RESEARCH ARTICLE

The Analysis of Music Emotion and Visualization Fusing Long Short-Term Memory Networks Under the Internet of Things

YUJING CAO[1] AND JINWAN PARK [iD]2
[1]Department of Video Art Media Art Direction New Media Production, Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University, Seoul 06911, South Korea
[2]Department of Image Science and Arts, Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University, Seoul 06911, South Korea

Corresponding author: Jinwan Park (jinpark@cau.ac.kr)

ABSTRACT This paper aims to provide an automatic analysis of music emotion, enhancing users' ability to perceive and comprehend the emotional nuances conveyed in music intuitively. It first reviews the limitations of traditional music emotion analysis methods, and highlights the transformative potential of Internet of Things (IoT) technology in addressing these limitations. Subsequently, the paper employs the Long Short-Term Memory (LSTM) network to model time series data in music, integrating it with the Sequence-to-Sequence (STS) framework to construct an advanced music emotion analysis model. Finally, several machine learning algorithms are adopted to train and evaluate the system. The findings underscore the efficacy of the music emotion analysis model based on the LSTM network fused with STS, demonstrating notable success in the task of music emotion prediction. Specifically, the fusion model achieves an average absolute error of Arousal value of 0.921, a root mean square error (RMSE) of 0.534, and an R square of 0.498. The average absolute error in the Valence value is 0.902, the RMSE is 0.575, and the R square is 0.478. This paper holds significant implications for a deeper understanding of music emotion and provides guidance for music recommendation and creation.

INDEX TERMS Music emotion, Internet of Things, long short-term memory network, sequence-to-sequence model, music emotion prediction.

I. INTRODUCTION
A. RESEARCH BACKGROUND AND MOTIVATIONS

As an artistic form of emotional expression, music can evoke emotional resonance, influencing individuals' psychological states and behaviors. Therefore, delving into the understanding and analysis of music emotion holds profound significance for fields like music psychology, the entertainment industry, and personalized recommendation systems [10], [21], [34]. However, the traditional music emotion analysis methods exhibit certain limitations, characterized by pronounced subjectivity and restricted accuracy. These challenges primarily emanate from the reliance of traditional methods on manual annotation or feature extraction based on audio signals, lacking the capacity to model the

The associate editor coordinating the review of this manuscript and approving it for publication was Giacomo Fiumara [iD].

dynamic and sequential nature inherent in music [32]. Predicting music emotions is a complex and challenging task, primarily shaped by the limitations of traditional methods and contemporary challenges. Traditional approaches encounter difficulties in handling time-series data, struggling to accurately capture the rich and intricate emotional expressions embedded in music. Typically, these methods cannot effectively handle the dynamic changes and intricate emotional information present in music, resulting in shortcomings in providing a comprehensive understanding of emotions. Simultaneously, the subjectivity and diversity inherent in music emotions add complexity to achieving accurate predictions. Music, being a highly subjective and culturally dependent art form, can elicit significantly different emotional experiences among individuals listening to the same piece. Moreover, modern music often encompasses a spectrum of emotions, and the expression of emotions may evolve

over time. Hence, comprehensively capturing and predicting these nuanced emotional changes is challenging. However, with the rapid advancement of Internet of Things (IoT) technology, the wide application of various sensors and devices provides more data sources for music emotion analysis [52]. Using IoT technology enables the acquisition of multimodal data, encompassing audio, text, and images, thereby offering a more comprehensive and multi-dimensional perspective on musical information. Analyzing and visualizing music emotions with rich data is a complex and challenging task [33].

The fusion of the Long Short-Term Memory (LSTM) network has made remarkable achievements in sequence modeling, especially in natural language processing and speech recognition. LSTM network has a memory unit and gating mechanism, enabling it to effectively model long-term dependence and is suitable for processing time series data in music [22], [26]. Modern solutions tend to employ deep learning (DL) models when facing the challenges of predicting musical emotions. This trend reflects an acknowledgment of the limitations of traditional methods, as DL excels in handling complex and multi-dimensional emotional information in music. However, current DL approaches still encounter a series of challenges, especially in the field of music, where the dynamic nature of time-series data and the complexity of emotional information require more nuanced handling. For the processing of time-series data, LSTM has become a popular choice. LSTM is renowned for its effective modeling of temporal relationships, enabling better capture of emotional changes in music. Nevertheless, further innovation is needed in music emotion analysis to enhance the DL model's ability to extract and understand emotional information. Additionally, modern solutions are gradually integrating multimodal information to achieve a more comprehensive understanding of musical emotion. In this regard, the fusion of Sequence-to-Sequence (STS) LSTM networks presented serves as an innovative approach. By integrating information from different modalities, this model can comprehensively grasp the emotions in music, improving the comprehensiveness of emotion analysis. It can overcome the limitations of traditional methods, enhancing the accuracy and comprehensiveness of music emotion analysis. Therefore, combining the IoT technology with the integrated LSTM network can effectively capture the emotional changes and contextual information in music, and improve the accuracy and effect of music emotional analysis and visualization.

This paper aims to explore a new framework for music emotion analysis and visualization by integrating the IoT and the LSTM network. It is hoped that the multimodal data collected by the IoT technology, combined with the sequence modeling ability of the LSTM network, can understand the emotional connotation of music more comprehensively. Moreover, it can provide more accurate and detailed emotional analysis results, and enable users to perceive

and understand the emotional expression of music more intuitively.

This paper has made several significant contributions to addressing issues in the field of music emotion prediction. First, by comprehensively examining the limitations of traditional methods for music emotion analysis, it clearly identifies challenges in handling time-series data and extracting emotional information. Then, this paper innovatively introduces IoT technology into the field of music emotion analysis, offering a fresh perspective on addressing these challenges. Methodologically, it proposes a music emotion analysis model that combines STS with LSTM networks, further emphasizing the importance of integrating multimodal information. This model considers time-series data and integrates information from different modalities, enhancing its comprehensive understanding of musical emotion. Ultimately, system training and evaluation validate the superiority of the proposed LSTM-based music emotion analysis model in emotion prediction tasks. Experimental results demonstrate the model's outstanding performance in predicting Arousal and Valence values, providing a novel approach for research and applications in the field of music emotion analysis. The research findings of this paper hold significant theoretical and practical value for gaining a deeper understanding of musical emotions, offering music recommendations, and facilitating music composition.

This paper is organized into six sections. Section I serves as an introduction, outlining the research objectives, addressing issues in predicting music emotion, discussing the shortcomings of existing models, and explaining the research motivation, contributions, and research framework. Section II provides the background, detailing the theoretical content of the LSTM model and STS model to offer comprehensive background knowledge. Section III presents a literature review, explaining and discussing relevant past research to provide motivation for the current study. Section IV outlines the research methods, including the establishment process of the STS-based LSTM music emotion analysis model. Section V covers experimental design and performance evaluation, encompassing the dataset, experimental environment, parameters, results, and discussions. Section VI concludes the study, highlighting research contributions, limitations, and future prospects.

B. RESEARCH OBJECTIVES
This paper aims to realize the automatic analysis and visualization of music emotion by using the integrated LSTM network supported by the IoT. The specific objectives are as follows:

1. Develop an end-to-end music emotion analysis system, which can automatically identify the emotional content in music, including happiness, sadness, and excitement.

2. Multi-modal data, including audio, text and images, are fused to improve the comprehensive effect of music emotion analysis and visualization.

3. Experiments are conducted to evaluate the system's performance, and the differences between the fusion LSTM network and the traditional method in music emotion analysis and visualization are compared.

II. BACKGROUND

LSTM is a Recurrent Neural Network (RNN) variant designed to process and predict sequence data [1], [38]. Its core innovation lies in the introduction of a memory cell and three gates (input gate, output gate, and forget gate) [39]. The memory cell captures and stores information over a long period, enabling LSTM to excel in handling dynamic temporal changes in music and better capture the evolution of emotions over time. Gangwani et al. presented a method using LSTM STS encoder-decoder neural network structure to predict geothermal energy production, explaining LSTM structure and providing technical insights for this paper [36]. LSTM incorporates a special cell state responsible for preserving long-term information, which can be precisely controlled, avoiding the issues of gradient vanishing or exploding. The gate mechanisms, comprising input gate, forget gate, and output gate, control the flow of information, allowing LSTM to selectively forget and remember information and effectively handle long-term dependencies in sequences. Thus, in music emotion analysis, LSTM can better capture the emotional evolution in music.

The STS model is a DL model consisting of an encoder and a decoder. It was initially designed for machine translation tasks and later successfully applied to various sequence generation tasks. The encoder receives input sequences and maps them to a fixed-length vector representation, capturing the semantic information of the input sequence. The decoder accepts the vector representation generated by the encoder and maps it to the output sequence, which is the ultimate target. In music emotion analysis, STS models can be used for the generation task of mapping music sequences to emotion labels. The model's flexibility in handling sequence data makes it possible to comprehensively understand musical emotion. Through training, STS models can learn complex mapping relationships between input and output sequences, automating music emotion analysis.

A model that integrates LSTM and STS is employed in music emotion analysis here. LSTM processes the input sequence (music features), while the STS model contributes to a better understanding and interpretation of the complexity of musical emotion. By leveraging the strengths of both LSTM and STS, the LSTM network learns patterns and rules in time-series data to extract emotional information from music. The introduction of the STS model allows for a more comprehensive consideration of multimodal information, such as additional data from IoT technology. This integrated architecture aims to address the shortcomings of traditional methods in handling music emotion, enabling the model to more accurately and comprehensively predict the emotional expression of music. The integration of LSTM and STS models enhances the accuracy of music

emotion analysis, making it better suited for the complex and multi-dimensional nature of musical emotions.

III. LITERATURE REVIEW

A. RESEARCH IN THE FIELD OF MUSIC EMOTION ANALYSIS

On the research of musical emotion, Hizlisoy et al. reviewed the application of DL in musical emotion recognition. They introduced various DL models, including the convolutional neural network (CNN), RNN, attention mechanisms, and their applications in music emotion analysis. The review also summarized current research challenges and future directions [41]. Pan et al. explored the method of identifying emotion from music by using time and frequency domain features. They extracted different features and utilized machine learning algorithms for music classification and emotion recognition [23]. Abdullah et al. combined audio and lyrics information and used multimodal emotion recognition technology to identify emotions in music. They proposed a DL-based approach that combined LSTM with other models to achieve more accurate emotion recognition [42]. Salakka et al. used LSTM networks to extract audio features, and used the emotional participation model to evaluate the emotional impact of background music on listeners [17].

B. DL ALGORITHMS

In the application of DL algorithms, Dai employed an improved CNN to infer the source and determine the destination of monetary funds. The findings revealed that the prediction accuracy of the proposed method had reached the expected level [50]. Chen and Du used adaptive momentum estimation to optimize the LSTM model. This paper found that the Adam-LSTM model presented the smallest Mean Relative Errors (MRE) in Online Public Sentiment (OPS) [28]. Chen et al. used the particle swarm optimization (PSO) algorithm to optimize the DL neural network, which enabled the company to make decisions according to market changes. Through experimental analysis, the trend parameter of the optimal normal distribution of the algorithm was determined to be 0.4433 [27].

C. THE APPLICATION OF IOT TECHNOLOGY

About the application of IoT technology, Meng et al. integrated big data to precisely assess the credit assets of enterprises under the influence of social stability risks. They proposed that the prediction accuracy of the integrated algorithm model could reach 83.5%, with a small standard deviation in data prediction, highlighting the model's high stability [47]. Lei et al. explored the relationship between the quality of accounting information and the innovation investment efficiency of enterprises. They found that the quality of accounting information could improve the innovation investment efficiency of enterprises with low information environments [57]. Wang and Dai built a centripetal force model of industrial clusters to examine the influence of

various factors on the development of the e-commerce industry [19]. Liu and Chen proposed the latent feature topic model, addressing the limitations of the traditional LDAA model in identifying topics in complex environments [54]. Feng and Chen analyzed the factors affecting the service quality of cross-border import e-commerce, identifying responsiveness as the most critical factor through Artificial Neural Network analysis [56]. Zhang et al. analyzed the early warning indicators of enterprises through the DL algorithm, and found that the procurement management process caused certain risks to the financial management level of enterprises [16]. Feng et al. explored the internal relationship between information sharing and investment performance in the venture capital network community. They found that information sharing in the venture capital network community was positively correlated with investment performance [4]. Ye and Chen introduced information processing theory into the study of team improvisation and contributed to effective team improvisation [46].

D. SUMMARY

The above literature highlights significant progress in the research domains of music emotion analysis, DL algorithm, and IoT technology. However, certain shortcomings persist in the current research. First, while researchers have experimented with various DL models for music emotion recognition, challenges remain, including issues with the efficacy of data preprocessing and feature extraction and concerns about the accuracy and consistency of emotion tags. Second, regarding the application of the DL algorithm, researchers have proposed some improved models and optimization methods. However, problems persist in terms of the training efficiency, the model's generalization ability, and the parameter tuning's complexity. In addition, regarding the application of IoT technology, researchers have used big data and integrated algorithms to realize the evaluation and prediction of enterprise credit assets and industrial development. Nevertheless, there are some problems, such as data security, privacy protection, interpretability, and operability of the algorithm. To sum up, this paper deeply explores the application of IoT technology in music emotion analysis and visualization, and further improves the efficiency and robustness of the LSTM algorithm. It explores models and methods more suitable for music emotion analysis to improve the accuracy and effect of music emotion understanding and presentation.

IV. RESEARCH METHODOLOGY

A. APPLICATION OF INTERNET OF THINGS TECHNOLOGY IN MUSIC EMOTION ANALYSIS AND VISUALIZATION

IoT is a technology that connects various physical devices, sensors, and objects to the Internet through a network connection and communication technology [8], [31]. Figure 1 displays the application process of IoT technology in music emotion analysis.

The concrete application of IoT technology in music emotion analysis and visualization in Figure 1 can improve the accuracy and effect of analysis and presentation by collecting, processing, and utilizing multimodal data [37], [43]. Figure 2 illustrates the detailed implementation of IoT technology in music emotion analysis and visualization.

The IoT technology in Figure 2 collects multimodal music-related data through various sensors and devices, encompassing audio, text, and images. The IoT technology fuses the collected multimodal data and extracts the characteristics of music emotion. More comprehensive and multi-dimensional musical emotional information can be obtained by converting audio signals into spectrograms or extracting audio features, combining the emotional content of lyrics and the emotional expression of images. An emotion classification model is constructed using the data collected by the IoT technology and the characteristics of fusion, and applying the DL algorithm. The model can automatically identify the emotional content in music, such as happiness, sadness, and excitement, and analyze the emotion of music. Through the classification and analysis of music emotion, people can better understand the influence of music on people's mood and psychological state [11], [35], [44], [45].

B. EDGE AUDIO PROCESSING IN MUSIC EMOTION ANALYSIS

With the rapid evolution of IoT technology, the application of edge computing in various fields has gradually emerged [13]. As a part of edge computing, edge audio processing provides a new perspective and method for music emotion analysis. In traditional music emotion analysis, audio data typically undergo transmission to a central server over the network for analysis and processing. However, this method introduces a certain delay, impacting users' immediate perception of music emotion. Edge audio processing realizes faster and real-time data analysis by transferring audio data processing and analysis tasks from the central server to the edge devices [18], [20]. In the context of music emotion analysis, this implies that users can experience the emotion conveyed by music more instantaneously, and have a richer experience for music appreciation. By using edge computing resources, audio data can be processed on edge devices in real-time, without transmitting a large number of audio data to the central server for analysis, thus reducing transmission delays and improving the data analysis efficiency. However, edge audio processing faces challenges, including limited computing resources on edge devices that may struggle with complex audio analysis tasks. Consequently, algorithm design optimization is necessary to ensure efficient audio emotion analysis on edge devices. In addition, energy consumption is a critical concern. Edge devices usually have limited battery life, necessitating effective energy use during task handling.

In the practical application of music emotion analysis, edge audio processing holds considerable promise. For instance, the intelligent music player can analyze the music emotion
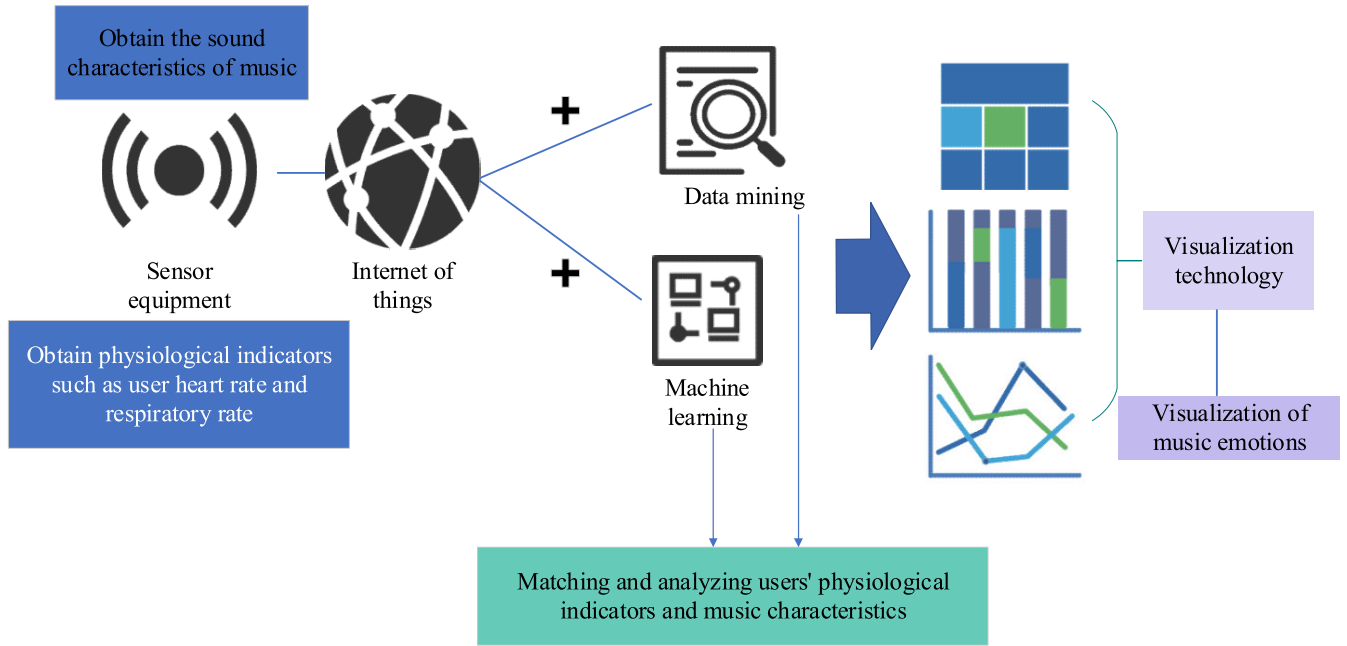
FIGURE 1. Application process of IoT technology in music emotion analysis.
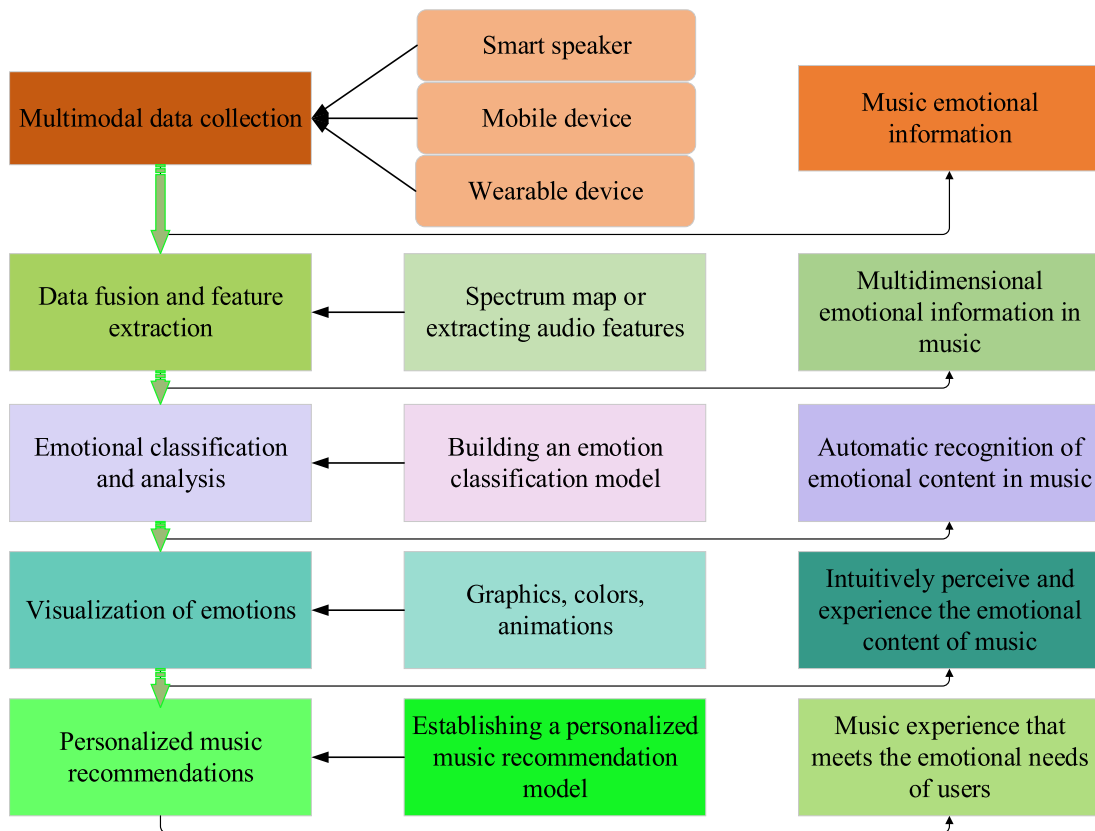

FIGURE 2. Application of IoT technology in music emotion analysis and visualization.

in real-time during the user's use, and automatically select suitable music tracks based on the user's emotional preferences. Personalized music recommendation systems can leverage edge audio processing technology to recommend the most suitable music for users according to their real-time emotional state. These application scenarios can realize more
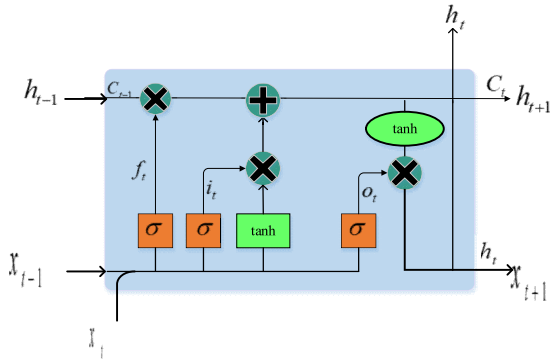
FIGURE 3. The algorithm principle and process of LSTM.

TABLE 1. Key elements of architecture design of integrated LSTM network.

Key elements	Network architecture characteristics
Stacked multilayer LSTM cells	Have different numbers of hidden units and time steps
Bidirectional LSTM	Considering the context information, the outputs are combined to get a more comprehensive sequence representation.
Attention mechanism	The network can automatically learn the importance weights of different time steps.
Sequence to sequence model	The encoder uses the LSTM unit to encode the input sequence into a fixed-length context vector, and the decoder uses the LSTM unit to generate the target sequence according to the context vector.
Parameter sharing and pre-training	The network is more robust and improves the performance and convergence speed of the network.

intelligent and real-time music emotion analysis experience through edge audio processing technology. The fusion of edge audio processing and IoT technology introduces new possibilities for music emotion analysis. In the IoT environment, diverse devices can interconnect and share audio data, enabling a more comprehensive and holistic analysis of musical emotions. Edge audio processing can accomplish audio analysis tasks among multiple devices, realize distributed music emotion analysis, and provide more accurate and personalized support for music recommendation, creation, and other fields.

In conclusion, edge audio processing technology holds significant promise in music emotion analysis. The task of audio analysis is migrated from the central server to the peripheral devices, which can realize more real-time and efficient music emotion analysis, enrich the music appreciation experience, and provide more in-depth guidance and support for music recommendation and creation. In the IoT era, marginal audio processing has injected new vitality into music emotion analysis, paving the way for fresh possibilities in the development of the music industry.

C. ARCHITECTURE DESIGN OF LONG SHORT-TERM MEMORY NETWORK

Figure 3 presents the algorithm principle and process of LSTM.

The basic unit of the LSTM network in Figure 3 is a memory unit with self-circulation mechanism.

Through a *sigmoid* function, the forget gate generates a forget gate output $f(t)$ [24] between 0 and 1 according to the hidden state $h_{(t-1)}$ of the last time step and the input $x_{(t)}$ of the current time step. Equation 1 shows the calculation of the forget gate:

$$f(t) = sigmoid(W_f * [h_{(t-1)}, x_{(t)}] + b_f \qquad (1)$$

The input gate consists of *sigmoid* function and *tanh* function. *sigmoid* function determines the degree of cell state to be updated according to $h_{(t-1)}$ and $x_{(t)}$ [25]. *tanh* generates a new candidate value $g_{(t)}$ based on the same input. The

calculation of the output gate reads:

$$i(t) = sigmoid(W_i * [h_{(t-1)}, x_{(t)}] + b_i \qquad (2)$$

$$g_{(t)} = tanh(W_c * [h_{(t-1)}, x_{(t)}] + b_c \qquad (3)$$

The cell state $C_{(t)}$ is updated according to the cell states $C_{(t-1)}$, $f(t)$ and $i(t)$ in the previous time step [29]. The calculation of cell state reads:

$$C_{(t)} = f(t) * C_{(t-1)} + i(t) * g_{(t)} \qquad (4)$$

The output gate multiplies the output of *sigmoid* function with the output of *tanh* function to get the final hidden state $h_{(t)}$. The calculation of the output gate reads [30]:

$$o_{(t)} = sigmoid(W_o * [h_{(t-1)}, x_{(t)}] + b_o \qquad (5)$$

$$h_{(t)} = o_{(t)} * tanhC_{(t)} \qquad (6)$$

b_f, b_i and b_o are bias vectors. W_f, W_i and W_o are weight matrices. $*$ means matrix multiplication. [] refers to vector connection operation. Equation 7 and Equation 8 show *sigmoid* function and *tanh* function:

$$\sigma_{(t)} = \frac{1}{1 + e^{(-x)}} \qquad (7)$$

$$tanh = \frac{e_x - e^{(-x)}}{e_x + e^{(-x)}} \qquad (8)$$

Table 1 displays the key elements of the architecture design of the integrated LSTM network.

In order to analyze music emotions more coherently and accurately, this paper fuses the LSTM network with the STS model to automatically identify music emotions [9], [52]. Figure 4 depicts the construction of the emotion analysis model integrating LSTM network music with STS.

In order to construct the LSTM network based on the STS model, the initial step involves mapping the music emotion tag sequence into a continuous vector representation through the Embedding Layer. The output of the embedded layer is input into the stacked multi-layer LSTM unit, and the input sequence is processed sequentially and its context information is extracted. Ultimately, the hidden state obtained from
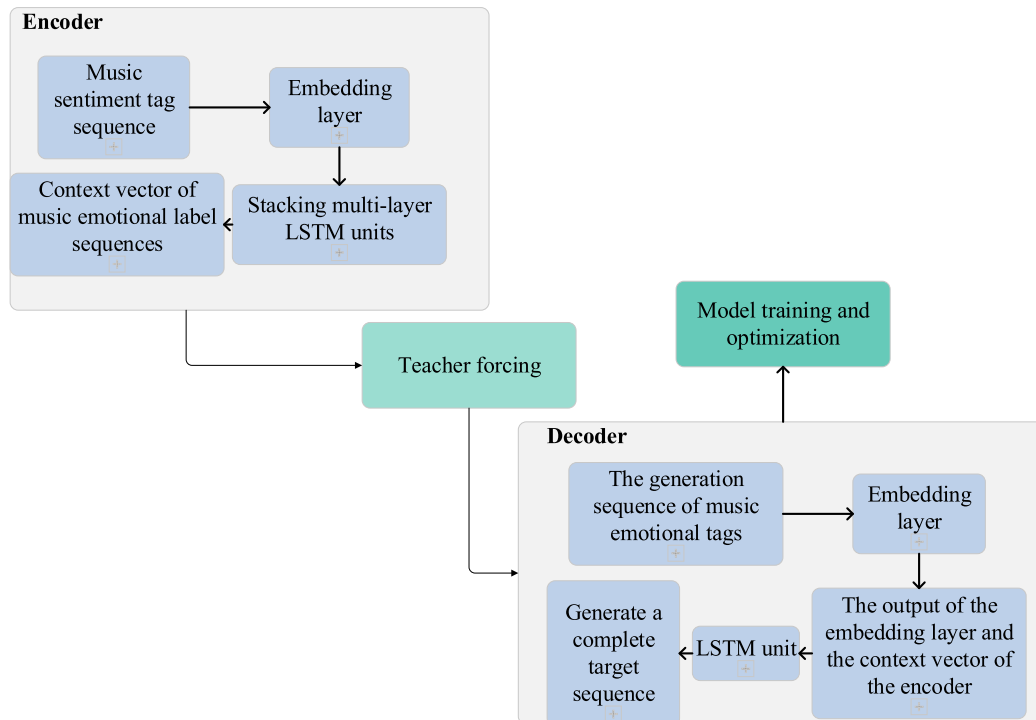
FIGURE 4. Emotional analysis model of LSTM network music based on STS.

the LSTM unit serves as the encoder's output. In the decoder, the generated sequence of music emotion tags is mapped into a continuous vector representation through the embedding layer. The embedding layer's output and the encoder's context vector are input into the LSTM unit. The LSTM unit is responsible for the stepwise generation of the target sequence, producing a label or symbol at each time step, which then serves as the input for the subsequent time step. At each time step, the hidden state and output of the LSTM unit are employed to generate the label for the next time step until the complete target sequence is generated. In the training process, techniques such as Teacher Forcing are adopted to provide the input of the decoder to help the model learn more accurate sequence generation [6], [51].

In addition, the music emotion analysis model combined with the STS LSTM network has a wide application prospect in practical application. This model gives users a more intuitive way to perceive and understand musical emotions by automatically recognizing emotional expressions in music. Such technology holds significant potential across various domains. First, it can enhance music recommendation systems. By analyzing users' emotional preferences for different types of music, the system can more accurately recommend musical works that align with users' emotional needs, thereby enhancing their overall musical experience. Second, the model can be used in the process of music creation and production. Given that music creation is often a vital avenue for emotional expression, the model's emotional analysis capability aids musicians in better understanding and capturing emotional elements, facilitating the creation of more resonant musical works. In addition, the music emotion

analysis model combined with the STS LSTM network can also contribute to the field of music education. Educators can tailor music teaching materials and methods to students based on their emotional states, thereby enhancing teaching effectiveness and the overall learning experience.

In summary, this model transcends its initial application in music and exhibits substantial potential across various domains. Its presence has the potential to significantly enrich people's lives and work, playing a pivotal role in the practical application of sentiment analysis technology. Table 2 presents the algorithm pseudocode for the STS LSTM network-based music emotion analysis model.

D. METHODS AND TECHNIQUES OF MUSIC EMOTION VISUALIZATION

Visualization technology is a method to present data or information through visual elements such as graphics, colors, and animations [2], [48]. In music emotion analysis, this paper uses timeline visualization technology to show the emotional state of music to users or audiences intuitively [7], [40]. Table 3 shows the application of different visualization techniques in music emotion analysis.

In addition, visualization serves several crucial functions in music emotion analysis: first, the intuitive presentation of emotional expression. Visualization technology employs visual elements such as color and time axes to intuitively convey the emotional state within music to users. This enables a quicker and deeper understanding of the emotional highs, lows, and changes in the music. Second, the display of emotional time series changes. Through time axis visualization technology, changes in music emotion can be clearly depicted

TABLE 2. The algorithm pseudocode for the STS LSTM network-based music emotion analysis model.

```
import numpy as np

import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense
model = Sequential()

model.add(LSTM(units=256,
input_shape=(sequence_length, feature_dim),
return_sequences=True))
model.add(LSTM(units=128, return_sequences=True))
model.add(tf.keras.layers.RepeatVector(sequence_length))
model.add(LSTM(units=128, return_sequences=True))
model.add(LSTM(units=256, return_sequences=True))
model.add(Dense(units=output_dim, activation='linear'))
model.compile(optimizer='adam', loss='mse')
model.fit(train_data, train_labels, epochs=num_epochs,
batch_size=batch_size, validation_data=(val_data,
val_labels))
predictions = model.predict(test_data)
evaluation = model.evaluate(test_data, test_labels)
```

TABLE 3. Application of different visualization techniques in music emotion analysis.

Visualization method	Principle
Color mapping	Different emotional intensities and changes can be expressed more accurately by changing the brightness, saturation, and transparency of colors.
Timeline visualization	Visualize the emotional state of music along the time axis.
Two-dimensional or three-dimensional space visualization	According to the different characteristics of music, the emotional state is mapped to different dimensions on the coordinate axis, thus forming an interactive emotional space.
Dynamic visualization	The change and intensity of emotion can be expressed by changing the shape, size, and movement path of graphics.

over time. This allows users to trace emotional shifts in different parts of the music, facilitating a comprehensive understanding of its emotional expression. Third, interactive emotional space. Visualization technology can also map musical emotions to coordinate axes of different dimensions to build an interactive emotional space. This empowers users to freely explore the relationship and changes of musical emotions in different dimensions.

In a word, the utilization of methods and techniques in music emotion visualization plays a pivotal role. It enables users to better understand music's emotional expression and changes through intuitive, interactive, and dynamic ways, and provides a powerful auxiliary tool for music recommendation and creation.

V. EXPERIMENTAL DESIGN AND PERFORMANCE EVALUATION

A. DATASETS COLLECTION

This paper uses the Million Song Dataset to train and evaluate the music emotion analysis. This database collects many music tracks, aiming to provide rich music data for music

TABLE 4. Explanation of dataset string parameters.

String parameters:	Parameter settings
dataset_dir (str)	Root directory path for the dataset files
usage (str)	None
num_samples (int)	None
num_parallel_workers (int)	8
shuffle (bool)	None
sampler (Sampler)	None
num_shards (int)	None
shard_id (int)	None
cache (DatasetCache)	None

TABLE 5. The expected sorting results obtained from different combinations of configuring the sampler and shuffle.

Parameter sampler	Parameter shuffle	Expected sorting results
None	None	Random Permutation
None	True	Random Permutation
None	False	Sequential Ordering
sampler instance	None	Order defined by the behavior of the sampler
sampler instance	True	Not allowed
sampler instance	False	Not allowed

researchers to conduct various analyses and research. It contains music tracks from different periods, styles, and artists. Each song within this dataset has rich musical features, covering rhythm, melody, harmony, and music type. In addition, there are emotional tags for each song, which indicate the emotional state expressed by each song, such as pleasure, sadness, excitement and so on. The data are stored in Hierarchical Data Format 5 (HDF5) format, totaling about 300GB. It comprises 10 genres (Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae, and Rock), each consisting of 100 tracks. These tracks are in the form of 16-bit, single-channel.wav format audio files with a sampling rate of 22050Hz [5]. The dataset can be accessed at gtzan | TensorFlow Datasets (google.cn). The generated dataset has three columns: [waveform, sample_rate, label]. The waveform column has a data type of float32, the sample_rate column has a data type of uint32, and the label column has a data type of string. Table 4 provides explanations for the string parameters in the dataset.

Table 5 illustrates the expected sorting results obtained from different combinations of configuring the sampler and shuffle.

Table 6 depicts the process and explanations of data preprocessing operations.

In order to deal with the large-scale dataset in music emotion analysis, this paper draws insights from relevant literature and employs data sampling methods to enhance analysis efficiency and ensure result quality. This strategy aligns with common practices in big data processing and holds significant implications for music emotion analysis.

TABLE 6. The process and explanations of data preprocessing operations.

Operations	Explanation
mindspore.dataset.Dataset.apply	Performing the given operation function on the dataset object
mindspore.dataset.Dataset.concat	Performing concatenation operations on multiple incoming dataset objects
mindspore.dataset.Dataset.filter	Filtering the data in the dataset object based on custom-defined criteria
mindspore.dataset.Dataset.flat_map	Applying the given data processing to each data point in the dataset object and flattening the results
mindspore.dataset.Dataset.map	Giving a set of data augmentation lists, and sequentially applying data augmentation to dataset objects
mindspore.dataset.Dataset.project	Selecting the desired columns from the dataset object and sorting them in the order of given column names
mindspore.dataset.Dataset.rename	Renaming the dataset object according to specified column names
mindspore.dataset.Dataset.repeat	Repeating this dataset count times
mindspore.dataset.Dataset.reset	Resetting the dataset object for the next epoch
mindspore.dataset.Dataset.save	Saving the currently processed data in the data processing pipeline as a common dataset format
mindspore.dataset.Dataset.shuffle	Shuffling the dataset by creating a buffer of size "buffer_size"
mindspore.dataset.Dataset.skip	Skipping the first count of data for this dataset object
mindspore.dataset.Dataset.split	Splitting the dataset into multiple non-overlapping sub-datasets
mindspore.dataset.Dataset.take	Retrieving the element with the highest count from the dataset.
mindspore.dataset.Dataset.zip	Merging and compressing multiple dataset objects along columns, ensuring that they do not have identical column names

TABLE 7. The validation results for the number of neurons in the LSTM network.

The number of neurons in LSTM	Valence (RMSE, %)	Arousal (RMSE, %)
25	48.52	59.54
50	45.32	48.63
100	42.47	42.52
150	37.81	40.15
250	35.16	35.65
256	32.39	33.12

significantly improve energy consumption in the case of uneven resource consumption.

Inspired by the above methods, this paper employs a similar idea, sampling and analyzing data by identifying the most crucial features and data blocks in music time series for emotional analysis. These methods aim to improve the quality and efficiency of music emotion analysis with limited resources.

B. EXPERIMENTAL ENVIRONMENT

Hardware configuration: Intel(R)Core (TM)i7 9750H CPU@2.60Ghz 2.59GHz, 16GB memory.

Software environment: install operating system Windows 10, graphics card RTX2080Ti (CPU), DL framework PyTorch, programming language Python, and audio dataset processing package Librosa.

C. PARAMETERS SETTING

It is uncertain which combination of structures for the STS-based LSTM music emotion analysis model can achieve the maximum regression prediction performance. Hence, this paper conducts experiments on the LSTM network structure to determine its network layers and the number of neurons. Table 7 presents the validation results for the number of hidden neurons in the LSTM network.

Table 7 reveals that as the number of LSTM neurons increases, the prediction error for Valence gradually decreases. Specifically, from 25 neurons to 256 neurons, the Valence error decreases from 48.52% to 32.39%. This indicates that increasing the number of neurons improves Valence's accuracy in music emotion analysis. Similarly, the prediction error for Arousal decreases with an increase in the number of neurons. From 25 neurons to 256 neurons, the Arousal error decreases from 59.54% to 33.12%. This suggests that increasing the number of neurons also positively impacts the accuracy of Arousal. Experimental verification indicates that the number of neurons in the LSTM network significantly influences the regression prediction performance of the music emotion analysis model. For both Valence and Arousal, the prediction error gradually decreases with an increase in the number of neurons, indicating that adopting 256 neurons can enhance the model's performance. This empirical evidence supports the selection of an appropriate network structure to achieve optimal results in music

Scaled Additional Interaction Regression (SAIL) method is based on significant differences and uses statistical methods and confidence intervals to identify significant parts of data [12]. This method can allocate significant data parts to the most efficient resources in time and cost, thus improving the quality of results under budget constraints. The Gapprox method considers the diversity of data and adopts cluster sampling to improve estimation accuracy. The block size and sample size are determined by dividing the input data into different blocks according to the data's internal and external cluster variance [14]. This method can reduce the amount of data to be processed to satisfy the acceptable confidence interval and error boundary, thus achieving the required result quality. The dynamic voltage-dynamic frequency regulation method combines data diversity and Dynamic Voltage and Frequency Scaling (DVFS) technology to manage the energy consumption of big data processing [13]. This method applies DVFS technology by estimating processing time and required frequency to reduce energy consumption. This method can

TABLE 8. Validation results for the number of layers in the LSTM network.

The number of layers in the LSTM network	Valence (RMSE, %)	Arousal (RMSE, %)
1	26.42	36.87
2	33.84	37.54
3	41.14	39.37
4	35.59	38.69

emotion analysis. Table 8 presents the validation results for the number of layers in the LSTM network.

The data in Table 8 indicate that as the number of layers in the LSTM network increases, the prediction error for Valence shows an upward trend. Specifically, the error for Valence increases from 26.42% to 35.59% as the number of layers goes from 1 to 4. This suggests that increasing the number of network layers has a negative impact on the prediction of Valence. Regarding Arousal, there is no clear trend in the impact of LSTM network layers on prediction performance. The error fluctuates between different numbers of layers, with the lowest value at 36.87% (1 layer) and the highest at 39.37% (3 layers). This indicates that increasing the number of network layers does not significantly improve the prediction performance for Arousal. Upon reanalyzing the data from the table, it is evident that increasing the number of layers in the LSTM network has a negative impact on Valence, while there is no clear trend for Arousal. Therefore, the LSTM network structure with 1 layer performs better, suggesting that choosing 1 layer achieves the optimal results for music emotion analysis.

Experimental parameter setting: pre-training LSTM layer number: 1; Number of LSTM hidden units: 256; Fusion features: 1 * 128; Learning rate: 0.001; Optimization algorithm: Adam; Batch size: 128; Training iterations: Early Stopping; Regularization intensity and discard rate: 0.2; Loss function: cross entropy loss function. Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and square coefficient are used to evaluate the results. Equations 9 to 11 show the calculation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(X_i - Y_i)^2}{N}} \qquad (9)$$

$$MAE = \frac{\sum_{i=1}^{N}|X_i - Y_i|}{N} \qquad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\widehat{e_l} - e_i)^2}{\sum_{i=1}^{N}(e_i - \overline{e_l})^2} \qquad (11)$$

N is the number of samples. X_i is the prediction label of samples. Y_i is the actual label of samples. $\widehat{e_l}$ is the predicted label of the test sample. e_i is the actual label, and $\overline{e_l}$ is the average value of the actual label of the test.

D. PERFORMANCE EVALUATION

Figure 5 displays the influence of different LSTM layers on model training. It demonstrates that the lower number of layers of LSTM leads to fluctuation of the Valence value
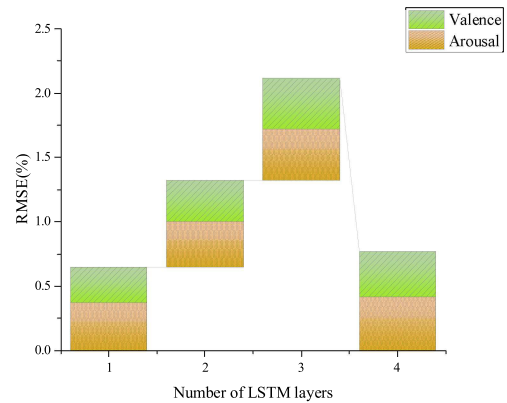
FIGURE 5. Influence of different LSTM layers on model training.
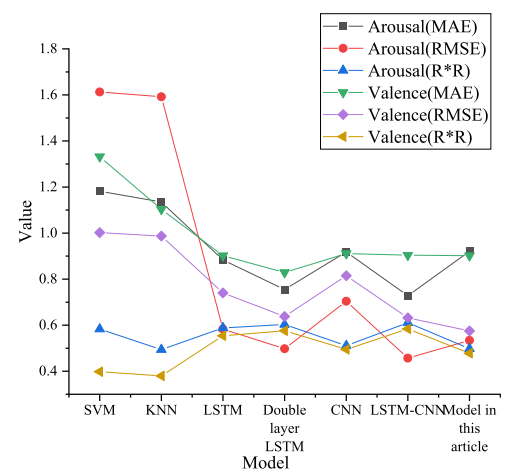
FIGURE 6. Comparison of experimental results of different DL models.

and Arousal value. The LSTM with 1 layer and LSTM with 2 layers have advantages in Valence value and Arousal value respectively. However, the RMSE range of the Arousal value is smaller than the mean square error range of the Valence value, that is |A1-A2l<lV1-V2|. Therefore, LSTM with 1 layer is selected to improve the prediction performance of both Valence and Arousal.

Figure 6 depicts the comparison of experimental results among various DL models. In Figure 6, Support Vector Machines (SVM) and K-Nearest Neighbor (KNN) based on the machine learning model exhibit inferior prediction results than DL models. In this paper, the LSTM network-based music emotion analysis model integrating STS is designed. Arousal (MAE) is 0.921. Arousal (RMSE) is 0.534. Arousal (R^2) is 0.498. Valence (MAE) is 0.902. Valence (RMSE) is 0.815. Valence(R^2) is 0.478, which makes the overall characteristics more perfect, thus further proving that combining the fusion model with IoT can effectively improve the prediction accuracy of Arousal and Valence values.

Figure 7 presents the comparative experimental results of different models on the same dataset. It suggests that compared with a single model, both the traditional machine
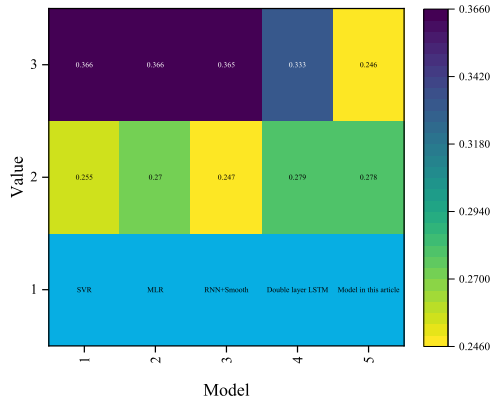
FIGURE 7. Comparative experimental results of different models.
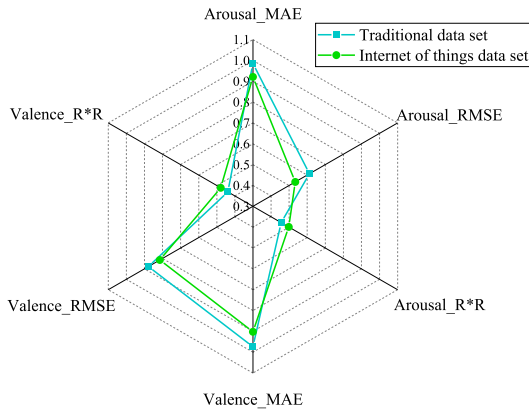

FIGURE 8. Comparison of model performance of different datasets.

learning model and the DL model can effectively improve the prediction performance of music emotions, and the DL model still has advantages. Additionally, the model's predictive performance for Arousal values surpasses that for Valence values, and this difference is significant. Despite the high complexity of the model proposed, it lacks an advantage in predicting Arousal values; however, it markedly improves the prediction performance of Valence values. Furthermore, the difference between the two is not significant, making it more versatile. This observation further underscores the effectiveness of fusing the LSTM network-based music emotion analysis model with STS in music emotion analysis and prediction.

Finally, the performance of the model using the IoT dataset and the traditional music feature data is compared. Figure 8 presents the results. It demonstrates that in the prediction of Arousal value, the MAE of the model decreases from 0.985 to 0.921, the RMSE decreases from 0.613 to 0.534, and R^2 increases from 0.456 to 0.498. In the prediction of Valence value, the MAE of the model decreases from 0.974 to 0.902, the RMSE decreases from 0.878 to 0.815, and R^2 increases from 0.439 to 0.478. This clearly shows that introducing IoT data has significantly improved the model's performance. In a word, in the task of music emotion analysis, the introduction of IoT data can effectively improve the prediction accuracy and performance of the model. This further underscores the crucial role of IoT data in the analysis of music emotions and

TABLE 9. The cross-validation results of different time-series data modeling and recognition models in the research on music emotion and visualization.

Methods	Valence (RMSE, %)	Arousal (RMSE, %)	Accuracy
Rough Autoencoder (RAE)	92.79	60.94	0.86
Interval Probability Distribution Learning (IPDL)	84.18	62.46	0.72
Deep Temporal Dictionary Learning (DTDL)	86.22	61.59	0.89
The model proposed	81.50	53.40	**0.98**

TABLE 10. The analysis results of different models' time complexity in the application of music emotion and visualization.

Model	Time complexity (O)	Model	Time complexity (O)
RAE	4.3	SVM	4.2
IPDL	4.3	KNN	4.5
DTDL	4.3	LSTM	4.8
LSTM-CNN	4.8	Double-layer LSTM	4.8
The model proposed	2.4	CNN	5.9

provides robust support for gaining a deeper understanding of musical emotional expression.

Table 9 presents the cross-validation results of different time-series data modeling and recognition models in the research on music emotion and visualization.

Table 9 suggests that, compared to other models, the proposed LSTM network model, integrated with STS, performs well in predicting Valence, with the lowest RMSE, indicating its more accurate capture of the music's pleasantness. DTDL follows closely in Valence, slightly outperforming both RAE and IPDL. In predicting Arousal, the proposed model stands out with the lowest RMSE, signifying its ability to more accurately capture the music's excitement. RAE performs poorly in Arousal, with the highest RMSE. Relatively, IPDL's performance is slightly inferior to the proposed model. DTDL's performance falls between RAE and IPDL. In terms of accuracy, the proposed model significantly outperforms other models, reaching a high accuracy of 0.98. RAE and DTDL have lower accuracies, at 0.86 and 0.89, respectively. IPDL has the lowest accuracy, at 0.72. Overall, according to the experimental data, the proposed model excels in predicting Valence and Arousal, achieving a high level of accuracy. This indicates that the LSTM network combined with STS has a significant advantage in music emotion analysis and visualization research. Table 10 presents the analysis results of different models' time complexity in the application of music emotion and visualization. It demonstrates that the proposed model has a relatively lower time complexity than others, suggesting higher efficiency in this application scenario.

E. DISCUSSION

In the research of music emotion analysis and visualization, worldwide researchers have conducted various studies. Sams and Zahr used convolutional LSTM networks to perform audio classification tasks. The results suggested that the multimodal method for music emotion recognition performed better than the single-modal method [3]. This is consistent with the research on music emotion and visualization of the fusion of LSTM networks supported by the IoT in this paper. It further emphasizes the importance of multimodal data for music emotion analysis and the effectiveness of the fusion model. Based on two-channel LSTM, Chen introduced the analytic hierarchy process (AHP) to fuse weighted features at the decision-making level, and applied it to multimodal music emotion analysis in emotion calculation. This method can effectively improve the recognition rate and save much training time [49]. These studies prove the effectiveness of multimodal data and the fusion model in music emotion analysis. Yu et al. proposed a speech emotion recognition model of attention-LSTM-attention, but the weighted accuracy of the model could not reach more than 68% in a simple dataset [55]. In contrast, this paper adopts the method of integrated LSTM networks, and combines with the IoT technology to analyze and visualize music emotions to improve the performance of emotion prediction.

VI. CONCLUSION

A. RESEARCH CONTRIBUTION

In this paper, the LSTM network-based music emotion analysis model combined with STS demonstrates effective performance in the task of music emotion prediction, and the DL model is more suitable for dealing with complex music emotion characteristics than the traditional machine learning model. These results provide a valuable reference for further research and application in musical emotion analysis.

B. FUTURE WORKS AND RESEARCH LIMITATIONS

There are also some research shortcomings. First, although various evaluation indexes are employed to gauge prediction performance, additional measures such as cross-validation or other statistical methods can be incorporated to further bolster the assessment of model effectiveness. Then, the research outcomes may be constrained by the sample dataset, prompting the need to expand the dataset for enhanced stability and reliability of experimental results. Furthermore, the proposed emotional analysis model requires practical application validation to affirm its effectiveness and feasibility. Future research should focus on refining research methodologies and enlarging the scale of experiments to advance the field of musical emotion analysis.

REFERENCES

[1] A. Iyer, S. S. Das, R. Teotia, S. Maheshwari, and R. R. Sharma, "CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings," *Multimedia Tools Appl.*, vol. 82, no. 4, pp. 4883–4896, Feb. 2023, doi: 10.1007/s11042-022-12310-7.

[2] A. S. Cowen, X. Fang, D. Sauter, and D. Keltner, "What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 4, pp. 1924–1934, Jan. 2020, doi: 10.1073/pnas.1910704117.

[3] A. S. Sams and A. Zahra, "Multimodal music emotion recognition in Indonesian songs based on CNN-LSTM, XLNet transformers," *Bull. Electr. Eng. Informat.*, vol. 12, no. 1, pp. 355–364, Feb. 2023, doi: 10.11591/eei.v12i1.4231.

[4] B. Feng, K. Sun, Z. Zhong, and M. Chen, "The internal connection analysis of information sharing and investment performance in the venture capital network community," *Int. J. Environ. Res. Public Health*, vol. 18, no. 22, p. 11943, Nov. 2021, doi: 10.3390/ijerph182211943.

[5] B. L. Sturm, "The GTZAN dataset: Its contents, its faults, their effects on evaluation, and its future use," 2013, *arXiv:1306.1461*.

[6] B. Naranjo, "Can music inspire translators? Using background music as a trigger for narrative engagement in literary translation," *Transl. Interpreting Student*, vol. 15, no. 2, pp. 280–303, Oct. 2020, doi: 10.1075/tis.18018.nar.

[7] C. M. Doyle, S. T. Lane, J. A. Brooks, R. W. Wilkins, K. M. Gates, and K. A. Lindquist, "Unsupervised classification reveals consistency and degeneracy in neural network patterns of emotion," *Social Cognit. Affect. Neurosci.*, vol. 17, no. 11, pp. 995–1006, Nov. 2022, doi: 10.1093/scan/nsac028.

[8] C. Wang and Y. C. Ko, "Emotional representation of music in multi-source data by the Internet of Things and deep learning," *J. Supercomput.*, vol. 79, no. 1, pp. 349–366, Jan. 2023, doi: 10.1007/s11227-022-04665-3.

[9] D. Han, Y. Kong, J. Han, and G. Wang, "A survey of music emotion recognition," *Frontiers Comput. Sci.*, vol. 16, no. 6, Dec. 2022, Art. no. 166335, doi: 10.1007/s11704-021-0569-4.

[10] E. A. Smit, A. J. Milne, H. S. Sarvasy, and R. T. Dean, "Emotional responses in Papua new Guinea show negligible evidence for a universal effect of major versus minor music," *PLoS ONE*, vol. 17, no. 6, Jun. 2022, Art. no. e0269597, doi: 10.1371/journal.pone.0269597.

[11] F. Abri, L. F. Gutiérrez, P. Datta, D. R. W. Sears, A. S. Namin, and K. S. Jones, "A comparative analysis of modeling and predicting perceived and induced emotions in sonification," *Electronics*, vol. 10, no. 20, p. 2519, Oct. 2021, doi: 10.3390/electronics10202519.

[12] H. Ahmadvand and M. Goudarzi, "SAIR: Significance-aware approach to improve QoR of big data processing in case of budget constraint," *J. Supercomput.*, vol. 75, no. 9, pp. 5760–5781, Sep. 2019, doi: 10.1007/s11227-019-02797-7.

[13] H. Ahmadvand, F. Foroutan, and M. Fathy, "DV-DVFS: Merging data variety and DVFS technique to manage the energy consumption of big data processing," *J. Big Data*, vol. 8, no. 1, pp. 1–16, Dec. 2021, doi: 10.1186/s40537-021-00437-7.

[14] H. Ahmadvand, M. Goudarzi, and F. Foroutan, "Gapprox: Using gallup approach for approximation in big data processing," *J. Big Data*, vol. 6, no. 1, pp. 1–24, Dec. 2019, doi: 10.1186/s40537-019-0185-4.

[15] H. Ahmadvand, T. Dargahi, F. Foroutan, P. Okorie, and F. Esposito, "Big data processing at the edge with data skew aware resource allocation," in *Proc. NFV-SDN*, 2021, pp. 81–86, doi: 10.1109/NFV-SDN53031.2021.9665051.

[16] H. Zhang, L. Fan, M. Chen, and C. Qiu, "The impact of SIPOC on process reengineering and sustainability of enterprise procurement management in e-commerce environments using deep learning," *J. Organizational End User Comput.*, vol. 34, no. 8, pp. 1–17, Jun. 2022, doi: 10.4018/JOEUC.306270.

[17] I. Salakka, A. Pitkäniemi, E. Pentikäinen, K. Mikkonen, P. Saari, P. Toiviainen, and T. Särkämö, "What makes music memorable? Relationships between acoustic musical features and music-evoked emotions and memories in older adults," *PLoS ONE*, vol. 16, no. 5, May 2021, Art. no. e0251692, doi: 10.1371/journal.pone.0251692.

[18] J. Miquel, L. Latorre, and S. Chamaillé-Jammes, "Energy-efficient audio processing at the edge for biologging applications," *J. Low Power Electron. Appl.*, vol. 13, no. 2, p. 30, Apr. 2023, doi: 10.3390/jlpea13020030.

[19] J. Wang and Y. Dai, "The agglomeration mechanism of network emerging e-commerce industry based on social science," *J. Organizational End User Comput.*, vol. 34, no. 3, pp. 1–16, Nov. 2021, doi: 10.4018/JOEUC.291561.

[20] K. Kim and S.-C. Liu, "Continuous-time analog filters for audio edge intelligence: Review on circuit designs [feature]," *IEEE Circuits Syst. Mag.*, vol. 23, no. 2, pp. 29–48, 2023, doi: 10.1109/MCAS.2023.3267893.

[21] K. R. M. Sanfilippo, N. Spiro, M. Molina-Solana, and A. Lamont, "Do the shuffle: Exploring reasons for music listening through shuffled play," *PLoS ONE*, vol. 15, no. 2, Feb. 2020, Art. no. e0228457, doi: 10.1371/journal.pone.0228457.

[22] L. Jin and E. Y. Kim, "Interpretable cross-subject EEG-based emotion recognition using channel-wise features," *Sensors*, vol. 20, no. 23, p. 6719, Nov. 2020, doi: 10.3390/s20236719.

[23] L. Pan, S. Wang, Z. Yin, and A. Song, "Recognition of human inner emotion based on two-stage FCA-ReliefF feature optimization," *Inf. Technol. Control*, vol. 51, no. 1, pp. 32–47, Mar. 2022, doi: 10.5755/j01.itc.51.1.29430.

[24] L. Singh, "Deep bi-directional LSTM network with CNN features for human emotion recognition in audio-video signals," *Int. J. Swarm Intell.*, vol. 7, no. 1, p. 110, 2022, doi: 10.1504/IJSI.2022.121102.

[25] M. Algarni, F. Saeed, T. Al-Hadhrami, F. Ghabban, and M. Al-Sarem, "Deep learning-based approach for emotion recognition using electroencephalography (EEG) signals using bi-directional long short-term memory (Bi-LSTM)," *Sensors*, vol. 22, no. 8, p. 2976, Apr. 2022, doi: 10.3390/s22082976.

[26] M. Awais, M. Raza, N. Singh, K. Bashir, U. Manzoor, S. U. Islam, and J. J. P. C. Rodrigues, "LSTM-based emotion detection using physiological signals: IoT framework for healthcare and distance learning in COVID-19," *IEEE Internet Things J.*, vol. 8, no. 23, pp. 16863–16871, Dec. 2021, doi: 10.1109/JIOT.2020.3044031.

[27] M. Chen and W. Du, "Dynamic relationship network and international management of enterprise supply chain by particle swarm optimization algorithm under deep learning," *Expert Syst.*, vol. 10, Jan. 2022, Art. no. e13081, doi: 10.1111/exsy.13081.

[28] M. Chen and W. Du, "The predicting public sentiment evolution on public emergencies under deep learning and Internet of Things," *J. Supercomput.*, vol. 79, no. 6, pp. 6452–6470, Apr. 2023, doi: 10.1007/s11227-022-04900-x.

[29] M. Russo, L. Kraljević, M. Stella, and M. Sikora, "Cochleogram-based approach for detecting perceived emotions in music," *Inf. Process. Manage.*, vol. 57, no. 5, Sep. 2020, Art. no. 102270, doi: 10.1016/j.ipm.2020.102270.

[30] M. S. K. Hosseini, S. M. Firoozabadi, K. Badie, and P. Azadfallah, "Personality-based emotion recognition using EEG signals with a CNN-LSTM network," *Brain Sci.*, vol. 13, no. 6, p. 947, Jun. 2023, doi: 10.3390/brainsci13060947.

[31] M. T. Quasim, E. H. Alkhammash, M. A. Khan, and M. Hadjouni, "Emotion-based music recommendation and classification using machine learning with IoT framework," *Soft Comput.*, vol. 25, no. 18, pp. 12249–12260, Sep. 2021, doi: 10.1007/s00500-021-05898-9.

[32] N. Fuentes-Sánchez, R. Pastor, M. A. Escrig, M. Elipe-Miravet, and M. C. Pastor, "Emotion elicitation during music listening: Subjective self-reports, facial expression, and autonomic reactivity," *Psychophysiology*, vol. 58, no. 9, Sep. 2021, Art. no. e13884, doi: 10.1111/psyp.13884.

[33] N. He and S. Ferguson, "Music emotion recognition based on segment-level two-stage learning," *Int. J. Multimedia Inf. Retr.*, vol. 11, no. 3, pp. 383–394, Sep. 2022, doi: 10.1007/s13735-022-00230-z.

[34] N. Liao, "Research on intelligent interactive music information based on visualization technology," *J. Intell. Syst.*, vol. 31, no. 1, pp. 289–297, Feb. 2022, doi: 10.1515/jisys-2022-0016.

[35] P. Dey, S. K. Chaulya, and S. Kumar, "Hybrid CNN-LSTM and IoT-based coal mine hazards monitoring and prediction system," *Process Saf. Environ. Protection*, vol. 152, pp. 249–263, Aug. 2021, doi: 10.1016/j.pSep.2021.06.005.

[36] P. Gangwani, J. Soni, H. Upadhyay, and S. Joshi, "A deep learning approach for modeling of geothermal energy prediction," *Int. J. Comput. Sci. Inf. Secur.*, vol. 18, no. 1, pp. 123–135, 2020.

[37] R. De Prisco, A. Guarino, D. Malandrino, and R. Zaccagnino, "Induced emotion-based music recommendation through reinforcement learning," *Appl. Sci.*, vol. 12, no. 21, p. 11209, Nov. 2022, doi: 10.3390/app122111209.

[38] R. Du, S. Zhu, H. Ni, T. Mao, J. Li, and R. Wei, "Valence-arousal classification of emotion evoked by Chinese ancient-style music using 1D-CNN-BiLSTM model on EEG signals for college students," *Multimedia Tools Appl.*, vol. 82, no. 10, pp. 15439–15456, Apr. 2023, doi: 10.1007/s11042-022-14011-7.

[39] R. Orjesek, R. Jarina, and M. Chmulik, "End-to-end music emotion variation detection using iteratively reconstructed deep features," *Multimedia Tools Appl.*, vol. 81, no. 4, pp. 5017–5031, Feb. 2022, doi: 10.1007/s11042-021-11584-7.

[40] S. A. Hossain, M. A. Rahman, A. Chakrabarty, M. A. Rashid, A. Kuwana, and H. Kobayashi, "Emotional state classification from MUSIC-based features of multichannel EEG signals," *Bioengineering*, vol. 10, no. 1, p. 99, Jan. 2023, doi: 10.3390/bioengineering10010099.

[41] S. Hizlisoy, S. Yildirim, and Z. Tufekci, "Music emotion recognition using convolutional long short term memory deep neural networks," *Eng. Sci. Technol., Int. J.*, vol. 24, no. 3, pp. 760–767, Jun. 2021, doi: 10.1016/j.jestch.2020.10.009.

[42] S. M. S. A. Abdullah, S. Y. A. Ameen, M. A. M. Sadeeq, and S. Zeebaree, "Multimodal emotion recognition using deep learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 2, pp. 52–58, Apr. 2021, doi: 10.38094/jastt20291.

[43] S. Pal, S. Mukhopadhyay, and N. Suryadevara, "Development and progress in sensors and technologies for human emotion recognition," *Sensors*, vol. 21, no. 16, p. 5554, Aug. 2021, doi: 10.3390/s21165554.

[44] S. Roy, A. Mukherjee, and D. De, "OrangeMusic: An orange computing-inspired recommender framework in Internet of Music Things," *Internet Technol. Lett.*, vol. 5, no. 3, p. e331, May 2022, doi: 10.1002/itl2.331.

[45] S. Siddiqui, R. Nesbitt, M. Z. Shakir, A. A. Khan, A. A. Khan, K. K. Khan, and N. Ramzan, "Artificial neural network (ANN) enabled Internet of Things (IoT) architecture for music therapy," *Electronics*, vol. 9, no. 12, p. 2019, Nov. 2020, doi: 10.3390/electronics9122019.

[46] S. Ye and M. Chen, "Leveraging team expertise location awareness in improving team improvisation: A dynamic knowledge integration perspective," *Psychol. Res. Behav. Manage.*, vol. 14, pp. 2135–2146, Dec. 2021, doi: 10.2147/PRBM.S341685.

[47] T. Meng, Q. Li, Z. Dong, and F. Zhao, "Research on the risk of social stability of enterprise credit supervision mechanism based on big data," *J. Organizational End User Comput.*, vol. 34, no. 3, pp. 1–16, Oct. 2021, doi: 10.4018/JOEUC.289223.

[48] V. Chaturvedi, A. B. Kaur, V. Varshney, A. Garg, G. S. Chhabra, and M. Kumar, "Music mood and human emotion recognition based on physiological signals: A systematic review," *Multimedia Syst.*, vol. 28, no. 1, pp. 21–44, Feb. 2022, doi: 10.1007/s00530-021-00786-6.

[49] W. Chen, "A novel long short-term memory network model for multimodal music emotion analysis in affective computing," *J. Appl. Sci. Eng.*, vol. 26, no. 3, pp. 367–376, 2022, doi: 10.6180/jase.202303_26(3).0008.

[50] W. Dai, "Application of improved convolution neural network in financial forecasting," *J. Organizational End User Comput.*, vol. 34, no. 3, pp. 1–16, Oct. 2021, doi: 10.4018/JOEUC.289222.

[51] W. Lin, C. Li, and Y. Zhang, "Emotion visualization system based on physiological signals combined with the picture and scene," *Inf. Visualizat.*, vol. 21, no. 4, pp. 393–404, Oct. 2022, doi: 10.1177/14738716221109146.

[52] W. Qian, J. Tan, Y. Jiang, and Y. Tian, "Deep learning with convolutional neural networks for EEG-based music emotion decoding and visualization," *Brain-App. Commun., J. Bacomics*, vol. 1, no. 1, pp. 38–49, Dec. 2022, doi: 10.1080/27706710.2022.2075241.

[53] X. Du, C. Ma, G. Zhang, J. Li, Y.-K. Lai, G. Zhao, X. Deng, Y.-J. Liu, and H. Wang, "An efficient LSTM network for emotion recognition from multichannel EEG signals," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1528–1540, Jul. 2022, doi: 10.1109/TAFFC.2020.3013711.

[54] Y. Liu and M. Chen, "The knowledge structure and development trend in artificial intelligence based on latent feature topic model," *IEEE Trans. Eng. Manag.*, vol. 21, no. 32, pp. 102–123, Jan. 2023, doi: 10.1109/TEM.2022.3232178.

[55] Y. Yu and Y.-J. Kim, "Attention-LSTM-Attention model for speech emotion recognition and analysis of IEMOCAP database," *Electronics*, vol. 9, no. 5, p. 713, Apr. 2020, doi: 10.3390/electronics9050713.

[56] Z. Feng and M. Chen, "Platformance-based cross-border import retail e-commerce service quality evaluation using an artificial neural network analysis," *J. Global Inf. Manage.*, vol. 30, no. 11, pp. 1–17, Jun. 2022, doi: 10.4018/JGIM.306271.

[57] Z. Lei, G. Gong, T. Wang, and W. Li, "Accounting information quality, financing constraints, and company innovation investment efficiency by big data analysis," *J. Organizational End User Comput.*, vol. 34, no. 3, pp. 1–21, Dec. 2021, doi: 10.4018/JOEUC.292525.

[58] Z. Tian, D. Li, Y. Yang, F. Hou, Z. Yang, Y. Song, and Q. Gao, "A novel domain adversarial networks based on 3D-LSTM and local domain discriminator for hearing-impaired emotion recognition," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 1, pp. 363–373, Jan. 2023, doi: 10.1109/JBHI.2022.3212475.

● ● ●