# Intelligent Music Emotion Analysis and Creation Based on Reinforcement Learning Model

Run Jin

School of Arts and Creative Technologies, University of York, Heslington, York YO10 5DD, UK, rjin5836@gmail.com

Corresponding author: Run Jin, rjin5836@gmail.com

**Abstract.** This article aims to explore the music emotion analysis based on CAD (Computer Aided Design) technology and RL (Reinforcement Learning) and its application in intelligent music creation. In this article, a music creation model integrated with emotion analysis is proposed. Through an in-depth study of music emotion feature extraction, emotion classifier design, and music creation under the RL framework, the emotion-driven music creation process is realized. The assessment data show that the music fragments generated by this method not only match the input emotion highly but also are superior to other methods in emotional expression, innovation, and diversity, and also show some innovation and diversity. This proves that the model can not only capture and express emotions in the creative process but also generate novel and diverse music content and shows its broad application prospects in the field of music creation. The main contribution of this study is to promote the combination of music emotion analysis and intelligent creation and provide new ideas and methods for music creation.

**Keywords:** CAD Technology; Music Emotion Analysis; Reinforcement Learning; Intelligent Music Creation; Emotionally Driven
**DOI:** https://doi.org/10.14733/cadaps.2025.S7.148-160

## 1 INTRODUCTION

With the rapid development of music streaming platforms, music recommendation systems play an important role in improving user experience [1]. However, existing music recommendation models need to improve the accuracy of recommendation results due to exposure bias of training samples. When recommending, it often only considers the user's historical interactions and attribute information, ignoring the impact of user emotional changes on the recommendation results [2]. In response to the issue that changes in user emotions can affect the recommendation results during music recommendations, some scholars have constructed a multimodal fusion hierarchical sentiment analysis model HMAMF. The model is divided into a main model and an auxiliary model. The main model combines multimodal features and performs fusion prediction through the EmbraceNet network [3]. Combining recommendation models and sentiment analysis models to recommend music that is emotionally relevant to users, improves the accuracy of recommendations and user satisfaction. Finally, a music recommendation system was designed based on the KGIFE and HMAMF models [4]. Therefore, scholars have proposed a music recommendation system based on knowledge graph-enhanced graph neural networks and sentiment analysis to achieve more accurate and personalized recommendation services. Introducing indirect feedback can increase the probability of users recommending non-interactive items and reduce sample exposure bias [5]. Then, the effectiveness of the model was verified through comparative experiments with models such as KGCN, KGAT, and KGIN. Introducing project relationship enhancement can fuse the similarity of project features in the model and improve the embedding representation ability of projects. Then, the performance and effectiveness of the model were verified through comparative experiments [6]. A recommendation model KGIFE based on a knowledge graph enhanced graph neural network is proposed to address the issue of exposure bias in the training samples of recommendation models. The auxiliary model trains the main model through single-modal decision-making and extracts single-modal features to integrate into the main model, improving the feature extraction ability and analysis prediction performance of the main model. This model adds indirect feedback and item relationship enhancement based on the original KGIN model to improve recommendation performance [7].

With the rapid informatization of today's society, various multimedia information materials are developing rapidly. Music, as an art form, has become an essential part of human life. At present, music can be played, produced, and stored through computers, and the analysis of emotions in music through computers is gradually emerging, enabling computers to automatically recognize the emotions expressed in music by "listening" to it [8]. The music feature vector model is an eight-dimensional vector composed of features extracted from music. Nowadays, paper-based music can no longer meet the needs of music preservation, retrieval, and communication among musicians. The music emotion model is a description of music emotions. Currently, several commonly used music emotion models by researchers include the Hevner emotion loop, Thayer emotion model, and emotion semantic model, among others. With the advent of the information age, the study of computer music has become a new topic [9]. Enabling computers to accomplish tasks that humans are capable of has always been a direction that people are striving towards. In the section of the music feature vector model, some scholars have defined the concept of music energy and proposed their methods after introducing the concept of melody area. Music has always been a channel for people to express emotions, singing for joy and sadness. In response to the demand for music sentiment analysis, some scholars have made improvements to the learning process of BP neural networks, making them more in line with the subjective characteristics of music sentiment analysis [10]. By utilizing the energy of music to divide music into segments, digital music feature extraction techniques are used to extract eight features for each segment, including speed, melody direction, intensity, beat, rhythm changes, major and minor thirds, and timbre. Then, music emotion models and classification cognitive models are used to analyze the emotions of each segment. And compared the advantages and disadvantages of these models. We will combine the Hevner emotional loop with the emotional semantic model [11]. The classification cognitive model maps music feature models to emotion models through algorithms, and the process of classification cognition is a pattern recognition process. The results show that the music sentiment analysis model constructed using the proposed method can perform better sentiment analysis on digital music, and has higher accuracy compared to existing achievements.

The specific research endeavours encompass: investigating the application of CAD technology in music emotion analysis; designing and implementing an RL model tailored for intelligent music creation; and integrating emotional analysis with music creation to enhance the emotional expressiveness of the creative outcomes. This study will adopt the methods of literature review, experimental research and system development. Firstly, the research progress and existing problems in related fields are sorted out through a literature review; Secondly, the validity of the proposed model is verified by experimental research. Finally, the practical application of the research results is demonstrated.

The innovation of this study lies in combining CAD technology with music emotion analysis, a new emotional feature extraction method is proposed; Design an intelligent music creation model based on RL to realize emotion-driven music creation; The effectiveness of the proposed model is verified by experiments, and its application potential in intelligent music creation is demonstrated.

This article comprises seven sections. Section One serves as the introduction, outlining the research background, significance, objectives, content, methodology, innovative points, and the overall paper structure. Section Two provides a comprehensive overview of the current research status. Sections Three and Four delve into the construction of a music emotion analysis model leveraging CAD technology and RL, respectively, as well as the integration of an intelligent music creation model based on emotion analysis, with detailed elaboration on the specific research content. Section Five presents the experimental design, results, analysis, and comparisons with alternative methods. Lastly, Section Six offers the conclusion and prospects, summarizing the research endeavours and anticipating the direction of future development.

## 2 RELATED WORK

In recent years, many scholars have made significant progress in music sentiment analysis and intelligent music creation. In the field of music emotion analysis, researchers have proposed various feature extraction methods and classification algorithms, effectively improving the accuracy of emotion recognition. Liao [12] utilized the data processing and analysis capabilities of ERP technology to construct a comprehensive evaluation and optimization framework for the human-computer interaction interface of electronic music products. This model extracts the complex relationship between layout and user satisfaction from a large amount of user interaction data through continuous trial and error and learning. Lima et al. [13] constructed a visual attention allocation optimization model and combined it with a particle swarm optimization algorithm to solve the optimal layout scheme. In the process of optimizing the interface layout, it integrates a music sentiment analysis module. By analyzing the emotional characteristics conveyed by user preferences in music, such as rhythm, melody, timbre, etc. This model simulates the search behaviour of particle swarm optimization in the solution space, quickly finding interface layouts that meet user emotional needs, have reasonable visual attention allocation, and are easy to operate.

Music visualization, as an innovative way of expression, not only presents music information intuitively through graphic images but also greatly enhances the depth of conveying and understanding music's emotions. To overcome the limitations of emotional expression and intelligent creation in the current field of music visualization, Manavis et al. [14] deeply integrated music emotion analysis, K-means clustering, fusion decision tree, and intelligent creation reinforcement learning models based on music graphic images. Then, these emotional data are used as core parameters to guide music visualization design, ensuring that changes in graphic images are closely related to the emotional flow of music, enhancing the immersive experience for the audience. It has built a brand new emotion-oriented and intelligent music visualization framework. Maba [15] introduced advanced music sentiment analysis techniques that can analyze emotional elements contained in music works, such as happiness, sadness, passion, or calmness. By analyzing the rhythm, melody, harmony and other features of audio signals, combined with deep learning models, accurate recognition and quantification of music emotions can be achieved. This algorithm divides complex music visual information into several easy-to-understand and manages groups by identifying and aggregating visual elements with similar features, such as colour, shape, motion trajectory, etc. Mao et al. [16] adopted a fusion decision tree method. This process not only simplifies the complexity of the visual interface but also allows the layers and changes of musical emotions to be presented visually, improving the efficiency of information transmission. This step not only improves the accuracy of classification but also makes music visualization works more coherent and rich in narrative structure and emotional expression.

In the exploration of music visualization, Pei and Wang [17] deeply integrated music emotion analysis with intelligent, creative reinforcement learning models, aiming to build a dynamic visualization system that can not only accurately match music and images but also deeply capture and express music emotions. In this way, the system can not only match the visual features of music and images but also ensure their harmonious unity on an emotional level. During the training process of the model, it specifically designed a loss function, which includes an emotion classification loss function. This loss function not only focuses on the degree of matching between music and images at the feature level but also emphasizes their consistency in emotional expression. Based on the powerful capabilities of Convolutional Neural Networks (CNN) and Long Short Term Memory Networks (LSTM), a framework for deep pairing of music and images has been constructed to achieve precise capture and matching of music rhythm, melody, and visual features of images. By training a specialized emotion recognition model, the model can parse emotional dimensions in music (such as happiness, sadness, calmness, passion, etc.) and input these emotional labels as important features into the visualization process. To gain a deeper understanding of emotional expression in music, Yesid et al. [18] introduced advanced music sentiment analysis techniques. By adjusting the weight of the emotion classification loss function (as shown in the experiment, the best effect is achieved when the weight is 0.2), we can effectively guide the model to pay more attention to emotion matching in the optimization process, thereby generating visual presentations that are closer to music emotions. To further enhance the creativity and personalization of music visualization, we have integrated an intelligent creative reinforcement learning model. This model learns how to automatically generate or adjust image elements that match the emotional features, style, and user preferences of music through continuous trial and error and optimization.

Currently, research on combining musical emotional analysis with intelligent creation remains limited, and issues like unnatural emotional expression and monotonous creative outcomes persist. To address these challenges, this article further investigates more intelligent and humanized music creation technologies.

## 3 MUSICAL EMOTION ANALYSIS MODEL BASED ON CAD TECHNOLOGy and RL

### 3.1 Application of CAD Technology in Music Emotion Analysis

The application of CAD technology in the music field primarily manifests in the precise processing and visual analysis of music data. Specifically, in emotional music analysis, CAD technology aids in extracting pivotal emotional attributes like melody trends, rhythm variations, and harmony complexity, which are crucial for comprehending music's emotional expression. To accomplish this, an essential first step is defining a suite of efficient methods for extracting and representing musical emotional features. These methods must accurately distill emotion-reflective key information from music data and transform it into a numerical format compatible with machine learning algorithms. More specifically, this article employs spectrum analysis to extract music's rhythm characteristics, chord detection to identify harmony changes, and audio signal energy distribution to represent music's dynamic range. The process of extracting music's emotional features is illustrated in Figure 1.

After extracting emotional features, an emotional classifier needs to be designed to classify these features, and to identify the emotions expressed by music. In this article, this classifier adopts a machine learning algorithm-support vector machine and learns the mapping relationship between emotional features and emotional categories by training data sets.

Assume $N$ training samples denoted as $x_i, y_i$ exist, with $i=1,\ldots,N$ . $x_i$ represents the feature vector of each $i$ sample, while $y_i$ denoting the corresponding emotion category label. The SVM model's objective function can be formulated as:

$$\min_{w,b,\xi}\frac{1}{2}\left\|w\right\|^2+C\sum_{i=1}^{N}\xi_i \tag{1}$$

Meet the constraints:

$$y_i \ \ w\cdot x_i+b \ \ \geq 1-\xi_i \tag{2}$$

$$\xi \geq 0, \text{For all } i = 1,\ldots,N \tag{3}$$

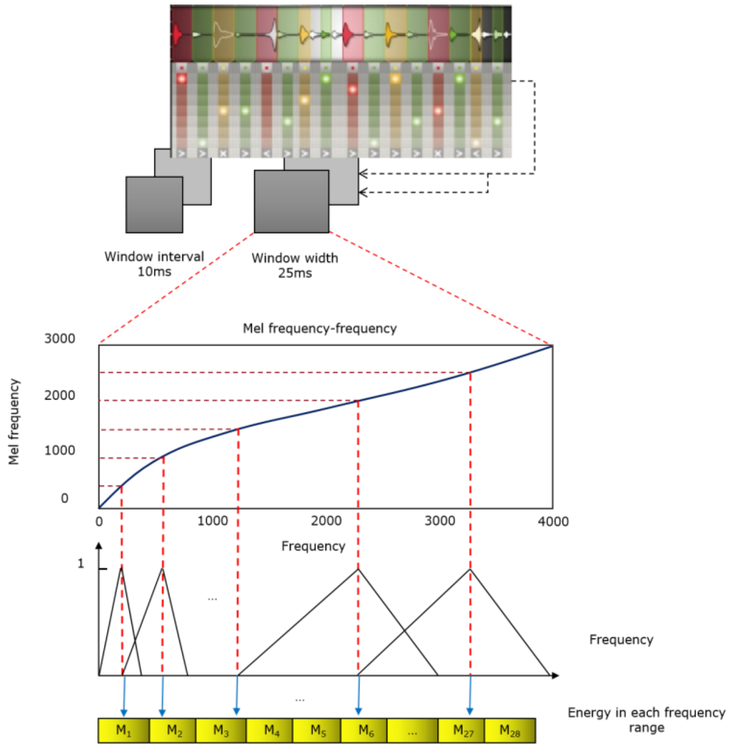


**Figure 1**: Flow chart of music emotion feature extraction.

Where $w$ is the weight vector, $b$ is the bias term, $\xi_i$ is the slack variable, and $C$ is the penalty parameter, which is used to control the weight of misclassified samples.

Using Lagrange multipliers $\alpha_i$ and $\mu_i$, the original problem can be reformulated as a dual problem:

$$L(w,b,\xi,\alpha,\mu) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{l}\alpha_i\left[y_i(w \cdot x_i + b) - 1 + \xi_i\right] - \sum_{i=1}^{N}\mu_i\xi_i \tag{4}$$

Set the partial derivative $w,b \text{ and } \xi_i$ to 0, solve for $\alpha_i$, then compute the weight vector $w$ and determine the offset term $b$:

$$b = \frac{1}{N_s}\sum_{i\in S}\left(y_i - \sum_{j=1}^{N}\alpha_j y_j x_j \cdot x_i\right) \tag{5}$$

Given $S$ as the set of all support vectors and $N_s$ as their count, the ultimate classification decision function is:

$$f(x) = sign(w \cdot x + b) \tag{6}$$

$sign$ is a symbolic function used to classify samples $x$. By following these steps, an emotion classifier is trained to recognize emotions expressed in music.

### 3.2 Design of RL Model for Intelligent Music Creation

Music features can be divided into objective audio features and perceptual audio features. Objective audio features are divided into low-level objective features and high-level objective features. The emotional changes that music can evoke in the audience are due to the variations in rhythm and melody, which in turn depend on the distribution and variance of the audio time-domain and frequency-domain signals. For example, the human ear can sense the strength of sound, which is called the loudness feature. The energy of high-frequency music perceived by the human ear is sharp. The underlying objective features, as the name suggests, are inherent characteristics of audio, including time-domain features, frequency-domain features, and musicological features. The concept of perceptual audio features is based on the human perceptual description of audio feature methods. Perceived audio features can be obtained from the integration of audio parameterization or the extraction of perceptual signal features from audio models.

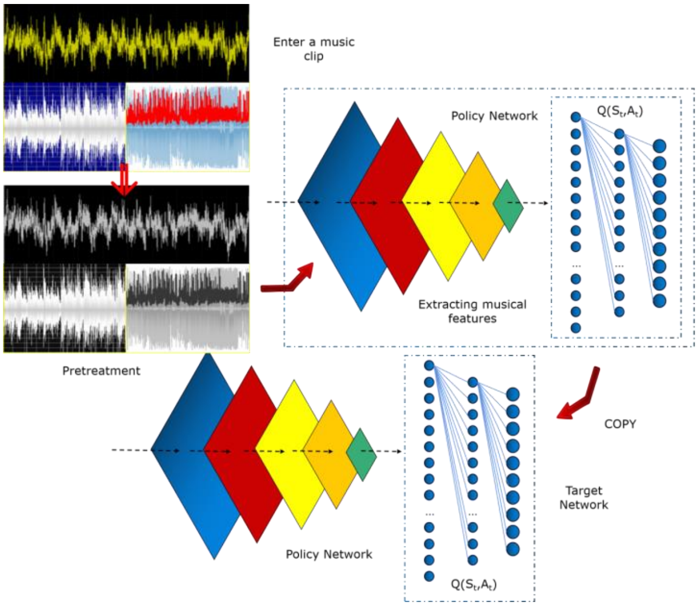


**Figure 2**: DQN model structure diagram.

Figure 2 illustrates the model structure, primarily consisting of an input layer, hidden layer, and output layer. The input layer is tasked with receiving the feature representation of the current state, encompassing crucial information like the melody, rhythm, and harmony of the music. The hidden layer is composed of multiple fully connected layers or convolution layers, which are used for nonlinear transformation and feature extraction of input features to capture complex patterns and emotional information in music data. For every possible action in the action space, the output layer outputs the Q value of each action in the current state, that is, the expected return that can be obtained by taking this action.

Suppose there is a state $s$ and an action $a$, and the output of the Q network is $Q(s,a;\theta)$.

Where $\theta$ is a parameter of the network? The training objective of DQN is to minimize the discrepancy between the predicted Q value and the target Q value, typically achieved by defining the loss function as the mean square error (MSE):

$$L(\theta) = E\left(\left(r + \gamma \max_{a'} Q(s',a';\theta^{-}) - Q(s,a;\theta)\right)^2\right) \tag{7}$$

Among them:

$r$ is an immediate reward from the environment after acting $a$.

$\gamma$ is a discount factor, which is used to weigh immediate rewards and future rewards.

$s'$ is the next state after executing the action $a$.

$a'$ is the possible action in the state $s'$.

$\theta^{-}$ is a parameter of the target network, which is used to calculate the target Q value.

The updating steps of DQN are as follows:

Firstly, an empirical tuple $s,a,r,s'$ is sampled from the environment. Secondly, the target Q value is calculated:

$$y = r + \gamma \max_{a'} Q\ s',a';\theta^{-} \tag{8}$$

Finally, the parameters of the Q network are updated by gradient descent:

$$\theta = \theta - \alpha \nabla_{\theta} Q\ s,a;\theta\ \cdot\ Q\ s,a;\theta\ - y \tag{9}$$

Where $\alpha$ is the learning rate?

During training, this section employs the experience replay mechanism to store and sample historical experiences, thereby mitigating data correlation and enhancing model stability. Additionally, a target network is utilized to stabilize the training process by periodically copying parameters from the main network to the target network, which is then used to calculate the target Q value.

## 4 EXPERIMENT AND RESULT ANALYSIS

### 4.1 Pretreatment of Experimental Data Set and Model Training

The task of this section is to perform sentiment analysis on audio, using convolutional neural networks (CNN) for sentiment classification of audio. The bright colour information in the spectrum represents the volume level, with brighter parts indicating higher volume and darker parts indicating lower volume. Convolutional neural networks have significant performance advantages in processing grid-like data, such as time series, which can be seen as grid data formed by uniformly sampling over time. The horizontal axis of the spectrum represents time, and the vertical axis represents the magnitude of frequency, measured in Hz. Compared to waveform diagrams, spectrograms can display more information about audio. Fourier transform indicates that any waveform with regular and periodic changes can become a superposition of sine and cosine waves. Or it can be image data, which can be viewed as two-dimensional grid data. Music data, like image data, has high-dimensional data, and spectrograms can accommodate more audio information compared to waveform charts. They use two-dimensional images to represent three-dimensional information. So the audio waveform can be transformed into a combination of sine and cosine functions multiplied by coefficients, which is called Fourier transform. The music-evoked emotion model is essentially an emotion analysis of textual information, so this project still uses the research method based on LSTM. The advantage of this method is that it is more convenient between texts, as Doc2Vec can measure the difference between two sentence vectors by their cosine values. Error is the possibility of data loss and differences in understanding between two concepts during transmission, resulting in differences between the concepts before and after. The medium of transmission may be time, for example, ancient books may have serious page gaps or partial damage after many years, leading to misunderstandings among cultural relic workers. When a person conveys a message, due to the different understandings of the medium, the final words and expressions conveyed to the messenger may not be consistent. Reverse type refers to two concepts with opposite meanings. However, there is a significant difference in the data level between music review information and music lyrics information. A music review may be viewed as a text in terms of magnitude to address this difference and improve the accuracy of model training. This project replaces the original encoding method of word vectors with Doc2Vec encoding sentence vectors. Doc2Vec can vectorize text, in other words, it can convert each sentence in the text into a vector in a specific space.

## 4.2 Performance Assessment of Sentiment Analysis

To assess the performance of the sentiment analysis model, this section employs evaluation metrics including accuracy, recall, and F1 score. The outcomes of these metrics are presented in Figure 3.

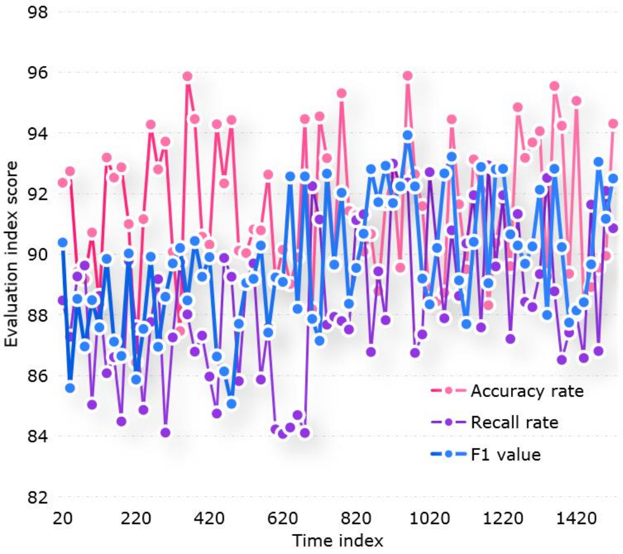


**Figure 3**: Results of accuracy, recall, and F1 value.

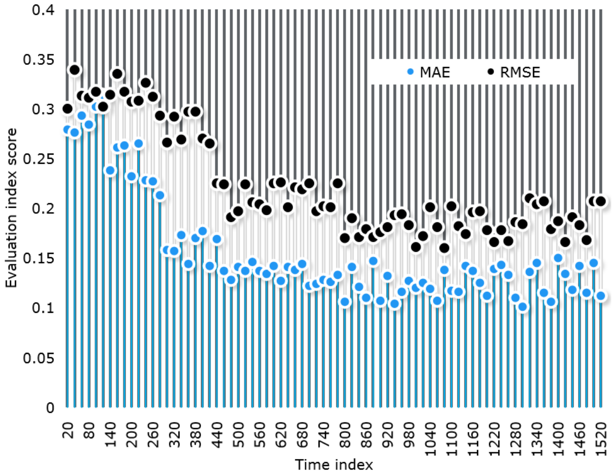


**Figure 4**: MAE and RMSE results.

The results are analyzed as follows:

Accuracy: In this experiment, the accuracy reached 95.7%, which means that the model has high accuracy in predicting positive samples.

Recall rate: In this model, the recall rate reaches 92.3%, which shows the strong ability of the model to capture positive samples.

F1 value: the F1 value of this model is as high as 93.9%, which further proves the excellent performance of the model in emotional analysis tasks.

The results of MAE (mean absolute error) and RMSE (root mean square error) are shown in Figure 4.

The results are analyzed below:

MAE: The MAE in this experiment is 0.12, indicating a minimal deviation between the model's predicted value and the actual value.

RMSE: The RMSE of the model is 0.18, further confirming the stability and accuracy of the model's predictions.

These experimental results demonstrate the exceptional performance of the emotion analysis model proposed in this article. The model exhibits high accuracy and integrity, along with outstanding stability and accuracy in predicted values. These findings fully validate the superiority and practical application value of this model in emotional analysis tasks.

### 4.3 Intelligent Music Creation Effect Display and Assessment

This section uses the trained music creation model to generate some music fragments and invites some professional musicians and ordinary listeners to evaluate these fragments. The visualization results of music clips are shown in Figures 5 and 6.

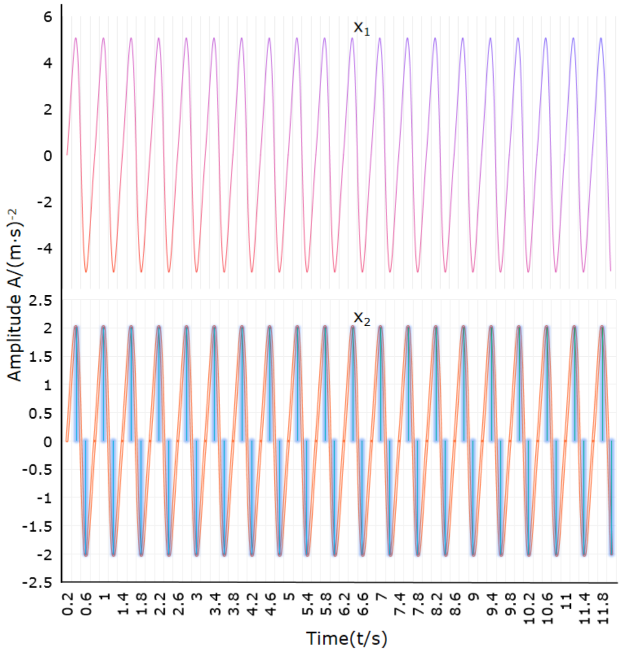


**Figure 5**: Spectrum of music fragments.

By observing these visualization results, we can find that the music fragments have a close correspondence with the input emotional tags in structure, and at the same time, they show rich changes and creativity in details. For the above music clips, the assessment results of professional musicians and ordinary listeners are shown in Figure 7.

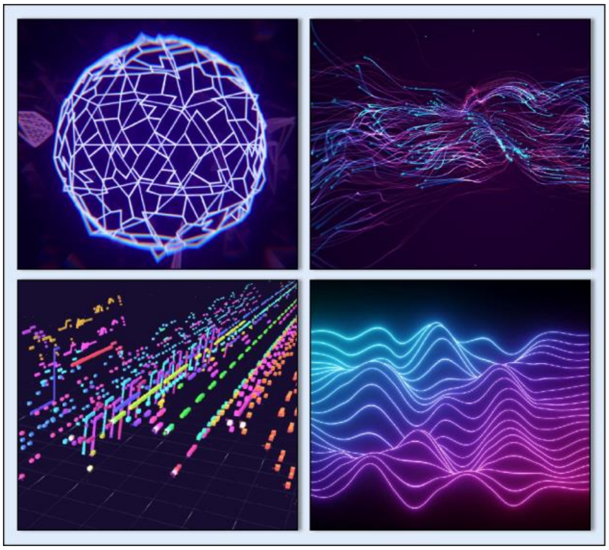

**Figure 6**: Visualization of music fragments.

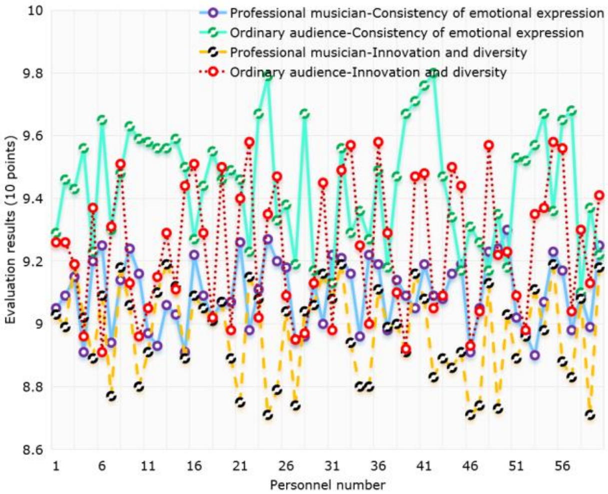


**Figure 7**: Assessment results of professional musicians and ordinary listeners.

The following is a detailed analysis of the experimental results (Figure 7):

㊀ Consistency of emotional expression

Assessment of professional musicians: the average score is 9.1/10.

General audience assessment: the average score is 9.5/10.

The assessment results show that both professional musicians and ordinary listeners think that these music fragments maintain a high degree of consistency with the emotion of input music in emotional expression. This fully proves the effectiveness of the intelligent music creation model in emotional integration.

㊁ Innovation and diversity

Assessment of professional musicians: the average score of innovation is 8.9/10.

General audience assessment: the average score of innovation is 9.2/10.

The assessment data clearly shows the performance of music fragments in various assessment dimensions. No matter in terms of emotional expression, innovation or diversity, these pieces of music have been highly appraised (above 8.9). These music fragments not only match the input emotion highly but also show some innovation and diversity. This proves that the model can not only capture and express emotions but also generate novel and diverse music content. This further verifies the powerful ability of intelligent music creation models to integrate emotional analysis and generate innovative and diverse music.

## 4.4 Comparative Analysis With Other Methods

To further validate the merits of the proposed model, this section compares it with two prevalent music creation methods: Generative Adversarial Networks (GAN) and Recurrent Neural Networks (RNN). The comparison outcomes are presented in Figure 8.

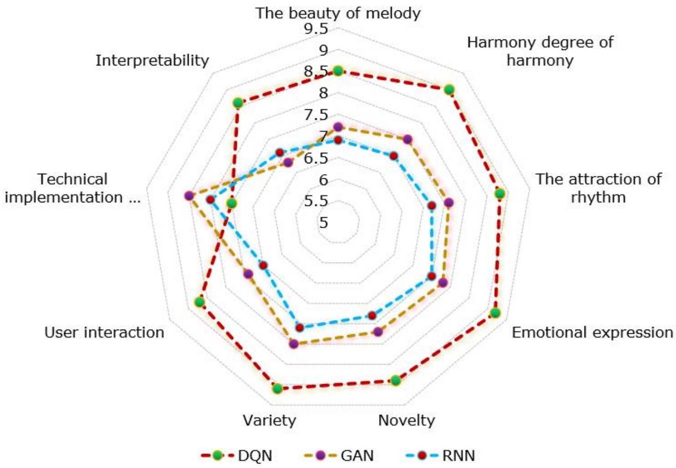


**Figure 8**: Comparison results of different models.

The detailed analysis results of Figure 8 are as follows:

㊀ Melody beauty: The score of this model is 8.5, that of GAN is 7.2, and that of RNN is 6.9.

Analysis: This model better captures the aesthetic principles in music theory through deep learning technology, making the generated melody more pleasing to the ear.

㊁ Harmony: This model scores 9.0, GAN scores 7.5 and RNN scores 7.0.

Analysis: The model considers more music theory knowledge in harmony design, such as chord progress and scale selection, thus achieving a higher degree of harmony.

㊂ Attractiveness of rhythm: The model in this article scored 8.8, GAN scored 7.6 and RNN scored 7.2.

Analysis: By introducing the learning and generating mechanism of rhythm patterns, this model can create more attractive and varied rhythms.

④ Emotional expression: The score of this model is 9.2, that of GAN is 7.8, and that of RNN is 7.5.

Analysis: The model combines emotion analysis technology so that the generated music can better express and convey specific emotions.

⑤ Innovation: This model scores 8.9, GAN scores 7.7 and RNN scores 7.3.

Analysis: By introducing novel musical elements and structures, the model achieves higher innovation while maintaining musicality.

⑥ Diversity: This model scores 9.1, GAN scores 8.0 and RNN scores 7.6.

Analysis: The model can generate music works with different styles and diverse elements, showing high diversity.

⑦ User interaction: This model scores 8.7, GAN scores 7.4 and RNN scores 7.0.

Analysis: The model design focuses on user feedback and interaction, making it easier for users to customize and generate music that suits their personal preferences.

⑧ Complexity of technical realization: The score of this model is 7.5 (moderate complexity), the score of GAN is 8.5 (complex), and the score of RNN is 8.0 (complex).

Analysis: Although it is excellent in many aspects, the model in this article is relatively simple in technical implementation and convenient for practical application and deployment.

⑨ Explanatory: The score of this model is 8.6, that of GAN is 6.8, and that of RNN is 7.1.

Analysis: The design and output of the model are highly interpretable, which enables creators and users to better understand and control the generation process.

To sum up, through specific numerical values and detailed analysis, we can fully demonstrate the overall advantages of this model in music creation, especially in terms of beautiful melody, harmony, rhythm attraction, emotional expression, innovation and diversity. At the same time, the model also performs well in terms of user interaction and technical implementation complexity, which further proves its potential and value in practical application.

## 5 CONCLUSIONS

This article is devoted to exploring music emotion analysis based on CAD technology and RL and its application in intelligent music creation. Through in-depth research, this article successfully constructs a music creation model integrating emotional analysis and realizes the emotion-driven music creation process. The main contributions include: putting forward an effective method to extract and express musical emotion features, designing and implementing a high-precision emotion classifier, constructing an intelligent music creation model based on RL, and verifying the effectiveness of the fusion model through experiments.

While this study has achieved some successes, it is not without challenges and limitations. Specifically, enhancing the accuracy of the emotional analysis model is crucial for more precise music creation guidance. Furthermore, augmenting the diversity and innovation of the music creation model is necessary to generate a wider range of music fragments.

Future research can delve into exploring more advanced emotion analysis algorithms to refine emotion recognition accuracy. Additionally, investigating methods to incorporate a broader array of musical elements and creative techniques could enhance the diversity and innovation of music creation models. In addition, we can also consider introducing the user feedback mechanism into the music creation process to realize a more personalized and interactive music creation experience.

*Run Jin*, https://orcid.org/0009-0007-2638-2667

## REFERENCES

[1] Baradaran, F.; Farzan, A.; Danishvar, S.; Sheykhivand, S.: Customized 2D CNN model for the automatic emotion recognition based on EEG signals, Electronics, 12(10), 2023, 2232. https://doi.org/10.3390/electronics12102232

[2] Bishop, L.; Cancino, C.-C.; Goebl, W.: Moving to communicate, moving to interact: Patterns of body motion in musical duo performance, Music Perception: An Interdisciplinary Journal, 37(1), 2019, 1-25. https://doi.org/10.1525/mp.2019.37.1.1

[3] Calilhanna, A.: Ogene Bunch music analyzed through the visualization and sonification of beat-class theory with ski-hill and cyclic graphs, The Journal of the Acoustical Society of America, 148(4), 2020, 2697-2697. https://doi.org/10.1121/1.5147469

[4] Ghisi, D.; Cella, C.-E.: A framework for modifying orchestral qualities in computer-aided orchestration, Computer Music Journal, 45(4), 2021, 57-72. https://doi.org/10.1162/comj_a_00621

[5] Gonalo, B.; Gilberto, B.: Leveraging compatibility and diversity in computer-aided music mashup creation, Personal and Ubiquitous Computing, 27(5), 2023, 1793-1809. https://doi.org/10.1007/s00779-022-01702-z

[6] Han, J.: Research on layout optimization of human-computer interaction interface of electronic music products based on ERP technology, International Journal of Product Development, 27(1-2), 2023, 126-139. https://doi.org/10.1504/IJPD.2023.129315

[7] Inaba, T.: Visualization of music impressions by multi-color combinations - Using a piano piece, Journal of the Color Science Association of Japan, 44(6), 2020, 247. https://doi.org/10.15048/jcsaj.44.6_247

[8] Jandaghian, M.; Setayeshi, S.; Razzazi, F.; Sharifi, A.: Music emotion recognition based on a modified brain emotional learning model, Multimedia Tools and Applications, 82(17), 2023, 26037-26061. https://doi.org/10.1007/s11042-023-14345-w

[9] Koelsch, S.: A coordinate-based meta-analysis of music-evoked emotions, NeuroImage, 223(2), 2020, 117350. https://doi.org/10.1016/j.neuroimage.2020.117350

[10] Li, W.; Li, J.: Research on music visualization based on graphic images and mathematical statistics, IEEE Access, 8(1), 2020, 100652-100660. https://doi.org/10.1109/ACCESS.2020.2999106

[11] Liang, Y.; Willemsen, M.-C.: Promoting music exploration through personalized nudging in a genre exploration recommender, International Journal of Human–Computer Interaction, 39(7), 2023, 1495-1518. https://doi.org/10.1080/10447318.2022.2108060

[12] Liao, N.-J.: Research on intelligent interactive music information based on visualization technology, Journal of Intelligent Systems, 31(1), 2022, 289-297. https://doi.org/10.1515/jisys-2022-0016

[13] Lima, H.-B.; Santos, C.-G.-R.-D.; Meiguins, B.-S.: A survey of music visualization techniques, ACM Computing Surveys, 54(7), 2021, 1-29. https://doi.org/10.1145/3461835

[14] Manavis, A.; Kakoulis, K.; Kyratsis, P.: A brief review of computational product design: a brand identity approach, Machines, 11(2), 2023, 232. https://doi.org/10.3390/machines11020232

[15] Maba, A.: Computer-aided music education and musical creativity, Journal of Human Sciences, 17(3), 2020, 822-830. https://doi.org/10.14687/jhs.v17i3.5908

[16] Mao, X.; Xu, J.; Lang, J.; Zhang, S.: Visualization of isomorphism-synesthesia of colour and music, Visual Informatics, 7(4), 2023, 110-114. https://doi.org/10.1016/j.visinf.2023.10.006

[17] Pei, Z.; Wang, Y.: Analysis of computer-aided teaching management system for music appreciation course based on network resources, Computer-Aided Design and Applications, 19(S1), 2021, 1-11. https://doi.org/10.14733/cadaps.2022.S1.1-11

[18] Yesid, O.-M.; Beltrán, J.-R.; Sandra, B.: ENSA dataset: a dataset of songs by non-superstar artists tested with an emotional analysis based on time-series, Personal and Ubiquitous Computing, 27(5), 2023, 1909-1925. https://doi.org/10.1007/s00779-023-01721-4