

Applied Mathematics and Nonlinear Sciences

<https://www.sciendo.com>

Teaching Integration of Piano and Traditional Music Elements in Colleges and Universities Based on Network Flow Optimization

Yufei Dai^{1,†}

1. Xinzhou Normal University, Xinzhou, Shanxi, 034000, China.

Submission Info

Communicated by Z. Sabir
 Received February 26, 2023
 Accepted June 28, 2023
 Available online December 18, 2023

Abstract

In this paper, from the basis of piano music theory, pitch and twelve equal temperaments are studied, and due to the characteristics of the piano's construction, there are different timbres and harmonic structures. In order to realize the digitization of music signals for piano teaching, the collected analog signals are converted into digital signals by Fourier transform, the samples are inputted into the neural network model to get the output results, and the loss function based on Markov's music language model in piano teaching is used to find the partial derivatives of the parameters to realize the maximization of the efficiency of the piano pitch and note recognition. According to the multi-task learning in neural networks and the envelope curve of notes, we construct the automatic music transcription model in piano teaching based on CNN-HMM and use the simulation experiment analysis to empirically analyze the piano teaching based on CNN-HMM. The results show that the dual-channel audio results are improved by 2.68% in F1 value, 3.18% in accuracy and 2.61% in recall than the mono audio results. That is, dual-channel audio has richer information than mono audio, which enables the CNN-HNN network to learn richer features and thus improve the pitch and note recognition results. In this paper, the reliability of the teaching quality evaluation results of the CNN-HMM-based quantitative assessment model for facilitating the combination of electronic organ and piano teaching concepts ranges from 78.1% to 85.4%. This study provides guiding value for the development of digital piano teaching, which is of great significance for the innovative reform of piano teaching in colleges and universities.

Keywords: Convolutional neural network; Markov algorithm; Envelope curve; Harmonic structure; Piano teaching.
AMS 2010 codes: 97C70

[†]Corresponding author.

Email address: Dlpiano123@163.com

ISSN 2444-8656



<https://doi.org/10.2478/amns.2023.2.01545>



© 2023 Yufei Dai, published by Sciendo.



This work is licensed under the Creative Commons Attribution alone 4.0 License.

1 Introduction

Western instrumental piano has been present in China for more than a century. After more than one hundred years of development, Chinese composers have continuously absorbed the purification of Western piano art and combined it with excellent Chinese traditional art to create many piano works with Chinese cultural characteristics, which has created a new road of development for the piano art [1-2]. Chinese traditional opera art is the excellent national culture of China and has a high artistic status [3-4]. Integrating traditional opera art into piano art, expanding the scope and space of piano creative art, and creating many works that are both in line with the style of traditional Chinese opera culture and show the characteristics of the piano are of great significance in promoting the development of piano culture [5-6].

In this paper, we first explore and explain the foundation of piano music theory from the three aspects of pitch and twelve equal temperaments, timbre and harmonic structure, and music signal conversion, and inductively derive the convolutional neural network flow optimization algorithm and Markov algorithm in deep learning theory. Secondly, in the multi-task learning in note-level transcription, the encoder constructed by the CNN-based multi-task learning model is used to simultaneously extract the feature mapping of the pitch, note onset and cutoff of piano notes. Then, according to the TDR envelope curve proposed above, the notes are divided into three phases: transient phase, decay phase, and release phase, respectively, and the music language model in piano teaching based on HMM is constructed. Finally, the simulation experiment settings and evaluation indexes, as well as the dataset, are determined, and the method of simulation experiment analysis is applied to empirically analyze the research on piano teaching based on CNN-HMM.

2 Literature Review

Composers can utilize big data technology to extract nutrients from traditional music elements for the creation of piano music works, thus advancing the continuous innovation of piano music in China. Literature [7] uses a regression fitting algorithm and decompression F-weighting algorithm to extract the validity of each feature variable. Next, under the guidance of the metric learning theory, a hierarchical identification of influencing features of piano teaching was proposed to be accomplished in the projected feature space (P-KNN) algorithm. Literature [8] introduces complex network and multimedia technology into piano teaching cases, classifies, analyzes, researches and evaluates them, eliminates the roughness, preserves the essence, and screens some typical teaching cases that meet the modern learning theory, teaching requirements and the characteristics of the piano discipline. Literature [9] comprehensively considered the complex factors affecting the quality of piano information teaching and proposed to use of the improved BP neural network algorithm to evaluate and predict the quality of piano information teaching, aiming at enriching the piano teaching methods in colleges and universities, and improving the students' learning interest and passion. Literature [10] studied the perceived status of information piano education of college teachers and students, which mainly included a summary of the status of piano teaching in universities, and the findings were analyzed and summarized. In addition, it also utilizes new media to establish a network piano learning environment and builds a piano teaching "MOOC" platform, which is of reference value to the reform and innovation of piano teaching in colleges and universities.

In recent years, with the rise of piano teaching, many people have begun to learn to play the piano. Literature [11] proposed the design of an intelligent piano teaching system based on neural networks and studied the implementation method of the piano teaching system. It also guides teachers to guide students to practice, aiming to solve the shortage of piano education resources so that learning to play the piano is no longer a luxury activity, which is important for piano teaching. Literature [12] edge-

based solutions provide computation, analysis, storage and control for online piano teaching systems, supporting efficient processing and decision-making. Machine learning has also gained much attention in this area because of its flexibility and ability to provide a variety of supervised, unsupervised, and semi-supervised techniques. Thus, a specific model for assessing the potential relevance of piano teaching using machine learning is proposed, which is of great importance for the improvement of the quality of piano teaching in colleges and universities. Literature [13] study compared the teaching strategies of a novice and two expert instructors who started children's group piano lessons, highlighting how the components of these themes manifested themselves in the classrooms of these novice and expert instructors and described how these piano teaching techniques affected students' interest in learning. Literature [14] explored the role of gestures in teacher communication interactions during one-on-one piano lessons, where three teachers were asked to teach a pre-selected repertoire with three students studying piano in grade 1. The data were collected from video recordings of piano lessons based on the type and frequency of gestures used by the teachers in the associations to guide the gestures that fit within (or escape from) the predefined categories. Looking at theories of motivation, piano and psycho-pedagogy, the literature [15] Practice-based research describes how the author engaged her students in a selection process, adapting the economy to their tastes and goals, and learning about the students' values, expectations, and influences on their learning, finding that adult piano students find motivation and musical fulfillment to be more complex than simply finding what is pleasurable.

3 Fundamentals of Piano Music Theory and Deep Learning Theory

3.1 Fundamentals of Piano Theory

3.1.1 Pitch and twelve equal temperament

Pitch is defined as the frequency of vibration of the articulating object, and is used to describe the frequency level of a sound, expressed in Hertz (Hz). A sequence of notes composed of different pitches according to a certain time value is a monophonic melody. In the musical sound system, the pitch of each tone and its interrelationship is called the meter, and the twelve equal temperaments are one of the meters, which is widely used in symphony orchestras and keyboard instruments, and the 88 keys of the piano are tuned according to the twelve equal temperament. Twelve equal temperament is defined as the division of a pure octave into twelve semitones in equal proportions of frequency, with the ratio of frequencies between the semitones being $2^{1/12}$, which is expressed mathematically as:

$$\frac{f_{k+1}}{f_k} = \frac{1}{2^{12}} \approx 1.06 \quad (1)$$

That is, the frequency of each semitone is roughly 1.06 times the frequency of the preceding semitone, and two notes twelve semitones apart are exactly one octave apart, doubling the difference in frequency.

In 1936, the American Standards Committee set the frequency of the A note on center C at 440 Hz, which also sets the absolute pitch. For the piano, there are 88 keys in total, and according to the standard, the 49th key corresponds to a pitch of 440Hz, and then the pitch of the other keys can be calculated according to the twelve equal temperament laws:

$$f_i = f_0 \times 2^{\frac{i-49}{12}}, i = 1, 2, \dots, 88, i \in N \quad (2)$$

Where f_0 is the frequency of the standard pitch 440Hz, when $i = 1$, the pitch of the first key of the piano is calculated to be 27.5Hz, and when $i = 88$, the pitch of the last key of the piano is calculated to be 4185Hz.

3.1.2 Tone and harmonic structure

A comparison of the spectra corresponding to the keys in the bass and treble regions is shown in Figure 1, where (a) is the spectrogram of note A0 and (b) is the spectrogram of note A5. In addition to the fundamental sound, the sound from the instrument is accompanied by overtones, of which the frequency of the fundamental is called the fundamental frequency, the frequency of the overtones is called the harmonic frequency, and the difference in the tone is reflected in the difference of the fundamental and the overtones in the frequency and the amplitude, so that the different musical instruments play the same pitch, but the tone has a great deal of difference, which is the principle of the human ear to distinguish between the musical instruments. Due to the characteristics of the piano's construction, the sound emitted by the keys in the bass region has fewer fundamental components and more harmonic components, while the sound emitted by the keys in the treble region has more fundamental components and fewer harmonic components.

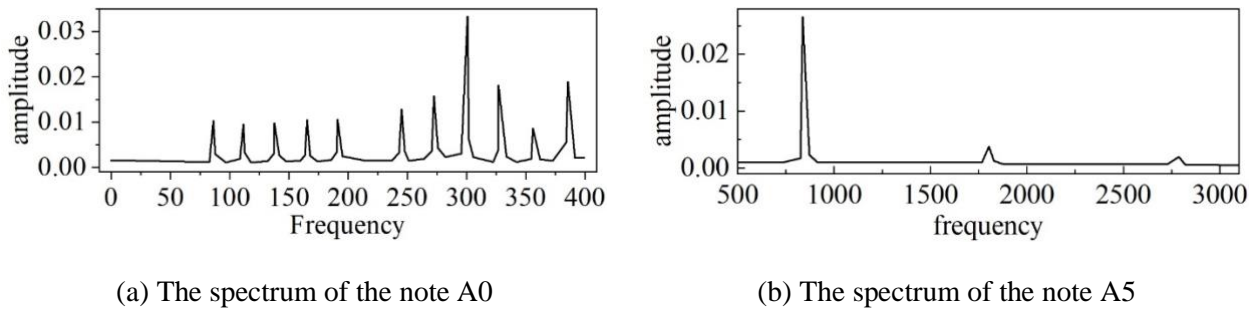


Figure 1. The contrast of the spectrum in the bass and the pitch

3.1.3 Music signal conversion

Piano music signals are non-stationary signals and cannot be analyzed by Fourier Transform, while STFT is based on the assumption that the signals are short-time stationary to analyze the signals in a steady state. Therefore, it can be assumed that the piano music has a short-time smoothness and is analyzed by STFT. The definition of STFT is as follows:

$$X_m(\omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n} \quad (3)$$

Where $x[n]$ represents the discrete music signal, $w[n-m]$ represents the window function, and $X_m(\omega)$ represents the spectrum at m moments. In the process of short-time Fourier transform, the window length determines the time resolution and frequency resolution. The longer the window length is, the longer the intercepted signal is, the lower the practical resolution and the higher the frequency resolution, and vice versa, the shorter the window length is, the shorter the intercepted signal is, the higher the time resolution and the lower the frequency resolution. Therefore, in the short-

time Fourier transform, the time resolution and frequency resolution are a contradiction, and the window length should be decided according to the actual situation.

The constant Q transform (CQT) is another method of frequency domain analysis, which is defined as follows:

$$X^{cq}(k) = \frac{1}{N_k} \sum_{n=0}^{N_k-1} x(n) w_{N_k}(n) e^{-j\frac{2\pi Q}{N_k}n} \quad (4)$$

Where k is the spectral line number, Q is the quality factor, the value of which is equal to the ratio of the center frequency to the bandwidth, due to the exponential distribution of the center frequency, Q is a constant, N_k is the window length of the window function $w_{N_k}(n)$, and N_k takes the value of:

$$N_k = \left\lceil Q \frac{f_s}{f_k} \right\rceil \quad (5)$$

$$f_k = f_{\min} \times 2^{\frac{k}{b}} \quad (6)$$

Where f_s is the sampling frequency, f_{\min} is the lowest frequency of the music signal, f_k is the frequency value of the k th spectral line, b is the number of spectral lines in an octave because the twelve equal temperament law divides an octave into twelve semitones, so b generally takes the value of 12 or multiples of 12, if it takes the value of 12, the frequency corresponding to each spectral line corresponds exactly to the scale frequency.

3.2 Deep Learning Theory

3.2.1 Convolutional Neural Network Flow Optimization Algorithm

A comparison of fully connected and convolutional layers is shown in Fig. 2, where (a) the fully connected layer and (b) the convolutional layer. Convolution is a crucial operation in analytical mathematics, and one- or two-dimensional convolution is commonly employed in signal processing or image processing. One-dimensional convolution is often used in signal processing to calculate the delay accumulation of the signal, and in the two-dimensional structure of the image, two-dimensional convolution is often used to smooth and denoise the image or extract the edge features and so on. In convolutional neural networks, the parameters are the weights in the convolution kernel and the bias. For a given data set, the samples are fed into the neural network model to get the output, and the parameters are determined by minimizing the structural risk function. In models with existing learning criteria and training samples, the network model parameters can be learned and updated by each iteration of the gradient descent method, which proceeds by taking the partial derivatives of the parameters over the loss function that constitutes the structural risk function to minimize the structural risk. Suppose the partial derivatives are computed for each parameter of each layer step by step by chain rule. In that case, the whole process becomes extremely inefficient, and the backpropagation algorithm is often used in the training of convolutional neural networks for efficient computation.

A convolutional neural network flow optimization algorithm consists of an input layer, a number of hidden layers and an output layer, in which the hidden layer usually consists of a convolutional layer,

pooling layer and fully connected layer. The hidden layers execute actions on the input data and use the result as input for the next layer. Each node of the fully connected layer is connected to each node of the previous layer with certain weights, where the operations performed can be expressed as:

$$h_{t+1} = f(W_l h_t + b_l) \quad (7)$$

Where h_t , W_l and b_l represent the output of layer l , the weight matrix and bias, respectively, the input layer represents $h_0 = x$, x is the input of the convolutional neural network flow optimization algorithm, and f represents the activation function that operates on the input data element by element. The commonly used activation functions are sigmoid function, softmax function, linear rectifier function, hyperbolic tangent function.

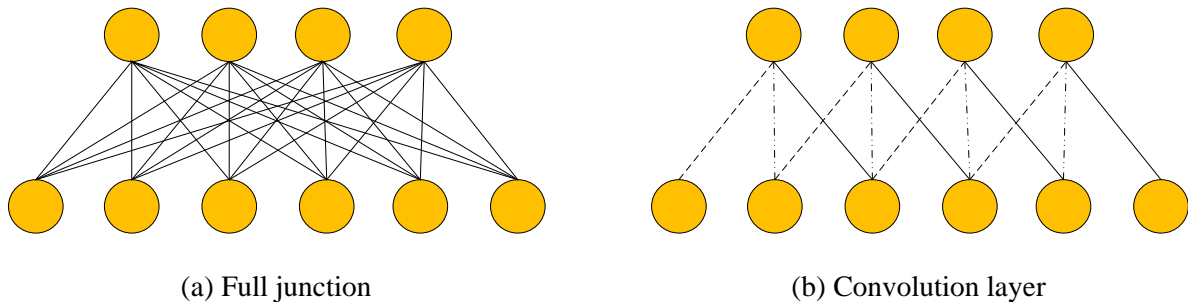


Figure 2. The whole connecting layer is compared to the convolution layer

3.2.2 Markov Algorithm (HMM)

The HMM observation variables and hidden states are shown in Fig. 3, with state sequences $S = \{S_1, S_2, \dots, S_N\}$, u_i denoting the system state values at moment t , i.e., hidden state $U \in \{u_1, u_2, \dots, u_n\}$, observation sequence $O = \{O_1, O_2, \dots, O_M\}$, and set of observations $V = \{v_1, v_2, \dots, v_N\}$. A complete HMM consists of the following five elements: hidden state S , observation state O , initial probability matrix π , hidden state probability transfer matrix A , and emission probability matrix B , where A and B are shown in Eqs. (8) and (9).) are shown. Further the HMM is abbreviated as:

$$A = \{a_{ij}\}, a_{ij} = P(u_{t+1} = S_j | u_t = S_i), 1 \leq i, j \leq N \quad (8)$$

$$B = \{b_j(k)\}, b_j(k) = P(O_t = v_k | S_t = u_j), 1 \leq j \leq N, 1 \leq k \leq M \quad (9)$$

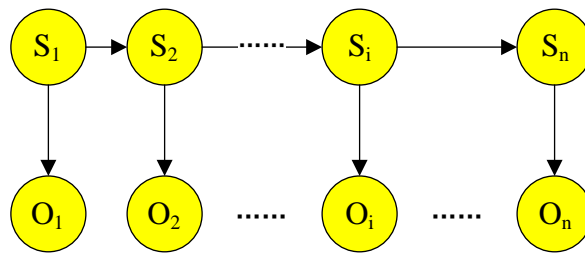


Figure 3. HMM observation and anonymity

First, define Viterbi variable $\delta_t(i)$, which indicates that after a certain sequence of hidden states is reached S_t at the moment of t passing through a sequence of hidden states to find the sequence of hidden states that maximizes the probability of its observation sequence $\{O_1, O_2, \dots, O_t\}$ as shown in Eq. (10). $\delta_t(i)$ The operation can be performed iteratively, as shown in equation (11), and finally the hidden state sequence with maximum probability is disciplined by equation (12):

$$\delta_t(i) = \max_{S_1, S_2, \dots, S_{t-1}} P(S_1, S_2, \dots, S_t = u_i, O_1, O_2, \dots, O_t | \lambda) \quad (10)$$

$$\delta_{t+1}(j) = \left[\max_{1 \leq i \leq N} \delta_t(i) a_{ij} \right] b_j(O_{t+1}) \quad (11)$$

$$\varphi_{t+1}(j) = \arg \max_{1 \leq i \leq N} \left[\delta_t(i) a_{ij} \right] \quad (12)$$

4 Research on automatic music transcription in piano teaching based on CNN-HMM

4.1 CNN-based acoustic model construction and implementation

4.1.1 Multi-task learning

The different structures of note-level transcription are shown in Fig. 4, where (a) ~ (c) is the note pitch, note onset and note cutoff, respectively. Feature triage can be done before the output layer if all targets have a network model. In multitask learning in note-level transcription, the three targets extracted are note pitch, note onset and note cutoff, and the combination of the three targets forms a complete note with closely linked relationships, and the labels of note onset and note cutoff are more sparsely labeled in the dataset compared to note pitch, and the dense labeling of the note pitch will be helpful in detecting note onset and cutoff.

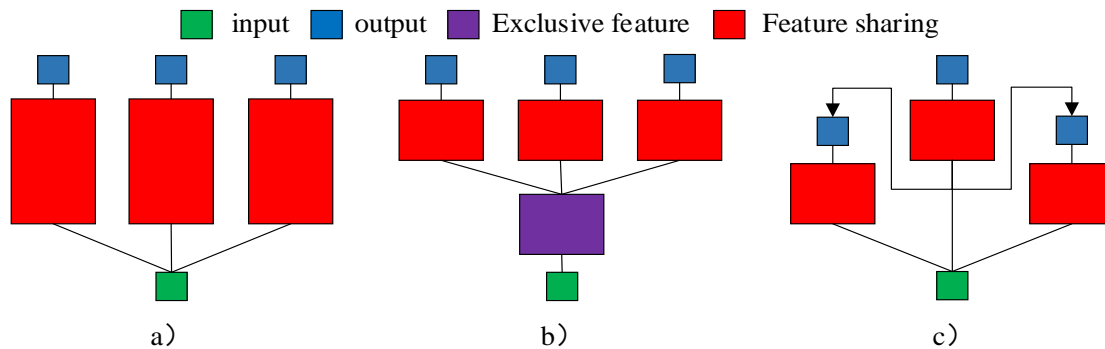


Figure 4. Different structures of the note level transcription

In this paper, we choose the hard sharing model in multi-task learning, where the shared features are diverted at some intermediate layer, the shared module at the lower level is used to compute the shared features, and the private module at the higher level is used to compute the exclusive features. The CNN model-based acoustic model coding process proposed in this paper is shown in Fig. 5, where the input is a spectrogram of a small segment, and an encoder constructed using the CNN-based multitask learning model is used to extract the feature mappings of the piano note's pitch, note onset and cutoff at the same time.

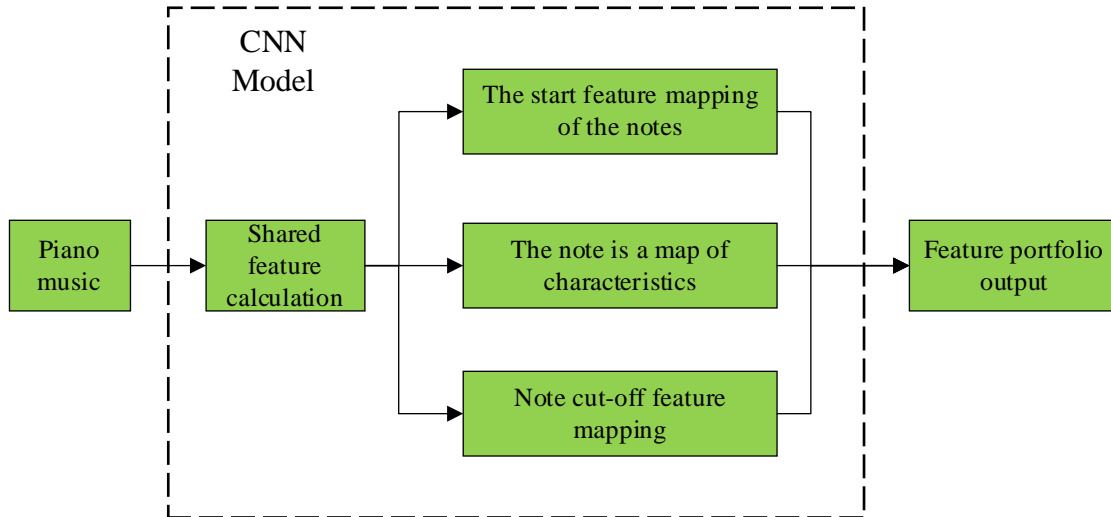


Figure 5. The acoustic model coding process based on CNN model

4.1.2 CNN network model structure

The structure of the CNN model is shown in Fig. 6, where (a) ~ (c) are the convolutional blocks, the average pooling layer, and the linear full connectivity, respectively. The shared feature computation part consists of five convolutional blocks. Each layer of convolution is followed by batch normalization, ReLU function activation and Dropout in order; the average pooling layer follows the second convolutional block, and the private feature computation part consists of one convolutional block and one layer of linear full connectivity.

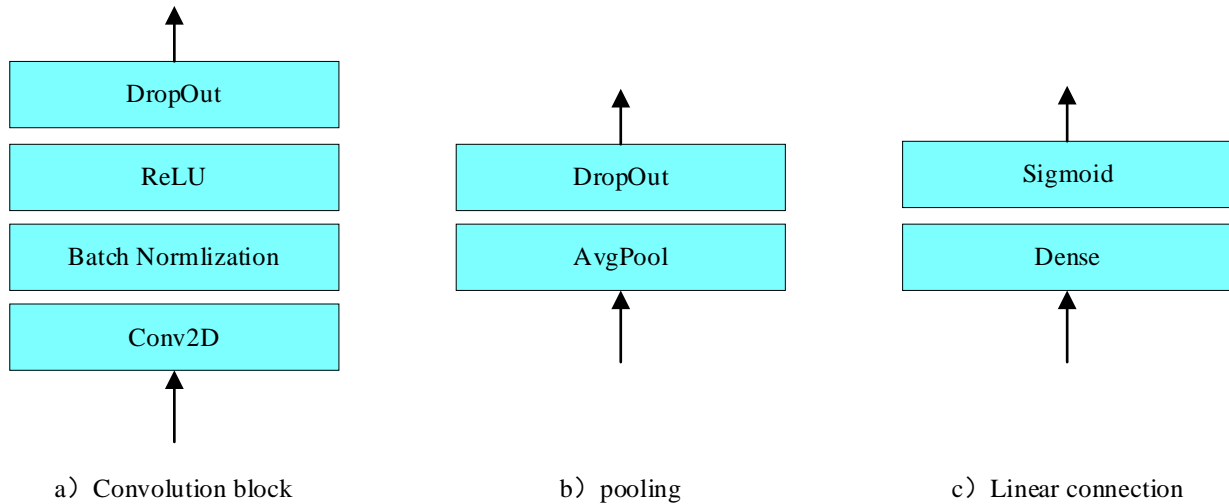


Figure 6. CNN model structure

The structure and parameters of the CNN model are shown in Table 1. The CNN model input $x_i \in \mathbb{R}^{13 \times 328}$ is a small segment of the spectrogram of the piano audio. The target matrix $y_i \in \{0,1\}^{88 \times 3}$ consists of vectors y_i^{on} , y_i^{int} , y_i^{off} , where t moments are the intermediate frame moments of the 13-frame spectrogram, i.e., the note onset, the note pitch, and the note cutoff state at the 7th frame, and for the piano note $k \in \{0, \dots, 87\}$, $y_i^k \in \{0,1\}^{1 \times 3}$ represent the value of the label of practical

information of the note at the t th moment, and the predicted value $y_t = \left\{ y_t^{on}, y_t^{int}, y_t^{off} \right\}$ is the activation output after the fully connected layer, where $y_t^k \in [0,1]$. In the shared feature layer, 30 convolution kernels are used for computation in each convolution, the first two layers are convolved using 3×3 convolution kernels, which are able to learn both temporal and frequency features simultaneously, and average pooling is used to reduce the data dimensionality to prevent overfitting, and 1×35 and 7×1 convolution kernels are used to learn the frequency features and temporal features, respectively. In the private feature layer, the learned time and frequency features are first fused using a convolutional kernel containing $10 \times 3 \times 7$, and then a linear fully connected layer is used to map the input data into a 1×88 vector output, which is squeezed between $[0, 1]$ by a Sigmoid activation function.

Table 1. CNN model structure and parameters

| Characteristic layer | input | CNN | output |
|-----------------------|---------------------------|---|---------------------------|
| Shared feature layer | $1 \times 13 \times 328$ | Conv $30 \times 3 \times 3$, BatchNorm, ReLU, Dropout | $30 \times 11 \times 326$ |
| | $30 \times 11 \times 326$ | Conv $30 \times 3 \times 3$, BatchNorm, ReLU, Dropout | $30 \times 9 \times 324$ |
| | $30 \times 9 \times 324$ | AvgPool 1×2 , Dropout | $30 \times 9 \times 162$ |
| | $30 \times 9 \times 162$ | Conv $30 \times 1 \times 35$, BatchNorm, ReLU, Dropout | $30 \times 9 \times 128$ |
| | $30 \times 9 \times 128$ | Conv $30 \times 7 \times 1$, BatchNorm, ReLU, Dropout | $30 \times 3 \times 128$ |
| | $30 \times 3 \times 128$ | Conv $30 \times 1 \times 35$, BatchNorm, ReLU, Dropout | $30 \times 3 \times 94$ |
| Private feature layer | $30 \times 3 \times 94$ | Conv $10 \times 3 \times 7$, BatchNorm, ReLU, Dropout | $10 \times 1 \times 88$ |
| | $30 \times 3 \times 88$ | Dense 88, Sigmoid | 1×88 |

4.2 HMM-based modeling of music language in piano teaching

4.2.1 Envelope curves of notes

When a piano key is struck, the sound produced is a superposition of the cents corresponding to the fundamental frequency and the overtones corresponding to its octave. When an individual note is played, a compound sound is produced that changes continuously over time. At the beginning of a note, there is often a sudden increase in energy, and during this short onset phase, the sound is gradually built up, with the ADSR envelope curve shown in Figure 7. In the case of a piano performance, hitting a piano key triggers the entire chain of mechanical actions, and then the percussion mallet strikes one or more strings. Starting from the finger touching the keys and ending with the percussion mallet hitting the strings, the generated mechanical noise will be acoustically blended with the musical notes emitted by the strings. After the onset phase, the pitch sound of the note slowly and steadily degrades until it reaches a stabilization phase with a cyclic pattern. The third phase, also known as the maintenance phase, constitutes the majority of the duration of the note, in which the energy remains more or less constant while the evolution of the piano note is slightly decreasing. The final stage of the note is referred to as the release stage, where the note's pitch fades away. The release phase begins when the finger pressing the key leaves the key.

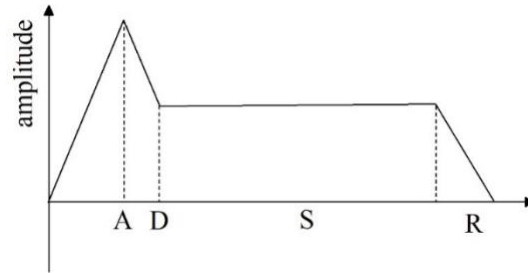


Figure 7. ADSR envelope curve

4.2.2 Modeling piano notes

In StbgTGd2, the amplitude envelope and energy envelope of a single note in which a sustain pedal was used were selected, and the waveform and envelope of the StbgTGd2 note are shown in Fig. 8, with the curve in the waveform as the energy envelope, and the audio sampling rate of 44.1 kHz. In calculating the energy envelope in the time domain, which was obtained by calculating the root-mean-square energy, the frame length of 4,095 samples was selected, and the time shift of 881 samples was chosen. 881 samples. It can be seen that, in contrast to the ADSR envelope of Fig. 7, there is no stable yet observable sustaining phase in the piano's single-note waveform plots for either the amplitude envelope or the energy envelope. In particular, the amplitude variations in the foremost and last parts of the waveforms of the notes with piano key values MIDI of 46, 49, 59, 66, 68, etc., are due to the collisions between the keys and the keyboard resulting from the touching of the keys between the years of the finger pressing the keys and the years of the finger releasing the keys. Ideally, this paper gives a note modeling that is more consistent with the evolution of piano notes over time, and the TDR envelope curve is shown in Fig. 9. After the note onset, the oscillatory waveform envelope continues to increase, and the audio signal evolves rapidly in some unpredictable way, which corresponds to a transient phase containing the onset of the note and the sharply decaying part. Even with the addition of a sustain pedal while playing, the sustain portion of the note is not particularly noticeable, so the decay and sustain phases of ADSR are collectively referred to as the decay (D) phase. The moment the piano keys are released, the strings gradually stop vibrating but still sound for a duration related to the use of the sustain pedal; this phase is called the release (R) phase when the sound gradually diminishes and tends to become silent.

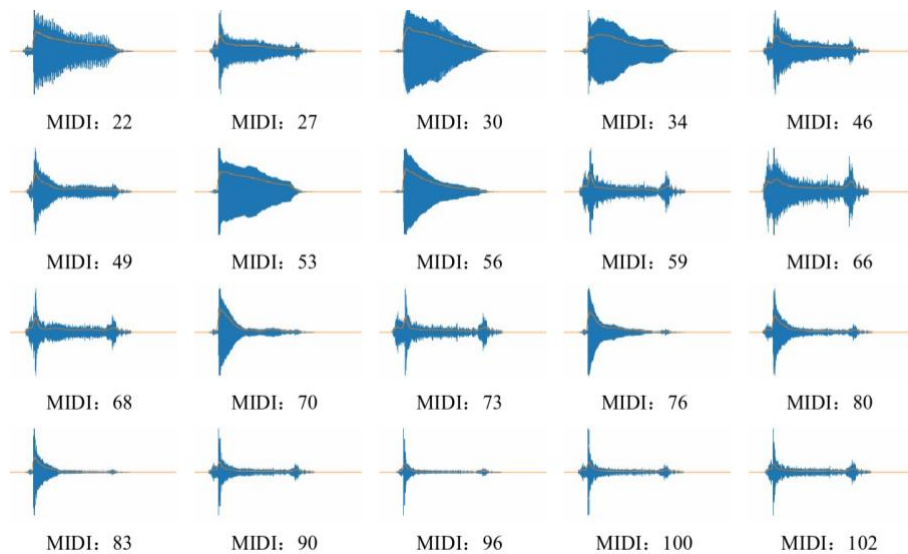


Figure 8. StbgTGd2 note waveform and envelope

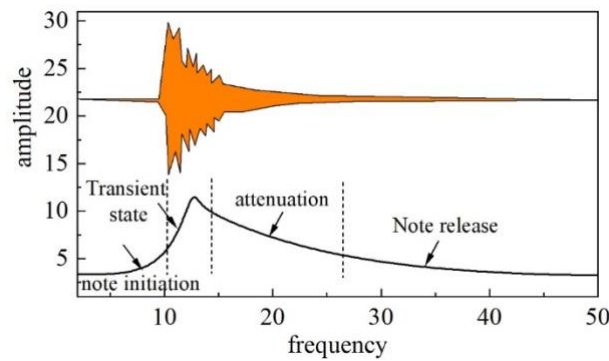


Figure 9. TDR envelope

4.2.3 HMM-based music language modeling

Based on the encoding of note TDR envelope curves and CNN acoustic models above, this section will illustrate the HMM-based music language model with a complete note. For a complete independent note k , based on the TDR envelope curve proposed above, the note can be divided into three stages, namely the transient stage, the decay stage, and the release stage. The HMM state transfer process of the note is shown in Fig. 10. The state transfer matrix needs to be manually adjusted and applied to all notes, and each piano note k will be individually decoded into a series of segments. For the decay and release phases, due to the large span of the decay phase, the release phase will have the effect of the sustain pedal, which is set to have a certain probability of returning to its state during the state transfer. When a note is pressed twice in a row, the first note is allowed to return directly to the beginning of the note after the decay phase because the first release phase will be covered by the second onset part, which is not easy to be detected, i.e., the state of the note is allowed to return directly to state 7, from state D, without going through the state R, because the probability of the emission corresponding to state T, at this point is much higher than that of state R. The probability of firing in the T, state is much higher than that in the R state.

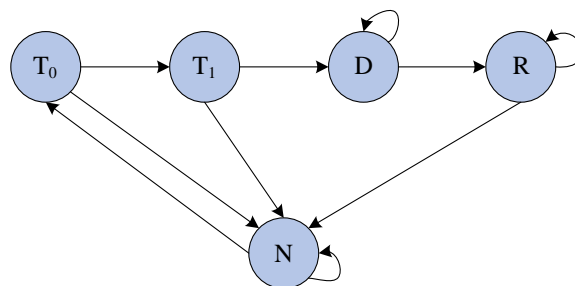


Figure 10. The process of transition on HMM states in one note

5 Empirical analysis of piano teaching based on CNN-HMM

5.1 Simulation experiment setup

5.1.1 Experimental configuration

The experiments are done on PCs with Windows 10 as the operating system and Inter(R) Core(TM) i5-9400F as the CPU. The GPU is NVIDIA GeForce RTX 2080Ti, invoking the parallel computing

architecture CUDA (Compute Unified Device Architecture) and the NVIDIA deep neural network library cuDNN (CUDA Deep Neural Network).

The inputs to the network model are small segments of spectrograms in spectrogram $x_i \in R^{c \times b}$ with a size of 128 per batch. The weights are initialized using the Xavier initialization method before the stochastic gradient descent method is used to learn the neural network parameters, with the momentum factor set to 0.9, the initial learning rate set to 0.1, and the weights decayed by 1e-5, and the learning rate halved when the performance of the network model is no longer improved in 10 iterations.

5.1.2 Evaluation indicators

Three evaluation metrics commonly used in the field of automatic music transcription are as follows: accuracy, recall, and F1 value. Each parameter used for evaluation is defined, and the formulas for several of them are given below:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (13)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (14)$$

$$F1 = \frac{2 + P + R}{P + R} \quad (15)$$

Where N_{TP} denotes the number of correctly predicted samples, N_{FP} denotes the number of positively predicted samples, and N_{FN} denotes the number of negatively predicted samples. The evaluation of the results of the following experiments is based on these parameters. However, for the following experimental results, the meaning of the parameters of N_{TP} , N_{FN} and N_{FP} in the calculation of evaluation indexes P , R and $F1$ is different in different cases, which needs to be treated differently.

5.1.3 Evaluation data sets

Real piano recordings have been used as test datasets in many MPAS datasets to validate the environmental independence of the algorithm. In this paper, we also follow this criterion and use real emotion recordings for testing and evaluation in all the experimental evaluations. Consistent with the generalized evaluation approach in MIREX competitions and academic papers, in this paper, we first calculate the value of each evaluation metric for each audio in the test set separately and then take its average as the final evaluation result of the whole test set, and then finally compare the performance of the final model or algorithm, usually based on the F1 value.

5.2 Model evaluation

5.2.1 Detection of polyphonic onset times in piano teaching

For the polyphonic onset time detection model, the performance of the model is a key factor in the performance of the polyphonic onset time detection module, and the complexity of the task, which detects both note onset time and pitch, requires maximizing the performance of the model. In order

to explore the optimal structure of the multi-tone onset detection model, we investigate the number of input data channels, the number of convolutional layers of the CNN-HNN network, and the positive sample weighting of the loss function, respectively. The experimental findings for various conditions are as follows:

For the effect of the number of input data channels on the model performance, Fig. 11 shows the evaluation results of the multi-tone onset time detection model under mono and dual-channel conditions, (a) and (b) for mono and dual-channel, respectively. It can be seen that the dual-channel audio results are improved by 2.68% in terms of F1 value, 3.18% in terms of accuracy, and 2.61% in terms of recall over the mono audio results. This is due to the fact that the dual-channel audio has richer information than the mono audio, which allows the CNN-HNN network to learn richer features, thus improving the pitch and note recognition results.

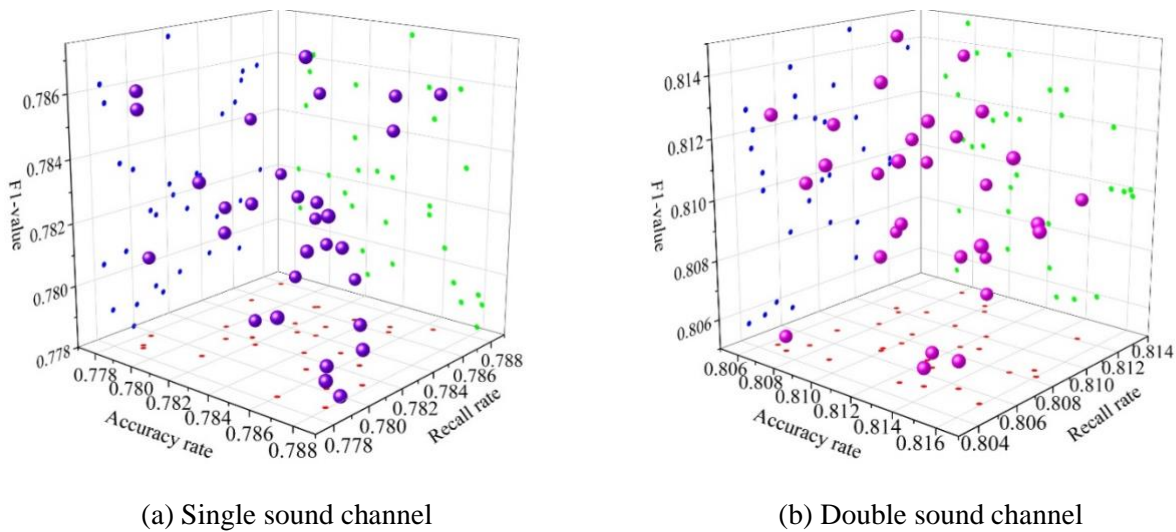


Figure 11. Evaluation of the sound and double sound conditions

5.2.2 Evaluation of public start time detection models

For the public start time detection model, we mainly explored the effect of the size of the sliding window on the model performance. Since we use the middle frame time of the sliding window to represent the input data time, it will cause ambiguity if the window length is even, so we choose odd window lengths such as 3, 5, 7, 9, etc. as the super reference experiment. Figure 12 shows the evaluation results of the public start time detection model with different sliding window lengths. The optimal sliding window length of 5 frames can be obtained. The public onset time detection detects the edge of the spectrogram by CNN, which requires sufficient length of spectral information. When the sliding window is 3 frames, the sliding window is too small, the information is insufficient, and it is easy to miss the note onset event, so its recall is small. When the sliding window increases from 5 to 9, their accuracy stays similar, but the recall decreases. The process of increasing the sliding window leads to similar starting notes being easily merged into one.

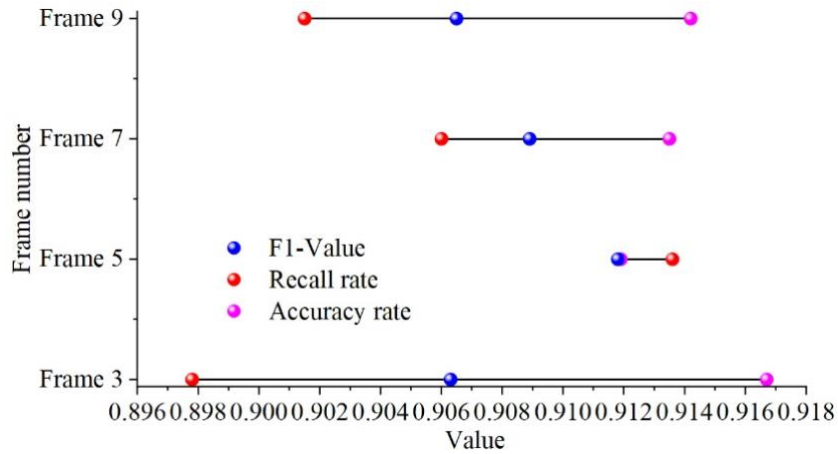


Figure 12. The results of the public start time test model

5.2.3 Evaluation of frame-level multi-tone detection models

In order to show the role of environmental information in frame-level polyphony detection, this paper demonstrates the comparison of the evaluation results of the frame-level polyphony model with a single-frame input and a 9-frame input. The evaluation results of the frame-level polyphony detection model are shown in Fig. 13, where (a) is a single frame and (b) 9 frames. When using a 9-frame spectrum as input, the frame-level evaluation results obtained from the training improved the accuracy by 2.83%, recall by 5.00%, and F1 value by 3.03% compared to when a single frame is used as input. This is because when multiple frames of spectrum are used to form a spectrogram as input, some of the additional spectral information can provide more features and also prevent the bias problem caused by less accurate labeling, and with these two aspects, better frame-level results can be achieved when 9 consecutive frames of spectrum are used as input.

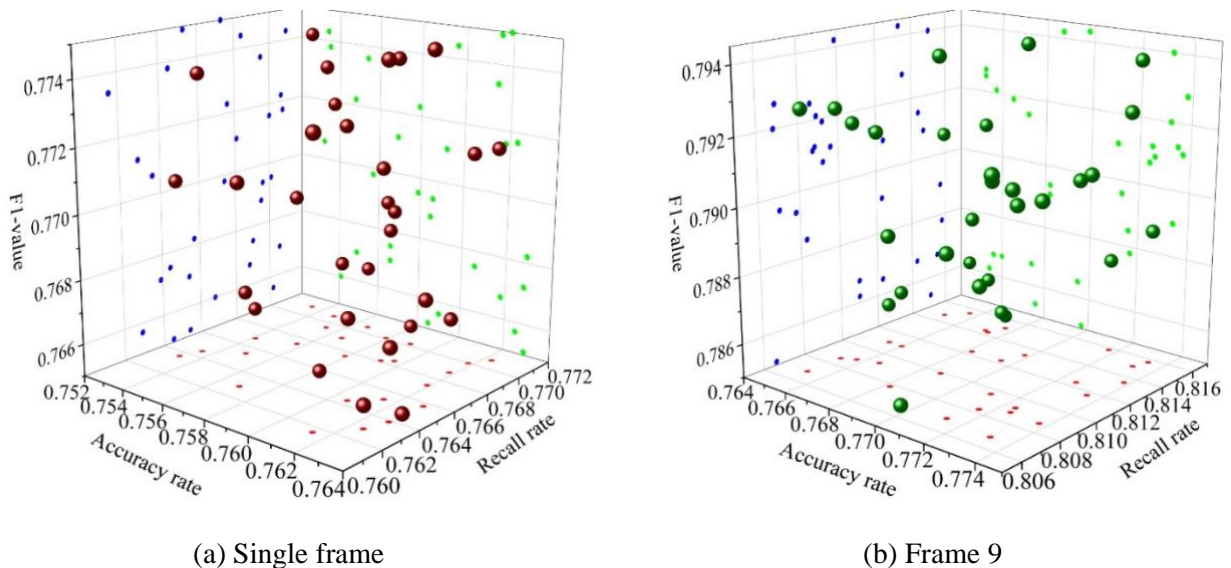


Figure 13. The evaluation results of the frame level multitone detection model

5.3 Analysis and Strategies for Evaluating the Effectiveness of Teaching the Integration of Electronic Organ and Piano

5.3.1 Evaluation and Analysis of the Effectiveness of Teaching the Integration of Electronic Organ and Piano

In order to verify the application performance of the model in this paper in realizing the quantitative evaluation of the effect of combining the teaching concepts of electronic pipe organ and piano, the time domain distribution of the quantitative analysis of the effect of combining the teaching concepts of the electronic pipe organ and piano is shown in Fig. 14, which can be seen that the quantitative value of the evaluation of the integration of the teaching of the electronic pipe organ and piano stays above 0.35 when the sampling points are in the range of 875 to 880, and the model can effectively promote the electronic organ and piano teaching concepts, the confidence level of quantitative regression analysis is high, the teaching quality evaluation results are good, and it provides a reliable model basis for promoting the cultivation of electronic organ performance talents.

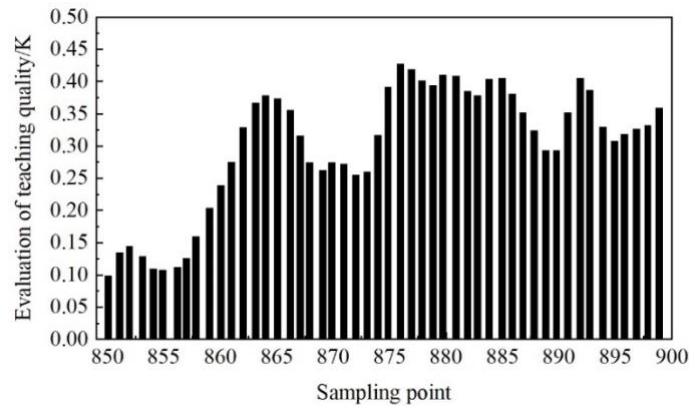


Figure 14. The time domain distribution of electronic tube organ and piano fusion teaching

Then take the data in Figure 14 as the test sample, analyze the confidence level of teaching quality evaluation of the combination of electronic organ and piano teaching concepts, and the confidence level results of the integration of teaching effect of electronic organ and piano are shown in Figure 15, which can be seen that with the increase of the sampling points the confidence level results stabilize between 0.23 and 0.38, i.e., it shows that the model can effectively promote the combination of teaching concepts of the electronic organ and the piano, the confidence level of quantitative regression analysis is high.

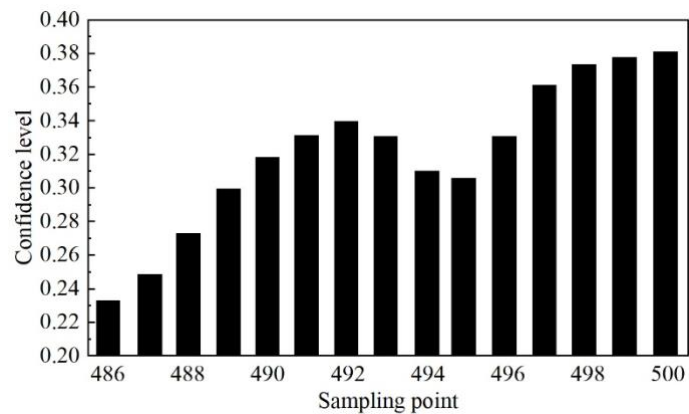


Figure 15. Results of electronic tube organ and piano fusion teaching effect

In order to further verify the validity of the model in this paper, the reliability of the teaching quality evaluation results was comparatively analyzed by using the CNN-HMM-based quantitative assessment model of promoting the combination of the teaching concepts of the electronic organ and the piano and the traditional model, and the comparative results of the reliability of the evaluation results are shown in Figure 16. It can be seen that the reliability of the teaching quality evaluation results in the traditional model ranges between 11.5% and 32.4%. The reliability of the teaching quality evaluation results of this paper's CNN-HMM-based quantitative assessment model for promoting the combination of teaching concepts of electronic organ and piano is between 78.1% and 85.4%. The reliability of the teaching quality evaluation results of this paper's model is higher than that of the traditional model, indicating that this paper's model has better teaching quality evaluation results.

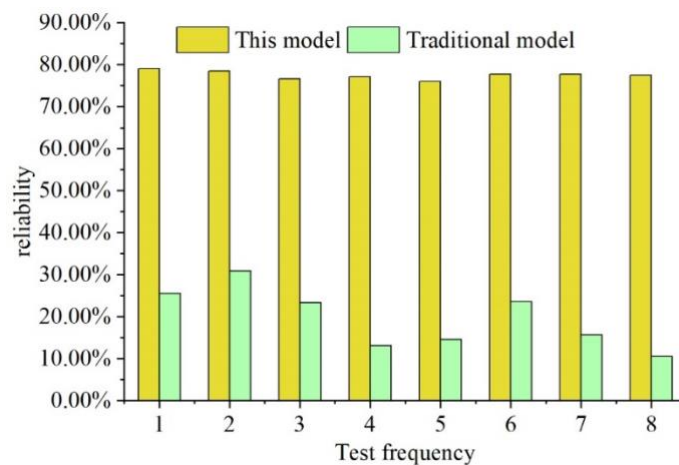


Figure 16. The reliability comparison results of the evaluation results

5.3.2 Teaching Strategies for the Integration of Electronic Organ and Piano

In the piano program, the practice of teaching materials as intensive training is very important to improve the students' playing ability. Piano playing is the process of practice making perfect. After having the basic playing ability, you still need to practice a lot to have the ability to play. Generally speaking, the first exercise is to practice the exercises, then to practice polyphony, music class training, and finally to practice large-scale sonatas and concertos. Exercises focus on the training of all kinds of basic playing skills. They generally require accurate intonation, finger placement, and skillful techniques. This is the basic training content. Polyphonic training is a relatively complex training after mastering the basic skills, training students to think logically about the level of music and the ability to control the fingering techniques of multiple parts. Music training involves mood training, which allows students to experience the emotion of music and exercise their imagination and musicality. Finally, the large-scale sonata and concerto training is mainly designed to exercise the students in the comprehensive application of various techniques and the ability to organize the performance. Through gradual practice from easy to difficult, piano majors eventually can perform independently.

6 Conclusion

Piano teaching in colleges and universities has a late start compared with the teaching of other professional courses and is under-appreciated in all kinds of majors and the results of the teachers' teaching quality evaluation are poor, and the students' learning effect is not very satisfactory. Therefore, this paper proposes a CNN-HMM-based quantitative assessment model for the promotion

of the combination of teaching concepts of electronic organ and piano, information collection and adaptive feature extraction of the feature volume of the quantitative assessment of the combination of teaching concepts of electronic organ and piano, combined with the segmented sample detection method for the quantitative assessment of the combination of teaching concepts of electronic organ and piano, the introduction of statistical analysis of quantitative assessment of the data of the combination of teaching concepts of electronic organ and piano by introducing the CNN-HMM method, to achieve the quantitative assessment of the combination of teaching concepts of electronic organ and piano. Understand the multi-parametric constraint statistical analysis of the combination of electronic organ and piano teaching concepts. The two-channel audio results improve the F1 value by 2.68%, the accuracy by 3.18%, and the recall by 2.61% over the mono audio results. This is due to the fact that dual-channel audio has richer information than mono audio, which allows the CNN-HNN network to learn richer features, thus improving the pitch and note recognition results. From 78.1% to 85.4%, the CNN-HMM-based quantitative assessment model's teaching quality evaluation results were reliable in facilitating the combination of electronic organ and piano teaching concepts. The model can effectively promote the combination of teaching concepts of electronic organ and piano, the confidence level of quantitative regression analysis is high, and the teaching quality evaluation results are good.

References

- [1] Yang, L. (2020). Reform of the piano teaching methods in the internet environment. *Basic & clinical pharmacology & toxicology*.(S3), 126.
- [2] Chen, L. (2019). Development of the indoor piano performance training in the piano teaching design of colleges and universities. *Basic & clinical pharmacology & toxicology*.(S2), 125.
- [3] Fang, X. (2019). Application strategy of the new media technology in the piano teaching. *Basic & clinical pharmacology & toxicology*.(S2), 125.
- [4] Ying, J. (2016). Methodical support of future music teachers' aesthetic development in the process of teaching piano. *Science & Education*, 30(1), 51-56.
- [5] Lui, C. (2011). Case study of teaching and playing piano duet to improve overall rhythmic sense for mental handicap student. *Geologica Carpathica*, 62(1), 77-90.
- [6] Elgersma, & K. (2012). First year teacher of first year teachers: a reflection on teacher training in the field of piano pedagogy. *International Journal of Music Education*, 30(4), 409-424.
- [7] Guo, R., Ding, J., & Zang, W. (2021). Music online education reform and wireless network optimization using artificial intelligence piano teaching. *Wireless Communications and Mobile Computing*.
- [8] Niu, Y. (2021). Penetration of multimedia technology in piano teaching and performance based on complex network. *Mathematical Problems in Engineering*, 2021.
- [9] Hou, Y. (2022). Research on piano informatization teaching strategy based on deep learning. *Mathematical Problems in Engineering*, 2022.
- [10] Chen, Y., & Zheng, N. (2020). Ai based research on exploration and innovation of development direction of piano performance teaching in university. *Journal of Intelligent and Fuzzy Systems*, 40(1), 1-7.
- [11] Liu, M., & Huang, J. (2020). Piano playing teaching system based on artificial intelligence – design and research. *Journal of Intelligent and Fuzzy Systems*, 40(1), 1-9.
- [12] Sun, S. (2021). Evaluation of potential correlation of piano teaching using edge-enabled data and machine learning. *Mobile Information Systems*.
- [13] Pike, P. D. (2013). The differences between novice and expert group-piano teaching strategies: a case study and comparison of beginning group piano classes. *International Journal of Music Education*, 32(2), 213-227.

- [14] Simones, L., Schroeder, F., & Rodger, M. (2015). Categorizations of physical gesture in piano teaching: a preliminary enquiry. *Psychology of Music*, 43(1), 103.
- [15] Coutts, L. (2018). Selecting motivating repertoire for adult piano students: a transformative pedagogical approach. *British Journal of Music Education*, 35.