International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 6880-6891

https://doi.org/10.70517/ijhsa463590

An Analytical Study of Machine Learning Modeling of the Relationship between Clarinet Technique and Symphonic Concerto in a Music Education Setting

Yuchen Wang[1] and Qin Shi[2,*]
[1] Music School and Dance, Chongqing Institute of Foreign Trade and Economics, Chongqing, 400000, China
[2] Chongqing College of International Business and Economics, Chongqing, 400000, China
Corresponding authors: (e-mail: shiqin543578241@sina.cn).

Abstract As an indispensable woodwind instrument in the symphony orchestra, the clarinet, with its wide range and rich timbral variations, is regarded as "the instrument closest to the human voice". In symphonic concerto, the clarinet plays an important role, not only in creating dramatic conflicts and interpreting the characteristics of woodwind instruments, but also in adding special colors to the music. This paper takes the relationship between clarinet technology and symphonic concerto in music education environment as the research object, and constructs a clarinet music classification model based on machine learning. The study adopts the feature extraction method combining Mel Frequency Cepstrum Coefficient (MFCC) and Perceptual Linear Prediction (PLP) to construct a self-constructed symphonic concerto dataset, which contains audio of six instruments, including violin, viola, cello, oboe, clarinet, flute, etc., with a total of 1.25 GB of data, and a playback time of about 7.46 hours. Subsequently, the Improved Particle Swarm Optimization algorithm (IPSO) is proposed to optimize the classification model of Support Vector Machine (SVM) to achieve accurate recognition of clarinet music in symphony concertos. The experimental results show that the MFCC-PLP algorithm is significantly better than other feature extraction methods with the number of times the feature items are selected close to 51 times in both the training set and the test set.The IPSO-SVM model achieves a classification accuracy of 99.09% for the clarinet music, and the mean of the overall music classification correctness is 97.43%, with a classification time of only 1.67 s. This method can be used to recognize the clarinet music in the symphony concerto, which is the most important music in the symphony concerto. The method provides an effective technical support for the intelligent recognition and performance optimization of clarinet music in symphonic concertos.

Index Terms Clarinet, Symphonic concerto, Machine learning, MFCC-PLP, IPSO-SVM, Music classification

I. Introduction

Symphonic orchestra performance, each instrument serves as a different performance work, through the cooperation between each other, interpretation of the connotation of the musical work [1]. The different timbre and expression of various instruments will make the music appear very obvious dramatic conflict, prompting the symphonic performance to be more vivid image and produce a sense of hierarchy [2], [3]. The components of the symphonic orchestra mainly include a variety of musical instruments, such as orchestral instruments, brass instruments, percussion instruments and woodwind instruments, etc. At the same time, in the process of the performance in order to be able to better present the effect of the performance, the orchestra in the original basis of the addition of a number of piano instruments, such as xylophones and harpsichords, etc. [4]-[6]. Among them, the clarinet has a very important position in the symphony because of its uniqueness of tone color and range that other instruments do not have [7]. The clarinet has different forms of musical expression in different registers and special sound quality, which can enhance dramatic conflicts and reflect the ups and downs of musical emotions in symphonic music, increasing the charm and uniqueness of musical performances [8]-[10].

At the same time, the clarinet's rich timbre, high and low sound strength, also provides power for symphonic performances, enriching the musical expression, so that more people feel the charm of the music and the embedded emotions [11]-[13]. As the saying goes, "Sharpening the sword is not a mistake", an excellent performer must have a full understanding of the instrument in his hands in order to bring out the best of both and to integrate with the other instruments of the symphony. Therefore, the performance of symphonic music should strengthen the understanding and application of the relationship between clarinet technology and symphonic concerto, to give full play to the advantages of the clarinet, to increase the infectious force of the music, and to enhance the performance effect.

As an important member of the woodwind family, the clarinet occupies a unique and important position in the symphony orchestra. With its wide tonal range and rich timbral variations, the clarinet is regarded as “the instrument closest to the human voice” by Mozart, and plays an irreplaceable role in the symphonic concerto. From the composer's point of view, the clarinet has the functions of color, discourse, balance and structure in the orchestra, which can effectively create musical dramatic conflicts and interpret the unique mellow and melodious characteristics of woodwind instruments. From the performer's point of view, the clarinet has a powerful musical expression, and is able to take on different “dramatic roles”, providing important character values for the orchestra's performance of the work. From the listener's point of view, the clarinet's presence in the orchestra enriches the color of the music, clearly portrays the musical roles, and makes the work closer to the heart, effectively realizing the communication and resonance between the listener and the music. The clarinet has a long history of development, from the first appearance of the baroque period to the flourishing development of the 18th and 19th centuries, many outstanding musicians for its classic works greatly expanded the expressive power of the clarinet. In the modern music education environment, the clarinet not only occupies an important position in the field of classical music, but also plays a unique role in other music styles such as jazz. However, in the complex acoustic environment of symphonic concerto, how to accurately identify and extract the musical features of the clarinet and scientifically analyze its expression has become an important issue in the field of music education and music information retrieval. Based on the above background, this study proposes a machine learning-based modeling analysis method for clarinet music classification. First, a symphonic concerto dataset containing multiple instruments is constructed, and a feature extraction method combining Mel frequency cepstrum coefficient and perceptual linear prediction is adopted to make full use of the high efficiency of the MFCC and the noise robustness of the PLP to realize the accurate extraction of features of the clarinet music. Secondly, the improved particle swarm optimization algorithm is proposed to optimize the classification model of support vector machine, and through the introduction of adaptive weighting coefficients, it effectively solves the problems of slow convergence speed and easy to fall into local optimum of the traditional particle swarm algorithm, so as to improve the classification accuracy and efficiency. Finally, through detailed experimental verification, the effectiveness and superiority of the method in clarinet music classification is proved, providing scientific basis and technical support for the in-depth study of the relationship between clarinet technology and symphonic concerto in music education environment.

II. Clarinet in Symphonic Concerto

The place and role of the clarinet in the symphony orchestra is indisputable. Facing the same basic point, i.e. the symphony orchestra, the clarinet can be recognized as valuable from at least three perspectives: the composer's, the performer's and the listener's. From the composer's point of view, the clarinet has a color function, a discourse function, a balancing function and a structuring function in the orchestra. From the perspective of the orchestra player, the clarinet has a powerful musical expression and is able to take on different “dramatic roles”, providing important character value to the orchestra's performance of the work. From the listener's point of view, the clarinet's presence in the orchestra enriches the colors, clearly portrays the musical roles, and brings the work closer to the heart, effectively reaching the communication and resonance between the listener and the music.

II. A. Clarinet technique

II. A. 1) Overview of clarinet development

The clarinet, also known as the black pipe, has a long history of development and deep historical roots, its origin can be traced back to ancient Europe. Early clarinets were relatively simple, mostly made of simple wooden tubes. With the passage of time, through the continuous exploration and improvement of countless craftsmen and musicians, the structure of the clarinet was gradually perfected. During the Baroque period, the clarinet began to make its debut on the musical stage. Its unique sound added a different color to the music of that time. In the eighteenth and nineteenth centuries, the clarinet ushered in a golden age of vigorous development. Many outstanding musicians composed classical works for it, greatly expanding the expressive power of the clarinet. With its wide range, rich tonal variations and strong expressive power, the clarinet occupies an important position in the field of classical music. At the same time, it also plays a unique role in other musical styles such as jazz. From court music to the modern stage, the clarinet has witnessed the changes in the history of music and become an indispensable treasure in the world of music [14].

In the vast world of music, the clarinet, with its unique woodwind sound, has become an important tool for expressing emotions and interpreting art. Clarinet playing is not only a test of the player's skills, but also a demonstration of his artistic creativity. From cheerful melody to deep sorrow, from lively section to serious expression, the clarinet can switch freely, showing rich musical expressiveness.

II. A. 2) Clarinet-related classifications

(1) Soprano Clarinet

Often referred to as the soprano bE key clarinet, also known as the small black tube, compared with the ordinary clarinet is shorter, the range is higher than the bB key clarinet four degrees. The tone is loud and bright, and it is a characteristic instrument in symphony orchestras. The fingering is the same as that of the common clarinet, and the pitch control is more difficult, which requires good playing skills and sound control skills.

(2)Soprano Clarinet

There are various kinds of clarinets, among which the bB key clarinet is the most widely used, with a luxurious and bright tone, strong expressive power, giving people a bright, moody and happy feeling. The A key clarinet has a gentle and delicate tone, with a somber atmosphere.

(3) Alto Clarinet

The alto clarinet in the key of bE is mainly used in clarinet orchestras, with a body similar to that of the bass clarinet, and serves as the alto voice in clarinet orchestras, bridging the gap between the soprano and bass voices, and serving to maintain a balance and fill in the harmony between the registers.The Basset Clarinet in the key of F has a sweet tone, which is in line with the tonal requirements of the Classical period, and is almost exclusively used in the performance of Mozart's works.

(4) Bass Clarinet

Bass clarinet in the key of bB. The rapid development of the instrument manufacturing industry during the Romantic period gave rise to the larger bass clarinet, which plays an important role in the performance of symphony orchestras, wind orchestras, and clarinet orchestras. By playing the bass clarinet, a low and thick, dark and sad atmosphere can be rendered.

(5) Double Alto and Double Bass Clarinets

The double alto and double bass clarinets are members of the bass clarinet family, and they mainly appear in a few works of the wind orchestra and clarinet orchestra, assuming the role of thickening the bass group of the orchestra. Especially in clarinet ensembles, the alto and bass clarinets are particularly important in rendering a tense, grave musical atmosphere.

II. A. 3) Clarinet instrument characteristics

(1) Construction and Sound Principle

The clarinet consists mainly of a mouthpiece, a single reed, a body, and a key system. The mouthpiece is where the player blows, and its design affects the airflow. The clarinet reed is the key to the sound, and is a thin, flexible piece of bamboo that is fixed to the mouthpiece. The body of the clarinet is the channel through which the airflow vibrates. Commonly, the body of the clarinet is cylindrical, with different lengths and diameters affecting the pitch. When playing, the player blows through the mouthpiece, and the airflow strikes the clarinet, causing it to vibrate at a high frequency, which in turn generates sound waves. The length of the tube determines the frequency of the sound waves, and a key system can change the length of the air column inside the tube, thus realizing a change in pitch. A long tube produces a low-frequency sound with a deep, mellow tone. A short tube corresponds to a high-frequency tone, which is bright and sharp. The material, thickness and elasticity of the single reed will directly affect the delicacy of the tone and the stability of the pitch. Just as a good paintbrush can draw more delicate lines, high-quality clarinet reeds are crucial to the tone of the clarinet [15].

(2) Tonal characteristics of the clarinet

The timbre characteristics of the clarinet are unique and charming, it has the softness and warmth of a woodwind instrument and the brightness and penetration of a metal instrument, and this timbre characteristic makes the clarinet play an irreplaceable role in the symphony orchestra. Its tone is warm and full, just like the first ray of sunshine in the morning sun, sprinkling the earth through the mist, bringing people endless hope and beauty. In classical symphonic works, the tone of the clarinet is often used to depict a peaceful and beautiful scene. In modern symphonies, the clarinet's tone is more often used to show complex and changing emotions. Composers use different playing techniques and timbral treatments to give the clarinet a unique and expressive sound in the orchestra, bringing a new musical experience to the listener. In addition, the clarinet's timbre has a very good integration, and can harmonize with other instruments to create a rich and colorful world of music. Whether it is playing with string instruments or with brass instruments, the clarinet can show its unique charm and add endless colors to the whole orchestra. The tone of the clarinet is like the brushstrokes of a poet, able to depict the most delicate and profound emotional world, allowing the listener to be immersed in the ocean of music, intoxicated.

II. B. Clarinet in the symphony

II. B. 1) Clarinet role performance

(1) Creating musical dramatic conflicts. From the symphony as a whole, all the instruments in the orchestra have their own functions, but also by virtue of their own characteristics of the interpretation of the musical works of the various parts of the use of multiple timbres of the interweaving of the music to enhance the dramatic conflict, but also to make the whole music has a storytelling, expressing a diverse range of emotional changes. Clarinet has a unique sound quality in all registers, which can bring out the music drama conflict effectively when cooperating with other instruments in the orchestra.

(2) Interpretation of woodwind characteristics. The clarinet has a very distinctive shape, and when played independently, it can bring out the unique mellowness and melodiousness of woodwinds, and make the whole melody sound like a sob. It can be said that it is because of the clarinet, so that many of the composer's artistic ideas can be cleverly expressed and greatly enhance the color.

(3) Add musical expression to the instrument. Symphony orchestra in the process of playing music through the instrument's own tone to present the melody of the music, but also through the different instruments between the strong and weak, high and low, fast and slow differences to increase the richness of the music, so that it presents a unique meaning and “expression” characteristics.

(4) Adding special colors to the music. Usually in the process of symphonic performance, performers want to create and present different environments or special characters when they will use different musical colors and means to make special expression. Clarinet through the use of its different scales and tones, can achieve a good narrative, so that the story has a richer plot and color.

II. B. 2) Clarinet status performance

The clarinet embodies a strong sense of place in the symphony due to its timbre and range, which is very different from that of other instruments, and plays an important role in the symphony. The clarinet's timbre is closer to the human voice, reflecting a strong sense of character and narrative. The musician Mozart spoke highly of the clarinet and called it “the instrument closest to the human voice”. The main reason why it is called the “orator” of the symphony orchestra is that the story behind it is more moving, and it has three scales of high, middle and low notes, which can reach a span of four octaves. The clarinet, which joined the symphony orchestra later, was introduced to everyone's view, and not only has a wider variety, but also a wider range.

In the development of the symphony orchestra, not only is the range wider, but also increases the dramatic expression and improves the acoustic effect. Categorized according to the soprano clarinet, it can be planned as the key of D, C, A, B-flat. According to the alto clarinet implementation classification can be divided into E-flat, F key, according to the bass clarinet implementation classification can be divided into B key. In practice, in general, the more frequent use of B-flat, E-flat and A-flat and other integral types. In essence, when using the soprano clarinet, the main purpose is to present the bright tones in it to everyone, mainly to express the grand scenes of heroes in general. The alto part brings a pure and beautiful tone to the public, and when used, is applied to more mournful and melancholic music. The bass part will give people a more full-bodied feeling, not only reflecting the mysterious and dramatic characteristics, but also will express the timbre and volume realistically.

III. A model for categorizing clarinet music in symphonic concertos

Clarinet is an important member of the wind instrument family that cannot be ignored, and at the same time, the clarinet itself also has a very strong musical expression, and the clarinet has its uniqueness. Whether in terms of playing technique or in terms of timbre characteristics, it can make the audience feel the unique artistic charm and artistic value. For the clarinet's range is analyzed, itself is relatively wide, in the current symphony orchestra, the clarinet needs to be further strengthened research.

III. A. Symphony concerto dataset

III. A. 1) Symphony Data Acquisition

This dataset recruited multidimensional players from a symphony orchestra to record symphony concerto music data. There are three clarinet players, the first player has a doctorate in clarinet performance, the second player is a graduate clarinet student enrolled in a conservatory, and the second player is a graduate engineering student at a university who has been studying clarinet for 18 years. In order to cover the diversity of clarinet expression, the dataset contains a number of different types of instruments, mainly violin, viola, cello, oboe, clarinet, flute, etc. The dataset also contains a number of other instruments, such as the clarinet, viola, cello, oboe, clarinet, and flute. , i.e., string, woodwind, brass, and percussion instruments included in symphonic music concertos, the recordings of symphonic concertos were collected over a period of 15 to 25 days.

The dataset contains raw audio clips, music segments sliced according to the score, sentiment scores for each track based on a Valence Arousal (VA) model, and a set of feature data designed for categorical identification, totaling 1.25 GB in size and approximately 7.46 hours of playback time. The selected musical pieces cover a wide range of instrumental emotional spectrums, and the recorded tracks are taken from orchestral concerto related textbooks, with music recorded in a quiet environment using a cell phone or professional equipment.

III. A. 2) Music data preprocessing

For the constructed symphony concerto music classification dataset, some data processing work is needed to facilitate the subsequent training and evaluation of the music classification model. The specific data processing steps are as follows:

(1) Data cleaning. The original data need to be cleaned first to remove duplicated, missing and unqualified data to ensure the quality and accuracy of the data.

(2) Data balancing. In order to ensure the training effect of the model, it is necessary to ensure that the dataset is balanced in terms of categories, i.e., the number of samples in each category should be the same or similar.

(3) Data preprocessing. For audio data, audio signal processing is required. Convert the sample rate from the original sample rate to the desired sample rate to be compatible with different processors and devices. Eliminate noise and clutter to improve the quality and clarity of the audio signal. Enhance bass and treble to improve the quality and clarity of audio signals. Make the volume of different audio signals the same by adjusting the volume so that they can be easily compared and analyzed. Remove the muted segments of the audio signal to reduce the amount of data and improve the processing efficiency, and avoid segmentation to keep the key information from being lost.

(4) Feature extraction. In this paper, the audio feature extraction using Mel's inverted spectral coefficient and linear perception, which will be described in detail later, is required to extract the corresponding multi-dimensional base features for each song.

(5) Data division. The dataset is divided into two parts: training set and test set, and this chapter adopts 70% of the data as the training set and 30% of the data as the test set to avoid data leakage.

III. B. Clarinet music classification model construction

III. B. 1) Clarinet music feature extraction

(1) Mel frequency cepstrum coefficient

A symphonic concerto can be divided into many frames, and each frame of speech can correspond to a spectrum after the fast Fourier transform calculation, reflecting the relationship between frequency and energy, i.e., the amplitude of different frequencies is different in size. The sound spectrogram has to reflect the relationship between all frequencies and energies, and the correspondence often comes out to be a very large one. In order to obtain sound features suitable for the final music classification, it is usually necessary to input them into a Mayer scale filter bank and transform them into a Mayer spectrum [16]. After transforming an ordinary frequency scale into a Mel frequency scale, the mapping relation can be expressed as:

$$Mel(f) = 2595 \cdot \log_{10}\left(1 + f / 700\right) \tag{1}$$

where f denotes the frequency (in Hz), and the specific calculation process is shown in Fig. 1.

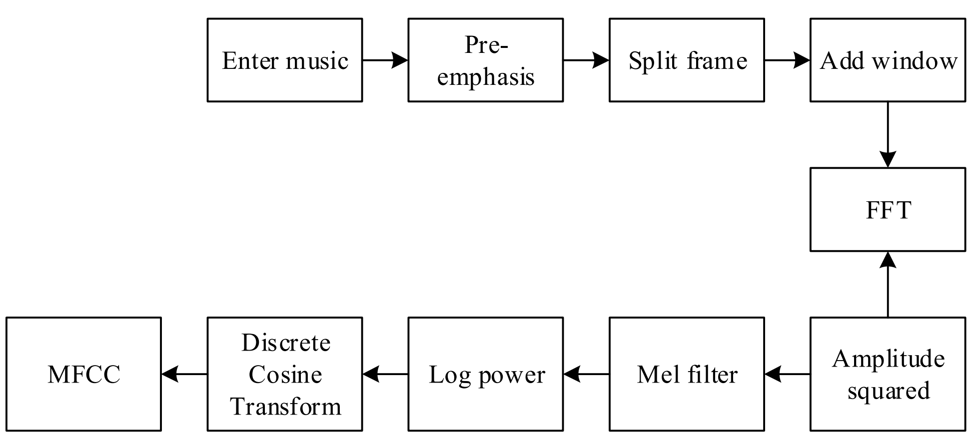


Figure 1: Music emotion recognition workflow

(2) Perceptual linear prediction

Clarinet music feature extraction in a symphonic concerto is not a simple splicing of frame-to-frame information, so we need to incorporate features that indicate that the cepstral coefficients are changing over time, i.e., when making the inherently static waveform graph move. We call Δ^*, Δ^{2*} the Δ feature, that is, the dynamic feature is derived from time, and the cepstrum feature Δ feature is calculated as follows:

$$d(t) = \frac{c(t+1) - c(t-1)}{2} \tag{2}$$

where $c(t)$ is the dynamic characteristics of music at moment t.

The dynamic characteristics of music information can be reflected to a certain extent by the way of derivation.

Compared with spectral features, MFCC is less correlated and easier to model. However, MFCC is not robust against noise, and it happens that PLP has better noise robustness compared to MFCC, so this paper combines two acoustic features as feature input. The combination of the two acoustic features as input is one of the innovations of this paper, which ensures that while the MFCC features are extracted to achieve the effective extraction of audio features, the PLP features are used to increase the noise robustness, which can optimize the extraction of the features of clarinet music.

III. B. 2) IPSO algorithm and SVM algorithm

(1) Improved Particle Swarm Optimization (IPSO) algorithm

The core of the particle swarm method is to do random initialization of some particles, treating all particles as feasible solutions to the problem, defining the particle velocity to ensure its normal movement in space. The velocity affects the particle movement orientation and distance, and the particle performance is judged to be good or bad by the fitness function, and then the best solution is obtained through repeated iterations. During the iteration process, the particle will follow the two solutions of individual optimal and population optimal, and control its own behavior by combining its own experience and that of the population [17].

Set up a population of size N' in a D -dimensional search space, with all particles moving at different speeds. Suppose the position and velocity vectors of particle i' in this space are denoted as $x_{i'} = (x_{i'1}, x_{i'2}, \cdots, x_{i'd}, \cdots, x_{i'D})$ and $v_{i'} = (v_{i'1}, v_{i'2}, \cdots, v_{i'd}, \cdots, v_{i'D})$, the current stage of adaptation is obtained by some adaptation value function. The best individual position found by the current particle is doing $p_{i'} = (p_{i'1}, p_{i'2}, \cdots, p_{i'd}, \cdots, p_{i'D})$, which is called p_{best}, and the best value of the individual opposite to it is f_{pbest}, which is obtained by the population search as $p_g = (p_{g1}, p_{g2}, \cdots, p_{gd}, \cdots, p_{gD})$, called g_{best}, with its relative population optimum being f_{gbest}.

During the flight, the particle velocity and position are continuously updated, and the process is expressed using the following equation:

$$v_{i'd}^{t+1} = v_{i'd}^{t} + c_1 r_1 (p_{best}^{t} - x_{i'd}^{t}) + c_2 r_2 (g_{best}^{t} - x_{gd}^{t}) \tag{3}$$

$$x_{i'd}^{t+1} = x_{i'd}^{t} + v_{i'd}^{t+1} \tag{4}$$

In the formula, t represents the number of iterations, c_1 and c_2 belong to the learning factor, and r_1, r_2 represent random numbers.

There are three parts in the particle swarm velocity update expression, the first is the particle swarm memory, which represents the inertial motion with its own velocity, the second is the particle cognitive module, which represents the repeated analysis of its own experience, and reflects the degree of influence of the individual best on the current particle, and the last is the “social” part, which characterizes the form of cooperation between the particles. During the whole search period, the particles save their own experience and also refer to the experience of their peers, constantly adjusting to achieve the optimal state.

The traditional particle swarm algorithm is relatively simple and low computational complexity, but the disadvantages are also more obvious, the convergence speed is slow, and it is easy to fall into the local optimal solution.

From the speed update formula of particle swarm algorithm, it can be seen that the speed of particle swarm particles is mainly determined by the individual cognitive factor c_1, social cognitive factor c_2 and inertia weight ω. It can be seen that when c_1 is larger, the individual cognition accounts for a larger proportion, which is more favorable to the global search, and the social cognition accounts for a larger proportion when c_2 is larger, which is more favorable to the local search. Therefore, it can be seen that it is not possible to achieve the effect of balancing the global search and local search by fixing the weight coefficients, so the introduction of a kind of adaptive weight coefficients related to the number of iterations can be a good solution to this problem. In this paper, the adaptive part of the parameters is realized by using a nonlinear function related to the exponential function. The inertia weights ω are updated as follows:

$$\omega = \omega_{\max} - \alpha e^{\frac{g}{G}} \tag{5}$$

where $\omega_{\max}$ is the maximum value of the global inertia weight coefficient, α is the influence factor of the iteration number function, g is the current iteration number, and G denotes the total global iteration number.

(2) Support Vector Machine (SVM)

Support Vector Machine (SVM) is a kind of machine learning classifier with strong generalization ability, which is especially suitable for obtaining the global optimal solution when the number of training samples is relatively small. In support vector machines, for low-dimensional spaces, it is generally not possible to describe the classification surface with a simple function due to the complexity of the sample distribution. And by mapping the low-dimensional space to the high-dimensional space, it is easy to find the optimal classification hyperplane between the training samples in the high-dimensional space, thus improving the accuracy and efficiency of classification [18].

In the process of data conversion from low-dimensional to high-dimensional, if we set the amount of original data as n, the amount of transformed data as d, and the dimension of feature space required for the conversion is $\binom{n+d-1}{d}$. It can be seen that even though the values of n and d are not large, the dimension of the feature space can become very large, which makes the computation difficult or even impossible, which is called dimensionality catastrophe.

Although the "dimensional catastrophe" problem is encountered in the dimension upgrading process, the kernel function $K(x_i, x_j)$ in the high-dimensional eigenspace can effectively overcome this problem and construct a nonlinear decision function as in the input space:

$$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) \tag{6}$$

$$f(x) = \operatorname{sgn}\left\{ \sum_{i=1}^{n} \alpha_i^* y_i K(x_i \cdot x) + b^* \right\} \tag{7}$$

Eq. (6) is the objective function of SVM and Eq. (2) is the decision function of SVM for predicting the class of new input data x. Where α_i is the kernel function parameter in SVM, y_i is the label of the i th training sample, which takes the value of +1 or -1, and $K(x_i, x_j)$ is the kernel function, which measures the similarity between the i th and j th training samples. $f(x)$ is the classification result for the new input data x, α_i^* is the optimized kernel function parameter, $K(x_i, x)$ calculates the similarity between the new input data x and all the training samples x_i, and b^* is the bias term, which is used to adjust the threshold of the classifier.

The kernel function is usually given directly, so it is crucial to choose a suitable kernel function, and there are three main types of kernel functions:

The polynomial (poly) kernel function can be expressed as:

$$K\left(x, x_i\right) = \left(x \cdot x_i + 1\right)^d \tag{8}$$

where d denotes the number of polynomials, usually a positive integer, and x and x_i denote the current input data and a piece of training data in the training set, respectively.

The radial basis (RBF) kernel function can be expressed as:

$$K\left(x, x_i\right) = \exp\left(-\frac{\left\|x - x_i\right\|^2}{2\sigma^2}\right) \tag{9}$$

where σ is a scalar parameter that controls the width of the Gaussian kernel, and $\left\|x - x_i\right\|$ denotes the Euclidean distance between x and x_i.

The two-layer feedforward neural network (Sigmoid) kernel function can be expressed as:

$$K(x,x_i)=\tanh\left(v(x-x_i)+c\right) \tag{10}$$

where v is a scalar parameter controlling the slope of the linear part, c is another scalar parameter controlling the center position of the hyperbolic tangent function, and $x \cdot x_i$ denotes the inner product of x and x_i.

There is no clear criterion for the selection of the kernel function, which is usually based on experience. Although the support vector machines constructed with different kernel functions may be similar in results, the choice of kernel function still has some influence on the final results.

III. B. 3) Clarinet music classification process

The penalty factor and kernel function parameters have a large impact on the SVM prediction accuracy, so this paper adopts the IPSO algorithm to optimize the penalty factor and kernel function parameters in the SVM, and proposes an optimization algorithm based on the Improved Particle Swarm Algorithm for Support Vector Machines (IPSO-SVM), and the steps of the algorithm are as follows:

Step1 Input the training samples containing clarinet music features extracted with MFCC-PLP algorithm.

Step2 Initialize the penalty factor and kernel function parameters of SVM.

Step3 Initialize the position and velocity of the population, and use the accuracy derived from SVM as the fitness function of PSO algorithm.

Step4 Update the particles and calculate the fitness function of the updated particles.

Step5 Re-initialize the particles with a certain probability.

Step6 Determine whether the individual extreme value of the current particle is the global optimal solution, if so, assign the current individual extreme value to the global optimal solution, if not, Step4.

Step7 Use the optimized parameters for SVM audio signal classifier for training, and test it with signal test set.

IV. Validation of the classification of clarinet music in symphonic concertos

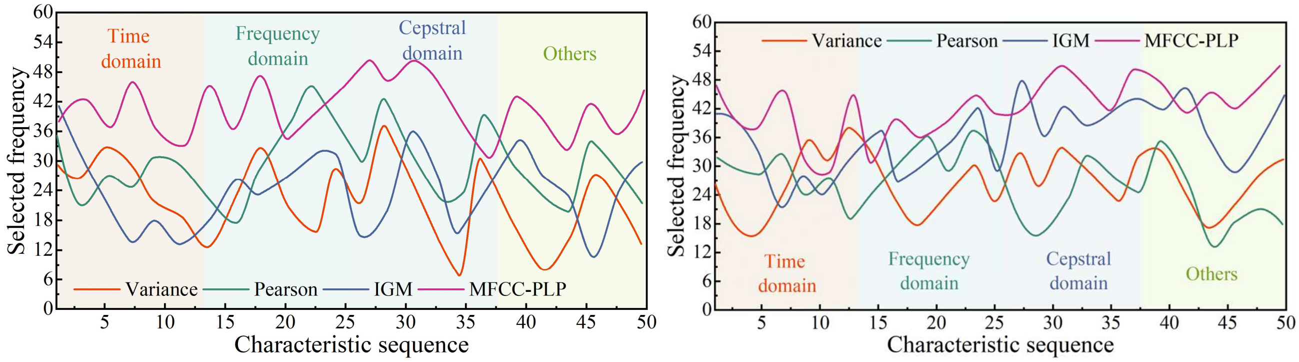
The clarinet is a Western wind instrument, belonging to the woodwind group, which has always played an important role in the development of music in the West. The clarinet is widely used in symphony orchestras and military bands, where it plays the melody of the voice. The clarinet's technique is rich and its tone is very moving, so it is loved by music lovers all over the world. In the current music education environment, how to achieve the accurate extraction of clarinet timbre from symphonic concertos provides reliable support for optimizing the effect of symphonic concertos.

IV. A. Feature Extraction and Parameter Optimization Verification

IV. A. 1) Clarinet audio feature extraction

The operating system selected is Windows 10, Python is selected for the audio processing library, and the deep learning framework is TensorFlow. The CPU is selected to be Intel Core i5 3.2 GHz with dual-core and four-threads, and 32 GB of RAM. the self-constructed symphonic concerto dataset from the previous section is selected as the data source for the test, which is collected through the performance of symphonic orchestra after the The relevant symphony dataset contains a variety of different types of audio samples, which cover a variety of different instrument types, such as violin, viola, cello, oboe, clarinet, flute, etc. The data source of the test is the self-constructed symphony concerto dataset, which is collected after the performance by a symphony orchestra. This dataset is divided into training set and test set in the ratio of 7:3. The test is conducted for the extraction of clarinet audio feature parameters in symphonic music, and the test results of clarinet audio feature extraction in symphonic music are shown in Fig. 2 in terms of the number of times the feature parameter items are selected as an index, comparing the variance method, Pearson's correlation coefficient method, the information gain method (IGM), and the method proposed in this paper (MFCC-PLP), respectively. Where Fig. 2(a)~(b) shows the extraction results of the training set and test set, respectively.

As can be seen from the figure, there are obvious fluctuations in the performance results of different types of clarinet audio feature extraction methods in different audio feature terms in the training set. Especially in the cepstrum domain features, the feature term of this paper combining the Mel cepstrum coefficients and the perceptual linear prediction algorithm is selected up to nearly 51 times, while the feature term of the variance method is selected up to about 38 times. In addition, the changes in the number of selections of different methods in the test set are obviously more stable than the training set, and the performance of the MFCC-PLP algorithm proposed in this paper is still optimal in extracting the audio features of the clarinet in the symphony concerto, when the maximum number of times of the feature items being selected is also close to 51 times. As a result, the combination of Mel's inverted spectral coefficients and the perceptual linear prediction algorithm can be utilized to increase the noise robustness of the PLP algorithm, which can better ensure the extraction of the musical features of the clarinet.



(a) Training set (b) Test set

Figure 2: Performance testing of feature extraction methods

IV. A. 2) Optimization of IPSO algorithm parameters

In carrying out the classification and recognition of clarinet music in symphonic concertos, this paper uses the IPSO algorithm to optimize the penalty factor and kernel parameters of the SVM model, so as to ensure the accurate classification of clarinet music. There are several classifiers included in the SVM classification model, and the penalty parameters and kernel parameters of the classifiers need to be optimized using the IPSO algorithm. First, set the population size of the IPSO algorithm to 30, initialize the position and speed of the particles, the maximum number of iterations is 200, and set the local and global acceleration factor coefficients to 1.45 and 1.65, respectively.Using the ten-fold cross-validation method, the average cross-validation classification recognition rate of the ten iterations is obtained after ten iterations, which improves the reliability of the algorithm's performance, and reduces the kernel function parameter error. Figure 3 shows the particle swarm adaptation convergence curve of the optimized SVM model of the IPSO algorithm.

As shown in the figure, with the increase of the number of evolution, the average fitness is between 70% and 97%, which verifies the optimization performance of the IPSO algorithm and illustrates the effect of the change of the model's kernel parameter on the accuracy of the classification model. The combinations of optimal penalty parameters and kernel parameters of the SVM model obtained by this experiment through the optimization search of the IPSO algorithm are {5.207,1.038}, {3.262,0.647}, and {6.279,1.021}, respectively. In addition, the relevant parameters of the SVM model optimized by the IPSO algorithm proposed in this paper have more obvious time advantages with higher classification accuracy. The accuracy and fastness of the SVM model classifier optimized by the IPSO algorithm are verified. Considering that there are more different types of timbres in the symphonic concerto, the optimization method page is also applicable to the occasions with a large number of categories and higher real-time requirements for classification, i.e., it can meet the demand of extracting the clarinet audio timbres from the symphonic concerto.

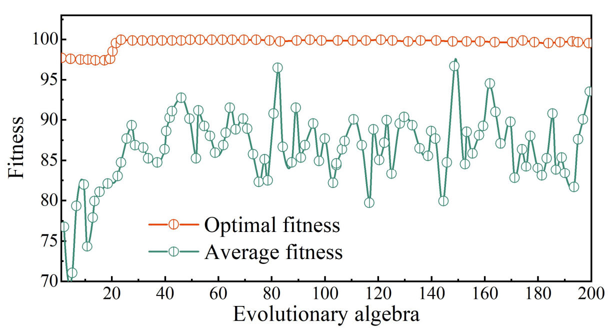


Figure 3: Fitness convergence curve

IV. B. Clarinet music classification model validation

IV. B. 1) Clarinet music classification results

In this paper, the self-constructed symphonic concerto dataset contains different instruments such as violin, viola, cello, oboe, clarinet, and flute. The IPSO-SVM model is used to classify the audio of the instruments in the symphonic concerto as an illustration of the effectiveness of the classification of clarinet music. The results of the symphonic concerto music classification are shown in Table 1.

From the table, it can be seen that the improved particle swarm optimization algorithm and the support vector machine of the symphony concerto music classification method has an average correct rate of 97.43%, and relatively speaking, the classification accuracy of the clarinet music features is the highest 99.09%. It shows that this method is an effective music feature classification method, which can accurately recognize different instrumental audio (especially clarinet audio) in symphonic concerto.

Table 1: Classification results of music signals

Type	Sample quantity	Correct classification number	Accuracy/%
Violin	240	237	98.75
Viola	180	174	96.67
Cello	120	112	93.33
Oboe	160	158	98.75
Clarinet	220	218	99.09
Flute	200	196	98.00
Means	186.7	182.5	97.43

On this basis, in order to test the superiority of the clarinet music feature classification model based on the IPSO-SVM algorithm, the same symphony concerto dataset is used to classify in the same environment, and SVM and BP neural network are used as a comparison to obtain the correct rate of the music classification of the different methods, as shown in Figure 4. From the data distribution in the figure, it can be seen that the correct rate of clarinet music signal classification of improved particle swarm optimization algorithm and support vector machine is higher than that of the standard support vector machine and BP neural network, and the probability of classification error of clarinet music signal has been effectively suppressed, and the ideal classification results have been obtained, which can be used as a reference for the optimization of the performance of clarinet music in the symphony concerto.

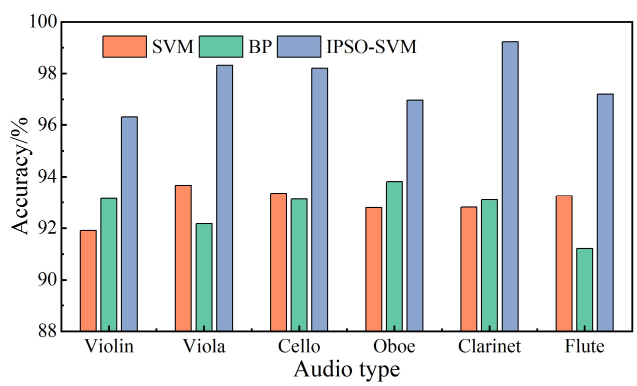


Figure 4: The accuracy rate of music classification by different methods

Table 2: Comparison of music signal classification time (s)

Type	IPSO-SVM	SVM	BP
Violin	1.79	7.32	6.12
Viola	2.14	8.43	6.34
Cello	2.06	8.65	6.06
Oboe	1.85	8.27	5.41
Clarinet	1.33	8.38	5.29
Flute	1.67	8.51	5.76
Means	1.81	8.26	5.83

In addition, in this paper, the music classification time of the improved particle swarm optimization algorithm and support vector machine for clarinet music classification method and the classical method were counted during the testing process, and the results of their classification time are shown in Table 2. As can be seen from the table, the average value of the music signal classification time of the improved particle swarm optimization algorithm and support vector machine is 1.81 s, which is 78.09% and 68.95% lower than the average value of the music signal classification time of the standard support vector machine and BP neural network, respectively. The classification time for clarinet music is only 1.67, which is lower than the average classification time, firstly, this paper focuses on the extraction method of clarinet audio features, and secondly, it is because the sample data in the dataset is less, but it also shows the feasibility of the model in clarinet music classification. In summary, the method in this paper has the least classification time and obtains more efficient results for clarinet music signal classification.

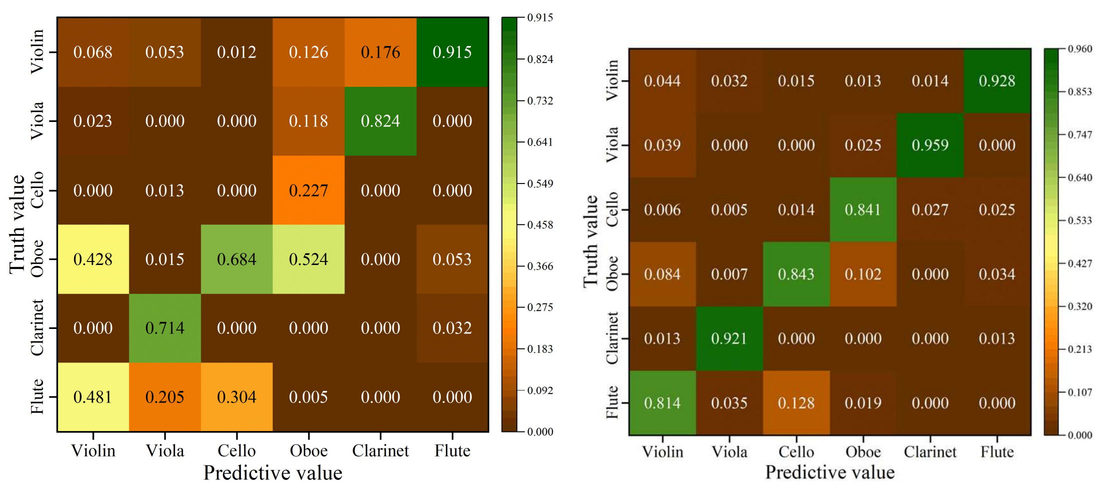
IV. B. 2) Confusion matrix for audio classification

Confusion matrix is a visualization tool that can provide easily observable experimental results, so in this paper, we use the confusion matrix to represent the results of various recognition classification methods for musical instrument kinds. In the confusion matrix, the rows represent the labels of the real musical instrument species, the columns represent the labels of the predicted musical instrument species, and the numbers in the confusion matrix represent the ratio of the number of samples of labels recognized as the predicted musical instrument species to the number of samples of labels of the real musical instrument species. The bolded value on the diagonal line is the ratio of the number of samples of labels identified as predicted musical instrument categories to the number of samples of labels of real musical instrument categories when correctly categorized, and the larger the number on the diagonal line, the higher the accuracy of the musical instrument category identification.

In order to verify the effectiveness of the IPSO algorithm proposed in this paper, the Gaussian white noise symphonic concerto with the addition of SNR are 15dB are feature extracted, respectively, using SVM and IPSO-SVM for classification and recognition, and get the average classification and recognition rate of the ten-fold cross-validation of the confusion matrix is shown in Figure 5. Where Fig. 5(a)~(b) shows the classification confusion matrices of SVM and IPSO-SVM, respectively.

From Fig. 5(a), it can be seen that the average classification recognition accuracy of the SVM model for both flute and clarinet is greater than 80%, and the classification accuracy of the SVM model for flute reaches 91.5%. The SVM model has the lowest average classification recognition rate for oboe, with a classification accuracy of only 22.7%, and the SVM model has the highest rate of misclassification of oboe as a cello, with a probability of 52.4% for both. This shows that the SVM model has good classification recognition ability for flute and clarinet, and poor recognition ability for oboe.

From Fig. 5(b), it can be seen that the classification recognition accuracy of the IPSO-SVM model for each instrument in the symphony concerto dataset is higher than 80%, with the classification accuracy for the clarinet instrument reaching 95.9%. In contrast, the IPSO-SVM model is good at correctly classifying all types of instruments in the symphony concerto, and is especially suitable for recognizing and classifying the audio signals of woodwind instruments.



(a) SVM (b) IPSO-SVM

Figure 5: Confusion matrix for audio classification

V. Conclusion

By constructing a clarinet music classification model based on IPSO-SVM, the accurate identification and classification of clarinet music in symphony concertos is successfully realized.The MFCC-PLP feature extraction method shows significant advantages in clarinet audio feature extraction, and its noise robustness improvement effectively guarantees the accuracy of feature extraction. The experimental results show that the overall classification accuracy of the method for the six instruments in the symphonic concerto reaches 97.43%, among which the classification accuracy of clarinet music is as high as 99.09%, and the classification error rate is effectively suppressed. In terms of classification efficiency, the average classification time of the IPSO-SVM method is only 1.81 seconds, which is 78.09% and 68.95% lower than that of the traditional SVM method and BP neural network, respectively, and significantly improves the classification efficiency. Confusion matrix analysis further validates the reliability of the method, especially in the recognition of woodwind instruments, where the IPSO-SVM model achieves a recognition accuracy of 95.9% for the clarinet. The study not only provides a scientific quantitative tool for the technical analysis of clarinet in the music education environment, but also provides technical support for the optimization of symphonic concerto effects. The proposed method has good real-time and accuracy, which is suitable for applications with a large number of categories and high real-time requirements for classification, and provides an important theoretical foundation and practical guidance for the development of music information retrieval, music education intelligence, and computer musicology and other related fields.

References

[1] Yoon, H., & Ra, J. (2024). The First Appearance of the Clarinet in the Mannheim Orchestra and Symphonies. International Review of the Aesthetics and Sociology of Music, 55(2), 219-238.

[2] Jia, M., Zhao, L., Li, G., & Zheng, Y. (2025, April). Orchestrating the Symphony of Prompt Distribution Learning for Human-Object Interaction Detection. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 4, pp. 3940-3948).

[3] Grachten, M., Cancino-Chacón, C. E., Gadermaier, T., & Widmer, G. (2017). Toward computer-assisted understanding of dynamics in symphonic music. IEEE MultiMedia, 24(1), 36-46.

[4] Bibu, N., Brancu, L., & Teohari, G. A. (2018). Managing a Symphony Orchestra in Times of Change: Behind the Curtains. Procedia-Social and Behavioral Sciences, 238, 507-516.

[5] Pellauer, M., Clemons, J., Balaji, V., Crago, N., Jaleel, A., Lee, D., ... & Emer, J. S. (2023). Symphony: Orchestrating sparse and dense tensors with hierarchical heterogeneous processing. ACM Transactions on Computer Systems, 41(1-4), 1-30.

[6] Marchenko, V., Yunjia, C., Yitong, L., Antonyuk, I., & Khovpun, O. S. (2021). Symphonic music in contemporary theatre. Linguistics and Culture Review, 162-170.

[7] Chernaya, M., & Zhao, Y. (2019). Peculiarities of Clarinet Concertos Form-Building in the Second Half of the 20th Century and the Beginning of the 21st Century. Agathos, 10(1), 177-188.

[8] Le, D. V. T., Giraud, M., Levé, F., & Maccarini, F. (2022, July). A corpus describing orchestral texture in first movements of classical and early-romantic symphonies. In Proceedings of the 9th International Conference on Digital Libraries for Musicology (pp. 27-35).

[9] Hao, Z. (2022). Ensemble and orchestral types of music-making by clarinetists as pedagogical technologies. Revista on line de Política e Gestão Educacional, e022016-e022016.

[10] Pesic, P. (2025). The Horn of Enlightenment: Mozart's Operatic Use of the Clarinet. Cambridge Opera Journal, 1-28.

[11] Weiss, A. E., Nusseck, M., & Spahn, C. (2018). Motion types of ancillary gestures in clarinet playing and their influence on the perception of musical performance. Journal of new Music research, 47(2), 129-142.

[12] Wang, Y., & Shi, Q. (2025). A Computational Simulation Study of the Relationship Between Clarinet Technique and Symphonic Concerto in a Music Education Environment. J. COMBIN. MATH. COMBIN. COMPUT, 127, 6797-6809.

[13] Hanudel, P. (2020). WEBER: Symphonies; Clarinet Concertos. American Record Guide, 83(5), 110-112.

[14] Yoon Hyunjoo & Ra Julie. (2024). The First Appearance of the Clarinet in the Mannheim Orchestra and Symphonies:Its Historical Significance and Effects. International Review of the Aesthetics and Sociology of Music,55(2),219-238.

[15] Pesic Peter. (2024). The Horn of Enlightenment: Mozart's Operatic Use of the Clarinet. Cambridge Opera Journal,36(3),267-294.

[16] Ricky K.W. Chan & Bruce X. Wang. (2024). Do long-term acoustic-phonetic features and mel-frequency cepstral coefficients provide complementary speaker-specific information for forensic voice comparison?. Forensic Science International,363,112199-112199.

[17] Dong Liangwei & Hu Yueli. (2024). Self-tuning control of steam sterilizer temperature based on fuzzy PID and IPSO algorithm. Journal of Measurements in Engineering,12(4),638-655.

[18] Mengyang Zhang,Xuyinglong Zhang,Shan Gao & Yujie Zhu. (2023). Comfort Study of General Aviation Pilot Seats Based on Improved Particle Swam Algorithm (IPSO) and Support Vector Machine Regression (SVR). Applied Sciences,13(15),9038-9038.