Research Article

Application of 5G Internet of Things Multisensor Information Fusion Model in Piano Performance

Tianshu Wang

Henan Vocational University of Science and Technology, Zhoukou, Henan 466000, China

Correspondence should be addressed to Tianshu Wang; wangts@henu.edu.cn

Received 18 May 2022; Revised 13 June 2022; Accepted 15 June 2022; Published 28 June 2022

Academic Editor: Gengxin Sun

Copyright © 2022 Tianshu Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we use melodic multisensor information fusion combined with 5G IoT to conduct an in-depth study and analysis of the experience model of piano performance. In the form of multimodal data, two main storage forms, audio and MIDI, are chosen. First, audio signal processing technology and deep learning technology are used to extract shallow and high-level feature sequences in turn, and then, the alignment of the two modal data is completed with the help of sequence alignment algorithm. For the problem that encrypted data cannot be queried by uploading blockchain, this paper proposes an IoT encrypted data query mechanism based on blockchain and Bloom’s filter. The blockchain stores IoT encrypted indexes by temporal attributes to ensure data consistency, tamper-evident, and traceability. A new loss function training multimodal model is designed for piano performance signals. The piano performance generated by this model differs from the traditional piano performance generation in that it does not need to add complex piano performance rules manually but generates piano performance directly with piano performance theory rules by training the initial piano performance dataset and improves the stability of the generated piano performance by chord constraints and enhances the note dependence on time. In the analysis of the experimental results, the generated melodies were invited 50 for evaluation and analysis. The overall style-based GAN network piano performance generation model proposed in the study makes the generated piano performance melodies more pleasing to the ear through chord constraints and the content of autonomous learning moments, which has important theoretical and practical implications for the creation and realization of mass and batch piano performances.

1. Introduction

In the traditional network management mechanism, the work of each network sensor is independent of each other, and the effective information is not extracted and integrated, so it cannot reflect the globality of the network status, which not only does not achieve the effect of strengthening management but also increases management. The burden on the staff needs to analyze the detection results of multiple sensors. Network management must be able to collect data from multiple sensors in a complex and changeable network environment, standardize, fuse, and evaluate the collected uncertain information, and present the evaluation results in a visual way to assist network management. People make quick decisions and fix cybersecurity issues as they arise. Every day in everyday life, massive amounts of data are constantly being generated, and the variety of data is becoming increasingly complex. At the same time, this wave is disrupting the piano performance market and the piano performance industry is undergoing a radical change [1]. From the original recordings to today’s digital piano playing, piano performance works have entered the era of big data. With the advent of increasingly powerful computing systems and larger storage media, the study of piano performance based on big data has become a possibility. In an era of rapid computer development, the form of storage for piano performance is also being revolutionized [2, 3]. The same piano piece can be stored in various forms: video, audio, lyrics, and pentameter. Two of the most popular forms of storage are the sampled audio file and the piano notation file such
as MIDI, which is a new form of storage for piano performances that records the pitch, start time, and duration of each note. Piano playing is an important form of artistic expression in our life, which is closely related to our life and becomes an indispensable part of it. It can soothe the mood and express emotions, and different tunes and melodies can express different emotions [4].

With the rapid development of electronic information technology, data fusion technology has also been introduced into the field of civil engineering as an intelligent processing technology to eliminate uncertain factors and provide accurate observation results and new observation information, such as: intelligent detection systems, industrial process monitoring, robots, air traffic control, and disease diagnosis.

The industrial wave represented by computers, the Internet, and mobile communication networks has passed, and now we are in a period of rapid development and wide application of the Internet of Things industry [5, 6]. The initial concept of IoI comes from a sensor network, that is, combining sensors and communication networks to connect real objects into the Internet and realize online management of objects. In short, IoT is a technology that maps objects in the real world into the virtual network world to achieve virtual integration of the entire physical world. This technology has grown rapidly in recent years and has even started the trend of interconnecting everything [7].

The common local differential privacy mechanisms for typed data are summarized, and the PK-RR mechanism is proposed to address the problem that the frequency estimation variance of the K-RR mechanism increases as the size of the attribute domain k increases, and the frequency estimation variance in the K-RR mechanism is improved by calculating the optimal solution of the frequency estimation, and the PK-RR mechanism is added to the proposed personalized local differential privacy method for wearable devices, so that the proposed method can achieve privacy protection for both numerical and subtyped data [8, 9]. Secondly, it requires a lot of human and material resources as well as geographic constraints and limitations in the creation process. By using deep neural networks to assist composers in creating piano performances, it can greatly save the creation time, the creative environment is not bound by geographic constraints and limitations, and inspire composers to be more inspired and better show their ideas in the creation process, to create more interesting and pleasant piano performances.

2. Related Works

There are two types of score alignment, online and offline. Online alignment, also known as score following, is usually applied in a real-time performance environment. Compared with online alignment, offline alignment is not limited to the current moment, and the feature data of the whole piano performance piece can be used in the alignment process [10]. In this paper, we focus on offline alignment, which has many applications in the field of piano performance information retrieval. In addition to "listening to songs" described above, offline alignment can also be applied to database search and electronic repositories. According to the modalities of the data to be aligned, score alignment can be classified as audio-to-audio alignment, audio-to-MIDI (MIDI) alignment, and audio-to-sheet alignment. To solve the problem that the charRNN method can only be trained for monophonic music, Chen and Zheng proposed a method to convert MIDI music into a music description language based on certain grammatical rules, which makes charRNN applicable to polyphonic music, and use charring to complete text training and obtain a music generation model [11]. The process of reconstructing music sequences can be mapped to deep learning algorithms. Except for audio-to-audio alignment, the other two alignment tasks involve multimodal data. Such musical features can be specially designed or postprocessed to extract the "Onset" information of the notes. For example, in the study by Mi et al., a chromatic onset-based feature was designed [12].

Information fusion technology has developed into a common key technology that attracts attention from many parties, and many hot research directions have emerged. Many scholars have devoted themselves to theoretical and applied research in the fields of maneuvering target tracking, distributed detection fusion, multisensor tracking and positioning, release information fusion, target recognition and decision-making information fusion, situational assessment, and threat estimation. A batch of multtarget tracking systems and multisensor information fusion systems with preliminary comprehensive capabilities have appeared one after another. With the gradual development of information fusion (data fusion), it is more and more used in more fields.

For the task of score alignment, the input to the model is often multiple modal data, so most studies use deep learning-based multimodal models to complete the combination and abstraction of shallow features. Based on neural networks, music reconstruction has been better developed in deep learning, and the combination of "MIDI + Recurrent Neural Network (RNN)" has become the mainstream of music generation methods. However, there is a problem of gradient disappearance when RNN generates long time sequences, and Takahashi et al. used long short-term memory recurrent neural network (LSTM) to generate drum rhythm sequences to solve this problem [13]. Xu et al. researched deep learning methods for the polyphony estimation problem and music generation problem of piano and proposed a nonnegative multimodal dictionary learning and sparse dictionary with encoded Lorentzian-Block Frobenius parametric constraints and incoherence/incoherence constraints for the polyphony estimation problem [14]. Qiu et al. further investigates deep learning-based methods for reconstructing melodies and arrangements of music. Such multimodal models are usually designed with two parallel deep neural networks, which extract the high-level abstract features of the two modal data separately and finally map the multimodal data onto a common subspace [15]. The multimodal model-based curvilinear alignment methods take advantage of the powerful feature abstraction capability of deep learning and the parallel network structure and
gradually replace the traditional alignment methods and become a mainstream model design. The ECG signals in eight emotional states, such as quiet, angry, sad, fear, disgust, calm, surprised, and funny, were collected. After filtering out the noise and extracting the features in the time and frequency domains, a support vector machine was used to build the emotion recognition model [16]. The results show that the recognition accuracy of the support vector machine can reach 60%~75% for different emotions after optimizing the parameters.

With the development of information fusion theory, this emerging interdisciplinary subject has developed from the initial multisensor fusion of robotic systems, processing signals from different or the same sensor to obtain global long-term fusion data of objects. From the perspective of more general application objects, it has gone through a long development process. In this process, it is found that the information from various sensors may have different characteristics, so a variety of different information fusion methods appear accordingly.

The proposed Generative Adversarial Network (GAN) model has caused quite a stir in the field of artificial intelligence. In the field of text generation, the relevant text description of images is further generated by generative adversarial networks after extracting the relevant features of images; in speech, the quality of speech generation is enhanced by generative adversarial networks in the presence of nonsmooth noise and unknown noise; in the field of piano performance, real piano performance data is learned by GAN models to generate pleasant and harmonious piano performance melodies.

3. Design of 5G IoT Multisensor Information Fusion Model

Multisensor information fusion is a common basic function in human or other logical systems. In the data fusion system that simulates the human brain to comprehensively deal with complex problems, the information of various sensors may have different characteristics, real-time or nonreal-time, fast or slow change, fuzzy or clear, and support or complement each other. The basic principle of multisensor information fusion is similar to the comprehensive processing of information by the human brain. It makes full use of various sensor resources. Redundant or complementary information is combined according to certain criteria to obtain a consistent interpretation or description of the object being measured.

Compared with all single-sensor signal processing or low-level multisensor data processing, multisensor information fusion system is a low-level imitation of human brain information processing. Multisensor resources cannot be utilized as effectively as multisensor information fusion systems. A multisensor system can obtain a greater degree of detected target and environmental information.

The framework of the proposed multimodal model-based song alignment algorithm is shown in Figure 1 in this paper: first, audio and MIDI clips are preprocessed to extract primary features; then, they are input to a multimodal model consisting of two parallel deep convolutional networks, respectively, to obtain the learned high-level abstract features; finally, the two feature sequences are aligned using a suitable alignment algorithm to obtain the alignment path.

For an audio signal with a sampling rate of 44.1 kHz, the signal is first framed with a frame length of 2048 46.4 ms and a frame shift of 10 ms; before the Fourier transforms, a window is added to the framed signal and a Hamming window is taken, after which the time domain signal is converted to the frequency domain using the Fourier transform to obtain the STFT spectrum; to mimic the human ear’s perception of sound intensity, the STFT spectrum amplitude is then multiplied and the new spectrum is obtained by logarithmic compression; to reduce information redundancy, the logarithmically compressed spectrum is passed through a set of chromatic filters [17]. To mimic the human ear’s perception of sound intensity, the STFT spectrum amplitude is then 1000 multiplied and logarithmically compressed; to reduce information redundancy, the logarithmically compressed spectrum is passed through a chromatic filter set to obtain the new spectrum; to increase the dynamics of the features, the first-order difference of the spectrum is finally calculated, and the original spectrum is combined with the difference spectrum to obtain the final STFT features.

The chromatic filter set involved in the above extraction process is composed of a series of overlapping triangular filters. The center frequency of each triangular filter is the frequency value of a standard pitch, and the cutoff frequency is the frequency value of the adjacent pitch. Since the pitch distribution is exponentially distributed, the bandwidth of the filter increases gradually. It is important to note that the number of filter banks does not exactly match the number of notes [18]. Due to the frequency resolution, there are duplicate filters at low frequencies, which need to be removed. From the above, the window function used by CQT is time-varying, i.e., the window length taken at the low and high-frequency parts is different. For low-frequency signals, it has a higher frequency domain resolution and lower time-domain resolution; for high-frequency signals, it has a higher time-domain resolution and lower frequency domain resolution. In addition, CQT uses a frequency distribution that matches the twelve mean rhythms commonly used in Western music. For these reasons, CQT has become an important feature of music signal processing and analysis.

For the problem that the data structure of IoT does not conform to the blockchain structure, the index of encrypted data on the chain can only be compared and verified, but not queried. In this chapter, a query method based on Bloom filter and double-layer combined Bloom filter with higher efficiency than traditional query is proposed. The IoT data index is stored in the block corresponding to the block timestamp by time attribute, and the encrypted index is queried by using spatial location attribute, and the index contains unique public key encoding, plaintext hash, and encrypted data hash. The corresponding private key and encrypted data can be queried by public key encoding and encrypted data hash, and the plaintext data can be obtained.
after decryption and verification, so this query structure can be queried by time attribute. This query structure can query the required IoT data by temporal and spatial location attributes. At the same time, we use the characteristics of blockchain to realize the third-party trust, put the encrypted query index on the blockchain, store the encrypted data on the cloud storage server, verify the integrity and consistency of the data through the encrypted index on the chain, and reduce the overhead of trust calculation.

When the IoT nodes upload the verification data to the blockchain, the verification data will not be able to be tampered with; the data of previous blocks cannot be modified at the current time, so the data of all blocks except the current block are deterministic and do not need to be dynamically modified; then, the hash calculation server can construct the Bloom filter with optimal performance based on the determined amount of data. The criteria for an optimal performance Bloom filter are the lowest probability of false normality and the least memory occupation. Suppose a block $B$ is a block in the whole blockchain $B = \{B_1, B_2, B_3, \cdots, B_s\}$ in which there $n_i$ are a total of definite index data, and the length of the Bloom over filter $BF_i$ constructed for this block is $m_i$ such that the pseudonormality probability can be deduced from the existing situation, $p_{f_i}$ denoted by

$$
p_{f_i} \approx 1 - \left(1 + e^{n_i \times k_i}\right)^{k_i}.
$$

Assuming that the $r$th spatial location attribute to be written to the $BF_i$ Bloom filter is mapped to $BF_i$ the element using $k_i$ a hash function, each hash function mapping data to the Bloom filter to be set to 1. The insertion of a spatial location attribute to the Bloom filter using all hash functions $BF_i$ can be expressed as

$$
BF_i(h_i, e_i) = 1, \; (0 \leq i \leq k_i).
$$

Inserting all the spatial location attributes in the block into the corresponding Bloom filter $BF_i$ can be expressed as

$$
BF_i(h_j, r_j) = BF_i(l) = 1(0 \leq j \leq k_i \text{ and } 0 \leq i).
$$

---

**Figure 1**: Curve alignment algorithm flow for IoT multisensor information fusion.
According to the characteristics of IoT, data time attributes corresponding to block timestamps, querying data within a certain time range only requires querying data in the corresponding time block. Querying spatial location attribute \( y \) requires first determining the query time range after which the range of the lookup block is determined and then checking the existence of spatial location attributes for each block in the range \( y \). Checking the block \( B \) corresponding to the Bloom over filter \( BF \), the result of the mapping position of the spatial location attribute \( y \) is calculated using all the hash functions in \( BF \), which can be expressed as

\[
BF_i(y) = h(x)^{1+e^{x+y}}. \tag{4}
\]

Equation (5) gives the mathematical way to calculate CQT, but the time complexity of direct calculation can be very high. The fast FFT is used in the computation of the discrete Fourier transform, and the fast FFT can be used to reduce the computational effort in the computation of the CQT, which includes both time and frequency domains.

\[
X_{CQT}(k) = \sum_{n=1}^{N+2} x(n)a_k(n), (k = 1, 2, 3, \ldots k + 1), \tag{5}
\]

\[
X_{CQT}(k) = N\sum_{i=1}^{N+1} (N+1)x(i)A_k(i), (k = 0, 1, 2 \ldots k). \tag{6}
\]

In this paper’s proposed song alignment algorithm, the extracted primary features of audio and MIDI are mapped into a common subspace by a multimodal model based on deep learning. In this space, the heterogeneous data expressing the same semantics will remain maximally correlated or closest in distance; the heterogeneous data expressing different semantics will be very low correlated or distant from each other. Meanwhile, the mapping space of features is the space represented by the similarity measure function. If the Euclidean distance is used to measure sequence similarity in the trained loss function, after training, the features are mapped to the Euclidean space. The main similarity measures commonly used are Euclidean distance, cosine similarity, Minkowski distance, Hamming distance, and Jaccard similarity.

The Delta and Theta subbands were acquired by the above method. The subbands of 5 EEG signals were extracted separately: Delta band (0.5-4 Hz), Theta band (4-8 Hz), the Alpha band (8-14 Hz), the Beta band (13-30 Hz), and Gamma band (30-45 Hz), as shown in Figure 2. The Delta band mainly reflects the EEG wave motion of children in a sleep state, which is generated in the back of the child’s brain. Theta band mainly expresses the EEG waveforms of teenagers between 10 and 17 years old, while the EEG waveforms of adults reach this band in small quantity and low amplitude. The Gamma band mainly reflects the connection between mental activity and perceptual activities such as attention; the Beta band describes the normal human brain in a state of arousal, which is not disturbed by electrooculography and is suitable for emotional analysis. In this study, the Alpha frequency band (8-14 Hz) and Beta (13-30 Hz) frequency band were selected as the EEG data for emotion recognition. It is determined that the goal of GAN is to be a high-level fake generator using a high-performance discriminator [19].

4. Experimental Design of Piano Performance in the Environment of Internet of Things Multisensor Information Fusion

ISAPI is a set of API interfaces for Internet services provided by Microsoft. It can implement all the functions provided by CGI and extend it on this basis, such as providing a filter application program interface. The working principle of ISAPI is basically the same as that of CGI. Both obtain the user’s input information through the interactive page, and then, hand it over to the server for background processing. But the implementation methods of the two are completely different. ISAPI uses DLLs (dynamic link libraries), not executable programs. The dynamic link library is loaded into the memory of the Web server and can be shared by multiple user requests without starting a separate process for each user request, thus greatly improving the performance of the
A song in MIDI format consists of a melody track and several accompaniment tracks. A rich song is usually accompanied by multiple instruments, and to satisfy the need for uniformity of pace among multiple instruments in the music, multitrack music requires the coordination of multiple instruments to maintain harmony. 12 tracks exist in MIDI format music, and the presence of the chord F in five of the tracks at the 10th second of the sequence indicates that these tracks all remain on the F chord, i.e., there is a harmonic match between the tracks. Harmony is a measure of the tonal quality of a computer reconstruction of music and is an important indicator for analyzing music.

The spectral center of mass is introduced to characterize the music sequence in the frequency domain, and the center of mass of the spectral envelope of the music signal is transformed from the time domain to the frequency domain by the fast Fourier transform (FFT). The spectral center of mass can be used to measure the proportion of high-frequency components and low-frequency components contained in the music. When the center of mass is low, it means that the music is more distributed in the low frequency, and the content of the music is low and gloomy; when the center of mass is high, it means that the music contains more high-frequency content, and it presents a bright and soothing music style. The formula for calculating the center of mass of the spectrum is as follows.

\[ p_i = \sum_{k=0}^{N-1} k|S_i(k) + 2, \]

where \( N \) is the frame length and \( k \) denotes the frequency subscript; \( S_i(k) \) represents the amplitude value of the frequency domain signal at \( k \) in the \( i \)-th frame signal position.

The stacked self-encoder is trained before it is used to extract potential features of music. The training process is divided into two steps: firstly, pretraining using the layer-by-layer greedy training method and secondly, the network fine-tuning process. The layer-by-layer greedy training method is an unsupervised training process that trains the self-encoder network with only one hidden layer at a time. The first hidden layer in the overall network is trained first, and then, this layer is used as the input layer for the next training, and the next hidden layer is trained. All layers are trained in turn according to the above process, and the goal is to keep the output as consistent as possible with the input by pretraining to prevent underfitting and overfitting, as shown in Figure 4.

Network fine-tuning is the process of fine-tuning to obtain the optimal parameters of the overall network after the unsupervised pretraining is completed because the layer-by-layer greedy training method only optimizes the parameters between two adjacent layers. Network fine-tuning requires connecting the whole network together, using the pretraining results as the initial, input to propagate forward. The new network-wide cost function is reconstructed, and the weights and biases of each hidden layer are adjusted by iterative iterations of the gradient descent method to update the parameters of the network until the global optimal parameters of the network are obtained.

The feature-level fusion model is to first extract the corresponding features from the respective data and then fuse these features by rules. The fusion process completes the
compression of many data features so that the features have high information content. Such joint features are more expressive than single features and improve the accuracy in classification recognition [21]. In the case of physiological signals expressing emotional states, the features of a single signal are one-sided and reduce the degree of classification accuracy, and the features of different physiological signals are complementary to emotions to achieve improved accuracy in recognizing emotions. This chapter investigates the fusion of physiological signal features by using feature-level fusion, which is input to the classifier, to improve the accuracy of recognizing emotions.

5. Analysis of the Performance Results of the Melodic Multisensor Information Fusion 5G IoT Model

According to Figure 5, a multilevel analysis can be performed: first, when the models are trained using the designed loss functions, the alignment results are all better than those of the transcription-based algorithm. Although the transcription model also implements the conversion of primary features to the new space, the difference in effectiveness is that the features learned by the transcription model is limited to the MIDI space of pitch and coloration. The transcription model-based score alignment divides the entire task into two separate tasks music transcription and alignment. Although to some extent, the better the transcription, the better the alignment, the alignment is always limited by the result of music alignment. The results of music transcription in academic research are not very good, and there are always difficulties in transcribing polyphonic music with multiple instruments [22]. The distance formula used in the loss function when training the multimodal-based model can also be used as a sequence similarity measure for the next alignment step. Based on this, the two tasks of feature transformation and sequence alignment are linked together.

Since there is a correlation between the accuracy of fragment retrieval and music recognition, this section measures the retrieval effectiveness mainly in the fragment retrieval task. After selecting the optimal fragment retrieval model, the evaluation results of music recognition are finally presented. Since there is a correlation between the accuracy of fragment retrieval and music recognition, this chapter measures the effectiveness of retrieval mainly in the fragment retrieval task [23]. After selecting the optimal fragment retrieval model, the evaluation results of music recognition are finally given.

The results of the analysis can be concluded that the fearfulness of the melodies generated by the four groups of models has a significant positive correlation with the coherence and creativity of the music, as shown in Figure 6.
the accuracy after encoding and decoding process. The trained stack self-encoder has a small error between output and input, and the reconstructed notes have high accuracy. It can be analyzed that the stack self-encoder can compress the high-dimensional data efficiently through the process of encoding and decoding, and the features of the music samples can be extracted effectively in the encoding process of the stack self-encoder because the activation function is a nonlinear function to compress the data to get the features with deeper information and can get the potential features.

The data of the second-order feature layer extracted by the stack self-encoder is the potential decompression music features.

The number of iterations of the LSTM network closely affects the quality of the music reconstructed by the SAE-LSTM network. The spectrograms of the music reconstructed by the SAE-LSTM network at 300 and 500 iterations of the LSTM network are shown in Figure 6. The similarity between the music reconstruction network and the sample music spectrogram at 300 iterations is low, and the music sequences reconstructed by the network at this time are cluttered in many places; the music reconstructed at 400 iterations learns the pattern of the sample sequence, and the distribution of the generated music spectrum is similar, which indicates that the dependence between the notes before and after the sample music sequence is learned at this time.

6. Analysis of Application Experimental Results

The experimental design of the resting-state paradigm used in this paper is to select a quiet and soundproof laboratory free from external noise interference, to keep the room quiet and tidy, and to set up physiological signal employing instruments: 32 channel EEG cap, ECG signal collector, respiratory sensor, server, audio, and headphone equipment.

The spectral center of mass of the music clips corresponding to the calm mood is calculated, and a training set of 200 music clips with a similar spectral center of mass is selected with a 10% similarity tolerance. The training set was preprocessed using the method in Chapter 4, and the music sequences were fed into the trained SAE-LATM music reconstruction network by one-hot coding.
The music generated by this network is considered stress-reducing music and is used as an emotional intervention for anxious subjects.

The characteristics of the EEG signals of subject 1 in calm and stressful emotional states are presented, and the entropy of the alignment of the eight electrode channels (Fp1, Fp2, F3, F4, C3, C4, O1, and O2) in each of the $\alpha$ and $\beta$ frequency bands of the EEG signals is calculated, as shown in Figure 7. The horizontal axis represents the eight channels in which EEG was detected, and the vertical axis represents the average alignment entropy of each channel. The results show that using the alignment entropy as a
feature of emotion in these eight channels can better distinguish between calm and stressful emotions, and the alignment entropy values in the calm emotion state are overall higher than those in the stressful emotion state.

The parameters $N$, $m$, and $r$ affect the final calculated value during the calculation of sample entropy, and in the experiment $N = 60$, $m$ was chosen to be 2-dimensional, and the similarity capacity was the standard deviation of the original sequence. To show the role of sample entropy in emotion classification, sample entropy was calculated for subjects in anxiety and calm emotion states according to $r = 0.2SD$. The sample entropy of the eight channels was calculated separately. The red line represents the sample entropy value under anxiety, and the blue line represents the sample entropy value under calm emotion; the results indicate that the sample entropy clearly distinguishes between anxiety and calm emotion, as shown in Table 2.

Subjects’ anxiety was significant before the start of the experiment and was relieved after listening to stress-reducing music, and the stress-reducing effect differed according to everyone’s perception of music. The statistical results showed that the stress-reducing effect of the experimental group was significantly improved than that of the control group, and the method of stress-reducing music reconstruction based on multichannel physiological signals proposed in this paper could effectively regulate anxiety and achieve the stress-reducing effect. Since different people have different music preferences, the physiological signals under the listening music were identified to identify the music fragments that soothe people’s emotions and reconstruct the targeted stress-reducing music to effectively help relieve anxiety [24, 25]. During the experiment, it was found that music samples with higher spectral quality showed bright and high-frequency quality, and the reconstructed music had bright and soothing characteristics, which were more effective in relieving anxiety.

The generated music in MIDI format is displayed in the MidiEditor software as a piano roll with each melody selected, the green keys indicate the chord part, the chord expression in columnar chords, and the red keys indicate the melody. The generation process tends to be smooth in both the chord and melody sections. The experimental results of the music generation model based on music theory rules are shown in Figure 8. The melody part is more richly varied, and the experimental results of DCC_GAN are shown in Figure 8. The chord part and the melody part are more richly varied, but the chord keys are still connected by two bars as seen in Figure 8. The experimental results of DCG_GAN are shown in Figure 8. Both the chord part and the melody part are more richly varied, the melody is more bound by the chords, and the resulting melody is more coherent.
7. Conclusion

Aiming at the characteristics of multisource data in the network and some deficiencies in the current network security situation research, this paper proposes a “network security situation assessment and trend prediction model based on multisensor data fusion,” which makes up for the traditional use of a single-sensor data source. Insufficient security posture value has the disadvantage of being comprehensive. Based on the music generation model based on music theory rules, three modules are added: chord coding, chord prediction, and chord contextualization. To encode chords of different pitches and major/minor keys, one-hot encoding is used for chord pitches and major/minor keys to facilitate the subsequent classification of the generated melodies. Chord prediction: to improve the realism of the generated chords, the real chords are used as labels and the cross-entropy loss function is used to achieve the effect that the prediction model makes the generated chords closer and closer to the real chords. Chord contextual correlation: to make the generated chords correlated before and after and thus enhance the coherence of the generated melodies, the chords of the moment are learned by combining the real chords with the generated chords as the next round of input to achieve increased musical coherence and pleasingness. In response to the limitations of artificial features, the development history of deep learning and related technical details are next elicited. Considering the limitations of artificial features and the powerful feature representation capability of deep learning, a multimodal model based on deep learning is finally introduced.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Social Sciences Federation of Henan Province Project (A Preliminary Study on Online Teaching of Sight Singing and Ear Training in Universities under the Perspective of Artificial Intelligence, No. SKL-2020-2325).

References

[1] J. S. Brumberg, K. M. Pitt, and J. D. Burnison, “A noninvasive brain-computer interface for real-time speech synthesis: the importance of multimodal feedback,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 26, no. 4, pp. 874–881, 2018.
[2] F. Xiao, “Evidence combination based on prospect theory for multi-sensor data fusion,” ISA Transactions, vol. 106, pp. 253–261, 2020.
[3] H. Yu, H. Li, Y. Li, and Y. Li, “A novel improved full vector spectrum algorithm and its application in multi-sensor data fusion for hydraulic pumps,” Measurement, vol. 133, pp. 145–161, 2019.
[4] D. Johnson, D. Damian, and G. Tzantakis, “Detecting hand posture in piano playing using depth data,” Computer Music Journal, vol. 43, no. 1, pp. 59–78, 2019.
[5] M. Muzammal, R. Talat, A. H. Sodhro, and S. Pirbhulal, “A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks,” Information Fusion, vol. 53, pp. 155–164, 2020.
[6] A. A. Zaidan, B. B. Zaidan, M. A. Alsalem, O. S. Albahri, A. S. Albahri, and M. Y. Qahtan, “Multi-agent learning neural network and Bayesian model for real-time IoT skin detectors: a new evaluation and benchmarking methodology,” Neural Computing and Applications, vol. 32, no. 12, pp. 8315–8366, 2020.
[7] J. Wu, Y. Su, Y. Cheng, X. Shao, C. Deng, and C. Liu, “Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system,” Applied Soft Computing, vol. 68, pp. 13–23, 2018.
[8] G. L. Stavrinides and H. D. Karatza, “A hybrid approach to scheduling real-time IoT workflows in fog and cloud environments,” Multimedia Tools and Applications, vol. 78, no. 17, pp. 24639–24655, 2019.
[9] W. Liang, W. Huang, J. Long, K. Zhang, K. C. Li, and D. Zhang, “Deep reinforcement learning for resource protection and real-time detection in IoT environment,” IEEE Internet of Things Journal, vol. 7, no. 7, pp. 6392–6401, 2020.
[10] M. A. Munnaf, G. Haesaert, M. Van Meirvenne, and A. M. Mouazen, “Site-specific seeding using multi-sensor and data fusion techniques: a review,” Advances in Agronomy, vol. 161, pp. 241–323, 2020.
[11] Y. Chen and N. Zheng, “AI based research on exploration and innovation of development direction of piano performance teaching in university,” Journal of Intelligent & Fuzzy Systems, vol. 40, no. 2, pp. 3681–3687, 2021.
[12] X. Mi, T. Lv, Y. Tian, and B. Kang, “Multi-sensor data fusion based on soft likelihood functions and OWA aggregation and its application in target recognition system,” ISA Transactions, vol. 112, pp. 137–149, 2021.
[13] N. Takahashi, S. Furuya, and H. Koike, “Soft exoskeleton glove with human anatomical architecture: production of dexterous finger movements and skillful piano performance,” IEEE Transactions on Haptics, vol. 13, no. 4, pp. 679–690, 2020.
[14] M. J. Xu, A. A. Lazaro, A. A. Garsa et al., “Major prognostic factors for recurrence and survival independent of the American Joint Committee on Cancer eighth edition staging system in patients with cutaneous squamous cell carcinoma treated with multimodality therapy,” Head & Neck, vol. 40, no. 7, pp. 1406–1414, 2018.
[15] S. Qiu, H. Zhao, N. Jiang et al., “Multi-sensor information fusion based on machine learning for real applications in human activity recognition: state-of-the-art and research challenges,” Information Fusion, vol. 80, pp. 241–265, 2022.
[16] E. Campayo-Muñoz, A. Cabeled-Mas, and D. Hargreaves, “Intrapersonal skills and music performance in elementary piano students in Spanish conservatories: three case studies,” *International Journal of Music Education*, vol. 38, no. 1, pp. 93–112, 2020.

[17] Q. Gu, S. Jiang, M. Lian, and C. Li, “Health and safety situation awareness model and emergency management based on multi-sensor signal fusion,” *IEEE Access*, vol. 7, pp. 958–968, 2019.

[18] S. Oore, I. Simon, S. Dieleman, D. Eck, and K. Simonyan, “This time with feeling: learning expressive musical performance,” *Neural Computing and Applications*, vol. 32, no. 4, pp. 955–967, 2020.

[19] J. Glück, S. Bluck, and N. M. Weststrate, “More on the MORE life experience model: what we have learned (so far),” *The Journal of Value Inquiry*, vol. 53, no. 3, pp. 349–370, 2019.

[20] F. Xiao, “Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy,” *Information Fusion*, vol. 46, pp. 23–32, 2019.

[21] M. A. Gocłowska, R. I. Damian, and S. Mor, “The diversifying experience model: taking a broader conceptual view of the multiculturalism–creativity link,” *Journal of Cross-Cultural Psychology*, vol. 49, no. 2, pp. 303–322, 2018.

[22] S. Magistretti, D. Trabucchi, C. Dell’Era, and T. Buganza, “A new path toward a hybrid model: insights from PwC’s Italian experience Centre,” *Research-Technology Management*, vol. 62, no. 5, pp. 30–37, 2019.

[23] J. Hussain, A. Ul Hassan, H. S. Muhammad Bilal et al., “Model-based adaptive user interface based on context and user experience evaluation,” *Journal on Multimodal User Interfaces*, vol. 12, no. 1, pp. 1–16, 2018.

[24] L. Wang and S. Li, “Enhanced multi-sensor data fusion methodology based on multiple model estimation for integrated navigation system,” *International Journal of Control, Automation and Systems*, vol. 16, no. 1, pp. 295–305, 2018.

[25] Q. Liu, “Intelligent water quality monitoring system based on multi-sensor data fusion technology,” *International Journal of Ambient Computing and Intelligence (IJACI)*, vol. 12, no. 4, pp. 43–63, 2021.