Analysis and Application of Big Data-Driven Visual Scene in Music Teaching

Dongxuan Wu[1], Cheng Zhou[2] and Xiaojing Wen[3]

[1,2,3] Hanshan Normal University, Chaozhou, GuangDong 521041, China, [1]wdx@hstc.edu.cn, [2]zc@hstc.edu.cn, [3]2428@hstc.edu.cn

Corresponding author: Dongxuan Wu, wdx@hstc.edu.cn

Abstract. Traditional music teaching is mainly dependent on the teacher's explanation, with more emphasis on the teaching of theoretical knowledge. The lack of practical opportunities causes many students in the process of learning to face a lot of difficulties. Therefore, this paper constructs a visual model of music teaching based on big data and random forest algorithm and improves the quality of music emotion classification by extracting and preprocessing the basic elements of music. At the same time, in order to improve the performance of the random forest algorithm, this paper introduces an ant colony algorithm for optimization to improve the performance of music emotion classification and provide reliable data support for music visualization. The experimental results show that compared with other algorithms, the proposed model shows good accuracy and stability of emotion classification and can basically meet the needs of applications. In the comparative experiment, the model presented in this paper can present the music knowledge and structure in a more intuitive visual way, help students understand the inner emotion of music at multiple levels, and enable them to better use the corresponding music knowledge in practice, so as to achieve the purpose of greatly improving the score of music test.

Keywords: Big Data; Music Teaching; Visualization; Random Forest; Optimize Ant Colony Algorithm
DOI: https://doi.org/10.14733/cadaps.2025.S9.264-278

1 INTRODUCTION

Music teaching is a process that needs to mobilize auditory, visual, tactile, and other sensory organs to achieve resonance with music content and then deconstruct music through professional knowledge so that students can re-experience the content and charm of music in combination with professional and emotional feelings [1]. Traditional music teaching often adopts the infusing teaching method in which teachers explain and students listen, focusing on the single sensory experience of hearing, ignoring the role of other senses, such as vision and touch, in music learning [2]. At the same time, in the teaching process, teachers may fail to fully create a situation that matches the music works, which makes it difficult for students to deeply understand the emotions and connotations of music [3]. Teaching without context makes it difficult for students to relate music to real life or cultural background, reducing the perceived effect of music [4]. The lack of diversified and innovative teaching methods also fails to fully demonstrate the charm and characteristics of music, making it difficult for students to generate strong resonance and interest [5]. In addition, traditional music teaching is based on a theoretical basis and lacks emphasis on music practice. Due to the lack of sufficient practical experience, it is difficult for students to convert the theoretical knowledge they have learned into actual musical performance ability, and many low-level problems cannot be solved in practice [6]. Therefore, how to improve students' perception and resonance of music content has become an important research direction in music teaching reform.

The development of big data technology and computer visualization technology has provided a new technical direction for education and teaching reform and achieved certain teaching results. Some researchers have applied big data technology to mathematics teaching, extracted the characteristic data of different students' mathematics learning through classification algorithms and association algorithms, and realized personalized recommendations of mathematics exercises for students in different learning states [7]. Some educators have also introduced virtual reality technology into art teaching, which presents the full picture of traditional artworks for students in multiple dimensions through AR, VR, and other technologies so that students can feel the charm of art in an immersive environment [8]. Some researchers combined neural networks and virtual reality technology with building an English dialogue environment for students, identifying the accuracy of students' English dialogue pronunciation and the correctness of grammar through algorithms, and enable students to deepen their understanding of English knowledge and improve their ability to use knowledge in practice. It can be seen that in the field of music teaching, the application of big data and computer vision technology can present complex music information in graphic, image, animation, and other intuitive and understandable forms, such as waveform diagrams, spectral diagrams, musical note animation, etc., to help students better understand the internal structure and emotional expression of music works. Through visualization, students can feel the rhythm, melody, harmony, and other elements of music more intuitively, thus deepening their understanding and perception of musical works [9]. In addition, in the explanation of music emotion, visualization technology can present music-related pictures for students based on big data analysis and visually deepen students' understanding and resonance of music emotion. Therefore, this paper builds a visualization model of music teaching driven by big data and combined with computer vision technology, realizes the recognition of music emotion through random forest technology, and lays a solid foundation for the visualization of music teaching.

With the development of big data technology, the music industry has accumulated massive amounts of data, including audio files, user behaviour data, social media feedback, and more [10]. By combining the music emotion detection method in the research with visual scene analysis technology, it is possible to achieve comprehensive emotional analysis of music teaching videos or live broadcasts. This big data-driven pre-training strategy lays a solid foundation for the accuracy and generalization ability of music emotion recognition and also provides possibilities for the application of emotion analysis in music teaching [11]. This provides powerful feature extraction capabilities for relatively small but emotion-focused datasets like EmoMusic. By analyzing students' learning feedback (including emotional reactions, skill mastery, etc.), combined with the results of music emotion detection, teachers can tailor teaching content and practice tracks for each student to meet their learning needs and interests better. Some research institutes have adopted transfer learning methods that utilize pre-trained convolutional neural networks on the Million Song dataset, effectively utilizing the knowledge from these large-scale datasets. In music teaching, visual scenes (such as the performer's facial expressions, body movements, stage arrangements, etc.) are often closely linked to music emotions [12]. For example, by analyzing the subtle changes in the facial expressions of performers and combining them with the emotion detection results of music, it is possible to more accurately evaluate how performers convey musical emotions through nonverbal means, thereby guiding students to express themselves in performances better. Based on big data-driven music emotion recognition and visual scene analysis, personalized music teaching plans can be constructed. At the same time, through visual scene analysis, teachers can also observe students' body language and facial expression changes during the practice process, discover and correct bad habits in a timely manner, and improve teaching effectiveness.

2 RELATED WORK

The introduction of visual scenes makes this customization more intuitive and precise, and Pei and Wang [13] enhance the learning effect through specific visual elements in video tutorials, such as instrument-playing techniques. Big data technology can collect and analyze various data generated by students during the process of learning music, including teaching scenes recorded in videos, students' facial expressions, body language, and other visual information. Combining these visual data with audio data can construct a more comprehensive learning behaviour model, helping teachers understand students' understanding and perception of music. A highly immersive music learning environment can be created by integrating virtual reality (VR) or augmented reality (AR) technologies, combined with big data analysis and visual scene recognition. Ríos et al. [14] can deeply explore a large amount of student learning data by combining big data analysis and machine learning algorithms to identify the strengths and weaknesses of different students in music learning. Students can freely explore virtual music scenes, interact with virtual instruments, and even participate in virtual concerts to experience music works of different styles and emotions. This teaching method can greatly stimulate students' interest and participation in learning and improve teaching effectiveness. Music and visual arts have a natural connection in emotional expression. Big data-driven visual scene analysis can capture and quantify emotional elements in videos. For example, by analyzing students' facial expressions while listening to specific music clips, their mastery of music emotions can be evaluated, and teaching strategies can be adjusted accordingly. Based on these analysis results, Rocamora et al. [15] customized personalized learning paths for students and recommended music works and teaching methods that are suitable for their level and interests. This cross-modal sentiment analysis not only helps improve the accuracy of music classification but also guides students to deeply understand the emotional connotations of music works in music teaching, enhancing their aesthetic and emotional resonance abilities. Conduct correlation analysis with the emotional characteristics of music works, such as colour, composition, and character expressions.

Big data-driven visual scene analysis technology can also be combined with virtual reality (VR) and augmented reality (AR) technologies to create highly immersive music teaching environments. Big data technology enables us to capture and analyze rich visual information of students during the process of learning music, such as facial expressions, body language, and interactions with the environment. Sarkar et al. [16] analyzed students' facial expressions and body postures during performances to help determine their mastery and expression of musical emotions, which can be mutually confirmed with the results of audio analysis and improve the accuracy of emotion classification. Combining this information with emotional features in audio data can provide a more comprehensive understanding of emotions. In this environment, students can not only hear music but also visually perceive scenes and atmospheres that match the emotions of the music. Emotion classification technology based on big data and deep learning can identify the emotional response patterns of different students in music learning. By combining visual scene analysis, teachers can more accurately understand each student's emotional needs and learning difficulties, and then develop personalized teaching plans. For example, for students who exhibit confusion or lack emotional engagement in specific music segments, teachers can stimulate emotional resonance among students by adjusting teaching methods, introducing visual aids, or adjusting the learning environment. Real-time analysis of students' visual and auditory data during the learning process can provide teachers with immediate emotional feedback. This multi-sensory experience can help students gain a deeper understanding of music emotions and improve learning outcomes. For example, when the system detects that students lack emotional involvement in a certain music segment, it can automatically recommend relevant visual materials or practice methods to help students better understand and express musical emotions. Van et al. [17] used deep learning algorithms to automatically identify and evaluate students' emotional responses, providing teachers with suggestions for adjusting teaching strategies.

Wang et al. [18] proposed an innovative percussion rhythm analysis method that is not only based on information theory principles but also cleverly integrates big data-driven visual scene analysis techniques to deepen its application and understanding in music teaching. This strategy provides a new perspective for shot detection problems by using weighted features for lossy encoding. Within the framework of rate-distortion theory, they explored how to balance information retention and distortion levels while preserving critical rhythm information by adjusting data compression rates. Simultaneously allowing for moderate information distortion to adapt to the dynamic changes in actual music performances. In the experiment, Yang et al. [19] not only analyzed the rhythm features in audio data but also combined visual scene data generated by students during the learning process (such as facial expressions, body movements, etc.) to extract key information related to rhythm learning through big data analysis techniques. In addition, to comprehensively evaluate the practicality and value of this new method in the field of music teaching, we combined it with big data-driven visual scene analysis technology and selected the Candombe drum beat recording dataset as the experimental object. When processing recordings of percussion performances, the algorithm designed an optimized lossy representation method aimed at capturing the rich rhythmic patterns contained in the audio. This method not only improves the accuracy of rhythm analysis but also enhances its robustness in practical applications. Assuming that when the audio signal is correctly aligned with the preset rhythm pattern, the most concise and effective data representation can be generated. This result indicates that combining big data-driven visual scene analysis, and percussion rhythm analysis methods not only improves the accuracy of rhythm recognition but also provides richer and more dimensional feedback information for music teaching. Zhang [20] found that the proposed method can accurately reflect the style differences and skill levels among performers by comparing the overall complexity measurement standards of different percussion performances under the rate-distortion curve. The evaluation results are highly consistent with subjective judgments and professional review opinions.

3 CONSTRUCTION OF VISUAL MUSIC TEACHING MODEL BASED ON BIG DATA

Music art is essentially the art of hearing and sound, which is the most remarkable feature of music itself. In music teaching, it is necessary to pay close attention to the characteristics of music itself, that is, the essence of music. Whether it is appreciation teaching or singing teaching, the language of music should always be penetrated, including melody, speed, strength, harmony, and other elements. Therefore, listening to the sound is the main way of teaching music. Teachers should avoid using too much non-musical language to explain music but should guide students to feel and understand music through the music itself. In terms of teaching, music is different from other disciplines in that it places special emphasis on the importance of the teaching process. The teaching process is the purpose of teaching and is unified. It can be seen that music teaching itself is highly abstract and non-intuitive, which brings specific difficulties to students' perception and understanding, especially for young students or students with a weak music foundation; they may have difficulty in directly grasping the connotation and emotion of music. The results of relevant questionnaires show that students have difficulty learning the emotional understanding, timbre change, music beat, and work analysis of music in the music learning process, as shown in Figure 1.

Therefore, in order to improve the quality of music teaching, some researchers put forward the concept of audio-visual teaching, that is, the use of human visual and auditory perceptual knowledge to deepen understanding and improve the teaching effect of teaching and learning methods. It was developed in the 1960s with the application of teaching machines and calculators in teaching and has been continuously improved with the development of modern educational technology.

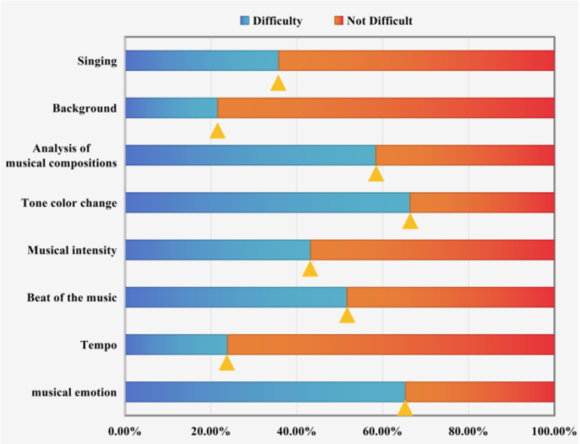

Figure 1: Students' difficulties in learning music.

In the audio-visual teaching theory, expert Edgar Dyer proposed the "tower of experience" theory, which classifies the basis of learning experience from concrete to abstract level, which is a teaching concept that deeply reflects the law of human cognitive development. This theoretical framework argues that the learning process should follow a gradual path, starting from intuitive, concrete experience and gradually climbing to more complex, abstract levels of understanding. Such a learning path is not only in line with the brain's natural way of processing information but also helps students build a solid and deep body of knowledge, as shown in Figure 2.

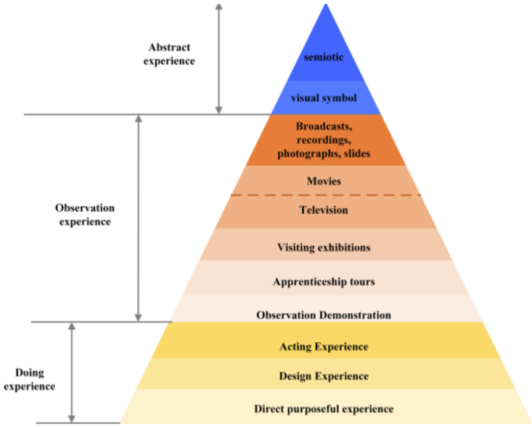

Figure 2: Theoretical diagram of "Tower of Experience."

In the field of music teaching, this theory is particularly important because it provides a solid theoretical basis for the implementation of visual strategies in music teaching. Visualization of music teaching, as an innovative teaching method, aims to enhance students' direct experience and perception ability through the integration of visual elements so that music concepts and emotions that may originally appear abstract and elusive become tangible and easy to understand.

The dual-channel hypothesis theory is one of the core theories of multimedia learning, which holds that the human information processing system consists of two independent and cooperative information processing channels: auditory/verbal channel and visual/image channel. This theory is an important theoretical basis for music teaching visualization. According to the dual channel hypothesis theory, information processing using both auditory and visual channels can significantly improve learning efficiency. In music teaching, combining abstract music concepts with intuitive visual images through visual means can help students receive and process information through auditory and visual channels at the same time, so as to improve the efficiency of information processing. Visualization of music teaching uses visual elements to present rhythm, melody, harmony and other elements of music works in an intuitive form, enabling students to experience and understand music through visual perception. This intuitive perception helps to stimulate students' interest and enthusiasm in learning and enhances their learning experience.

Visualization of music teaching is the construction of scenes that can help students intuitively understand the content of music teaching through computer technology such as big data, so as to improve students' sense of participation in the teaching process. Therefore, the presentation of visual teaching scenes is based on musical content and emotions. In order to build a better teaching visualization scene, the model in this paper first realizes the reception of music information source through sensor technology, extracts and preprocesses music emotion information, expression features, and other feature data during the pre-processing process, and then completes emotion recognition through random forest model to provide corresponding data for the subsequent visualization scene.

3.1 Extraction of Basic Music Elements and Audio Signal Preprocessing

Although music has high abstraction, it contains many information elements, and each element has its own characteristic information. According to micro, conventional and macro scales, music elements can be divided into three levels: low, medium and high, as shown in Table 1. As can be seen from the table, low-level music information is audio data, which is specifically described as music signals. Music time domain information and frequency domain information can be obtained through the analysis of music signals.

	Measure	*type*	*Give an example*
Low-level music information	Microscopic	Music data analysis	Energy information, spectrum information, etc.
Middle music information	routine	Analysis of basic theory structure of music	Melody, rhythm, harmony, mode, etc.
High-level music information	macroscopic	Music emotion analysis	Mood, musical style, etc.

Table 1: Results of the hierarchical division of music information.

Music intermediate information, namely music infrastructure, is also an important theoretical knowledge category in music teaching. In a broad sense, it is an important attribute index to describe music characteristics, as shown in Table 2.

	Makeup	*Characteristic description*
melody	The tendency and repetition of musical sounds	Density, tightness, harmony, conflict
rhythm	Speed, beat, stress	Contrast and harmony in time dimension
harmony	Superimposed combinations of musical sounds	Density, tightness, harmony, conflict
pitch	Frequency, scale	Conflict and reconciliation
timbre	Harmonic spectrum, envelope	Music, noise, brightness

The intensity of a sound	Amplitude, decibel	Strength, size, gradient, mutation

Table 2: Composition and characteristics of music infrastructure elements.

Compared with low and middle-level information, high-level information is more complex, and its content is more extensive, which contains music emotion information with subjective components.

In the visualization process of music teaching, an emotion recognition model should be used to extract music emotion feature information. However, this process cannot be completed directly from high-level information but needs to be extracted indirectly and quantitatively through low-level and middle-level information. In this paper, Fourier and Maier scales are used to transform the music sound signal processing, and then the residual fragrance and Maier frequency cepstrum coefficient are expressed. The description of the music sample is shown in formula (1):

$$\dot{s}(t)=\sum_{g=1}^{p}a_g s(t-g) \tag{1}$$

Where the predicted time sequence is expressed as p, the predicted music sample is described as $\dot{s}(t)$, the actual value is expressed as $s(t)$, the set of linear prediction coefficients is denoted by a_g and g.

The error formula is shown in (2):

$$e(t)=s(t)-\dot{s}(t) \tag{2}$$

The residual linear prediction of music signals is shown in (3):

$$\begin{cases} r_a(t)=r(t)+jr_h(t) \\ r_h(t)=IFT[R_h(w)] \end{cases} \tag{3}$$

Where the predicted residual of the music signal is expressed as $r(t)$, the corresponding Fourier transform result is expressed as $R_h(w)$, the Hilbert transformation result is expressed as $r_h(t)$, and The analytic signal is expressed as $r_a(t)$.

The analytical signal is shown in (4):

$$h_e(t)=\left|r_a(t)\right|=\sqrt{r_h^2(t)+r^2(t)} \tag{4}$$

The remaining phase analytic signal is shown in (5):

$$\cos(\theta(t))=\frac{R_e[r_a(t)]}{\left|r_a(t)\right|}=\frac{r(t)}{h_e(t)} \tag{5}$$

Where the analytic signal phase cosine is expressed as $\cos(\theta(t))$.

3.2 Music Emotion Recognition Based on Random Forest

Random forests improve the accuracy and robustness of models by combining predictions from multiple decision trees. Each decision tree learns and makes predictions independently, and the final result is reached by a majority vote or average. This integrated approach effectively reduces the overfitting or underfitting problems caused by a single decision tree, thus improving the accuracy of music emotion recognition. In music emotion recognition, audio signals contain rich features, such as rhythm, timbre, spectrum, etc. Random forests are able to randomly select feature subsets to split during the construction of each decision tree, which not only increases the diversity of the model but also helps to deal with complex musical features. In the practical application of music emotion recognition, there may be an imbalance in the number of samples of different emotion categories. By integrating the prediction results of multiple decision trees, random forest can alleviate the impact of unbalanced data sets on model performance to a certain extent and make the model prediction results more balanced in various categories.

Let the corresponding node be represented as T, after segmentation, the node is represented as t_1, t_2, classification type representation C, T the number of sample types included is j, the probability of a class in the split node $P(C_n | t_m)$ indicates. When the split point is selected as a specific value in the indicator, the Split node t_m the Gini index is calculated as shown in (6):

$$Gini(t_m) = 1 - \sum_{n=1}^{2} [P(C_n | t_m)]^2 \tag{6}$$

The Gini impurity of any index attribute divided by the node can be calculated by (7):

$$Gini(T) = \sum_{m=1}^{2} \frac{i_m}{i} Gini(t_m) \tag{7}$$

Where the number of split nodes is m, the number of samples in the split node i_m, i represents the total number of samples of the node.

The prediction result of each classification tree represents a type of vote, and the category with the highest number of votes is the prediction classification label. For example, (8) indicates the calculation of prediction results:

$$Y_f = \arg\max_{y} \sum_{k=1}^{ntree} \delta(h(x, \theta_k) = y) \tag{8}$$

The formula $ntree$ represents the number of decision trees, $\delta(\bullet)$ representational function.

In order to prevent the random forest model from falling into local optimal or overfitting problems, an ant colony algorithm is introduced to optimize the model. Ant colony algorithm has a strong global search ability and can search widely in the solution space to find a better solution. Applying it to the optimization of random forest models can help find better decision tree combinations or parameter settings to improve the prediction accuracy of the models. In the optimization process, the ant colony algorithm can increase the diversity of decision trees by adjusting the generation strategy or selection mechanism of decision trees in random forests. The enhancement of diversity helps to improve the stability of the model so that it can maintain good performance on different data sets. Figure 3 shows the schematic diagram of the operation flow of the improved random forest model based on the ant colony algorithm.

Let the music be represented as X, put every element in the music X_i as an individual ant, the distance between the element and the cluster centre is expressed as l_{ij}, it is calculated by Euclidean distance, as shown in formula (9):

$$l_{ij} = \sqrt{\sum_{k=1}^{m} p_k (X_{ik} - C_{jk})^2} \tag{9}$$

Among them, the basic characteristic dimension of ants is expressed as m the weighting factor of each dimension is expressed as p_k.

The amount of information contained in a region is expressed as ph_{ij} the influence range of individual ants on the amount of information in the surrounding local area is expressed as r the degree of influence is shown in formula (10):

$$ph_{ij} = \begin{cases} 1, & d_{ij} \le r; \\ 0, & otherwise \end{cases} \quad (10)$$

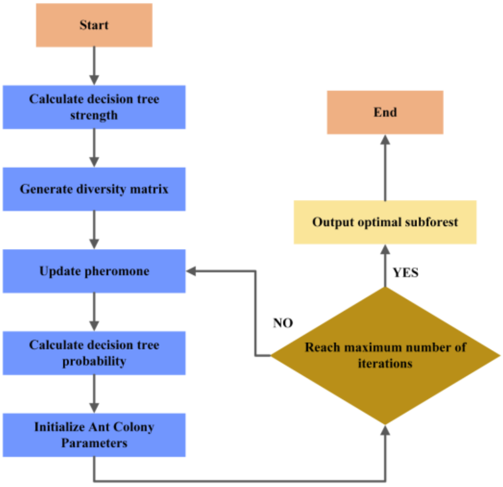

Figure 3: Schematic diagram of the operation flow of the improved random forest model based on the ant colony algorithm.

When individual ants make route selection, let the attraction factor between the current element point and the cluster centre in the iteration number of sequence number be expressed as $p_{ij}^{k}(t)$ its expression is shown in formula (11):

$$p_{ij}^{k}t = \begin{cases} \dfrac{[p_{ij}^{k}t]^{\alpha}[\eta_{ij}t]^{\beta}}{\sum\limits_{s=allowed_k}[p_{ij}^{k}t]^{\alpha}[\eta_{ij}t]^{\beta}}, if \ j \in allowed_k; \\ 0, \ or \ other \end{cases} \quad (11)$$

Where the residual information factor is expressed as a, heuristic information factor is expressed as β, the set of ant paths is represented as $allowed_k$, and as shown in (12):

$$allowed_k \in \{X_s \mid d_{sj} \le r, s = 1,2,\ldots,N\} \quad (12)$$

The heuristic function is expressed as $\eta_{ij}(t)$ and its formula is shown in formula (13):

$$\eta_{ij}t = \frac{r}{d_{ij}} = \frac{r}{\sqrt{\sum\limits_{k=1}^{m} p_k (X_{ik} - C_{jk})^2}} \quad (13)$$

Before the validity of the model is verified, the algorithm performance of the optimized random forest model needs to be checked first. In this paper, two other algorithm models are selected for comparison. Figure 4 shows the comparison results of classification accuracy of the three algorithms in the same data set. The results show that the classification accuracy of the three algorithm models will first improve with the increase in the number of iterations. When the number of iterations reaches more than 100, the improvement speed of the classification accuracy of the algorithm will obviously slow down and gradually enter a stable state. In the process of improving the classification accuracy of its algorithm, the other two algorithms have relatively obvious fluctuations, among which the ant colony algorithm has the largest fluctuation range. The proposed algorithm shows good stability in the iterative process, and the classification accuracy is the highest.

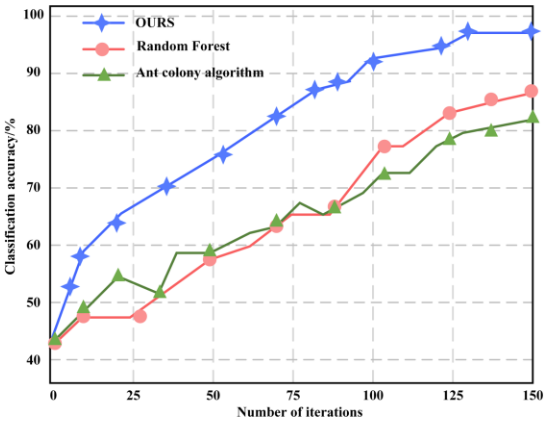

Figure 4: Classification accuracy pairs of three algorithms in the same data set.

In order to test the music visualization performance of the model, this paper compares its performance with that of other common algorithms, and the results are shown in Figure 5. In the whole feature space of music, the classification accuracy of the three models has varying degrees of volatility, which is due to the high similarity and abstractness of music features, and the difficulty of distinguishing some features is relatively large. In terms of accuracy, the accuracy of the algorithm model in this paper has the best performance in all the ten groups of experiments, and the accuracy rate is above 82%, which is significantly higher than that of the other two algorithm models, and the fluctuation range of accuracy is the smallest. This shows that the model can provide more accurate music emotion classification results in practice.

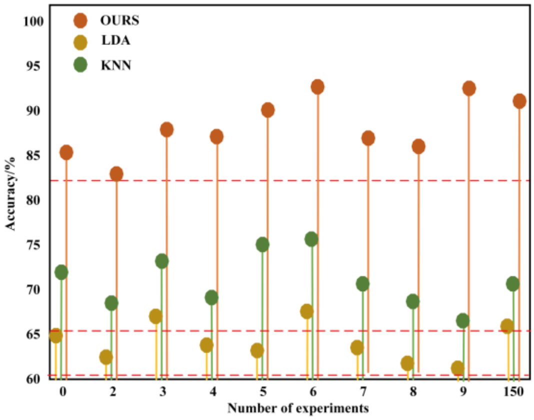

Figure 5: Comparison of music emotion classification accuracy in all feature Spaces of the three models.

Considering the effect of music emotion classification on visualization, the accuracy rate of emotion classification of the model was tested in this paper, and the results are shown in Figure 6. The main musical emotions in the test data are divided into four categories, namely irritability, depression, excitement, and relaxation, which are also the most common emotional categories in musical emotional expression. The results of four kinds of emotion classification showed that the accuracy of depression emotion classification was the highest, which basically remained above 90%, and the highest could reach 98%. The stability of the classification accuracy of excitement emotion was the worst, which was mainly due to the high subjectivity of emotion expression, and the difference of excitement expression factors was the highest among different music. The classification accuracy of the model is the lowest, but it can reach more than 81% on the whole, which can achieve the expected effect of practical application. It can be seen that this model can provide reliable and accurate emotion classification results for music visualization in practical applications.

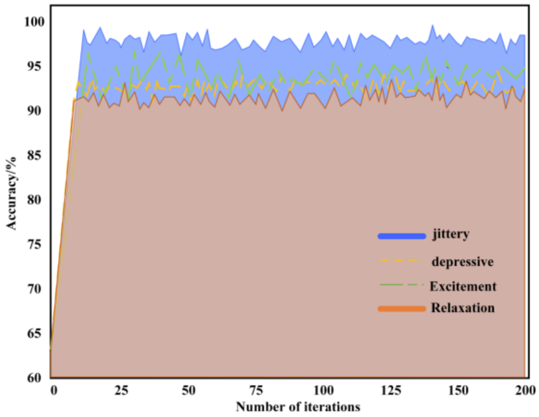

Figure 6: Emotional accuracy of music.

4 EXPERIMENTAL RESULTS AND ANALYSIS OF VISUAL MUSIC TEACHING MODEL BASED ON BIG DATA

The purpose of music teaching visualization is to help students better learn and understand music knowledge and connotation, so in the application experiment, this paper randomly selected two classes in a school to carry out a music teaching comparison experiment. The number of students in the two classes is 43 and 42 respectively. According to the previous analysis results of music performance, there is no obvious difference between the two classes. In the experiment, Class A is the comparison class, which adopts the traditional music teaching method; Class B is an experimental class, which uses visual scene teaching of music teaching. The comparison experiment lasted for one month. Before and after the experiment, the two classes took the same music test.

According to the composition of the basic elements of music above, students in music teaching need to understand the overall structure and internal emotion of music through pitch, speed, strength, and other aspects so as to express themselves better in their music performance. In traditional music teaching, students are mainly helped to understand these basic elements of music through hearing and the teacher's language description. In the visual scene of music teaching, the model will combine visual and auditory aspects to help students understand the changes in the basic elements of music more intuitively. Figure 7 shows the visual result of a piece of music. It can be seen from the figure that the change in the speed of music is shown through the solid line, and the intensity of music playing at different speeds is intuitively shown through the colour range. In addition, in order to help students understand better, the figure also marked the beat and average speed so that students can have a more effective reference in the learning process.

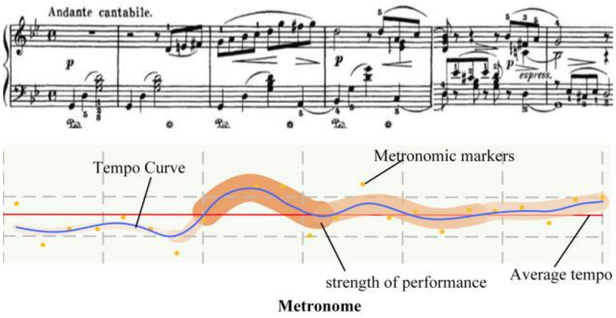

Figure 7: Visual results of basic elements of music fragments.

The inner emotion of music is dynamic. In the teaching process, students can distinguish simple musical emotions, but in difficult music, many students can not clearly distinguish the change of musical emotion. Therefore, the model in this paper marks different emotions in the music through different colours, as shown in Figure 8. As can be seen from the results in the figure, the visual results can intuitively present the melody, rhythm, and emotional changes of the music fragments. From the perspective of emotion, the emotional tone of segment A is richer and brighter, and the emotional atmosphere displayed by the whole segment is more open and bright. The emotional colour richness of segment B is relatively weak, and the colour is greener, which indicates that the emotional atmosphere of this segment is relatively reserved and peaceful.

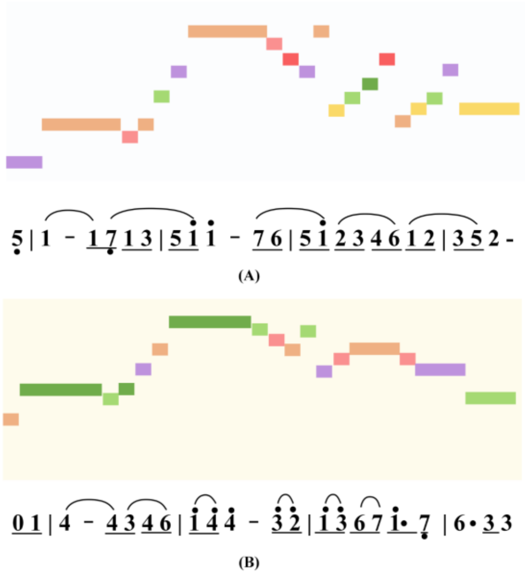

Figure 8: Visual results of music colour corresponding to music fragments.

The whole music shows different inner emotions without using sections. Therefore, in the teaching process, teachers will help students deconstruct the music structure through sections to help students better understand. However, in the traditional music teaching method, the teacher mainly tells the students to mark the sections directly and then makes the corresponding explanation and analysis, which is easy to cause the problem of wrong marking. At the same time, when students follow the teacher to explain and observe the music, they are also prone to problems such as wrong lines or lack of concentration. Therefore, different music segments will be marked with different colours in the visualization of music teaching, as shown in Figure 9.

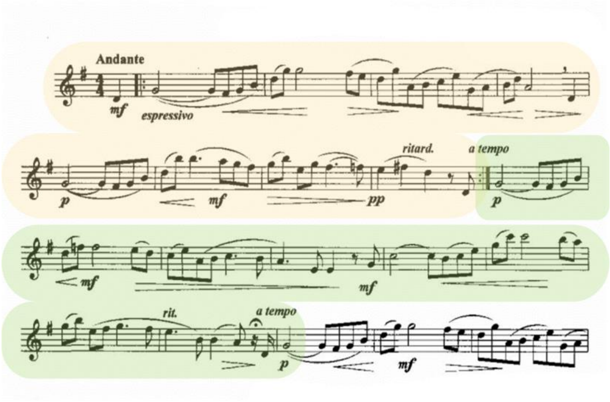

Figure 9: Visual results of segment labelling of a certain music segment.

Figure 10 shows the comparison results of the scores of the two classes before and after the experiment. The music test mainly includes theory tests and skill tests. The theory test includes three separate tests: music theory knowledge, music history and music appreciation; the skill test includes vocal music and solfeggio and ear training. The results showed that before the experiment, the theoretical test results of the students in both classes were not satisfactory, and the scores of music theory knowledge and music appreciation were low. In terms of skill tests, the average score for vocal music is relatively good, but the score for solfeggio training is poor. After the experiment, the control class had a small improvement in theory and skill, but the improvement in solfeggio and ear training was not ideal. The overall test results of the experimental class have improved significantly, especially the scores of music theory knowledge and solfeggio and ear training. This shows that the teaching visual scene constructed by the model in this paper can effectively and intuitively present the corresponding knowledge points, help students better understand the changes in the basic elements of the music, structure the music composition from a deep level, and help students improve their weak points in a targeted way, and ultimately improve their test scores.

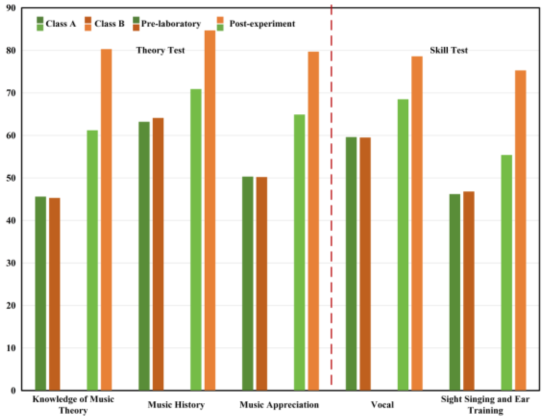

Figure 10: Comparison of results before and after the experiment between two classes.

5 CONCLUSIONS

Music knowledge is highly abstract, and traditional music teaching mainly explains music knowledge through hearing and language description, which is still a great learning difficulty for students. Therefore, this paper builds a visual music teaching model driven by big data and combined with a random forest algorithm. In order to improve the model's performance of music emotion classification, this paper also introduces an ant colony algorithm for optimization. Performance test results show that compared with other algorithms, the optimized algorithm in this paper has the highest classification accuracy after sufficient iterations and shows better stability and higher accuracy in all feature Spaces. In addition, the classification accuracy of different music emotions can reach more than 81%, which can basically meet the needs of the application of the model. The experimental results show that the model can visualize the changes in basic musical elements and the dynamics of emotional changes according to the results of music emotion classification and music signal processing. At the same time, the model can also mark fragments with different colours to help students better understand the structure of the music. The comparative experimental results show that the model in this paper can effectively help students learn better music knowledge, help students understand the changes in the basic elements of the music through an intuitive visual way, understand the structure of the music from multiple angles, and help students improve their weak points and greatly improve the final test scores.

Dongxuan Wu, https://orcid.org/ 0000-0002-2973-9816
Cheng Zhou, https://orcid.org/0000-0003-4836-3217
Xiaojing Wen, https://orcid.org/0000-0002-8967-5500

REFERENCES

[1] Alfaro, C.-M.; Valero, M.-J.-J.: Exploiting the two-dimensional nature of agnostic music notation for neural optical music recognition, Applied Sciences, 11(8), 2021, 3621. https://doi.org/10.3390/app11083621
[2] Andrea, P.; Zahra, A.: Music note position recognition in optical music recognition using convolutional neural network, International Journal of Arts and Technology, 13(1), 2021, 45-60. https://doi.org/10.1504/IJART.2021.115764
[3] Baro, A.; Riba, P.; Calvo, Z.-J.; Fornés, A.: From optical music recognition to handwritten music recognition: a baseline, Pattern Recognition Letters, 123(5), 2019, 1-8. https://doi.org/10.1016/j.patrec.2019.02.029
[4] Bhattarai, B.; Lee, J.: Automatic music mood detection using transfer learning and multilayer perceptron, International Journal of Fuzzy Logic and Intelligent Systems, 19(2), 2019, 88-96. https://doi.org/10.5391/IJFIS.2019.19.2.88
[5] Calvo, Z.-J.; Jr, J.-H.; Pacha, A.: Understanding optical music recognition, CM Computing Surveys (CSUR), 53(4), 2020, 1-35. https://doi.org/10.1145/3397499
[6] Cavdir, D.; Wang, G.: Borrowed Gestures: The body as an extension of the musical instrument, Computer Music Journal, 45(3), 2021, 58-80. https://doi.org/10.1162/comj_a_00617
[7] Garrido, M.-C.; Rios, V.-A.; Calvo, Z.-J.: A holistic approach for image-to-graph: application to optical music recognition, International Journal on Document Analysis and Recognition (IJDAR), 25(4), 2022, 293-303. https://doi.org/10.1007/s10032-022-00417-4
[8] Ge, M.; Tian, Y.; Ge, Y.: Optimization of computer aided design system for music automatic classification based on feature analysis, Computer-Aided Design and Applications, 19(S3), 2021, 153-163. https://doi.org/10.14733/cadaps.2022.S3.153-163
[9] Kim, H.-R.: Development of the artwork using music visualization based on sentiment analysis of lyrics, The Journal of the Korea Contents Association, 20(10), 2020, 89-99. https://doi.org/10.5392/JKCA.2020.20.10.089
[10] Lima, H.-B.; Santos, C.-G.-D.; Meiguins, B.-S.: A survey of music visualization techniques, ACM Computing Surveys (CSUR), 54(7), 2021, 1-29. https://doi.org/10.1145/3461835

[11] 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
[12] Paul, A.; Pramanik, R.; Malakar, S.; Sarkar, R.: An ensemble of deep transfer learning models for handwritten music symbol recognition, Neural Computing and Applications, 34(13), 2022, 10409-10427. https://doi.org/10.1007/s00521-021-06629-9
[13] 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
[14] Ríos, V.-A.; Esplà, G.-M.; Rizo, D.; Ponce, L.-P.-J.; Iñesta, J.-M.: Applying automatic translation for optical music recognition's encoding step, Applied Sciences, 11(9), 2021, 3890. https://doi.org/10.3390/app11093890
[15] Rocamora, M.; Cancela, P.; Biscainho, L.-W.-P.: Information theory concepts applied to the analysis of rhythm in recorded music with recurrent rhythmic patterns, Journal of the Audio Engineering Society, 67(4), 2019, 160-173. https://doi.org/10.17743/jaes.2019.0003
[16] Sarkar, R.; Choudhury, S.; Dutta, S.; Roy, A.; Saha, S.-K.: Recognition of emotion in music based on deep convolutional neural network, Multimedia Tools and Applications, 79(1), 2020, 765-783. https://doi.org/10.1007/s11042-019-08192-x
[17] Van, B.-S.-M.; Grajek, J.-S.; Rettschlag, S.; Schraven, S.-P.; Mlynski, R.: Interactive electronic visualization formats in student teaching, HNO, 72(5), 2024, 341-349. https://doi.org/10.1007/s00106-024-01436-9
[18] Wang, Q.; Zhou, L.; Chen, X.: Kernel density estimation and convolutional neural networks for the recognition of multi-font numbered musical notation, Electronics, 11(21), 2022, 3592. https://doi.org/10.3390/electronics11213592
[19] Yang, D.; Goutam, A.; Ji, K.; Tsai, T.-J.: Large-scale multimodal piano music identification using marketplace fingerprinting, Algorithms, 15(5), 2022, 146. https://doi.org/10.3390/a15050146
[20] Zhang, Y.: Music emotion recognition method based on multi-feature fusion, International Journal of Arts and Technology, 14(1), 2022, 10-23. https://doi.org/10.1504/IJART.2022.122447