

Retraction

Retracted: Waveform Feature Extraction of Intelligent Singing Skills under the Background of Internet of Things

Mobile Information Systems

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

Copyright © 2023 Mobile Information Systems. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] C. Luo, "Waveform Feature Extraction of Intelligent Singing Skills under the Background of Internet of Things," *Mobile Information Systems*, vol. 2022, Article ID 4638801, 11 pages, 2022.

Research Article

Waveform Feature Extraction of Intelligent Singing Skills under the Background of Internet of Things

Chengping Luo 

Conservatory of Music, Jiangxi Normal University, Nanchang 330000, Jiangxi, China

Correspondence should be addressed to Chengping Luo; 003772@jxnu.edu.cn

Received 30 March 2022; Accepted 13 June 2022; Published 28 June 2022

Academic Editor: Yang Gao

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

Aiming at the complex and changeable characteristics of intelligent singing skills in the context of Internet of Things, this paper proposes a feature extraction method suitable for intelligent singing skills in this context. Firstly, focusing on vocal features, the time-domain algorithm based on open-loop and closed-loop gene extraction extracts the genetic features of songs with accompaniment; then, the section and its features are extracted by using the windowed moving matching algorithm, and the segments are divided by using the similarity between adjacent segments to obtain the segment features with emotional factors. The segment features are input into the improved BP emotion recognizer for emotion recognition. Finally, the intelligent singing skills of the whole music are determined. The experimental results show that, with the increase in feature extraction time, the accuracy of the extraction results of the existing methods changes little, which is basically maintained at a low level between 15% and 30%. When the proposed method is for feature extraction of intelligent singing skill information, the accuracy shows a continuous growth trend, and with the growth of time, its accuracy is significantly higher than the existing methods, indicating that the proposed method has significant advantages in the accuracy of feature extraction. Because this waveform feature extraction method is applied to the intelligent singing skills under the background of the Internet of Things, it has the advantages of high extraction efficiency, high accuracy, and reliability.

1. Introduction

With the development of the Internet of Things, the smart singing art based on the Internet of Things is also developing. IoT is a network that allows you to connect to the Internet, exchange information, communicate, and intelligently identify any product using data sensors such as radio frequency identification, infrared sensors, global positioning systems and laser scanners, control, and management [1]. Internet-based intelligent singing ability is the basis for people to master and recreate music. How to extract the waveform quickly becomes the key skill of the Internet of Things [2]. Intelligent singing skills include two aspects: one is to extract the singer's song from the accompaniment sound of existing musical instruments, and the second is to study the singer's singing ability from the perspective of basic characteristics, complex characteristics, and general characteristics. Its main properties are resin length and strength. Complexity is

analyzed by the basic characteristics of the passage, such as interval, rhythm, and melody, while general characteristics are analyzed by the complex characteristics of music, such as emotion and style. The main characteristics of loudness, length, and intensity are common to music [3]. On this basis, it can include three aspects: the stage of discriminating the genetic (note) features of sounds, the stage of distinguishing fragment and fragment features, and the stage of unpacking the emotional features of music. Based on these three aspects [4], this paper uses different extraction methods to extract the waveform features of intelligent singing skills under the background of the Internet of Things to obtain an extraction method with high extraction efficiency and accurate and reliable extraction results and boost the development of intelligent singing skills in the context of the Internet of Things. At present, research has been done to extract waveform features. Figure 1 shows the multimodal significance waveform detection process for the sleep stage.

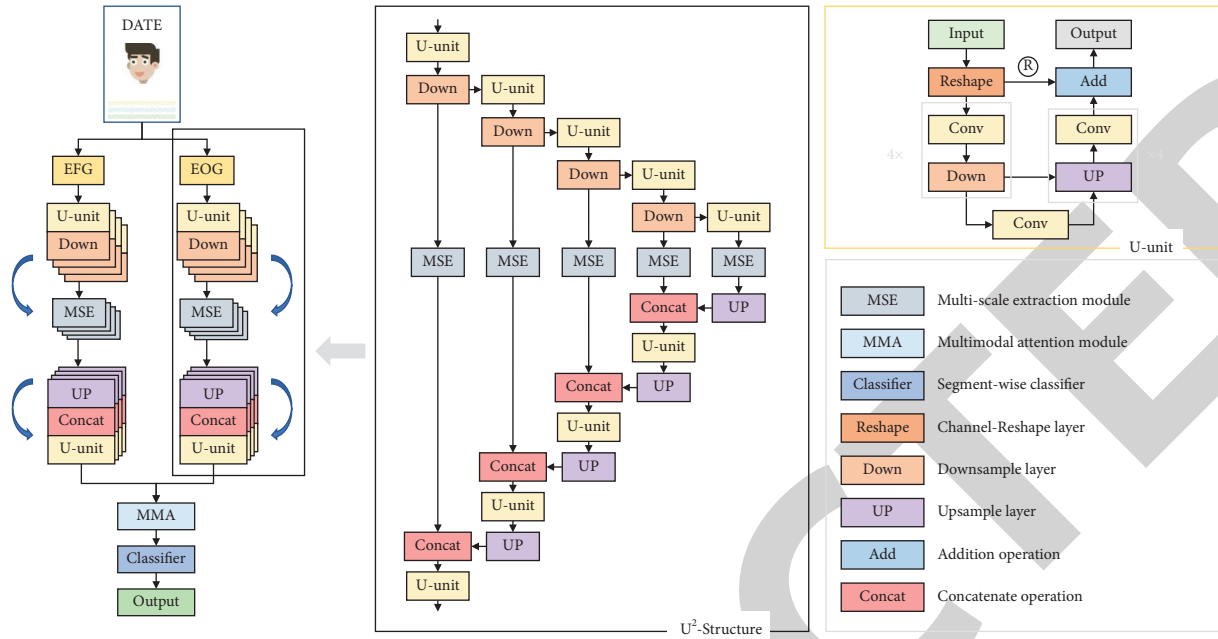


FIGURE 1: Multimodal significance waveform detection process.

2. Literature Review

Since the 1960s, a variety of note detection algorithms have been proposed, such as ACF (autocorrelation function) method, AMDF (average amplitude difference function) method, and SIFT (simplified inverse filtering) method. These traditional algorithms have high error rate and poor noise resistance. Many improved algorithms were proposed later, such as the modified autocorrelation algorithm and the cyclic AMDF method based on the AMDF algorithm. These algorithms have made significant progress compared with the traditional algorithms, but they are still difficult to achieve good performance in the case of strong background noise, which is highlighted by the error in the judgment of the relationship such as frequency doubling/half-frequency [5]. Verena and others have designed a system to separate vocal and piano accompaniment, but it is necessary to predict a priori information including approximate pitch [6]. Anshakov and others proposed to separate singing and music background by predicting the pitch contour and amplitude modulation of the sound signal. The progress of this series of research provides convenience for note extraction, but these systems also depend on the accurate extraction of notes to a considerable extent [7]. However, the frequency-domain algorithm and training model adopted by the above system have problems such as high computational complexity, so it is difficult to realize its function in hardware. In the stage of bar and segment feature extraction, Zheng and others proposed a dynamic bar pointer model based on Bayesian theory, which can distinguish the bar position of music, but it is only applicable to MIDI music with relatively constant music speed [8]. Xiao and others' sequential Monte Carlo method based on Bayesian theory

can extract bar information, but it is only applicable to single-tone audio music [9]. Goux and others proposed a method of dividing music segments by using the similarity between bars, which judges the similarity of bars by calculating the cosine of the vector including the angle between adjacent small knots, to divide the music into multiple independent segments [10]. In the stage of music emotion feature recognition, Tan and others believe that what inspires people's inner emotions is the characteristics of music, such as tonal interval, melody intensity, beat, rhythm, speed, and timbre, which are reflected through music segments. These features are combined into a vector to represent the characteristics of segments, which can not only effectively distinguish the differences between segments but also retain the music information of segments [11]. Hanifa and others used BP neural network to establish an automatic emotion recognition model for emotion feature recognition but used MIDI format music files [12]. In addition, for the Internet of Things itself, Ding and others proposed a structured feature extraction method for mobile application recognition. According to the change characteristics of mobile application traffic, this method proposes an HTTP streaming structured feature extraction method to accurately identify the applications to which the mobile network traffic belongs. The traffic collection tool is used to collect the required data for feature extraction and traffic clustering. In this process, the common subsequence can be obtained, and the structural features can be obtained by replacing characters. Experiments show that this method can realize structured feature recognition for most mobile applications, involving a wide range, but the accuracy of feature recognition results is not high [13]. Chu and others proposed an efficient extraction method of key features of big data based on cloud

computing. This method obtains the task execution process of different stages through the MapReduce programming model in cloud computing, extracts the key features in big data through programming, evaluates the local features, selects the key features according to the evaluation results, and proposes a phase space reconstruction method to ensure the invariance of data features. The experimental verification shows that the feature extraction results of this method can meet the general requirements, but in the face of massive big data, the feature extraction process will be complex and the processing efficiency will be low [14]. In view of the above, this paper proposes a time-domain algorithm based on human voice features and the framework of open-loop and closed-loop gene extraction to extract the gene features of songs with accompaniment; then, the section and its features are extracted by using the windowed moving matching algorithm, and the segments are divided by using the similarity between adjacent segments to obtain the segment features with emotional factors. The segment features are input into the improved BP emotion recognizer for emotion recognition. Finally, the intelligent singing skills of the whole music are determined.

3. Research Methods

3.1. Vocal Gene (Note) Feature Extraction Stage. The key to extract songs from songs with accompaniment is to use the fundamental difference between human voice and music background to strengthen voice features to facilitate pitch extraction. We studied the vocal characteristics of human beings, focused on the periodic distribution of voiced energy in the frequency domain, and proposed a voiced energy judgment method based on SIFT to solve the common problem of frequency doubling/half-frequency judgment errors in traditional time-domain algorithms [15]. Only one decision method is still not enough to ensure the accuracy of pitch extraction. Therefore, we learn from the pitch extraction part of 3GPPAMR-WB + standard coding and use the open-loop closed-loop method to determine the pitch, that is, first determine the approximate value of pitch and then make an adjustment and accurate values, to improve the accuracy of pitch extraction and the noise resistance of the system.

3.1.1. Voiced Energy Decision Method Based on SIFT. Sift method is a classical traditional algorithm. Its core is to use LPC (linear predictive coding) to offset the modulation effect of human channel parameters. In the digital model of speech signal, the excitation function of voiced voice is periodic, and its period is $1/F_0$, that is, pitch period. Speech can be regarded as a pitch pulse, which is generated after channel modulation. Based on the speech signal model of LPC, the combined effects of lip radiation, vocal tract, and glottal excitation are represented by a short-time constant all-pole digital filter.

$$H(z) = \frac{1}{A(z)}. \quad (1)$$

$A(z)$ is the inverse filter.

$$A(z) = 1 - \sum_{m=1}^p a_m z^{-m}, \quad (2)$$

where p is the order of LPC analysis. Let the speech signal pass through $A(z)$ to obtain the input signal of the filter $H(z)$, that is, the excitation function of voiced sound, which is used for pitch detection.

In the traditional SIFT algorithm, after the data are filtered by the inverse filter, the extreme value of the autocorrelation function is calculated as the final pitch output. However, due to the strong interference of background noise in music, the periodicity of the signal cannot be well reflected, so the error rate of extracting the pitch value by the autocorrelation function is large. Based on the periodic distribution of voiced energy in the frequency domain, we use the voiced energy decision method to extract the pitch value instead of the direct autocorrelation function. Multiple peaks on the autocorrelation spectrum are selected as the candidate pitch, and the peak filter is used at each frequency doubling point of the candidate pitch to obtain the signal energy of the voiced pulse sequence corresponding to the candidate pitch. The peak filter here is a narrowband bandpass filter, and the bandwidth is set according to the voiced pulse bandwidth of the human voice. Select the pitch value with the largest corresponding energy as the output pitch [16].

3.1.2. Open Loop, Closed-Loop Pitch Extraction. Using the voiced energy decision method based on SIFT described above, to solve the judgment error of the octave/half-frequency relationship commonly existing in the traditional pitch extraction algorithm, and to reduce the complexity of the voiced energy decision method, the pitch period is roughly estimated through open-loop pitch analysis and recorded as top [17]. On this basis, the candidate closed-loop pitch set δ is set by using the information of the first and last frames of the frame in the UN subtracted sampling data and the values around the open-loop pitch value. The 21 pitch candidate values included in δ are used to further ensure the smoothness and correctness of the pitch trajectory, which extracts the closed-loop pitch and records it as TCL, and define the conversion relationship between closed-loop pitch and open-loop pitch of a frame as follows:

$$T_{cl} = nT_{OP}, \quad (3)$$

where n is the multiple of minus sampling. (derived from the time-domain algorithm for extracting the pitch of songs with accompaniment).

3.1.3. Pretreatment. The noise mainly comes from the interference of electrical equipment, such as peak pulses and ripple voltage. These interferences cannot be filtered and eliminated through the power supply circuit. The interference of the power frequency signal is the most serious noise. It has different characteristics such as period, frequency, and amplitude of the signal of music, so it will have a great impact on the analysis. PCM-encoded music signal has the

characteristics of high sampling density and large amount of data, so it can truly restore the music signal, but it also increases the amount of data calculation for extracting the beat [18]. To filter or reduce the interference of 50 Hz power frequency signal and reduce the impact of large amount of calculation, the method of accumulating all sampling data in 20 ms is adopted. The reason 20 ms is used as the accumulation unit here is that the period of 50 Hz interference signal is 20 ms. The formula is

$$S_j = \sum_{i=-441}^{i+441} D_m. \quad (4)$$

Figure 2 shows the waveform of the compressed signal. The peaks and peak points of the PCM signal waveform are the same, maintaining the characteristics of the energy peaks. However, there are still a lot of local bumps, which affect the accuracy of extraction.

3.1.4. Extract Envelope. After signal preprocessing, the computation is greatly reduced [19]. However, the filtered signal not only contains beat information but also contains some information irrelevant to the beat. It can be seen from the above that the beat information is hidden in the 50 Hz~200 Hz frequency band of the music signal, while the frequency band above 200 Hz is occupied by those nonbeat information. Therefore, in the music signal, the main component is the low-frequency component. However, a variety of approximation polynomials can be used to approximate the design of this filter. The time-domain expression of the mean filter is

$$h_v(t) = \begin{cases} \frac{1}{D}, & |t| \leq \frac{D}{2} \\ 0, & |t| > \frac{D}{2} \end{cases}, \quad (5)$$

$$H_v(\Omega) = \int_{-\infty}^{+\infty} H_v(t) e^{-j\Omega/\Omega_c t} dt,$$

$$\int_{-\infty}^{+\infty} H_v(t) e^{-j\Omega/\Omega_c t} dt = Sa\left(\frac{D\Omega}{2\Omega_c}\right),$$

where Ω_c is the cut-off frequency. The approximation filter of Gauss filter can be constructed according to the cascade of multistage mean filter. The self-multiplication of the frequency response of the mean filter is its frequency response:

$$H_{vn}(\Omega) = Sa^n\left(\frac{D\Omega}{2\Omega_c}\right). \quad (6)$$

When $n \rightarrow \infty$, H_{vn} is the Gauss filter. Its limit expression is

$$\lim_{n \rightarrow \infty} Sa^n\left(\frac{D\Omega}{2\Omega_c}\right). \quad (7)$$

A Gaussian low-pass filter is introduced to filter the preprocessed PCM signal, and finally, a smooth envelope

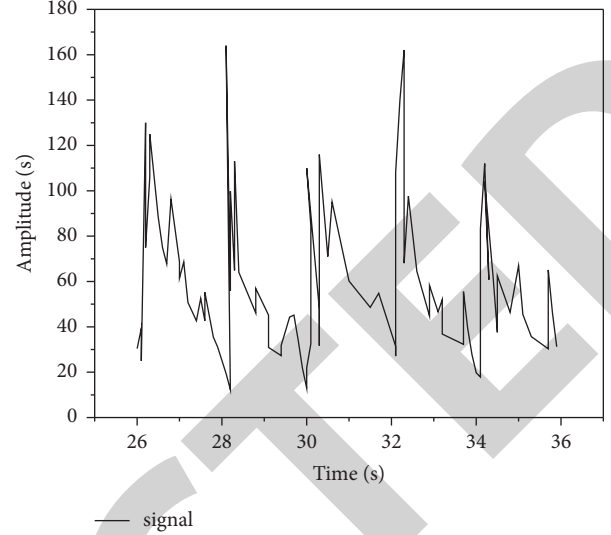


FIGURE 2: Compressed signal.

containing only low-frequency signals is obtained as shown in Figure 3.

The research object of this paper is the music file with a sampling frequency of 44100 Hz and sampling bits of 16 bits. This format music file is the most widely used music file at present. Through the analysis and research of WAV files, the sampling data representing music information is extracted. Through the above algorithm research, the envelope of PCM signal energy change is obtained. Peak detection in the time domain has the advantage of intuitive image, but some peak points with small changes will be omitted [20, 21]. For peak detection in the frequency domain, although it is not easy to omit the peak points at the time points with large change rate, it will also include some nonpeak points with large change rate. Here, we can extract the peak points by combining the time domain and frequency domain. Although the amount of operation has increased, the accuracy can be greatly improved. The time interval of each two notes and make statistics are calculated as shown in Figure 4.

As can be seen from the figure, the recording interval is mainly concentrated on the interval of 0.6~0.9 seconds over the peak. If the point is the peak and the time interval from the previous note is 0.6 seconds to 0.9 seconds, then the point is the starting point; otherwise, it is an error. Once the starting point of a note is determined, the fundamental characteristics of the note, such as length, frequency, and intensity, can be derived from it.

3.2. Feature Extraction Stage of Sections and Passages. The music section is composed of the sequence of notes extracted in the previous step, and the music section is composed of multiple sections. According to the strong and weak law of notes, the music is divided into several sections, and the characteristic information of sections is extracted. Then, according to the similarity of adjacent sections, the music is divided into several segments, and the characteristic information of the segments is extracted [22].

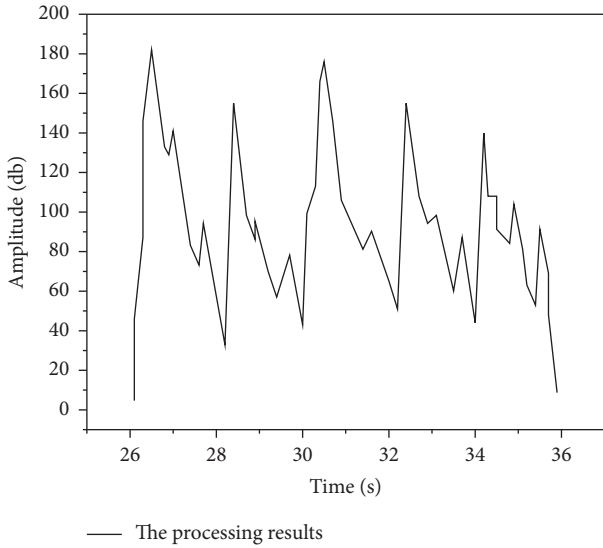


FIGURE 3: Processing results after Gaussian filtering.

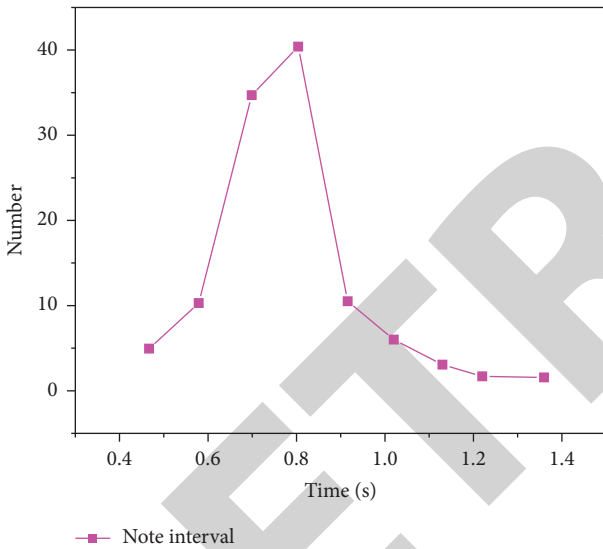


FIGURE 4: Notes interval statistics.

3.2.1. *Subsection Feature Extraction.* Bars in music show the law of alternating strength, which is fixed in a certain time range. This law can be used to extract the position of the section. In music, time is divided into equal basic units, and each unit is the length of time of a note [23]. Therefore, the time length of each note is the same, and the time length of sections containing the same number of notes is the same. Music is always composed of the alternation of strong beats and weak beats. This alternation cannot be disorderly and arbitrarily arranged but can organize a section according to a certain law of strength and weakness and then cycle on this basis. The strength law of music has been set in the production process of music scores, that is, as long as the music score creation of a piece of music is completed, its strength law is fixed. In front of the score, it will be marked by 2/4, 3/

4, and other ways. A concert has a variety of strength laws, but its strength laws are fixed within a certain time range.

Figure 5 is the envelope after signal sampling data compression, and Figure 6 is the position of the extracted notes. In the 10 seconds of Figure 6, the note shows such a form: “strong, weak, weak, strong, weak, weak, strong, weak, strong, weak, weak, strong, weak, weak, strong, weak, weak, weak.” It can be seen that during this period, it has been repeating the rhythm law of “strong, weak, and weak.” Only between 30 seconds and 32.5 seconds, there is a “strong and weak.” There are two interpretation methods here, which change the rhythm law from “strong, weak, and weak” to “strong and weak,” because there is only one “strong and weak,” which is unlikely. There is an error, and a weak beat is not extracted. On the contrary, looking at Figure 5, there was originally a local maximum at the position of 30 seconds, but it was not distinguished because the front-to-back difference was too small. Therefore, there is an error [24]. It can be seen that if the bar lines are divided, in turn, the accuracy of extracting notes in front can be optimized.

The position of the note has been extracted, which is based on the time point. Therefore, first locate the time points of these notes and read the pitch of these time points. The template set $M = \{m_1, m_2, \dots, m_n\}$ is established according to several strength laws of music. Set a time window W with a length of m and step length of $m/2$ and process according to the following algorithm:

In the time window, match the data with the template M . If the probability of the occurrence of a beat strength law is greater than 50%, the law in the first $m/2$ time has been determined, and mark and divide the bar line. Adjust the size of m as a multiple of the length of the bar line according to the length and law of the bar line.

Move the time window forward by $M/2$ steps for template matching. If the probability of occurrence of a beat strength law is greater than 50% and the law is the same as that in the previous step, repeat this step; otherwise, skip to the next step.

If the probability of occurrence of a rhythm strength law is greater than 50%, but is different from the previous step, then it is deemed that there is a dividing point of regular transformation in this time window. The length of the bar line of the strength law determined in the previous step is the step length to move the window and match until the matching is unsuccessful. The length of the bar line of the later law is matched. If it is successful, it indicates that this point is the dividing point of the conversion.

Resize m and repeat step 2.

At the end of the window, the bar line extraction is completed.

Bars are composed of note sequences, so the feature vector of bars can be constructed by the characteristics of notes. The basic characteristics of notes are pitch, intensity, and length. Therefore, the characteristics of these characteristics can be used to represent the feature vector of bars. The formula is as follows:

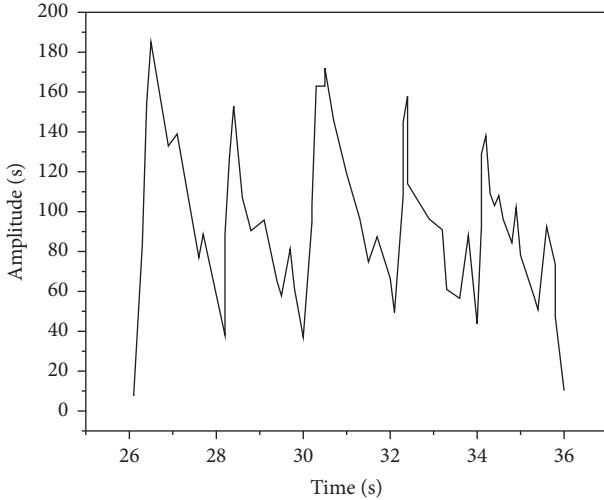


FIGURE 5: Envelope signal compression.

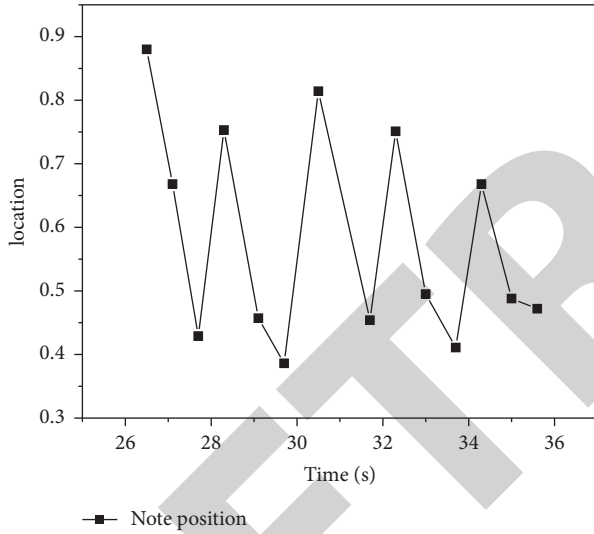


FIGURE 6: Note position.

$$Pa = \frac{1}{M} \sum_{i=1}^m P_i,$$

$$Ps = \sqrt{\frac{1}{M} \sum_{i=1}^m (P_i - Pa)^2},$$

$$Ia = \frac{1}{M} \sum_{i=1}^m I_i,$$

$$Is = \sqrt{\frac{1}{M} \sum_{i=1}^m (I_i - Ia)^2},$$

$$Da = \frac{1}{M} \sum_{i=1}^m D_i, \quad (8)$$

$$Ds = \sqrt{\frac{1}{M} \sum_{i=1}^m (D_i - Da)^2},$$

where Pa is the average value of pitch; Ps is the stability of pitch, Ia is the average value of sound intensity, Is is the stability of sound intensity, Da is the average value of sound length, and the stability of Ds is the sound length. Then, the eigenvector of the section is

$$BV = \{Pa, Ps, Ia, Is, Da, Ds\}. \quad (9)$$

3.2.2. Segment Feature Extraction. A passage is made up of subsections, and the adjacent subsections in the same passage are similar. Therefore, by judging the similarity of adjacent subsections, the contiguous subsections with similarity can be connected to form a passage. The similarity value of adjacent sections and the number of sections in a segment have a certain range, so three thresholds are set, namely, eigenvector similarity threshold EST (eigenvector similar threshold old), maximum length threshold malt (maximum length threshold), and minimum length threshold MNLT (minimum length threshold). There is no strict standard for the selection of the threshold, and the empirical value is usually selected. Whether the threshold is correct or not will affect the accuracy of the algorithm, and the most influential is the similar threshold. Some bars in the music play a transitional role, which can be divided into the next paragraph. Pb is used to represent the number of small stanzas in the passage, MB is used to represent the total number of small stanzas in the music, $\delta(BV_i, BV_{i+1})$ is used to represent the similarity of adjacent stanzas, and BV_i and BV_{i+1} are used to represent the eigenvectors of two adjacent stanzas B_i and B_{i+1} , respectively. The following algorithm is adopted:

- (1) Judge the size of i and MB . If $i \geq MB$, skip to 5; otherwise, use the cosine of the included angle of the vector to calculate the distance measurement between the two eigenvectors, as shown in formula (10), and continue with 2.

$$\delta(BV_i, BV_{i+1}) = \frac{BV_i \times BV_{i+1}}{\|BV_i\| \times \|BV_{i+1}\|}. \quad (10)$$

- (2) If $\delta(BV_i, BV_{i+1}) > EST$, continue with 3; otherwise, jump to 4.
- (3) If $PB < MALT$, connect B_i and B_{i+1} ; $i++$; $PB++$, and repeat 1; otherwise, start a new paragraph, $i++$; $PB=0$ and repeat 1.

- (4) If $PB < MNLT$, connect B_i and B_{i+1} ; $i++$; $PB++$, and repeat 1; otherwise, start a new paragraph, $i++$; $PB = 0$ and repeat 1.
- (5) Exit the loop, and the segment extraction is completed.

The reason music can stimulate people's inner feelings and arouse people's resonance is that music has different characteristics such as melody, beat, intensity, rhythm, speed, timbre, and tonal interval. These characteristics are reflected by the passage. Therefore, we can use these features to express the passage.

3.3. Music Emotion Feature Extraction Stage. To use the segment feature vector extracted by the above process to recognize the emotion of the segment, it is necessary to construct the mapping from the feature vector to the emotion of the segment. Therefore, an automatic emotion recognizer based on the improved BP neural network is constructed to recognize the emotion of the passage.

First, the weights and thresholds of the BP neural network are decimals between 1 and 1, which are large numbers and are not suitable for binary encoding. This document specifies the rules for encoding real numbers. The algorithm uses the standard squared error answer as a criterion for evaluating people during the search process. The selection operator determines the initial package and calculates the fitness value of each population; the individuals in the population are arranged in descending order from large to small; the first two people are selected to be directly inherited to the next generation, and the others are equally divided into excellent, good, medium, and poor. There are no two great individuals, one great individual, one average individual, and one poor individual. Adaptive Genetic Algorithm (AGA) is the crossover ability of individuals in a population and the probability of mutation according to the environment. The improved BP neural network algorithm can be used as a classifier to identify the emotional features of music clips. The algorithm can use the sample error information in the calibration process, avoid overadjustment and insufficient connection caused by the fixed structure of the neural network, and improve the training accuracy. To solve these problems, this paper adopts a human-computer interaction method for the emotion annotation of learning examples, and the number of initial potential centers is determined by the number of emotion types in the training samples. To improve the training efficiency and reduce the network training time and complexity, the widths of the center of the first underground layer and the center of the underground layer are adjusted for each training sample package.

First, an automatic emotion recognizer of the BP neural network is developed. Since the segment privately vector in the input layer has seven dimensions, the number of neuron nodes in the input layer is set to 7. In the output layer, the emotion representation space is set as a four-dimensional vector. According to Hewner's model, a four-dimensional vector is chosen that contains inspiration, joy, sadness, and

lyrical expressions as components of the emotional features. (1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), and (0, 0, 0, 1) represent the four emotions of inspiration, joy, sadness, and sad lyrics. Therefore, the number of neuron nodes in the output layer is set to 4. Hidden layer: the Gaussian function has the advantages of simple imaging, good smoothness, and good analytical performance, so the activation function of the hidden layer of the neural network is set to Gaussian type function as shown in the following formula:

$$\varphi(x, c_i) = \exp\left(-\frac{\|x - c_i\|^2}{\sigma^2}\right), \quad i = 1, 2, \dots, k, \quad (11)$$

where σ is the corresponding width, c_i is the i th hidden layer node, and $\|x - c_i\|^2$ is the Euclidean norm.

4. Result Analysis

4.1. Analysis of Feature Extraction Results. Under the guidance of the music teacher, the song samples are divided into five categories: fast, fast three, middle three, slow three, and waltz. The note-taking algorithm in this paper is used to decompose Ganku's notes. The results compared with other logging algorithms are shown in Table 1 and Figures 7–9.

Table 1 and Figures 7–9 show the average percentage of false detections, the average percentage of missed detections, and the average cost of several logging algorithms. The results show that the average false detection rate and average missed detection rate of this algorithm are higher than other algorithms. Although the average time is longer than the other two algorithms, it is acceptable. Therefore, the algorithm can make recording more efficient and accurate.

We have collected 201 pieces of music of various types and lengths, along with associated notes, in the form of wave music files. Music is divided into five categories: march, fast three, middle three, slow three, and waltz. Eighty pieces of music were randomly selected, and the notes and lines of each paragraph were manually marked. The program is designed to extract banknotes, obtain the banknote location, and then compare it with handwritten banknote location information. The results are shown in Table 2. For waltzes and decelerated thirds, the boundaries between the low-frequency signal notes are sharper, and the envelope peaks are more easily distributed, resulting in a higher level of note identification. For fast-paced music, such as fast march and march, the peak of the envelope is uncertain, so the level of note recognition is relatively low, and the number of false detections and errors is relatively high. However, the overall acceptance remains high.

Through the pitch, length, and intensity of notes, the feature vector of sections is analyzed, and then, the segments are divided according to the similarity of adjacent sections to obtain the seven-dimensional feature vector of the segments. 80 pieces of music are extracted to train the improved BP neural network. After the training, the remaining 121 song samples were labeled with emotions. Section extraction, segment extraction, and segment feature extraction are carried out, respectively. Take the seven-dimensional

TABLE 1: Comparison of note extraction algorithms.

Extraction method	Average false detection rate	Average missed detection rate	Average time
Algorithm in this paper	0.07	0.08	1.62
Parallel processing algorithm	0.09	0.11	1.12
Harmonic peak method	0.20	0.09	1.30

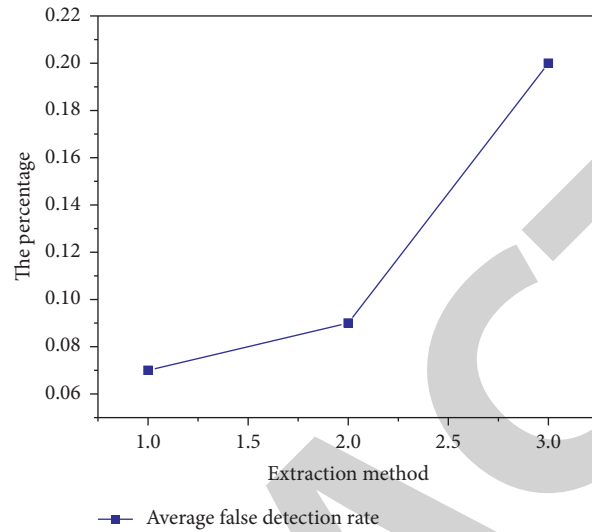


FIGURE 7: Average false detection rate of note extraction algorithm.

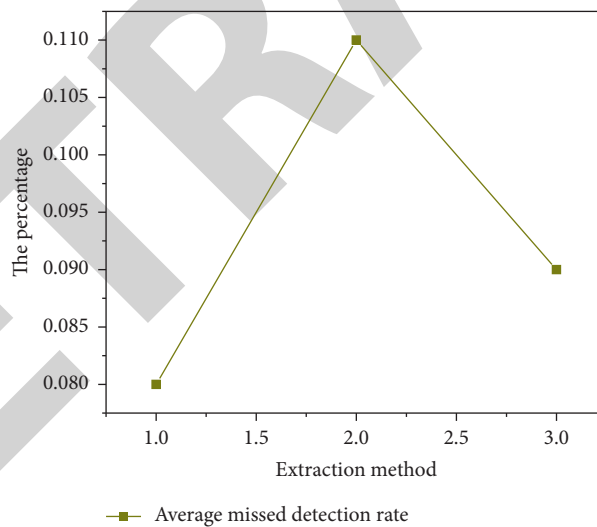


FIGURE 8: Average missed detection rate of note extraction algorithm.

segment feature vector as the input, input it into the improved BP neural network emotion recognition model, output the emotion types of the segment, count the proportion of each of the four music types, and finally get the main emotion of the music. The results are shown in Figure 10 and Table 3.

As can be seen from Figure 10 and Table 3, there is a small but acceptable error in correctly identifying most musical emotions and identifying musical emotions. Therefore, the method can extract musical features from different types of waveform music files.

4.2. Waveform Feature Extraction Performance Analysis.

To put the application value of the automatic feature extraction method of intelligent singing skill information into practice, the round AMDF method feature extraction method (method 1) based on the AMDF algorithm is an efficient key extraction method. The cloud computing-based big data features (method 2) and the Bayesian theory-based dynamic bar pointer design are compared with the proposed unpacking method (method 3). Taking the extraction accuracy as the index, compare the extraction effects of different methods, and the results are shown in Figure 11.

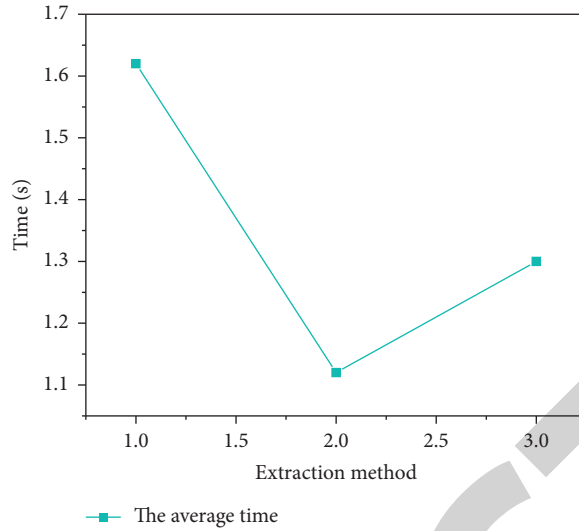


FIGURE 9: Average time consumption of note extraction algorithm.

TABLE 2: Pitch extraction results of different music.

Song name	Category	Music length (s)	Actual number of notes	Detect the number of notes	Number of misdetected notes	Number of missing notes
Chopin valse	Waltz	47	67	67	1	1
Waltz of flowers	Waltz	19	27	28	0	1
The last waltz	Slow three	17	25	27	1	1
Soul broken blue bridge	Slow three	27	41	41	2	0
Hawthorn	Middle three	41	60	63	1	0
I care about you	Middle three	29	46	45	1	2
Vienna forest	Fast three	19	32	33	3	2
Danube night	Fast three	48	75	74	1	2
Patrol	March	30	49	48	2	3
Schubert	March	50	83	83	3	3

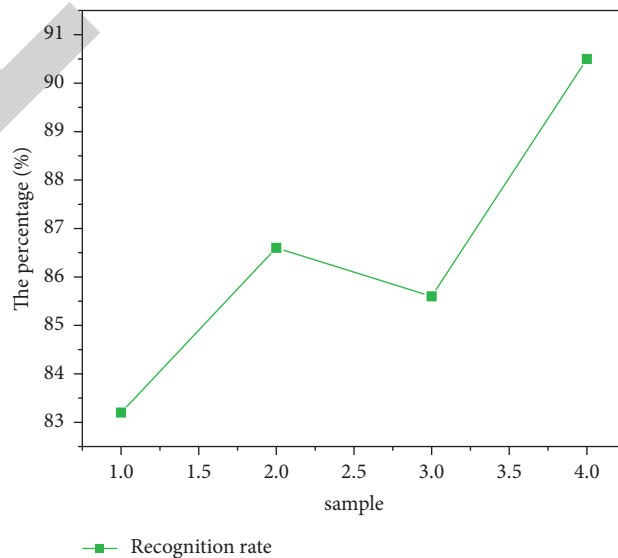


FIGURE 10: Music emotion recognition rate.

TABLE 3: Emotion recognition results.

Sample	Number	Number of successful identifications	Number of false identifications	Recognition rate
Enthusiasm	24	20	4	83.2
Cheerful	30	26	4	86.6
Sad	35	30	5	85.6
Lyric	32	29	3	90.5

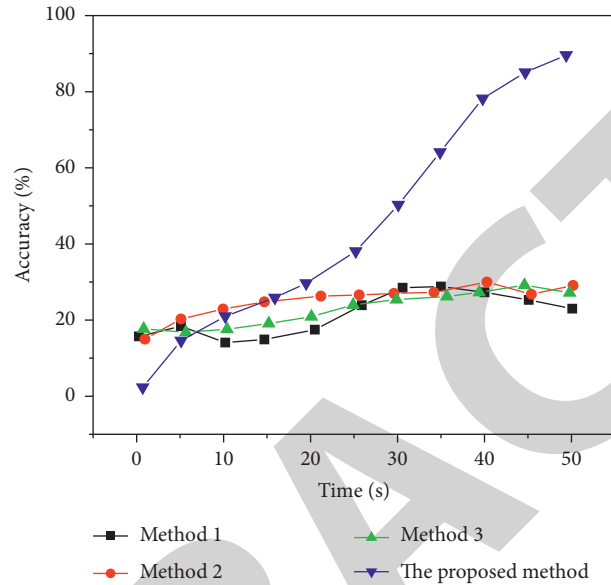


FIGURE 11: Extraction accuracy of different extraction methods.

According to the analysis of Figure 11, with the increase of feature extraction time, the accuracy of extraction results of existing methods changes little, which is basically maintained between 15% and 30% at a low level. When the proposed method is for feature extraction of intelligent singing skill information, the accuracy shows a continuous growth trend, and with the growth of time, its accuracy is significantly higher than the existing methods, indicating that the proposed method has significant advantages in the accuracy of feature extraction. This method uses the multifeature discrimination method to extract the features and realizes the accurate extraction of the valuable features by distinguishing the valuable features and worthless features in the intelligent singing skill information. It can be applied to the field of Internet of Things information processing to provide technical support for relevant information processing.

5. Conclusion

Based on the characteristics of the wave of singing skills under the Internet, this paper successfully identifies the characteristics of notes, measures, segments, and tunes in three stages: dismantling the tonal characteristics, dissecting the segment and segment characteristics, and dissecting the emotional characteristics. Experiments show that the method of extracting music features can effectively generate different types of music features, and the accuracy of feature

recognition can meet the requirements of unraveling the waveform features of singing skills in the Internet context. Compared with other algorithms, it has higher accuracy, reliability, and higher extraction speed. Music is complex and varied. Different music has different music theories and signals. Even a piece of music can have different emotional expressions in different environments. To accurately and quickly extract the characteristics of music, not only the improvement and development of computer-related technologies but also the support of music theory, psychology, and other fields is needed. This research is still in its infancy, and there is still a vast area of research waiting for the efforts of researchers. Hope that this article can provide a reference for other researchers' research.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this article.

References

- [1] T. Lin, "Application of feature extraction method based on support vector machine in Internet of things," *Journal of*

- Intelligent and Fuzzy Systems*, vol. 39, no. 6, pp. 8623–8632, 2020.
- [2] A. Yaseen, M. Nazir, A. Sabah et al., “Dimensionality reduction for internet of things using the cuckoo search algorithm: reduced implications of mesh sensor technologies,” *Wireless Communications and Mobile Computing*, vol. 2020, pp. 1–21, Article ID 8897026, 2011.
 - [3] Y. Wang, J. Yan, Z. Yang, Y. Zhao, and T. Liu, “Optimizing gis partial discharge pattern recognition in the ubiquitous power Internet of things context: a mixnet deep learning model,” *International Journal of Electrical Power & Energy Systems*, vol. 125, no. 4, Article ID 106484, 2021.
 - [4] W.-L. Chen, Y.-B. Lin, F.-L. Ng, C.-Y. Liu, and Y.-W. Lin, “Ricetalk: rice blast detection using Internet of things and artificial intelligence technologies,” *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1001–1010, 2020.
 - [5] G. Oliva, T. Zarra, G. Pittoni et al., “Next-generation of instrumental odour monitoring system (IOMS) for the gaseous emissions control in complex industrial plants,” *Chemosphere*, vol. 271, no. 9, Article ID 129768, 2021.
 - [6] V. Epp, “Beth spacey, the miraculous and the writing of crusade narrative. (Crusading in context.) woodbridge, boydell press 2020,” *Historische Zeitschrift*, vol. 312, no. 2, pp. 499–500, 2021.
 - [7] G. P. Anshakov, A. I. Belousov, A. V. Sedelnikov, and A. S. Taneeva, “Analysis of micro-acceleration requirements in the context of designing a small technological spacecraft,” *Journal of Physics: Conference Series*, vol. 1791, no. 1, Article ID 012001, 2021.
 - [8] T. Zheng, B. Peng, and G. Zhou, “Industrial internet of things for mobile phone shell intelligent detection in smart cities,” *Wireless Communications and Mobile Computing*, vol. 2021, no. 4, pp. 1–12, Article ID 5567017, 2021.
 - [9] Z. Xiao, X. Xu, H. Xing, F. Song, X. Wang, and B. Zhao, “A federated learning system with enhanced feature extraction for human activity recognition,” *Knowledge-Based Systems*, vol. 229, no. 5, Article ID 107338, 2021.
 - [10] N. Goux, J. B. Casanova, G. Pillonnet, and F. Badets, “A 6nw 0.0013mm² ilo bandpass filter for time-based feature extraction,” *IEEE Solid-State Circuits Letters*, vol. 99, p. 1, 2020.
 - [11] J. Tan, D. Xia, S. Dong, H. Zhu, and B. Xu, “Research on pre-training method and generalization ability of big data recognition model of the internet of things,” *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 5, pp. 1–15, 2021.
 - [12] R. M. Hanifa, K. Isa, and S. Mohamad, “A review on speaker recognition: technology and challenges,” *Computers & Electrical Engineering*, vol. 90, no. 4, Article ID 107005, 2021.
 - [13] X. Ding, Y. Xie, P. Li, M. Cui, and J. Chen, “Image Steganography Based on Artificial Immune in mobile Edge Computing on Internet of Things,” *IEEE Access*, vol. 99, p. 1, 2020.
 - [14] C.-H. Chu, C.-J. Lee, and H.-Y. Yeh, “Developing deep survival model for remaining useful life estimation based on convolutional and long short-term memory neural networks,” *Wireless Communications and Mobile Computing*, vol. 2020, no. 6, pp. 1–12, Article ID 8814658, 2020.
 - [15] C. Tan, Y. Sun, G. Li, G. Jiang, D. Chen, and H. Liu, “Research on gesture recognition of smart data fusion features in the IoT,” *Neural Computing & Applications*, vol. 32, no. 22, pp. 16917–16929, 2020.
 - [16] M. M. Hassan, S. Huda, S. Sharmeen, J. Abawajy, and G. Fortino, “Adaptive trust boundary protection for iiot networks using deep-learning feature-extraction-based semisupervised model,” *IEEE Transactions on Industrial Informatics*, vol. 99, p. 1, 2020.
 - [17] L. Chen and S. Li, “Human motion target posture detection algorithm using semi-supervised learning in Internet of things,” *IEEE Access*, vol. 99, p. 1, 2021.
 - [18] W. Yang, G. Zhang, T. Zhu, M. Cai, H. Zhao, and J. Yan, “GIS partial discharge patterns recognition with spherical convolutional neural network,” *IOP Conference Series: Materials Science and Engineering*, vol. 853, no. 1, Article ID 012001, 2020.
 - [19] R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar, and B. Qureshi, “An overview of lot sensor data processing, fusion, and analysis techniques,” *Sensors*, vol. 20, no. 21, p. 6076, 2020.
 - [20] C. Yin, S. Zhang, J. Wang, and N. N. Xiong, “Anomaly detection based on convolutional recurrent autoencoder for iot time series,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 1, pp. 112–122, 2022.
 - [21] R. Kaluri, D. Singh Rajput, Q. Xin et al., “Roughsets-based approach for predicting battery life in Iot,” *Intelligent Automation & Soft Computing*, vol. 27, no. 2, pp. 453–469, 2021.
 - [22] M. A. Asghar, M. J. Khan, M. Rizwan, M. Shorfuazzaman, and R. M. Mehmood, “AI inspired EEG-based spatial feature selection method using multivariate empirical mode decomposition for emotion classification,” *Multimedia Systems*, vol. 99, pp. 1–14, 2021.
 - [23] S. Ahmadi-Asl, S. Abukhovich, M. G. Asante-Mensah, A. Cichocki, and I. Oseledets, “Randomized algorithms for the computation of Tucker decomposition and higher order svd (hosvd),” *IEEE Access*, vol. 99, p. 1, 2021.
 - [24] Y. Liu, J. Chen, N. Bao, B. B. Gupta, and Z. Lv, “Survey on atrial fibrillation detection from a single-lead ECG wave for Internet of Medical Things,” *Computer Communications*, vol. 178, no. 1, pp. 245–258, 2021.