Research Article

Music Emotion Classification Method Using Improved Deep Belief Network

Guiying Tong ⬤

College of Music, Huizhou University, Huizhou, Guangdong 516000, China

Correspondence should be addressed to Guiying Tong; tgy@hzu.edu.cn

Aiming at the problems of difficult data feature selection and low classification accuracy in music emotion classification, this study proposes a music emotion classification algorithm based on deep belief network (DBN). The traditional DBN network is improved by adding fine-tuning nodes to enhance the adjustability of the model. Then, two music data features, pitch frequency and band energy distribution, are fused as the input of the model. Finally, the support vector machine (SVM) classification algorithm is used as a classifier to realize music emotion classification. The fusion algorithm is tested on real datasets. The results show that the fused feature data of pitch frequency and band energy distribution can effectively represent music emotion. The accuracy of the improved DBN network fused with the SVM classification algorithm for music emotion classification can reach 88.31%, which shows good classification accuracy compared with the existing classification methods.

1. Introduction

In the Internet era, the increasing demand for music has promoted the vigorous development of the music market, and a large number of online music resources continue to flow into the market. In order to better organize, retrieve, and recommend music, various music applications launched by many companies, such as QQ Music, Netease Cloud Music, and KuGou Music, have adopted intelligent classification algorithms to classify music in multiple dimensions and launched music recommendation services according to users' listening habits. Common classification methods use music metadata as classification labels, such as year, singer name, song name, etc., but these classification methods do not involve the audio content information of the music itself, so users cannot search for music through feelings. Therefore, emotion, as an important factor of song expression, has become an important recommendation index in music classification.

Some studies use wearable devices to record the specific psychological feelings generated by users listening to different kinds of music and use a classification method based on user information to identify music emotions [1]. With the development of machine learning and other related technologies, computers can realize complex emotional analysis and calculation. Automatic output of emotion analysis results through algorithms has gradually become a method adopted by many scholars. Wen [2] used BP neural network to classify the emotional features of music, but the BP neural network model had the inherent defects of being very sensitive to the initial weight and easy to converge to the local minimum. Because deep learning algorithms such as convolution neural network (CNN) and recurrent neural network (RNN) have the advantages of strong learning ability and good portability, the research on emotion classification methods has moved from shallow learning methods to deep neural network classification methods. Combining feature extraction and time series data classification, Yang et al. [3] proposed a hybrid architecture called the parallel recurrent convolution neural network. Parallel CNN and Bi-RNN focus on extracting spatial features and time frame sequences, respectively, fusing the outputs of the two modules into a powerful time series vector, and finally sending the fusion vector to the Softmax function for classification.

In terms of feature selection of music classification, Lee et al. [4] put forward a feature set suitable for the original

audio for the performance of traditional Korean musical instruments and summarized a variety of common classification methods. But its method is only for Korean musical instruments, and its application has limitations. Kim et al. [5] improved the one-dimensional CNN architecture for automatic music annotation, adopted building blocks from the most advanced image classification models, ResNet and SENET, and added multilevel feature aggregation. However, because too many features are selected, it is easy to lead to overfitting in training. From the perspective of the frequency domain, references [6, 7] use Mel frequency as a data feature and adopt a deep learning method to classify music emotions, but the classification accuracy needs to be improved.

Music classification is an important method to deal with massive amounts of music information. Emotion, as an important factor of expression of song, has become an important recommendation index in music classification. The traditional music emotion classification method has the defects of long time-consuming and low accuracy. Therefore, it is of great practical value and practical significance to study how to design a music emotion classification method with fast learning speed and high classification accuracy. Therefore, this study improves the traditional deep belief network (DBN) and proposes a fusion classification algorithm combined with the support vector machine (SVM) classification algorithm. Its basic ideas are as follows: (1) select the two features of music pitch frequency and band energy distribution to preprocess the music data and fuse the feature data. (2) Based on the improved deep belief network, a training framework for fused feature data has been designed. (3) Support vector machine classification algorithm is used to complete the final music emotion classification.

2. Music Emotion Classification Algorithm

2.1. Overall Framework. Source judgment of feature data and classifier selection are two important links in music emotion classification [8, 9]. In particular, the source judgment of feature data needs to explore multifeature data so as to avoid the problem that music emotions cannot be fully expressed due to single-feature data. Therefore, this study extracts 10 dimensions of data based on pitch frequency and band energy distribution to ensure the diversity of feature data, selects the improved deep belief network as the feature data training model, and finally uses the SVM algorithm for music classification. As shown in Figure 1, the algorithm proposed in this paper mainly includes constructing multifeature data, extracting and fusing feature vectors, constructing a deep belief network model suitable for music emotion classification, and SVM algorithm classification.

2.2. Music Emotion Personalization. Music emotion itself is subjective. Individual differences such as gender, personality, and cultural background will lead to different individuals having different emotional cognition of the same song [10]. The commonly used music emotion classification
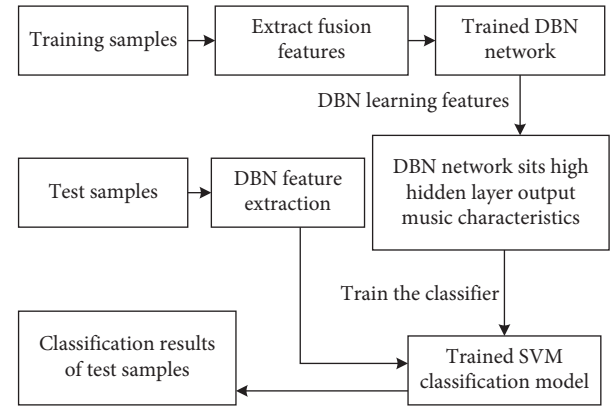

Figure 1: Algorithm structure block diagram.

is the classifier obtained by training the emotional data of music by different individuals. However, music emotions are themselves subjective, and the classification accuracy is not ideal for each different individual. Establishing an emotion classifier for each individual user to classify the same music dataset is a solution to the problem of personalized classification of music emotions proposed by many scholars.

2.2.1. Music Emotion Personalization Based on Individual User. The most direct way to deal with individual differences is to individualize the system. Grimaldi M and others have used the historical records of users listening to music to mine users' music tastes and achieved good results [11]. If this method is extended to music emotion classification, obtaining the emotional features of users listening to the same type of music over a certain period of time can improve the effect of music emotion classification. Yang et al. [12] proposed a two-layer classifier to realize music emotion personalized classification. The first layer used a bag-of-users model to train music general emotion classification, and the second layer used a residual model to predict the difference between general emotion cognition and a specific individual user's emotion cognition. The research showed that this method treated the music content itself and a specific individual user separately, and this method was better than the traditional single-layer classification method. Su and Fung [13] proposed a method based on active learning to realize the personalized classification of music emotions. After labeling the music data set, the experimenters were invited to listen to the music and judge whether the music was consistent with the given classification label. A key problem with this method based on user feedback is the burden problem. Too much participation is a burden for users [14]. In order to avoid excessive user participation, less music and user interaction should be designed. At the same time, in order to ensure the amount of data, a general classification method should be introduced to form a mixed classification method as follows:

$$C(x_i) = (1 - \omega) * C_{\text{popular}}(x_i) + \omega * C_{\text{individualization}}(x_i), \quad (1)$$

where, C is a classifier, C_{popular} is a general classifier, $C_{\text{individualization}}$ is a personalized classifier, and $\omega \in [0, 1]$ represents the weight.

2.2.2. Personalization Based on User Groups. User group-based personalization technology improves the personalization performance by extracting users' personal information (such as age, gender, etc.). For general emotion classification, the truth value of data often takes the average value of all test users' voting labels. For groupwise music emotion classification (GWMEC), the processing of data truth values is different. All users are grouped according to individual information. For example, all users can be divided into male and female groups according to gender, and users can be divided into eastern and western cultural groups according to cultural background. After grouping, calculate the average value of each group according to the voting labels of all members in the group. Finally, based on the truth value of each group, a music emotion classifier is calculated for each group. Yeh et al. [15] have divided this method into five steps:

(1) *Data collection.* We collect user attributes, music attributes, and music emotions, which are expressed by vectors respectively;

(2) *User emotion group clustering.* We cluster all emotions into more representative emotion groups;

(3) *User group classification.* We group users according to user attributes;

(4) *Music emotion classification.* For each user group, music emotion classification is carried out once;

(5) *Personalized emotion classification.* We calculate and determine the group information of the user and generate personalized emotion classification through emotion classifiers of the group.

2.2.3. Personalization through Social Information. An important problem of music emotional personalization is to reduce the participation burden of users [16]. Generating social information (such as focusing on others) is a very natural operation for users on the online music platform. Even without the personalized classification of music emotions, many users will still generate a lot of social data on the online music platform. This method is to realize the personalization of music emotion based on social information. The personalization method of social information has been studied in many fields and achieved good results, such as friend recommendations, music recommendations, film recommendations in social media such as Facebook, Twitter, Sina Weibo, Renren, Douban, etc. Ma et al. [17] proposed the Geolife system, which established a user map representing social relations for users. Through the travel track shared by users, the system recommends places and friends for other users of social networks. This application not only facilitates users' lives to a great extent but also increases their interest in their lives. Bu et al. [18] combined social media information with music content to create a unified hypergraph, on which training calculations were carried out. A personalized music recommendation was carried out for users, and good results were achieved. The Film-Trust system proposed in the study by Golbeck and Hendler [19] recommended the movies that users have seen and may be most interested in, and the algorithm part also involved the user's social network information. Therefore, the personalized classification of music emotions through social networks is also a feasible method.

3. Multifeature Fusion and Emotional Feature Extraction

3.1. Audio Segmentation Preprocessing. The audio dataset used in traditional audio emotion classification research is a pure music segment or voice segment, with short audio time and a single composition. However, the audio composition of modern digital music, especially musical instruments, vocals, and effectors in pop music, is complex [20], and the time is as long as 3–4 minutes, which makes feature extraction difficult. Therefore, this study proposes two audio segmentation preprocessing methods to solve the above problems.

3.1.1. Vocal Separation. In the traditional music emotion classification methods, the audio feature classification performance of pure music segments is outstanding. For complex music composed of audio fused with human voice and background sound, the music is preprocessed by human voice separation, and the classification effects of human voice and background sound are studied, respectively. In the actual process, four levels of segmentation are used to construct the dataset for the whole music. The first is dividing the music into 30 s segments on average; the second is fine-grained segmentation to 15 s sentence level, and the other two are pure human voice and pure background sound fragments extracted by audio processing tools. Through experiments, the classification effects of different preprocessing methods are compared to improve the performance of the emotion classification system.

3.1.2. Fine-Grained Segmentation. The long duration of music will lead to the phenomenon of large feature dimension, slow training speed, and overfitting of classifiers. Considering that the same music shows different emotional tendencies in different periods, there may be local problems in the emotional classification of the whole music. Therefore, in order to improve the speed and accuracy of classification, the overall emotion of music is integrated to make fine-grained segmentation of music data in the classification process, and the classification results are obtained by voting decision-making, which can improve the accuracy of classification.

3.2. Extraction of Fusion Features. The sensitivity of different kinds of features to emotion is different. The traditional single-feature data can be better classified by music types,

musical instruments, and other content. But when classifying the emotions expressed by music, the sensitivity of the single feature to music emotions will lead to lower classification accuracy. Although there are many kinds of music, emotions can be expressed in the frequency domain. Therefore, this paper chooses to fuse the music features from the two aspects of pitch frequency and band energy distribution, and use the feature matching method for classification.

In the song, the pitch frequency of the singer's voice will change with the singer's state of mind and emotion, and it is also the lowest frequency in the whole music spectrum. The frequency band energy distribution describes the energy distribution of music signals in the frequency domain, and the energy level directly reflects the music emotion type and change trend. In this study, the band energy is divided into 6 groups, with frequencies of 0~250 Hz, 250~500 Hz, 500~1000 Hz, 1000~2000 Hz, 2000~3000 Hz, and 3000~5000 Hz, respectively. During classification, each piece of music is converted into an audio frame, and the maximum, minimum, standard deviation, and average value of pitch frequency are extracted.

4. Classifier Based on Deep Belief Network

4.1. Improvement of Deep Belief Network. DBN, like CNN, is a typical deep learning model that can learn the corresponding input and obtain more abstract and higher-level features. DBN is mainly a machine learning model under unsupervised learning, but in practical applications, DBN and the BP algorithm are often combined to realize the unsupervised forward propagation of sample data features through the restricted Boltzmann machine (RBM), and then the reverse fine-tuning through the supervised BP algorithm is realized. As shown in Figure 2, a typical DBN model structure diagram mainly includes forward propagation and backpropagation processes. It can be seen from the model diagram of DBN that it is mainly formed by the superposition of multiple RBMs, so the reconstruction process of the visible layer and hidden layer in RBM and the training process of RBM are also applicable to the DBN model structure.

In view of the fact that deep learning related models can highly abstract and learn the features of sample data when studying the recognition and classification of music emotions, this paper considers the use of deep learning related technologies for music emotion classification. Although many major breakthroughs in deep learning in recent years are based on CNN, compared with the advantages of CNN in processing two-dimensional data, DBN is more suitable for processing one-dimensional data. A music signal is a typical one-dimensional data. To sum up, it is undoubtedly advantageous to use DBN in the research of music emotion classification in this paper.

DBN is composed of multilayered RBM. The insufficient network training accuracy and overfitting phenomenon are affected by the number of layers [21, 22]. The training process of DBN is progressive layer by layer, which is called generative pretraining [23]. After a RBM is trained, the
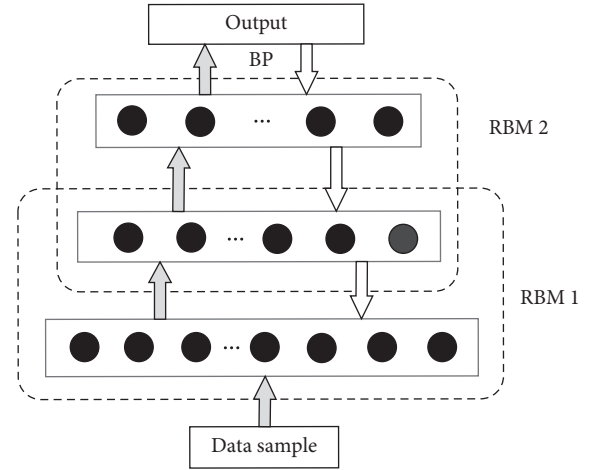

Figure 2: DBN model structure.

activation probability of its hidden unit is used as the input data of the next layer, and it is trained layer by layer until the end of all model training. According to the characteristics of music emotion classification feature data, this study adds discrimination and fine-tuning operations on the basis of generating pretraining method. That is, in the training process, a hidden layer node is added to each layer of the RBM model to provide it with the label of training data, and the values of all the weights are fine-tuned to increase the training accuracy of the model. As shown in Figure 3, the DBN network is composed of an n-layer improved RBM model, a one-layer traditional RBM model, and a Softmax layer. The first layer of traditional RBM is the input layer with n input vectors, and the Softmax is the output layer with m nodes.

4.2. Classification Model Based on Improved DBN and SVM. Based on the feature analysis of the training sample set, the SVM algorithm uses a small number of support vectors that can better represent the classification information of the whole training sample set to participate in training, which can effectively reduce the training time and further accelerate the convergence speed. Therefore, this study integrates the SVM classification algorithm [24] with the DBN network; that is, the output node of the Softmax layer (the last layer of the improved DBN network) is used as the input of the SVM classifier to realize music emotion classification. The whole classification process first preprocesses the audio segmentation, extracts the fusion feature data of music as the input of the DBN network model, obtains the music features extracted by DBN through deep learning, and finally trains through the SVM classification algorithm to realize the music emotion classification. The model integrating the improved DBN network and SVM algorithm can retain the original features and have the advantages of an SVM classifier [25]. The music emotion classification process based on the DBN and SVM fusion algorithm is shown in Figure 1.

As can be seen from Figure 1, the fusion algorithm function of music emotion classification is divided into two parts: feature extraction and emotion classification. The
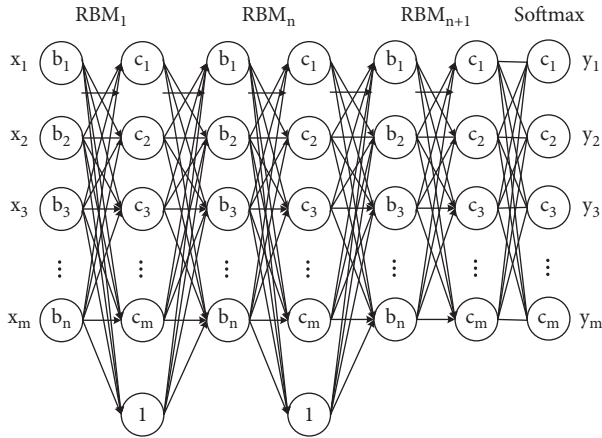
FIGURE 3: Improved DBN model structure.

FIGURE 4: Relationship between classification error and number of iterations.

extracted fusion music features are input into the DBN network. Each layer of RBM in the DBN network is used to learn the original data. The output of the Softmax layer in DBN is extracted as the feature data used for classification. Finally, the trained SVM classifier is used to obtain the classification accuracy from test sample data. The training data of the classifier adopts the feature samples extracted by DBN. The influence factors of the fusion algorithm on the accuracy of music emotion classification mainly include the feature parameters input into the DBN network, the features extracted by DBN and the training effect of the SVM classifier.

5. Experiments and Analysis

5.1. Music Data. In this study, a small dataset in the FMA music analysis dataset launched in 2017 is used as the original data for music emotion classification [26]. The FMA data set contains 163 genres, 106574 songs, and information such as song ID, name, singer, genre, and times of play, which can meet most music information retrieval tasks. However, the lengths of songs in this data set are different, and the classification of each genre is unbalanced, so the data processing is difficult, and the test results are easily affected. Therefore, 8 genres and 8000 songs in the FMA-Small dataset are selected as the test data. The duration of each song is 30 s and is stored in MP3 format. The sampling frequency is 44100 Hz and the sampling rate is 320 kb/s.

5.2. Music Emotion Classification Experiments

5.2.1. Classification Error. As shown in Figure 4, the classification error is mainly affected by the number of iterations. With the increase in the number of iterations, the classification error decreases gradually. However, when the number of iterations increases to 150, the error gradually tends to be stable. Therefore, in the case of meeting the error requirements, try to set a small number of iterations to reduce the training time and increase the training efficiency.
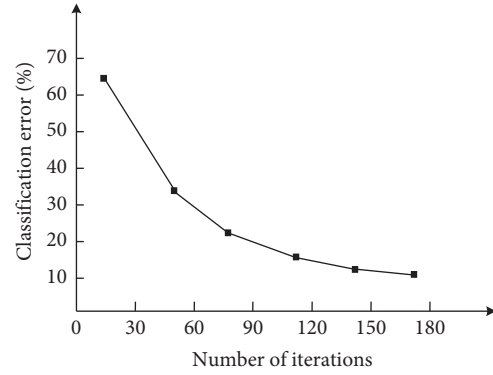
5.2.2. Classification Accuracy of Different Classification Algorithms. In order to verify the effectiveness of the proposed algorithm for music emotion classification, the SVM algorithm, DBN, and SVM fusion algorithm are selected for comparison. Except for the algorithm, the other test settings are consistent. The test results are shown in Table 1. When the algorithm combining DBN and SVM is used for music emotion classification, its classification accuracy can reach 68.4%, which improves the accuracy by 17.6% compared with the simple SVM algorithm and shows that the DBN network has strong learning ability for music features. After improving the DBN network and integrating the SVM classification algorithm, the classification accuracy was further improved to 83.35%, which proves the effectiveness of the improved algorithm in this study. Compared with the traditional DBN network, adding a hidden layer node to each layer of the RBM model to provide it with the label of training data, and fine-tuning all the values of weights can effectively increase the training accuracy of the model.

5.2.3. Accuracy of Classification Results. As shown in Table 2, the classification accuracy of the proposed algorithm for different music emotions is higher than 75.65%, and the highest classification accuracy for "Passionate" music is 88.31%. The main reason for the large fluctuation in the classification results of different emotional music is that the classification is based on the two features of pitch frequency and band energy distribution, which have certain limitations. However, compared with the algorithm using only Softmax as the classifier model, the accuracy of different emotional classifications is comprehensively better. The experimental results show that the classification model based on improved DBN and SVM can significantly improve the accuracy of music emotion classification.

5.2.4. Performance Comparison with Existing Research Methods. In order to further verify the performance of the fusion classification algorithm proposed in this study, three classification models of CNN + LSTM network proposed in reference [27], LeNet network and ResNet network proposed in reference [28] are selected for classification performance comparison experiments. The experimental environment and

TABLE 1: Accuracy of classification results of different classification algorithms.

Algorithm	Accuracy
SVM	50.8%
DBN + SVM	68.4%
Improved DBN + SVM	83.35%

TABLE 2: Accuracy of music emotion classification results.

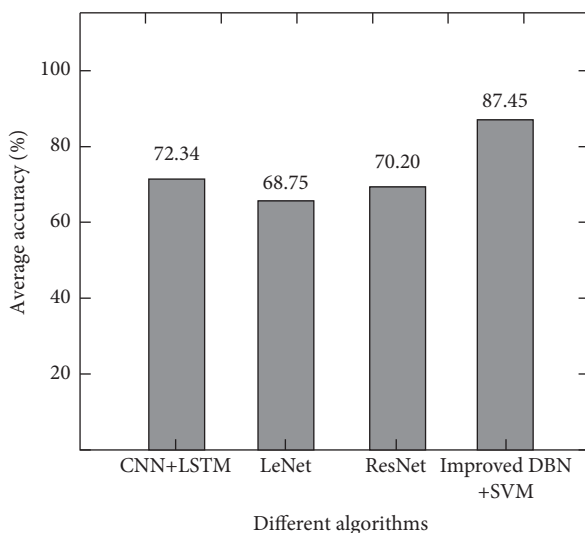| Classification	Accuracy (Improved DBN + SVM)	Softmax
Sad	79.58%	69.68%
Happy	86.26%	75.28%
Quiet	78.26%	66.94%
Passionate	88.31%	80.13%
Romantic	75.65%	68.99%
Thrilling	82.64%	78.48%

FIGURE 5: Comparison with the average accuracy of the existing research methods.

dataset are consistent with this study. The experimental results are shown in Figure 5. The average accuracy of the three selected algorithms is 72.34%, 68.75%, and 70.20%, respectively. The average classification accuracy of the fusion classification algorithm proposed in this study reaches 87.45%, which is the best performance. In addition, the experimental results show that the music data fusion feature proposed in this study can be effectively applied to the classification of music emotions. This is because DBN can extract features from sample data through autonomous learning, which is suitable for tasks requiring highly abstract and complex features. Compared with CNN, which is good at processing two-dimensional pictures, DBN can process one-dimensional data well. Considering that music signals are typically one-dimensional data, compared with the CNN + LSTM algorithm, the improved DBN shows stronger applicability to music emotion classification. During the process of experiments, the LeNet network and ResNet network are easy to confuse the music styles of neutral and negative emotions, so they have low average accuracy, and the training time in the recognition process is longer than the other two algorithms.

6. Conclusion

Aiming at the problem of music emotion classification, this study proposes a fusion classification algorithm based on the DBN network. The pitch frequency and band energy distribution, which can best reflect the singer's state and music emotion change in the song, are used as the original feature data, and the two feature data are fused, which can simply and effectively express the emotional style represented by music. The traditional DBN network is improved, and the pitch frequency and band energy distribution are fused as the input of the model. Finally, the SVM algorithm is used as the classifier to realize the emotion classification of music. Experiments on real data sets show that the fusion feature data of pitch frequency and band energy distribution can effectively represent music emotions, and the improved DBN network fused with the SVM classification algorithm can effectively improve the accuracy of music emotion classification.

In future work, the additional features of music that are not considered in this paper should be deeply excavated, and various features of music should be integrated as comprehensively as possible on the premise of avoiding overfitting. In addition, considering that the data of online music platforms are generated at any time and the music style and quantity will continue to increase with the passage of time. In the follow-up research, richer music datasets should be selected for model training to improve the model training accuracy and optimize the algorithm.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] H. Tong, M. Zhang, P. Soleimaninejadian et al., "Music Mood Classification Based on Lifelog," in *Proceedings of the 24th China Conference on Information Retrieval (CCIR)*, pp. 55–66, Guilin, China, September 2018.

[2] Y. Wen, "Personal Music Emotion Analysis Based on BP Neural Model," in *Proceedings of the 5th International Conference on Electrical & Electronics Engineering and Computer Science (ICEEECS)*, pp. 285–290, Bali, Indonesia, November 2018.

[3] R. Yang, L. Feng, H. Wang, J. Yao, and S. Luo, "Parallel recurrent convolutional neural networks-based music genre classification method for mobile devices," *IEEE Access*, vol. 8, no. 99, pp. 19629–19637, 2020.

 [4] J. Lee, M. Lee, D. Jang, and K. Yoon, "Korean traditional music genre classification using sample and MIDI phrases," *KSII Transactions on Internet and Information Systems*, vol. 12, no. 4, pp. 1869–1886, 2017.

 [5] T. Kim, J. Lee, and J. Nam, "Sample-Level CNN Architectures for Music Auto-Tagging Using Raw Waveforms," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 366–370, IEEE, Calgary, Canada, April 2018.

 [6] R. Geetha Ramani and K. Priya, "Improvised emotion and genre detection for songs through signal processing and genetic algorithm," *Concurrency and Computation: Practice and Experience*, vol. 31, no. 14, pp. e5065–5062, 2019.

 [7] R. Delbouys, R. Hennequin, F. Piccoli, J. Royo-Letelier, and M. Moussallam, "Music mood detection based on audio and lyrics with deep neural net," in *Proceedings of the International Society for Music Information Retrieval Conference*, vol. 48, pp. 1465–1472, Paris, France, September 2018.

 [8] Y. Shu and G. Xu, "Emotion recognition from music enhanced by domain knowledge," in *Proceedings of the Pacific Rim International Conference on Artificial Intelligence 2019*, vol. 11670, pp. 121–134, Cuvu, Fiji, Auguest 2019.

 [9] X. Zhang, F.-L. Chung, and S. Wang, "An interpretable fuzzy DBN-based classifier for indoor user movement prediction in ambient assisted living applications," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 42–53, 2020.

[10] X. Yang, Y. Dong, and J. Li, "Review of data features-based music emotion recognition methods," *Multimedia Systems*, vol. 24, no. 4, pp. 365–389, 2018.

[11] M. Grimaldi and P. Cunningham, "Experimenting with Music Taste Prediction by User profiling," in *Proceedings of the Acm Sigmm International Workshop on Multimedia Information Retrieval*, pp. 1733–2180, DBLP, New York, NY, USA, October 2004.

[12] Y. H. Yang, Y. C. Lin, and H. Chen, "Personalized music emotion recognition via model adaptation," in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pp. 748-749, Boston, MA, USA, July 2009.

[13] D. Su and P. Fung, "Personalized music emotion recognition via active learning," in *Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies*, pp. 57–62, New York, NY, USA, November 2012.

[14] V. Gupta, M. D. Chopda, and R. B. Pachori, "Cross-subject emotion recognition using flexible analytic wavelet transform from EEG signals," *IEEE Sensors Journal*, vol. 19, no. 6, pp. 2266–2274, 2019.

[15] C.-C. Yeh, S.-S. Tseng, P.-C. Tsai, and J.-F. Weng, "Building a personalized music emotion prediction system," *Advances in Multimedia Information Processing - PCM 2006*, Springer, Berlin, Germany, pp. 730–739, 2006.

[16] S. Hizlisoy, S. Yildirim, and Z. Tufekci, "Music emotion recognition using convolutional long short term memory deep neural networks," *Engineering Science and Technology an International Journal*, vol. 24, no. 3, pp. 760–767, 2020.

[17] H. Ma, H. Yang, M. R. Lyu, and I. King, "Sorec: social recommendation using probabilistic matrix factorization," in *Proceedings of the 17th ACM conference on information and knowledge management. ACM*, pp. 931–940, Napa valley, CA, USA, October 2008.

[18] J. Bu, S. Tan, C. Chen et al., "Music recommendation by unified hypergraph: combining social media information and music content," in *Proceedings of the international conference on Multimedia. ACM*, pp. 391–400, Firenze, Italy, October 2010.

[19] J. Golbeck and J. Hendler, "Movie recommendations using trust in web-based social networks,"vol. 96, pp. 282–286, in *Proceedings of the Consumer Communications & Networking conference*, vol. 96, pp. 282–286, IEEE, Las Vegas, NV, USA, January 2006.

[20] R. Sutcliffe, P. G. Rendell, J. D. Henry, P. E. Bailey, and T. Ruffman, "Music to my ears: age-related decline in musical and facial emotion recognition," *Psychology and Aging*, vol. 32, no. 8, pp. 698–709, 2018.

[21] J. Liu, N. Q. Wu, Y. Qiao, and Z. Li, "Short-Term traffic flow forecasting using ensemble approach based on deep Belief networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 404–417, 2020.

[22] Z. Tang, Y. Wang, Y. He, X. Wu, and S. Cao, "Modeling of boiler-turbine unit with two-phase feature selection and deep Belief network," *Journal of Chemical Engineering of Japan*, vol. 51, no. 10, pp. 865–873, 2018.

[23] B. T. Pham, D. Tien Bui, and I. Prakash, "Bagging based Support Vector Machines for spatial prediction of landslides," *Environmental Earth Sciences*, vol. 77, no. 4, pp. 146–153, 2018.

[24] S. Sun, X. Xie, and C. Dong, "Multiview learning with generalized eigenvalue proximal support vector machines," *IEEE Transactions on Cybernetics*, vol. 49, no. 2, pp. 688–697, 2019.

[25] M. Defferrard, K. Benzi, P. Vandergheynst, and X. Bresson, "FMA: a dataset for music analysis," *Computer Science*, vol. 67, no. 5, pp. 386–370, 2016.

[26] M. Jakubec and M. Chmulik, "Automatic music genre recognition for in-car infotainment," *Transportation Research Procedia*, vol. 40, no. 1, pp. 1364–1371, 2019.

[27] Z. Zhao, Y. Zhao, Z. Bao, H. Wang, Z. Zhang, and C. Li, "Deep Spectrum Feature Representations for Speech Emotion Recognition," in *Proceedings of the Joint Workshop of the 4th Workshop on Affective Social Multimedia Computing and first Multi-Modal Affective Computing of Large-Scale Multimedia Data*, pp. 27–33, Seoul, Republic of Korea, October 2018.

[28] W. Fang, S. H. Zhong, J. Peng, J. Jiang, and Y. Liu, "Data Augmentation for EEG-Based Emotion Recognition with Deep Convolutional Neural Networks," in *Proceedings of the International Conference on Multimedia Modeling*, pp. 82–93, Springer, Bangkok, Thailand, February 2018.