

Music Classification Based on Genre and Mood

Ayush Shakya, Bijay Gurung, Mahendra Singh Thapa, Mehang Rai,
and Basanta Joshi^(✉)

Institute of Engineering, Pulchowk, Lalitpur, Nepal
qwerty.ayush5@gmail.com, bjgurung10@gmail.com,
mahendrasinghthapa27@gmail.com, mehanga.raii007@gmail.com,
basanta@ioe.edu.np

Abstract. The advent of internet and the growing number of digital media has increased the necessity of Music Information Retrieval systems within which Music Classification is a prominent task. Here, we present methods to perform genre based classification over five different genre and mood based classification using a mood model that maps mood onto a two-dimensional space along axes of arousal and valence. Support vector machine and a feed-forward artificial neural network are applied to achieve an overall accuracy of 88% for genre based classification and 73% and 67% accuracy for the arousal and valence axes respectively in mood based classification.

Keywords: Music classification · Genre · Mood · Artificial neural network · Support vector machine

1 Introduction

With the advent of the internet, the number of songs being created as well as the number available to the average person has grown a lot. Simply put, it's overwhelming. Sifting through the deluge of songs manually isn't practical or appealing. It needs to be automated.

Automatic classification of music is a growing field with the primary goal of making it easier for people to find songs they like and for vendors to present those songs to their listeners. It can also lay the foundation for figuring out ways to represent similarity between two musical pieces and in the making of a good recommendation system.

Given the perplexing nature of music, its classification requires specialized representations, abstraction and processing techniques for effective analysis, evaluation and classification that are fundamentally different from those used for other mediums and tasks.

2 Literature Review

2.1 History of MIR and Music Classification

The field of Music Information Retrieval (MIR) can be traced back to the 60s with reference to the works done by Kassler in [1]. Even Automatic Transcrip-

tion of Music was attempted as early as the 70s [2]. However, there were two limiting factors that prevented progress in the field at the time. Firstly, the high computational requirements of the problem domain was simply not available. And secondly, other related fields of study such as Digital Signal Processing, Speech Processing, and Machine Learning were also not advanced enough. So, the field stalled for the next few decades.

In the 1990s, the field regained prominence as computational resources improved greatly and the rise of the internet resulted in massive online music collection. So, there was both an opportunity and demand for MIR systems. The organization of the first International Symposium on Music Information Retrieval (ISMIR 1) in 2000 highlights this resurgence of interest in the field. 280 people from 25 different countries participated in ISMIR Conference Malaga 2015.

As for the methodologies used, MIR in the 90s was influenced by the field of Text Information Retrieval (IR), techniques for searching and retrieving text documents based on user queries. So, most of the algorithms were developed based on symbolic representations such as MIDI files [3]. One such method is described in [4].

However, recognizing approximate units of meaning in MIR, like it is done in a lot of text-IR methods was difficult [5].

Instead, statistical non-transcriptive approaches for non-speech audio signals started being adopted in the second half of the 90s [3]. This was probably influenced by progress of such methods in other fields of speech processing. For example, in [6], the authors reported 98% accuracy in distinguishing music from speech in commercial radio broadcasts. This was based on the statistics of the energy contour and the zero-crossing rate.

In [7], the authors introduced similar statistical methods for retrieval and classification of isolated sounds. Similarly, in [8], an algorithm for music-speech classification based on spectral feature was introduced. It was trained using supervised learning.

And so, starting in the 2000s, instead of methods attempting note-level transcriptions, researchers focused on direct extraction of information of audio signals using Signal Processing and Machine Learning techniques.

Currently, three basic strategies are being applied in MIR: [9]

- Based on Conceptual Metadata - Suited for low-specificity queries.
- Using High-level Descriptions - Suited for mid-specificity queries.
- Using Low-level Signal-based Properties - Used for all specificity.

But still most of the MIR techniques being employed at present use low-level signal features instead of high-level descriptors [10]. Thus, there exists a semantic gap between human perception of music and how MIR systems work.

2.2 Audio Processing

Particularly speaking, music signal processing may appear to be the junior relation of the large and mature field of speech signal processing, not least because

many techniques and representations originally developed for speech have been applied to music, often with good results. However, music signals have certain characteristics that are different from spoken language and other signals [11].

2.3 Genre Based Classification

In [12], Scaringella et al. discuss how and why musical genres are a poorly defined concept making the task of automatic classification non-trivial. Still, although the boundaries between genre are fuzzy and there are no well-defined definitions, it is still one of the widely used method of classification of music. If we look at human capability in genre classification, Perrot et al. [13] found that people classified songs—in a ten-way classification setup—with an accuracy of 70% after listening to 3 s excerpts.

The features used for genre based classification have been heavily influenced by the related field of speech recognition. For instance, Mel-frequency Cepstral Coefficients (MFCC), a set of perceptually motivated features that is widely used in music classification, was first used in speech recognition.

The seminal paper on musical genre classification by Tzanetakis et al. [14] presented three feature sets for representing timbral texture, rhythmic content and pitch content. With the proposed feature set, they achieved a classification accuracy of 61% for ten musical genre.

Timbral features are usually calculated for every short-time frame of sound based on the Short Time Fourier Transform (STFT). So, these are low-level features. Typical examples are Spectral Centroid, Spectral Rolloff, Spectral Flux, Energy, Zero Crossings, and the afore-mentioned Mel-Frequency Cepstral Coefficients (MFCCs). Among these, MFCC is the most widely preferred feature [15,16]. Logan [17] investigated the applicability of MFCCs to music modeling and found it to be “at least not harmful”.

Rhythmic features capture the recurring pattern of tension and release in music while pitch is the perceived fundamental frequency of the sound. These are usually termed as mid-level features.

Apart from these, many non-standard features have been proposed in the literature.

Li et al. [18] proposed a new set of features based on Daubechies Wavelet Coefficient Histograms (DWCH), and also presented a comparative study with the features included in the MARSYAS framework. They showed that it significantly increased the classifier accuracy.

Anglade, Amélie, et al. [19] propose the use of Harmony as a high-level descriptor of music.

Music classification has been attempted through a variety of methods. Some of the popular ones are SVM, K Nearest Neighbours and variants of Neural Networks. The results are also widely different. In [20], 61% accuracy has been achieved using a Multilayer Perceptron based approach. While in [21], the authors have achieved 71% accuracy through the use of an additional rejection and verification stage. Haggblade et al. [22], compared simpler and more

naive approaches (k-NN and k-Means) with more sophisticated neural networks and SVMs. They found that the latter gave better results.

However, lots of unique methods – either completely novel or a variation of a standard method – have been put into use too. In [23], the authors propose a method that uses Chord labeling (ones and zeros) in conjunction with a k-window subsequence matching algorithm used to find subsequence in music sequence and a Decision tree for the actual genre classification.

It is also noted that high-level and contextual concepts can be as important as low-level content descriptors [19].

2.4 Mood Based Classification

As mood is a very human thing, Mood Based Classification, also known as Mood Emotion Recognition (MER), requires knowledge of both technical aspects as well as the human emotional system.

Generally, emotions are conceptualized in two ways:

Categorical Conceptualization. This approach to MER categorizes emotions into distinct classes. It requires a set of base emotions (happiness, anger, sadness, etc.) from which other secondary emotion classes can be derived [24]. However, this approach runs into the problem that the whole spectrum of human emotions cannot be captured by a small number of classes.

Dimensional Conceptualization. It defines Musical Values as numerical values over a number of emotion dimensions. So, the focus is on distinguishing emotions based on their position on a predefined space. Most of these conceptualizations map to three axes of emotions: valence (pleasantness), arousal (activation) and potency (dominance). By placing emotions on a continuum instead of trying to label them as discrete, this approach can encompass a wide variety of general emotions.

Thayer [25] proposed a similar two-dimensional approach that adopts the theory that mood is entailed from two factors: -Stress (happy/anxious) -Energy (calm/energetic). This divides music mood into four clusters: Contentment, Depression, Exuberance and Anxious/Frantic (Fig. 1).

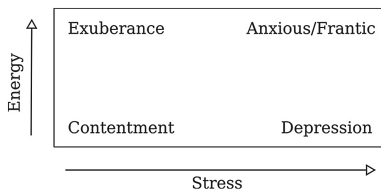


Fig. 1. Thayer’s two-dimensional model of mood

Although, the two-dimensional approach has been criticized as deficient (leading to a proposal of the third dimension of potency), it seems to offer the right balance between sufficient “verbosity” and low complexity [26].

So, we use a similar simplified two-dimensional model based on *arousal* and *valence* (Fig. 2):

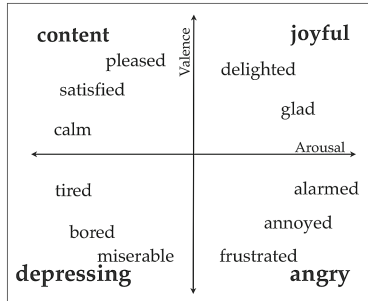


Fig. 2. Two-dimensional model of mood based on Arousal and Valence

2.5 Features in MER

Some of the commonly used features in MER are:

- Energy: Energy related features such as audio power, loudness, specific loudness sensation coefficients (SONE), are correlated to the perception of arousal. Lu et al. [27] used it to classify arousal.
- Melody: These include features such as Pitch (perceived fundamental frequency), chromogram centroid, etc.
- Timbre: As with the AMGC problem, MFCC is widely used in MER too. Apart from MFCC, octave-based spectral contrast as well as DWCH (Daubechies wavelets coefficient histogram) are also proposed in literature.

So, we see that the features used in MER are almost the same as those in AMGC. However, Fu et al. note in their extensive survey on Audio-based Music Classification [28] that although their effectiveness is debatable, mid-level features such as Rhythm seem to be more popular in MER.

The algorithms used in AMGC are also popular in MER. So, support vector machines, Gaussian mixture models, neural networks, and k-nearest neighbor are the ones regularly used.

3 Methodology

3.1 Audio Signal Pre-processing

The pre-processing in music classification systems is used in order to increase the efficiency of subsequent feature extraction and classification stages and therefore

to improve the over-all system performance. Commonly pre-processing includes framing and windowing of the input signal. At the end of pre-processing, the smoothed frame are forwarded to the feature extraction stage.

Framing. Framing is the process of dividing the whole audio sample into frames. Although an audio signal changes continuously, the assumption that on short time scales it remains statistically stationary can be made. So, we frame the signal into 20–40 ms frames. A shorter frame gives too few samples while in a longer one, the signal varies too much.

Windowing. Windowing is necessary because whenever we do a finite Fourier transform, it is implicitly being applied to an infinite repeating signal. So, for instance if the start and end of a finite sample doesn't match then that will look just like a discontinuity in the signal, and show up a lots of high-frequency noise in the Fourier Transform, which is harmful. If the sample happens to be a perfect sinusoid but with an integer number of periods then it doesn't fit exactly into the finite sample and the FT will show appreciable energy in all sorts of places nowhere near the real frequency.

3.2 Feature Integration

As the features are temporal, the feature integration is also temporal. We used the mean and variance of the features for temporal feature integration although they capture neither the temporal dynamics nor dependencies among the individual feature dimensions. As seen below, the mean and standard deviation of MFCCs for a classical and a hiphop songs are sufficiently distinguishable. So, this representation of the features can be used to separate classes of music (Fig. 3).

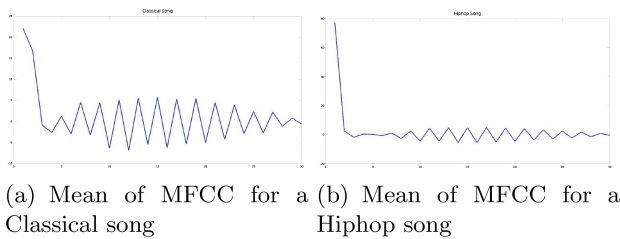


Fig. 3. Comparison of means for Classical and Hiphop songs

3.3 Dataset

The publicly available GTZAN dataset introduced in [14] has become one of the standard datasets for Music Genre Classification used by researchers across the world. We too used this dataset. The dataset contains 100 representative

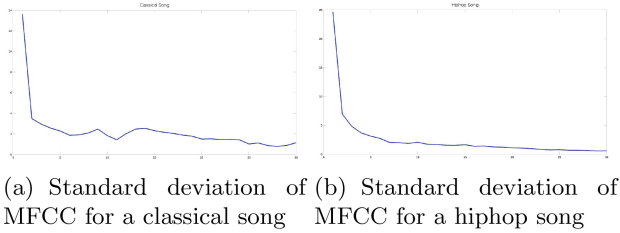


Fig. 4. Comparison of standard deviations for Classical and Hip-hop songs

excerpts from ten different genre. They were taken from radio, compact disk, and MP3 compressed audio files. All the files are stored as 22 050 Hz, 16-bit, mono audio files. The Genres dataset has the following classes: classical, country, disco, hip-hop, jazz, rock, blues, reggae, pop, metal (Fig. 4).

For mood based classification, in 2013, Soleymani et al. [29] created a 1000 songs dataset for emotional analysis of music which uses the Valence-Arousal axes for representing emotional values for songs. The songs, in the dataset, each 45 s long, were collected from FMA. They used Amazon Mechanical Turk as a crowd sourcing platform for collecting more than 20,000 annotations on the 1,000 songs.

Furthermore, their analysis on the annotations revealed a higher agreement in arousal ratings compared to the valence ratings. We have used a filtered version (with some redundancies removed) of that dataset resulting in a final set of 744 songs. We further labeled them as high/low arousal and high/low valence songs based on the numerical values in the dataset. To achieve equal number of songs in each class, we finally used 600 songs of those 744 songs.

3.4 Classifier

For classification process, we used Support Vector Machine and Feed-Forward Artificial Neural Network.

Support Vector Machine. Support vector machines (SVM) are supervised learning models with associated learning algorithms that analyze data used for classification analysis. The popularity of Support Vector machine is huge as a lot of research papers [16, 19, 21, 22] shows its implementation. For the construction of multi-class SVM, we use one vs one SVM classifier. This leads to $\frac{N(N-1)}{2}$ classifiers.

Genre Based Classification: In genre based classification linear kernel is used with the soft margin method.

Mood Based Classification: In mood based classification gaussian kernel and laplacian, the kernel is used which are the non linear type.

Feed-Forward Neural Network. A Feed-Forward Neural Network is a type of Neural Network architecture where the connections are “fed forward”. Research papers [16, 19–22] shows the implementation of artificial neural network in the field of music classification. For training, Backpropagation algorithm is used which calculates the error at a layer and propagates it back to the earlier layers.

Genre Based Classification: In genre based classification we used Cross-entropy error function for output as probabilities and softmax activation function.

Mood Based Classification: In mood based classification we used Least mean squares error function and logistic sigmoid activation function.

4 Results and Discussion

4.1 Effect of Frame Size

As seen in the figure, frame size of 11.5 ms and 23 ms performed considerably better than the bigger frame sizes. We chose the 23 ms (1024 samples) frame size because the smaller 512 sample frame size would lead to higher number of frames and hence necessitate more computation (Fig. 5).

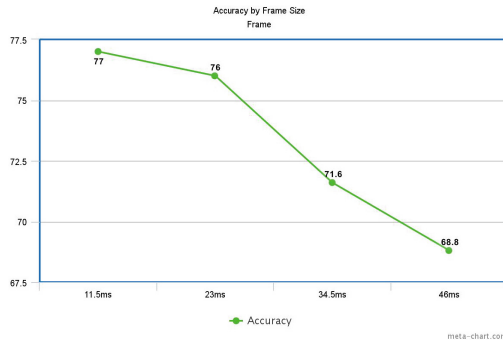


Fig. 5. Effect of frame size

4.2 Effect of Frame Overlap

We explored four different overlapping schemes: 0%, 25%, 50% and 75% overlap. In each of the cases, we received almost the same accuracy (75.4% on No-overlap, 75.8% on Quarter overlap, 76% on half-overlap, and 75.2% on three-quarters overlap). And so, as it seemed to indicate that the overlapping didn’t have any bearing on our results, we chose the less computationally intensive option of using no overlap at all.

4.3 Comparison of Features

Genre Based Classification. MFCC was found to be the best feature for genre classification (in fact, it was found to do well in mood classification too) (Table 1).

Table 1. Genre classification using ANN and SVM

Algorithm	Feature	Classical	Hiphop	Jazz	Pop	Rock	Overall
ANN	Spectral Centroid	47.54	11.92	11.61	51.25	18.89	28.40
	MFCC	92.45	83.42	91.57	82.98	74.44	84.20
	Zero Crossing	63.29	48.31	39.83	52.65	52.00	51.20
	Pitch	37.83	15.00	34.80	61.73	1.67	28.00
	Compactness	81.53	81.75	57.67	28.39	45.55	58.60
	Timbre	6.25	20.00	30.00	20.00	28.46	20.60
	RMS	85.52	34.99	19.85	47.89	70.46	51.20
	Spectral Flux	87.94	26.71	19.77	43.63	57.45	46.40
	Spectral Roll off point	84.10	54.11	22.14	43.74	13.18	41.60
	Spectral Variability	83.10	32.98	25.98	51.24	71.76	52.40
SVM	Spectral Centroid	58.95	1.11	69.09	7.50	0.00	26.40
	MFCC	91.79	85.25	87.98	85.62	77.61	85.80
	Zero Crossing	63.20	48.86	41.47	58.51	44.31	50.00
	Pitch	59.95	38.37	36.04	35.77	13.64	36.20
	Compactness	67.18	66.13	42.30	47.93	53.60	55.60
	Timbre	1.67	37.83	34.80	61.73	15.00	28.00
	RMS	20.00	40.00	30.00	20.00	0.00	21.60
	Spectral Flux	59.17	10	20.91	33.06	0.00	27.20
	Spectral Roll off point	34.34	52.60	29.07	14.29	0.00	24.40
	Spectral Variability	28.46	20.00	30.00	20.00	6.25	20.60

Mood Based Classification. Results favor MFCC here too (Tables 2 and 3).

4.4 Effect of MFCCs on the Result

The results indicate that once we use more than 10 MFCC Coefficients, the accuracy plateaus and doesn't increase at all. So, using around 15 coefficients is found to be good enough for the problem domain (Fig. 6).

4.5 Effect of Number of Hidden Nodes

We used only one hidden layer as it should be enough for our problem domain. As seen in the figure, for any number of hidden numbers after six or so, we get

Table 2. Mood classification(Arousal) using ANN and SVM

Feature	ANN			SVM		
	Low arousal	High arousal	Overall	Low arousal	High arousal	Overall
Spectral Centroid	70.07	26.20	50.34	50.00	50.00	44.14
MFCC	69.32	75.09	71.38	73.22	71.77	72.41
Zero Crossing	70.70	64.03	67.24	74.00	67.07	70.69
Pitch	44.27	64.62	54.83	59.00	55.01	56.55
Compactness	59.73	51.71	57.24	47.22	78.45	62.76
Timbre	58.96	61.10	58.28	62.58	53.87	56.55
RMS	70.76	67.28	68.97	50.00	50.00	42.76
Spectral Flux	76.58	58.87	67.93	40.00	60.00	45.86
Spectral Roll off point	50.79	47.67	50.34	70.00	30.00	41.38
Spectral Variability	73.28	62.51	67.24	50.00	50.00	40.00

Table 3. Mood classification(Valence) using ANN and SVM

Feature	ANN			SVM		
	Low valence	High valence	Overall	Low valence	High valence	Overall
Spectral Centroid	37.69	60.63	51.38	40.00	60.00	44.83
MFCC	60.04	65.45	63.79	45.37	72.06	58.62
Zero Crossing	62.74	57.80	59.66	70.34	52.46	60.34
Pitch	66.29	35.46	50.69	53.56	52.53	49.66
Compactness	50.67	58.79	52.76	57.25	62.40	58.97
Timbre	61.75	57.69	56.90	60.50	61.19	60.34
RMS	63.85	44.87	53.79	50.00	50.00	44.83
Spectral Flux	70.19	47.50	59.31	60.00	40.00	40.34
Spectral Roll off point	57.56	44.94	48.28	60.00	40.00	41.72
Spectral Variability	63.99	45.13	51.03	50.00	50.00	41.83

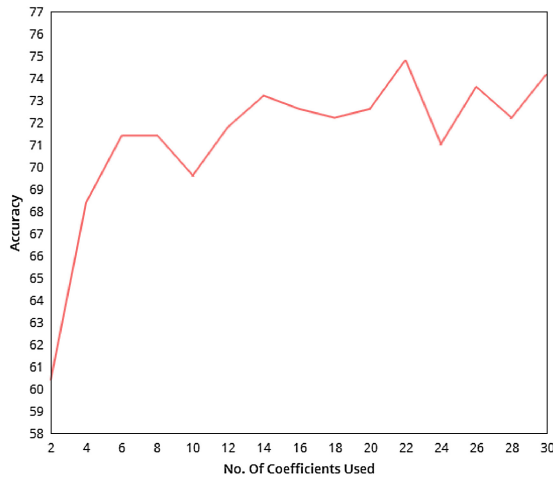


Fig. 6. Effect of number of MFCCs on result

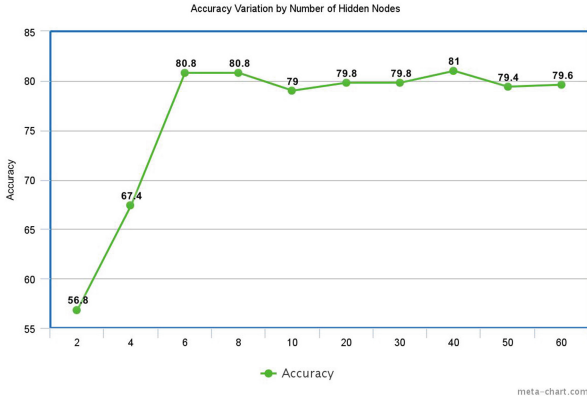


Fig. 7. Effect of number of hidden nodes

almost the same accuracy. As a rule of thumb, it is usually recommended that the number of nodes be around the mean of the number of inputs and outputs, so we chose 30 as our final number of hidden nodes (Fig. 7).

The number of nodes had minimal effect in regard to mood classification.

4.6 Effect of Number of Iterations

As seen in the figure, for genre classification, the number of iterations has an effect on the accuracy up to a certain point (around 20 iterations) (Fig. 8).

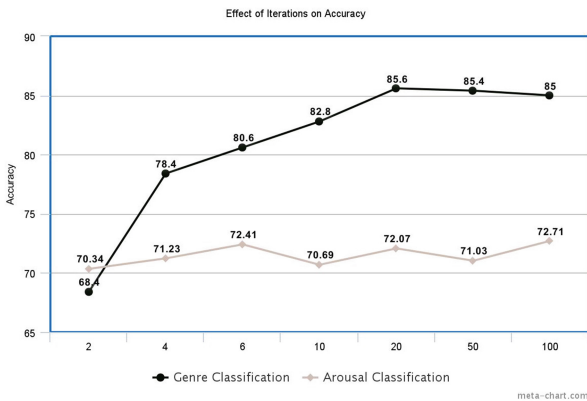


Fig. 8. Effect of number of iterations

As for Arousal, the increase in iterations had no effect on the accuracy (Fig. 9).

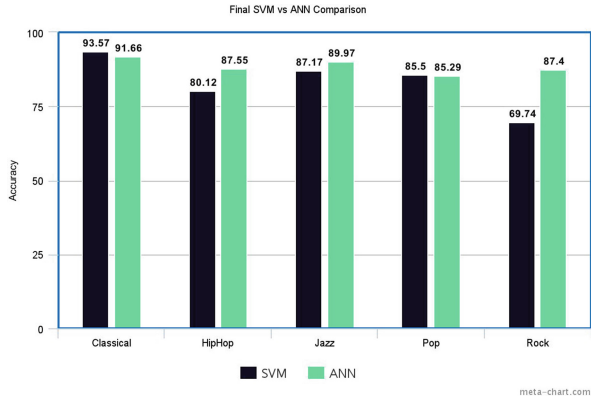


Fig. 9. Final SVM and ANN comparison based on genre

4.7 Final Results

Genre Classification. For our final model we used ANN with these feature: MFCC, Spectral Centroid, Zero Crossing, Compactness and RMS (Fig. 10 and Tables 4, 5 and 6).

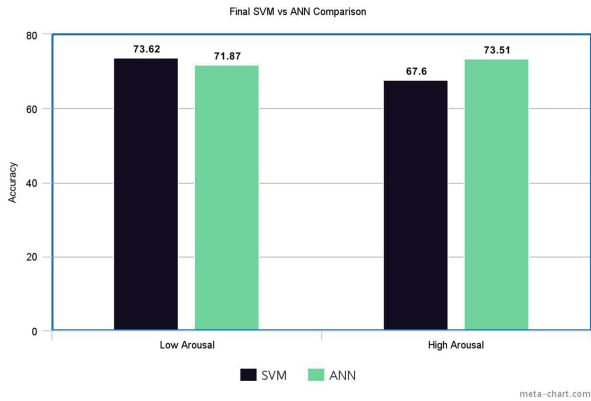


Fig. 10. Final SVM and ANN comparison based on arousal

Table 4. Genre classification performance measure

Classifier	Precision	Recall	F-Measure	Accuracy
SVM	0.87	0.94	0.89	83.00
ANN	0.94	0.92	0.92	88.80

Table 5. Genre classification confusion matrix

	Classical	Hiphop	Jazz	Pop	Rock
Classical	93	1	4	0	2
Hiphop	0	88	0	3	9
Jazz	5	0	90	1	4
Pop	2	7	0	86	5
Rock	0	4	4	5	87

Table 6. Arousal classification performance measure

Classifier	Precision	Recall	F-Measure	Accuracy
SVM	0.70	0.74	0.72	70.69
ANN	0.75	0.72	0.72	73.10

Mood Classification. For our final model we used ANN with these feature: Spectral centroid, MFCC, Zero Crossing, Compactness, Rhythm, Spectral Flux, RMS and Spectral Variability (Fig. 11 and Tables 7, 8 and 9).

For our final model we used ANN with these feature: Spectral centroid, MFCC, Zero Crossing, Compactness, Rhythm, Spectral Flux, RMS and Spectral Variability.

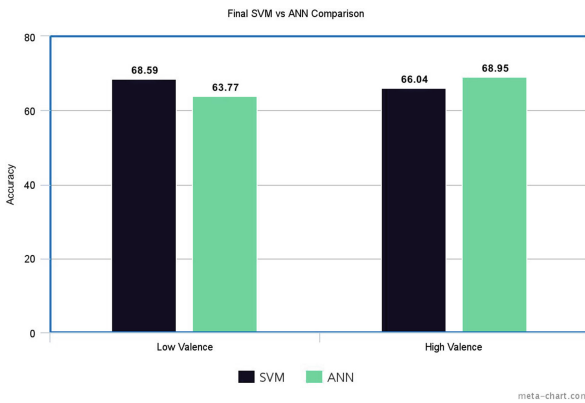


Fig. 11. Final SVM and ANN comparison based on valence

Table 7. Arousal classification confusion matrix

	Low arousal	High arousal
Low arousal	105	40
High arousal	38	107

Table 8. Valence performance measure

Classifier	Precision	Recall	F-Measure	Accuracy
SVM	0.68	0.69	0.68	67.59
ANN	0.68	0.64	0.65	67.24

Table 9. Valence confusion matrix

	Low valence	High valence
Low valence	95	52
High valence	43	100

5 Conclusion

Any type of classification of music is difficult simply because the classifications themselves don't have a clear definition. Still, we can work with fuzzy boundaries between these classes to get good enough results with Music Classification Systems.

In this paper, we studied many such components and approaches such as: types and combinations of features for proper representation of songs, feature integration approaches, classifier types, and their parameters, etc.

All these studies were done in order to tackle two related but distinct problems:

- In Automatic Music Genre Classification (AMGC), good performances were achieved with both of the classifiers employed: the final SVM model got 83% accuracy while the ANN model got 88% accuracy for five genres. These results are comparable with the state-of-the-art results, especially involving the same dataset.
- In Music Mood Classification however, the good results couldn't be replicated. The result along both axes of the music mood model used (arousal and valence) were underwhelming. Around 73% accuracy was achieved using ANN for the binary low/high arousal classification. SVM did even worse with around 70% accuracy. For low/high valence classification, both of the classifiers settled on 67% accuracy.

5.1 Limitations and Future Work

Distance Measure for Songs. One way to achieve song clustering or even classification is to develop distance measures to figure out the "distance" or difference between any two given songs. So, we tried to do the same. However, our initial attempts at using a simple Euclidean Distance measure were unsuccessful and later attempts using Gaussian Mixture Models proved to be too computationally intensive to be useful.

Future work could focus on figuring out appropriate distance measures for specific types of music being compared.

Feature Learning. Filtering and pre-processing might result in better high-level features. Or perhaps unsupervised feature learning methods as done in [30] might yield even better features. These approaches weren't explored in this paper.

Deep Learning. Future work could involve application of deep learning techniques in the problem domain.

Multi-tagging. A song can belong to multiple genre. So it is sure to consist of features characterizing multiple genre. Future work can be done to resolve this ambiguity.

References

1. Kassler, M.: Toward musical information retrieval. *Perspect. New Music* **4**, 59–67 (1966)
2. Andel, J.: On the segmentation and analysis of continuous musical sound by digital computer. Dissertation, Stanford University (1975)
3. Tzanetakis, G., Cook, P.: Manipulation, analysis and retrieval systems for audio signals. Princeton University, Princeton, NJ, USA (2002)
4. Alghoniemy, M., Tewfik, A.H.: Rhythm and periodicity detection in polyphonic music. In: 1999 IEEE 3rd Workshop on Multimedia Signal Processing. IEEE (1999)
5. Byrd, D., Crawford, T.: Problems of music information retrieval in the real world. *Inf. Process. Manag.* **38**(2), 249–272 (2002)
6. Saunders, J.: Real-time discrimination of broadcast speech/music. In: ICASSP, vol. 96 (1996)
7. Wold, E., et al.: Content-based classification, search, and retrieval of audio. *IEEE Multimed.* **3**(3), 27–36 (1996)
8. Scheirer, E., Slaney, M.: Construction and evaluation of a robust multifeature speech/music discriminator. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-97, vol. 2. IEEE (1997)
9. Casey, M.A., et al.: Content-based music information retrieval: current directions and future challenges. *Proc. IEEE* **96**(4), 668–696 (2008)
10. Kaminskas, M., Ricci, F.: Contextual music information retrieval and recommendation: state of the art and challenges. *Comput. Sci. Rev.* **6**(2), 89–119 (2012)
11. Muller, M., et al.: Signal processing for music analysis. *IEEE J. Sel. Top. Sig. Process.* **5**(6), 1088–1110 (2011)
12. Scaringella, N., Zoia, G., Mlynek, D.: Automatic genre classification of music content: a survey. *IEEE Sig. Process. Mag.* **23**(2), 133–141 (2006)
13. Perrot, D., Gjerdigen, R.: Scanning the dial: an exploration of factors in the identification of musical style. In: Proceedings of the 1999 Society for Music Perception and Cognition (1999)
14. Tzanetakis, G., Cook, P.: Musical genre classification based on audio signals. *IEEE Trans. Speech Audio Process.* **10**(5), 293–302 (2002)
15. Lippens, S., Martens, J.-P., De Mulder, T.: A comparison of human and automatic musical genre classification. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings. (ICASSP 2004), vol. 4. IEEE (2004)

16. Kour, G., Mehan, N.: Music genre classification using MFCC, SVM and BPNN. *Int. J. Comput. Appl.* **112**, 12–14 (2015). ISSN 0975-8887
17. Logan, B.: Mel frequency cepstral coefficients for music modeling. In: ISMIR (2000)
18. Li, T., Ogihara, M., Li, Q.: A comparative study on content-based music genre classification. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development In Informaion Retrieval. ACM (2003)
19. Anglade, A., et al.: Improving music genre classification using automatically induced harmony rules. *J. New Music Res.* **39**(4), 349–361 (2010)
20. Neumayer, R.: Musical genre classification (2004)
21. Koerich, A.L.: Improving the reliability of music genre classification using rejection and verification. In: ISMIR (2013)
22. Haggblade, M., Hong, Y., Kao, K.: Music genere classification. Stanford University, Department of Computer Science (2011)
23. Nasridinov, A., Park, Y.-H.: A study on music genre recognition and classification techniques. *Int. J. Multimed. Ubiquitous Eng.* **9**(4), 31–42 (2014)
24. Ekman, P.: Are there basic emotions? *Psychol. Rev.* **99**, 550 (1992)
25. Thayer, R.E.: *The Biopsychology of Mood and Arousal*. Oxford University Press, New York (1990)
26. Juslin, P.N., Sloboda, J.A.: *Music and Emotion: Theory and Research*. Oxford University Press, Oxford (2001)
27. Lu, L., Liu, D., Zhang, H.-J.: Automatic mood detection and tracking of music audio signals. *IEEE Trans. Audio Speech Lang. Process.* **14**(1), 5–18 (2006)
28. Fu, Z., et al.: A survey of audio-based music classification and annotation. *IEEE Trans. Multimed.* **13**(2), 303–319 (2011)
29. Soleymani, M., et al.: 1000 songs for emotional analysis of music. In: Proceedings of the 2nd ACM International Workshop on Crowdsourcing for Multimedia. ACM (2013)
30. Lee, H., et al.: Unsupervised feature learning for audio classification using convolutional deep belief networks. In: *Advances in Neural Information Processing Systems* (2009)