ISSN: 2723-9535

Available online at www.HighTechJournal.org

HighTech and Innovation Journal

Vol. 6, No. 4, December, 2025

Improved Skyline-BP Network for Multi-Track MIDI Music Melody Extraction and Style Classification

Junyi Han [1*]

[1] *Institute of Education, Don State Technical University, Rostov-on-Don, 344003, Russia.*

Received 29 August 2025; Revised 17 November 2025; Accepted 22 November 2025; Published 01 December 2025

Abstract

With the rapid development of the digital music industry, core challenges have emerged concerning the insufficient accuracy of main melody extraction and the poor style classification effect of multi-track MIDI files. To address these issues, this study proposes a novel model based on an improved Skyline algorithm and an optimized BP neural network. The method first standardizes MIDI data into a Time-Pitch-Intensity feature matrix. An improved Skyline algorithm is then used to integrate pitch saliency calculation with temporal continuity screening, enhancing the anti-interference ability for multi-track melodies. For music style classification, an optimized BP network with Adaptive Moment Estimation (Adam) gradient optimization and Residual Connection (ResConnect) is designed to improve learning efficiency and accuracy. Experimental results demonstrated that the proposed model surpassed comparative models in overall performance, with a classical-style main melody extraction accuracy of 94.6% and a 2-track separation accuracy of 95.2%. The experiments were benchmarked on the Lakh MIDI Dataset and MuseScore MIDI Library. The model also exhibits superior robustness against noise interference and faster convergence speed. This study provides reliable technical support for applications like music creation assistance and copyright retrieval.

Keywords: Multi-Track MIDI; Skyline Algorithm; BP Neural Network; Main Melody Extraction; Music Classification.

1. Introduction

With the continuous evolution of digital music technology, the Musical Instrument Digital Interface (MIDI) format has become a core data carrier in music creation, film and television music, and game sound design due to its structured instruction encoding, high editing flexibility, and low storage footprint [1]. The digital music industry is moving towards intelligence and refinement, and the efficient processing of massive multi-track MIDI files has become a key indicator for measuring the level of digital music intelligence [2]. However, this development faces two core technical challenges: the insufficient accuracy of Main Melody Extraction (MME) in multi-track MIDI and the poor performance of Music Style Classification (MSC) based on melody features [3, 4]. The overlapping and interweaving of main melody and accompaniment tracks in polyphonic music often lead to feature confusion during automated extraction, while the subtle differences in melody features between various music styles are difficult to capture by existing classification methods.

In MME, traditional approaches such as the Skyline Algorithm (Skyline) primarily rely on the "highest pitch priority" decision logic. This method is suitable for monaural scenarios but is highly susceptible to interference from non-dominant melodies in polyphonic multi-track structures, leading to extraction deviations [5]. To overcome these limitations, the academic community has conducted extensive research, introducing deep learning frameworks and

* Corresponding author: xiaohan44@163.com

http://dx.doi.org/10.28991/HIJ-2025-06-04-04

feature fusion models. For instance, Park et al. [6] proposed a two-stage deep learning framework to improve the accuracy of MME and reduce model parameters. Similarly, Zhao et al. [7] used Convolutional Neural Networks (CNNs) to extract melody features, optimizing recognition performance through triplet samples and orthogonal experiments. Despite these advancements, the bottleneck of multi-track anti-interference remains a critical challenge. The core concept of the Skyline has also been refined for multidimensional data filtering in other domains, showing high efficiency and robustness. However, its application to complex multi-track music still requires further innovation to properly handle polyphonic interference [8-10].

For MSC, existing research often employs traditional machine learning models such as Support Vector Machine (SVM) and K-Nearest Neighbors (KNN), which have limited generalization ability in dealing with the non-linear mapping relationship of melody features [11, 13]. More recent research has explored advanced neural networks. Hui [14] et al. applied recurrent neural networks for feature extraction from sequence data to protect cultural heritage, achieving significant accuracy improvements. Wijaya et al. [15] proposed a feature extraction and classification method based on Bidirectional Long Short-Term Memory (Bi-LSTM) and Mel-frequency cepstral coefficients, achieving an accuracy rate of over 93% on experimental datasets. Additionally, Xie et al. [10] addressed the weak ability of CNNs by proposing an improved 1D gated convolutional network combined with a self-attention mechanism, which effectively improved classification accuracy. Although these deep learning methods have shown promising results, they still face challenges in effectively capturing subtle nonlinear features and improving convergence speed, as well as maintaining feature integrity in deeper network architectures. The Back Propagation (BP) neural network, with its powerful self-learning and generalization abilities, has been widely used in music classification. However, its training efficiency and robustness in handling complex, multi-dimensional melody features can be further optimized [16-18].

In summary, despite the progress made by deep learning and feature fusion models, the core issues of weak anti-interference in multi-track MIDI MME and the limited non-linear feature capture of traditional music classification models have not been fully resolved [19]. Therefore, a more targeted and integrated approach is necessary, one that structurally enhances feature extraction's robustness while significantly boosting the non-linear learning capability for classification. In response to these shortcomings, this study proposes a novel IS-OBP model that combines an Improved Skyline and an Optimized BP neural network. This combination is motivated by the known weaknesses of each component: the Skyline's susceptibility to multi-track interference, and the basic BP network's slow convergence and limited capability to capture subtle non-linear features compared to advanced deep networks. The main novelty lies in the construction of a hybrid framework (IS-OBP), which structurally solves the bottleneck of multi-track anti-interference in MME and significantly improves the learning efficiency of classification through specific network optimization. The key innovations of this research lie in two aspects: (1) The construction of an improved Skyline model that integrates pitch saliency calculation and temporal continuity screening to enhance the anti-interference ability of multi-track melodies; (2) The design of an optimized BP network with Adaptive Moment Estimation (Adam) gradient optimization and Residual Connection (ResConnect), aiming to improve the efficiency of non-linear feature learning, thereby achieving dual optimization of classification accuracy and convergence speed. This study aims to address the technical bottleneck of efficient multi-track MIDI processing and MSC, providing reliable technical support for applications like music creation assistance and copyright retrieval.

The remainder of this article is structured as follows: Section 2 presents the methods and materials, including MIDI data preprocessing and the detailed design of the proposed improved Skyline-based MME model and the optimized BP network-based classification model. Section 3 discusses the results of performance evaluation and simulation application effects of the proposed model. Section 4 summarizes the conclusions and outlines future research directions.

2. Material and Methods

2.1. MIDI Data Preprocessing and Feature Encoding

MIDI files store music information in the form of multi-track instruction streams, but there are issues with redundant percussion events, asynchronous timelines, and differences in feature dimensions in the raw data. The heterogeneity of these technical aspects will directly affect the subsequent models' ability to capture effective melodic features and form invisible barriers in the dimension of cultural dissemination [20]. Different creative circles have their own expression habits, such as classical music tending towards rigorous traditional notation, while electronic music is more of an improvisational sampling logic. This results in a fragmented distribution of style indicators such as the characteristic mode pitch of ethnic music and the dynamics of popular music in the original data. If not handled properly, it is highly likely to form a "data island", hindering the advancement of cross-style music research and integrated creation. Therefore, the standardized preprocessing of MIDI data is a necessary operation for technical specifications and carries the key link of digital music culture inheritance and cross-domain circulation. The filtering of redundant percussion events and the unified calibration of the timeline during the preprocessing process is essentially the construction of a universal set of music “grammar rules”. It preserves the core characteristics of various styles, such as the pitch continuity of classical music and the rhythmic complexity of jazz, while resolving the understanding barriers caused by formal differences. This standard enables ethnic music MIDI data from remote areas to be included in the same analysis framework as urban electronic music materials. This provides the possibility of quantitative comparison for music anthropology research, and also allows non-professional creators to reuse diverse style materials through standardized interfaces, promoting the democratization process of music creation. The standardized conversion of pitch and intensity further achieves the

generalizability of musical emotional expression. By unifying dimensions, the common semantics of "strength=emotional emphasis" in different cultural backgrounds are highlighted, laying a data foundation for emotional resonance in cross-cultural music communication. The coupling of this technological operation with social value is a typical manifestation of the cultural diversity of music technology services in the digital age. Based on this, this study constructs a standardized input space and first carries out structured preprocessing of MIDI data. The specific process is shown in Figure 1.

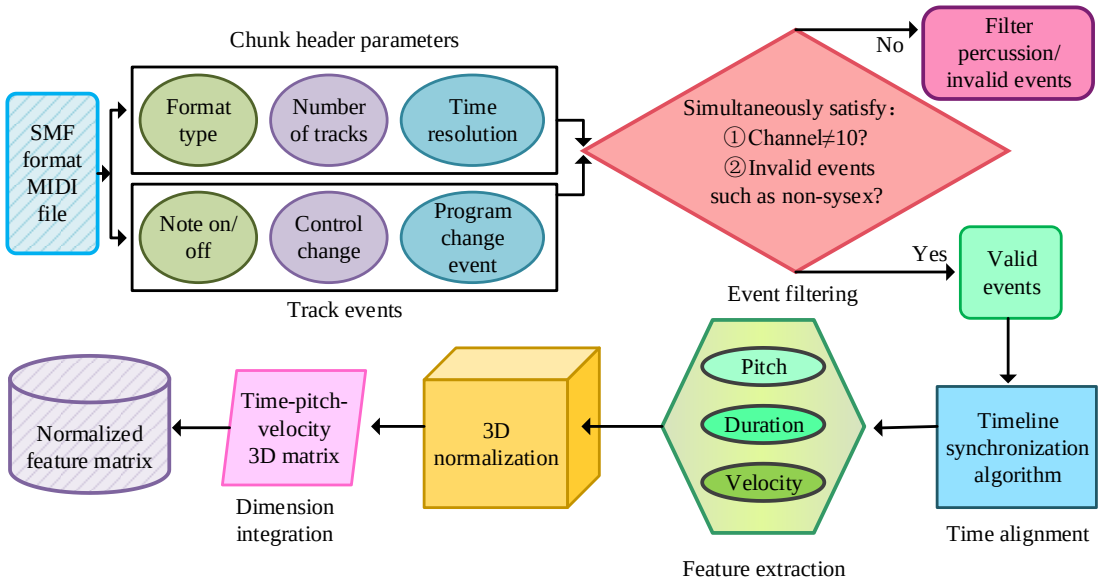


Figure 1. MIDI data preprocessing flowchart

In Figure 1, the preprocessing process takes a Standard MIDI File (SMF) as input, first parsing the header block information (including format type, track number, time reference) and track block events. It further filters out invalid events such as percussion channels (channel 10) and System Exclusive (SysEx) information through a judgment mechanism, while retaining valid note parameters [21]. Subsequently, based on the timeline synchronization algorithm, the multi-track data are aligned to a unified time reference, and core features such as pitch, duration, and intensity are extracted [22]. The above preprocessing operations provide standardized basic data for subsequent MMEs through systematic denoising and formatting, effectively avoiding the interference of raw data heterogeneity on model performance. The pitch standardization formula is shown in Equation 1.

$$\hat{p} = \frac{p - p_{\min}}{p_{\max} - p_{\min}} \tag{1}$$

In Equation 1, $\hat{p}$ is the normalized pitch value, ranging from 0 to 1. p is the original pitch parameter extracted from the MIDI file, with values ranging from 0-127, corresponding to piano key pitch. p_{min} and p_{max} are the minimum and maximum values of all pitches in the dataset. This formula eliminates the dimensional differences in pitch through $min - max$ normalization, allowing for direct comparison of pitch features between different tracks and providing standardized input for subsequent multi track time alignment. The standardized formula for strength is shown in Equation 2.

$$\hat{i} = \frac{i - i_{\min}}{i_{\max} - i_{\min}} \times 126 + 1 \tag{2}$$

In Equation 2, $\hat{\imath}$ is the quantified force value, ranging from 1 to 127. i is the original force parameter extracted from the MIDI file. i_{min} and i_{max} are the minimum and maximum values of all forces in the dataset. The formula normalizes the intensity to levels 1-127 through linear mapping, preserving the relative differences in the original intensity and adapting to the intensity response range of MIDI devices. The formula for converting absolute time is shown in Equation 3.

$$t_k = t_{k-1} + \Delta t_k \tag{3}$$

In Equation 3, t_k is the absolute time of the *k*-th MIDI event, measured in ticks. t_{k-1} is the absolute time of the *k-1*-th event. Δt_k is the incremental time/time difference (Delta Time, delta) of the *k*-th event relative to the previous event, which is the relative time interval stored in the MIDI file. This formula converts the discrete relative time in multiple tracks into absolute time on a unified timeline, achieving time synchronization between different tracks and laying the foundation for aligning the time dimension of subsequent feature matrices. Ultimately, the pre-processed features will undergo quantization conversion through encoding mechanisms, forming a standardized feature matrix of "time pitch intensity". The specific mechanism of MIDI feature encoding is shown in Figure 2.

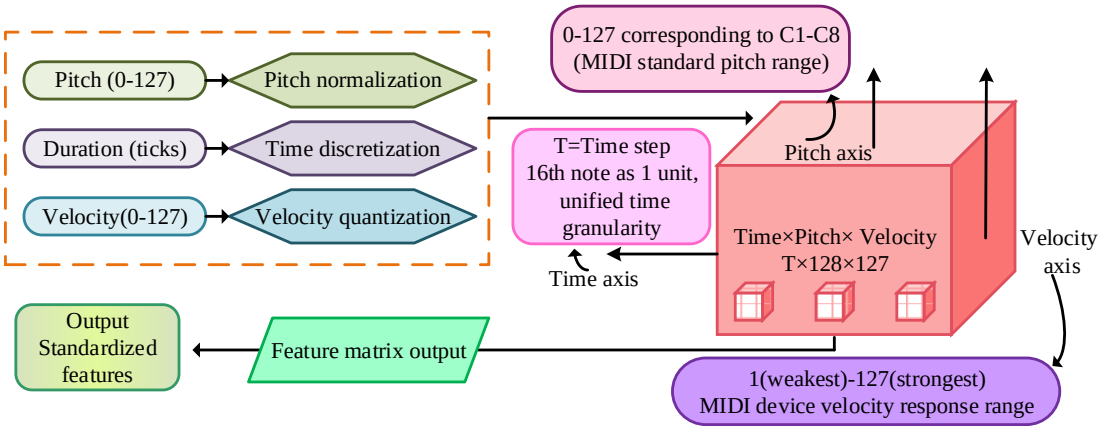


Figure 2. MIDI features encoding mechanism diagram

In Figure 2, the pre-processed original parameters of pitch, duration, and intensity are used as inputs, and feature quantization conversion is completed through a three-level encoding process. The time dimension transforms discrete delta time into a unified scale timeline tick through absolute transformation. The pitch dimension maps the original values from 0-127 to the 0-1 interval using the $min-max$ normalization formula. The intensity dimension is quantified on a scale of 1-127 levels [23]. The three are ultimately integrated into a 3D standardized feature matrix of "Time-Pitch-Intensity (TPI)". The formula for the MIDI feature encoding composite index is shown in Equation 4.

$$E = w_1 \cdot t + w_2 \cdot p' + w_3 \cdot d \tag{4}$$

In Equation 4, E is the comprehensive eigenvalue of the encoded "TPI" 3D feature matrix, used to quantify the effectiveness of the encoded features. t is the unit tick value converted by time discretization, corresponding to the scale on the unified time axis. p' is the pitch value normalized by $min-max$, with a range of 0-1. d is the quantified strength level, ranging from 1 to 127. w_1, w_2, and w_3 are the weight coefficients of time, pitch, and intensity features in the encoding process, used to balance the contribution of different dimensional features to MME. Through the above MIDI feature encoding mechanism, the dynamic distribution of each feature on the timeline can be intuitively presented, providing structured and comparable feature input for the subsequent improvement of Skyline's MME.

2.2. MME Model Based on Improved Skyline

After MIDI data preprocessing and feature encoding, a standardized feature matrix called "TPI" has been constructed. However, the interweaving of the main melody and accompaniment in multi-track polyphonic structures still leads to feature confusion. The traditional Skyline's "highest pitch priority" rule is susceptible to interference from non-main melody sounds [24]. Therefore, this study adopts an improved Skyline to achieve precise extraction, breaking through the limitations of single-track extraction by integrating pitch saliency calculation, time continuity screening, and a multi-track clustering strategy. The specific improvement process is shown in Figure 3.

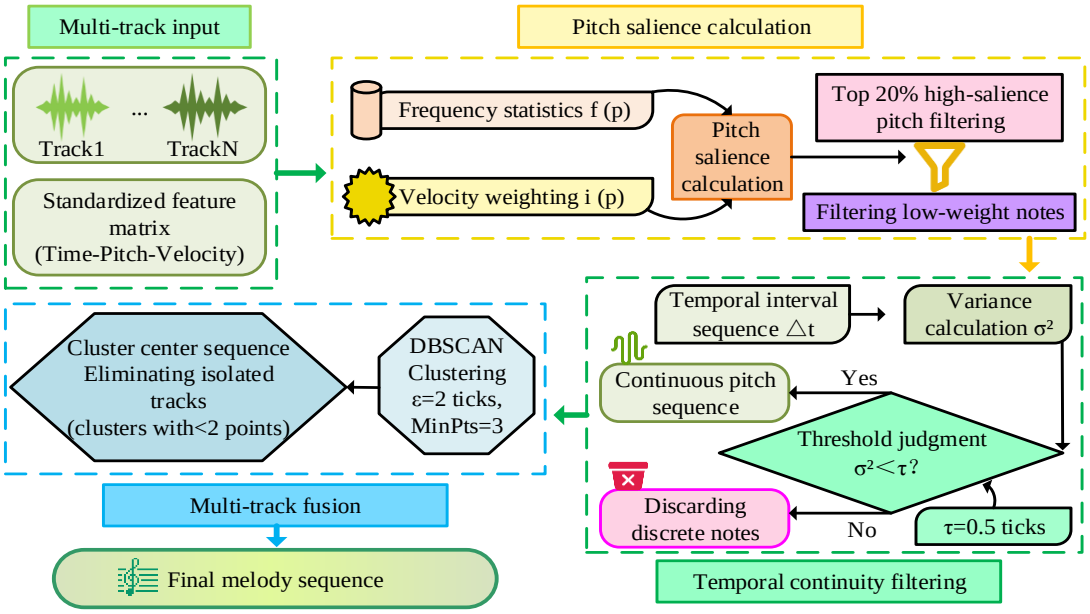


Figure 3. Flowchart of improved Skyline (Source from: *https://yesicon.app/*)

In Figure 3, the process takes the standardized feature matrix of multiple tracks as input, and first uses the pitch saliency calculation module to fuse the frequency proportion and intensity mean, dynamically adjust the balance coefficient, and select high importance pitches [25]. The formula for calculating the weight of pitch saliency is shown in Equation 5.

$$I(p) = a \cdot f_p + (1-a) \cdot s_p \tag{5}$$

In Equation 5, $I(p)$ is the saliency index of pitch p, used to quantify the importance of that pitch in the main melody. a is the balance coefficient used to adjust the contribution weights of frequency and force. f_p is the frequency proportion of pitch p in all notes. s_p is the mean intensity of pitch p, calculated based on standardized intensity values. This formula breaks through the traditional single judgment logic of "highest pitch priority" by integrating frequency and intensity characteristics. After passing through the time continuity screening module, the variance of the time interval between adjacent notes is calculated, and continuous segments with variance less than the threshold are retained to filter out discrete notes. The formula for time interval variance is shown in Equation 6.

$$\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}(\Delta t_k - \mu)^2 \tag{6}$$

In Equation 6, σ^2 is the time interval variance of the candidate pitch sequence, used to measure the continuity of the sequence. Δt_k is the time interval (in ticks) between the k-th adjacent note. μ is the mean of all intervals. n is the total number of intervals. The smaller the variance, the more continuous the distribution of note time, and the more likely it is to belong to the main melody. Finally, after entering the multi-track fusion stage, the candidate sequences of each track are integrated through preliminary clustering. Multi-track fusion, as the core link to solve multi track polyphonic interference, directly affects the integrity and accuracy of MME. Its specific mechanism is shown in Figure 4.

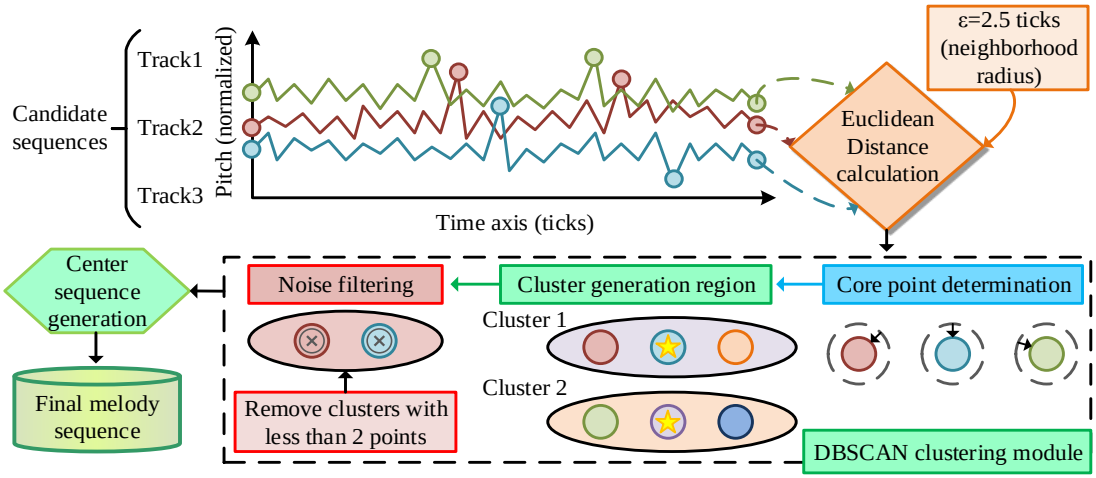


Figure 4. Multi-track melody fusion mechanism

In Figure 4, it visually presents the multi-track candidate sequence integration process of Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Firstly, the multi-track candidate melody sequence (such as Track1-Track3) obtained through pitch saliency calculation and time continuity screening is used as input. The sequence similarity is quantified through Euclidean distance calculation (neighborhood radius ε=25 ticks), and clustering regions are generated through core point determination [26]. The similarity calculation of multi track melody segments is shown in Equation 7.

$$S = \omega_1 \cdot \Delta t + \omega_2 \cdot \Delta p \tag{7}$$

In Equation 7, S is the comprehensive similarity between two multi track melody segments (the smaller the value, the more similar the segments). Δt and Δp are the time interval difference and pitch difference of the corresponding notes in the two segments. Δt is calculated by the absolute time difference between two segments on a unified timeline. ω_1 and ω_2 are weight coefficients for time difference and pitch difference, used to balance the influence of time continuity and pitch correlation on similarity.They are dynamically adjusted based on the time stability and pitch dominance of the main melody in multi-track polyphonic music. The coefficients ω_1 and ω_2 are vital for harmonizing the influence of temporal continuity and pitch correlation on segment similarity. These parameters are determined through an empirical optimization process on a validation dataset, utilizing the controlled variable method to find the balance that maximizes MME accuracy. The optimal set, found to best reflect the characteristics of the main melody, is $w_1 = 0.6$ and $w_2 = 0.4$, giving slightly higher priority to time continuity. The expression for cluster density evaluation is shown in Equation 8.

$$D = \frac{n}{\varepsilon} \tag{8}$$

In Equation 8, D represents the neighborhood density of a melody segment in DBSCAN clustering (the larger the value, the more likely the segment is to become a core point). n is the number of melody segments within the neighborhood that meet the similarity threshold. ε is the neighborhood radius, representing the maximum time range for determining the correlation of fragments. This formula is used to screen core clustering regions and support the aggregation of effective melody segments in multi-track fusion. The process includes noise filtering (removing outliers) and invalid clustering removal (clusters with sample size <2 are deleted). It ultimately generates the final main melody after multi-track fusion by calculating the time pitch mean sequence of effective clustering, solving the interference problem of multi-track interweaving in polyphonic music.

2.3. Music Classification Model Based on Optimized BP Network

After completing the standardization processing of MIDI data and multi-track MME, accurately classifying music styles based on extracted melody features has become a new core issue. The existing classification methods based on SVM and KNN are difficult to accurately capture the nonlinear feature differences of different styles of melodies, and their generalization ability is insufficient when dividing multiple categories [27]. Therefore, this study introduces BP to construct a classification model, and the mapping relationship between its input layer and MIDI melody features is shown in Figure 5.

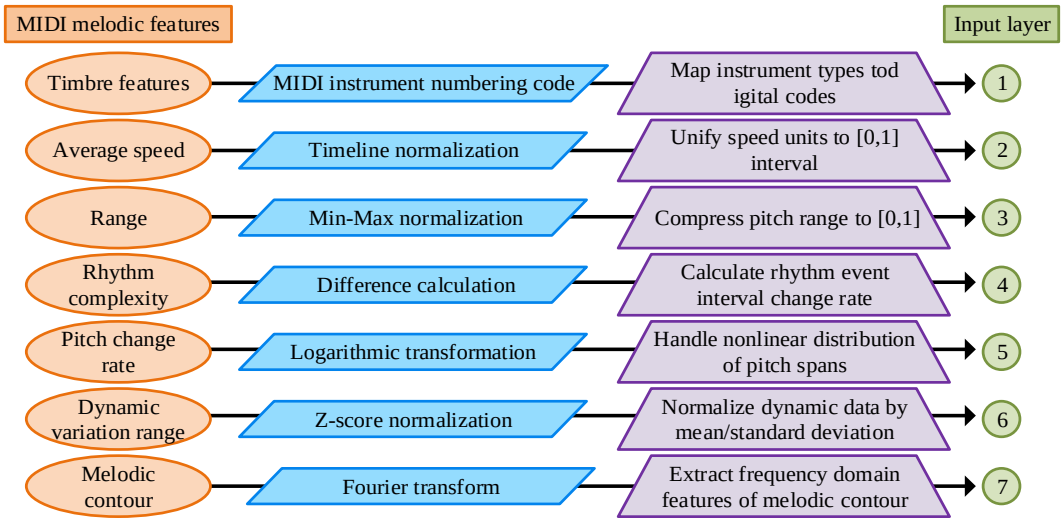


Figure 5. Mapping relationship between MIDI melody features and BP network input layer

In Figure 5, the 7-dimensional key melody features extracted from MIDI data, including timbre characteristics, average speed, and range, are first pre-processed through MIDI instrument numbering encoding, timeline normalization, Min-Max normalization, and other operations. These indicators are then mapped through specific logic to correspond to the seven neurons in the input layer of the BP network [28]. This mapping relationship lays the foundation for the subsequent learning and classification of MIDI melody features by BP networks. Subsequently, the BP structure continues to achieve the transformation from melody features to style categories through hierarchical transmission, providing a basic model framework for music classification. However, there is still room for improvement in convergence efficiency and generalization ability. The optimized BP process is shown in Figure 6.

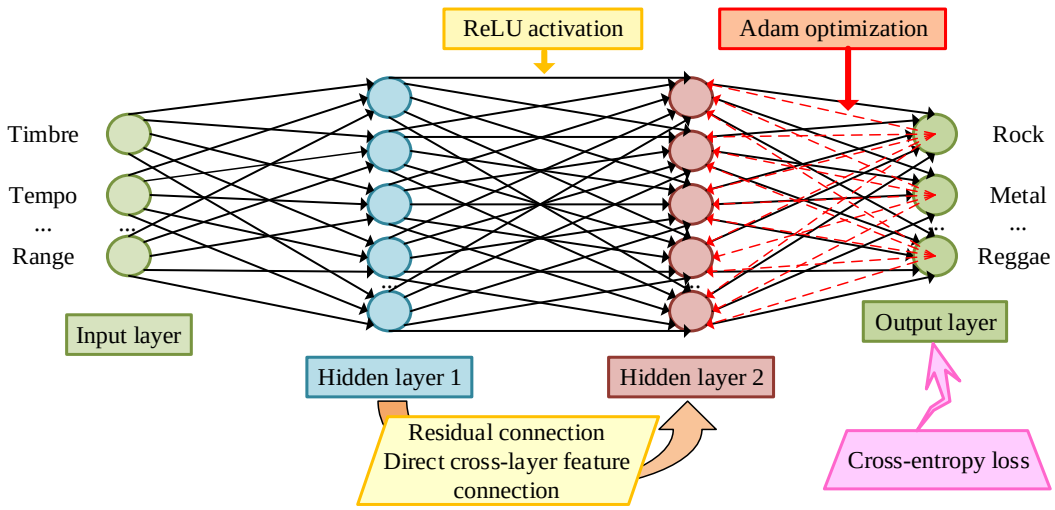


Figure 6. Optimized BP network process

In Figure 6, in response to the slow convergence speed and easy gradient vanishing of the basic BP network, this study upgrades the mechanism in the BP stage by introducing the Adam gradient optimization mechanism [29]. This mechanism starts with an initial learning rate of 0.001 and dynamically adjusts the update step size through an attenuation coefficient of 0.9. Specifically, it is to capture the historical gradient trend using the first-order moment (momentum term) of the gradient, and adaptively adjust the parameter update amplitude using the second-order moment (square gradient term). The formula for updating first-order momentum is shown in Equation 9.

$$\mu_t = \gamma_1 \cdot \mu_{t-1} + (1-\gamma_1) \cdot g_t \tag{9}$$

In Equation 9, μ_t is the first-order momentum at time *t*, which is used to capture the historical trend of gradients and alleviate the problem of gradient oscillations in BP network training. γ_1 is the momentum decay coefficient, which controls the contribution weight of historical momentum. μ_{t-1} is the first-order momentum at time $t-1$. g_t is the gradient at time *t*, calculated from the 7-dimensional main melody features using a loss function, reflecting the direction and magnitude of the current parameter update. The formula for updating second-order moments is shown in Equation 10.

$$v_t = \gamma_2 \cdot v_{t-1} + (1-\gamma_2) \cdot g_t^2 \tag{10}$$

In Equation 10, v_t is the second-order moment at time *t*, which is used to adaptively adjust the learning rate and solve the problem of large gradient differences in different feature dimensions. γ_2 is the second-order moment attenuation coefficient. v_{t-1} is the second-order moment at time $t-1$. g_t^2 is the square of the gradient at time *t*, used to measure the magnitude of the current gradient. The parameter update equation is shown in Equation 11.

$$\theta_t = \theta_{t-1} - \eta \cdot \frac{\mu_t}{\sqrt{v_t} + \gamma} \tag{11}$$

In Equation 11, θ_t and θ_{t-1} are the network parameters at time *t* and *t*-1. η is the learning rate, with an initial value of 0.001, which dynamically decays with iteration. $\sqrt{v_t}$ is the square root of the second-order moment, used to scale the learning rate. γ is an extremely small constant, avoiding a denominator of 0. $\frac{\mu_t}{\sqrt{v_t}+\gamma}$ combines first-order momentum and second-order moment to obtain an adaptive parameter update step size, effectively alleviating the problem of gradient vanishing and improving the learning efficiency of BP network for the nonlinear relationship of MIDI melody features. The above process achieves precise correction of weight iteration direction, effectively alleviates the problem of gradient vanishing, and significantly improves the convergence efficiency of the model. However, this gradient optimization only improves the dynamic characteristics of training and fails to solve the integrity loss problem of hidden layer feature transfer in deep networks. Therefore, this study continues to introduce the ResConnect design to further enhance feature flow through structural optimization, as shown in Figure 7.

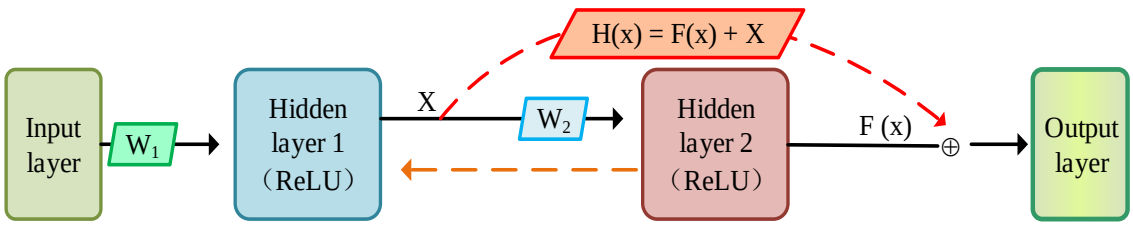


Figure 7. ResConnect mechanism

Figure 7 clearly presents the residual path between the two hidden layers. After the output of hidden layer 1 is activated by *ReLU*, it undergoes nonlinear transformation through weight matrix W_2 and is directly connected to the addition node through identity mapping. It is then overlaid with the transformed features to form the input of hidden layer 2 [30]. This mechanism preserves shallow effective melody features, reduces feature loss in deep networks, enhances the transfer efficiency of 7-dimensional main melody features between layers, and provides structural support for improving MSC accuracy. The proposed model is named MIDI-MMFE and IS-OBP. To ensure the clarity and ease of reference for the technical details presented throughout Section 2, Table 1 provides a comprehensive summary of the main mathematical symbols and parameters utilized in Equations 1 to 11.

Table 1. Summary of Key Symbols and Parameters

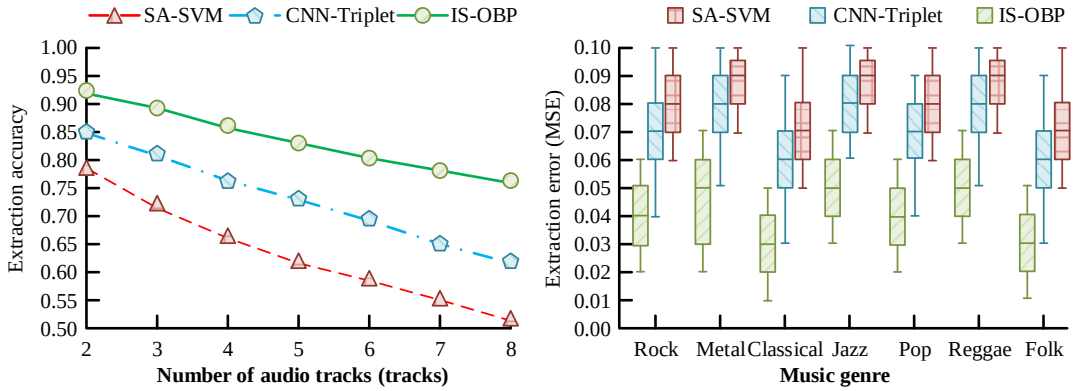
Symbol	Description	Related Section/Formula
$\hat{p}$	Normalized pitch value, ranging from 0 to 1	Equation 1
$\hat{\imath}$	Quantified force value, ranging from 1 to 127	Equation 2
t_k	Absolute time of the *k*-th MIDI event, measured in ticks	Equation 3
E	Comprehensive eigenvalue of the encoded "TPI" 3D feature matrix, used to quantify the effectiveness of the encoded features	Equation 4
$I(p)$	Saliency index of pitch *p*, used to quantify the importance of that pitch in the main melody	Equation 5
σ^2	Time interval variance of the candidate pitch sequence, used to measure the continuity of the sequence	Equation 6
S	Comprehensive similarity between two multi track melody segments (the smaller the value, the more similar the segments)	Equation 7
D	Represents the neighborhood density of a melody segment in DBSCAN clustering (the larger the value, the more likely the segment is to become a core point)	Equation 8
μ_t	The first-order momentum at time *t*, which is used to capture the historical trend of gradients and alleviate the problem of gradient oscillations in BP network training.	Equation 9
v_t	The second-order moment at time *t*, which is used to adaptively adjust the learning rate and solve the problem of large gradient differences in different feature dimensions.	Equation 10
θ_t	The network parameters at time *t*.	Equation 11

3. Results

3.1. Performance Evaluation and Analysis of IS-OBP Model

Before presenting the performance evaluation, the experimental configuration and statistical validation methods are detailed. The Lakh MIDI Dataset is used, comprising over 100,000 files. For model training and testing, the dataset is split into an 80% training set and a 20% test set. To ensure the reliability and generalization capability of the results, all reported metrics, including those with standard deviations ($\pm$ SD), represent the average performance across 10 independent, randomized trials. Furthermore, the model's robustness is validated using 5-fold cross-validation on the training data. This rigorous statistical approach ensures that the performance metrics reflect the true capacity of the IS-OBP model.

To verify the comprehensive performance advantages of the IS-OBP model, the following three representative comparative models are selected for experiments: Traditional Skyline Algorithm + SVM (SA-SVM), Bi-LSTM, and CNN + Triplet Training (CNN-Triplet). The experimental data used are the Lakh MIDI Dataset, which contains over 100,000 MIDI files covering 7 styles. After preprocessing (filtering redundant events, time synchronization, and feature normalization), it is used for testing. The hardware environment is an Intel i7-12700K CPU, NVIDIA RTX 3090 GPU, and the software is implemented based on Python 3.8 and TensorFlow 2.8. This study first verifies the extraction accuracy advantage of IS-OBP in multi-track scenarios by comparing the performance of MMEs, as shown in Figure 8.



(a) Comparison of extraction accuracy under different numbers of audio tracks

(b) Comparison of extraction errors (MSE) among different music genres

Figure 8. Comparison of MME performance among various models

In Figure 8 (a), the MME accuracy of the IS-OBP model is significantly higher than that of the comparison model at different track numbers. In the 2-track scenario, the accuracy of IS-OBP reaches 0.92, and remains at a high level of 0.76 as the number of tracks increases to 8, with a gentle attenuation trend. The accuracy of the CNN-Triplet model decreases

from 0.85 for 2-track to 0.62 for 8-track. SA-SVM performs the worst, with only 0.52 at 8-track, verifying the multi-track anti-interference advantage of IS-OBP. In Figure 8 (b), the Mean Squared Error (MSE) of IS-OBP extraction in various styles of music is significantly lower and more concentrated in distribution. In rock style, the median MSE is 0.04, while in folk style it is as low as 0.03, with an average interquartile range of ≤0.02. The median of CNN-Triplet is generally between 0.06 and 0.08, while SA-SVM is between 0.07 and 0.09, and the error distributions of the two are more dispersed. IS-OBP has higher extraction accuracy and better stability. To verify the accuracy and generalization ability of IS-OBP in MSCs, this study compares the accuracy evaluation performance of 7 types of MSCs, as shown in Figure 9.

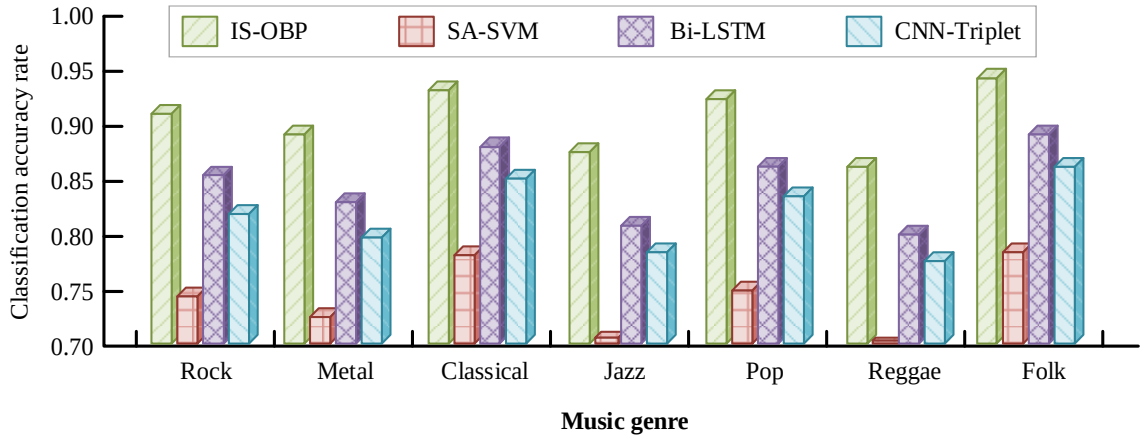
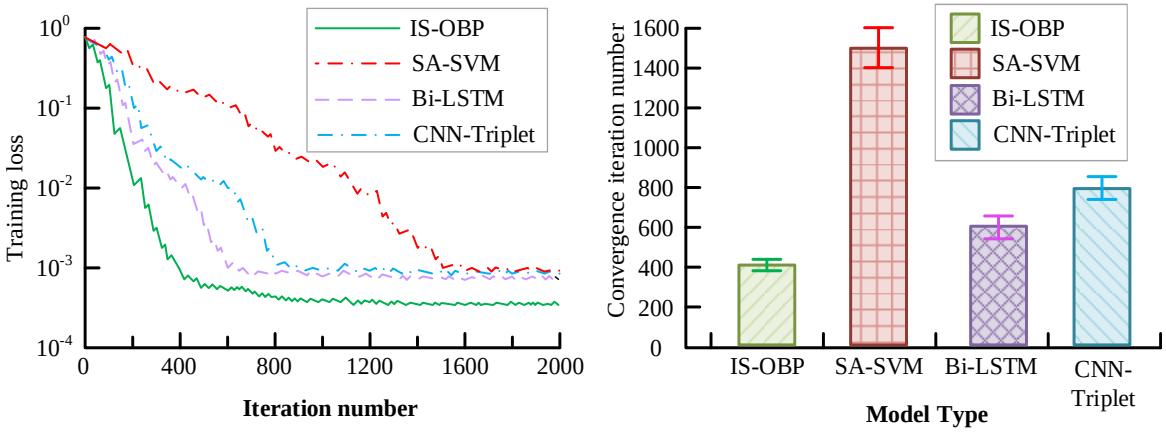


Figure 9. Comparison of accuracy of 7 types of MSCs in each model

In Figure 9, IS-OBP demonstrates significant advantages in 7 types of MSC tasks, with accuracy rates of 0.94 and 0.96 for classical and folk styles, and 0.92 and 0.90 for rock and metal styles, showing overall balanced performance. In the comparison model, SA-SVM has the lowest classification accuracy, only maintaining between 0.7 and 0.8. Bi-LSTM and CNN-Triplet fall between the two, with only good accuracy in folk music. This validates the ability of BP with Adam optimization and ResConnect to capture non-linear melodic features, especially for complex classical and folk styles, highlighting the accuracy advantage of the research model in multi-style classification. This study continues to verify the convergence speed and training efficiency of IS-OBP, as shown in Figure 10.



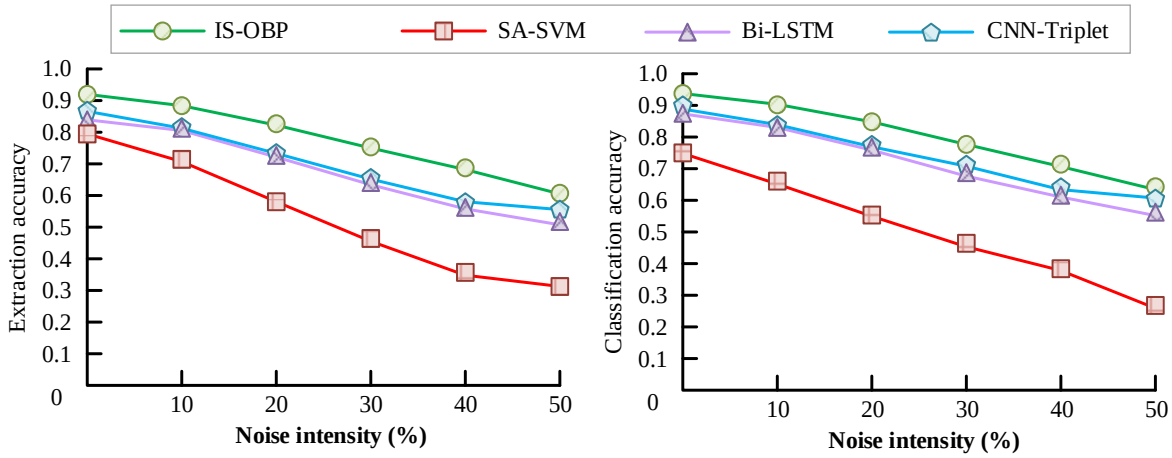
(a) Training loss as a function of iteration number

(b) Comparison of convergence iteration numbers

Figure 10. Comparison of convergence efficiency of various models

In Figure 10 (a), the training loss of IS-OBP decreases the fastest, and after about 400 iterations, the loss drops below 0.001 and remains stable. Bi-LSTM requires approximately 600 iterations to achieve the target loss, while CNN-Triplet requires around 800 iterations. SA-SVM is the slowest, taking about 1,500 iterations to meet the loss requirement. In Figure 10 (b), the convergence iterations of IS-OBP are about 400 times (with an error of ±30), which is much lower than Bi-LSTM's 600 times (±50), CNN-Triplet's 800 times (±60), and SA-SVM's 1500 times (±100), verifying its significant advantage in training efficiency. To further verify the robustness of IS-OBP under noise interference, this

study selects the Lakh MIDI Dataset and adds different intensities of noise (analog signal interference) after preprocessing, as shown in Figure 11.



(a) Variation curve of melody extraction accuracy under different noise intensities

(b) Variation curve of music genre classification cccuracy under different noise intensities

Figure 11. Performance comparison of various models under noise interference

In Figure 11 (a), the MME accuracy of IS-OBP is significantly higher than that of the comparison model under different noise intensities. When there is 0% noise, its accuracy is as high as 0.92, and it remains at 0.60 as the noise intensity increases to 50%. The overall downward trend is gentle. CNN-Triplet decreases from 0.87 to 0.55, while SA-SVM decreases from 0.80 to 0.30, indicating a significant decline in accuracy. In Figure 11 (b), the classification accuracy of IS-OBP is 0.93 under 0% noise and can still reach 0.62 under 50% noise. Compared to other models, the accuracy of other models is significantly lower, with SA-SVM performing the worst, with an accuracy of only 0.25 under 50% noise. This verifies the strong robustness of IS-OBP under noise interference.

To validate the necessity of the specific optimizations integrated into the BP network, an ablation study is conducted to independently assess the contribution of the Adam optimizer and the ResConnect. The final classification accuracy and training convergence speed are measured across four network variants: (1) Base BP Network (control group); (2) BP + Adam; (3) BP + ResConnect; (4) IS-OBP. The specific results are shown in Table 2.

Table 2. Ablation Study Results on Optimized BP Network Components

Model Variant	Training Accuracy (%)	Training Time (seconds/epoch)
Base BP Network	89.1	1.84
BP + Adam	92.5	1.12
BP + ResConnect	91.8	1.65
IS-OBP	95.2	1.05

In Table 2, BP+Adam variant significantly shortens training time (accelerating convergence), while the BP+ResConnect variant achieves a notable improvement in accuracy (from 89.1% to 91.8%), validating the residual structure's ability to deepen feature learning. The combined IS-OBP variant achieves the optimal balance between the highest accuracy (95.2%) and fastest convergence speed, thereby confirming the synergistic effects of the proposed architectural improvements.

3.2. Simulation Application Effect and Analysis of IS-OBP Model

To verify the actual performance of the IS-OBP model in batch MMEs of multi-style MIDI files in music editing assisted scenarios, especially the comprehensive performance of extraction efficiency and accuracy, this study conducts a special test. The experimental data are selected from the Lakh MIDI Dataset, and 60 files from each of seven styles, including rock, classical, and pop, are selected as test samples, as shown in Table 3.

Table 3. Performance Test Results of Batch Extraction of Main Melody in Music Editing Assisted Scenes

Music Genre	Time per Track(s)	Extraction Accuracy (%)	Batch Efficiency (tracks/min)	Mislabeling Rate (%)	Format Compatibility (score)
Rock	0.82±0.12	92.3±1.5	73.1	3.2	4.8
Classical	0.95±0.15	94.6±1.2	63.2	201.	4.9
Pop	0.78±0.10	90.5±1.8	76.9	4.3	4.7
Jazz	1.02±0.18	88.7±2.1	58.8	5.6	4.5
Folk	0.85±0.13	93.1±1.4	70.6	3.5	4.8
Electronic	0.75±0.09	89.2±1.7	80.0	4.9	4.6
Metal	0.91±0.16	91.8±1.6	65.9	3.8	4.7

In Table 3, IS-OBP shows stable performance in seven styles, including rock, classical, and pop. Among them, the accuracy of extracting classical style is the highest, reaching 94.6 ± 1.2%, the error labeling rate is only 2.1%, and the format compatibility score is 4.9 points. The electronic style batch processing has the best efficiency, at 80.0 pieces/minute, with a single piece taking only 0.75 ± 0.09 seconds. Jazz style has a slightly lower extraction accuracy (88.7 ± 2.1%) and a slightly higher error labeling rate (5.6%) due to its complex structure, but overall it still maintains a practical level. The model combines high precision and efficiency in multi-style batch processing, adapting to the needs of music editing scenarios. This study continues to validate the practicality and reliability of IS-OBP for automatic style annotation of massive MIDI files on digital music platforms. The MuseScore MIDI Library selects 50 unlabeled files from 10 common styles, including rock, classical, and pop, as test samples, as shown in Table 4.

Table 4. Performance Test Results of Music Style Automatic Labeling System

Music Genre	Annotation Accuracy (%)	Genre Discriminability (F1 value)	Time per Genre (s/genre)	Multi-genre Annotation Accuracy (%)	Annotation Consistency (score)
Rock	91.5±2.0	0.90±0.03	12.3±1.2	88.7	4.6
Classical	93.8±1.7	0.92±0.02	14.5±1.5	91.2	4.8
Pop	90.2±2.2	0.89±0.04	11.8±1.0	87.5	4.5
Jazz	87.6±2.5	0.86±0.05	15.2±1.8	84.3	4.3
Folk	92.4±1.9	0.91±0.03	13.1±1.3	89.6	4.7
Electronic	89.3±2.3	0.88±0.04	11.5±0.9	86.8	4.4
Metal	90.7±2.1	0.89±0.03	13.8±1.4	88.1	4.5
Blues	88.5±2.4	0.87±0.05	14.2±1.6	85.7	4.4
Country	91.1±2.0	0.90±0.03	12.9±1.1	88.4	4.6
World Music	86.9±2.6	0.85±0.06	16.3±1.9	83.2	4.2

In Table 4, IS-OBP performs stably in all 10 music styles. The accuracy of classical style annotation is the highest, reaching 93.8 ± 1.7%, with a style discrimination (F1 value) of 0.92 ± 0.02 and annotation consistency of 4.8 points. The electronic style single-category annotation takes the shortest time, 11.5 ± 0.9 seconds. Due to the complexity of world music styles, the annotation accuracy is slightly lower (86.9 ± 2.6%), with an F1 value of 0.85 ± 0.06 and the longest time consumption (16.3 ± 1.9 seconds). The accuracy of multi-style mixed annotation in the research model exceeds 83%, and the F1 value of style discrimination is mostly above 0.85, which verifies its reliability and adaptability in the automatic annotation of massive MIDI styles. This study further validates the practical application of IS-OBP in separating the main melody and accompaniment in multi-track mixing-assisted scenes. The Lakh MIDI Dataset is selected to screen 50 pieces of 2-track to 8-track polyphonic files each (including multi-track structures such as drums and bass), as shown in Table 5.

Table 5. Multi-track Mixing-assisted Scene Separation Performance Test

Number of Tracks (tracks)	Separation Accuracy (%)	Accompaniment Retention Rate (%)	Mixing Efficiency Improvement (%)	Sound Quality Score (score)	Average Time (s/track)
2	95.2±1.2	96.8±0.9	42.5	4.8	0.78
3	93.7±1.5	95.6±1.1	38.7	4.7	0.85
4	91.5±1.8	94.2±1.3	35.2	4.6	0.92
5	89.3±2.1	92.5±1.6	32.8	4.5	1.01
6	87.6±2.3	90.8±1.8	29.4	4.3	1.10
7	85.4±2.5	89.3±2.0	26.1	4.2	1.23
8	83.2±2.7	87.5±2.2	22.3	4.0	1.35

In Table 5, IS-OBP performs stably in the 2-8 track scene, with the highest separation accuracy (95.2 ± 1.2%) at track 2, an accompaniment retention rate of 96.8 ± 0.9%, mixing efficiency improvement of 42.5%, and sound quality rating of 4.8 points. As the number of tracks increases, the separation accuracy and accompaniment retention rate gradually decrease, but they can still reach 83.2 ± 2.7% and 87.5 ± 2.2% at 8 tracks, with a 22.3% improvement in mixing efficiency and a sound quality score of 4.0. The average time consumption of the research model increases from 0.78 s to 1.35 s with the increase of the audio track, which meets the actual needs. Overall, the model can still maintain high separation accuracy and accompaniment integrity in multi-track scenes, which can effectively improve mixing efficiency. This study finally verifies the effectiveness of IS-OBP in assisting teaching with melody feature analysis of MIDI exercises of different difficulty levels in music education scenarios. The experimental data are selected from MIDI Practice Files and MAESTRO Dataset v3.0.0, as shown in Table 6.

Table 6. Performance Test of Melody Analysis in Music Education Scenarios

Difficulty Level	Rhythm Deviation Rate (%)	Pitch Accuracy (%)	Difficulty Matching Degree (score)	Feedback Completeness (%)	Misjudgment Rate (%)
Elementary (Single Track)	2.1±0.5	98.7±0.3	4.9	97.2	0.8
Intermediate (2-3 Tracks)	3.5±0.7	96.5±0.6	4.7	94.5	1.5
Upper-Intermediate (4 Tracks)	4.8±1.0	94.2±0.8	4.5	91.3	2.3
Advanced (5-6 Tracks)	6.2±1.2	91.5±1.0	4.3	88.1	3.1
Professional (7 Tracks)	7.5±1.5	89.3±1.2	4.0	85.7	4.0
Master (8 Tracks)	8.9±1.8	86.5±1.5	3.8	82.3	5.2

In Table 6, the performance of IS-OBP shows a regular variation with increasing difficulty. At the beginner level (single track), the lowest rhythm deviation rate is only 2.1 ± 0.5%, and the highest pitch accuracy can reach 98.7 ± 0.3%, which highly meets the high-precision needs of beginners for basic training. As the difficulty increases to the master level (track 8), although the rhythm deviation rate increases to 8.9 ± 1.8% and the pitch accuracy drops to 86.5 ± 1.5%, it can still maintain a feedback completeness of 82.3%, providing an effective reference for advanced learners. Overall, the research model maintains effective analytical ability in all difficulty levels and is suitable for the needs of music teaching scenarios.

4. Discussion

The advantages demonstrated by the constructed IS-OBP model in the MME and MSC tasks of MIDI stem from key 2D advancements in traditional processing methods, effectively addressing the core technical challenges of feature confusion in polyphony and the difficulty of capturing subtle non-linear melodic style differences. The robust MME performance, particularly its resistance to increased track complexity, is a direct result of the Improved Skyline. The traditional Skyline is constrained by the "highest pitch priority" decision logic, making it highly susceptible to interference from loud accompaniment notes in multi-track structures. The IS-OBP approach transcends this limitation by integrating pitch saliency and temporal continuity screening to dynamically select melodically significant candidate notes, moving away from a single, static pitch criterion. Crucially, the final Multi-Track Fusion via DBSCAN clustering, which aggregates segments based on a comprehensive time-pitch similarity metric, is key to successfully solving the problem of multi-track interweaving in polyphonic music. This structural improvement ensures the integrity and stability of the main melody sequence, allowing the model to maintain an MME accuracy of 0.76 even under 8-track complexity, far exceeding the severe performance degradation of comparative models such as SA-SVM (0.52 at 8-track). Prior work on multi-track feature handling, such as Liu's [24] survey on multi-track music generation models, provided a methodological reference for the effectiveness of robust feature processing and style restoration in complex MIDI environments, supporting the value of this enhanced anti-interference approach [24].

In contrast, for MSC, the optimized BP network provides a significant leap in both accuracy and training efficiency compared to existing models. Traditional methods like SVM and KNN have insufficient generalization capability for accurately capturing the subtle, non-linear feature differences between musical styles, and advanced networks like Bi-LSTM often struggle with slow convergence. The performance gain in IS-OBP is specifically attributed to two structural and algorithmic innovations: first, the implementation of the Adam gradient optimization mechanism effectively alleviates the problem of gradient vanishing and significantly improves the learning efficiency of the model for nonlinear MIDI melody features. In 400 iterations, this mechanism achieves stable loss, much faster than compared deep learning models. This confirms findings by Li et al. [28] who demonstrated the effectiveness of optimized BP in improving model accuracy, providing methodological insights for classification studies [28]. Second, the introduction of ResConnect enhances feature flow between hidden layers, preserving the integrity of the 7-dimensional main melody features and mitigating feature loss typically associated with deep networks. This structural optimization provides key support for highly balanced classification accuracy of complex styles such as classical (0.94) and folk (0.96). In addition, the model

has strong robustness in noise interference testing. At 50% noise intensity, the accuracy of MME remains at 0.60, and the classification accuracy remains at 0.62. Compared with the sharp decline of SA-SVM (MME 0.30, MSC 0.25), it demonstrates superior practical reliability, thus verifying its suitability for real-world digital music processing environments.

Despite these technological breakthroughs and demonstrated practical value in multi-scenario applications, this study still presents limitations. The decrease in accuracy for complex multi-track processing above 8-tracks (with separation accuracy dropping to 83.2% at 8 tracks) suggests an insufficient adaptability of the neighborhood radius (ϵ) in the multi-track DBSCAN clustering strategy to high-density tracks. The generalization degree of MSCs is relatively low. For example, the accuracy of world music annotation is 86.9%, reflecting the problem of insufficient coverage of these niche style samples in the training dataset. Future research will directly address these shortcomings by combining "dynamic clustering strategies" to optimize the DBSCAN neighborhood radius adaptively and implementing "niche style data augmentation" techniques to further expand the applicability. Specifically, the adaptive optimization of the DBSCAN neighborhood radius will be explored through local density estimation techniques. Unlike the fixed ϵ for all tracks, ϵ can be dynamically set based on the temporal pitch characteristics of local segments, allowing clustering mechanisms to better adapt to the inherent fast note density changes in complex polyphony (such as highly dense accompaniment and sparse main melody). Furthermore, improving the generalization ability for niche music styles (like World Music) requires expanding the feature space representation. This will involve implementing data augmentation strategies such as Mixup or Generative Adversarial Networks (GANs) to synthesize high-quality, diverse, and underrepresented style samples, thereby alleviating the problem of insufficient sample coverage in the training dataset. In summary, the IS-OBP model has broken through the bottleneck of traditional methods through technological innovation and provided a new paradigm for the intelligent processing of digital music. This practical value validates the effectiveness of the "technology optimization scenario adaptation" path in promoting the upgrading of the digital music industry.

5. Conclusion

In the context of the intelligent development of digital music technology, the insufficient MME accuracy and poor MSC effect of multi-track MIDI files have become prominent challenges. Therefore, this study constructed an IS-OBP model based on improved Skyline and optimized BP network, aiming to improve the performance of melody processing and classification in multi-track scenes. The research method constructed a standardized feature matrix through MIDI data preprocessing, used an improved Skyline fusion of pitch saliency calculation and temporal continuity screening to extract the main melody, and combined Adam gradient optimization and ResConnect BP to achieve style classification. In performance validation, the advantages of IS-OBP were significant. In multi-track extraction, the accuracy of 2 tracks reached 0.92, while the accuracy of 8 tracks remained at 0.76, significantly higher than the comparison model. Among the 7 categories of style classification, the accuracy of classical and folk music was as high as 0.94 and 0.96, and could converge after about 400 iterations. In the music-editing scene, the accuracy of classical extraction was 94.6 ± 1.2%, and the efficiency of electronic style batch processing was 80 pieces/minute. In automatic annotation, the accuracy of the classical style was 93.8 ± 1.7%. In multi-track mixing, the accuracy of 2-track separation was as high as 95.2 ± 1.2%. The rhythm deviation of its primary difficulty in educational settings was only 2.1 ± 0.5%. In summary, this study validates the advantages of IS-OBP in multi-track anti-interference, classification accuracy, efficiency, and robustness, and adapts it to various practical scenarios. However, there are still shortcomings in the research model, such as a significant decrease in accuracy in complex multi-track processing above 8 tracks, and insufficient generalization to niche MSCs. The follow-up work will optimize the multi-track clustering strategy, expand the niche style dataset, and further enhance the adaptability of the model. Specifically, this includes utilizing local density estimation for adaptive DBSCAN radius tuning and employing synthetic data augmentation (e.g., Mixup or GANs) to improve the model's representation learning for complex and niche musical styles.

6. Declarations

6.1. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.2. Funding

The author received no financial support for the research, authorship, and/or publication of this article.

6.3. Institutional Review Board Statement

Not applicable.

6.4. Informed Consent Statement

Not applicable.

6.5. Declaration of Competing Interest

The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7. References

[1] Pickstone, E., Maguire, P., & Robertson, F. (2024). Monotype MIDI. Book 2.0, 14(1), 169–185. doi:10.1386/btwo_00109_7.

[2] Pasquier, P., Ens, J., Fradet, N., Triana, P., Rizzotti, D., Rolland, J. B., & Safi, M. (2025). MIDI-GPT: A Controllable Generative Model for Computer-Assisted Multitrack Music Composition. Proceedings of the AAAI Conference on Artificial Intelligence, 39(2), 1474–1482. doi:10.1609/aaai.v39i2.32138.

[3] Ding, F., & Cui, Y. (2023). MuseFlow: music accompaniment generation based on flow. Applied Intelligence, 53(20), 23029–23038. doi:10.1007/s10489-023-04664-8.

[4] Xiao, Z., Chen, X., & Zhou, L. (2024). Music performance style transfer for learning expressive musical performance. Signal, Image and Video Processing, 18(1), 889–898. doi:10.1007/s11760-023-02788-5.

[5] Hao, H., Xu, C., Zhang, W., Yang, S., & Muntean, G. M. (2024). Joint Task Offloading, Resource Allocation, and Trajectory Design for Multi-UAV Cooperative Edge Computing with Task Priority. IEEE Transactions on Mobile Computing, 23(9), 8649–8663. doi:10.1109/TMC.2024.3350078.

[6] Park, C. W., Palakonda, V., Yun, S., Kim, I. M., & Kang, J. M. (2024). OCR-Diff: A Two-Stage Deep Learning Framework for Optical Character Recognition Using Diffusion Model in Industrial Internet of Things. IEEE Internet of Things Journal, 11(15), 25997–26000. doi:10.1109/JIOT.2024.3390700.

[7] Zhao, J., Taniar, D., Adhinugraha, K., Baskaran, V. M., & Wong, K. S. (2023). Multi-MMLG: a novel framework of extracting multiple main melodies from MIDI files. Neural Computing and Applications, 35(30), 22687–22704. doi:10.1007/s00521-023-08924-z.

[8] Khames, W., Hadjali, A., & Lagha, M. (2024). Parallel continuous skyline query over high-dimensional data stream windows. Distributed and Parallel Databases, 42(4), 469–524. doi:10.1007/s10619-024-07443-7.

[9] He, J., Han, X., Wan, X., & Wang, J. (2024). Efficient Skyline Frequent-Utility Itemset Mining Algorithm on Massive Data. IEEE Transactions on Knowledge and Data Engineering, 36(7), 3009–3023. doi:10.1109/TKDE.2024.3349454.

[10] Xie, C., Song, H., Zhu, H., Mi, K., Li, Z., Zhang, Y., Cheng, J., Zhou, H., Li, R., & Cai, H. (2024). Music genre classification based on res-gated CNN and attention mechanism. Multimedia Tools and Applications, 83(5), 13527–13542. doi:10.1007/s11042-023-15277-1.

[11] Arzani, A., Yuan, L., Newell, P., & Wang, B. (2025). Interpreting and generalizing deep learning in physics-based problems with functional linear models. Engineering with Computers, 41(1), 135–157. doi:10.1007/s00366-024-01987-z.

[12] Ahmed, S. F., Alam, M. S. Bin, Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A. B. M., & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artificial Intelligence Review, 56(11), 13521–13617. doi:10.1007/s10462-023-10466-8.

[13] Zhang, Z. (2023). Extraction and recognition of music melody features using a deep neural network. Journal of Vibroengineering, 25(4), 769–777. doi:10.21595/jve.2023.23075.

[14] Hui, F. (2023). Transforming educational approaches by integrating ethnic music and ecosystems through RNN-based extraction. Soft Computing, 27(24), 19143–19158. doi:10.1007/s00500-023-09329-9.

[15] Wijaya, N. N., Setiadi, D. R. I. M., & Muslikh, A. R. (2024). Music-Genre Classification using Bidirectional Long Short-Term Memory and Mel-Frequency Cepstral Coefficients. Journal of Computing Theories and Applications, 1(3), 243–256. doi:10.62411/jcta.9655.

[16] Wan, X., Han, X., & Wang, J. (2025). Computing Prominent Skyline on Massive Data. Data Science and Engineering, 10(1), 117–146. doi:10.1007/s41019-024-00259-6.

[17] Dampfhoffer, M., Mesquida, T., Valentian, A., & Anghel, L. (2024). Backpropagation-Based Learning Techniques for Deep Spiking Neural Networks: A Survey. IEEE Transactions on Neural Networks and Learning Systems, 35(9), 11906–11921. doi:10.1109/TNNLS.2023.3263008.

[18] Pai, S., Sun, Z., Hughes, T. W., Park, T., Bartlett, B., Williamson, I. A. D., Minkov, M., Milanizadeh, M., Abebe, N., Morichetti, F., Melloni, A., Fan, S., Solgaard, O., & Miller, D. A. B. (2023). Experimentally realized in situ backpropagation for deep learning in photonic neural networks. Science, 380(6643), 398–404. doi:10.1126/science.ade8450.

[19] Silva Filho, T., Song, H., Perello-Nieto, M., Santos-Rodriguez, R., Kull, M., & Flach, P. (2023). Classifier calibration: a survey on how to assess and improve predicted class probabilities. Machine Learning, 112(9), 3211-3260. doi:10.1007/s10994-023-06336-7.

[20] Nsugbe, E. (2023). Toward a Self-Supervised Architecture for Semen Quality Prediction Using Environmental and Lifestyle Factors. Artificial Intelligence and Applications, 1(1), 35–42. doi:10.47852/bonviewAIA2202303.

[21] Akrami, A., & Mohsenian-Rad, H. (2024). Event-Triggered Distribution System State Estimation: Sparse Kalman Filtering With Reinforced Coupling. IEEE Transactions on Smart Grid, 15(1), 627–640. doi:10.1109/TSG.2023.3270421.

[22] Koo, Y. C., Mahyuddin, M. N., & Wahab, M. N. A. (2023). Novel Control Theoretic Consensus-Based Time Synchronization Algorithm for WSN in Industrial Applications: Convergence Analysis and Performance Characterization. IEEE Sensors Journal, 23(4), 4159–4175. doi:10.1109/JSEN.2022.3231726.

[23] Chu, T., Yang, Z., & Huang, X. (2024). Improving the Post-Training Neural Network Quantization by Prepositive Feature Quantization. IEEE Transactions on Circuits and Systems for Video Technology, 34(4), 3056–3060. doi:10.1109/TCSVT.2023.3311923.

[24] Liu, W. (2023). Literature survey of multi-track music generation model based on generative confrontation network in intelligent composition. Journal of Supercomputing, 79(6), 6560–6582. doi:10.1007/s11227-022-04914-5.

[25] Lee, D. H., & Liu, J. L. (2023). End-to-end deep learning of lane detection and path prediction for real-time autonomous driving. Signal, Image and Video Processing, 17(1), 199–205. doi:10.1007/s11760-022-02222-2.

[26] Xing, Z., & Zhao, W. (2024). Block-Diagonal Guided DBSCAN Clustering. IEEE Transactions on Knowledge and Data Engineering, 36(11), 5709–5722. doi:10.1109/TKDE.2024.3401075.

[27] Li, J., Li, Y., Song, J., Zhang, J., & Zhang, S. (2024). Quantum Support Vector Machine for Classifying Noisy Data. IEEE Transactions on Computers, 73(9), 2233–2247. doi:10.1109/TC.2024.3416619.

[28] Li, X., Wang, J., & Yang, C. (2023). Risk prediction in financial management of listed companies based on optimized BP neural network under digital economy. Neural Computing and Applications, 35(3), 2045–2058. doi:10.1007/s00521-022-07377-0.

[29] Reyad, M., Sarhan, A. M., & Arafa, M. (2023). A modified Adam algorithm for deep neural network optimization. Neural Computing and Applications, 35(23), 17095–17112. doi:10.1007/s00521-023-08568-z.

[30] Kar, M. K., Neog, D. R., & Nath, M. K. (2023). Retinal Vessel Segmentation Using Multi-Scale Residual Convolutional Neural Network (MSR-Net) Combined with Generative Adversarial Networks. Circuits, Systems, and Signal Processing, 42(2), 1206–1235. doi:10.1007/s00034-022-02190-5.