SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2024, 21(2), 370.
https://doi.org/10.62617/mcb.v21i2.370

Article

Establish a novel framework for enhancing minority music genre identification

Lili Yan

College of Dance, Sichuan Film and Television University, Chengdu 610000, Sichuan, China; yll110480789@163.com

CITATION

Yan L. Establish a novel framework for enhancing minority music genre identification. Molecular & Cellular Biomechanics. 2024; 21(2): 370. https://doi.org/10.62617/mcb.v21i2.370

ARTICLE INFO

Received: 14 September 2024
Accepted: 30 September 2024
Available online: 6 November 2024

Abstract: The objective of this study is to develop a new framework based on Waterwheel Plant optimization for improving minority music genre classification using Layer-tuned Long Short-Term Memory (WP-LT-LSTM). Chinese minority music includes various musical styles of different ethnic groups in China. It depends on the specific instrumentalities, the distribution of pitch classes and rhythms, and the culture. Specifically, the proposed framework will enhance the ability to efficiently detect under represented music genres, which could have applicability for cultural sustainability and more personalized music recommendation services. For this, we collected a dataset that includes a wide range of different minority music samples in audio format. These include genre labels, artist information and audio features necessary for the training of our suggested model. Using K-fold cross validation to enhances the accuracy. Min-max Normalization is used on the obtained data to perform pre-processing. To extract the important features from the processed data, we used Mel-frequency cepstral coefficients (MFCCs). In our proposed model, the WP algorithm dynamically adjusts LT-LSTM's internal parameters, enhancing model adaptability. LT-LSTM processes sequential audio data, capturing temporal dependencies crucial for genre classification in minority music genres. The implemented model is executed in Python software. It evaluates the model's performance across a range of parameters throughout the result analysis phase. We also performed comparison studies using standard methods. The results collected indicate the excellence and effectiveness of the proposed framework for music genre identification.

Keywords: minority music; music genre identification; Waterwheel Plant optimization-driven Layer-tuned Long Short-Term Memory (WP-LT-LSTM); audio processing

1. Introduction

The process of defining and acknowledging ethnicity in generic studies of music involves certain issues, at the same time, it offers a perspective for more cultural understanding and research [1]. Music as a form of self-expression is a complicated language that offers great and diverse multiplication of emotions and experiences around the world. However, in this richness, there are subgenres that can be regarded as oppressed or not given much attention in everyday practice as well as in scholarly production [2]. Minority music genres refer to a multitude of musical styles that may have their origin in certain cultural, geographic, ethnic, or economic parameters. Many of these genres struggle to get recognition and find ways to preserve their music due to the aspects like poor documentation, lack of support from music institutions, or competition with popular genres [3].

Minority music genres significantly contribute to cultural memory, building community cohesion, and resisting the standardization of world music preferences masking the multifaceted features of post-minority genres [4]. Thus, through employing and synthesizing the approaches of ethnomusicology, cultural studies, and digital humanities. Music genre identification is accessible to many new forms of expansion because of technological progress [5]. There is the ability to turn to digital archives or employ computational analysis or machine learning to help compile and categorize these manifold forms of musical expression to make them more easily disseminated and understood [6]. This makes their identification and documentation informed by actual interaction with the musicians themselves, culture bearers, and other cultural leaders of the community. This not only helps to strengthen music's credibility and reliability, but also develops the concept of co-operation based on confidence and trust [7].This study's primary goal is to create a novel framework called Waterwheel Plant optimization-driven Layer-tuned Long Short-Term Memory (WP-LT-LSTM) that will improve minority music genre identification accuracy and efficiency, supporting music recommendation systems and cultural preservation.

The remaining research falls under the following categories: Section 2 discusses related works. We provide our suggested methodology in Section 3. In section 4, the outcomes of our techniques will be evaluated. A conclusion based on the data from the study is given in Section 5.

2. Related works

A customized deep neural network-based technique called PMG-Net was presented by the study [8] for automatically categorizing Persian music genres. Additionally, a dataset called PMG-Data from several Rap, Monody, Rock, Pop & Traditional genres was gathered and categorized to evaluate the PMG-Net. In study [9], content-based characteristics and a 1D-CNN were used to classify Nigerian songs into genres. Four distinct styles of Nigerian music Apala, Afro, Juju, and Fuji were included in the dataset. To improve classification accuracy, paper [10] developed a double-weighted KNN approach for music genre automatic categorization. The method addressed the shortcomings of existing algorithms by focusing on the closest samples and ignoring the association between attributes and categories. A self-supervised learning approach for music representation was presented in the research [11]. To learn about music structure, they initially used a beat-level audio pre-training model. Next, model simultaneously musical co-relations & self-representation using a multi-task learning architecture. In addition, a variety of downstream tasks, including music genre categorization, music emphasis, and music similarity retrieval, were suggested for evaluating music representation.

A music recognition approach mixed with hashing learning was proposed by the study [12], which offered the protection of copyright mechanisms for authentic digital music in the context of big data by integrating deep learning and blockchain-based systems to identify and categorize various music features. A revolutionary method known as BMNet-5 was presented bya study [13] to categorize Bangla genres of music as Palligeeti, Rabindra Sangeet, Bangla Band Music, Nazrulgeeti, Bangla Hip-hop, and Bangla Adhunik. By taking traits out of Bangla music arrangements and putting their automated classification to the test. Using an improved deep learning model, a novel automatic music genre classification technique was proposed in the paper [14]. The pair of main stages in the suggested model were extraction of features (a) and categorization (b). The most pertinent aspects, such as pitch attributes, STFT traits, and NMF amenities, were picked from the given music stream during the extraction of features. LSTM, Bi-LSTM, GRU, and Bi-GRU are examples of RNN variations and CNN hybrid architectures that were proposed in paper [15]. The performance was further contrasted using features from the Mel-spectrogram &MFCC. A hybrid model based on deep learning was presented in paper [16] to examine various genres of music. For classification, the suggested approach primarily combines the multimodal &transfer learning-based models. Research [17] proposed a suitable deep learning model as well as an efficient data augmentation strategy to attain high accuracy in classification for music genre categorization utilizing the Minimal FMA dataset. Instead of pitch shifting, it was more effective to generate an echo to enhance the data for Small FMA.A novel strategy for classifying music was proposed in study [18] that used deep neural networks for classification and music extraction. The timbre feature and melodic feature are the first two categories of characteristics that the algorithm retrieves and uses as the classification variables for music.

3. Methodology

The methodology of this study involves collecting the dataset, cleaning the data and normalizing them through Z-score normalization and extracting features by using MFCC. Using the Waterwheel Plant optimization-driven Layer-tuned Long Short-Term Memory (WP-LT-LSTM) to enhance identification for minority music genre classification. **Figure 1** shows the process of minority music genre identification.

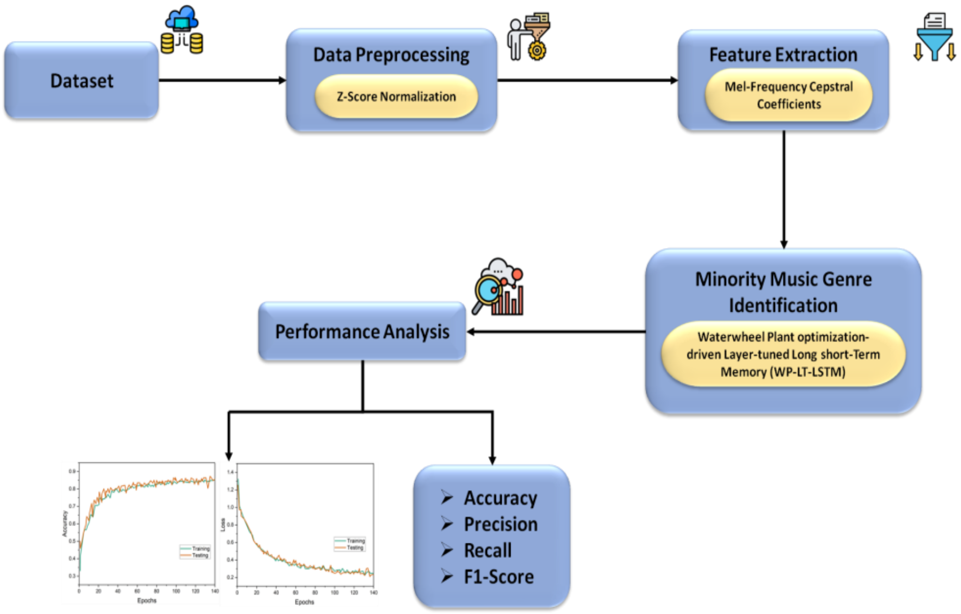

Figure 1. Minority music genre identification process.

3.1. Dataset

The dataset contains 15 ethnic minority musical genres, offering a diverse collection of choices for this analysis. Each of the given genres comprises 200 audio files, which enables examining different aspects of musicality more comprehensively. All the recordings are 30 s long thus providing standard length for comparison and analysis. overall, total 3000 music event selections which therefore provides a solid grounding for research and analysis. Procedures such as time stretching and pitch shifting can be applied to the dataset in order to introduce more changes to the audio recordings without losing important musical aspects.

3.2. Pre-processing the data

The application of Min-Max normalization for feature scaling is well regarded in the context of minority music genre identification. This method scales the values of each audio feature into the range of [0, 1] to make sure that all the features are in the same range. This is especially valid because of differences in audio features which may be highly different in different music genres in the context of music genre classification.

Through Min-Max normalization, we ensure that while preserving the relationships which exist between the data features, the features are adjusted such that they form beneficial distribution to help feed those learning algorithms. Each value in the audio feature is mapped to a new normalized value using the following Equation (1):

$$u' = \frac{u - min_X}{min_X - max_X}(new\_max_X - new\_min_X) + new\_min_X \quad (1)$$

In Min-Max normalization it is a number that resulted from the normalization of the original value u of an audio feature, and $u'$is the new normalize value. max_X is the maximum value of feature, min_X is the maximum value of feature. The range of the feature set is estimated by finding the maximum max_X and minimum min_X of new range.

3.3. K-fold cross validation

K-fold cross validation is a method that randomly divides a dataset into two parts: Training and Testing. The dataset is divided into k sub-samples, with one used for testing and the rest for training. The process is repeated *k*-times by changing the datasets, and the best model is selected based on minimum error using various statistical tools. The steps involved in k-fold cross validation are as follows.

1) Split the Data: the total of 3000 audio files that have been collected for 15 different genres of ethnic minorities. It is necessary to divide these files into *K* equal parts (folds).
2) Train and Validate: For each of the *K*-iterations, use the *K*-1 folds (2400 files) for training and the remaining fold (600 files) is used for testing the model. Repeat this all *k*-folds.
3) Calculate Average Metrics: Perform *K* iterations of the algorithm, and at the end of each iteration collect the performance outcomes such as accuracy, precision, and recall. To assess the performance of the model in detecting minority music genres.

3.4. Feature extraction

MFCC is used for feature extraction. The Mel filter serves as the basis for the MFCC concept. The ear's auditory mechanism is taken into consideration when designing the Mel filter. It takes into the manner and traits by which people produce voices and the human ear perceives them. The sound frequency of the Mel filter is known as the "Mel frequency," and it is closely associated with the actual voice rate in accordance with the auditory properties of the human ear. There is a nonlinear relationship between frequency and the Mel frequency. The precise correlation between the true frequency and the Mel frequency is as follows:

$$e_{mel} = 2595 \times \lg\left(1 + \frac{e}{700}\right) \tag{2}$$

The real voice frequency (in hertz) is denoted by e. The prediction cepstral coefficient based on Mel frequency found is called MFCC. Here's how to extract MFCC:

Pre-processing: after framing, the finite signal $w_j(m)$ is obtained, where j is the jth frame, and then next applying the FFT to every frame.

$$W(j,l) = FFT\left[w_j(m)\right] = \sum_{m=0}^{M-1} w_j(m) X_M^{kn} \tag{3}$$

The number of samples in every frame is denoted as M. After that, the power spectral is acquired:

$$F(j,l) = [W(j,l)]^2 \tag{4}$$

Next, determine the energy using the Mel filter and use DCT to get the MFCC.

$$T(j,n) = \sum_{l=0}^{M-1} F(j,l) G_n(l) \tag{5}$$

$$\sqrt{\frac{2}{N}} \sum_{l=0}^{M-1} \log[T(j,n)]\, cos\left(\frac{\pi m(n-0.5)}{N}\right)\; m = 1,2,\ldots\ldots,K \tag{6}$$

where L is typically assumed to be 12, and n is the MFCC order. The quantity of Mel filters is M.

To improve minority music genre identification, use more audio features like spectral contrast whereby small, medium and large range of amplitude differences between the peaks and valleys of the audio spectrum is calculated to detect the texture and the timbre of the music. The tonal centroid, which determines pitch and perceived harmonics, is also used in differentiating genres based on tonal comportment. Chroma descriptors store the amount of energy on each pitch class allowing for the analysis of the kind of chords and harmonic patterns typical to different styles. These features coupled with MFCC will enhance the predictions of the genre classification in addition to offering a broader view of the minority music genre.

3.5. Minority music genre identification using Waterwheel Plant optimization-driven Layer-tuned Long Short-Term Memory (WP-LT-LSTM)

Improving the Layer-tuned Long Short-Term Memory (LT-LSTM) model, the WP algorithm adapts the weights and biases in the layers using a water flow-inspired optimization mechanism. This optimization helps to improve the hyperparameters like learning rate and layer size, so that model works on temporal dependencies of audio data, and also to reduce the risk of overfitting. It thus optimizes the tuning of the model to work as a better identification system for various types of minority music genres.

3.5.1. Waterwheel Plant (WP)optimization

The foundation of the WP is the notion that we should attempt to simulate the behavior of the waterwheel plant during its hunting season in its native habitat. The way waterwheel plants locate, catch and then relocate their insect prey to a better location for feeding served as the inspiration for the main concept of WP. The concepts that guided the development of method and the mathematical framework that underpins its operations will be covered in the next section.

WP inspiration: With the help of a mucus sealant and interlock teeth, once apprehended, the victim is brought to the trap's foot, which is close to the hinge. As it digests the leftover water, the trap absorbs the majority of the nutrients from the prey. Aldrovanda traps can devour two to four meals before going inactive, just like flytraps.

WP mathematical model: The WP is a population-based strategy that is iterative and uses the members' ability to search the universe for viable solutions to find the right one. The values of the WP demographic waterwheels problem variables are determined by their position inside the search space. Every waterwheel, represents a vector-based solution. WP generates a random water wheel site assignment at the beginning of the search space using Equations (7) and (8).

$$O = \begin{bmatrix} o1,1 & o1,2 & \cdots & o1,i & \cdots & o1,n \\ o2,1 & o2,2 & \cdots & o2,i & \cdots & o2,n \\ oj,1 & oj,2 & \cdots & oj,i & \cdots & oj,n \\ oM,1 & oM,2 & \cdots & oM,i & \cdots & oM,n \end{bmatrix} \tag{7}$$

$$o_{j,i} = Ib_i + q_{j,i}(ub_i - Ib_i), j = 1{:}M, i = 1{:}n \tag{8}$$

In this case, Mrepresents the number of waterwheels, while m represents the number of variables. A selection ranging from zero to one represents the amount of $q_{j,i}$for the jth element in the problem, which has two limits: Ib_i and ub_i. O Stands for population matrix, which contains the waterwheel positions. Where each row o_j indicates a potential solution.

Equation (9) has established in prior work that the values that make up the problem's objective function can be efficiently expressed using a vector.

$$E = \begin{bmatrix} E_1 \\ \vdots \\ E_j \\ \vdots \\ E_M \end{bmatrix} = \begin{bmatrix} E(W_1) \\ \cdots \\ E(Wj) \\ \cdots \\ E(W_M) \end{bmatrix} \tag{9}$$

The objective function values are displayed as a vector E, in which the expected outcome for the jth waterwheel is denoted by E_j. The best responses are determined by analyses of the objective functions. As a consequence, the optimal candidate solution is represented by the function's goal with the highest value, while the worst candidate solution is represented by the function with the lowest value. The optimal response varies as more iteration is carried out due to the waterwheels' variable rates of movement throughout the search space.

Exploration and Exploitation: Waterwheels have an innate aggressiveness that helps them tofind and efficiently hunt pests due to their advanced sense of smell. When the waterwheel is detected, a bug will immediately attack it and use its location to monitor its victim. The WP models initial phase of its population updating mechanism using a hunting behavior simulation. The WP is better able to discover the optimal region and stays out of local optima when the waterwheel is used to combat the insect. Equations (10) and (11) show that if moving the waterwheel to a new position increases the value of the goal function, the old site disappears in preference to the new one.

$$\vec{X} = \overrightarrow{q_1}(\vec{O}(s) + 2L) \tag{10}$$

$$\vec{O}_{s+1} = \vec{O}_{(s)} + \vec{X}(2L + \overrightarrow{q_2}) \tag{11}$$

Equation (12) can be utilized to modify the waterwheel's location if the solution is not improved upon after three consecutive attempts.

$$\vec{O}_{s+1} = Gaussian(\mu_o, \sigma) + \overrightarrow{q_1}\vec{O}(s) + 2L_{\vec{X}} \tag{12}$$

The random variables $\overrightarrow{q_1}$and $\overrightarrow{q_2}$in this context have values between 0 and 2 and between 0 and 1, respectively. Lalso has values between 0 and 1, making it an exponential variable. The waterwheel module is going to examine an annulus with a circumference of $\vec{X}$ to find potentially feasible places.

The second phase of the WP population update model mimics waterwheels gathering and moving insects into a feeding tube. To converge on acceptable answers that are close to found ones, WP makes use of this simulated trend during local search. By mimicking the insect's path to the tube, the waterwheel adjusts the position of its search space. To replicate the normal actions of waterwheels, the WP manufacturers initiate the water wheel in a group with a random "good location for consuming insects." If the desired feature value is higher, the waterwheel is shifted to the new location, as shown by Equations (13) and (14) where the previous location is kept.

$$\vec{X} = \overrightarrow{q_3} \times \left(L\overrightarrow{O_{best}}(s) + q_3\vec{O}(s)\right) \tag{13}$$

$$\vec{O}(s+1) = \vec{O}(s) + L\,\vec{X} \tag{14}$$

The random variable in this case is denoted as $\overrightarrow{q_3}$ whose values alternate between 0 and 2. The answer at iteration $s$is denoted by $\vec{O}(s)$, and the best solution thus far is represented by $\overrightarrow{O_{best}}$. Equation (15) illustrates the next mutation that is performed to ensure that local minima are prevented, similar to the exploration phase that was previously described if the solution is unable to advance after three iterations.

$$O(s+1) = (q_1 + L)\sin\left(\frac{ED}{\theta}\right) \tag{15}$$

when the values of E and D, two independent random variables, range from $[-5,5]$. Furthermore, the exponential decrease of L, represented by Equation (16), can be demonstrated using the following equation.

$$L = \left(1 + \frac{2s^2}{S_{max}^3} + E\right) \tag{16}$$

The WP is offered an approach that can be used again. Rearranging all of the waterwheels is the third and last step in the WP implementation process. Following the completion of the first two stages comes this phase. The applicants for the best solution are refined when the target function's values are compared. The procedure proceeds until the procedure reaches its final iteration, during the position of the waterwheels is adjusted for the next iteration.

3.5.2. Layer-tuned Long Short-Term Memory (LT-LSTM)

An LSTM is a variation that is dependent on RNN. To accomplish long-term transmission and information memorization, this model's hidden layer structure is enhanced with gate units and cell states that regulate the flow of memory data. **Figure 2** presents the flow of LT-LSTM.

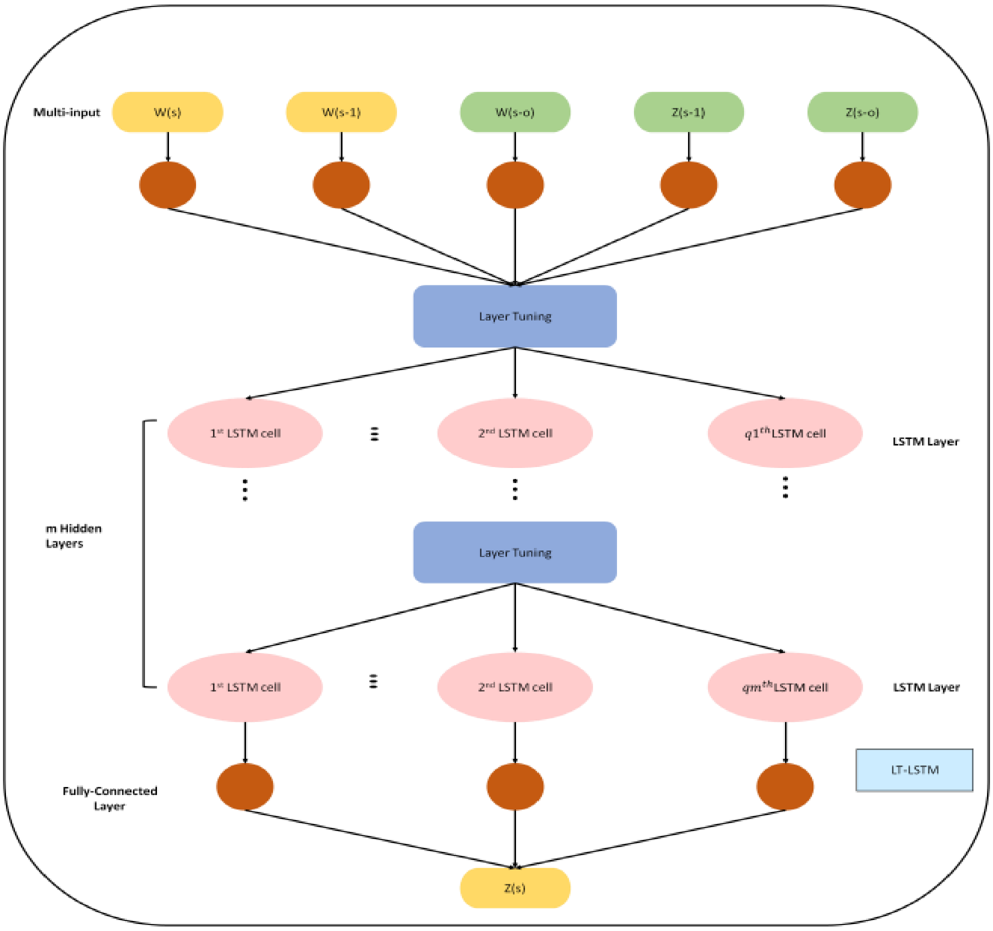

Figure 2. Flow of the LT-LSTM.

This model can carry information over long time steps and more appropriate than RNN to process sequences containing long time information. When the input variable at time step $s$is represented by w_s, the framework of the LSTM model is revealed by the time dimension. The hidden & cell state vectors are indicated, respectively, by g_s and d_s.

Filtering and selectively managing the information that flows through the gate structure, which is a completely connected layer. In the model, the gate units regulate the insertion, removal, modification and transmission of data to other cells. The following are the fundamental equations governing the LSTM network's unit updating procedure:

$$\begin{pmatrix} e_s \\ j_s \\ p_s \\ h_s \end{pmatrix} = X_g g_{s-1} + X_w W_s + a \tag{17}$$

$$d_s = \sigma(e_s) \odot d_{s-1} + \sigma(j_s) \odot \tanh(h_s) \tag{18}$$

$$g_s = \sigma(p_s) \odot \tanh(d_s) \tag{19}$$

Input, forget, & output gates are represented by the symbols j_j, e_j, and p_s, respectively. g_s Stands for hidden state; d_s for cell state; h_s for candidate cell state; s is the current time step. a is bias, σ is a sigmoidal function, X_g is the recurrent weights from a specific hidden layer to a different hidden layer, X_w is the bottom-up inputs to hidden weight, ʘ and stands for element-wise multiplication.

Layer Tuning in the LSTM System: The development of layer Tuning (LT) aims to overcome the problem of internal correlate displacement in the RNN model and speed up network convergence. LT is a horizontal normalizing technique that relies on batch normalization. It takes into account all input dimensions in each network layer, and it ensures that neurons in the same layer have identical mean and variance. Every dimension's input is converted using the same normalization process. Definition of LT: A function mappings LT: using two adaptive parameter sets, gains (α) & biases (β), is used to demonstrate the LT is applied to an LSTM model.

$$LN(y, \alpha, \beta) = \frac{(y-\mu)}{\sigma} \odot \alpha + \beta \tag{20}$$

$$\mu = \frac{1}{C} \sum_{j=1}^{C} y_j \tag{21}$$

$$\sigma = \sqrt{\frac{1}{C} \sum_{j=1}^{C} (y_j - \mu)^2} \tag{22}$$

where the element-wise multiplication of two vectors is represented by the symbol $\odot$, C is the number of hidden units, and y_j is the jth element of vector 2.

The information from the nodes will always become normalized statistics in an LSTM time step since LT can be applied to each time step. The expression for LSTM integrated with LT is as follows:

$$\begin{pmatrix} e_s \\ j_s \\ p_s \\ h_s \end{pmatrix} = LN\left(X_g g_{s-1}; \propto_1, \beta_1\right) = LN(X_w W_s; \alpha_2, \beta_2) + a \tag{23}$$

$$d_s = \sigma(e_s) \odot d_{s-1} + \sigma(j_s) \odot \tanh(h_s) \tag{24}$$

$$g_s = \sigma(p_s) \tanh \odot \left(LN(d_s; \alpha_3, \beta_3)\right) \tag{25}$$

The values β_j and $\propto_j$ are multiplicative and additive, respectively. A vector is initialized for each β_j, and an array of zeros is initialized for each $\propto_j$. The LSTM network's other parameters are comparable to its own.

3.5.3. Proposed method

The 'Waterwheel Plant optimization-driven Layer-tuned Long Short-Term Memory (WP-LT-LSTM) for enhancing minority music genre identification' offers a solution for accurate classifications of minority music genres which are usually overlooked in the classification of music genres. The method involves an optimization algorithm known as the Waterwheel Plant (WP) combined with a Layer-tuned LSTM neural network. The WP algorithm modifies the arrangement of the LSTM layers to improve their functions. RNN is one of the most commonly used architectures for sequence-to-sequence tasks, one of the most effective of them being the LSTM network, particularly for the identification of the music genre because it can retain and learn temporal dependencies in a music track. In this approach, the WP algorithm is employed to make a fine tune of the LSTM model in terms of the number of layers, number of neurons per layer, and learning rate. The specific focus on the architecture of the LSTM is the WP algorithm's goal of developing the ideal model for distinguishing between minority genres, which differ in small details. To develop the WP-LT-LSTM model, various types of data are used, which allows to broaden the coverage of songs with a diverse spectrum of genres but also to maximize the model's exposure to different features of musical output. This is necessary so that the model can train from this exposure and evaluate its generalization on the features of minority genres. WP-LT-LSTM method was developed to enhance the performance of the classification of music genres since it deals with the low rate of identification of the minority genre as a result of overfitting problems in traditional models of music analysis. It could be useful for increasing the accessibility and the quality of recommendations of music and the identification of genres.

4. Result

Use evaluation criteria like recall, accuracy, F1-score, and precision to assess the efficacy of our suggested strategy. In comparison with existing techniques such as NB, RF, LR, & DT [19].

4.1. Experimental setup

Tasks were completed with Python 3.12 on Windows 11. As the processor, it was an Intel Core i7 12th Gen, while the memory amounted to 32 GB. The tested hardware comprised a modern laptop configuration that would allow for capturing performance data for heavy multitasking and development workloads.

The **Figure 3** shows the dominant features for different music genres and which features contribute most to a model that can predict the minority music genres. The importance scores for genre classification indicate that genres are essential for achieving good classification accuracy. It is thus proposed that a framework of such features can help improve the accuracy by which minority music genres can be characterized.

Predictive accuracy measures how accurate the model is matching its estimated values to its actual values. The model's extremely accurate prediction rate indicates a safe and accurate prognosis. Losses are defined as the discrepancy between expected and actual results. It determines how erroneous the model is. Through the use of a predetermined metric to minimize loss, the model can generate projections that are more accurate. The accuracy and loss outcomes are displayed in **Figure 4**.

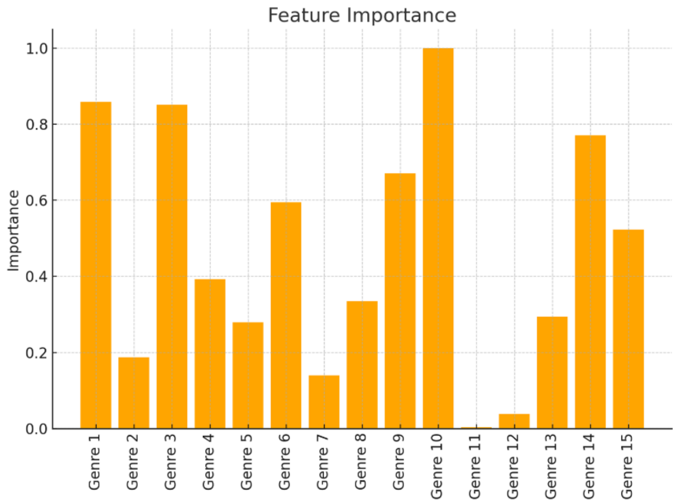

Figure 3. Outcome of feature importance score.

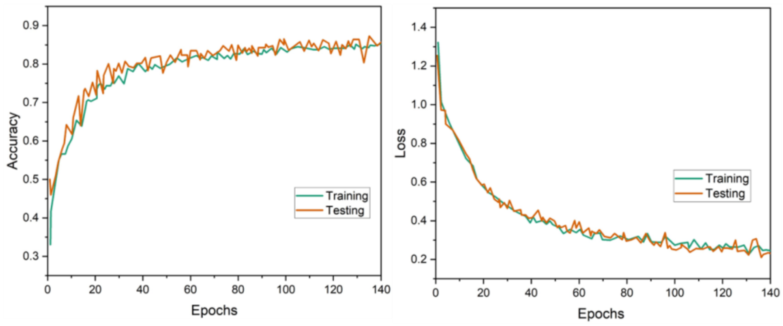

Figure 4. Accuracy and loss results.

The ROC (Receiver Operating Characteristic) is the graphical representation of real positive rate and false positive rate within the model classification at varying thresholds. For improving minority music genre identification, assist in qualifying or quantifying its effectiveness in identifying the various genres. **Figure 5** result of ROC.

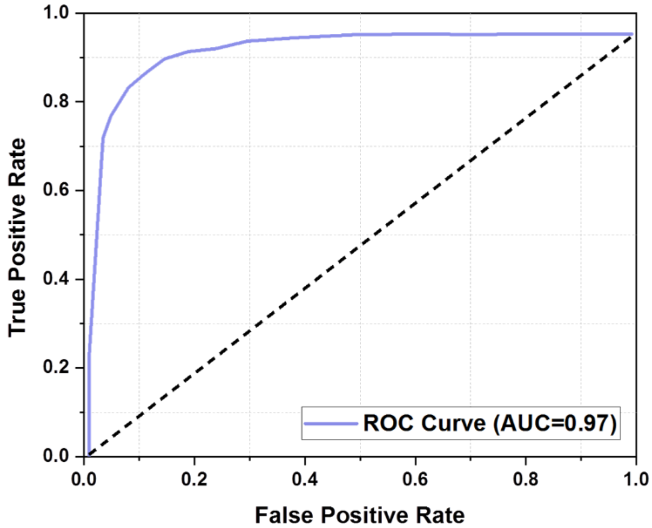

Figure 5. Outcome of ROC.

Accuracy is calculated as the percentage of correct minority music genre forecasts. It is calculated as the proportion of all forecasts that are accurate. **Figure 6** shows the accuracy outcome. The WP-LT-LSTM (0.93), significantly outperforms traditional approaches such NB (0.42), RF (0.9), LR (0.72), and DT (0.61) minority music genre identification. This demonstrate the effectiveness of WP optimization in improving classification performance.

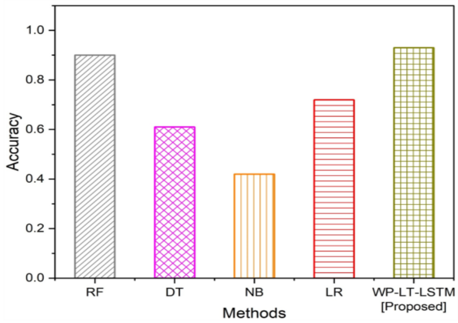

Figure 6. Outcomes of accuracy.

Precision is the percentage of minority music genre occurrences accurately detected out of all instances the model classifies as belonging to that genre. **Figure 7** displays the precision result. In comparison to traditional methods like NB (0.28), RF (0.9), LR (0.75), and DT (0.67), the suggested WP-LT-LSTM achieved a precision rate (0.91) that is significantly higher. this demonstrates the WP-LT-LSTM model's better performance in precisely recognising minor music genres.

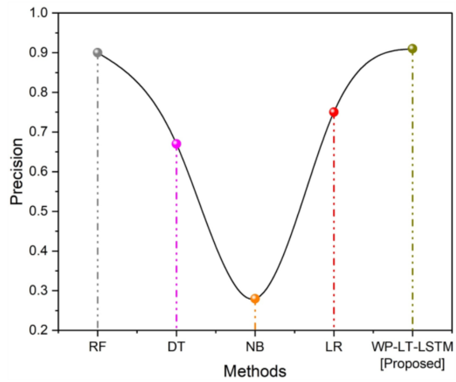

Figure 7. Outcomes of precision.

The percentage of accurately detected examples of a minority music genre out of all instances that belong to that genre is known as recall. Conventional techniques such as NB (0.34), RF (0.91), LR (0.84), and DT (0.67), respectively, were discovered to be effective. By contrast, the recall rate (0.92) of the suggested approach is noticeably greater. This illustrates how well the WP-LT-LSTM model performs in accurately identifying small musical genres. **Figure 8** shows the result of Recall.

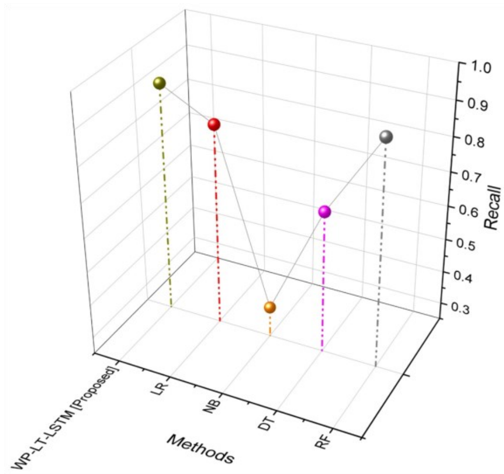

Figure 8. Outcomes of recall.

The F1-score, which is the harmonic mean of recall and precision, is a statistic that offers a balance between two. The results demonstrated the effectiveness of conventional methods such as NB (0.31), RF (0.91), LR (0.78), and DT (0.67). In comparison, the recommended approach achieves a substantially higher recall rate (0.92). This illustrates how well the WP-LT-LSTM model can distinguish minor musical genres. **Figure 9** displays the result of F1-Score. **Table 1** shows that the overall outcomes of various parameters.

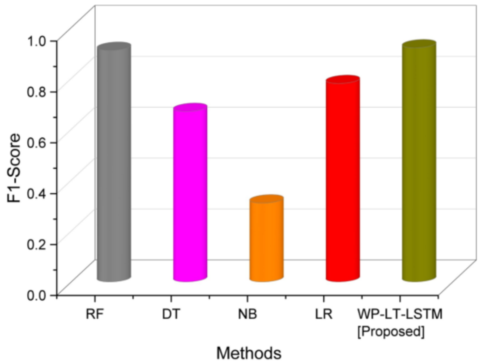

Figure 9. Outcomes of F1-score.

Table 1. Numerical outcomes of parameters.

Methods	Recall	Accuracy	F1-Score	Precision
RF	0.91	0.9	0.91	0.9
DT	0.67	0.61	0.67	0.67
NB	0.34	0.42	0.31	0.28
LR	0.84	0.72	0.78	0.75
WP-LT-LSTM [Proposed]	0.92	0.93	0.92	0.91

5. Discussion

When comparing traditional methods such as NB, RF, LR, and DT have numerous significant differences. Even though methods such as NB and LR [19] were easily interpretable and intuitive, they are not suitable when it comes to the complex structures in the minority music genrs and high-dimensional audio characteristics. RF [19], while highly reliable and natural to deal with the problem of overfitting using ensemble learning, could be inferior in terms of temporal dynamics in sequential data such as audio. DT [19], were quite easy to implement because of their simple decision making strategy, but they over-fit the training data and do not generalize well on the new unseen data. However, our proposed WP-LT-LSTM model not only retains temporal dependencies of the audio sequences due to the LT-LSTM layer but also the internal parameters of the model could be optimized by the Waterwheel Plant optimization algorithm for the better classification of the audio signals. The results in significantly improved detection of underrepresented music genres showing the effectiveness of combining optimization approaches with deep learning to improve the music genre classification, especially in culturally diverse but less represented areas.

6. Conclusion

Minority music genre identification refers to the identification, categorization, and analysis of distinct and scarce music genres. It promotes multiculturalism in musical databases and serves the cause of conserving and popularizing specific genres or national cultures. The proposed WP-LT-LSTM framework introduced in this research greatly contributes to enhancing minority music genre identification. When improving LT-LSTM using WP optimization, we use the distinction of the Multi-dimensional model WP-LT-LSTM to identify various and minority music in Chinese music with higher accuracy and efficiency. The effectiveness of Mel-frequency cepstral coefficients (MFCCs) and Z-score Normalization in the pre-processing of data provides accurate extraction of features and normalization of data respectively. Further, the dynamic update of LT-LSTM parameters using the WP algorithm allows for better adaptability while addressing complex temporal structures that are fundamental for genre distinction. Python is used to implement the proposed approach. Compared with the traditional approach, our approach is superior based on qualitative measures of recall (0.92), accuracy (0.93), F1-score (0.92), and precision (0.91). Some of the limitations that affect the WP-LT-LSTM method are that it may not perform well in cross-genres and it may solve a subset of simple cases. Future work can extend WP-LT-LSTM to other genres as well, and such extensions would further increase the overall effectiveness and usability of the system.

Ethical approval: No applicable.

Conflict of interest: The author declares no conflict of interest.

Nomenclature

KNN	k-nearest neighbor
BMNet-5	Bengali Music Classification Using Neural Network
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
BI-GRU	Bidirectional Gated Recurrent Unit
NB	Naïve Bayes
MFCC	Mel-frequency cepstral coefficient
STFT	Short-Time Fourier Transform
FMA	Free Music Archive
DCT	Discrete Cosine Transform
RNN	Recurrent Neural Network
FFT	Fast Fourier Transform
NMF	Non-Negative Matrix Factorization
DT	Decision Tree
Bi-LSTM	Bidirectional Long Short-Term Memory
LR	Logistic Regression
GRU	Gated Recurrent Unit
1D-CNN	1 dimensional- Convolutional Neural Network
RF	Random Forest
PMG-Net	Persian Music Genre Using Neural Networks

References

1. Curran, G. and Radhakrishnan, M. (2021). The Value of Ethnographic Research on Music: An Introduction. The Asia Pacific Journal of Anthropology, 22(2-3), pp.101-118. https://doi.org/10.1080/14442213.2021.1913511
2. Kaimal, G., Carroll-Haskins, K., Ramakrishnan, A., Magsamen, S., Arslanbek, A. and Herres, J.(2020). Outcomes of visual self-expression in virtual reality on psychosocial well-being with the inclusion of a fragrance stimulus: A pilot mixed-methods study. Frontiers in Psychology, 11, p.589461.https://doi.org/10.3389/fpsyg.2020.589461

3. Hoad, C. and Hoad, C. (2021). Mapping Representation in Metal Music Studies. Heavy Metal Music, Texts, and Nationhood: (Re) sounding Whiteness, pp.17-58.https://doi.org/10.1007/978-3-030-67619-3_2
4. Lerch, A., Arthur, C., Pati, A. and Gururani, S.(2021). An interdisciplinary review of music performance analysis. arXiv preprint arXiv:2104.09018.https://doi.org/10.48550/arXiv.2104.09018
5. Singh, J. and Bohat, V.K.(2021). A neural network model for recommending music based on music genres. In 2021 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-6). IEEE.https://doi.org/10.1109/ICCCI50826.2021.9402621
6. Georges, P. and Seckin, A.(2022). Music information visualization and classical composers discovery: an application of network graphs, multidimensional scaling, and support vector machines. Scientometrics, 127(5), pp.2277-2311.https://doi.org/10.1007/s11192-022-04331-8
7. MacHaffie, J.(2021). Mutual trust without a strong collective identity? Examining the Shanghai cooperation organization as a nascent security community. Asian Security, 17(3), pp.349-365.https://doi.org/10.1080/14799855.2021.1895115
8. Farajzadeh, N., Sadeghzadeh, N. and Hashemzadeh, M.(2023). PMG-Net: Persian music genre classification using deep neural networks. Entertainment Computing, 44, p.100518. https://doi.org/10.1016/j.entcom.2022.100518
9. Falola, P.B. and Akinola, S.O.(2021). Music genre classification using 1D convolution neural network. International Journal of Human Computing Studies, 3(6), pp.3-21. https://journals.researchparks.org/index.php/IJHCS
10. Wu, M. and Liu, X.(2020). A double-weighted KNN algorithm and its application in the music genre classification. In 2019 6th International Conference on Dependable Systems and Their Applications (DSA) (pp. 335-340). IEEE. https://doi.org/10.1109/DSA.2019.00051
11. Zhu, H., Niu, Y., Fu, D. and Wang, H.(2021). MusicBERT: A self-supervised learning of music representation. In Proceedings of the 29th ACM International Conference on Multimedia (pp. 3955-3963). https://doi.org/10.1145/3474085.3475576
12. Chen, X., Qu, X., Qian, Y. and Zhang, Y.(2022). Music recognition using blockchain technology and deep learning. Computational Intelligence and Neuroscience, 2022(1), p.7025338. https://doi.org/10.1155/2022/7025338
13. Hasib, K.M., Tanzim, A., Shin, J., Faruk, K.O., Al Mahmud, J. and Mridha, M.F.(2022). Bmnet-5: A novel approach of a neural network to classify the genre of Bengali music based on audio features. IEEE Access, 10, pp.108545-108563. https://doi.org/10.1109/ACCESS.2022.3213818
14. Kumaraswamy, B.(2022). Optimized deep learning for genre classification via improved moth flame algorithm. Multimedia Tools and Applications, 81(12), pp.17071-17093. https://doi.org/10.1007/s11042-022-12254-y
15. Ashraf, M., Abid, F., Din, I.U., Rasheed, J., Yesiltepe, M., Yeo, S.F. and Ersoy, M.T.(2023). A hybrid cnn and rnn variant model for music classification. Applied Sciences, 13(3), p.1476. https://doi.org/10.3390/app13031476
16. Jena, K.K., Bhoi, S.K., Mohapatra, S. and Bakshi, S.(2023). A hybrid deep learning approach for classification of music genres using wavelet and spectrogram analysis. Neural Computing and Applications, 35(15), pp.11223-11248.https://doi.org/10.1007/s00521-023-08294-6
17. Le Thuy, D.T., Van Loan, T., Thanh, C.B. and Cuong, N.H.(2023). Music Genre Classification Using DenseNet and Data Augmentation. Computer Systems Science & Engineering, 47(1).https://doi.org/10.32604/csse.2023.036858
18. Zhang, K.(2021). Music style classification algorithm based on music feature extraction and deep neural network. Wireless Communications and Mobile Computing, 2021(1), p.9298654.https://doi.org/10.1155/2021/9298654
19. Chaudhury, M., Karami, A. and Ghazanfar, M.A.(2022). Large-Scale Music Genre Analysis and Classification Using Machine Learning with Apache Spark. Electronics, 11(16), p.2567.https://doi.org/10.3390/electronics11162567