

# Analysis of the implementation methods for emotion recognition in music based on machine learning

**Gengyu Zheng**

School of Computing and Data Science, Xiamen University Malaysia, Sepang  
Selangor 43900, Malaysia

MBT2009454@xum.edu.my

**Abstract.** Music is a carrier of emotions, capable of triggering, expressing, and conveying people's innermost emotional experiences. With the development of digital technology, the recognition of emotions in music has been the subject of extensive research. From the viewpoint of application, MER is of great significance in music personalized recommendation services, psychotherapy, music creation, and music visualization. Currently, MER implementation methods based on machine learning are gradually becoming mainstream. This study starts from the basic knowledge of MER, first introduces the relevant research directions and research background of MER, and then proposes a three-part research framework combining MER and MEC. Based on this research framework, the machine learning algorithms and specific applications involved in the framework are introduced. Finally, the challenges and future paths facing the implementation of MER technology are proposed. Provide a reference for improving the accuracy and quality of using artificial intelligence methods to capture music characteristics and expressiveness.

**Keywords:** Music emotion recognition, music emotion classification, machine learning, affective computing.

## 1. Introduction

Computer music, which was born in the latter part of the 20th century, uses computer technology to comprehensively process various musical elements (such as pitch, rhythm, dynamics, timbre, etc.). It uses electronic audio processing technology to effectively combine with music in a new form. [1] In 1982, with the rapid development of digital technology, MIDI (Musical Instrument Digital Interface) was born, which solved the "compatibility problem" between devices, thus making computer music begin to reach the public and show a commercial trend. As one enters the twenty-first century, computer music has advanced significantly. Its high-performance digital data processing capabilities enrich the expressive power of music, allowing it to penetrate into various fields of composition of music, television and movie soundtracks, sound publishing, the Internet, mass entertainment, and education and teaching [1].

Music is the language of feeling, and emotion is one of the indispensable elements that constitute a musical work. Computer music is a combination of science and technology and music art. The expression and perception of emotions are becoming more and more crucial for personal choice and

human-computer interaction [2]. How to accurately identify and classify emotions in music has become a matter of great concern in the field of computer music.

Though they are not able to "hear" music the way humans can, computers can process audio to extract content aspects like rhythm, spectrum, and Mel Frequency Cepstrum Coefficients (MFCCs) [2]. This is music emotion recognition (MER). As a subtask of music information retrieval (MIR), computers are used to extract and evaluate musical features, map the relationship between musical aspects and emotional space, and portrayed the emotions expressed by music [3]. The majority of MER research to date uses supervised machine learning to train models, which then use listeners' emotion classifications from a huge collection of songs to build models for music emotion prediction. Machine learning often requires extensive emotional annotations of songs to create precise models for predicting emotions.

Numerous studies covering various facets of MER have surfaced in recent years. For example, data sampling can be used to create a data set based on music lyrics and tags. After single-label classification, multi-label emotion classification based on online Genrm label grouping appeared to improve the performance of emotion classification [4]. In order to quantify emotions and define the scope of research, researchers have used various data sets and proposed various emotion models. Kang et al. reviewed and summarized common and common centralized data sets and emotion models. The data sets include CAL500 data set, emoMusic data set, AMG1608 data set, DEAP data set, and emotional models include Hevner model, Thayer model, TWC model and PAD model, etc. [5]. In addition, more research is on the continuous in-depth exploration of classic machine learning and deep learning methods to improve the accuracy and practicality of MER. As music emotion recognition is applied and developed in a range of fields, it is of great significance to various fields such as recommendation systems, automatic music creation, psychotherapy, and music visualization.

This article initially provides an overview of the basic definition of music emotions as well as the research background on computer music and music emotion recognition. Based on previous research, it introduces the classification of music emotions and the music characteristics based on the classification. Secondly, a research framework is proposed based on MER and MEC, and commonly used machine learning related models (support vector machine, random forest, etc.) are explained. Then the research applications based on this framework are introduced, and lastly an analysis and summary of the issues facing this field now and its predicted future directions.

## **2. Basic descriptions of music emotion**

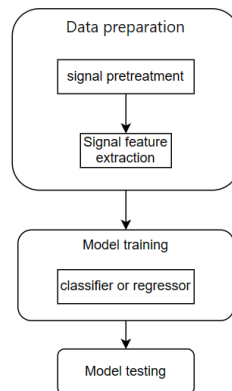
The structural aspects of music itself influence the emotional interpretation of it in different ways. In music, musical codes can generally communicate on an emotional level. From a structural point of view, the elements mentioned in The Aesthetic Mind include harmony, loudness, proximity to the main key, and maintenance of the overall structure. The elements are related to rhythm. For example, when listening to slow-paced music, one often feels calm or sad, while fast-paced music makes people happy and restless [6]. On the other hand, the emotion that humans generate from audio-visual language is elusive not only because it is nebulous and subjective, and because the assessment of musical emotion is influenced to differing degrees by human cognition and environmental shifts [7]. In comparison, the mechanisms by which music is perceived by humans and emotions are created are both intricate and unclear. Because pure machine classification is objective, it makes subjective music emotion classification more challenging [2].

Several contemporary research efforts have identified musical elements related to emotion, such as time, dynamics, pitch, timbre, intervals, rhythm, melody, articulation, harmony, pattern, musical form, vibrato, or sonority [8]. Research has examined the relationship between eight musical dimensions (melody, rhythm, harmony, timbre, dynamics, form, texture, and expressiveness) and specific emotions [8]. An article believes that using rough vocabulary classification is more convenient for users to search for emotion-related keywords in music stores. They divide emotions into six basic emotion categories, Exciting, Fear, Neutral, Relaxation, Sad, and Tension [7]. Land, the vast diversity of languages, cultures, and musical instruments brings blurred boundaries to this emotion classification. In this study, stimulating music generally featured irregular rhythms, fast tempo, and high loudness. Slow rhythms

and harmonizing harmonies characterize relaxation. Typically, sad music has a lower tone and a sluggish tempo. Tension music is usually high-pitched, fast-paced, and has conflicting harmonies. Neutral music expresses a mixture of other emotions [7].

### 3. Model

Currently, the traditional methods of music emotion classification (MEC) are divided into audio-based classification, lyrics-based text classification and multi-modal classification. Because the classification method of audio classification contains richer audio information, classification based on audio features has become a more common choice in MEC work. It is further divided into machine learning and deep learning [5]. In the practice of many studies, Music Emotion Classification (MEC) and MER often complement each other. One can divide MEC and MER work based on machine learning into three parts, namely data preparation, model training, and model testing. The overall framework is shown in Fig. 1. First, due to the high-dimensional and high-redundancy characteristics of music signals, signal preprocessing and signal feature extraction are required in the data preparation stage. Signal preprocessing can also be called domain definition. In this case, dataset and emotion model are selected. Suitable musical features are then extracted through subjective testing based on knowledge of music and psychology. [9] Apply machine learning in the model training phase, input the prepared music features into the model to train the classifier, use the classification algorithm to identify music emotion categories, and learn the connection between emotion labels and music features. Several techniques that are frequently used for classification are decision trees, K-nearest neighbors (KNN), support vector machines (SVM), and Bayesian. In this area, self-organizing maps, regression analysis, and artificial neural networks are a few machine learning techniques that are frequently employed.



**Figure 1.** Overall framework [5].

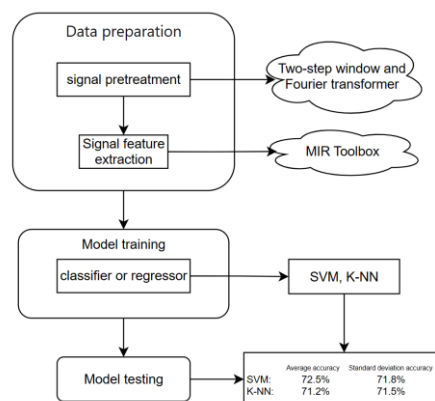
Among the machine learning models with good results in music emotion classification, SVM is a supervised machine learning technique, mainly used to solve classification and regression analysis. It can handle both linear and nonlinear models. In the linear model, the trained labels are divided into categories, and the ideal classification hyperplane is found in the high space of the feature vector, so that different types can get the largest interval and avoid over-fitting. Compared with other models, it is better at dealing with high latitudes. Data is even more advantageous. If a nonlinear condition is encountered, it is converted to linear through the kernel function [5]. Random forest classifier, which functions like a decision tree, is a strategy that combines multiple classifiers to solve difficult problems [10]. As a top-down tree structure, decision trees recursively represent feature attributes, results, and categories. Because they do not need to set parameters, they have low dependence on data, and their simple structure is widely used. But it is not suitable for situations where the amount of data is large or there are many categories. Researchers should adapt the models used to different situations, purposes, and indicators. Finally, in the model testing phase, the test data set is used to test the trained machine learning model to evaluate its performance and accuracy in emotion recognition. Han divides commonly

used indicators into classification indicators and regression model indicators. Among them, classification problems use accuracy and precision as indicators, and regression problems use  $R^2$  and root mean square error (RMSE) as indicators.

#### 4. Application

MER is a branch of study that combines artificial intelligence, natural language processing (NLP), audio signal processing, and music psychology. It determines the emotional trend of music through music intelligent computing. It can be used for a variety of tasks, including music tag classification, music information retrieval, music recommendation and other fields to improve the personalization and quality of listeners' music experience. In addition, it is also used in psychotherapy, music visualization, automatic composition of music and other domains [3]. The traditional application of MER is information retrieval. In many online music streaming services, such as Apple, Spotify, Genesis, Last FM, iTunes, etc., they provide users with automatically classified playlists based on emotions for the purpose of organization, management, and recommendation to solve the problem of individual differences [10]. It is also applied to emotion perception and music recommendation in the game, video, and movie industries, allowing music to be used for learning and health.

The research by Deepti Chaudhary and others performs music emotion recognition on Indian songs, making full use of and comparing machine learning models, which is in line with the framework mentioned above. As shown in Fig. 2, by using 3150 Indian songs as a data set, 70% is used to train the classifier and 30% is used for model testing. Classify them into different categories using two-step windowing and Fourier transformer framing. Use the MIR toolbox to extract 9 emotional features. Then, one can use SVM and K-NN classifiers for training. The test results obtained are as shown in the Fig. 2, and one concludes that SVM is better than K-NN [10].



**Figure 2.** The research by Deepti Chaudhary et al. corresponds to the MER working framework based on machine models [10].

The article by Jessica et al. describes the application of MER in the field of psychotherapy [11]. Music is a popular choice of stimulation, and the emotions they evoke can be reflected through human physiological signals, such as galvanic skin responses, whose correlation properties such as chills and microtremors reflect the intensity of emotions elicited by music. Music therapy has significant effects on improving mental health, such as reducing anxiety, improving sleep quality, and reducing epileptic seizures. Researchers use the MER calculation method to analyze the physiological signals of subjects after listening to different types of music to help understand which music should be used to face and treat different diseases. Subjects subjectively classify emotions, and use classification methods such as neural network (NN), K-NN, and SVM to classify signal features under different conditions. According to experimental findings, NN can discern between musical genres with an accuracy of up to 99.2% [10], which provides a great scientific basis for the connection between physiological signals and music. Research could improve music therapy by revealing which parts may trigger epilepsy, thereby reducing

musical epilepsy and improving patients' mental health. Research by Anupam et al. is similar. They proposed a model for calibrating MER using machine learning methods, combining physiological signals with human brain electrical activity. In addition to the preparation, training, and testing and evaluation stages of extracting music emotional features, they also generate music playlists based on the features and real-time emotions generated by users based on physiological sensors to test the practicality of the work and combine psychotherapy with music recommendations [12].

In addition to basic machine learning methods, MER based on deep learning has received a lot of attention because it is suitable for situations with huge data and does not require audio feature extraction. Commonly used methods include recurrent neural network (RNN), convolutional neural network (CNN) and Gaussian mixture model (GMM). GMM is often used in popular music emotion recognition research, while CNN is suitable for spatial data processing and is widely used in image processing. It can be used CNN recognizes Mel spectrograms to achieve music emotion classification and is applied to automatic labeling and music feature learning [5]. Research on music creation through music emotion recognition mostly uses deep learning models, which have obvious advantages in automatically creating high-quality polyphonic music pieces by learning music features from data sets [13]. In the research of Lucas et al., they created a symbolic music data set. 30 subjects annotated each musical work according to the model, and mapped the annotation results to emotions to extract the emotion of the work. They use a deep learning model such as generative Long short-term memory (LSTM) with logistic regression to create music with specific emotions. At the same time, this model can also be applied to symbolic music sentiment analysis [13]. Yu's research method also uses the neural network (NN) algorithm and LSTM model, uses music theory laws as filter restrictions, improves the accuracy of model creation through continuous iteration, and generates images that include sadness, happiness, loneliness, and relaxation. music [14]. These technologies can be used to provide soundtrack generation systems for movies, games, audiobooks, etc., which is an intelligent development of traditional music creation models.

## 5. Limitations and prospects

Although many scholars have conducted in-depth research on the MER problem in many aspects, and the accuracy of classification has been continuously improved so far, and good results have been obtained, there is still a lot of room for improvement overall. First, the subjectivity and complexity of music emotions make it difficult to quantify. It is difficult for researchers to objectively classify based on manually extracted features and use appropriate adjectives as emotional annotations. Secondly, in order to fully tap the problem-solving capabilities of machine learning and deep learning, high-quality, diverse and large-scale music open data is needed. But so far, relatively few data sets on music emotion have been publicly provided to the research community, and many data sets are only suitable for single experiments or benchmark tests. There is currently insufficient research on the emotional characteristics of music. Generally, classification analysis is performed by extracting the physical characteristics of music, but the relationship between characteristics and emotions requires further research. In order to solve the problems of which characteristics can accurately convey musical emotions and how to evaluate the extracted features, a reasonable music feature model needs to be established. Starting from the application level, there are few studies on using MER based on machine learning for music creation. Modeling musical style without replicating the musical content is a challenge for large machine learning projects.

In the future, new high-quality data collection emotional models will continue to appear. For music emotion data sets, in addition to opening anonymous user data and maintaining experimental repeatability, discrete emotion models and continuous emotion models should also be integrated during label annotation to reduce errors caused by subjective factors. For the annotation of the data set, because the musical emotion changes dynamically in the music, dynamic annotations and annotations that are more relevant to the context will be more accurate, so adding a dynamic recognition model is necessary. In order to improve the performance and accuracy of recognition, using deep learning methods and integrating them with other methods for MER is a major research trend. A major problem with deep learning models is that their opacity hinders the interpretation of correlations between emotions and

features. Deep learning improves the interpretability of extracted features, which is an important issue that needs to be solved in the future. There are ethical issues in the application of MER. One needs MER to be applied in a fair, safe and non-harmful way to the public interest.

## 6. Conclusion

To sum up, this study provides a review of current MER research based on machine learning. First, the research background and definition of computer music and MER are introduced, and the related research directions involved in MER are briefly introduced. Then it introduces the research framework combining MER and MEC under the current situation, explaining the relevant knowledge of each part and the machine learning methods used. Then, different fields of application of MER based on machine learning are explained based on the framework and actual research. Finally, the current problems and prospects for future research and development are summarized. This article explores the future path of MER by sorting out its development route and framework. MER is in a vigorous stage of development. The basic need is to increase the accuracy of emotion recognition through machine learning or deep learning. Large-scale, accurate, and publicly available data sets, as well as emotion categorization models, are urgently needed. In addition, there are Many music-related aspects such as MVs, music videos, and music creation can be explored. MER has important research significance and huge development potential, and researchers still need to continue to innovate and improve.

## References

- [1] Wang L 2021 Research on the reform of music education mode in colleges and universities based on computer music technology, *Journal of Physics: Conference Series* vol 1744(3) p 032149.
- [2] Xia Y and Xu F 2022 Study on Music Emotion Recognition Based on the Machine Learning Model Clustering Algorithm, *Mathematical Problems in Engineering* vol 2022 p 9256586.
- [3] Han D, Kong Y, Han J, et al. 2022 A survey of music emotion recognition, *Front. Comput. Sci.* vol 16 p 166335.
- [4] Lin Y C, Yang Y H and Chen H H 2011 Exploiting online music tags for music emotion classification, *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* vol 7(1) pp 1-16.
- [5] Kang J, Wang H, Su G and Liu L 2022 Survey of Music Emotion Recognition, *Computer Engineering and Applications* vol 58(4) pp 64-72.
- [6] Schellekens E and Peter 2012 *The Aesthetic Mind: Philosophy and Psychology* (Oxford).
- [7] Pandeya Y R and Lee J 2021 Deep learning-based late fusion of multimodal information for emotion classification of music video, *Multimed Tools Appl* vol 80 pp 2887–2905.
- [8] Panda R, Malheiro R and Paiva R P 2020. Audio features for music emotion recognition: a survey, *IEEE Transactions on Affective Computing* vol 14(1) pp 68-88.
- [9] Han D, Kong Y, Han J, et al. 2022 A survey of music emotion recognition, *Frontiers of Computer Science* vol 16(6) p 166335.
- [10] Shashikumar G T, Spoorthi N P and Geeta R B 2022 Emotion Based Music Classification, *Journal of Next Generation Technology* vol 2(1) pp 22-29.
- [11] Rahman J, Gedeon T, Caldwell S, Jones R and Jin Z 2021 Towards Effective Music Therapy for Mental Health Care Using Machine Learning Tools: Human Affective Reasoning and Music Genres, *Journal of Artificial Intelligence and Soft Computing Research* vol 11(1) pp 5-20.
- [12] Garg A, Chaturvedi V, Kaur A B, et al. 2022 Machine learning model for mapping of music mood and human emotion based on physiological signals, *Multimed Tools Appl* vol 81 pp 5137–5177.
- [13] Ferreira L N and Whitehead J 2021 Whitehead J. Learning to generate music with sentiment, *Learning to generate music with sentiment*. arXiv preprint arXiv:2103.06125.
- [14] Yu W 2021 Music Composition and Emotion Recognition Using Big Data Technology and Neural Network Algorithm, *Computational Intelligence and Neuroscience* vol 2021 p 5398922.