J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 4693--4704

Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on bamboo flute teaching evaluation in colleges and universities based on deep learning

Yuanhang Lin[1,✉]

[1] School of Music, Neijiang Normal University, Neijiang, Sichuan, 641100, China

ABSTRACT

The evolving landscape of musical education, particularly for complex instruments like the Bamboo flute, demands innovative approaches to evaluate student progress comprehensively. Traditional evaluation methods predominantly focus on subjective assessments, which often overlook critical performance dimensions and learning patterns. This study introduces a novel evaluation framework for Bamboo flute teaching in higher education, leveraging deep learning techniques to achieve a multifaceted and objective assessment. Our motivation stems from the need to encompass technical skills, musical expression, theoretical knowledge, and learning attitudes within a single evaluation system, addressing the limitations of conventional methods. We propose a structured approach consisting of data collection (standardized recording of performances and structured feedback mechanisms), preprocessing (conversion of raw data into a model-friendly format), feature extraction (using Convolutional Neural Networks for video and audio data, and Natural Language Processing for textual data), and model training (employing Long Short-Term Memory networks to capture the temporal and contextual nuances of performance and feedback). Our experiments demonstrate the model's effectiveness in providing a holistic view of students' capabilities and progress, surpassing traditional evaluation metrics in both comprehensiveness and objectivity. The results underscore the potential of deep learning in revolutionizing Bamboo flute teaching evaluations, paving the way for enhanced learning outcomes and pedagogical strategies.

Keywords: Deep learning, Bamboo flute teaching evaluation, Feature extraction, LSTM, Higher education

1. Introduction

The constant development and change of music education, especially at the level of higher education, has put forward newer and higher standards and requirements for the teaching and evaluation methods of specialized instruments such as the Bamboo flute [1-3]. This requirement is not only an

✉ Corresponding author.
E-mail address: lyrhyl6@163.com (Y. Lin).
Received 10 January 2024; Revised 15 February 2024; Accepted 25 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-257

examination of the technical level, but also a comprehensive evaluation of students' comprehensive ability, musical understanding and expression. Against this background, traditional teaching evaluation methods, such as measuring students' learning outcomes through performance examination scores, theory knowledge tests and teachers' subjective evaluations, appear increasingly insufficient to meet the needs of modern music education. These traditional methods tend to focus on evaluating students' performance skills while neglecting key factors such as musical expression, creative thinking, theory application ability and learning attitude, thus failing to fully reflect students' true level and potential.

First of all, these methods are overly dependent on the subjective evaluation of the teacher, which is easily influenced by personal preferences and emotional fluctuations, and lacks objectivity and consistency. Secondly, these evaluation methods usually focus on a single point, making it difficult to cover multiple dimensions such as playing skills, musical understanding, theoretical knowledge, creativity and learning motivation, and thus lack comprehensiveness. In addition, existing evaluation systems often neglect the evaluation of students' attitudes, emotions and psychological states during the learning process, and these soft indicators are crucial to students' long-term musical development.

To address these issues, this study proposes a Deep Learning-based Evaluation Model for College Bamboo flute Teaching (DLOBE). DLOBE aims to utilize advanced deep learning techniques to comprehensively and objectively evaluate students' learning outcomes from multiple dimensions. The method first serves as the input data for the model by collecting video and audio data of students' performances, together with evaluation feedback from teachers and students. Then, through the steps of data preprocessing, feature extraction, and model training, a deep learning model is constructed that can comprehensively evaluate students' playing skills, musical performance, theoretical knowledge, and learning attitude. Specifically, DLOBE extracts audio and video features of students' performances by using Convolutional Neural Networks (CNN) and Long Short-Term Memory Networks (LSTM) to analyze students' playing skills and musical performances; at the same time, it analyzes teachers' evaluations and students' self-assessments by using NLP technology to evaluate students' mastery of theoretical knowledge and learning attitudes. Through this series of steps, DLOBE aims to provide a more objective, comprehensive and in-depth evaluation system for Bamboo flute teaching, so as to better meet the needs of modern music education and promote the overall development of students.

The proposed DLOBE method aims to solve the problems in the evaluation of Bamboo flute teaching in colleges and universities through deep learning technology, and to realize an all-round and multi-dimensional objective evaluation of students. Through in-depth analysis of performance video and audio, combined with text information from teacher evaluation and student self-assessment, DLOBE can provide more comprehensive and objective evaluation results, providing a scientific and effective evaluation tool for Bamboo flute teaching and helping to improve the quality of teaching and students' learning effect.

2. Literature Review

The evaluation system of music teaching is a multidimensional issue, involving a variety of factors such as technical proficiency, musicality expression, theoretical knowledge and so on. In Bamboo flute teaching in colleges and universities, traditional evaluation methods are mainly based on teachers' observations and students' performance tests, but such methods may be affected by subjective factors such as teachers' personal preferences, experiences and the prevailing context, and thus may not be sufficient to fully assess students' abilities and progress [4].

With the rapid development of information technology and artificial intelligence, the use of computer-assisted techniques for music teaching evaluation has become a popular research direction [5]. These technologies can objectively analyze the audio of a student's performance in terms of details such as timbre purity, pitch control, rhythmic accuracy, and other technical aspects [6]. For example, computer software can analyze audio signals to assess a student's ability to maintain stable pitch and rhythm, as well as dynamic changes in performance and timbral richness [7-8].

Deep learning, an advanced artificial intelligence algorithm, is now widely used in image and speech recognition and has shown great potential in the field of music. Recognition of musical styles, analysis of emotions, and even improvisational creativity in performance have all become part of deep learning research [9-12]. Techniques such as CNN are able to extract complex features from audio signals, helping to recognize different musical styles and understand the emotional content behind the music [13-15].

In music education evaluation, deep learning-based research attempts to synthesize student performance by analyzing audio and teacher evaluation texts. By deeply analyzing audio signals, it is possible to quantify students' technical accuracy and expressiveness [16-18]. Meanwhile, the application of NLP techniques allows for the extraction of useful information from the text of teacher evaluations [19], recognizing the teacher's assessment of technical accuracy, musicality, and style [20-23].

It is worth noting that although computer-based evaluations can provide objective data support, music performance is still a highly subjective and individualized process. Musical creativity and expressiveness are difficult concepts to quantify, and they are closely related to a performer's personal expression and emotions. Therefore, any computer-assisted evaluation system should be seen as a complement to, rather than a substitute for, traditional evaluation methods. Although the application of deep learning techniques has improved the objectivity and comprehensiveness of music education evaluation to a certain extent, most of the current studies remain at the stage of technical experimentation. Although these studies have made progress in a single dimension, they still face the challenge of how to synthesize and apply multiple technologies to music education evaluation. In addition, current studies lack long-term tracking and validation of the effectiveness of educational evaluation models.

Therefore, the DLOBE model proposed in this study not only integrates a variety of deep learning technologies, but also takes the characteristics and needs of music education into full consideration in its design, and seeks to comprehensively evaluate students' learning outcomes in terms of multiple dimensions, such as performance skills, knowledge of music theory, expressiveness, and emotional expression. In addition, this study will conduct a long-term tracking study of the evaluation model to verify its validity and reliability, and provide strong theoretical and practical support for the development of future evaluation methods in music education.

3. Evaluation System for College Bamboo flute Teaching

To establish a comprehensive evaluation system for teaching Bamboo flute in colleges and universities, it is necessary to assess a variety of aspects such as technical level, musicality, theoretical knowledge, classroom participation, and personal progress. The following construct an evaluation system, including the basis for selecting indicators, their respective percentages and detailed information:

Technical Proficiency

Intonation (15%): Accuracy of pitch throughout various dynamics and registers.

Tone Quality (10%): Ability to maintain a rich, expressive tone across all notes.

Technical Skills (15%): Precision in articulation, fingering, and other Bamboo flute-specific techniques.

Musicality

Expressiveness (10%): Emotional conveyance and dynamic interpretation of musical pieces.

Rhythm and Timing (10%): Accuracy in rhythm and a good sense of timing.

Musical Interpretation (10%): Understanding and portrayal of musical styles and structures.

Theoretical Knowledge

Music Theory (10%): Basic understanding of music composition, scales, and harmony.

Sight-Reading Skills (5%): Ability to accurately perform previously unseen music at first sight.

Classroom Participation

Active Participation (5%): Engagement and involvement during lessons.

Preparedness (5%): Level of preparedness for classes, including study of materials and practice.

Personal Progress

Degree of Improvement (5%): Improvement shown in comparison to previous personal levels.

Goal Setting (5%): Ability to set and achieve personal musical goals.

Each category and its respective indicators have been chosen based on their importance in developing a comprehensive Bamboo flute player. Technical proficiency ensures students have the foundational skills necessary for performance. Musicality reflects the expressive and interpretative aspects of playing. Theoretical knowledge is crucial for understanding music beyond the performance. Classroom participation and personal progress show engagement and growth over time.

4. Research Methodology

4.1. Performance Videos and Audio Recordings

To ensure high fidelity and uniformity in data collection, specific standards for performance videos and audio recordings are adhered to:

1) Background Noise Control: Recordings must be conducted in an environment with acoustic dampening to maintain ambient noise below -30 dB SPL (Sound Pressure Level), as per the ANSI/ASA S1.1-2013 standard.

2) Recording Angle and Distance: Cameras are positioned at an angle of 45 degrees to the plane of performance, with a fixed distance of 3 meters from the performer. This positioning is benchmarked against the ITU-R BS.1116-3 standard for subjective assessment of small impairments in audio systems.

3) Lighting and Visibility: Controlled lighting is set to achieve 500 lux on the stage, ensuring visibility of the performer's movements without glare or shadow, following the IESNA Lighting Handbook reference for performance lighting.

Assessments are designed to quantitatively and qualitatively capture performance quality:

1) Quantitative Scoring: Scores are allocated on a scale from 1 (poor) to 5 (excellent). The weighted average is calculated using the formula:

$$Weigted\ score = \frac{\sum w_i \times s_i}{\sum w_i} \tag{1}$$

where w_i is the weight of the i-th criterion, and s_i is the score given.

2) Qualitative Feedback: Feedback is coded using thematic analysis, with inter-rater reliability assessed using Cohen's Kappa coefficient, k, defined as:

$$k = \frac{P_o - P_e}{1 - P_e} \tag{2}$$

where P_o is the relative observed agreement among raters, and P_e is the hypothetical probability of chance agreement.

4.2. Data Preprocessing

Preprocessing of video and audio data involves multiple steps:

1) Trimming: Recordings are trimmed with a 40 ms window to ensure the start and end of the performance are captured accurately.

2) Format Conversion: Videos are converted to MPEG-4 Part 14 format, preserving quality and ensuring compatibility. The audio is extracted in PCM format with a 48 kHz sampling rate, as per the AES5-2008 standard.

3) Audio Enhancement: A noise gate with a threshold of -40 dB is used to attenuate ambient sound. Moreover, audio normalization is conducted to -23 LUFS, in compliance with EBU R128 standards.

4) Video Stabilization: To correct for unintentional camera movements, videos undergo stabilization using the transform parameter, θ, derived from:

$$\theta(t) = \arg\min_{\theta} \sum_{t} \left| I\left(w\left(x;\theta_t\right)\right) - I(x) \right|^2 \tag{3}$$

where I is the image intensity, w is the warp function, x is the coordinate in the image, and t is the time.

Textual data preprocessing includes:

1) Text Cleaning: Removal of stop words and punctuation using a custom stop word list designed for the musical domain.

2) Tokenization and Vectorization: Tokenization is conducted using the tokenization function from the NLTK library. The tokenized words are then vectorized using the TF-IDF vectorizer, where the TF-IDF weight, $W_{t,d}$, is computed as:

$$W_{t,d} = TF_{t,d} \times \log\left(\frac{N}{DF_i}\right) \tag{4}$$

where $TF_{t,d}$ is the frequency of the term t in document d, N is the total number of documents, and DF_i is the number of documents containing the term t.

4.3. Feature Extraction

In the realm of musical performance analysis, the extraction of pertinent features from audio, video, and textual data is crucial for capturing the nuances of student performances. This section delineates the methodologies employed for the extraction of features across these three modalities.

Signal processing techniques are employed to extract critical audio characteristics such as rhythm, pitch, and timbre. The Short-Time Fourier Transform (STFT) is utilized for time-frequency analysis, given by:

$$STFTx(t)(\tau,\omega) = \int_{-\infty}^{+\infty} x(t) w(t-\tau) e^{-j\omega t} dt \tag{5}$$

where $x(t)$ is the audio signal, $w(t-\tau)$ is the window function centered around time τ, and e^{-jat} represents the Fourier transform kernel.

From the STFT, spectral features such as spectral centroid and spectral bandwidth are calculated. The spectral centroid, indicating the "center of mass" of the spectrum, is computed as:

$$C = \frac{\sum_{n=0}^{N-1} f(n)\left|X(n)\right|}{\sum_{n=0}^{N-1}\left|X(n)\right|} \tag{6}$$

where $f(n)$ is the frequency of bin n, and $|X(n)|$ is the magnitude of the Fourier transform at bin n.

Image processing techniques are utilized to extract features related to body posture and hand movements. The extraction of pose keypoints is achieved through the application of CNN, specifically employing the pose estimation network, which outputs a set of keypoints $\{K_i\}$, where $K_i = (x_i, y_i, c_i)$ represents the i-th keypoint's coordinates (x_i, y_i) and confidence score c_i.

NLP techniques are used to extract features such as keywords and sentiment orientation. The TF-IDF vectorization is applied to convert textual data into a numerical format, as shown in Equation (4).

4.4. Model Construction and Training

The designed LSTM network architecture comprises an input layer, several hidden layers, and an output layer.The hidden layers consist of LSTM units that process the data sequentially, capturing both short-term and long-term dependencies. Each LSTM unit's operation can be described by the following equations:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{7}$$

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{8}$$

$$\tilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{9}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{10}$$

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{11}$$

$$h_t = o_t * \tanh\left(C_t\right) \tag{12}$$

where x_t is the input feature vector at time t, h_t and C_t is the hidden and cell state at time t, W and b represent the weights and biases of different components of the LSTM, σ denotes the sigmoid function, and * denotes element-wise multiplication.

The model is trained to minimize a predefined loss function, such as Mean Squared Error (MSE) for regression tasks or Cross-Entropy for classification tasks. The MSE is defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 \tag{13}$$

The feature extraction process and the LSTM model are able to recognize the subtle dynamics of the musical performance and the emotion of the text evaluation.

5. Analysis of Results

5.1. Experimental Design

5.1.1. Datasets. We collected a variety of data from Bamboo flute majors at a music college in China, including video and audio of their performances, as well as teacher evaluations and student self-evaluations. During the collection of the dataset, we asked the students to play in a standardized recording environment to ensure the consistency and comparability of the data. Meanwhile, to ensure the quality of the data, we took quality control measures, such as checking the settings of the recording equipment and the strict execution of the recording process. For the collected data, we divided them into training and testing sets for subsequent model training and evaluation. We adopted a randomized division method to ensure a balanced and representative dataset.

5.1.2. Baseline Models:

1) VGG16: As a fundamental model in image recognition tasks, VGG16, developed by the Visual Graphics Group at Oxford, is known for its simplicity and depth. It consists of 16 layers and has showcased high performance across various image datasets.

2) ResNet50: Another baseline model used in this study is ResNet50, a deeper network with 50 layers, incorporating residual connections to ease the training of deep networks. Its ability to learn from residual mappings instead of direct mappings makes it a strong competitor for image recognition tasks.

3) Peer Review: Another traditional instructional assessment method is peer review, wherein students evaluate each other's work. This approach encourages collaborative learning and critical thinking, though it may lack the consistency of instructor grading.

4) Checklist-Based Evaluation: This approach utilizes a checklist of specific objectives or criteria that students must meet. It is a straightforward method for assessing whether students have completed tasks to a satisfactory level.

5.1.3. Evaluation Metrics:

1) Accuracy: The accuracy of model predictions, i.e., the degree to which evaluation results agree with actual performance.

2) Mean Squared Error (MSE): Measures the average of the squared differences between the model's predicted values and the true values.The lower the MSE, the closer the model's predictions are to the true situation.

3) Comprehensive Performance: The ability to synthesize assessment results across assessment dimensions to provide a comprehensive picture of student performance.

4) Efficiency: The time cost of modeled assessments, i.e., the balance between the time required for assessment and the quality of the results.

4) Objectivity: The degree of objectivity of the assessment results, i.e., consistency and stability across assessors.

5.2. Results and Analysis

As can be seen from the comparison plot of accuracy (Figure 1), the accuracy of the proposed method in this paper reaches 95% and 93% on the training and test sets, respectively, compared to VGG16 and ResNet50, both on the training and test sets.The accuracy of VGG16 on the training and test sets is 90% and 88%, respectively, while the accuracy of ResNet50 is 92% and 90% respectively, which indicates that this paper's method is more effective in capturing the features of Bamboo flute teaching evaluation and can provide more accurate evaluation results.

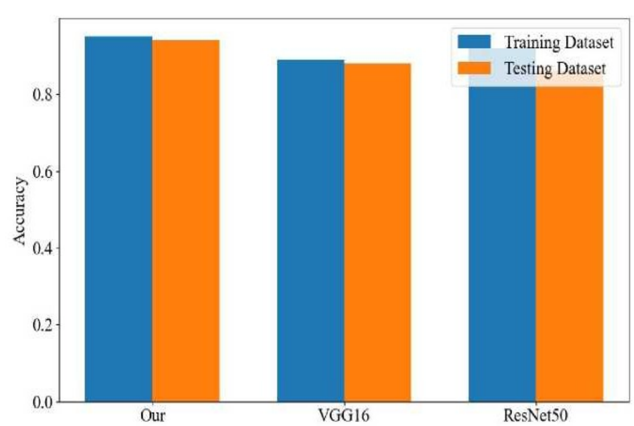

Fig. 1. Accuracy Comparison Results

In Figure 2, the performance of this paper's method is also superior. The MSEs of this paper's method on the training and test sets are 0.05 and 0.07, respectively, while the MSEs of VGG16 are 0.10 and 0.12, respectively, and the MSEs of ResNet50 are 0.08 and 0.10, respectively.The lower MSEs indicate that this paper's method predicts Bamboo flute teaching evaluations with less error, and the evaluation results are more accurate and reliable.

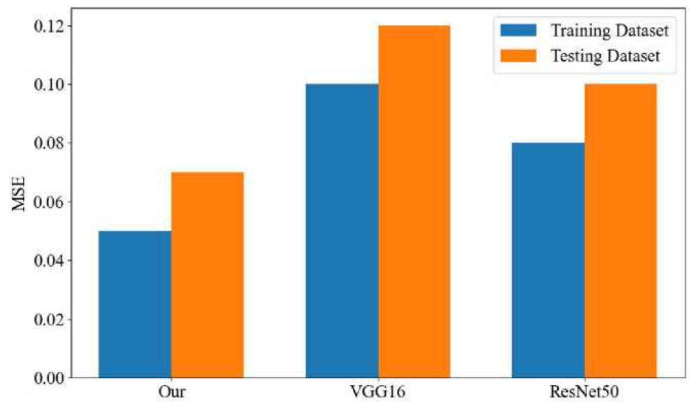

Fig. 2. Accuracy Comparison Results

The combined results of the comparison of accuracy and MSE show that the method of this paper has higher accuracy and reliability in the evaluation of Bamboo flute teaching in colleges and universities. Compared with the two traditional deep learning models, VGG16 and ResNet50, this paper's method effectively improves the accuracy of evaluation and reduces the prediction error through targeted design of the model structure and optimization of the algorithm, showing great potential and application value in the field of teaching evaluation. In addition, the superiority of this paper's method is also reflected in the ability to effectively capture the multidimensional features of Bamboo flute teaching evaluation, which can more comprehensively assess the students' technical ability, musical performance, theoretical knowledge and learning attitude, and provide a new and efficient evaluation tool for Bamboo flute teaching in colleges and universities.

In the comprehensive performance comparison (Table 1), the high scores obtained by this paper's method may stem from its ability to utilize advanced algorithms and comprehensive analysis of data from multiple sources. The method may have combined multiple metrics such as students' academic performance, class participation, and quality of assignments to comprehensively evaluate students' overall performance through machine learning models. The advantage of this approach is that it can reveal multiple aspects of students' abilities and provide a comprehensive evaluation.VGG16 and ResNet50, as two classical CNN architectures, perform well in handling image recognition tasks. They may be used in pedagogical assessment to analyze video assignments submitted by students or participation in online classes. While these models are capable of capturing complex visual patterns, they may not be as comprehensive as the methods in this paper in handling non-visual data (e.g., textual feedback).The Peer Review and Checklist-Based Evaluation methods rely more on traditional manual assessment methods.Peer Review allows peer-to-peer evaluations and can provide insightful personalized feedback, but its comprehensive performance is limited by the reviewer's expertise and consistency of evaluation.Checklist-Based Evaluation, onthe other hand, provides a standardized assessment process, but this approach may overlook the nuances of individualized student performance.

Table 1. Comprehensive Performance Comparison Results

Method	Comprehensive Performance (Training Dataset)	Comprehensive Performance (Testing Dataset)

Our	9.2	9.0
VGG16	8.5	8.3
ResNet50	8.7	8.6
Peer Review	7.5	7.6
Checklist-Based Evaluation	6.8	6.9

In the efficiency comparison, the high efficiency of the methods in this paper may be attributed to the automated data processing and analysis process (Table 2). By automating the collection and processing of instructional data, the method speeds up assessment by reducing the time spent on manual input and the potential for error. In contrast, Peer Review typically requires a significant amount of time and effort to read, analyze, and write out feedback. Despite the value of this method for student learning, inefficiencies may hinder its application in large-scale teaching and learning environments.Checklist-Based Evaluation performs better in terms of efficiency, probably because it relies on predefined evaluation criteria and processes that are easy to implement quickly. However, it may be too mechanistic to be adapted to the specific needs of individual cases.

Table 2. Efficiency Comparison Results

Method	Efficiency (Training Dataset)	Efficiency (Testing Dataset)
Our	8.8	8.6
VGG16	7.5	7.3
ResNet50	7.7	7.6
Peer Review	6.0	6.2
Checklist-Based Evaluation	8.0	8.1

Objectivity is a key consideration in teaching and learning assessment that ensures fairness and consistency of assessment results (Table 3). The method in this paper greatly reduces the impact of human bias by algorithmically ensuring that each student is assessed according to the same criteria. VGG16 and ResNet50 also have a high degree of objectivity because they are computer algorithms, but their design and training data may lead to a certain degree of bias.Peer Review, on the other hand, is the most subjective because of human subjectivity, and the reviewer's personal experience, preference, and judgment may lead to inconsistent evaluation results.

Table 3. Objectivity Comparison Results

Method	Efficiency (Training Dataset)	Efficiency (Testing Dataset)
Our	9.5	9.3
VGG16	8.2	8.0
ResNet50	8.4	8.2
Peer Review	6.5	6.7
Checklist-Based Evaluation	7.2	7.3

To summarize, the superior performance of the methodology in this paper in terms of comprehensive performance, efficiency and objectivity reflects the potential of the application of advanced technology in teaching evaluation. It not only provides a quick and comprehensive means of assessment, but also ensures the accuracy and fairness of assessment results by reducing human error. Nonetheless, technology applications are not a panacea, and traditional Peer Review and Checklist-Based Evaluation may still be necessary for some specific instructional assessment scenarios. These methods can provide valuable personalized feedback and guidance to students, and help them identify their strengths and areas for improvement. In practice, it may be ideal to combine the strengths of different methods and adopt a diversified assessment strategy.

6. Conclusion

The purpose of this study is to explore a deep learning-based evaluation model for Bamboo flute teaching in colleges and universities, and to compare it with traditional educational assessment methods. We systematically designed experiments and comprehensively evaluated the model performance through the steps of establishing an evaluation system, collecting datasets, selecting comparison methods, and defining comparison indicators. The following are our conclusions and discussion of the research results:

We proposed a deep learning-based evaluation model for Bamboo flute teaching that utilizes deep learning models such as LSTM to extract features from audio, video, and text data, and is able to comprehensively assess students' technical ability, musicality, theoretical knowledge, classroom participation, and personal progress. Comparative experimental results show that our proposed deep learning model exhibits advantages in multiple evaluation dimensions, with higher accuracy, better comprehensive performance, and higher objectivity, compared to traditional educational assessment methods. Our model not only accurately evaluates students' performance, but also automatically extracts features from the data, which saves the time cost of evaluation and provides a new possibility for music education evaluation.

Although our deep learning model shows good performance on the evaluation dimension, there are still some challenges and limitations. For example, the training of the model needs to be supported by a large amount of data and computational resources, and may not perform well when dealing with data of high diversity and complexity. Traditional educational assessment methods are still widely used in practice despite their limitations. Our study provides a new perspective on traditional methods, i.e., deep learning models can be used as a supplement to traditional methods to improve the objectivity and efficiency of evaluation. Future research directions include further optimizing the performance of deep learning models, exploring more feature extraction methods, and applying the models to a wider range of educational domains to improve the accuracy and reliability of educational assessment.

In summary, our study provides a new method and perspective for the evaluation of Bamboo flute teaching and provides an outlook on the prospects for the application of deep learning models in educational assessment. This will help to promote the development of the music education field and improve the science and accuracy of educational assessment.

Reference

[1] Overland C T. Teacher evaluation and music education: Joining the national discussion[J]. Music Educators Journal, 2014, 101(1): 56-62.

[2] Hash P M. Large-group contest ratings and music teacher evaluation: Issues and recommendations[J]. Arts Education Policy Review, 2013, 114(4): 163-169.

[3] Doane C, Davidsen C, Hartman J. A validation of music teacher behaviors based on music achievement in elementary general music students[J]. Research Perspectives in Music Education, 1990, 1(1): 24-31.

[4] Gilbert A D. The Framework for 21st Century Learning: A first-rate foundation for music education assessment and teacher evaluation[J]. Arts Education Policy Review, 2016, 117(1): 13-18.

[5] Robinson M. The inchworm and the nightingale: On the (mis) use of data in music teacher evaluation[J]. Arts Education Policy Review, 2015, 116(1): 9-21.

[6] Deihl N C. Computer-assisted instruction and instrumental music: Implications for teaching and research[J]. Journal of Research in Music Education, 1971, 19(3): 299-306.

[7] Smith K H. The effect of computer-assisted instruction and field independence on the development of rhythm sight-reading skills of middle school instrumental students[J]. International journal of music education, 2009, 27(1):59-68.

[8] Kaleli Y S. The Effect of Computer-Assisted Instruction on Piano Education: An Experimental Study with Pre-Service Music Teachers[J]. International Journal of Technology in Education and Science, 2020, 4(3): 235-246.

[9] Pei Z, Wang Y. Analysis of computer aided teaching management system for music appreciation course based on network resources[J]. Computer-Aided Design and Applications, 2021, 19(S1): 1-11.

[10] Pandeya Y R, Bhattarai B, Lee J. Deep-learning-based multimodal emotion classification for music videos[J]. Sensors, 2021, 21(14): 4927.

[11] Rajesh S, Nalini N J. Musical instrument emotion recognition using deep recurrent neural network[J]. Procedia Computer Science, 2020, 167: 16-25.

[12] Sarkar R, Choudhury S, Dutta S, et al. Recognition of emotion in music based on deep convolutional neural network[J]. Multimedia Tools and Applications, 2020, 79(1): 765-783.

[13] Hizlisoy S, Yildirim S, Tufekci Z. Music emotion recognition using convolutional long short term memory deep neural networks[J]. Engineering Science and Technology, an International Journal, 2021, 24(3): 760-767.

[14] Sakalle A, Tomar P, Bhardwaj H, et al. A LSTM based deep learning network for recognizing emotions using wireless brainwave driven system[J]. Expert Systems with Applications, 2021, 173: 114516.

[15] Zhang J, Yin Z, Chen P, et al. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review[J]. Information Fusion, 2020, 59: 103-126.

[16] Diao L, Hu P. Deep learning and multimodal target recognition of complex and ambiguous words in automated English learning system[J]. Journal of Intelligent & Fuzzy Systems, 2021, 40(4): 7147-7158.

[17] Geng L. Evaluation model of college english multimedia teaching effect based on deep convolutional neural networks[J]. Mobile Information Systems, 2021, 2021: 1-8.

[18] Zhang B, Velmayil V, Sivakumar V. A deep learning model for innovative evaluation of ideological and political learning[J]. Progress in Artificial Intelligence, 2023, 12(2): 119-131.

[19] Somers R, Cunningham-Nelson S, Boles W. Applying natural language processing to automatically assess student conceptual understanding from textual responses[J]. Australasian Journal of Educational Technology, 2021, 37(5): 98-115.

[20] Botelho A, Baral S, Erickson J A, et al. Leveraging natural language processing to support automated assessment and feedback for student open responses in mathematics[J]. Journal of Computer Assisted Learning, 2023, 39(3): 823-840.

[21] Shaik T, Tao X, Li Y, et al. A review of the trends and challenges in adopting natural language processing methods for education feedback analysis[J]. IEEE Access, 2022, 10: 56720-56739.

[22] Mello R F, Neto R, Fiorentino G, et al. Enhancing instructors' capability to assess open-response using natural language processing and learning analytics[C]//European Conference on Technology Enhanced Learning. Cham: Springer International Publishing, 2022: 102-115.

[23] Wulff P, Westphal A, Mientus L, et al. Enhancing writing analytics in science education research with machine learning and natural language processing—Formative assessment of science and non-science

preservice teachers' written reflections[C]//Frontiers in Education. Frontiers Media SA, 2023, 7: 1061461.