scientific reports

Check for updates

OPEN

The analysis of optimization in music aesthetic education under artificial intelligence

Yixuan Peng

In the artificial intelligence (AI) domain, effectively integrating deep learning (DL) technology with the content, teaching methodologies, and learning processes of music aesthetic education remains a subject worthy of in-depth exploration and discussion. The aim is to meet to the music aesthetic needs of students across different age groups and levels of musical literacy. In this paper, the concepts of AI and DL algorithm are first introduced, and their algorithm principles and application status are understood. Then, they are integrated into the application of music aesthetic education, and the algorithm principles and running codes are designed. Finally, experiments are carried out to verify the accuracy of music emotion recognition based on DL algorithm in AI environment to verify the effectiveness of music aesthetic education method based on DL. The results show that the algorithm proposed in this paper has higher accuracy, which combines the advantages of AI and DL algorithm, and obtains higher recognition accuracy. It provides more possibilities for future music aesthetic teaching activities. This paper is dedicated to investigating the feasibility and approach to optimizing the method of music aesthetic education through DL. Its objective is to chart a new developmental direction and practical pathway for music aesthetic education in the era of AI.

Keywords Artificial intelligence, Deep learning, Music aesthetic education, Algorithm optimization, Emotion recognition

Research background and motivations

Under the background of the rapid development of artificial intelligence (AI), the field of music education is also undergoing unprecedented changes. The breakthrough of AI technology, especially deep learning (DL) technology, has brought new opportunities and challenges to music aesthetic education[1,2]. Music aesthetic education is a form of aesthetic education that uses music as a medium. It aims to cultivate students' aesthetic abilities, humanistic qualities, and emotional experiences through learning, appreciating, and creating music. It not only focuses on the mastery of musical knowledge and skills but also emphasizes the impact of music on individual emotions, thoughts, and personality development. It helps students feel, understand, and create beauty through music. Unlike traditional music education, which primarily focuses on music theory, performance techniques, and the reproduction of musical works, music aesthetic education places greater emphasis on the aesthetic value and cultural significance of music. It prioritizes the cultivation of students' emotional experiences and artistic perception. In contrast, traditional music education is more goal-oriented toward developing professional skills. Music aesthetic education, however, strives to promote holistic personal development through music, making it an essential part of an individual's spiritual world[3].

Deep learning is a branch of machine learning that uses multi-layer neural networks to simulate the way the human brain learns, automatically extracting features from large amounts of data for pattern recognition and prediction. Integrating AI and deep learning into education faces many challenges. One is the accessibility and affordability of AI-driven platforms. Due to high technology costs and significant hardware requirements, many schools and educational institutions struggle to provide universal access. To address this, more cost-effective, cloud-based solutions or open-source platforms can be developed to lower the barriers to use. Another challenge is data privacy and ethical considerations. The education field involves a large amount of personal data. When using AI, it is essential to ensure student privacy is protected and to comply with relevant laws and regulations, such as GDPR or the U.S. FERPA. Data encryption and de-identification techniques can be used to enhance security. At the same time, biases may exist in deep learning models, which can lead to unfair assessment results, especially in diverse student populations. To tackle this issue, fairness adjustments need to be made in the training datasets to ensure they include representatives of various groups and to continuously optimize

School of Preschool Education, Hunan College for Preschool Education, Changde 415000, China. email: pengyixuan0426@163.com

the models to reduce biases. Therefore, this paper is committed to exploring the possibilities and pathways for optimizing music aesthetic education methods based on deep learning, hoping to bring new directions and practical paths to music education in the age of AI.

The purpose of this paper is to explore how to make full use of the advantages of DL technology to optimize and innovate the methods of music aesthetic education to enhance students' appreciation, understanding and creativity of music, cultivate more comprehensive music literacy, and meanwhile provide more efficient and scientific teaching tools and strategies for music educators, and promote the modernization and intelligent development of music education. Through this paper, it is expected to make up for the possible shortcomings in the existing music education system and provide strong theoretical support and practical reference for the future development of music aesthetic education in China.

Research objectives

The motivation of this paper stems from AI technology, especially the increasingly significant influence of DL technology in the field of education, and its potential great application value in music aesthetic education. The research goal is to use DL technology to solve the bottleneck problems of insufficient personalization, single teaching method and imperfect feedback mechanism in traditional music aesthetic education[4].

Firstly, this paper introduces DL technology to accurately analyze and identify the emotional characteristics in music works. This technical breakthrough makes the analysis of music emotion no longer limited to the traditional subjective evaluation, but relies on data-driven methods to achieve more objective and accurate emotion recognition. Secondly, the experimental method adopts intelligent data feedback mechanism, which transforms students' emotional reaction in music aesthetic education into visual data, and adjusts teaching strategies in real time according to these data to provide students with personalized aesthetic education guidance. This intelligent feedback mechanism enhances the interaction between teachers and students and makes teaching more targeted[5]. Finally, the experimental method pays attention to the combination of quantitative and qualitative methods. Additionally, through the quantitative analysis of students' emotional response, it supplements the qualitative evaluation method in traditional education, making music aesthetic education more scientific. This method not only enriches the means of education evaluation, but also provides data support for optimizing the teaching content and methods, and promotes the development of music aesthetic education in the direction of intelligence and personalization[6].

Literature review

In recent years, with the rapid development of AI technology, especially the major breakthrough in the field of DL, its application in various educational fields has become more and more extensive. Music aesthetic education, as an important branch full of artistry and creativity, has also ushered in unprecedented development opportunities[7–9]. Many researchers have started actively exploring the integration of AI and DL technology into music aesthetic education methods and practices. This integration aims to optimize the teaching process, enhance teaching effectiveness, and address personalized and intelligent educational needs in the modern era[10].

Bharadiya studied how to use DL technology to automatically analyze and mark music materials to realize intelligent analysis and personalized teaching of music elements. A music aesthetic teaching platform based on DL was developed. It could intelligently recommend learning content according to students' music preferences and learning progress, and optimize the teaching path through real-time feedback mechanism. Experiments showed that this platform had significant advantages in improving students' musical aesthetic ability and learning enthusiasm[11]. Moysis focuses on how AI technology was transforming the traditional music aesthetic education mode. This included utilizing DL algorithms to identify and convey music emotions, as well as integrating virtual reality (VR) technology to establish an immersive music learning environment. The research results showed that the music aesthetic education method optimized by AI not only improved students' cognitive understanding and emotional resonance of music, but also greatly enhanced their innovative thinking and practical ability[12]. Shukla put forward a new teaching mode of cultivating musical aesthetic ability based on DL technology, and carried out experiments in the actual teaching environment. By constructing a deep neural network model, students' musical aesthetic ability was quantitatively evaluated, and the teaching content and strategies are dynamically adjusted according to the evaluation results. The research showed that this model could meet the individual differentiated learning needs of students more accurately, thus effectively improving the effect and efficiency of music aesthetic education[13].

In the field of emotion recognition research, international studies have been exploring deep learning-based methods for some time. Oksanen et al. proposed the basic emotion theory, which laid the theoretical foundation for facial expression-based emotion recognition[14]. Shukla emphasized the importance of multimodal emotion recognition, advocating for the combination of voice, text, and facial expression data to improve recognition accuracy[13]. Borkowski used Generative Adversarial Networks (GANs) to optimize facial expression data, enhancing the generalization ability of recognition systems[15]. In China, researchers have built on international achievements to optimize emotion recognition methods in the context of the Chinese language and local culture. For example, Cui (2015) introduced a Mandarin speech emotion recognition method based on Deep Belief Network (DBN), which improved feature extraction[16]. Huang et al. developed a multimodal emotion recognition model incorporating an attention mechanism to enhance emotional perception in music education. The research noted that the application of emotion recognition technology in music aesthetic education could optimize intelligent teaching feedback, increase the precision of personalized music recommendations, and enhance learners' emotional experiences[17].

In summary, researchers have developed several deep learning-based teaching platforms and models. These tools significantly enhance students' musical aesthetic abilities and learning enthusiasm by automatically analyzing musical materials, recognizing emotions, and using VR. However, existing studies have some

potential limitations and biases. For example, although deep learning excels at extracting musical features, it falls short in nurturing the subjectivity of emotional expression and aesthetic awareness in music aesthetic education. Many studies focus on quantifying students' learning progress and personalized needs but fall short in comprehensively considering complex factors such as emotional responses and cultural backgrounds, which may lead to biases in personalized teaching. Moreover, existing deep learning models often rely on large amounts of data and high computational power, which may pose scalability issues, especially in educational environments with limited resources. Future research needs to further enhance the models' emotional perception capabilities and strengthen their applications in unstructured data and subjective experiences to overcome the limitations of current methods and promote the comprehensive development of music aesthetic education.

Research methodology

The combination of AI and music aesthetic education

AI is a science and technology field dedicated to understanding and creating intelligent behaviors. AI integrates the knowledge of computer science, statistics, cognitive science and other disciplines, aiming at researching, designing and developing intelligent agents. These agents can perceive the environment, learn, reason, solve problems and make decisions independently. The application forms of AI are diverse, covering but not limited to robotics, speech recognition, image recognition, natural language processing (NLP), expert system and machine learning. According to its function and complexity, AI can be further divided into three levels: weak AI, strong AI and super AI[18]. Its hierarchical division is shown in Fig. 1 below:

Weak AI focuses on and is good at the performance of specific tasks. While strong AI has cross-disciplinary and comprehensive human-level intelligence and can perform any cognitive task that human beings can complete. As for strong AI, it is assumed that it exists beyond the limits of human intelligence and can be far superior to the best human beings in all cognitive abilities[12,15,19]. With the development of technology, AI is increasingly infiltrating into all aspects of social life, and continues to promote scientific and technological progress and industrial upgrading[20].

With the continuous development of AI technology, reshaping music provides more support and creativity for cultivating students' academic accomplishment, aesthetic perception, artistic expression and cultural understanding. On this basis, the possibility of using AI technology to promote the reform and development of music aesthetic teaching is discussed[21,22]. By applying advanced AI technology, music aesthetic education can realize more refined and personalized teaching strategies. The relationship between them is shown in Table 1:

On one hand, AI can accurately identify and assess students' musical aesthetic tendencies, learning styles, and ability levels through big data analysis and algorithmic models. This enables the provision of personalized teaching content and learning paths, enhancing the relevance and effectiveness of teaching. On the other hand, AI can also be employed in emotional analysis, style identification, creative assistance, and other aspects of music composition. This aids students in gaining a deeper understanding of the internal structure and external expression of music, thereby broadening their aesthetic horizons and enhancing their appreciation of music. Furthermore, intelligent music teaching systems can facilitate students' active engagement and immersive learning experiences through real-time feedback and interaction. This shift from traditional indoctrination to a new teaching mode focused on inquiry and experiential learning promotes both the quality and efficiency of music aesthetic education, leading to a dual improvement in educational outcomes[23].

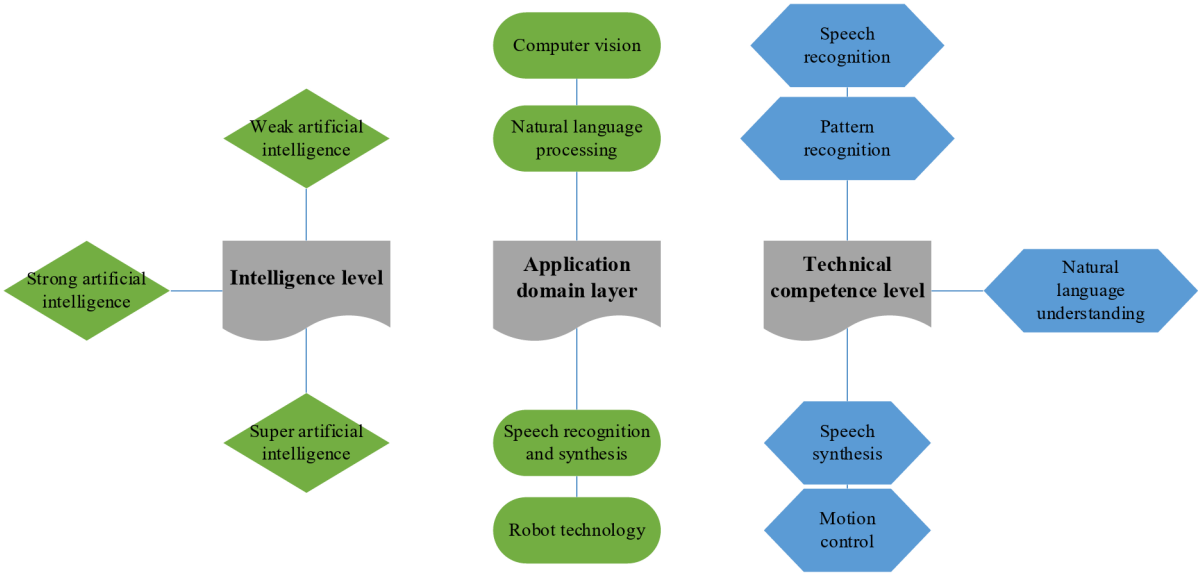

Fig. 1. AI hierarchy.

Centre	Bonding point	Specific description
The combination of AI and music aesthetic education	Individualized teaching strategy	Use AI big data to analyze the characteristics of students
		Provide students with personalized teaching content and learning ways
		Improve the pertinence and effectiveness of teaching
	Analysis and assistance of music works	AI is used in music emotion analysis and style recognition
		Make students have a deeper understanding of the meaning and form of music
		Broaden aesthetic vision and enhance appreciation level
	Real-time feedback and interaction	Real-time feedback of learning effect in intelligent music class
		Encourage students to actively participate in immersion learning.
		Reform of teaching methods
	Education quality and efficiency improvement	Application of AI technology in music aesthetic education teaching

Table 1. Relationship between AI and music aesthetic education.

DL algorithm

DL is a machine learning method, which is inspired by the working principle of human brain neural network, and carries out high-level abstraction and learning of data by constructing a series of interconnected artificial neural network (ANN) levels[24]. In the DL model, the original input (such as image pixels, text characters or sound waveforms) first passes through a multi-level nonlinear transformation. Each layer will extract more and more complex feature representations from the input data. This layer-by-layer processing mechanism enables the model to automatically learn and explore useful features from the original data without artificially designing features[25–27]. The core technologies of DL include deep neural networks (such as convolutional neural networks, recurrent neural networks, long and short-term memory networks, etc.). They are excellent in solving many complex problems such as image recognition, speech recognition, NLP, machine translation, recommendation system, game strategy, and automatic driving. By updating model parameters through optimization algorithms (such as back propagation algorithm), the DL model can self-train and improve on large-scale datasets, thus achieving high-precision prediction and decision-making ability[28]. With the enhancement of computing power and the popularity of big data, DL has become one of the key technologies in the field of modern AI and has made revolutionary breakthroughs in many industries[29].

The unsupervised learning algorithm of DL can be used to learn the intrinsic feature representation of music signals. By employing dimensionality reduction and reconstructing music data, self-encoders are capable of extracting meaningful music features from the original audio signal. These features may include rhythm, timbre, and harmonic structure, among others. Such extracted features can aid teachers in designing and implementing targeted teaching activities[30]. A typical unsupervised learning method is to find the "best" representation for data. "Best" can be articulated differently, but it typically implies certain constraints or limitations compared to the information it inherently conveys. In essence, it signifies having fewer advantages or constraints. Therefore, it is essential to retain as much information about X as possible.

The matrix X is designed as a combination of $m \times n$. The average value of the data is 0, that is, $E[x] = 0$. If it is not 0, the average value is subtracted from all samples in the preprocessing step, and the data is easily centralized[31]. The unbiased sample covariance matrix corresponding to X is given as follows:

$$\mathrm{Var}\,[\mathrm{x}] = \frac{1}{m-1} X^T X \tag{1}$$

Var[z] is a diagonal matrix which is transformed into z= $W^T x$ by linear transformation. The principal component of the design matrix X is given by the eigenvector of $X^T X$[32], as follows:

$$X^T X = W \wedge W^T \tag{2}$$

Assume that W is the right singular vector of singular value decomposition $X = U \sum W^T$, and taking W as the basis of eigenvector, the original eigenvalue equation can be obtained as follows:

$$X^T X = \left(U \sum W^T\right)^T U \sum W^T = W \sum {}^2 W^T \tag{3}$$

This can fully explain that Var[z] in the algorithm is a diagonal matrix. Using the principal component decomposition of X, the variance of X can be expressed as:

$$\mathrm{Var}\,[\mathrm{x}] = \frac{1}{\mathrm{m}-1} \mathrm{X^T X} \tag{4}$$

$$= \frac{1}{m-1} (U \sum W^T)^T U \sum W^T \tag{5}$$

$$= \frac{1}{m-1} W \sum^{T} U^T U \sum W^T \tag{6}$$

$$= \frac{1}{m-1} W \sum^{2} W^T \tag{7}$$

$U^T U = I$ is used, because the matrix U is orthogonal according to the definition of singular value, which shows that the covariance of z meets the diagonal requirements, as shown below:

$$\mathrm{Var}\,[z] = \frac{1}{m-1} Z^T Z \tag{8}$$

$$= \frac{1}{m-1} W^T X^T X^T W \tag{9}$$

$$= \frac{1}{m-1} W^T W \sum^{2} W^T W \tag{10}$$

$$= \frac{1}{m-1} \sum^{2} \tag{11}$$

The above analysis shows that when the data x is projected to z through linear transformation w. The covariance matrix represented by the obtained data is diagonal $\left(\sum^2\right)$, which means that the elements in Z are independent of each other[1].

Through these unsupervised learning techniques, music aesthetic education can become more intelligent and personalized. Meanwhile, it can help teachers and researchers to examine and optimize teaching methods from a new perspective, thus improving the quality and effect of music education.

Music aesthetic education method based on DL in AI environment

On this basis, this paper puts forward a new music aesthetic teaching mode based on DL. Based on AI, the system provides various supporting services for students' learning[33,34]. The functional design of the system is shown in Fig. 2:

The role of machine learning in the music industry is to extract "data" from a piece of music, input it into a specific pattern, learn "features" from it, and analyze and sort it out. The specific pseudo code is arranged as follows:

Initialize AI model with multimodal deep learning framework.

Input music features (MFCC, PLP, melody) and emotional data (Arousal, Valence).

Train and optimize model for personalized music aesthetic education feedback.

With the help of DL technology, AI can deeply understand each student's learning style, interest tendency and skill level through the analysis of a large number of user data. Then, AI provides customized music appreciation and learning content. The purpose is to ensure that music aesthetic education meets individual needs more

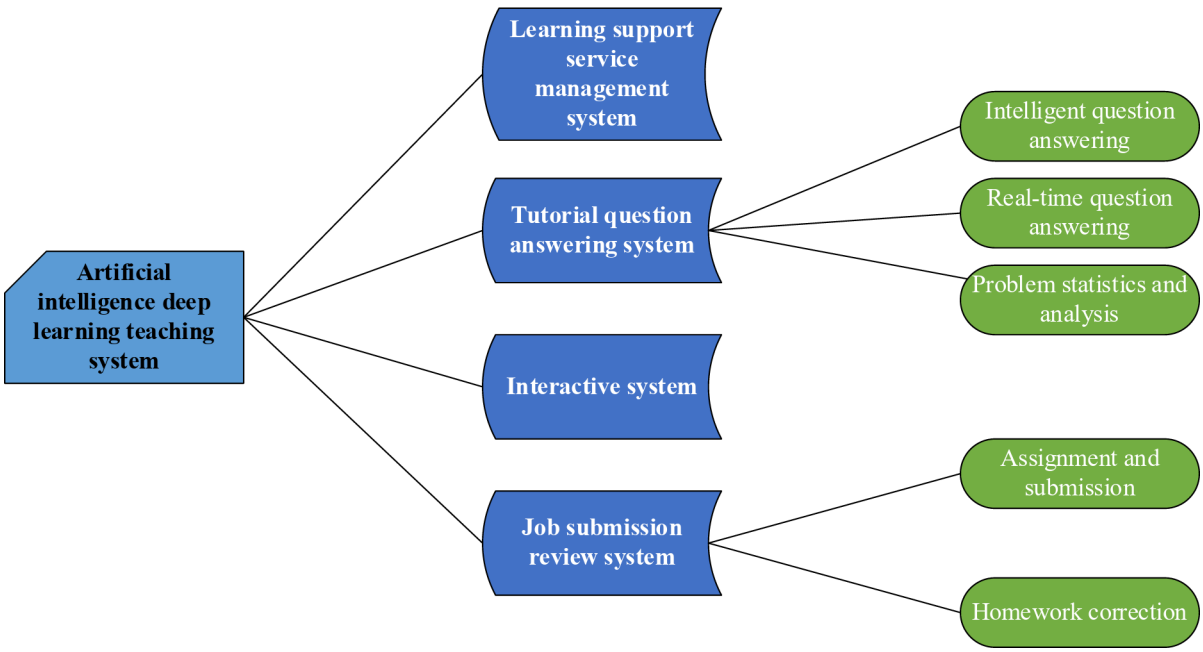

Fig. 2. System function design.

accurately and truly teach students in accordance with their aptitude[35]. Meanwhile, AI can monitor and respond to students' learning behavior and feedback in real time. It can dynamically adjust the teaching plan to ensure that the educational process is efficient and targeted, thereby enhancing the effectiveness of the learning experience. Secondly, DL strengthens the interactive experience in music education. By integrating VR/AR technology and DL algorithm, a highly immersive music teaching environment is constructed, so that students can directly feel the rhythm, melody, harmony and other aesthetic elements in music works in VR.

Music features and music emotion model based on DL in AI environment

The musical features studied here are mainly divided into Mel-frequency cepstral coefficient (MFCC) features and Perceptual Linear Prediction (PLP) features. Two kinds: MFCC is a feature parameter widely used in speech signal processing, especially in speech recognition and speaker recognition. The core idea of MFCC is to simulate the sound perception mechanism of human auditory system. PLP feature is also a feature parameter commonly used in speech recognition, but unlike MFCC, PLP feature is more directly based on human auditory perception model. PLP enhances speech recognition through a series of auditory perception simulations of speech signal spectrum.

The musical emotion model in this paper is Long Short-Term Memory (LSTM) model. LSTM is a variant of RNN and one of the important components of the neural network model designed and constructed in this paper. LSTM not only inherits the advantages of traditional RNN, but also overcomes the problem of gradient explosion or disappearance of RNN. It can effectively process arbitrary length series data and capture long-term dependence of data, and the total amount of data processed and training speed have been greatly improved compared with traditional machine learning models. LSTM consists of many repeating units, which are called memory blocks. Each memory block contains three gates and one memory unit, and the three gates are forget gate, input gate and output gate respectively. The specific calculation flow of the memory block is shown in Fig. 3:

The equation for calculating the input layer information for obtaining the upper input unit data is as follows:

$$a_t = \tanh\left(w_{xa}x_t + w_{ha}h_{t-1} + b_a\right) \tag{12}$$

v_t: The value of the candidate memory cell.

tanh: Hyperbolic tangent activation function, which is used to compress the input into [-1, 1] interval.

w_{xa}: The weight input to the candidate memory unit.

w_{ha}: Weight of hidden state to candidate memory cells at the previous moment.

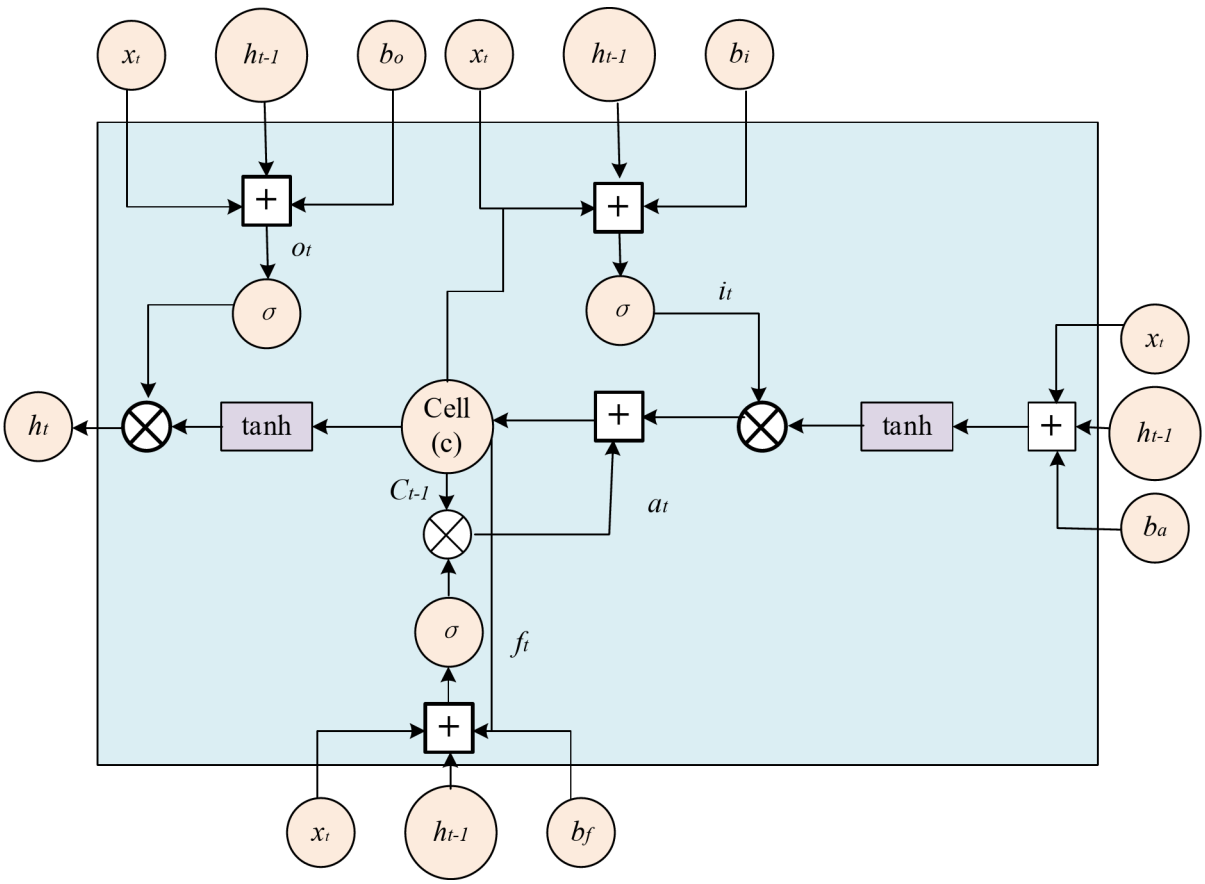

Fig. 3. Network structure diagram of LSTM.

x_t: Input at the current moment.
h_{t-1}: The hidden state of the previous moment.
b_a: Bias of candidate memory cells.
Input gate, forget gate and output gate are controlled by Sigmoid function, and the specific calculation equation is as follows:

$$i_t = \sigma\left(w_{xi}x_t + w_{hi}h_{t-1} + b_i\right) \tag{13}$$

i_t: Input the activation value of the gate to control the degree to which the current input information enters the memory unit.
σ: Sigmoid activates the function, which compresses the input into the interval of [0, 1] and is used to control the flow of information.
w_{xi}: The weight input to the input gate.
w_{hi}: The weight of the hidden state to the input gate at the previous moment.
b_i: Offset of input gate.

$$f_t = \sigma\left(w_{xf}x_t + w_f h_{t-1} + b_f\right) \tag{14}$$

f_t: The activation value of the forgetting gate controls the degree of information retention of the memory cell at the previous moment.
w_{xf}: The weight entered the forget gate.
w_{hf}: The weight from the hidden state to the forgotten gate at the previous moment.
b_f: Offset of forget gate.

$$o_t = \sigma\left(w_{x0}x_t + w_{h0}h_{t-1} + b_0\right) \tag{15}$$

o_t: The activation value of the output gate controls the information output degree of the memory unit at the current moment.
w_{x0}: The weight input to the output gate.
w_{h0}: The weight of the hidden state to the output gate at the previous moment.
b_0: Offset of output gate.
σ represents Sigmoid function, and the calculation equation is:

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{16}$$

The state values of the memory cell and the output gate are calculated as follows:

$$c_t = f_t \odot c_{t-1} + i_t \odot a_t \tag{17}$$

c_t: Memory cell state at the current moment.
⊙: Multiplication at the element level (Hadamard product).
c_{t-1}: The state of the memory cell at the previous moment.
a_t: Candidate memory cell states at the current moment.

$$h_t = o_t \odot \tanh(c_t) \tag{18}$$

h_t: The hidden state of the current moment.
$\tanh(c_t)$: Apply hyperbolic tangent activation function to the memory cell state at the current moment.

Experimental design and performance evaluation

Datasets collection

The University Students (whose age was above 18) of College of Music, The Yeungnam University of Korea are taken as the research sample, and 100 students are randomly selected.

The specific steps of the experiment are as follows: First, participants will listen to a series of carefully selected music excerpts, each lasting about 3–5 min. The music varies in style and emotional tone, covering emotions such as joy, melancholy, and excitement. During the experiment, participants' physiological responses (such as heart rate and skin conductance) will be recorded in real time using wearable devices. At the same time, their facial expressions will be captured by a camera for further analysis of emotional states. After the experiment, the model's predicted emotional states will be compared with the participants' actual emotional responses to assess the model's accuracy and effectiveness.

(1) Dataset construction.

The dataset should include multimodal data sources, primarily consisting of physiological responses, facial expressions, eye-tracking data, and music audio features generated by participants during the experiment. For physiological response data, devices such as heart rate monitors and skin conductance sensors can be used for collection. Facial expression data can be captured by high-precision cameras and extracted using facial expression recognition algorithms. Eye-tracking data can be recorded using an eye tracker to measure participants' attention levels. Music audio data should be categorized based on known emotional tones.

(2) Annotation method.

Hardware configuration	GPU1	Inter Xeon E5-2620 v4 @2.1 GHz
	GPU2	NVIDA GeForce GTX1080Ti 11GB
	Internal storage	128G
Software environment	Language	PyThon 3.7+
	Operating system	Linux Centos7
	DL framework	Pytorch 1.3
	Common library	Numpy, sklearn, opencv, etc.

Table 2. Experimental environment.

Backbone network model	B0	B1	B2	B3	B4
Network width (w)	1.0	1.0	1.1	1.2	1.4
Network depth (d)	1.0	1.1	1.2	1.4	1.8
Resolution (r)	543	637	775	904	1173

Table 3. Parameter settings.

To ensure the accuracy and consistency of the data, annotations of emotional responses and attention states should be completed by multiple annotators using a double-blind method to reduce bias. Emotional response annotations can be based on participants' physiological reactions and facial expressions, with clear criteria set for each emotional category (such as joy, sadness, anxiety, excitement, etc.) to ensure consistent labeling. Attention states are assessed through the analysis of eye-tracking data, considering metrics like eye movement frequency and fixation duration, focusing on capturing participants' concentration and fluctuations in attention towards the music.

(3) The definition of emotional response.

Emotional responses can be defined across multiple dimensions, including but not limited to: emotional valence (such as joy, sadness, anger), emotional intensity (low, medium, or high intensity), and emotional volatility (the amplitude and speed of emotional changes). The prediction of emotions should not only rely on physiological signals but also consider the emotional characteristics of the music and the participants' personal emotional inclinations.

(4) Definition of attention state.

Attention status is defined as the degree of concentration a participant has while listening to music and can be categorized into several types: highly focused, slightly distracted, and severely distracted. This state should be determined by combining eye-tracking data, reaction times, and behavioral performance (such as signs of attention drifting). Changes in attention status may interact with changes in emotional responses, especially when music evokes emotional fluctuations, causing corresponding variations in the participants' attention.

Experimental environment

This paper designs an unsupervised DL network model, which needs a large number of labeled samples for training. Consequently, it needs to use high-performance GPU for training to improve the training speed and testing speed. The built platform development environment[36] is shown in Table 2:

Parameters setting

This paper takes the collected dataset as the sample, the batch_size is 4, and the initial learning rate is 0.005, including 24 epochs cycles. Based on the initial learning rate, the learning rate is adjusted to 0.0005 in the 17th epoch and 0.00005 in the 22nd epoch, and the construction parameters[37] are shown in Table 3:

Performance evaluation

In music aesthetic education, music emotion is the core of aesthetic experience. Through emotion recognition technology, students can be provided with richer emotional experience, which can help them deeply understand the emotional expression in music and enhance their aesthetic sensibility. This technical means can assist teachers to implement personalized teaching according to students' reaction to music emotion, optimize the effect of music aesthetic education, and enhance students' emotional resonance and music understanding. Music emotion analysis based on DL can accurately identify the emotional features in music, such as joy, sadness and calmness, through AI technology, thus helping educators to better understand and convey the emotional connotation of music works. Therefore, this paper analyzes the music emotion based on DL in the AI environment to show the effectiveness of the music aesthetic education method.

The experiments are conducted under three conditions: single-input continuous emotional features, single-input discrete emotional features, and a 1:1 ratio of continuous to discrete emotional features. Under these conditions, this paper measures the MFCC and PLP features extracted from the continuous emotional space and the main melody features extracted from the discrete emotional space. The experimental results are shown in Figs. 4 and 5, and 6.

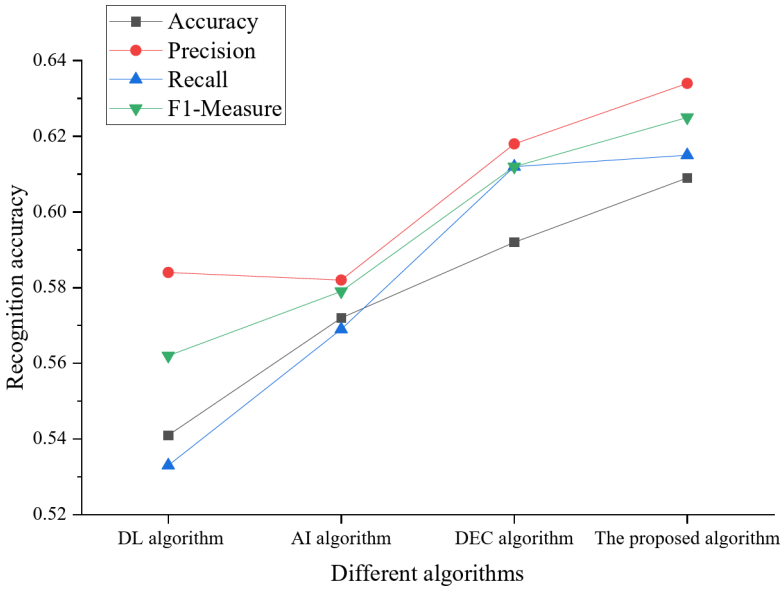

Fig. 4. Model recognition accuracy comparison with all continuous emotional features as input.

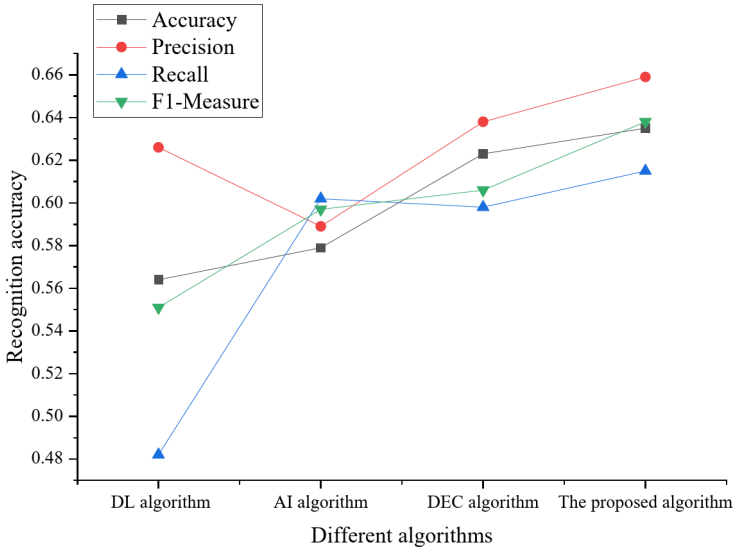

Fig. 5. Model recognition accuracy comparison with all discrete emotional features as input.

As Table 4; Fig. 4 show, the proposed model achieves the highest accuracy compared to other mainstream emotion recognition models when the number of MFCC and PLP features extracted from the continuous emotional space is equal to the number of main melody features extracted from the discrete emotional space (1:1). This demonstrates that multimodal feature input has a positive effect on music emotion recognition.

Table 5; Fig. 5 show that the proposed model performs better in the continuous emotional space on the premise that only a single feature is input, which shows that the model is more suitable for content recognition in the continuous emotional space.

Table 6; Fig. 6 show that when the same feature is input, the recognition accuracy is not as accurate as other mainstream models on the premise of only a single feature, which shows that this model is not the optimal solution when only a single feature is input.

In practice, interviews with teachers and students reveal that teachers can obtain real-time analysis reports on students' emotional responses and learning progress through the model, thereby optimizing teaching methods and content. Students, on the other hand, can adjust their performance based on model feedback, enhancing their understanding and expression of musical emotions. The practical experience assessment of the model includes feedback from teachers on its teaching assistance capabilities and evaluations from students on their personalized learning experiences. The model continuously refines its feedback accuracy and practicality through data analysis. Additionally, the model is integrated with a music interaction application, embedding

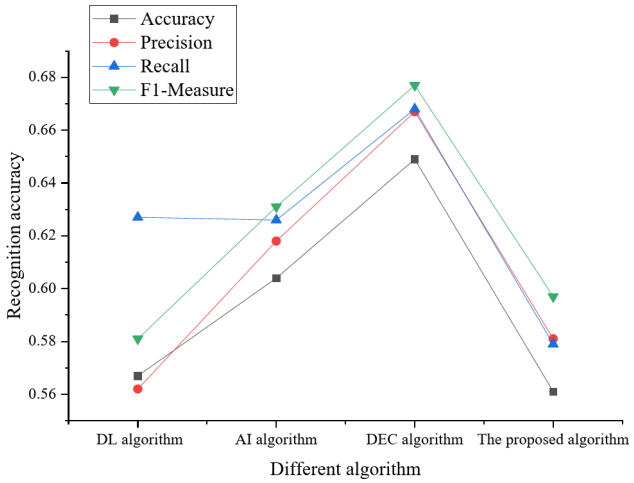

Fig. 6. Model recognition accuracy comparison with input features of continuous and discrete emotional features in a 1:1 ratio.

Model category	Accuracy	Precision	Recall	F1-Measure
DL algorithm	0.541	0.584	0.533	0.562
AI algorithm	0.572	0.582	0.569	0.579
DEC algorithm	0.592	0.618	0.612	0.612
The proposed algorithm	0.609	0.634	0.615	0.625

Table 4. Input the comparison table of model recognition accuracy with all continuous emotional features.

Model category	Accuracy	Precision	Recall	F1-Measure
DL algorithm	0.564	0.626	0.482	0.551
AI algorithm	0.579	0.589	0.602	0.597
DEC algorithm	0.623	0.638	0.598	0.606
The proposed algorithm	0.635	0.659	0.615	0.638

Table 5. Comparison table of model recognition accuracy with all discrete emotional features as input.

Model category	Accuracy	Precision	Recall	F1-Measure
DL algorithm	0.567	0.562	0.627	0.581
AI algorithm	0.604	0.618	0.626	0.631
DEC algorithm	0.649	0.667	0.668	0.677
The proposed algorithm	0.561	0.581	0.579	0.597

Table 6. Comparison table of model recognition with input features of continuous and discrete emotional features in a 1:1 ratio.

functions such as emotion analysis, audio feature extraction, and mood regulation into the app. This allows teachers and students to interact in real-time in a virtual learning environment. The feedback provided by the model dynamically adjusts according to student performance, thereby increasing students' interest in music learning and their aesthetic abilities.

Discussion

The research shows that the design of music interactive application with DL function can capture and analyze students' playing actions and performances in real time, provide immediate feedback and guidance, and promote the development of their music practical skills and aesthetic perception. Furthermore, DL has greatly improved the accuracy of evaluation and feedback in music aesthetic education. It can deeply analyze students' unstructured data (such as emotional reaction and attention state) in music activities through complex pattern

recognition and signal processing technology, and form an objective and detailed evaluation report. For specific singing or performance exercises, AI can accurately judge the nuances of pitch, rhythm and musical expression, and give specific suggestions for improvement in time.

DL is also applied to the generation and mining of innovative music education resources. Zhuang research uses AI music generation model to create new music materials, which not only enriches the teaching content, but also stimulates students' creative thinking. At the same time, through the intelligent screening and classification of massive music materials by DL technology, teachers can obtain and use the teaching materials that are most suitable for the curriculum objectives more conveniently. The construction of intelligent auxiliary teaching system and big data-driven teaching decision-making are also important ways for DL to optimize music aesthetic education[38]. Zhang researched and built an online education platform integrating DL module, which realized the automatic tutoring of students' music reading ability, auditory training and music theory knowledge, and significantly improved the teaching efficiency. Meanwhile, by analyzing the big data of music aesthetic education through DL, educators can obtain a macro view of teaching effectiveness, clarify the progress of student groups, and scientifically verify and optimize teaching strategies and methods[39]. In this paper, the combination of the two is applied to music aesthetic teaching, which not only greatly enriches the teaching methods and contents, but also effectively promotes the development of individualization, intelligence and effectiveness of music aesthetic education, making it closer to students' needs and more conducive to cultivating a new generation with high-level music aesthetic literacy.

Conclusion

Research contribution

The main contribution of this paper is to extract multimodal music features from music data, build its neural network model, and carry out emotion recognition and related application research. On this basis, this paper aims to extract various types of musical emotions from multiple perspectives. This paper seeks to integrate them with existing musical features and model musical emotions through preprocessing, feature extraction, model optimization, and other steps. The ultimate goal is to enhance the accuracy of musical emotion recognition. This paper also tests the speech recognition system based on AI and DL, which integrates the functions of emotion recognition and music production into the teaching of music aesthetics. Experiments show that this method can better realize the recognition of musical emotions and has broad application prospects in future music teaching. In real-world educational settings, the method proposed in this paper can ensure scalability and accessibility for schools or educators with limited technical resources through various strategies. First, the method can rely on cloud computing and edge computing technologies to deploy deep learning models in the cloud, allowing educational institutions to use intelligent music aesthetic teaching tools without the need for high-performance local computing devices. Second, by using lightweight deep learning models (such as knowledge distillation and pruning techniques) to optimize computational efficiency, core functions can be run on low-power devices like regular PCs, tablets, and even smartphones. Third, standardized music datasets and pre-trained models can be provided through open-source platforms and shared databases, reducing the technical barriers for schools to build their own data and train AI models. Fourth, the method involves combining modular teaching tools with tiered AI functions. This allows for the provision of different levels of AI-assisted functions based on the infrastructure conditions of individual schools. The range of functions includes simple music recommendations as well as more advanced capabilities such as emotion recognition and personalized teaching adjustments. This ensures that educators of all types can flexibly apply this method.

In addition, the implementation strategies for personalized teaching in a deep learning-based music aesthetic education model in an AI environment are as follows: First, the system can analyze students' emotional recognition results based on the deep learning model. For example, when it detects that a student shows positive emotions towards a certain music style, it can enhance the teaching content related to that style. Conversely, for music types that students find difficult to understand or have less interest in, the teaching method can be adjusted, such as by increasing interactive experiences or using different explanation methods. Second, intelligent data feedback can be used for personalized teaching path planning. For example, by predicting students' suitable learning pace based on their learning performance, the system can adjust music work recommendations, practice difficulty, or evaluation criteria to ensure that teaching meets students' ability levels and promotes the improvement of their aesthetic literacy. Third, the system can also use group data analysis to identify learning patterns of different types of students and provide teachers with suggestions for teaching optimization, achieving precise teaching that combines intelligence and human effort. Through the implementation of these strategies, students can obtain music aesthetic education that better meets their personal needs, thereby increasing their interest in learning, enhancing their music perception abilities, and promoting the cultivation of personalized aesthetic literacy.

Future works and research limitations

The emotion recognition method of DL music aesthetic teaching under the background of AI constructed in this paper has achieved good results, but there are still some shortcomings. The constructed model is complex in structure, too dependent on the original data scale and the accuracy of music emotion labeling, and has some problems such as weak adaptive ability. Therefore, how to realize the lightweight promotion of the model and make it have better adaptive ability in a low sample environment should be studied. The proportion used in this paper needs to be further improved and optimized.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author Yixuan Peng on reasonable request via e-mail pengyixuan0426@163.com.

Received: 16 September 2024; Accepted: 28 March 2025
Published online: 04 April 2025

References

1. Li, F. Chord-based music generation using long short-term memory neural networks in the context of artificial intelligence[J]. *J. Supercomputing*. **80** (5), 6068–6092 (2024).
2. Li, P. & Wang, B. Artificial intelligence in music education. *Int. J. Human–Computer Interact.* **40**(16), 4183–4192 (2023).
3. Zhang, L. Z. L. Fusion artificial intelligence technology in music education Teaching[J]. *J. Electr. Syst.* **19** (4), 178–195 (2023).
4. Paroiu, R. & Trausan-Matu, S. Measurement of music aesthetics using deep neural networks and Dissonances[J]. *Information* **14** (7), 358 (2023).
5. Iffath, F. & Gavrilova, M. RAIF: A deep learning-based architecture for multi-modal aesthetic biometric system. *Comput. Animat. Virtual Worlds* **34**(3–4), e2163 (2023).
6. Ji, S., Yang, X. & Luo, J. A survey on deep learning for symbolic music generation: representations, algorithms, evaluations, and challenges[J]. *ACM Comput. Surveys*. **56** (1), 1–39 (2023).
7. Oruganti, R. K. et al. Artificial intelligence and machine learning tools for high-performance microalgal wastewater treatment and algal biorefinery: A critical review[J]. *Sci. Total Environ.* **876**, 162797 (2023).
8. Yin, Z. et al. Deep learning's shallow gains: A comparative evaluation of algorithms for automatic music generation[J]. *Mach. Learn.* **112** (5), 1785–1822 (2023).
9. Gomes, B. & Ashley, E. A. Artificial intelligence in molecular medicine[J]. *N. Engl. J. Med.* **388** (26), 2456–2465 (2023).
10. Soori, M., Arezoo, B. & Dastres, R. Machine learning and artificial intelligence in CNC machine tools, a review[J]. *Sustainable Manuf. Service Econ.* **2**, 100009 (2023).
11. Bharadiya, J. P. A comparative study of business intelligence and artificial intelligence with big data analytics[J]. *Am. J. Artif. Intell.* **7** (1), 24 (2023).
12. Moysis, L. et al. Music deep learning: deep learning methods for music signal processing-a review of the state-of-the-art. *IEEE Access* **11**, 17031–17052 (2023).
13. Shukla, S. Creative computing and Harnessing the power of generative artificial Intelligence[J]. *J. Environ. Sci. Technol.* **2** (1), 556–579 (2023).
14. Oksanen, A. et al. Artificial intelligence in fine arts: A systematic review of empirical research. *Comput. Hum. Behav. Artif. Hum.* **20**, 100004 (2023).
15. Borkowski, A., Vocal Aesthetics, A. I. & Imaginaries *Reconfiguring Smart Interfaces[J] Afterimage*, **50**(2): 129–149. (2023).
16. Cui, K. Artificial intelligence and creativity: piano teaching with augmented reality applications[J]. *Interact. Learn. Environ.* **31** (10), 7017–7028 (2023).
17. Huang, X. et al. *Trends, Research Issues and Applications of Artificial Intelligence in Language education[J]*26112–131 (Educational Technology & Society, 2023). 1.
18. Rathore, B. Digital transformation 4.0: integration of artificial intelligence & metaverse in marketing[J]. *Eduzone: Int. Peer Reviewed/Refereed Multidisciplinary J.* **12** (1), 42–48 (2023).
19. Liu, B. Arguments for the rise of artificial intelligence art: does AI Art have creativity, motivation, Self-awareness and Emotion?[J]. *Arte Individuo Y Sociedad*. **35** (3), 811 (2023).
20. Kim, T. W. Application of artificial intelligence chatbot, including ChatGPT in education, scholarly work, programming, and content generation and its prospects: a narrative review[J]. *J. Educational Evaluation Health Professions*. **20**, 38 (2023).
21. Wang, H. et al. Scientific discovery in the age of artificial intelligence[J]. *Nature* **620** (7972), 47–60 (2023).
22. Liu, W. Literature survey of multi-track music generation model based on generative confrontation network in intelligent composition[J]. *J. Supercomputing*. **79** (6), 6560–6582 (2023).
23. Mosqueira-Rey, E. et al. Human-in-the-loop machine learning: a state of the art[J]. *Artif. Intell. Rev.* **56** (4), 3005–3054 (2023).
24. Alkayyali, Z. K., Idris, S. A. B. & Abu-Naser, S. S. A systematic literature review of deep and machine learning algorithms in cardiovascular diseases Diagnosis[J]. *J. Theoretical Appl. Inform. Technol.* **101** (4), 1353–1365 (2023).
25. Tufail, S. et al. Advancements and challenges in machine learning: A comprehensive review of models, libraries, applications, and algorithms[J]. *Electronics* **12** (8), 1789 (2023).
26. Ahmadi, S. Optimizing data warehousing performance through machine learning algorithms in the Cloud[J]. *Int. J. Sci. Res. (IJSR)*. **12** (12), 1859–1867 (2023).
27. Sonkavde, G. et al. Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications[J]. *Int. J. Financial Stud.* **11** (3), 94 (2023).
28. Paturi, U. M. R., Palakurthy, S. T. & Reddy, N. S. The role of machine learning in tribology: a systematic review[J]. *Arch. Comput. Methods Eng.* **30** (2), 1345–1397 (2023).
29. Zhang, J. et al. Public cloud networks oriented deep neural networks for effective intrusion detection in online music education[J]. *Comput. Electr. Eng.* **115**, 109095 (2024).
30. Zhang, W., Shankar, A. & Antonidoss, A. Modern Art education and teaching based on Artificial intelligence[J]. *J. Interconnect. Networks*. **22** (Supp01), 2141005 (2022).
31. Dai, D. D. Artificial intelligence technology assisted music teaching design[J]. Scientific programming, 2021: 1–10. (2021).
32. Ferreira, P., Limongi, R. & Fávero, L. P. Generating music with data: application of deep learning models for symbolic music Composition[J]. *Appl. Sci.* **13** (7), 4543 (2023).
33. Saberi-Movahed, F. et al. Deep metric learning with soft orthogonal Proxies[J]. arXiv preprint arXiv:2306.13055, 2023.
34. Dan, X. Social robot assisted music course based on speech sensing and deep learning algorithms[J]. *Entertainment Comput.* **52**, 100814 (2025).
35. Fang, J. Artificial intelligence robots based on machine learning and visual algorithms for interactive experience assistance in music classrooms. *Entertainment Comput.*, **52**, 100779 (2025).
36. Li, X. et al. Learning a convolutional neural network for propagation-based stereo image segmentation[J]. *Visual Comput.* **36**, 39–52 (2020).
37. Zhang, X. et al. Multi-level fusion and attention-guided CNN for image dehazing[J]. *IEEE Trans. Circuits Syst. Video Technol.* **31** (11), 4162–4173 (2020).
38. Zhuang, X. et al. Artificial multi-verse optimisation for predicting the effect of ideological and political theory course. *Heliyon* **10**(9), 1–14 (2024).
39. Zhang, L. et al. Bioinspired scene classification by deep active learning with remote sensing applications[J]. *IEEE Trans. Cybernetics*. **52** (7), 5682–5694 (2021).

Author contributions

Yixuan Peng: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

The studies involving human participants were reviewed and approved by College of Music, The Yeungnam University of Korea Ethics Committee (Approval Number: 2021.02039345). The participants provided their written informed consent to participate in this study. All methods were performed in accordance with relevant guidelines and regulations.

Additional information

Correspondence and requests for materials should be addressed to Y.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.