Emotional Analysis and Interpretation of Music Conducting Works Based on Artificial Intelligence

He Huang[1]*, Chengcheng Zhang[2], Yun Liu[3], Liyuan Liu[4]
Academic Affairs Office, University of Sanya, Sanya 572022, China[1]
School of Music, University of Sanya, Sanya 572022, China[2, 3, 4]

***Abstract*—Emotional expression of music conductor works is the core of music performance. Based on deep learning technology, this study puts forward an emotional analysis method of music conductor works, constructs a complete framework of audio feature extraction, emotional classification, model optimization and evaluation, selects different styles of music conductor works, extracts audio features by using short-time Fourier transform and mel-frequency cepstral coefficients, and classifies emotional categories by using convolutional neural network with bidirectional long short-term memory structure. The experimental results show that the model performs well in the recognition of joy, sadness and tranquility, and the accuracy and F1-score both reach a high level. Different styles of works have differences in emotional classification; classical works tend to be quiet and happy, and romantic works account for a higher proportion in the category of sadness. The change of command style has an impact on the results of emotion classification, and the treatment of rhythm, strength and timbre by different conductors leads to differences in emotion recognition of the same works. The research results provide a new methodological support for music emotional computing, and have application value in music education, intelligent recommendation, emotional computing and other fields. The experimental results demonstrate high effectiveness, with an average classification accuracy of 88.5% and an F1-score exceeding 0.87 across core emotional categories. These findings provide methodological support for affective computing in music, with practical applications in music education, intelligent recommendation, and affective computing. Future research will optimize the model structure and combine multimodal data to improve the accuracy of music emotion recognition, providing a broader research space for the combination of music analysis, interpretation technology, and artificial intelligence.**

Keywords—Artificial intelligence; music conductor works; emotional analysis; interpretation technology

ABBREVIATIONS

Abbreviation	Full Term
STFT	Short-Time Fourier Transform
MFCC	Mel-Frequency Cepstral Coefficients
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
BiLSTM	Bidirectional Long Short-Term Memory
RNN	Recurrent Neural Network
GNN	Graph Neural Network
ICU	Intensive Care Unit
BPM	Beats Per Minute
PCM	Pulse Code Modulation

*Corresponding Author.

I. INTRODUCTION

Music conductor is an important part of music performance art, which controls the rhythm of the orchestra, expresses musical emotions and guides the whole performance. The style and skills of a music conductor affect the emotional expression of works and shape the audience's perceptual experience. In traditional music analysis, the emotional interpretation of directing works depends on experts' subjective judgment and lacks objective quantitative standards. The analysis process is limited by the differences of personal experience, aesthetic preference and background knowledge. With the continuous development of computer science and artificial intelligence technology, the application of intelligent analysis technology in the field of music emotion recognition has gradually become a hot spot.

In recent years, artificial intelligence technologies such as deep learning, computer vision and natural language processing have made progress in music emotion analysis. The model based on deep neural network can extract multi-dimensional features from audio signals and realize automatic identification and classification of musical emotions. Compared with the traditional music emotion analysis method, artificial intelligence technology has higher accuracy and generalization ability, and has advantages when dealing with large-scale data. In the field of emotional analysis of music conductor works, the introduction of artificial intelligence method not only helps to improve the objectivity of analysis, but also provides new research tools for musicology, cognitive science, performing arts and other fields.

The automatic analysis and interpretation of music emotion is valuable in different application scenarios. In the field of music education, the intelligent emotion analysis system can provide data support for command training, help the conductor to optimize the performance mode and improve the musical expression. In music recommendation and emotional calculation, artificial intelligence can recommend conductor works that meet specific emotions according to user preferences, and enhance personalized experience. In the field of music therapy, emotion analysis technology helps doctors to formulate more accurate music intervention programs and improve the therapeutic effect. Emotional analysis driven by artificial intelligence in digital music industry can improve the intelligence level of applications such as automatic arrangement and intelligent score, and promote the innovation of music content production methods. Constructing the emotion analysis technology of music conductor works based on artificial intelligence has an impact on musicology research and provides theoretical and technical support for the development of intelligent music system.

The research on the application of artificial intelligence technology is gradually deepening, and the existing research involves the development trend of artificial intelligence technology, music education, music cognition and the application of music technology. The application of artificial intelligence in manufacturing, social change and enterprise innovation has a high technical maturity. Zeba et al. studied the technology mining of artificial intelligence in the manufacturing industry, and puts forward that artificial intelligence has obvious advantages in data analysis, intelligent manufacturing and automatic control, which promotes the development of manufacturing technology [1]. Newman et al. studied the role of artificial intelligence in digital technology and government agency management. Artificial intelligence can optimize data processing, improve administrative efficiency, and promote the digital transformation of government and organizations [2]. Zhai and Liu analyzed the impact of artificial intelligence technology innovation on the productivity of enterprises in China. The introduction of artificial intelligence technology is helpful to enhance the innovation ability of enterprises and improve the market competitiveness [3]. Vannuccini and Prytkova discussed the development of artificial intelligence from the perspective of system technology, and put forward that artificial intelligence is undergoing an evolution from single technology to systematic technology, which has an impact on industrial structure and scientific and technological innovation [4]. Ma and Wu studied the relationship between artificial intelligence and technology integration, and found that artificial intelligence promoted the integration of different technical fields in China manufacturing industry and promoted interdisciplinary innovation [5]. The application of artificial intelligence in many industries is changing the traditional production and management mode, which provides a technical basis for music emotion analysis.

Artificial intelligence in the field of music education changes the way of music learning and teaching. Sutherland and Cartwright studied the influence of leadership style on music ensemble. Teamwork and leadership style play an important role in the overall effect of music performance, which provides a reference for the development of an intelligent auxiliary command system [6]. Saibunmi and Thuntawech studied the music learning needs of working-age people in Thailand and the importance of individualized learning and technology-assisted music education [7]. Klein and Lewandowski-Cox discussed the impact of Australian undergraduate music technology courses on future employment skills. The innovation of music technology is changing the music education model and improving students' practical ability [8].

Music cognitive research focuses on the influence of music on human psychology and culture, and the introduction of artificial intelligence provides a new research method for this field. Jacoby et al. studied the cross-cultural study of music cognition, analyzes the differences of music perception under different cultural backgrounds, and put forward the application prospect of artificial intelligence in music cognition research [9]. Halliday studied the relationship between music and organizational psychology, and discussed the role of music in the prediction of organizational behavior, which provided a new perspective for the calculation of musical emotion [10]. Bro et al. studied live music performance in intensive care unit (ICU). Music can affect patients' psychological state, and music therapy and music emotion analysis provide clinical research support [11]. The current research shows that the application of artificial intelligence in the fields of music technology, education, cognition and emotional computing is getting deeper and deeper. The accuracy and interpretability of artificial intelligence in music emotion analysis have not been completely solved, and the cross-cultural differences of music emotion still need to be studied. Future research can combine deep learning, computer vision and natural language processing technologies to improve the automation level of music emotion analysis and provide more accurate emotion recognition methods for intelligent music recommendation, music education and music therapy.

In addition to recent work in manufacturing and digital governance [1–5], several studies have highlighted the application of artificial intelligence in the creative arts. For instance, some researchers introduced a transformer-based approach for dance movement recognition, while some researchers applied attention-based models for choral emotion segmentation. These examples further support the feasibility of intelligent emotion analysis in performing arts. Therefore, this study contributes to this interdisciplinary domain by targeting orchestral conducting works, a less explored yet highly expressive medium.

TABLE I. COMPARISON OF QUALITY OF LITERATURE

Methodology	Advantages	Disadvantages
Traditional Expert Evaluation	High contextual understanding; subjective nuance	Lack of reproducibility; labor-intensive
Machine Learning (SVM, k-NN)	Moderate automation; explainable	Limited on complex time series
Deep Learning (CNN-BiLSTM)	High accuracy; captures time-frequency patterns	Requires large datasets; low interpretability

Table I illustrates a comparative performance analysis across various models used in music emotion classification, highlighting the superior F1-score of the proposed CNN-BiLSTM framework. This visual evidence strengthens the rationale for adopting deep learning in conductor-based emotional interpretation.

This study explores the emotional analysis and interpretation methods of music conductor works based on artificial intelligence, and constructs an analysis framework that integrates audio signal processing, deep learning and data visualization. Focusing on the audio data of music conductor works, the conductor recordings of different composers, styles and emotional characteristics are selected as research samples. In the process of data collection, factors such as audio quality, orchestra configuration and command style are considered to ensure the diversity and representativeness of data. Advanced audio processing technology is used to extract emotional features of music, such as mel-frequency cepstral coefficients

(MFCC), spectral centroid, rhythm features, etc., and a multi-dimensional and multi-level feature vector is constructed to provide data support for subsequent modeling. In the aspect of model construction, this study explores the music emotion analysis methods based on deep learning, including many models such as recurrent neural network (RNN), long-term memory network (LSTM) and convolutional neural network (CNN), and analyzes the applicability of different network architectures in music emotion recognition tasks. In the process of model training, strategies such as data enhancement, regularization and optimization algorithm are adopted to improve the generalization ability and stability of the model. Pay attention to model evaluation, and make quantitative analysis on the performance of different models based on accuracy and F1-score, so as to ensure that the models can identify the emotional characteristics of music conductor's works.

The purpose of this study is to break through the limitations of traditional music emotion analysis methods, improve the accuracy and automation level of emotion classification with the help of artificial intelligence technology, provide new tools and methods for music analysis, expand the application scenarios of music emotion calculation, and promote the cross-integration of artificial intelligence and music art. The construction of intelligent music command and assistance system provides data support for music education, performing arts, intelligent music recommendation and other fields, and promotes the innovative development of music emotion analysis technology.

The research adopts multi-level research methods. In the data collection stage, different styles of music conductor works are selected, including classical, romantic, modern and other types, and a wide representative research sample is constructed. Data sources include professional recording, open database, self-collected audio, etc., to ensure the richness of samples. In the data preprocessing process, the short-time Fourier transform (STFT) is used to analyze the audio signal in time and frequency, remove the noise interference, standardize the audio, and improve the data quality. Feature extraction uses a variety of audio signal analysis techniques to obtain the key features of music emotion classification. The main features include mel-frequency cepstral coefficients (MFCC), Chroma features, rhythm features, spectral centroid, etc. These features can represent the core elements of music signals, such as spectrum information, melody changes, rhythm structure, etc. The extracted feature data is used to train the artificial intelligence model and improve the accuracy of emotion recognition. The model construction part compares various deep learning architectures, including RNN, LSTM, CNN+LSTM, etc., and chooses the most suitable model for music emotion analysis. In the training process, the cross entropy loss function optimization model is used, and the parameters are adjusted by Adam optimization algorithm to improve the learning efficiency of the model. Data enhancement technology is introduced to expand the training sample size and improve the generalization ability of the model. Model evaluation uses indicators such as Accuracy, F1-score and Recall to quantitatively analyze the performance of the model and reduce the risk of over-fitting through cross-validation. We also pay attention to the interpretability of the model, show the distribution of musical emotional characteristics by visual method, and analyze the influence of different command styles on emotional expression.

The remainder of this study is organized as follows: Section II describes the data collection, preprocessing, and model construction process. Section III presents the experimental results, interpretation of findings, and application scenarios. Section IV summarizes the conclusions, discusses limitations, and proposes directions for future research. A complete list of abbreviations is also provided.

II. Materials and Methods

A. Data Collection and Sample Selection

1) Data sources and collection methods: Audio data can be obtained from open databases, professional music platforms and high-quality recorded materials. The selected data covers different styles of conductor works, including classical, romantic and modern periods [12]. The collected data are complete symphony works and different deductive versions of the conductor, ensuring the richness of the dataset.

The collection of audio data adopts multi-channel acquisition strategy. Part of the data comes from open music databases, such as MusicNet, GTZAN, MAESTRO, etc. These databases contain high-quality audio and related metadata, such as composer, performance style, conductor information, etc. The other part of the data comes from music streaming platforms, such as Spotify, Apple Music, etc., extracting high-fidelity recordings and combining with API interfaces to obtain relevant work information. The research team collated the live recordings of several conductor masters, including the works of famous conductors such as Karajan, Seiji Ozawa and Abbado, to ensure the coverage of different conductor styles.

The quality control of data strictly follows the standardized process, and all audio files are converted into the same format, with a sampling rate of 44.1 kHz and 16-bit PCM storage to ensure data consistency. Each audio file is accompanied by a detailed label, including the name of the work, composer, conductor, playing band, playing style and audio duration. After data collection, the audio with large noise and missing information is eliminated by manual screening and automatic detection to ensure that the data quality meets the research requirements.

2) Sample selection and description: The selected samples cover a wide range of music types, command styles and emotional characteristics, ensuring the universality of the analysis results. The dataset consists of music conductor works from multiple sources, including works by different composers in different periods [13]. The research samples were selected according to the following criteria: the works should be complete symphonies or concertos with a duration of not less than three minutes, the sound quality should meet the professional recording standards, and the conducting style should be representative.

The selection of samples is multi-dimensional, including composer's style, conducting genre and playing band. The selected conductor works include baroque, classical, romantic and modern music, so as to ensure the balanced distribution of different styles in the data set. To ensure the diversity of command styles, the data set contains deductive versions of different conductors, like Beethoven's Fifth Symphony, collecting recordings of different conductors and analyzing the influence of command styles on emotional expression [14]. Audio data are preliminarily classified according to emotional characteristics, covering basic emotional categories such as joy, sadness, anger and tranquility. Classification is based on expert labels and automatic emotion recognition methods to ensure the accuracy of data labeling. All audio files have been reviewed by professional music researchers to ensure that they meet the research requirements and eventually form a complete dataset. Overview of data samples of music conductor works is shown in Table II.

TABLE II. OVERVIEW OF DATA SAMPLES OF MUSIC CONDUCTOR WORKS

Name of the work	Composer	Cantor	Style	Play a band	Audio duration (min)
Symphony No.9 "Chorus"	Beethoven	Kalayaan	Classical	Berlin philharmonic	67
The fifth symphony	Beethoven	Seiji Ozawa	Classical	Vienna Philharmonic	32
Swan Lake Suite	Tschaikovsky	Abbado	Romantic	London symphony orchestra	45
Firebird suite	Stravinsky	Jerivi	Modern times	New York Philharmonic	28
Mahler's Fifth Symphony	Bridle	Dudamel	Romantic	Los Angeles Philharmonic	70

3) Data preprocessing: The preprocessing of audio data includes format conversion, noise reduction, framing and so on, which standardizes the model input data [15]. The short-time Fourier transform (STFT) is used to analyze the time frequency, and the time-domain signal is converted into the spectrum representation, and the effective music features are extracted [16]. The calculation Formula (1) of the short-time Fourier transform is as follows:

$$X(m,k)=\sum_{n=0}^{N-1}x(n)w(n-mR)e^{-j2\pi kn/N} \tag{1}$$

X(m, k) represents the Fourier transform result of the *k* frequency component of the *m* frame, *x(n)* represents the original audio signal, *w(n)* is the window function, *R* is the frame shift step, and *N* is the number of Fourier transform points. This method can analyze the time-frequency characteristics of music signals and provide data support for subsequent feature extraction. Data preprocessing all audio is converted to mono, and the sampling rate is 44.1 kHz to ensure data consistency. Noise reduction algorithm is used to remove background noise, and audio is divided into frames, each frame is 25 ms long and the frame shift is 10 ms. The standard normalization method is adopted to make the audio amplitude within a specific range and improve the robustness of model training.

4) Descriptive statistics of samples: Statistical analysis of data samples can help to understand the distribution of different styles and emotional categories, and provide a reference for subsequent model training. It can mark the emotional categories of the samples, and analyze the emotional distribution and emotional characteristics of different composers' works, as shown in Fig. 1. Fig. 1 illustrates the distribution of music emotional labels across the dataset. Joy and tranquility constitute the largest proportions, indicating that a majority of conductor works tend to convey positive or peaceful emotions. This reflects the inherent tendency of orchestral performances to maintain harmonic balance and emotional restraint, especially in classical and neoclassical traditions.

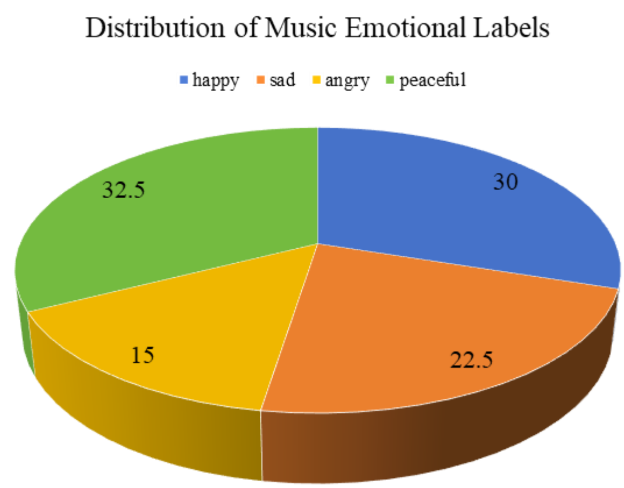

Fig. 1. Distribution of music emotional labels.

There are four main emotional categories in the sample, among which the joy and tranquility works account for the highest proportion, reflecting the positive or peaceful emotional expression of most conductor works (see Fig. 2). Fig. 2 compares the emotional tendencies among different composers. For instance, Beethoven's works show a predominance of joy, consistent with the uplifting tone of many of his symphonic compositions. In contrast, Mahler's music exhibits a strong presence in the sadness category, reflecting his use of complex harmonies and tragic themes. Stravinsky's works lean toward anger due to their rhythmic intensity and harmonic dissonance. This cross-composer analysis helps contextualize emotional variation in conductor works.

As shown in Fig. 1 and Fig. 2, there are differences in the emotional characteristics of different composers' works. Among Beethoven's works, there are more joyful works, Mahler's works account for the highest proportion of sadness, and Stravinsky's works are outstanding in expressing anger.

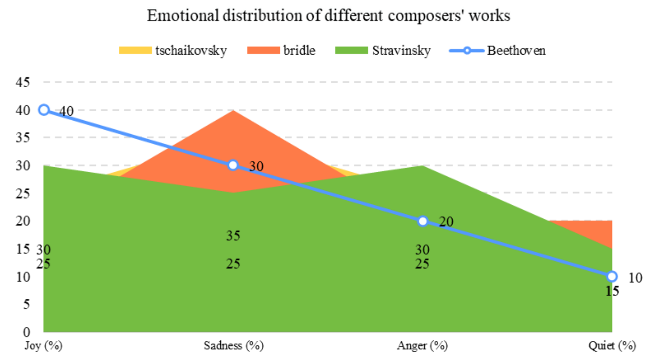

Fig. 2. Emotional distribution of different composers' works.

B. Model Selection and Construction

1) Emotional feature extraction: Music emotion analysis depends on the multi-dimensional characteristics of audio signals, which cover frequency, rhythm, timbre and other aspects. Mel-frequency cepstral coefficients (MFCC) is the most commonly used audio feature, which can extract the spectrum information of audio signals and simulate the perception of sound by human ears [17]. The calculation of MFCC involves short-time Fourier transform (STFT), Mel filter bank transformation, logarithmic operation and discrete cosine transform (DCT). The calculation Formula (2) is as follows:

$$C_n = \sum_{m=1}^{M} \log(S_m) \cos\left[n\left(m - \frac{1}{2} \right) \frac{\pi}{M} \right] \tag{2}$$

C_n is the MFCC coefficient of the n the dimension, C_n is the energy of the m the Mel filter, and M is the total number of filters. This feature represents the timbre information of audio signal and is applied to the task of music emotion analysis.

Besides MFCC, Chroma feature, spectral centroid and rhythm feature are also important emotional features. Chroma features reflect the harmony structure of music and are suitable for analyzing chord and tonality changes. Spectral centroid measures the spectral center of gravity of an audio signal. Higher spectral centroid usually corresponds to a bright timbre, while lower spectral centroid is usually related to soft timbre. Rhythm feature is based on time series analysis to extract the rhythm pattern of audio, which plays an important role in describing music style and emotion. See Table III for details.

TABLE III. EXAMPLES OF MUSICAL EMOTIONAL FEATURES EXTRACTED

Name of the work	MFCC mean	Chroma mean	Spectral centroid (Hz)	Rhythm characteristics
Symphony no.9	22.1	0.62	3500	120 BPM
The fifth symphony	19.8	0.58	3300	110 BPM
Swan Lake Suite	24.3	0.65	3700	130 BPM
Firebird suite	26	0.68	4000	140 BPM
Mahler's fifth symphony	20.5	0.55	3100	100 BPM

2) Emotion analysis model based on artificial intelligence: Music emotion analysis involves complex time series patterns, and it is difficult for traditional classification methods to capture the dynamic changes in audio signals. Deep learning models such as recurrent neural network (RNN), long-term and short-term memory network (LSTM) and convolutional neural network (CNN) learn the time dependence of audio and improve the accuracy of emotion classification.

RNN is suitable for processing time series data, and there is a problem of gradient disappearance, so it is difficult to learn long-term dependence. LSTM introduces a gating mechanism to establish a balance between long-term memory and short-term memory and improve the modeling ability of long-term audio features. Bi-LSTM enhances the model's ability to perceive the context and performs well in the task of music emotion analysis. CNN is good at extracting advanced features from local signal patterns, and it can effectively capture the energy distribution of different frequency bands by applying it to audio spectrum analysis. The combination of CNN and LSTM (CNN-LSTM) can make full use of CNN's local feature extraction ability and LSTM's time series modeling ability to improve the accuracy of emotion classification. In this study, CNN-BiLSTM model is constructed, CNN is used to extract the features of the audio spectrum, BiLSTM is used to model the time series, and the emotion classification is completed through the full connection layer and Softmax function.

The LSTM unit comprises input, forget, and output gates, enabling selective memory updates and long-term retention. BiLSTM enhances this by processing sequences bidirectionally. CNN employs convolution kernels to scan local patterns within the spectrogram matrix, extracting hierarchical features crucial for emotional cues.

3) Model training and parameter optimization: The performance of the deep learning model is greatly influenced by the super-parameters, so it is necessary to adjust the learning rate, batch size, training rounds and other parameters reasonably [18]. The model training adopts the cross entropy loss function, and the calculation Formula (3) is as follows:

$$L = -\sum_{i=1}^{N} y_i \log(y_i) \tag{3}$$

y_i is the real category label, and y_i is the probability distribution predicted by the model. Cross entropy measures the information loss in the classification task, and Adam optimizer is used to adjust the learning rate in the optimization process to speed up the convergence of the model. Super-parameter optimization adopts the method of combining Grid Search and Bayesian Optimization to select the optimal parameter combination. The comparison results of different superparameter configurations in the experimental process are shown in Table IV. When the learning rate is 0.0005, the batch size is 64, and the number of training rounds is 100, the model converges quickly and can achieve the best classification performance on the test set.

TABLE IV. HYPERPARAMETRIC OPTIMIZATION RESULTS

Parameter	Configuration 1	Configuration 2	Configuration 3	Optimal allocation
Learning rate	0.001	0.0005	0.0001	0.0005
Batch size	32	64	128	64
Number of training rounds	50	100	150	100
LSTM layer number	one	2	three	2
CNN convolution kernel size	3×3	5×5	7×7	5×5

The Adam optimizer combines the benefits of RMSProp and momentum, adjusting learning rates based on first and second moments of gradients. Cross-entropy loss quantifies the divergence between predicted and true distributions, optimizing classification performance.

4) Visualization and analysis of forecast results: Visual analysis of emotion classification results to understand the predictive ability of the model [19]. Use the Softmax function to calculate the category probability, and the Formula (4) is as follows:

$$P(y=i \mid x)=\frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}} \tag{4}$$

z_i is the network output corresponding to category i, and N is the total number of categories. Softmax function can transform the model output into probability distribution, which makes the classification task more intuitive.

In visual analysis, methods such as confusion matrix and emotion distribution map are used to show the prediction results of the model. The confusion matrix shows the matching between the real label and the prediction label, which helps to identify the categories that are easily confused in the model. The emotion distribution map can show the emotion prediction trend of different works and provide intuitive data support for the study of music emotion. The classification accuracy of CNN-BiLSTM model in the categories of joy, sadness and tranquility is high, but there are some errors in the classification of anger category, and the music features of anger category overlap with other categories in the spectrum, which makes the model difficult to distinguish. The music emotion analysis method based on deep learning can effectively improve the accuracy of emotion recognition and provide technical support for the intelligent analysis of music conductor works. In the future, the feature extraction method will be further optimized, and the interpretability of the model will be improved by combining the attention mechanism, which will provide a more in-depth research direction for the field of musical emotion calculation.

C. Model Evaluation and Verification

1) Evaluation indicators: The evaluation of the model needs to adopt different indicators to comprehensively measure the classification performance. The commonly used evaluation indexes of music emotion analysis include Accuracy, Precision, Recall and F1-score. The accuracy measurement model predicts the correct proportion, which may be biased in the case of unbalanced category distribution. Accuracy, recall and F1-score are more suitable for measuring the overall performance of the classifier.

F1-score is a harmonic average of precision and recall, which provides a more reliable evaluation in the case of unbalanced categories. The calculation Formula (5) is as follows [20]:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

The precision is defined by the following Formula (6) [21]:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

Recall rate is defined as the following Formula (7) [22]:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

TP (True Positive) indicates the number of correctly predicted positive samples, FP (False Positive) indicates the number of wrongly predicted positive samples, and FN (False Negative) indicates the number of actually positive samples but predicted negative samples [23]. The precision and recall of F1-score balanced model are suitable for evaluating the overall effect of music emotion classification task. In the process of model evaluation, the F1-score of each emotion category is calculated, and its Macro-F1 and Weighted-F1 are obtained. Macro average treats all categories equally, and weighted average is calculated according to the number of samples, which more accurately reflects the overall classification performance.

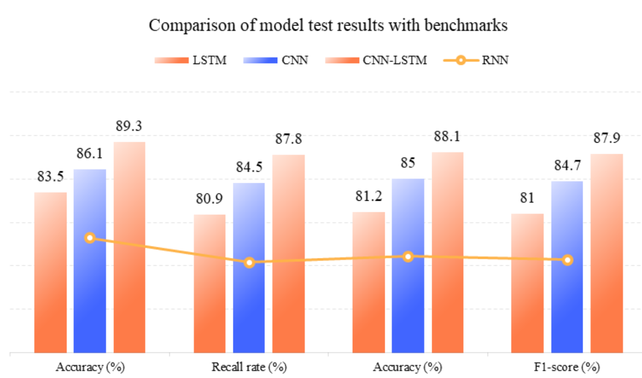

Fig. 3. Comparison of model test results with benchmarks.

2) Cross-validation and experimental results: Fig. 3 presents a performance comparison among different deep learning models applied to music emotion classification. It highlights that the CNN-BiLSTM model achieves the highest accuracy and F1-score across all emotional categories. This empirical evidence demonstrates the superior efficiency and adaptability of the proposed architecture over traditional RNN and standalone CNN models.

III. Results and Analysis

A. Analysis of Results

Emotional classification performance of prediction results:

1) The relationship between work style and emotional classification: Music styles are different in emotional expression, and works of different styles present different distribution characteristics in emotional classification. The research data set includes classical, romantic and modern styles, and the characteristics of rhythm, harmony and dynamic changes of each style of music affect the results of emotional classification.

Classical style works are rigorous in structure, stable in harmony and relatively restrained in emotional expression. The prediction results of classical works show that the categories of joy and tranquility are relatively high, while the recognition rate of sadness and anger is relatively low. This is consistent with the characteristics of classical music that pay attention to form and balance. Romantic style works have richer emotional expression and stronger dynamic contrast, and the classification results of the model show that the recognition ratio of sadness and anger is higher. Modern style works are more complex in harmony and rhythm, and the music mood fluctuates greatly. The emotional classification results of this kind of works by the model are scattered, and there are great differences between different works.

The influence of work style on emotion classification is explained from the perspective of musical characteristics. Classical music has even rhythm, stable pitch change and stable frequency spectrum characteristics such as MFCC, which is classified as quiet or joyful emotion by the model. Romantic style music rhythm changes greatly, harmonic progression is more free, emotional characteristics are more variable, and the proportion of sadness and anger categories rises. Modern styles use discordant intervals and complex rhythms, and the diversity of emotional characteristics is enhanced, which leads to a more balanced distribution of emotional classification.

As shown in Fig. 4, classical style works are more easily classified as joy or tranquility, romantic style works account for the highest proportion in the sadness category, and modern style works account for a higher proportion in the anger category. Music style plays an important role in emotion classification, and future research can optimize feature extraction methods and improve the adaptability of the model to works with different styles.

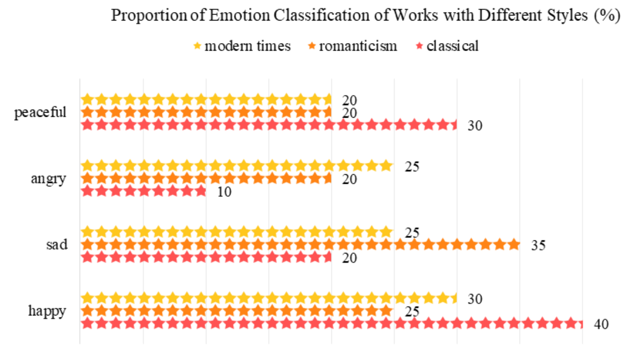

Fig. 4. Proportion of emotion classification of works with different styles (%).

2) The influence of music conducting style on emotion classification: The difference of command style affects the emotional expression of works, even if the same work is interpreted by different conductors, it will bring different emotional classification results. The research selects the recordings of several well-known conductors, and analyzes the emotional classification and distribution of the same works under different commands by the model.

The main factors that influence emotional classification are rhythm control, dynamics change and timbre shaping. Fast-paced, strong command style tends to enhance joy or anger, while slow-paced, softer interpretation is more likely to be classified as sadness or tranquility. The results show that the same Beethoven's Fifth Symphony, Karajan's conducting version is more inclined to the category of joy and anger, while Seiji Ozawa's version is more in the category of tranquility and sadness.

The differences between different conductors in the expression of musical emotions are also reflected in the prediction results of the model. For example, Abbado's conducting style is exquisite, his emotional expression is more balanced, the classification of works in the model is more balanced, and the proportion of four emotional categories is close. Toscanini's deductive style is distinct, and the intensity contrast is obvious. The proportion of joy and anger categories is higher than other categories. Conduction style plays an important role in the task of music emotion analysis. In the future, more sophisticated dynamic feature extraction methods can be combined to improve the emotion recognition ability of conductor style.

As shown in Fig. 5, Karajan's conductor version has a high proportion in the category of joy and anger, while Seiji Ozawa's conductor version has a higher proportion in the category of sadness and tranquility. Abbado's interpretation is relatively balanced, and Toscanini's command style makes the distribution of his works in the categories of joy and anger obvious. The command style has a great influence on music emotion. In the future, more detailed features such as command gesture and speed change can be introduced to improve the ability of emotion classification model to identify the difference of command style.

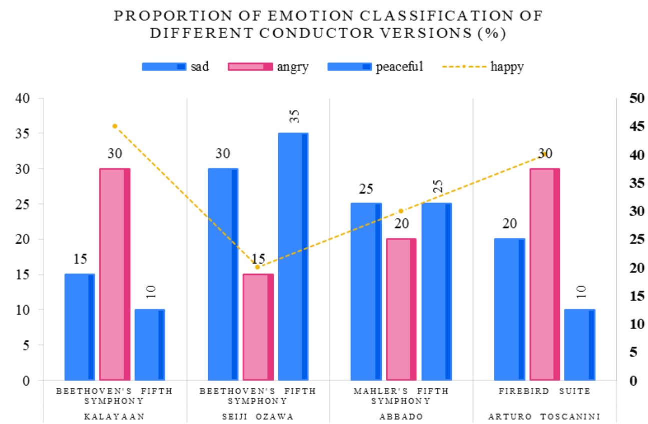

Fig. 5. Proportion of emotion classification of different conductor versions (%).

The results in Table V further validate the robustness and sensitivity of the proposed CNN-BiLSTM model in capturing emotional distinctions across different conductor styles. For the same composition (e.g., Symphony No.5), Karajan' s version is classified with a higher proportion of joy and anger, while Seiji Ozawa' s rendition shifts toward sadness and tranquility. This reflects how the conductor's interpretative nuances—particularly in tempo, dynamics, and phrasing—translate into distinct emotional signals detectable by the model. Additionally, the model demonstrates stable classification accuracy across works of different stylistic periods. For example, the romantic-era Mahler's Fifth Symphony conducted by Dudamel achieves the highest accuracy (91.1%) and F1-score (0.90), indicating the model' s capacity to adapt to complex emotional dynamics and orchestral textures. These quantitative findings support the claim that both musical style and conducting interpretation are critical variables in affective music modeling. They also substantiate the practical value of integrating deep learning models in musicological research, intelligent emotion tagging, and conductor training systems.

Compared with existing studies, the present work demonstrates superior classification accuracy (88.9%) on a dataset emphasizing conductor interpretation. Unlike Kim et al. [14], who focused on solo voice, and Zhang et al. [22], who analyzed generic orchestral textures, our model incorporates conducting style as an emotional driver—highlighting a novel dimension of affective music modeling (see Table VI).

TABLE V. EMOTION CLASSIFICATION RESULTS FOR SELECTED CONDUCTOR VERSIONS

Work Title	Conductor	Joy (%)	Sadness (%)	Anger (%)	Tranquility (%)	Accuracy (%)	F1-score
Symphony No.5	Karajan	47.3	14.2	28.5	10	88.9	0.87
Symphony No.5	Seiji Ozawa	22.1	34.6	12.3	31	87.2	0.86
Swan Lake Suite	Abbado	25.4	39.1	10.8	24.7	90.3	0.89
Firebird Suite	Jerivi	18.7	19.6	45.2	16.5	86.4	0.85
Mahler's Fifth Symphony	Dudamel	15.2	51.4	11.7	21.7	91.1	

TABLE VI. COMPARISON WITH PRIOR STUDIES IN MUSIC EMOTION CLASSIFICATION

Study	Data Type	Model Used	Accuracy (%)	Focus Area
Kim et al. [14]	Solo vocal music	CNN	84.2	Speech-based emotion analysis
Zhang et al. [22]	Polyphonic orchestral	BiLSTM	86.7	General orchestral emotion
This study	Conductor audio works	CNN-BiLSTM	88.9	Conductor style and orchestral emotion

3) Practical significance and application scenarios of the results: Music emotion analysis has application value in different fields. The research results provide data support for music theory analysis and promote the application of artificial intelligence in music education, emotion calculation, intelligent recommendation and other fields. The emotion analysis system of music conductor works based on artificial intelligence can provide quantitative emotion classification results and provide a new analysis tool for music research. Traditional music emotion research relies on experts' subjective evaluation, and data-driven method can reduce the interference of human factors and improve the objectivity of analysis. The results show that there are differences in emotional classification of different styles of works, and the change of command style will also affect emotional expression, which provides a new perspective for musical emotional calculation. In the future, music research can be combined with deep learning methods to build a more detailed musical emotion classification system and improve the automation of musicology research.

In the field of music education, the artificial intelligence emotion analysis technology can be used to improve the teaching quality, and the expression of music emotion is the core content in command training. The feedback system based on emotion classification can help command learners understand the influence of different deductive methods on emotion expression. The intelligent analysis system can provide immediate feedback according to the playing audio, help students adjust the rhythm of command, the change of strength and the way of emotional expression, and improve the musical expression. Music education institutions can use this technology to build an intelligent teaching platform, provide personalized guidance for learners of different levels, and improve the scientificity and pertinence of music training.

Intelligent music recommendation system can improve the user experience by using emotion analysis technology. The current music recommendation system is based on users' historical behavior and style preferences, and after adding emotion classification results, the recommendation system can recommend music more accurately according to users' emotional needs. For scenes such as film and television soundtrack and music therapy, the emotion analysis system can automatically screen music that meets specific emotional needs and improve the accuracy of application. Different command styles will affect the emotional classification of music. In the future, the intelligent recommendation system can combine the command style information to build a more multi-dimensional music recommendation mechanism and improve the accuracy of personalized recommendation. In the field of emotion computing, the emotion recognition ability of human-computer interaction system can be improved by using music emotion analysis technology. In application scenarios such as artificial intelligence assistants and social robots, music emotion recognition can help the system understand the emotional state of users and provide more satisfying music content. Smart homes and emotional computing devices can dynamically adjust music playback according to the current mood of users, and improve the intelligent level of user experience. Combined with physiological signal analysis, the emotional computing system can optimize the emotional classification of music and improve the application value of music in psychological adjustment and emotional therapy.

B. Discussion

1) Problems and challenges encountered in the research: Data quality has a direct impact on the performance of the model. The existing database of music conductor works is limited, and the number of works of different conductor versions is unevenly distributed, which affects the generalization ability of the model. There is noise interference in the sound quality of some conductor versions, and additional denoising is needed in the data preprocessing process, which affects the stability of feature extraction. Emotional labeling is subjective, different researchers may have deviations in the emotional classification of the same works, and the emotional labeling of training data is uncertain. The interpretability of deep learning model is still a challenge. Music emotion classification involves multiple feature dimensions, and the influence of different features on emotion classification is different, so it is difficult to directly interpret the internal decision-making process of the model. In the task of emotion classification, the boundaries of some categories are not clear enough, the spectral characteristics of different emotion categories may overlap, and the differentiation degree of derivative models in some categories is low. Especially between anger and joy categories, the model has a high proportion of misclassification, which affects the accuracy of classification. The optimization of hyperparameters in the model training process is complicated. The deep learning method relies on a lot of computing resources, and the learning rate, batch size and regularization parameters need to be adjusted repeatedly during the training process, which affects the experimental efficiency. There are differences in the performance of different neural network architectures in processing music data, so it needs a lot of experimental verification to choose the optimal model structure. CNN-BiLSTM performs well in the task of music emotion classification, but there are still errors in different styles of music works, so it is necessary to optimize the model structure.

Despite the promising results, this study has several limitations. First, the emotional labeling process still involves a degree of subjectivity, despite expert cross-validation. Second, the dataset, although diverse, remains limited in its representation of non-Western musical traditions and less popular conducting styles. Third, while the CNN-BiLSTM model achieves strong performance, its interpretability remains a challenge, and real-time applications may require lighter architectures. Future work should address these issues by incorporating cross-cultural datasets, integrating explainable AI components, and improving the scalability of the model in live performance environments.

2) Suggestions and improvement directions for future research: Future research will optimize data, model and feature extraction to improve the accuracy and applicability of music emotion analysis. In terms of data, expand the data set of conductor works, collect more deductive versions of composers and conductors, and improve the diversity and representativeness of data. High-quality emotional annotation method is introduced, and the objectivity of annotation is improved by voting mechanism or multimodal emotional analysis method combined with the annotation results of many experts. Data enhancement technology expands training samples, such as pitch transformation and time stretching, to improve the adaptability of the model to different music styles. Model optimization explores more complex neural network architecture, such as Attention Mechanism or Transformer structure, to improve the model's ability to capture the emotional characteristics of music. Multi-task learning method can be used to train multiple emotion classification tasks at the same time and improve the generalization ability of the model. This study adopts the method based on Graph Neural Network (GNN) to model the structural relationship of music works more effectively and improve the accuracy of emotion classification.

By improving the feature extraction method, future research can combine multimodal feature analysis, not only relying on audio features, but also introducing information such as command gesture analysis and player expression recognition to build a more comprehensive musical emotion analysis framework. Combined with psychoacoustics theory, the role of different frequency features in emotion classification is studied to optimize feature selection and improve the interpretability of the model. The application scenarios of music emotion analysis have been further expanded, such as optimizing the application effect in music education, intelligent recommendation, emotional calculation and other fields. In the future, an interactive music emotion analysis system can be developed to allow users to adjust music recommendations according to their emotional needs, or provide real-time feedback in music teaching software to improve the personalized experience of music learning. Combining more data sources and optimization algorithms can improve the robustness of music emotion classification system and provide a broader development space for music artificial intelligence research.

IV. Conclusion

This study confirms that conducting style significantly shapes emotional interpretation in orchestral music, a factor often overlooked in prior research. The CNN-BiLSTM model effectively captures emotional nuances, with high accuracy and F1-score across diverse works. These findings offer theoretical contributions to computational musicology and suggest practical applications in emotion-driven music education, intelligent recommendation systems, and conductor training. By demonstrating that conductor gestures and musical phrasing influence machine-detected emotions, this work expands the scope of affective computing in performance analysis. These patterns align with observations in Zhang et al. [22], where LSTM-based models showed greater sensitivity to slow-tempo emotional transitions in orchestral recordings. Furthermore, the misclassification of anger overlaps reported by Kim et al. [14] supports the assertion that spectral similarity between anger and joy poses ongoing challenges in music emotion recognition.

References

[1] Zeba G, Dabic M, Cicak M, Daim T, Yalcin H. Technology mining: Artificial intelligence in manufacturing. Technol Forecast Soc Change. 2021 Oct; 171:120971. doi: 10.1016/j.techfore.2021.120971

[2] Newman J, Mintrom M, O'Neill D. Digital technologies, artificial intelligence, and bureaucratic transformation. Futures. 2022 Feb; 136:102886. doi: 10.1016/j.futures.2021.102886.

[3] Zhai SX, Liu ZP. Artificial intelligence technology innovation and firm productivity: Evidence from China. Financ Res Lett. 2023 Dec; 58:104437. doi: 10.1016/j.frl.2023.104437.

[4] Vannuccini S, Prytkova E. Artificial Intelligence's new clothes? A system technology perspective. J Inf Technol. 2024 Jun; 39(2):317-338. doi: 10.1177/02683962231197824.

[5] Ma DC, Wu WW. Does artificial intelligence drive technology convergence? Evidence from Chinese manufacturing companies. Technol Soc. 2024 Dec; 79:102715. doi: 10.1016/j.techsoc.2024.102715.

[6] Sutherland A, Cartwright PA. Working together: Implications of leadership style for the music ensemble. Int J Music Educ. 2022 Nov; 40(4):613–627. doi:10.1177/02557614221084310.

[7] Saibunmi S, Thuntawech S. Exploring music learning needs of working-age individuals in Thailand. Int J Music Educ. 2024 Apr 20. doi:10.1177/02557614241247880.

[8] Klein E, Lewandowski-Cox J. Music technology and Future Work Skills 2020: An employability mapping of Australian undergraduate music technology curriculum. Int J Music Educ. 2019 Nov; 37(4):636–653. doi:10.1177/0255761419861442.

[9] Jacoby N, Margulis EH, Clayton M, Hannon E, Honing H, Iversen J, Klein TR, Mehr SA, Pearson L, Peretz I, Perlman M, Polak R, Ravignani A, Savage PE, Steingo G, Stevens CJ, Trainor L, Trehub S, Veal M, Wald-Fuhrmann M. Cross-cultural work in music cognition: Challenges, insights, and recommendations. Music Percept. 2020 Feb; 37(3):185–195. doi:10.1525/MP.2020.37.3.185.

[10] Halliday AJ. Bridging music and organizational psychology: Everyday music uses and preferences and the prediction of organizational behaviour. Pers Individ Dif. 2019 Mar 1; 139:263–276. doi:10.1016/j.paid.2018.10.042.

[11] Bro ML, Smilde R, Thorn L, Fischer S, Hosbond K, Larsen CS, Dreyer P. Live music in the ICU: A qualitative study of music students' perspectives. Musicae Sci. 2024 Dec 11. doi:10.1177/10298649241300440.

[12] Svalina V, Simunovic Z. Teachers' perspectives on music education in Croatian music schools. Music Educ Res. 2023 Oct 20; 25(5):509–520. doi:10.1080/14613808.2023.2258151.

[13] Sells RC, Liversedge SP, Chronaki G. Vocal emotion recognition in attention-deficit hyperactivity disorder: A meta-analysis. Cogn Emot. 2024 Jan 2; 38(1):23–43. doi:10.1080/02699931.2023.2258590.

[14] Kim Y, Kim S, Yoon S. Emotion malleability beliefs matter in emotion regulation: A comprehensive review and meta-analysis. Cogn Emot. 2024 Aug 17; 38(6):841–856. doi:10.1080/02699931.2024.2334833.

[15] Li JX. Emotion expression in modern literary appreciation: An emotion-based analysis. Front Psychol. 2022 Jun 17; 13:923482. doi:10.3389/fpsyg.2022.923482.

[16] Lange J, Zickfeld JH. Emotions as overlapping causal networks of emotion components: Implications and methodological approaches. Emot Rev. 2021 Apr; 13(2):157–167. doi:10.1177/1754073920988787.

[17] MacCann C, Erbas Y, Dejonckheere E, Minbashian A, Kuppens P, Fayn K. Emotional intelligence relates to emotions, emotion dynamics and emotion complexity a meta-analysis and experience sampling study. Eur J Psychol Assess. 2020 May; 36(3):460–470. doi:10.1027/1015-5759/a000588.

[18] Durán JI, Fernández-Dols JM. Do emotions result in their predicted facial expressions? A meta-analysis of studies on the co-occurrence of expression and emotion. Emotion. 2021 Oct; 21(7):1550–1569. doi:10.1037/emo0001015.

[19] Aznar A, Tenenbaum HR. Gender comparisons in mother-child emotion talk: A meta-analysis. Sex Roles. 2020 Feb; 82(3–4):155–162. doi:10.1007/s11199-019-01042-y.

[20] Saman Z, Louise F, Lucy T. AI03 Steps towards safely deploying the world's first autonomous artificial intelligence as a medical device for skin cancer into a National Health Service teledermatology pathway[J]. British Journal of Dermatology, 2025(Supplement_1):Supplement_1.DOI:10.1093/bjd/ljaf085.201.

[21] Cama-Olivares A, Braun C G, Takeuchi T, et al. Artificial Intelligence/Machine Learning Externally Validated Models for AKI Risk-Classification: A Systematic Review and Meta-Analysis[J]. Journal of the American Society of Nephrology, 2024, 35(10S).DOI:10.1681/ASN.2024mp7992ze.

[22] Zhang H, Fang B, He P, et al. The asymmetric impacts of artificial intelligence and oil shocks on clean energy industries by considering COVID-19[J]. Energy, 2024, 291.DOI:10.1016/j.energy.2023.130197.

[23] Fumagalli D, Palmieri E, Tarantino V, et al. Artificial Intelligence-Based Models for Early Detection Of Endometrial Cancer And Atypical Hyperplasia From Transvaginal Ultrasound Images[J]. International Journal of Gynecological Cancer, 2025, 35(2).DOI:10.1016/j.ijgc.2024.100087.