International Journal of Web-Based Learning and Teaching Technologies
Volume 18 • Issue 2

Application of Machine Learning Technology in Classical Music Education

Dongfang Wang, XinXiang University, China*

ABSTRACT

The goal is to promote the healthy and stable development of music education in China. The time-frequency sequence topology in frequency domain can improve the effect of convolution operation. Therefore, this paper applies the above algorithms to classical music education, including the recognition of classical instruments, the feature extraction and recognition of classical music, and the quality evaluation of classical music education. The quality of the music quality evaluation system can be judged according to the correlation between the output results and the subjective evaluation. The higher the correlation, the better the music quality evaluation method. Through relevant experiments, it is proved that DTW score alignment and end-to-end are more successful in extracting the features of classical music, and more accurate in identifying classical instruments. The objective evaluation method of pronunciation teaching quality is more objective and accurate than P.563 music teaching quality evaluation.

KEYWORDS

classical music, feature extraction, machine learning, musical instrument recognition, teaching quality evaluation

INTRODUCTION

In machine learning, ‘data mining’ and ‘data analysis’ are similar terms that connote a process of recognizing meaningful, effective, special, and valuable facts from abundant data (Gupta et al., 2021). Before the application of information technology, people could only mine and analyze data manually. In the era of data, information has seen extraordinary growth. Individuals are continuously generating and leveraging data to function and thrive in their day-to-day lives. Through a combination of data storage technology and advanced machine learning algorithms, the field of data mining and analysis has been greatly expanded (Domashova & Zabelina, 2021). Data can be read and written efficiently through the current efficient data storage technology (Gupta et al., 2012). Afterward, data mining and analysis are optimized through the deployment of knowledge discovery, data statistics, and machine learning technology. The utilization of such technology brings forth undeniable advantages in terms of data processing and evaluation (Islam et al., 2021).

Machine learning plays a crucial role in music education, primarily as follows. First, recent developments in machine learning and artificial intelligence (AI) have the potential to optimize the aptitude of music teachers (Walker, 2021). AI offers a viable solution for replacing staff members who do not specialize in music education, thereby elevating the capabilities of existing music educators. By utilizing AI, music teachers acquire the advantage of an effective supplementary aid, resulting in an improved standard of expertise across the board. Students and parents will continue to improve their

DOI: 10.4018/IJWLTT.320490
*Corresponding Author

recognition of intelligent machines. In addition, with the help of artificial intelligence, music teachers can carry out self-study more efficiently and conveniently, thus continuously optimizing the teachers' level. Second, it can promote the improvement of teachers' teaching quality and efficiency. Relying on AI and big data analysis, teachers can quickly understand issues such as the students' learning level or background. In this way, teachers can quickly become effective carry out effective teaching for students. Meeting each student's educational needs can improve teaching quality and efficiency. Third, it can enhance students' learning efficiency. Music learning is not always fun and to master certain music skills, learners must invest considerable time and energy; however, not every learner persists. The introduction of artificial intelligence can mobilize the students' subjective initiative in learning music, help them realize the shortcomings in their own learning, urge them to learn, and effectively improve their learning efficiency.

Machine learning is an interdisciplinary field, comprising elements of science, psychology, biology, systems science, cognitive science, and information science. Through incorporating robotic technology, classical music education can be advanced with features such as recognition of instruments, feature extraction, and recognizing classical tunes. Consequently, intelligent instruments gain additional useful features, creating a personalized learning environment. Furthermore, machine learning technology enables observation of classroom instruction, analysis of melody and rhythm, making evaluations of teaching proficiency more precise and accurate, ultimately creating an atmosphere to enhance instructors' creativity while they use artificial intelligence to innovatively present the discipline through modern means.

Musical instruments are important tools for learning music. The instrument is embedded with machine learning technology, forming an intelligent instrument. It can store many music data of various industrial instruments. Especially when it is used in the music keyboard, the function of the music keyboard has been improved significantly, which has triggered a variety of teaching modes in music teaching. After relevant experiments, it is proved that the score alignment and the end-to-end neural network used in this paper are relatively successful in extracting the features of classical music. The two-level classification model used in this paper is more accurate in identifying classical musical instruments. The objective evaluation method of pronunciation teaching quality is more objective and accurate than P.563 music teaching quality evaluation. The deep network can learn high-level features that are beneficial for classification from the input primary features; on the other hand, there is only one playing instrument in monophonic music, using efficient preprocessing. After relevant experiments, such a neural network as we used in this paper can be relatively successful in extracting the features of classical music. The two-level classification model used in this paper is more accurate in identifying classical musical instruments. It is easy to extract accurate harmonic structure from the spectrogram with mean value and time-frequency transformation. Analyzing the energy distribution of the harmonic can effectively distinguish various musical instruments.

MATERIALS AND METHODS

The Development and Research Status of Machine Learning in Classical Music Education

Using synthesizers in music teaching of machine learning has become one of the foundational teaching methods in today's primary and secondary school classrooms. Music is one subject with the highest informatization in the primary and secondary school classroom. The multifarious multimedia courseware and music video make the music classroom eliminate the limitation of the old recorder plus tape mode, and improve classroom efficiency (Ariza-Colpas et al., 2021). It is not uncommon to replace the audio with video in today's high school music classes, and video editing and processing have become one teaching skills for music teachers besides professional skills. Video technology has the potential to render abstract music tangible and comprehensible for students (Abdelkader et al., 2021).

I personally have been eager to develop my understanding of music composition. Nevertheless, when I started my post, I encountered fresh issues in both attending lectures and listening to the recordings thereof. In the classroom, the electronic synthesizer mainly plays the role of instant demonstration of phrases. It has the same usage as the piano but has more functions than the piano. However, there are also shortcomings. For example, the synthesizer cannot imitate multiple timbres simultaneously. When it is necessary to demonstrate the combined effect of two or more instruments, a simple performance-type synthesizer divides the keyboard into two areas with a certain tone as the boundary. While it can imitate two timbres, its available sound area is extremely limited, and it is difficult for the real effect to be displayed.

The utilization of one individual's composition or rendering on musical instruments has supplanted the traditional practice which necessitated collaboration amongst multiple people, improving learning and practice efficacy. Students can play their songs and practice musical works immediately after class, and teachers can understand the teaching objectives and difficulties in the teaching content in the music practice. Artificial intelligence technology continues to make breakthrough progress, which makes intelligent and humanized electronic instruments constantly innovate (Kolodziev et al., 2020). These intelligent electronic instruments can not only store a wide range of instrument timbres, but also realize the effective arrangement of various timbres, so that all kinds of timbres can perform orderly music according to the corresponding behavioral instructions. This kind of instrument function is difficult for traditional instruments to achieve.

As far as the curriculum is concerned, changes are also taking place according to the situation of the music classroom and the different cultures of the society (Faradhillah & Zahara, 2021). What is unusual is that this is the first time a top orchestra has played a work composed entirely of machines (Xiong et al., 2022). It can create incredibly complex tracks without human intervention (Haq et al., 2020). It usually resonates with the audience's works, and may try other music types in the future (Remolina et al., 2022). Everyone likes music, and music lessons should also be a favorite course for every student (Gupta, et al., 2021). Every high school music teacher constantly seeks a teaching mode that is easy to understand and absorb by students, can attract students' interest in learning, and can improve students' music core literacy (Banulescu-Radu & Yankol-Schalck, 2021). For curriculum, changes between the music classroom situation and the different cultures of the society are also happening (Tang et al.,2022).

RESEARCH STATUS OF CLASSICAL MUSICAL INSTRUMENT RECOGNITION

Identification is already possible for note-level audio segments or continuous audio signals played by a solo instrument. For example, to classify solo passages clip dataset of 24 instruments, use sparse coding to learn the features in the spectrogram, and then train a support vector machine to classify the instruments using the learned features, 24 instruments The classification accuracy of the categories reaches around 0.95. classification recordings in the MedleyDB multi-track dataset (Weng & Chen, 2020).

There has been some progress in primary instrument identification. A 'primary' instrument is defined as one that is present continuously in the audio segment and easily heard by a human listener. Multi-instrumental recognition in polyphonic music is challenging.

Application of Machine Learning in Classical Music

Convolutional neural networks have been widely used in instrument recognition work in monophonic music, and their instrument recognition of monophonic music, the classification features used are generally common time-frequency features or cepstral domain features based on auditory characteristics (Wang & Guo, 2022). It may be because of these two reasons: on the one hand, the multi-layer structure of the deep network can learn high-level features that are beneficial for classification from the primary input features; there is only one playing an instrument in monophonic music, using efficient preprocessing (Aróstegui, 2010). After relevant experiments, it is proved that the DTW score alignment and the end-to-end neural network used in this paper are relatively successful in extracting the features of classical music. The two-level classification model used in this paper is more accurate in identifying classical musical instruments (Yang, 2021). It is easy to extract the precise harmonic structure in the spectrogram through means and time-frequency transformation, and then analyzing the energy distribution of the harmonics can effectively distinguish various musical instruments. On the other hand, multi-instrumental recognition in polyphonic audio is challenging (Nag et al., 2022). Identification is already possible for note-level audio segments or for continuous audio signals played by a solo instrument. To classify solo passages clip dataset of 24 instruments, a sparse coding approach was used to extract features from the spectrogram. Afterwards, a support vector machine was used to classify the instruments based on the features derived. This process is necessary for multi-instrument recordings as methods such as autocorrelation coefficient, cepstral coefficient, linear prediction coefficient, MFCC, and short-term energy analysis for melody, rhythm, harmony, timbre, pitch do not produce the desired results in polyphonic audio (Kruthika et al., 2021).

(1) Pitch feature extraction

Machine learning is very effective in improving the final classification performance. In this respect, it is proven that when a neural network is used to learn the intermediate features of audio signals for classification, the spectral map is effective as a low-level feature. On the other hand, deep end-to-end models may overfit datasets of real-world singing audio (Chan et al.,2006). So, we reconsider using the filter bank to extract the time-frequency features of the audio as the primary features of the input convolutional network. In this paper, the above algorithm is used in classical music education, identifying classical musical instruments, the feature extraction and identification of classical music, and the quality evaluation of classical music education. After relevant experiments, it is proven that the DTW score alignment and the end-to-end neural network used in this paper are relatively successful in extracting the features of classical music. The two-level classification model used in this paper is more accurate in identifying classical musical instruments. The objective evaluation method of pronunciation teaching quality is more objective and accurate than P.563 music teaching quality evaluation.

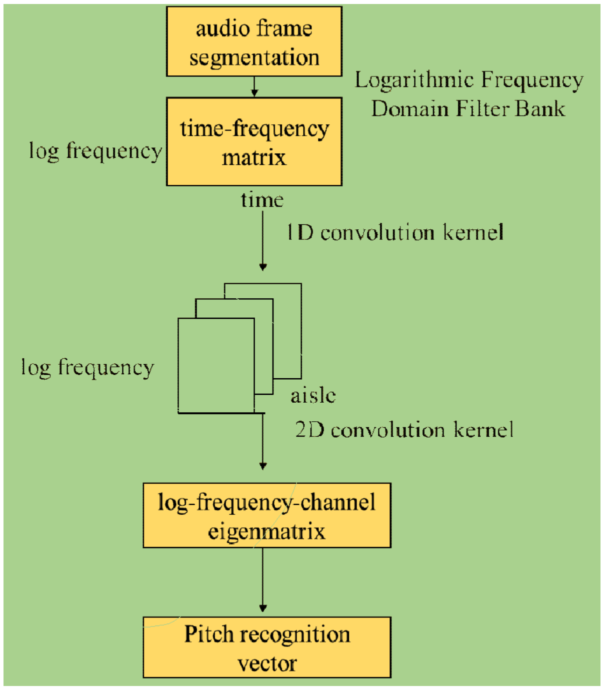
The time-frequency sequential topology in the frequency domain may improve the effect of the convolution operation. In addition, the risk of overfitting is greatly reduced due to the replacement of the first layer number of model parameters.

Given an audio frame X, we seek to predict the pitch combinations corresponding to the sung notes in X, We encode the pitch combination as a binary label vector $y \in \{0,1\}^{128}$, with 128 dimensions corresponding to the 128 pitch frequencies, and the nth component in vector y if the nth tone $y_n = 1$ is present in the frame. We learn a feature map $f_\theta : X \to P$ via a multilayer convolutional network, and then train a multiple linear regression to predict a given f. The label vector j of which the parameters are optimized using a cross-discipline loss function.

We replaced the first layer of the end-to-end network with a filter bank, and we show the constructed network model structure in Figure 1. The specific details are as follows.

The normalization $X \to X\left\|X\right\|^2$ of audio frame X, which can be physically interpreted as normalizing the audible volume of each frame, T_p is then divided into segments with a sliding window, each segment is denoted as $\boldsymbol{x}_t$, and there are s samples $x_t = \left(x_{t_1}, x_{t_2}, \ldots\ldots x_{t_s}\right)$. As a segment of an audio frame, $x_t \in X$. The audio frame X is passed through a set of logarithmic frequency domain filter banks to map it to the logarithmic frequency domain, and the logarithmic frequency-time matrix $\left(n_p \times T_p\right)$ is obtained. The filter $n_p = 511$ bank consists of sine and cosine filters, the logarithmic domain frequency range is $\log f_L$ to $\log f_H \left(f_L = 20\,\text{Hz}, \quad f_H = 6\text{kHz}\right)$, and the parameters of the i-th sine filter are:

Figure 1. Pitch Feature Extraction Process



$$w\left(\sin 2\pi f_i t_1, \sin 2\pi f_i t_2, \ldots\ldots \cdot \sin 2\pi f_i t_s\right)_{i,sin} \tag{1}$$

The parameters of the i-th cosine filter are:

$$w\left(\cos 2\pi f_i t_1, \cos 2\pi f_i t_2, \ldots\ldots \cdot \cos 2\pi f_i t_s\right)_{i,cos} \tag{2}$$

The output of the i-th filter is:

$$\text{filter}_i = \left(w_{i,sin}^{T_t}()^2 \left(w_{i,cos}^{T_t}()^2\right)\right) \tag{3}$$

The precision rate P, recall rate R, and F1 score are defined as:

$$\begin{aligned} P &= \frac{U_{TP}}{U_{TP} + U_{FP}} \\ R &= \frac{U_{TP}}{U_{TP} + U_{FN}} \\ F_1 &= 2\frac{P \times R}{P + R} \end{aligned} \tag{4}$$

$$x_t = \left(x_{t_1}, x_{t_2}, \ldots\ldots x_{t_s}\right) \quad (5)$$

$$K(1.1) = S(1,1) \tag{6}$$

(2) Classical musical instrument recognition algorithm

The relevant formula is as follows:

1) Loss function

$$\ell'_{ban} = -\sum_{k=1}^{K} \omega_k \left[\hat{y}_k \log y_k + \left(1 - \hat{y}_k\right) \log\left(1 - y_k\right)\right] \tag{7}$$

2) Attention Score

$$a_n = \sum_{m=1}^{M\Sigma} \frac{\exp\left(v_{n,m}\right)}{\sum_{n=1}^{N\Sigma\left(v_{n,m}\right)} exp} \tag{8}$$

3) Attention weights

$$\partial_n = \frac{\exp\left(a_n\right)}{\sum_{n=1}^{N\Sigma\left(a_n\right)} exp \sum_{n=1}^{N} \partial_n} \tag{9}$$

Score Alignment Algorithm

Hidden Markov Model (HMM) is a statistical model, used to describe a Markov process with hidden unknown parameters. The difficulty is to determine the implicit parameters of the process from the observable parameters. These parameters are then used for further analysis, such as pattern recognition. For HMM, simulation is quite easy if the transition probability between all hidden states and the output probability from all hidden states to all visible states are known in advance. On the problem of aligning musical scores to performance audio, Model (HMM), and the fusion of these two algorithms.

(1) Hidden Markov Model

Hidden Markov Models were originally proposed in the second half of the 1960s by Leonard E. Baum and other authors of statistical papers. In the mid-1970s, HMM was first used in speech recognition. In the second half of the 1980s, HMM was applied in genome sequence analysis and prediction of protein coding regions. Predicting the state sequence prediction of random processes such as weather and humidity is something it can do. Essentially, under the condition of a known observation sequence, the most likely corresponding hidden state sequence, or state sequence for short, is solved. Specific to the problem of aligning audio in MIDI scores, the alignment function is implemented through the decoding process of the HMM model.

(2) Dynamic time-warping algorithm

In the 1960s, the Japanese scholar Itakura first proposed the DTW algorithm. To solve the matching problem of two template sequences of different lengths, it was primarily used in speech recognition. For example, two people say a sentence with the same meaning, generating two speech sequences. However, the length of the generated sequences is also different due to the speaker's emotion, word usage habits, and different speech rates. A regular algorithm is needed to combine the two speech sequences in which the elements in the sequence correspond. DTW algorithm can apply to this scenario.

Define the loss accumulation matrix K, K is calculated using the trick of a recursive algorithm, where K(1.1) = S(1,1), as in formula (10):

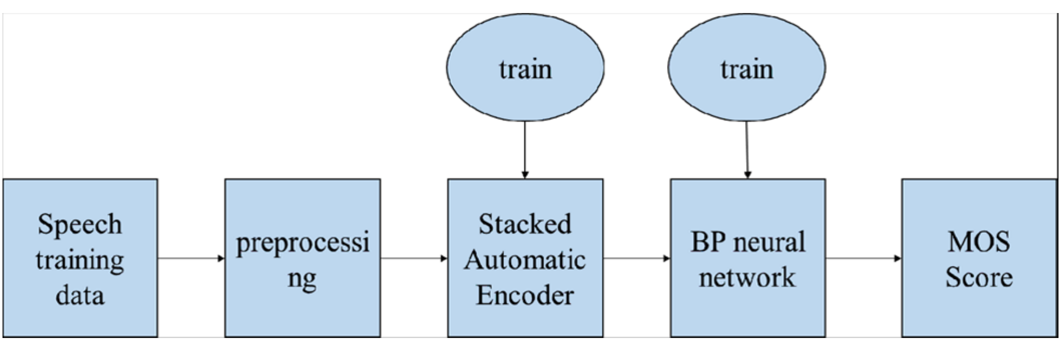
$$K\left(m,n\right) = S\left(m,n\right) + \min\{K\left(m,n-1\right), K\left(m-1,n\right)\} \tag{10}$$

From the description of the above two alignment algorithms, we can know that the HMM model needs to train the model to learn the parameters suitable for the scene of large-scale data volume, and the complexity is also high. Moreover, the amount of MIDI score and original audio data we get is limited, which is insufficient to train an HMM model with strong generalization ability. Therefore, the DTW algorithm that is less complex and easy to implement is finally selected.

Classical Music Teaching Evaluation Algorithm

The digital processing of the sound signal of classical music serves human hearing, and the human auditory system has its unique perceptual characteristics to the sound signal. Therefore, to accurately predict the real MOS score, the objective evaluation method of music teaching quality needs to select speech features that can fully characterize speech and human perception characteristics as parameters. In this paper, the parameters of the stacked auto-encoder are trained by selecting a database containing multiple impairment types of speech so that SAE can extract the essential features that can better characterize speech quality. During the training process, a single auto-encoder is trained. When the training of one auto-encoder is completed, the next auto-encoder takes the output of the hidden layer of the first auto-encoder as input, and then repeats the previous step until all auto-encoders are trained. We illustrate the testing and training framework for the objective evaluation method of speech teaching quality based on stack auto-encoder in Figure 2.

Figure 2. Frame Diagram of Music Teaching Quality Evaluation System



The performance of the music teaching quality evaluation system can be evaluated using two metrics: accuracy and correlation. The accuracy is judged by calculating the root mean square error (RMSE) between the objective evaluation results and subjective evaluation outcomes of music teaching quality; the correlation is described using the Karl-Pearson correlation coefficient. Furthermore, this system can be used to the state sequence in random processes, such as weather and humidity. In essence, given a known observation sequence, given a known observation sequence, the model will solve for the most probable hidden state sequence or state sequence. In particular, to the problem of aligning audio in scores, alignment is enabled via model decoding.

The accuracy of the objective evaluation system of music teaching quality is often measured by the mean square error value index. The mean square error (RMSE) is also called the standard error. It is the difference between the output score of the music teaching quality objective evaluation system and the subjective MOS score value. Take the square root of the result of the ratio of the sum of squares to the number of tests N.

The essence of the objective evaluation system of music teaching quality is to use computer simulation system to evaluate the quality of output music teaching. The absolute prediction error is the difference between the result value output by the computer and the true MOS score, which can be calculated by formula (11):

$$P_{\text{error}}\left(i\right) = \mathrm{MOS}_{s}\left(i\right) - \mathrm{MOS}_{o}\left(i\right) \tag{11}$$

Among them, $\mathrm{MOS}_o(i)$ is the real subjective MOS score value of the i-th corpus, and $\mathrm{MOS}_o(i)$ is the objective evaluation prediction value of the i-th corpus corresponding to it. The root mean square error RMSE is calculated by $P_{\text{error}}\left(i\right)$, as shown in formula (12):

$$RMSE = \sqrt{\left(\frac{1}{N-1}\sum\nolimits_{i=1}^{N} P_{\text{error}}\left(i\right)^{2}\right)} \tag{12}$$

Given N as the total number of the test corpus, we use N-1 in formula (12) to ensure the root mean square error is an unbiased estimate. Also referred to as the standard error, it denotes the difference between the output score of the music teaching quality objective evaluation system and the subjective MOS score value. Thus, we can compute the assessment by taking the square root of the ratio resulting from the sum of squares and dividing by the number of tests.

The correlation index between subjective and objective evaluations is analyzed below. Karl-Pearson correlation coefficient is usually used to represent the correlation between the music teaching quality output score and the subjective real value, the linear correlation between them. Therefore, the size of the correlation coefficient R is used in this paper to indicate the strength of the correlation between the objective evaluation result and the subjective real value. The larger the value of R, the stronger the correlation between the objective evaluation result value and the subjective MOS real value. sex is weaker. The calculation of R is shown in formula (13):

$$R = \frac{\sum_{i=1}^{N}\left(\left(\mathrm{MOS}_{o}\left(i\right) - \overline{MOS_{o}}\right)\left(\mathrm{MOS}_{s}\left(i\right) - \overline{MOS_{s}}\right)\right)}{\sqrt{\sum_{i=1}^{N}\left(MOS_{o}\left(i\right) - \overline{MOS_{o}}\right)^{2}\sum_{i=1}^{N}\left(MOS_{s}\left(i\right) - \overline{MOS_{s}}\right)^{2}}} \tag{13}$$

Among them, $MOS_s\left(i\right)$ is the real subjective MOS score value of the i-th corpus, and $\overline{MOS_s}$ is the average value of the subjective MOS score values of all corpora. $\overline{MOS_o}$ and $MOS_o\left(i\right)$ are the prediction score of the i-th corpus and the average of the prediction scores of all corpora.

The two performance indicators for judging the quality of the objective evaluation system of music teaching quality have been analyzed before. Equations (14) and (15) are the performance difference parameters of the music teaching quality evaluation system:

$$\Delta_R = \frac{R_{\text{proposed}} - R_{\text{compare}}}{1 - R_{\text{compare}}} \times 100\% \tag{14}$$

$$\nabla_{RMSE} = \frac{RMSE_{\text{compare}} - RMSE_{\text{proposed}}}{RMSE_{\text{compare}}} \times 100\% \tag{15}$$

Among them, Δ_R is the growth rate of the correlation coefficient of the research method in the text compared with the comparison method, ∇_{RMSE} is the decrease rate of the mean square error of the method in the text compared with the comparison method, the subscript "proposed" indicates the method studied in the paper, and "compare" means to compare methods, namely P.563 method and music teaching quality evaluation method based on FDGSVM. Δ_R or ∇_{RMSE} the value of is positive, which proves that the method in this study improves the correlation with subjective quality evaluation compared with the comparison method. On the other hand, the value is negative, showing that the method in this paper reduces the correlation with subjective quality evaluation.

RESULTS AND DISCUSSIONS

Application Experiment of Machine Learning in Classical Music

This article edited two annotated datasets. It includes WCMED and CCMED, respectively, including recordings of Chinese and Western music. Subsequently, they trained the pre-trained SED and pre-trained SER models on these datasets, and combined them with the support vector regression (SVR) model. The SED model aims to detect sound events in audio signals, while the SER model is trained to recognize the emotion conveyed by the sound scene recording. Figure 3 shows the results of tone recognition models under different frame lengths when using double frame lengths (global threshold is 0.4). These three are also multi-tone estimation standards used by international MIREX).

As can be seen from Figure 3, the classification effect and cost performance are relatively good when the frame length is 8192 sample points. Therefore, we use 8192 sampling points per frame to extract pitch features; considering the trade-off between shorter and longer frames, shorter frames may miss important contextual information of the classification task, resulting in a decrease in the recognition effect, and longer frames. As a result, there is a decreasing return in computational cost, and the number of parameters and training time increase, but the recognition effect is not significantly improved. It proves that the method in this paper improves the correlation with subjective quality evaluation compared with the comparison method, and the value is negative, indicating that the method in this paper reduces the correlation with subjective quality evaluation.

Identifying solo musical passages or note-level audio segments of continuous signals produced by a single instrument is now feasible. As such, it is possible to classify individual melodies.

Figure 3. Recognition Results of Pitch Recognition Models Under Different Frame Lengths

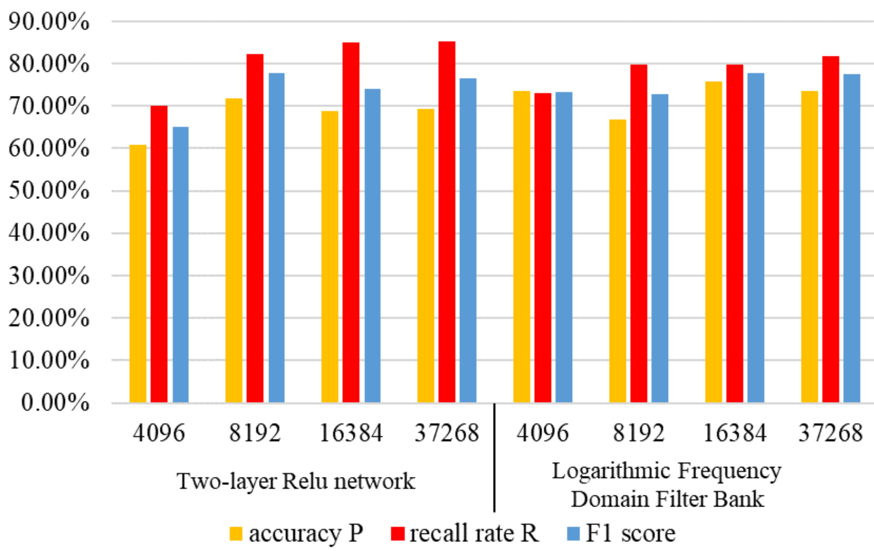
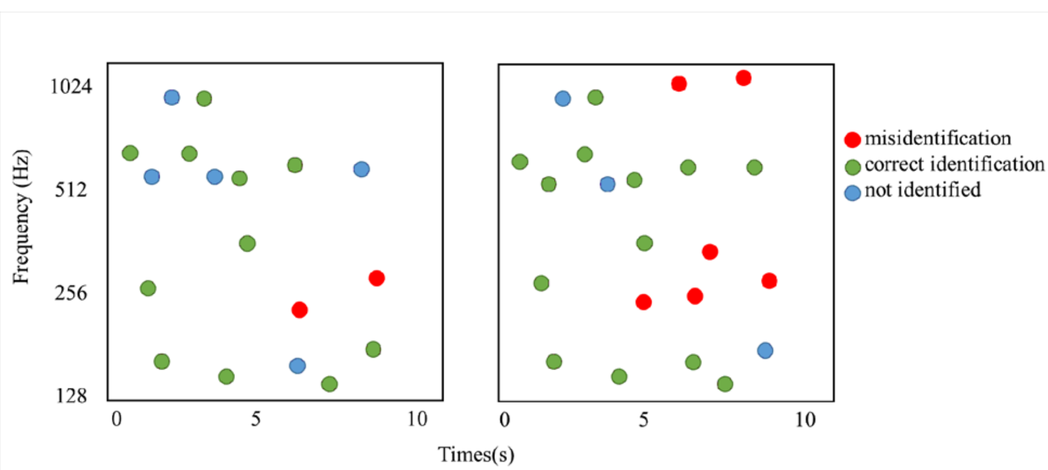


Figure 4 shows the performance of the recognition model for pitch recognition at a frame length of 8192 when using a filter bank (left) and a two-layer network (right) to extract the primary audio features. The blue dots indicate the correctly identified pitch frequencies, black dots indicate unidentified pitch frequencies, and the red dots indicate misidentified pitch frequencies. Figure 3 shows that the recall rate is higher when using the two-layer network, which is consistent with the situation reflected in Figure 4. Using the two-layer network can identify most of the pitches of specific frequencies that need to be identified, but those pitches that do not exist are also identified, resulting in obvious over-fitting. This phenomenon may be because when using a two-layer network, the number of learned parameters of the model has increased to the point of over-fitting compared to using a log-frequency domain filter bank with manually defined parameters. Two-layer network,

Figure 4. The Effect of the Recognition Model for Pitch Recognition Under the Frame Length of 8192



filter, and end-to-end neural network, two-level classification model, objective evaluation method of speech teaching quality based on stack auto-encoder and P.563 music teaching quality evaluation are all part of machine learning.

The use of this algorithm in classical music education has been investigated, encompassing the identification of classical musical instruments, the feature extraction and identification of classical music, and the quality evaluation of classical music education. Upon testing, we established that the DTW score alignment and the end-to-end neural network are effective tools for extracting the distinguishing features of classical music, while a two-level classifier is more accurate in recognizing classical instruments. The objective evaluation method of pronunciation teaching quality is more objective and accurate than P.563 music teaching quality evaluation. It is consistent with the situation reflected in Figure 4. Using the two-layer network can identify most of the pitches of specific frequencies that need to be identified; however, those that do not exist are also identified, leading to over-fitting. This phenomenon could be attributed to the increase in the number of learned parameters in such a model compared to using a log-frequency domain filter bank with manually defined parameters. Alternatives, like two-layer networks, filters, and end-to-end neural networks, two-level classification models, objective evaluation methods, must still be considered.

In addition, it can also be seen from Figure 4 that when the model is wrongly identified, the correct pitch is often identified as the pitch at the frequency corresponding to twice or half, which is called octave error (octave error). errors). The octave error means that when some musical instruments are playing a certain note, due to their rich overtones (harmonics), the energy of the harmonics and the fundamental frequency (pitch) is equal, and the algorithm will put the pitch candidate value. The frequency of the harmonics is mistakenly taken as the pitch frequency, resulting in a displacement of the output pitch by an octave. This issue should be addressed to improve pitch estimation and bolster the recognition efficacy of the model going forward.

The precision of the identification of various musical instruments shown in Figure 5 is demonstrated by the experimental results detailed in Figure 6, which depict the results of the Dynamic Time Warping (DTW) alignment of both sequences.

It shows that the alignment algorithm used in this paper meets the requirements.

Figure 5. Comparison of Recognition Scores of the Three Models

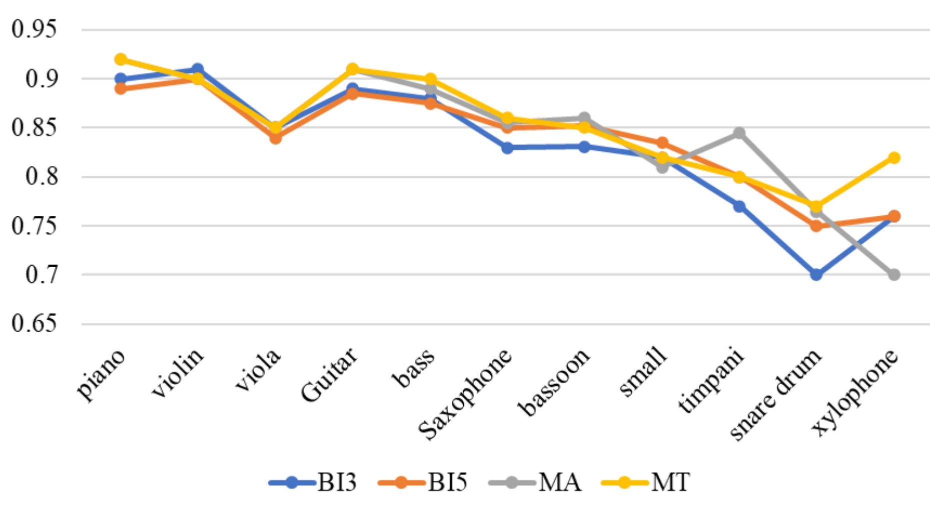
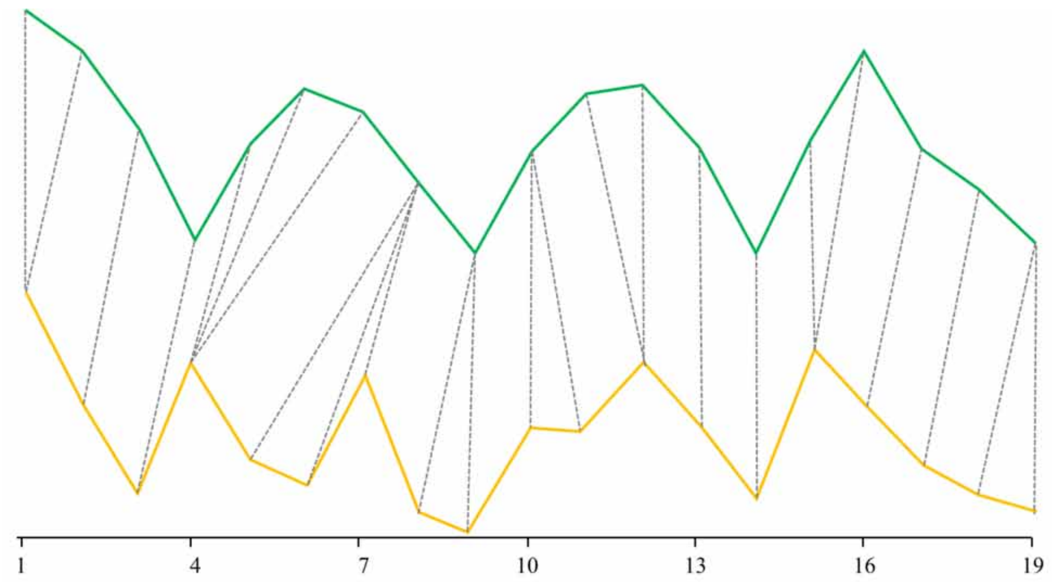


Figure 6. DTW Alignment Between Two 1D Sequences



EXPERIMENT RESULTS OF THE CLASSICAL MUSIC TEACHING QUALITY EVALUATION SYSTEM

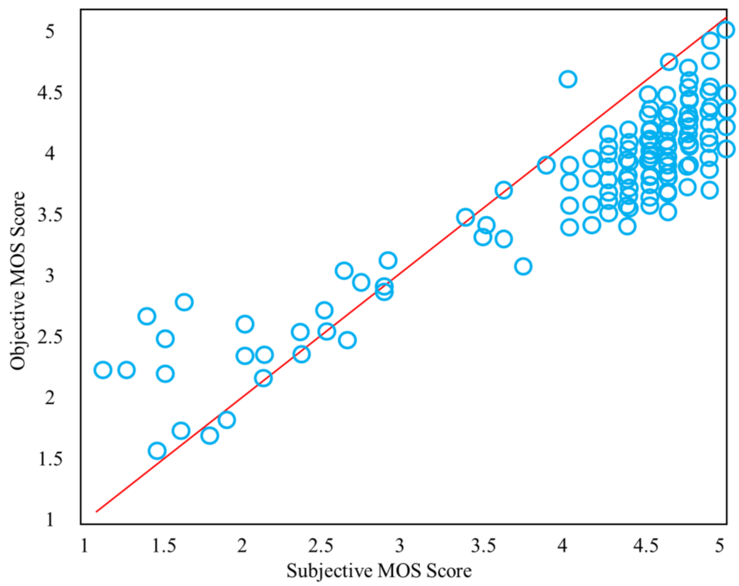
In Figures 7(a) and (b), the scatter plots illustrate the music quality evaluation methods used in this study and the P.563 music teaching quality evaluation method, respectively. We see the implementation of the above algorithm in classical music education, which involves identification of classical musical instruments, the feature extraction and identification of classical music, and the quality evaluation of classical music education. Through research, core alignment and end-to-end neural networks are comparatively successful in extracting the features represented by the circle points. We should note that the horizontal axis corresponds to the subjective MOS score. In contrast, the vertical axis is the objective prediction score, and the diagonal line reflects the ideal situation where the objective prediction score and subjective MOS score are perfectly compatible.

By assessing the correlation between output results and subjective evaluations, we can gauge the efficacy of a music quality evaluation system. The higher the correlation, the more ideal the music quality evaluation method is. A diagonal line in a scatterplot illustrates a perfect match between the objective and subjective ratings; the closer the data points are to this line, the more accurate the evaluation result is. Our method, as seen in Figure 7(a), provides a score distribution that is much nearer to the ideal than the P.563 music quality evaluation method shown in Figure 7(b), revealing how much better our approach is at evaluating the quality of musical instruction.

CONCLUSION

Combining machine learning and classical music will accelerate the development of the entire music industry and realize the upgrading of the music industry. Artificial intelligence composition has greatly improved the efficiency of musicians and eliminated many time-consuming parts of music production. If we know ‘what’ according to the analyzed data, we can directly know users’ music preferences and make or push the music that users want. This paper applies the above algorithm to classical music education, including the recognition of classical instruments, the feature extraction

Figure 7. Comparison Between the Method in this Paper and Various Music Quality Evaluation Methods (a) Evaluation method of music quality in this paper (b) P.563 Music Quality Evaluation Method



and recognition of classical music, and the quality evaluation of classical music education. Through relevant experiments, we have proven that DTW score alignment and the end-to-end neural network used in this paper are more successful in extracting classical music features. The two-level classification model used in this paper is more accurate in identifying classical musical instruments. The objective evaluation method of phonetics teaching quality is better than P 563, with the evaluation of music teaching quality being both more objective and accurate. In the future, artificial intelligence machine learning should pay more attention to emotional learning and establish more and more complete music emotion recognition models. In this way, the songs created by artificial intelligence will be full of feelings and emotions which listeners will resonate with. The application of AI in music education mainly plays an auxiliary role. However, for the perceptual problems in music teaching, AI has certain limitations. Therefore, further analysis is needed in this domain.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING STATEMENT

This work was not supported by any funds.

ACKNOWLEDGMENTS

The authors would like to show sincere thanks to those techniques who have contributed to this research.

REFERENCES

Abdelkader, W., Navarro, T., Parrish, R., Cotoi, C., Germini, F., Linkins, L. A., Iorio, A., Haynes, R. B., Ananiadou, S., Chu, L., & Lokker, C. (2021). A deep learning approach to refine the identification of high-quality clinical research articles from the biomedical literature: Protocol for algorithm development and validation. *JMIR Research Protocols*, *10*(11), e29398. doi:10.2196/29398 PMID:34847061

Ariza-Colpas, P. P., Ayala-Mantilla, C. E., Shaheen, Q., Piñeres-Melo, M. A., Villate-Daza, D. A., Morales-Ortega, R. C., De-la-Hoz-Franco, E., Sanchez-Moreno, H., Aziz, B. S., & Afzal, M. (2021). SISME, estuarine monitoring system based on IOT and machine learning for the detection of salt wedge in aquifers: Case study of the Magdalena River estuary. *Sensors (Basel)*, *21*(7), 2374. doi:10.3390/s21072374 PMID:33805544

Aróstegui, J. L. (2010). Risks and promises of ICT for Music Education. *Hellenic Journal of Music, Education and Culture*, *1*(1), 17–31.

Banulescu-Radu, D., & Yankol-Schalck, M. (2021). *Fraud detection in the era of Machine Learning: A household insurance case* (Report No. 2904). Orleans Economics Laboratory/Laboratoire d'Economie d'Orleans (LEO), University of Orleans.

Carneiro, D., Guimarães, M., & Sousa, M. (2021). Optimizing instance selection strategies in interactive machine learning: An application to fraud detection. In A. Abraham, T. Hanne, O. Castillo, N. Gandhi, T. Nogueira Rios, & T. P. Hong (Eds.), *Hybrid Intelligent Systems. HIS 2020. Advances in Intelligent Systems and Computing* (Vol. 1375, pp. 124–133). Springer. doi:10.1007/978-3-030-73050-5_13

Chan, M., Potter, J., & Schubert, E. (2006). Improving algorithmic music composition with machine learning. *Proceedings of the 9th International Conference on Music Perception and Cognition, ICMPC.* IEEE.

Domashova, J., & Zabelina, O. (2021). Detection of fraudulent transactions using SAS Viya machine learning algorithms. *Procedia Computer Science*, *190*, 204–209. doi:10.1016/j.procs.2021.06.025

Faradhillah, F., & Zahara, S. R. (2021). The aplication of learning models of project based learning to improve students' learning outcomes in post-legal materials. *International Journal for Educational and Vocational Studies*, *3*(3), 186–195. doi:10.29103/ijevs.v3i3.4308

Gupta, R. Y., Mudigonda, S. S., & Baruah, P. K. (2012). TGANs with machine learning models in automobile insurance fraud detection and comparative study with other data imbalance techniques. *International Journal of Recent Technology and Engineering*, *9*(5), 236–244. doi:10.35940/ijrte.E5277.019521

Gupta, R. Y., Mudigonda, S. S., & Baruah, P. K. (2021). A comparative study of using various machine learning and deep learning-based fraud detection models for universal health coverage schemes. *International Journal of Engineering Trends and Technology*, *69*(3), 96–102. doi:10.14445/22315381/IJETT-V69I3P216

Gupta, R. Y., Mudigonda, S. S., Baruah, P. K., & Kandala, P. K. (2021). Markov model with machine learning integration for fraud detection in health insurance. aarXiv.2102.10978

Haq, A. U., Li, J. P., Khan, J., Memon, M. H., Nazir, S., Ahmad, S., Khan, G. A., & Ali, A. (2020). Intelligent machine learning approach for effective recognition of diabetes in E-healthcare using clinical data. *Sensors (Basel)*, *20*(9), 2649. doi:10.3390/s20092649 PMID:32384737

Islam, M., Ripan, R. C., Roy, S., & Rahat, F. (2021, December). Feature engineering based credit card fraud detection for risk minimization in e-commerce. *International Conference on Intelligent Computing & Optimization,* (pp. 217-226).

Kolodiziev, O., Mints, A., Sidelov, P., Pleskun, I., & Lozynska, O. (2020). Automatic machine learning algorithms for fraud detection in digital payment systems. *Восточно-Европейский журнал передовых технологий, 5*(9-107), 14-26.

Kruthika, G., Kuruba, P., & Dushyantha, N. D. (2021). A system for anxiety prediction and treatment using indian classical music therapy with the application of machine learning. *Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020*, 345-359.

Nag, S., Basu, M., Sanyal, S., Banerjee, A., & Ghosh, D. (2022). On the application of deep learning and multifractal techniques to classify emotions and instruments using Indian Classical Music. *Physica A*, *597*, 127261. doi:10.1016/j.physa.2022.127261

Remolina, M. C. R., Li, Z., & Peleato, N. M. (2022). Application of machine learning methods for rapid fluorescence-based detection of naphthenic acids and phenol in natural surface waters. *Journal of Hazardous Materials*, *430*, 128491. doi:10.1016/j.jhazmat.2022.128491 PMID:35739672

Szelogowski, D. (2021). Generative deep learning for virtuosic classical music: Generative adversarial networks as renowned composers. arXiv.2101.00169

Tang, H., Zhang, Y., & Zhang, Q. (2022). The use of deep learning-based intelligent music signal identification and generation technology in national music teaching. *Frontiers in Psychology*, *13*, 762402. doi:10.3389/fpsyg.2022.762402 PMID:35814087

Walker, S. (2021). Rejoinder to the critique of an article on machine learning in the detection of accounting fraud. *Econ Journal Watch*, *18*(2), 230–234.

Wang, D., & Guo, X. (2022). Research on evaluation model of music education informatization system based on machine learning. *Scientific Programming*, *9658735*, 1–12. Advance online publication. doi:10.1155/2022/9658735

Weng, S.-S., & Chen, H.-C. (2020). Exploring the role of deep learning technology in the sustainable development of the music production industry. *Sustainability*, *12*(2), 625. doi:10.3390/su12020625

Xiong, T., Ma, Z., Li, Z., & Dai, J. (2022). The analysis of influence mechanism for internet financial fraud identification and user behavior based on machine learning approaches. *International Journal of System Assurance Engineering and Management*, *13*(3), 996–1007. doi:10.1007/s13198-021-01181-0

Yang, J. (2021). Research on the artificial intelligence teaching system model for online teaching of classical music under the support of wireless networks. *Wireless Communications and Mobile Computing*, *4298439*, 1–11. Advance online publication. doi:10.1155/2021/4298439

Dongfang Wang, a professor at Xinxiang University, mainly studies the application of machine learning in music education.