is done by using K-Nearest Neighbor (K-NN), Naive Bayes (NB) and Support vector machine (SVM). The classification process is conducted for all the four genres and also for arousal and valence classes. The accuracy, precision and recall are considered as evaluation parameter in this work. The evaluation parameters of all the genres and classification results of all the classifiers used are compared in the proposed work. The accuracy, precision and recall achieved by classical songs by using SVM are 0.91, 0.81 and 0.83. The songs are also categorized on the basis of arousal and valence. The accuracy achieved by positive arousal is more than negative arousal and accuracy achieved by positive valence is more than negative valence. The accuracy achieved by positive arousal is 0.86 and by positive valence is 0.83 by using SVM. Results reveal that SVM classifier outperforms other two classifiers in terms of the parameters considered.

In the proposed work dataset consisting of Hindi songs is considered and five types of features are extracted using MIRtoolbox. In this work music signals belonging to different genres are categorized in four quadrants belonging to Arousal-Valence (AV) plane of emotion categorization i.e. Positive Valence-Positive Arousal (PV-PA), Positive Valence- Negative Arousal (PV-NA), Negative Valence-Positive Arousal (NV-PA) and Negative Valence-Negative Arousal (NV-NA). In the related work mentioned above the annotation process is human defined and subjective. Various issues that arise with SBA as described below.

i. This process of annotation is less trustworthy because in this approach average opinion of a group of people is considered and finding specialized group of people in the field of music and psychology is time consuming and costly.

ii. This method is costly as researchers have to pay the people that take part in annotation process.

iii. The emotion of a person is instinctive feeling, thus it can vary for the same person for the same song in future also.

Thus the proposed ABA approach is proposed to deal with the subjective issues by considering following points.

1) Annotation process is implemented by algorithm based approach ABA [54] and compared with subjective approach SBA.

2) The model is proposed for categorization of emotion classes in the proposed work is 2-d and four basic classes are considered for both the approaches. The database used in the proposed work is Hindi songs

3) The SVM, NB and KNN are used as classifier for both the techniques to validate the results in terms of accuracy, precision and recall.

3. METHODOLOGY AND DISCUSSION OF PROPOSED APPROACH

Methodology adopted for proposed approach consists of five basic steps described in section 1. In this section stepwise methodology of proposed work is represented in Fig.3 and discussed thoroughly in this section.

3.1 DATABASE COLLECTION AND TRAINING PROCESS

The database for this work is Hindi songs of various genres such as pop, classical, patriotic, and jazz and ghazals is collected from freely available online resources 1, 2. The researcher can consider 'n' number of songs from the online music sources. Let is a database with huge collection of music. 70% of collected dataset that is are used as training samples and 30% of the dataset is considered as test dataset. The songs are collected from the online source at the sampling rate of 44100Hz. The whole song does not represent the emotion of the song equally. Thus 30 seconds segment representing most effective emotion of songs is considered in this work. In this work total 1000 songs of different genres are collected from online sources 1, 2 and out of which 700 songs are considered for training purpose and 300 songs are considered for testing purpose. The sample dataset considered in this work is given in Table.1. The whole song does not represent the emotion of the song equally. Thus 30 seconds segment representing most effective emotion of songs is considered in this work. In this work total 1000 songs of different genres are considered out of which 700 songs are considered for training purpose and 300 songs are considered for testing purpose. The sample dataset considered in this work is given in Table.1.

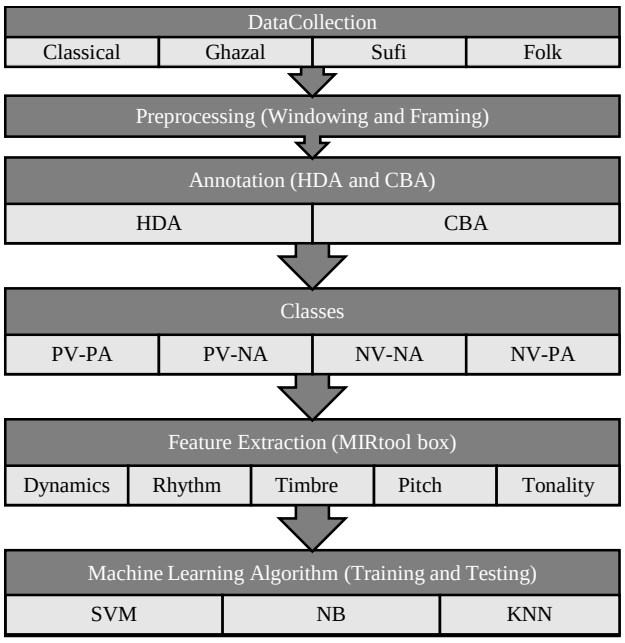

Fig.3. Basic steps of AMER

Pre-processing mainly consists of two steps they are windowing and framing. Windowing is directly in co-operated with the Fourier transform function. Hanning window is used to preprocess the signal. The sound signals are non-stationary, thus the analysis of sound signals is carried out by considering short time signals. The process of transforming the sound signal in short time signals is framing. Frame length of 25ms is considered in the proposed work.

Table.1. Types of dataset

Songs Genre		Example
Ghazal	Positive Arousal	Bahut pahle se un kadmo ki aahat
	Negative Arousal	Pahle jo dard tha wahi pyara hai in dino
	Positive Valence	Apni marji se hawao ka safar
	Negative Valence	Ae khuda reeet k sehaaroon ko samandar kar de
Folk	Positive Arousal	Aai aai basanti mela
	Negative Arousal	Dharti kahe pukar ke
	Positive Valence	Gaon Taraane mann ka
	Negative Valence	Phir tumhari yaad
Sufi	Positive Arousal	Allah Hoo Allah Hoo
	Negative Arousal	Mere Sahiba
	Positive Valence	Tu mane ya na mane
	Negative Valence	Tumhain Dillagi Bhool
Classical	Positive Arousal	Ajori badhavara
	Negative Arousal	Gunaun gaun tumharo
	Positive Valence	Bajan laagi bansri kaanhaki
	Negative Valence	Kate na Biraha ki raat

3.2 ANNOTATION

Annotation is the process of categorizing the songs in different classes. In this paper the music clips are categorized in four main categories Arousal-Valence (AV) plane of emotion categorization i.e. Positive Valence-Positive Arousal (PV-PA), Positive Valence-Negative Arousal (PV-NA), Negative Valence-Positive Arousal (NV-PA) and Negative Valence-Negative Arousal (NV-NA) as described in section 1.3.2 shown in Fig.4.

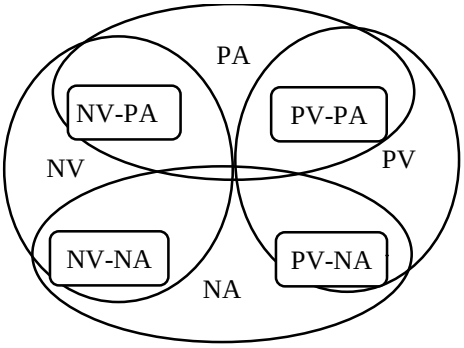

Fig.4. AV emotion plane

As described in section 1.3.2 the categories described by Russel's model are described in four main classes PV-PA, NV-PA, PV-NA and NV-NA. In this proposed work, annotation is done two using subjective and algorithm based annotation techniques termed as SBA and ABA are described below.

3.2.1 SBA:

SBA is subjective and is done before feature extraction. In this method the songs are distributed to four groups of eight people each. These groups are asked to categorize the songs in four categories i.e. PV-PA, NV-PA, PV-NA and NV-NA. The average of the decision of all the groups is considered as final class of the songs. The Table.2 shows the description of groups whose opinion is used to categorize the training data.

3.2.2 ABA:

This annotation is carried out after the feature extraction process has been completed. Annotation process algorithm based and it is incorporated with the correlation technique. In this approach the linear dependency between the various features of signals is measured using the correlation coefficient [54]. Annotation process is considered as the multiclass classification problem where each song in the database Annotation process is considered as the multiclass classification problem where each song in the database $s_i \in \upsilon$ represents a certain emotion represents a certain emotion. The correlation between any two signals or vectors is given as represented in Eq.(2).

$$x_i = s_i^{(k)} v^{(k)} \quad (1)$$

where, $x > 0$ signifies that the two signals are associated with the same half-space, $x = \pm 1$ indicates that both signals may be in opposite direction or sometimes it can be in same direction at times it can also in the form of orthogonal. The correlation value for all the feature vectors is extracted. The correlation values obtained for distinct class lies in different ranges. The average correlation value for different songs is taken. The average correlation values of songs belonging to different classes are used to train the SVM.

Table.2. Group's description

Group	Age	Occupation
Group-I	14-18 years	School Students
Group-II	18-24 years	College Students
Group-III	25-35 years	Service
Group-IV	50-60 years	Middle aged persons at home

The training process by using ABA is shown in Fig.5(a) and by using SBA is shown in Fig.5(b).

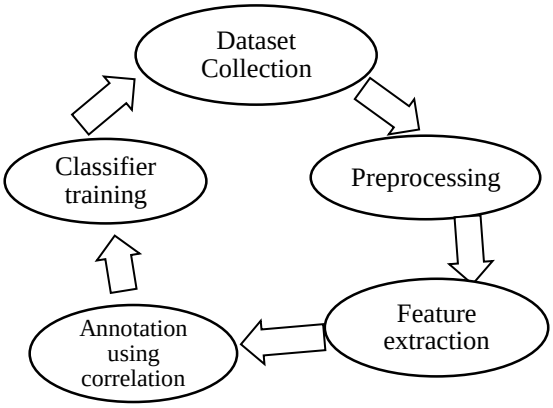

Fig.5(a). Block diagram representing ABA

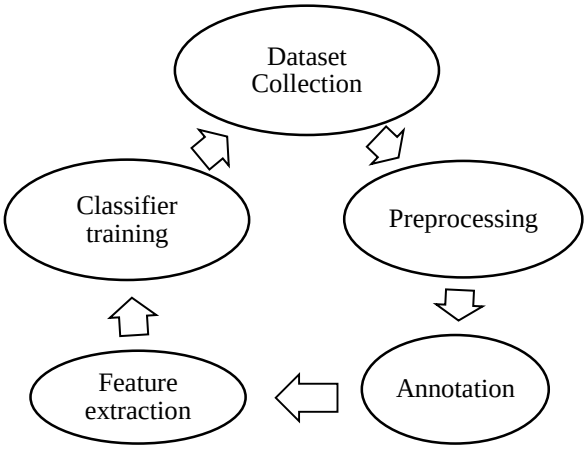

Fig.5(b). Block diagram representing SBA

The ABA is conducted by calculating the correlation values of feature vectors calculated for songs. The correlation values are unique for all the categories of songs. With the help of those correlation values of feature vectors it is possible to categorize the signals in four groups i.e. PV-PA, NV-PA, PV-NA and NV-NA. The range of correlation values of all the categories considered in this approach is shown in the Table.3.

Table.3. Correlation range for different categories of songs

Valence	Arousal	Correlation range
Positive	Positive	$0.86 \leq corr \leq 0.90$
Positive	Negative	$0.80 \leq corr \leq 0.85$
Negative	Negative	$0.91 \leq corr \leq 0.95$
Negative	Positive	$0.96 \leq corr \leq 1$

3.3 FEATURE EXTRACTION

Feature extraction is mainly used to extract various features that are associated with the audio signal. In this research the features are evaluated by using MIR toolbox [28]. Basically five types of features dynamics, rhythm, timbre, pitch and tonality can be evaluated using MIR toolbox as mention in Table.4.

Table.4. Definition of types of features

Type	Definition
Dynamics	Dynamics of a song clip is the alteration in loudness of notes
Rhythm	Rhythm represents the pattern of notes of different strength in a song.
Timbre	Timbre defines the quality of notes of a song.
Pitch	Pitch represents the perceived frequency of song clips.
Tonality	Tonality represents the arrangement of pitches in a proper order

In the proposed work the features are extracted by using MIRtoolbox 1.6.1 in Matlab. Total 25 features are extracted by using this toolbox and mean of the values of features of songs of all the categories is shown in Table.5. All the features calculated above are not equally important, thus feature redundancy is used to choose most effective features.

Relieff feature reduction technique [35] is used in the proposed work. It identifies the difference between feature values of *k*-Nearest Neighbors. A rank is specified to each feature indicating its importance by this technique. In this work $k = 10$ is considered for reduction technique.

The top five ranked features have been considered and are mentioned with weights and rank in Table.6. Based on these five features training process of data is carried out.

In ABA the correlation values of all the categories is of major importance. The songs belonging to different categories attain different set of correlation values.

Table.5. Mean of features

Features	PV-PA	NV-PA	PV-NA	NV-NA
RMS	0.94905	0.9396	0.9222	0.9339
Peak (pos)	4.28261	5.6222	3.9387	4.1777
Tempo	128.660	129.64	129.36	128.57
Attack time	0.04772	0.0455	0.0467	0.0470
Attack slope	4.31941	5.8283	4.4020	4.3568
Centroid	2559.17	3179.8	2672.7	2096.9
Brightness	0.4758	0.5551	0.4486	0.3951
Spread	2781.14	3178.3	3106.1	2583.9
Skewness	1.99787	1.7498	2.2244	2.8217
Kurtosis	14.0295	8.1954	10.424	15.917
Rolloff 95	8067.64	9585.8	9118.2	7257.9
Rolloff 85	5168.72	6425.5	5399.1	4069.0
Spectral entropy	0.75632	0.7881	0.7434	0.7139
Flatness	0.08671	0.1061	0.1019	0.0731
MFCC	0.22047	0.1753	0.2903	0.2735
Roughness	0.98478	2.0828	1.7829	0.6493
Irregularity	0.85192	0.8337	0.8003	0.8855
Zero Crossing Rate	1340.63	1814.4	1280.1	1078.7
Low energy	0.48777	0.4760	0.4708	0.4824
Spectral flux	734.934	795.43	647.85	652.94
Chromagram (PPM)	58.6201	60.370	58.453	60.095
Chromagram (CM)	63.0098	64.963	62.563	63.544
Key clarity	0.49267	0.5311	0.5134	0.5336
Mode	0.01851	-0.0303	-0.0281	-0.0180
HCDF	0.18556	0.1860	0.1689	0.1714

Table.6. Top 5 features with weights

Feature	Rank	Weight
MFCC	1	0. 17
HCDF	2	0. 11
ZCR	3	0. 07
Irregularity	4	0. 01
Spectral roll off	5	0. 007

The training songs were annotated in four categories by SBA and ABA. Number of songs that are predicted in particular category using SBA and ABA is shown in Table.7.

Table.7. Number of songs in different categories

Category of songs		Training data annotation	
Valence	Arousal	SBA	ABA
Positive	Positive	130	125
Negative	Positive	110	118
Positive	Negative	150	140
Negative	Negative	110	117

The distribution of data by SBA and ABA are shown graphically in Fig.6. The Fig.7(a), Fig.7(b), Fig.7(c) and Fig.7(d) represents songs in all the four categories based on correlation values of feature vectors.

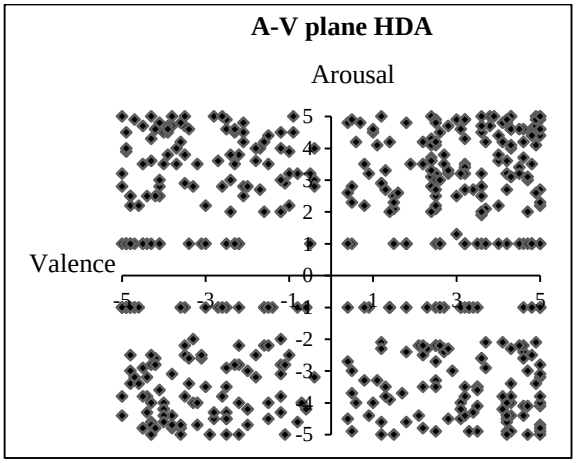

Fig.6. AV - plane SBA

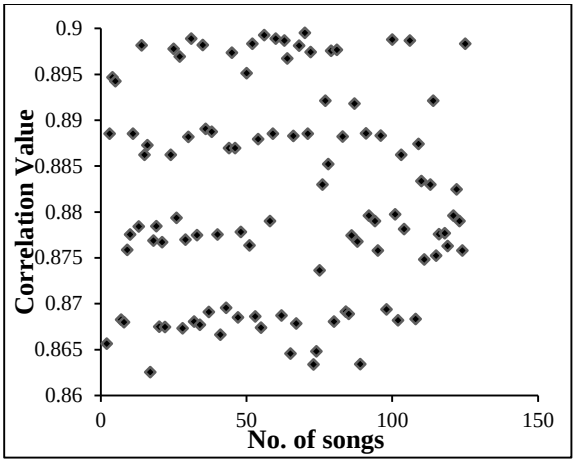

Fig.7(a). Correlation values of PV-PA category

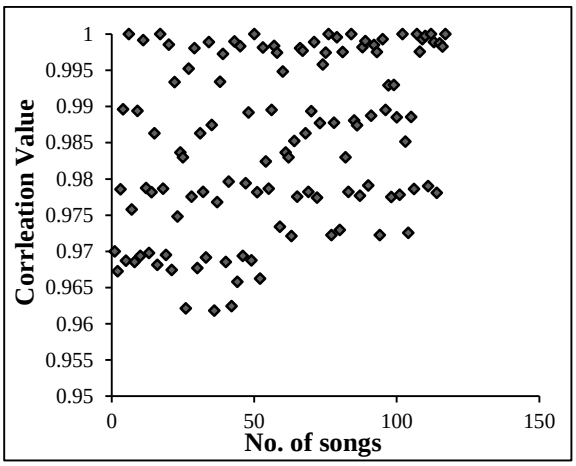

Fig.7(b). Correlation values of NV-PA category

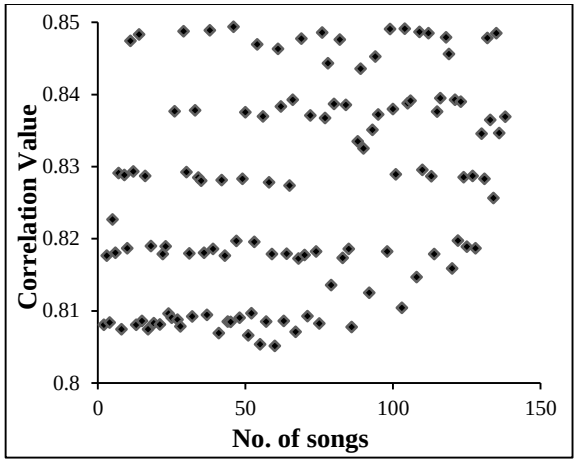

Fig.7(c). Correlation values of PV-NA category

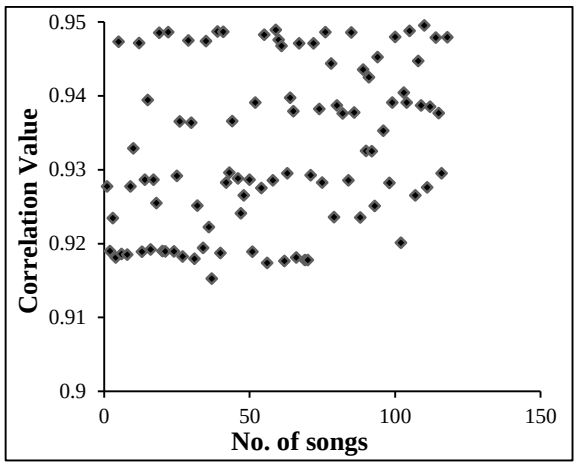

Fig.7(d). Correlation values of NV-NA category

3.4 CLASSIFICATION

Consider $Y = [Y_1]\ [Y_2]\ [Y_3]\ \ldots\ [Y_K]$ are various types of music available in the database X which is used for testing purpose. Y is exactly drawn from a union of 'K' independent linear subspace of unknown dimensions. For this music samples the preprocessed according to the Eq.(1). For both the types of annotation the classifiers used in the proposed work are SVM, NB and KNN. The results of all the three classifiers using both the annotation methods are compared. In ABA average correlation value extracted between feature vectors of test data is equated with the train data and result is computed on its basis. The testing process and dataset considered for both the approaches are similar for efficient analysis.

4. PERFORMANCE EVALUATION

The implementation is done in the MATLAB platform. The songs are collected from the online source at the sampling rate of 44100 Hz. For fair comparison of both the annotation methods classifiers are trained for same dataset. In SBA the classifiers are trained on the basis of feature values of dataset that is annotated by a group of subjects and in ABA the classifiers are trained on the basis of dataset annotated on the basis of correlation values attained for feature vectors of the songs. Test data of 300 songs is considered for both the approaches. In this to analyze the implementation of ABA, the evaluation parameters accuracy, precision and recall are computed and the obtained results are compared with SBA. The performance of SBA and ABA approaches using SVM, K-NN and NB has been shown in Table.8 and Table.9. The comparison of evaluation parameters is represented in Fig.10 and Fig.11 by using SBA and proposed technique ABA for Hindi songs using SVM, K-NN and NB.

The classification procedure for classifiers makes use of true positives (*TP*), false positives (*FP*), true negatives (*TN*) and false negatives (*FN*) for determining accuracy, precision and recall. The positive and negative terms deals with the predicted results of the classifier, whereas true and false deals with the trusted prediction provided by the subjects during annotation for SBA and correlation value range for ABA. Confusion matrix shows the

relationship between actual classification and the classifications that are projected by the classifier [55]. The confusion matrix shows the relationship between TP's and TN's. The various terms related to confusion matrix are explained below. The diagonal matrix of multiple class confusion matrices shows the TP's. The TN's for a particular class is the sum of all the elements of a matrix by factoring out the column and row of a particular class. False FN's are calculated as the sum of elements of the row of particular class factoring out the TP of that class. FP's is calculated as the sum of the elements of column of particular class by factoring out the TP of that class. Accuracy, precision and recall used in the proposed work are described below in terms of TP's, TN's, FP's and FN's.

4.1 ACCURACY

Accuracy is described as correctly projected outcomes out of total test class [55]. Accuracy is an important factor for any work. The greater accuracy percentage is important for any implementation. The accuracy is computed using the formulas given below,

$$Accuracy = (TP+TN) / (TP+FP+TN+FN) \quad (2)$$

The comparison between the existing and proposed work in terms of accuracy is shown in Fig.8 and Fig.9. The proposed implementation is highly efficient in terms of accuracy. The accuracy obtained by using SBA is 86.3%, 83% and 81% for SVM, NB and KNN, whereas the accuracy obtained by using ABA is 90.2%, 85% and 83% for SVM, NB and KNN as represented in Table.8 and Table.9. Thus the accuracy of SVM is better than NB and KNN and ABA outperforms in terms of accuracy for all the classifiers.

4.2 PRECISION

Precision is the fraction of relevant projected data out of projected data by the classifier [55].

$$Precision = TP / (TP+FP) \quad (3)$$

The precision value obtained by adopting SBA is 0.72, 0.63 and 0.6 for SVM, NB and KNN, whereas while using ABA the precision values obtained are 0.79, 0.68 and 0.63 for SVM, NB and KNN as represented in Table.8 and Table.9. Therefore it has been evaluated that the precision results of SVM are also better than NB and KNN and performance is better for ABA.

4.3 RECALL

Recall is the fraction of relevant projected data out of total relevant samples belonging to particular category present in the database [55].

$$Recall = TP/(TP + FN) \quad (4)$$

The recall value obtained by adopting SBA is 0.83 for SVM and 0.78 NB and KNN, whereas while using ABA the recall values obtained are 0.88 for SVM and 0.78 for NB and KNN as represented in Table.8 and Table.9. Therefore, it has been evaluated that the recall results of SVM are also better than NB and KNN and performance is better for ABA.

Table.8. Comparison of evaluation parameters for SBA

Dataset	Classifiers	Accuracy	Precision	Recall
Hindi Songs	SVM	86.3%	0.72	0.83
	NB	83%	0.63	0.78
	KNN	81%	0.60	0.78

Table.9. Comparison of evaluation parameters for ABA

Dataset	Classifiers	Accuracy	Precision	Recall
Hindi Songs	SVM	90.2%	0.79	0.88
	NB	85%	0.68	0.76
	KNN	83%	0.63	0.76

The graphical representation of above mentioned evaluation parameters is shown in Fig.8.

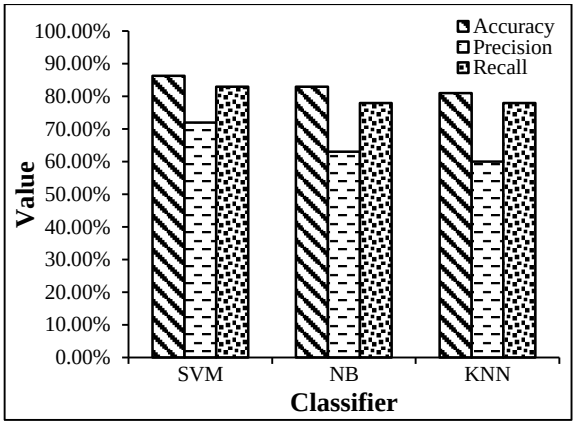

Fig 8. Graphical representation of evaluation parameters for SBA

The results discussed above reveals that the performance of SVM for ABA is also better than other two and it is also proved that the results of ABA for all the three classifiers is better than SBA.

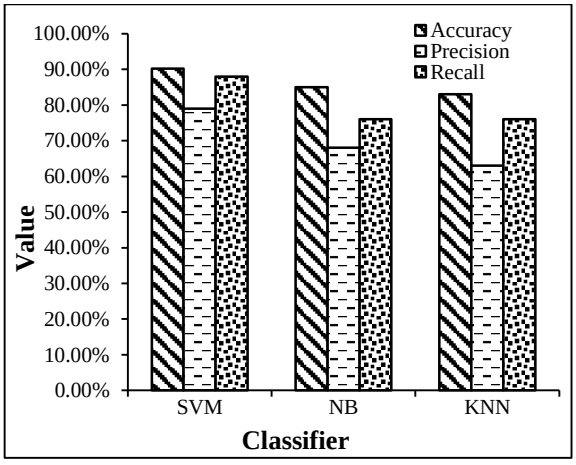

Fig.9. Graphical representation of evaluation parameters for ABA

5. ADVANTAGES AND LIMITATIONS OF PROPOSED WORK

The main advantage of this proposed work is to eliminate the disadvantages of subjective annotation. The ABA is based on correlation of features and is helpful in overcoming the granularity issue of SBA. The correlation annotation is limited to divide the database in less number of categories in comparison to SBA. The proposed work performs better for less number of categories but for large number of categories correlation values may overlap.

6. CONCLUSION

In this work the two methods of annotation are compared by using SVM, NB and KNN. In SBA approach average review of person for the emotion of song is considered for training process and mean of values of features is used to train SVM. In ABA features for all the songs is first calculated and the correlation values of feature vectors is used for annotation process and is further used to train the data. The training as well as testing of both the methods is carried out by using same data for fair comparison. SVM, NB and KNN are used in both the cases for classification and the data set is divided in four main categories PV-PA, NV-PA, PV-NA and NV-NA. The results show that ABA performs better than SBA in terms of accuracy precision and recall. The accuracy achieved by SBA approach is 86.3% for SVM, 83% for NB and 81% for KNN. The precision achieved by SBA is 0.72 for SVM, 0.63 for NB and 0.6 for KNN. The recall value obtained for SBA is 0.83 for SVM, 0.78 for NB and KNN. In comparison to SBA the accuracy obtained by implementing ABA is 90.2% for SVM, 85% for NB and 83% for KNN. The precision values obtained by using ABA are 0.79 for SVM, 0.68 for NB and 0.63 for KNN. The recall values obtained for ABA by using SVM are 0.88 and 0.76 for NB and KNN. Thus it is concluded that the SVM performs better than NB and KNN for both the cases and ABA is more efficient than SBA as ABA is algorithm based and achieves better performance results.

REFERENCES

[1] S. Shandilya and P. Rao, "Detection of the Singing Voice in Musical Audio", *Proceedings of 114th Convention of Audio Engineering Society*, pp. 1-6, 2003.

[2] X. Hu, "A Framework for Evaluating Multimodal Music Mood Classification", *Journal of the Association for Information Science and Technology*, Vol. 68, No. 2, pp. 273-285, 2017.

[3] Y.H. Yang, Y.C. Lin, Y.F. Su and H.H. Chen, "A Regression Approach to Music Emotion Recognition", *IEEE Transactions on Audio Speech and Language Processing*, Vol. 16, No. 2, pp. 448-457, 2008.

[4] X. Hu and Y.H. Yang, "The Mood of Chinese Pop Music: Representation and Recognition Xiao", *Journal of the Association for Information Science and Technology*, Vol. 68, No. 8, pp. 1899-1910, 2017.

[5] V. Hampiholi, "A Method for Music Classification based on Perceived Mood Detection for Indian Bollywood Music", *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, Vol. 6, No. 12, pp. 507-514, 2012.

[6] J. Lee and J. Nam, "Multi-Level and Multi-Scale Feature Aggregation Using Pretrained Convolutional Neural Networks for Music Auto-Tagging", *IEEE Signal Processing Letters*, Vol. 24, No. 8, pp. 1208-1212, 2017.

[7] Y.H. Yang, Y.F. Su, Y.C. Lin and H.H. Chen, "*Music Emotion Recognition*", CRC Press, 2011.

[8] Y.H. Yang and H.H. Chen, "Ranking-based Emotion Recognition for Music Organization and Retrieval", *IEEE Transactions on Audio Speech and Language Processing*, Vol. 19, No. 4, pp. 762-774, 2011.

[9] A. Shakya, B. Gurung, M.S. Thapa and M. Rai, "Music Classification based on Genre and Mood", *Proceedings of International Conference on Computational Intelligence, Communications and Bussiness Analytics*, pp. 168-183, 2017.

[10] X. Hu and J.S. Downie, "Exploring Mood Metadata: Relationships with Genre, Artist and usage Metadata", *Proceedings of International Conference on Music Information Retrieval*, pp. 23-27, 2007.

[11] X. Hu, J. S. Downie, C. Laurier, M. Bay and A.F. Ehmann, "The 2007 MIREX Audio Mood Classification Task: Lessons Learned", *Proceedings of International Conference on Music Information Retrieval*, pp. 462-467, 2008.

[12] S.O. Karl, F. Macdorman and C.C. Ho, "Automatic Emotion Prediction of Song Excerpts: Index Construction, Algorithm Design, and Empirical Comparison", *Journal of New Music Research*, Vol. 36, No. 4, pp. 281-299, 2007.

[13] T. Li and M. Ogihara, "Detecting Emotion in Music", *Proceedings of International Conference on Music Information Retrieval*, pp. 239-240, 2003.

[14] D. Huron, "Perceptual and Cognitive Applications in Music Information Retrieval", *Proceedings of International Conference on Music Information Retrieval*, pp. 23-25, 2000.

[15] D. Huron, "*Sweet Anticipation: Music and the Psychology of Expectation*", MIT Press, 2006.

[16] A. Gabrielsson," Emotion Perceived and Emotion Felt: Same or Different", *Musicae Scientiae*, Vol. 5, No. 1, pp. 123-147, 2002.

[17] S. Hallam, I. Cross and M. Thaut, "*The Oxford Handbook of Music Psychology*", Oxford University Press, 2008.

[18] K. Hevner, "Expression in Music: A Discussion of Experimental Studies and Theories", *Psychological Review*, Vol. 48, No. 2, pp. 186-204, 1935.

[19] K. Hevner, "Experimental Studies of the Elements of Expression in Music", *American Journal of Psychology*, Vol. 48, No. 2, pp. 246-268, 1936

[20] P. Ekman, "An Argument for Basic Emotions", *Cognition and Emotion*, Vol. 6, No. 3, pp. 169-200, 1992.

[21] R. Plutchik, "*Emotion A Psychoevolutionary Synthesis*", Harper and Row, 1980.

[22] A. Mehrabian and J.A. Russell, "*An Approach to Environmental Psychology*", MIT Press, 1974.

[23] J.A. Russell, "A Circumplex Model of Affect", *Journal of Personality and Social Psychology*, Vol. 39, No. 6, pp. 1161-1178, 1980.

[24] J.A. Russell, "Core Affect and the Psychological Construction of Emotion", *Psychological Review*, Vol. 110, No. 1, pp. 145-172, 2003.

[25] R.E. Thayer, "*The Biopsychology of Mood and Arousal*", Oxford University Press, 1989.

[26] E. Bigand, S. Vieillard, F. Madurell, J. Marozeau and A. Dacquet, "Multidimensional Scaling of Emotional Responses to Music: The Effect of Musical Expertise and of the Duration of the Excerpts", *Cognition and Emotion*, Vol. 19, No. 8, pp. 1113-1139, 2005.

[27] J. Fontaine, K. Scherer, E. Roesch and P. Ellsworth, "The World of Emotions is not Two Dimensional", *Psychological Science*, Vol. 18, No. 12, pp. 1050-1057, 2007.

[28] O. Lartillot and P. Toiviainen, "MIR in MATLAB (II): A Toolbox for Musical Feature Extraction from Audio", *Proceedings of International Conference on Music Information Retrieval*, pp. 127-130, 2007.

[29] D. Cabrera, "Psysound: A Computer Program for Psycho-Acoustical Analysis", *Proceedings of International Conference on Australian Acoustic Society*, pp. 47-54, 2007.

[30] G. Tzanetakis, "MARSYAS submissions to MIREX 2007", *Proceedings of International Conference on Music Information Retrieval Evaluation eXchange*, pp. 1-6, 2007.

[31] A. Sen and M. Srivastava, "*Regression Analysis: Theory, Methods, and Applications*", Springer, 1990.

[32] F. Sebastiani, "Machine Learning in Automated Text Categorization", *ACM Computing Surveys*, Vol. 34, No. 1, pp. 1-47, 2002.

[33] C. Laurier and P. Herrera, "Audio Music Mood Classification using Support Vector Machine", *Proceedings of International Conference on Music Information Retrieval Evaluation eXchange*, pp. 501-507, 2007.

[34] M.L. Zhang and Z.H. Zhou, "ML-KNN: A Lazy Learning Approach to Multi-Label Learning", *Pattern Recognition*, Vol. 40, No. 7, pp. 2038-2048, 2007.

[35] K. Trohidis, G. Tsoumakas, G. Kalliris and I. Vlahavas, "Multi-Label Classification of Music into Emotions", *Proceedings of International Conference on Music Information Retrieval*, pp. 325-330, 2008.

[36] G. Peeters, "A Generic Training and Classification System for MIREX08 Classification Tasks: Audio Music Mood, Audio Genre, Audio Artist and Audio Tag", *Proceedings of International Symposium on Music Information Retrieval*, pp. 1-6, 2008.

[37] Q. Pu and G.W. Yang, "Short-Text Classification based on ICA and LSA", *Proceedings of International Conference on Advances in Neural Networks*, pp. 265-270, 2006.

[38] K. Mannepalli, P.N. Sastry and M. Suman, "Analysis of Emotion Recognition System for Telugu using Prosodic and Formant Features", *Proceedings of International Conference on Speech Language Processing and Human-Machine Communications*, pp. 137-144, 2018.

[39] N. Sebe, M.S. Lew, I. Cohen, A. Garg and T.S. Huang, "Emotion Recognition using a Cauchy Naive Bayes Classifier", *Proceedings of International Conference on Pattern Recognition*, pp. 1-15, 2002.

[40] Y. Feng, Y. Zhuang and Y. Pan, "Popular Music Retrieval by Detecting Mood", *Proceedings of 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 375-376, 2009.

[41] W. Muyuan, Z. Naiyao, Z. Hancheng, M. Wang, N. Zhang and H. Zhu, "User-Adaptive Music Emotion Recognition", *Proceedings of 7th International Conference on Signal Processing*, pp. 1352-1355, 2004.

[42] M. Ogihara, "Content-Based Music Similarity Search and Emotion Detection", *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing*, pp. 505-508, 2004..

[43] L. Lu, D. Liu and H.J. Zhang, "Automatic Mood Detection and Tracking of Music Audio Signals", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 1, pp. 5-18, 2006.

[44] P. Saari, T. Eerola and O. Lartillot, "Generalizability and Simplicity as Criteria in Feature Selection: Application to Mood Classification in Music", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 19, No. 6, pp. 1802-1812, 2011.

[45] B. Den Brinker, R. Van Dinther and J. Skowronek, "Expressed Music Mood Classification Compared with Valence and Arousal Ratings", *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 2012, No. 1, pp. 1-14, 2012.

[46] J. Wang, Y. Yang, H. Wang and S. Jeng, "Personalized Music Emotion Recognition Via Model Adaptation", *Proceedings of Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 3-6, 2012.

[47] K. Markov and T. Matsui, "Music Genre and Emotion Recognition Using Gaussian Processes", *IEEE Access*, Vol. 2, pp. 688-697, 2014.

[48] H. Ahsan, V. Kumar and C.V. Jawahar, "Multi-Label Annotation of Music", *Proceedings of 8th International Conference on Advances in Pattern Recognition*, pp. 4-7, 2015.

[49] J.C. Wang, Y.S. Lee, Y.H. Chin, Y.R. Chen and W.C. Hsieh, "Hierarchical Dirichlet Process Mixture Model for Music Emotion Recognition", *IEEE Transactions on Affective Computing*, Vol. 6, No. 3, pp. 261-271, 2015.

[50] X. Hu and Y.H. Yang, "Cross-Dataset and Cross-Cultural Music Mood Prediction: A Case on Western and Chinese Pop Songs", *IEEE Transactions on Affective Computing*, Vol. 8, No. 2, pp. 228-240, 2017.

[51] S. Mo and J. Niu, "A Novel Method based on OMPGW Method for Feature Extraction in Automatic Music Mood Classification", *IEEE Transactions on Affective Computing*, pp. 1, 2017.

[52] B.G. Patra, D. Das and S. Bandyopadhyay, "Multimodal mood classification of Hindi and Western songs", *Journal of Intelligent Information Systems*, Vol. 51, No. 8, pp. 1-18, 2018.

[53] D. Chaudhary, N.P. Singh and S. Singh, "Genre based Classification of Hindi Songs", *Proceedings of* 9th *International Conf. on Innovations in Bio-Inspired Computations and Applications*, pp. 15-19, 2019.

[54] P.K. Jao and Y.H. Yang, "Music Annotation and Retrieval using Unlabeled Exemplars: Correlation and Sparse Codes", *IEEE Signal Processing Letters*, Vol. 22, No. 10, pp. 1771-1775, 2015.

[55] T. Fawcett, "An Introduction to ROC Analysis", *Pattern Recognition Letters*, Vol. 27, No. 8, pp. 861-874, 2006.