Research on Neural Network-based Automatic Music Multi-Instrument Classification Approach

Ribin Guo
Shanxi College of Applied Science and Technology, Taiyuan, Shanxi 030062, China

***Abstract*—The automatic classification of multi-instruments plays a crucial role in providing services for music retrieval and recommendation. This paper focuses on automatic multi-instrument classification. Firstly, instrument features were analyzed, and Mel-frequency cepstral coefficient (MFCC) and perceptual linear predictive coefficient (PLPC) were extracted from instrument signals. Features were selected using the entropy weight method. The optimal initial weight threshold of a back-propagation neural network (BPNN) was obtained by utilizing the sparrow search algorithm (SSA), achieving a SSA-BPNN classifier. Experiments were conducted using the IRMAS dataset. The results demonstrated that the combination of MFCC and PLPC selected through the entropy weight method achieved the best performance in automatic multi-instrument classification. The method yielded high P value, recall rate, and F1 value, 0.72, 0.71, and 0.71, respectively. Moreover, it outperformed other algorithms such as support vector machine and XGBoost. These results confirm the reliability of the automatic multi-instrument classification method proposed in this paper, making it suitable for practical applications.**

Keywords*—*Neural network; musical instrument; automatic classification; auditory feature; sparrow search algorithm

I. Introduction

Music serves as a medium for conveying emotions [1]. With the continuous progress and popularization of computer technology, an increasing amount of music information circulates and spreads on the internet, offering users a more convenient means to enjoy music. In order to enhance the user experience further, music information retrieval (MIR) has gradually emerged as a crucial area of research [2]. Through MIR, users can efficiently discover their preferred music. MIR research includes the identification and classification of musical instruments, genres, and styles [3]. Automatic classification of musical instruments refers to the use of intelligent algorithms to automatically classify different musical instruments through processing their signals. In multi-instrument automatic classification, different instrument signals can easily interfere with each other, resulting in a decrease in classification effectiveness.

Neural networks are a type of machine learning that possess strong self-learning capabilities, enabling them to derive useful conclusions from a series of complex and seemingly unrelated data. They have wide applications in fields such as speech recognition, image processing, and automation control, where they demonstrate certain advantages in handling complex data. Therefore, in order to further improve the accuracy of multi-instrument automatic classification, this study focuses on researching neural networks.

Based on the analysis of musical instrument signals, this article selected back-propagation neural network (BPNN) as the main algorithm. In order to further improve its performance, the sparrow search algorithm (SSA) was used to optimize BPNN. Finally, the performance of the proposed method was verified on the IRMAS dataset.

The main contribution of this article lies in providing a more accurate method for automatic classification of multiple musical instruments, which also offers some new references for music signal classification and even speech signal classification. This is conducive to promoting the further application of neural network algorithms in the field of music research.

II. Literature Review

With the continuous development of machine learning and other technologies, an increasing number of algorithms have been applied in MIR research.

In terms of music popularity prediction, Martin-Gutierrez et al. [4] introduced a deep learning architecture called HitMusicNet for predicting the popularity of music recordings. The experimental results demonstrated its superior predictive capabilities compared to existing techniques. Voetter et al. [5] presented two novel datasets to predict song popularity based on the data from AcousticBrainz, Billboard Hot 100, the Million Song dataset, and last.fm. They verified the usability of the designed dataset by performing experiments on different models.

In terms of singer recognition, Rajesh and Nalini [6] conducted experiments to validate the effectiveness of their approach in singer recognition, which involved the integration of MFCC with chroma DCT-reduced pitch features. Kooshan et al. [7] developed a singer recognition system by integrating deep learning and feedforward neural networks. They utilized long short-term memory (LSTM) to identify vocal frames within music segments, followed by the application of another LSTM for singer identification. Experimental validation confirmed the superior performance of this approach compared to existing methods.

In terms of music recommendation, Feng et al. [8] proposed to model and combine melody, chord, and rhythm features and utilized a multilayer perceptron for music recommendation. The experimental results indicated a 3.52% improvement over the best baseline. Elbir et al. [9] developed an innovative deep neural network that utilizes the acoustic attributes of music to classify genres and provide music recommendations.

In the domain of instrument recognition and classification, Chatterjee et al. [10] proposed employing a convolutional siamese networks along with its residual variation to identify instruments based on scalograms derived from audio recordings. They conducted experiments using two publicly accessible datasets to validate their approach. Wise et al. [11] developed an attention-enhanced convolutional neural network specifically designed for accurately classifying a wide range of 19 orchestral instruments, and they substantiated the effectiveness of this technique through experimental evaluations.

III. Instrument Feature Extraction

Different musical instruments emit distinct vibration frequencies, which, in turn, yield different sounds. Timbre represents the subjective, natural perception of these vibrations by humans. To achieve the automatic classification of various musical instruments, it is essential to extract features that reflect the timbral characteristics of the musical instrument sound signal. Currently, in the realm of speech signal recognition, a predominant reliance on time-domain and frequency-domain features is observed. These feature types are relatively straightforward and easy to analysis. However, they fall short in adequately characterizing timbre, resulting in suboptimal classification performance. This paper primarily focuses on the analysis of two categories of features derived from the human auditory system for musical instrument feature extraction.

MFCC is a feature obtained by simulating the auditory characteristics of the human ear [12], which has a good performance in speech recognition [13]. It is extracted in the following way.

1) Pre-emphasis, frame-by-frame, and windowing operations are performed on the original instrument audio signal to get pre-processed signal $\mathrm{x_i(m)}$.

2) A fast Fourier transform (FFT) is performed on $x_i(m)$, and

$$X(i,t) = FFT[x_i(m)] \tag{1}$$

is obtained.

3) The energy spectrum is calculated:

$$E(i,t) = [x_i(m)]^2. \tag{2}$$

4) A filtering operation is performed on $E(i,t)$. The spectrum energy is obtained through a group of triangular filters:

$$S(i,m) = \sum_{t-0}^{N-1} E(i,t) H_m(t), \tag{3}$$

where, $H_m(t)$ represents the transfer function of the triangular filter.

5) The final MFCC is obtained by discrete cosine transform (DCT):

$$X(i) = \sqrt{\frac{2}{M}} \sum_{m=0}^{M-1} lg[S(i,m)] \cos\left[\frac{\pi(2m-1)}{2M}\right], \tag{4}$$

where, M is the filter order.

PLPC is also a feature that mimics the auditory characteristics of the human ear [14] and is extracted as follows:

1) FFT is also performed on $x_i(m)$ to convert it to the frequency domain to obtain $X(i,t)$. $E(i,t)$ is calculated.

2) $E(i,t)$ is converted to the Bark domain:

$$\varphi(f) = 6\ln\left\{\frac{f}{600} + \left[\left(\frac{f}{600}\right)^2 + 1\right]^{0.5}\right\}, \tag{5}$$

where, φ is the value of the signal frequency in the Bark domain and f is the angular frequency.

3) The critical bands in the Bark domain are considered as a group of filters, and the function of each filter is:

$$\varphi\big(Z - Z_0(k)\big) = \begin{cases} 0, Z - Z_0(k) < -1.3 \\ 10^{Z-Z_0(k)+0.5}, -1.3 \le Z - Z_0(k) \le -0.5 \\ 1, -0.5 < Z - Z_0(k) < 0.5 \\ 10^{-2.5(Z-Z_0(k)-0.5)}, 0.5 \le Z - Z_0(k) \le 2.5 \\ 0, 2.5 < Z - Z_0(k) \end{cases}. \tag{6}$$

The output of each critical band is obtained:

$$\theta(k) = \sum_{Z-Z_0(k)=-1.3}^{2.5} E(\varphi - \varphi_i)\, \varphi\big(Z - Z_0(k)\big). \tag{7}$$

4) Equal loudness curve pre-emphasis is carried out on $\theta(k)$:

$$\Gamma(k) = D[f_0(k)]\theta(k), \tag{8}$$

where, $f_0(k)$ corresponds to central frequency $Z_0(k)$ of each filter, $Z_0(k)$ refers to the equal loudness curve,

$$D[f_0(k)] = \frac{16\pi^4 f_0^4(k)\left(4\pi^2 f_0^2(k) + 5.68\times10^7\right)}{\left(4\pi^2 f_0^2(k) + 6.3\times10^6\right)^2 \times \left(4\pi^2 f_0^2(k) + 3.8\times10^8\right)}. \tag{9}$$

5) Finally, the simulation of human ear hearing is realized by converting from intensity to loudness:

$$\Psi(k) = \Gamma(k)^{0.33}. \tag{10}$$

In the context of speech recognition tasks, MFCC is usually set as 13 dimensions along with their corresponding first-order and second-order difference coefficients, resulting in a total of 39 dimensions. Additionally, PLPC usually employs 24 filters to yield 24 dimensions. If both the 39-dimensional MFCC and the 24-dimensional PLPC features are utilized for automatic multi-instrument classification, the cumulative dimensionality reaches 63 dimensions. To mitigate the potential complexity and computational load associated with this higher dimensionality, this paper uses an entropy weight method [15] to optimize these features.

Information entropy is related to the probability of an event occurring. When it is applied to feature optimization, the greater the information entropy, the more disordered the distribution of that feature value is. For a $T_n \times T_m$ original feature set, it is assumed that the feature vector corresponding to the t_m-th feature item is Z_{t_m}, the process of feature optimization is presented in Fig. 1.

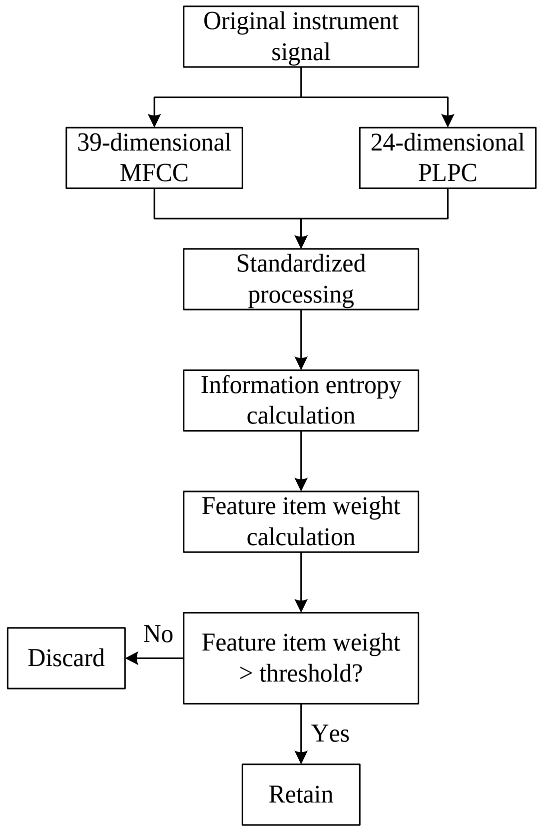

Fig. 1. The feature optimization process based on entropy weight method.

The specific calculation process is as follows:

1) Standardized eigenvalue:

$$Z'_{t_n,t_m} = \frac{z_{t_n,t_m} - min(z_{t_m})}{max(z_{t_m}) - min(z_{t_m})}. \quad (11)$$

2) The information entropy is calculated:

$$E(t_m) = -\sum_{t_n=1}^{T_n} p_x(t_n, t_m) \ln p_x(t_n, t_m), \quad (12)$$

where, $p_x(t_n, t_m)$ refers to the percentage of different eigenvalues in the corresponding term, $p_x(t_n, t_m) = \frac{X\prime(t_n,t_m)}{P_z(t_m)}$, $P_z(t_m) = \sum_{t_n=1}^{T_n} X'(t_n, t_m)$.

3) The feature term weight is calculated:

$$W_{t_m} = \frac{1-E(t_m)}{T_m - \sum_{t_m=1}^{T_m} E(t_m)}. \quad (13)$$

4) A threshold is set, and feature items with weights greater than the threshold are retained as subsequent features used for automatic multi-instrument classification.

IV. Automatic Classification Method Based on Neural Networks

Back-propagation neural network (BPNN) has excellent performance and is most widely used in automatic classification tasks [16]. Therefore, in this paper, BPNN is chosen as a classifier for automatic classification of multi-instrument. A simple BPNN structure is presented in Fig. 2.

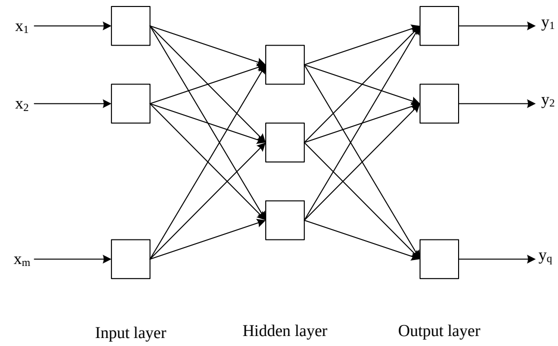

Fig. 2. The BPNN structure.

It is hypothesized that the number of nodes in the input layer, hidden layer, and output layer of BPNN is $m - p - q$. The output of the hidden layer during training can be written as:

$$h_j = f\left(\sum_{i=1}^{m} w_{ij}^{h} i_i - b_j\right). \quad (14)$$

The output of the output layer can be written as:

$$o_k = f\left(\sum_{j=1}^{p} w_{jk}^{o} h_j - b_k\right). \quad (15)$$

where, w and b denote the weight and threshold of each layer.

The global error function can be written as:

$$E = \frac{1}{2}\sum_{k=1}^{q}(y_k - o_k)^2, \quad (16)$$

where, y_k is the desired output. The BPNN continuously corrects the weight threshold according to the error until the error meets the requirements. However, the performance of BPNN is easily affected by the initial weight threshold, so it can be tempting to become trapped in the local optimum. In order to solve this problem, this paper uses the sparrow search algorithm (SSA) [17] to optimize the BPNN.

SSA is a sparrow-inspired algorithm that utilizes the foraging behavior of sparrows, which exhibits excellent global optimization performance and is used to calculate the optimal initial weight thresholds for BPNN. During the foraging process, some sparrows are responsible for finding the foraging area and direction, the rest of the sparrows feed, and some sparrows will sound an alarm when danger is detected. The population is categorized into discoverers, followers, and alerts, and their positions are updated as follows.

1) Discoverer:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} \cdot exp\left(-\frac{i}{\alpha \cdot t_{max}}\right), R < A_{ST} \\ X_{i,j}^{t} + QL, R \geq A_{ST} \end{cases}, \quad (17)$$

where, α is a random number in [0,1], R is an alert value in (0,1], and A_{ST} is the threshold at which the shift occurs, Q is a random number obeying a normal distribution, L is a $1 \times d$ matrix, $R < A_{ST}$ means that the area is dangerous and the finder needs to move to a safe area, and $R \geq A_{ST}$ means that the area is safe and sparrows can look for food.

3) Follower:

$$X_{i,j}^{t+1}=\begin{cases} Q\cdot exp\left(\frac{X_{worst}^{t}-X_{i,j}^{t}}{i^{2}}\right), i>\frac{n}{2} \\ X_{b}^{t+1}+\left|X_{i,j}^{t}-X_{b}^{t+1}\right|\cdot A^{+}\cdot L, i\leq\frac{n}{2} \end{cases}, \tag{18}$$

where, X_{worst} is the worst position of the individual, X_b is the best position of the discoverer, $A^+ = A^T(AA^T)^{-1}$, A is a $1 \times d$ matrix, randomly assigned as -1 or 1.

4) Watcher:

$$X_{i,j}^{t+1}=\begin{cases} X_{best}^{t}+\beta\cdot\left|X_{i,j}^{t}-X_{best}^{t}\right|, f_i>f_g \\ X_{i,j}^{t}+K\cdot\left(\frac{\left|X_{worst}^{t}-X_{i,j}^{t}\right|}{f_i-f_w+\varepsilon}\right), f_i=f_g \end{cases}, \tag{19}$$

where, X_{best} is the global optimal position, β is the step control parameter, f_i refers to the fitness of the i-th sparrow, f_g and f_w are the current best and worst fitness, K is a random number in (-1,1), and ε is a constant.

SSA finds the optimal weight threshold of BPNN by constantly updating the three positions. To improve the population diversity, this paper uses cubic mapping [18] to obtain the initialized population:

$$x_{n+1} = \rho x_n(1 - x_n^2), \tag{20}$$

where, ρ is the control parameter. The SSA-BPNN algorithm is used to realize the classification of different musical instruments, and the flow chart is presented in Fig. 3.

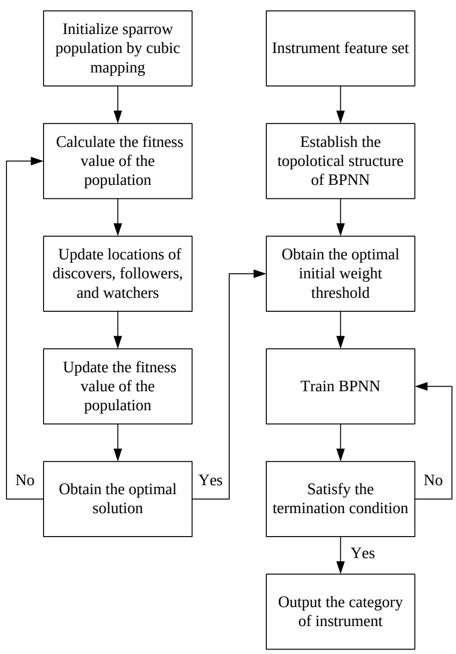

Fig. 3. The SSA-BPNN algorithm-based automatic multi-instrument classification method.

V. Results and Analysis

A. Experimental Setup

The algorithm was developed and trained in the MATLAB 2018b environment. The dataset used for the experiment was the IRMAS dataset [19], comprising a total of 6,705 WAV audio files. All audio files were in 16-bit stereo format and had a sampling rate of 44.1 kHz. A single performance clip of ten instruments, including cello, clarinet, and others, each with a duration of 3 s, was selected. The training set and test set configurations are detailed in Table I.

TABLE I. EXPERIMENTAL DATA SETS

Musical Instrument	Number in the Training Set	Number in the Test Set
Flute	451	163
Organ	682	361
Piano	721	995
Trumpet	577	167
Cello	388	111
Clarinet	505	62
Electric guitar	760	942
Violin	580	211
Saxophone	626	326
Acoustic guitar	637	535

The classification results were assessed using the precision (P), recall rate (R), and F1 value, calculated by:

$$P = \frac{TP}{TP+FP}, \tag{21}$$

$$R = \frac{TP}{TP+FN}, \tag{22}$$

$$F_1 = \frac{2PR}{P+R}, \tag{23}$$

where:

TP: number of samples retrieved and belonging to the positive category,

FP: number of samples retrieved but belonging to the negative category,

FN: number of samples not retrieved but belonging to the positive category.

In automatic multi-instrument classification, the macroscopic values of P and R need to be considered:

$$P_{marco} = \frac{1}{|L|}\sum_{l=1}^{L} P_l, \tag{24}$$

$$R_{marco} = \frac{1}{|L|}\sum_{l=1}^{L} R_l, \tag{25}$$

where, L refers to the number of labels, P_l and R_l are the corresponding P and R values calculated for each label l. By further calculation, the corresponding macro F1 value can be obtained.

B. Result Analysis

Firstly, the effect of feature selection on the multi-instrument classification results was analyzed. In the case of using the SSA-BPNN model as a classifier but changing the algorithm's feature input, the classification results were compared, as shown in Table II.

TABLE II. EFFECT OF FEATURE SELECTION ON MULTI-INSTRUMENT CLASSIFICATION RESULTS

Musical Instrument		Flute	Organ	Piano	Trumpet	Cello	Clarinet	Electric Guitar	Violin	Saxophone	Acoustic Guitar	**Macro-value**
MFCC	P	0.69	0.61	0.55	0.66	0.69	0.61	0.62	0.63	0.59	0.71	0.64
	R	0.66	0.64	0.65	0.62	0.66	0.71	0.65	0.61	0.71	0.61	0.65
	F1	0.67	0.62	0.60	0.64	0.67	0.66	0.63	0.62	0.64	0.66	0.64
PLPC	P	0.71	0.65	0.66	0.65	0.68	0.65	0.62	0.65	0.67	0.68	0.66
	R	0.65	0.67	0.66	0.64	0.66	0.68	0.71	0.68	0.72	0.66	0.67
	F1	0.68	0.66	0.66	0.64	0.67	0.66	0.66	0.66	0.69	0.67	0.67
MFCC+PLPC	P	0.71	0.65	0.65	0.68	0.74	0.68	0.71	0.61	0.68	0.68	0.68
	R	0.65	0.61	0.65	0.65	0.71	0.71	0.65	0.61	0.75	0.74	0.67
	F1	0.68	0.63	0.65	0.66	0.72	0.69	0.68	0.61	0.71	0.71	0.68
MFCC + PLPC (entropy weight method)	P	0.75	0.66	0.68	0.79	0.81	0.72	0.71	0.64	0.71	0.71	0.72
	R	0.68	0.62	0.68	0.71	0.75	0.75	0.75	0.66	0.78	0.75	0.71
	F1	0.71	0.64	0.68	0.75	0.78	0.73	0.73	0.65	0.74	0.73	0.71

According to the data in Table II, when MFCC was used as the feature, the multi-instrument classification yielded a P value of 0.71, an R value of 0.65, and an F1 value of 0.64. In contrast, when PLPC was employed as the feature, the multi-instrument classification produced a P value of 0.66, which was 0.02 larger than that when using MFCC. The R value was 0.67, which was 0.02 larger than that when using MFCC. The F1 value was 0.67, which was 0.03 larger than that when using MFCC. These results suggested that PLPC was more effective in the context of automatic multi-instrument classification when compared to MFCC. It can be attributed to the fact that PLPC is better at simulating the human ear's perception of sound, making it closer to actual music perception.

Subsequently, both MFCC and PLPC were simultaneously input into the SSA-BPNN model, resulting in a P value of 0.68, which showed an increase of 0.02 compared to inputting PLPC alone. The R value was 0.67, the same as when inputting PLPC alone, and the F1 value reached 0.68, which was an increase of 0.01 compared to inputting PLPC alone. These outcomes suggested that the enhancement in classification performance achieved by inputting both sets of features into the SSA-BPNN model simultaneously was not substantial. It may be because the large number of features affects the algorithm classification performance.

In the final step, the MFCC+PLPC features, selected through the entropy weight method, were employed. This configuration yielded a P value of 0.72, an R value of 0.71, and an F1 value of 0.71. These results surpassed the previous three feature input combinations, underscoring the effectiveness of utilizing the entropy weighting method for feature optimization. This approach enhances the automatic classification capability of the SSA-BPNN method for multi-instrumental instruments while simultaneously reducing the number of feature dimensions.

Then, also using MFCC+PLPC after the entropy weighting method selection as features, the effect of different classifiers on the multi-instrument classification results was compared. The SSA-BPNN algorithm was compared with other classifiers, including:

1) Support vector machine (SVM) [20],

2) XGBoost [21],

3) Traditional BPNN [22],

4) BPNN optimized using genetic algorithms (GA): GA-BPNN [23],

5) BPNN optimized using particle swarm algorithm (PSO): PSO-BPNN [24].

The results are presented in Fig. 4.

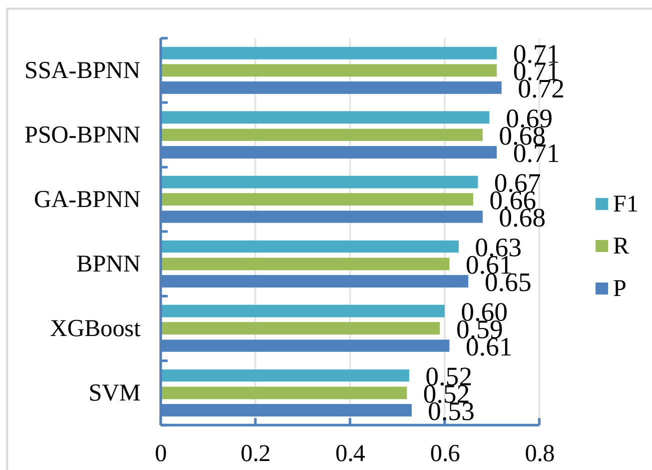

Fig. 4. The effect of the classifier on the automatic multi-instrument classification results.

From Fig. 4, it is evident that the BPNN algorithm exhibited superior classification performance compared to the SVM and XGBoost algorithms. The BPNN algorithm achieved a P value of 0.65, which were improved by 0.12 compared to the SVM algorithm and 0.04 compared to the XGBoost algorithm. The R value for the BPNN algorithm was 0.61, representing a 0.09 increase compared to the SVM algorithm and a 0.02 increase compared to the XGBoost algorithm. The F1 value for the BPNN algorithm was 0.63, indicating an improvement of 0.11 compared to the SVM algorithm and an increase of 0.11 compared to the XGBoost algorithm. These results underscored the advantages of using BPNN as a classifier. In the comparison of different weight threshold optimization methods, the GA-BPNN algorithm showed P value, R value, and F1 value scores below 0.7 for automatic multi-instrument classification, which did not demonstrate substantial improvement over the BPNN algorithm. The PSO-BPNN algorithm achieved only a P value of 0.71, an R value of 0.68, and an F1 value of 0.69. Compared to the PSO-BPNN algorithm, the SSA-BPNN algorithm showed an increase of 0.01 in P value, 0.03 in R value, and 0.02 in F1 value. This result highlighted the reliability of the classifier developed in this paper for automatic multi-instrument classification.

VI. Discussion

The automatic classification of multiple musical instruments is an important area of research in audio signal processing [25], and it also has significant implications for both theoretical and practical applications. Through the automatic classification of multiple musical instruments, it can provide strong support for MIR [26], helping users conveniently find music they are interested in. In music composition, it can be used to automatically separate different instrument tracks, providing flexible post-production techniques for music producers. In cultural preservation, it enables the analysis of multiple instruments in traditional music [27], contributing to better protection and inheritance of musical culture. However, music performance often involves a wide variety of complex instruments, which poses significant challenges for automatic classification. Further research is needed to enhance the effectiveness of multi-instrument automatic classification.

In current research on the automatic classification of multiple musical instruments, improving classification accuracy mainly relies on optimizing feature extraction and classification algorithms. In terms of feature extraction, this paper utilized entropy weighting to optimize MFCC and PLPC features, reducing dimensionality while enhancing accuracy. Regarding the classification algorithm, SSA was employed to optimize BPNN parameters for better performance. The experiments conducted on the IRMAS dataset revealed that the feature selection optimization method, based on entropy weighting, effectively enhanced the accuracy of multi-instrument automatic classification. The obtained P value was 0.72, the R value was 0.71, and the F1 value was 0.71, which were higher than those achieved by other feature combinations. This result indicated that the features selected through entropy weighting could better represent the characteristics of different instruments, thereby improving the discrimination effect of the SSA-BPNN method for various instruments. The analysis of the classifier revealed that the SSA demonstrated excellent performance in optimizing the parameters of BPNN and it could enhance classification accuracy and obtain results better than the other classifiers. This makes it applicable in practical scenarios.

VII. Conclusion

This paper proposed an approach for the automatic classification of multiple instruments. The MFCC and PLPC features were optimized using the entropy weighting method. An SSA-BPNN method was designed as the classifier. Through experiments, it was found that the features optimized by the entropy weighting method delivered optimal performance in the automatic classification of multiple instruments, with a P value of 0.72, an R value of 0.71, and an F1 value of 0.71, outperforming alternative methods like the SVM algorithm.

This method can accurately classify multiple musical instruments and can be further applied in practical music data processing.

The findings of this study demonstrate that the SSA-BPNN method, as designed, can achieve a relatively high level of accuracy in automatically classifying multiple instruments. This provides valuable theoretical insights for optimizing features, refining neural network algorithms, and improving parameter optimization methods. Moreover, it opens up new avenues for future research on automatic classification of multiple instruments. From a practical standpoint, the proposed method holds potential for application in real-world scenarios involving music data processing and even speech data processing.

However, this study also has certain limitations. For instance, the effectiveness of the proposed method has not been validated on a larger dataset, and there is potential for further optimization in feature selection. In future research, it is imperative to conduct additional investigations into feature combinations and optimization methods, explore strategies to enhance the classification accuracy of the model, and examine the applicability of this approach in domains such as artist classification and music genre categorization.

References

[1] S. Rajesh, and N. J. Nalini, “Musical instrument emotion recognition using deep recurrent neural network,” Proc. Comput. Sci., vol. 167, pp. 16-25, January 2020.

[2] C. E. Cella, “Music information retrieval and contemporary classical music: a successful failure,” Trans. Int. Soc. Music Inform. Retriev., vol. 3, pp. 126-136, September 2020.

[3] X. Fu, H. Deng, and J. Hu, “Automatic label calibration for singing annotation using fully convolutional neural network,” IEEJ T. Electr. Electr., vol. 18, pp. 945-952, April 2023.

[4] D. Martin-Gutierrez, G. H. Penaloza, A. Belmonte-Hernandez, and F. Alvarez Garcia, “A multimodal end-to-end deep learning architecture for music popularity prediction,” IEEE Access, vol. 8, pp. 39361-39374, February 2020.

[5] M. Voetter, M. Mayerl, G. Specht, and E. Zangerle, “HSP datasets: insights on song popularity prediction,” Int. J. Semant. Comput., vol. 16, pp. 233-255, June 2022.

[6] S. Rajesh, and N. J. Nalini, “Combined evidence of MFCC and CRP features using machine learning algorithms for singer identification,” Int. J. Pattern Recogn., vol. 35, pp. 2158001.1-2158001.21, July 2020.

[7] S. Kooshan, H. Fard, and R. M. Toroghi, “Singer identification by vocal parts detection and singer classification using LSTM neural networks,” 2019 4th Int. Conf. on Pattern Recognition and Image Analysis (IPRIA), pp. 246-250, March 2019.

[8] W. Feng, J. Liu, T. Li, Z. Yang, and D. Wu, “FAC: a music recommendation model based on fusing audio and chord features (115),” Int. J. Softw. Eng. Know., vol. 32, pp. 1753-1770, October 2022.

[9] A. Elbir and N. Aydin, “Music genre classification and music recommendation by using deep learning,” Electron. Lett., vol. 56, pp. 627-629, June 2020.

[10] D. Chatterjee, A. Dutta, D. Sil, and A. Chandra, “Deep single shot musical instrument identification using scalograms,” 2023 Int. Conf. on

Artificial Intelligence in Information and Communication (ICAIIC), pp. 386-389, August 2023.

[11] A. Wise, A. S. Maida, and A. Kumar, “Attention augmented CNNs for musical instrument identification,” 2021 29th Eur. Signal Process. Conf. (EUSIPCO), pp. 376-380, August 2021.

[12] S. K. Mahanta, A. F. U. R. Khilji, and P. Pakray, “Deep neural network for musical instrument recognition using MFCCs,” Comput. Sist., vol. 25, pp. 351-360, May 2021.

[13] E. Nakamura, Y. Kageyama, and S. Hirose, “LSTM-based japanese speaker identification using an omnidirectional camera and voice information,” IEEJ T. Electr. Electr., vol. 17, pp. 674-684, January 2022.

[14] V. V. Yerigeri, and L. K. Ragha, “Speech stress recognition using semi-eager learning,” Cogn. Syst. Res., vol. 65, pp. 79-97, January 2021.

[15] T. Xu, C. Qin, H. Zhang, Y. Qu, and W. Fang, “Study on petroleum standard attention index calculation based on the entropy weight method,” IOP Conf. Ser.: Earth Environ. Sci., vol. 514, pp. 1-7, July 2020.

[16] Y. Liu, L. Nie, R. Dong, and G. Chen, “BP neural network-Kalman filter fusion method for unmanned aerial vehicle target tracking,” Proc. Inst. Mech. Eng. C: J. Mec. Eng. Sci., vol. 237, pp. 4203-4212, July 2023.

[17] G. F. Fan, Y. Li, X. Y. Zhang, Y. H. Yeh, and W. C. Hong, “Short-term load forecasting based on a generalized regression neural network optimized by an improved sparrow search algorithm using the empirical wavelet decomposition method,” Energy Sci. Eng., vol. 11, pp. 2444-2468, April 2023.

[18] A. Zahid, and M. Arshad, “An innovative design of substitution-boxes using cubic polynomial mapping,” Symmetry, vol. 11, pp. 1-10, March 2019.

[19] J. J. Bosch, J. Janer, F. Fuhrmann, and P. Herrera, “A comparison of sound segregation techniques for predominant instrument recognition in musical audio signals,” Proc. ISMIR, pp. 559-564, January 2012.

[20] F. Camastra, V. Capone, A. Ciaramella, A. Riccio, and A. Staiano, “Prediction of environmental missing data time series by support vector machine regression and correlation dimension estimation,” Environ. Modell. Softw., vol. 150, pp. 1-7, April 2022.

[21] Y. Wu, Q. Zhang, Y. Hu, S. W. Ko, X. Zhang, H. Zhu, J. Liu, and S. Li, “Novel binary logistic regression model based on feature transformation of XGBoost for type 2 Diabetes Mellitus prediction in healthcare systems,” Future Gener. Comp. Sy., vol. 129, pp. 1-12, November 2022.

[22] M. Wei, X. Hu, and H. Yuan, “Residual displacement estimation of the bilinear SDOF systems under the near-fault ground motions using the BP neural network,” Adv. Struct. Eng., vol. 25, pp. 552-571, December 2021.

[23] J. He, X. Li, and Y. Zhao, “The fault diagnosis of diesel fuel supply system based on BP neural network optimized by genetic algorithm,” J. Phys.: Conf. Ser., vol. 1732, pp. 1-6, January 2021.

[24] Y. Wang, and Y. Zhao, “Predicting bedrock depth under asphalt pavement through a data-driven method based on particle swarm optimization-back propagation neural network,” Constr. Build. Mater., vol. 354, pp. 1-16, November 2022.

[25] S. R. Gulhane, S. D. Shirbahadurkar, and S. S. Badhe, “Indian classical musical instrument classification using timbral features,” Concurr. Comp. Pract. E., vol. 33, pp. e6414.1-e6418.19, May 2021.

[26] Y. H. Chen, M. Ceccarelli, and H. S. Yan, “A historical study and mechanical classification of ancient music-playing automata,” Mech. Mach. Theory, vol. 121, pp. 273-285, March 2018.

[27] S. R. Chaudhary and S. N. Kakarwal, “Various approaches in musical instrument identification: a review,” Int. J. Appl. Evol. Comput., vol. 10, pp. 1-7, April 2019.