

# Improved Broad Learning System for Birdsong Recognition

Jing Lu <sup>1</sup>, Yan Zhang <sup>2,\*</sup>, Danjv Lv <sup>1</sup>, Shanshan Xie <sup>3</sup>, Yixing Fu <sup>1</sup>, Dan Lv <sup>1</sup>, Youjie Zhao <sup>1</sup> and Zhun Li <sup>1</sup>

<sup>1</sup> College of Big Data and Intelligent Engineering, Southwest Forestry University, Kunming 650224, China; lujing@swfu.edu.cn (J.L.); lvdanjv@swfu.edu.cn (D.L.); fuyixing@swfu.edu.cn (Y.F.); ld583265@swfu.edu.cn (D.L.); bioala@swfu.edu.cn (Y.Z.); 839383767@swfu.edu.cn (Z.L.)

<sup>2</sup> College of Mathematics and Physics, Southwest Forestry University, Kunming 650224, China

<sup>3</sup> College of Engineering, Beijing Forestry University, Beijing 100091, China; xieshanshan@swfu.edu.cn

\* Correspondence: zhangyan@swfu.edu.cn

**Abstract:** Birds play a vital and indispensable role in biodiversity and environmental conservation. Protecting bird diversity is crucial for maintaining the balance of nature, promoting ecosystem health, and ensuring sustainable development. The Broad Learning System (BLS) exhibits an excellent ability to extract highly discriminative features from raw inputs and construct complex feature representations by combining feature nodes and enhancement nodes, thereby enabling effective recognition and classification of various birdsongs. However, within the BLS, the selection of feature nodes and enhancement nodes assumes critical significance, yet the model lacks the capability to identify high quality network nodes. To address this issue, this paper proposes a novel method that introduces residual blocks and Mutual Similarity Criterion (MSC) layers into BLS to form an improved BLS (RMSC-BLS), which makes it easier for BLS to automatically select optimal features related to output. Experimental results demonstrate the accuracy of the RMSC-BLS model for the three construction features of MFCC,  $d_{MFCC}$ , and  $d_{sequence}$  is 78.85%, 79.29%, and 92.37%, respectively, which is 4.08%, 4.50%, and 2.38% higher than that of original BLS model. In addition, compared with other models, our RMSC-BLS model shows superior recognition performance, has higher stability and better generalization ability, and provides an effective solution for birdsong recognition.

**Keywords:** broad learning system; birdsong recognition; feature sequence; residual blocks; mutual similarity criterion



**Citation:** Lu, J.; Zhang, Y.; Lv, D.; Xie, S.; Fu, Y.; Lv, D.; Zhao, Y.; Li, Z. Improved Broad Learning System for Birdsong Recognition. *Appl. Sci.* **2023**, *13*, 11009. <https://doi.org/10.3390/app131911009>

Academic Editor: Antonio Fernández-Caballero

Received: 29 August 2023  
Revised: 29 September 2023  
Accepted: 3 October 2023  
Published: 6 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Biodiversity is a fundamental and critical biological component of our planet, including all forms of life in ecosystems, from microorganisms to large organisms. The study of biodiversity is of great significance as it helps to maintain ecological balance, provides key ecosystem services, and promotes sustainable development [1]. Ecological monitoring plays an indispensable role in understanding, conserving, and advancing biodiversity. It also provides essential scientific evidence and decision support for environmental protection and the maintenance of ecological equilibrium. In this regard, birds are highly valuable indicators of ecosystem health due to their sensitivity to climate, habitat quality, and ecological disturbances. Compared to other animal groups, birds are relatively accessible for observation and research, which makes them particularly suitable for ecological assessments [2]. As an important part of the ecosystem, birds play a vital role in ecological balance by controlling pests and participating in pollination and seed dispersal. Bird diversity monitoring and protection can assess the overall health status of ecosystem, help to resist species invasion, protect endangered species, and provide scientific basis for formulating environmental protection policies and management measures. Birdsong is a form of vocal expression in birds, containing rich information such as species identification, individual identification, gender, and behavior. The study of birdsong holds significant importance in the fields of biodiversity, ecology, and behavior [3]. Birdsong classification

can accurately grasp the status and dynamic changes of bird populations, allowing for biodiversity monitoring and species identification, evaluation of ecosystem health and function, prioritization of conservation work, and detection of invasive bird species with unique songs. Ultimately, the goal of protecting biodiversity and environmental sustainability is achieved.

Feature extraction and pattern recognition techniques have found widespread application in the classification of birdsongs for achieving automated bird recognition [4–8]. This approach has achieved some success to a certain extent, but it also has some limitations. Manual feature extraction requires the expertise and experience of domain experts, and it is difficult to capture the underlying complex patterns and information in birdsongs [9]. Traditional machine learning algorithms also have certain limitations in dealing with high dimensional, non-linear, and dynamic features. In recent years, deep learning has made significant breakthroughs and has demonstrated powerful expressive and generalization abilities in areas such as image and speech recognition [10,11]. Models based on deep learning are suitable for extracting features from large-scale datasets, usually consisting of a large number of neurons and multi-layer networks. This requires a large amount of training data and computing resources for training and reasoning, which is challenging in resource constrained or real-time applications [12].

The Broad Learning System (BLS) provides an alternative approach to deep learning networks [13]. By expanding the width of network layers, the expressive capacity of the model is enhanced, thus improving its ability to learn complex patterns and information. Compared with deep networks, BLS has fewer parameters and a lower computational burden, is easier to train and optimize, and can exhibit better performance using limited data [14]. BLS has been widely applied in machine learning for tasks such as image recognition [15,16], food safety [17], and biomedicine [18]. However, in research fields such as text and speech processing, the application of BLS is relatively limited. BLS learns the input features using a hierarchical architecture, where each layer processes the features and extracts higher-level representations. In addition, BLS can process a large amount of audio data, has good scalability and efficiency, and is suitable for real-time or large-scale bird species recognition applications. By fully utilizing the capabilities of BLS, researchers can enhance their work in monitoring bird populations and studying their behavior, thereby contributing to the protection of biodiversity and ecosystem health. While BLS has been proven to be a powerful modeling structure, it still faces some challenges. The performance of BLS is influenced by the number and selection of nodes, and it remains to be solved [19]. Furthermore, adding more nodes to the intermediate hidden layers of BLS models for optimization purposes could lead to overfitting issues.

To address the limitations of the BLS model, this study focused on analyzing 16 types of birdsongs. Differential Mel-Frequency Cepstral Coefficients (MFCC) sequence feature ( $d_{sequence}$ ) is extracted and constructed. To further enhance the performance of the BLS model, residual blocks are incorporated. Moreover, nodes selection is performed by measuring the correlation between nodes and class. Subsequently, a comprehensive birdsong recognition study is conducted utilizing the proposed approach. The aims and contributions of this paper can be summarized as follows:

- (1) Build the improved BLS model for birdsongs recognition.
- (2) Integrate residual block into the BLS, enabling the model to learn the residual information of the differences between feature nodes and enhancement nodes instead of directly learning the mapping. This makes the network more capable of learning and optimizing complex functions.
- (3) Employ a mutual similarity criterion to measure the correlation between nodes and the class in BLS for nodes selection, enhancing the quality of node subset.

The rest of this paper is organized as follows. Section 2 presents the related work conducted by various authors for birdsong classification. Section 3 describes the data used in the experiment. Section 4 introduces the proposed method in detail. Then, the

experimental results and discussion are analyzed in Section 5. Section 6 gives limitations and scope. Finally, Section 7 concludes the study.

## 2. Related Work

Birdsong recognition mainly relies on the selection of specific audio features from bird vocalizations for identification purposes. These audio features capture distinctive patterns and characteristics inherent in the birdsongs, enabling the recognition system to differentiate between different bird species based on their vocalizations. Common audio features for birdsong recognition include Mel-Frequency Cepstral Coefficients (MFCC) [20], Mel-Frequency Spectrum Coefficients (MFC), Linear Prediction Coefficient (LPC) [21] and deep features [22,23], which represent the spectral content of audio signals. For example, Stowell et al. conducted unsupervised learning on MFCC, MFC, and other features extracted from birdsongs, and used a random forest classifier for classification [24]. Wang et al. focused on eight bird species, extracted MFCC feature parameters, and utilized the dual Gaussian mixture model for training and recognition [25]. In addition, Xu et al. studied 11 bird species and employed DTW templates based on syllable length, MFCC, and LPC, combining these with time-frequency texture features and multi-label classifiers for birdsong recognition [26]. With the widespread application of deep learning in the field of audio processing, researchers have begun to explore the application of deep learning technology for birdsong recognition. Yan et al. combined chroma, logmel spectrogram, and MFCC features, and used three-dimensional convolutional neural networks and long short-term memory networks as classifiers to realize birdsong classification [27]. Jisheng Bai et al. extracted logmel or loglinear spectra as features in BirdCLEF 2020 and used data augmentation technology to improve the detection performance of birdsongs, achieving good results [28]. Gupta et al. proposed a deep learning method for large-scale prediction and analysis of the acoustics of 100 different bird species [29]. Overall, these studies have achieved good results in birdsong recognition through the strategic use of audio features and classification techniques. The research progress in the field of birdsong recognition highlights the significance of using audio feature extraction and classification techniques for automatic identification of birdsongs. However, there are still some challenges in this area. Although neural networks perform well in processing complex data and tasks, their interpretability is relatively poor due to their highly nonlinear structure and large number of parameters. Thus, the neural network requires a large number of training sets to obtain a good model.

There is a significant difference in the structure between BLS and deep neural networks, as BLS tends to construct the network horizontally compared to deep structures. As a shallow neural network structure, BLS has been widely applied in various research fields since its inception. For instance, it performs well in tasks such as shrimp freshness detection [30], mushroom toxicity assessment [17], fatigue driving detection [31], and material recognition [32]. Nevertheless, with the widespread application of BLS, some issues have been discovered, and the method needs to be improved. Many researchers have made significant contributions in this regard. Huang et al. proposed a feature selection algorithm of an orthogonal BLS model based on mutual information by analyzing the mutual information between feature nodes and output nodes [33]. Zheng et al. introduced the maximum correlated entropy criterion into the BLS to train output weights, including the impact of outliers on the modeling performance and enhancing the robustness of the model [34]. Zhang et al. proposed an effective BLS with rich feature combinations, utilizing unsupervised grouping encoding for feature extraction and extensive feature fusion [35]. Ye et al. presented a regularized deep cascaded BLS, consisting of a cascaded feature mapping node layer and a cascaded enhancement node layer. They also designed a parallel framework to make it compatible with large-scale data and achieved significant success in image denoising [36]. Han et al. combined structured manifold learning with BLS to propose an efficient framework for large-scale time series processing. This framework can discover crucial intrinsic structures in time series, making the modeling more interpretable

than traditional methods [37]. Ding et al. developed a BLS architecture search method, which not only improved the efficiency of network architecture search but also showed good performance. Additionally, they analyzed its general approximation ability and provided two extended variants, enriching the neural network architecture search system [38]. Xie et al. proposed a stacked structure of the BLS, which utilized residual expansion of the original input space. This stacked structure can enhance system performance and simplify its architecture [39].

Despite some success with BLS, little attention and research have been paid to the issues of information redundancy and loss during model expansion. Ma J et al. have demonstrated that BLS suffers from information redundancy and loss between original and newly added nodes, which effects its performance in large scale expansion [40]. These issues may lead to overfitting on data features during model design and training, resulting in reduced generalization performance of the model on new data. Therefore, to further improve and optimize BLS, it is essential to conduct in-depth research on these problems and propose corresponding solutions to enhance its performance and robustness. Addressing these challenges is critical to achieving better performance and broader applicability of the BLS.

### 3. Dataset

The dataset used in this study mainly comes from the online resource <https://www.xeno-canto.org/> and <http://www.birder.cn/> (accessed on 13 October 2021) A total of 271 audio samples were collected, representing 16 bird species distributed in 7 orders, 9 families, and 15 genera. Each dataset is divided into a training set and a test set at the ratio of 8:2. Detailed information is presented in Table 1 below.

**Table 1.** The information of dataset.

ID	Genus	Family	Order	Latin Name	Training Set	Test Set	Total
1	<i>Francolinus</i>	<i>Phasianidae</i>	<i>Galliformes</i>	<i>Francolinus pintadeanus</i>	552	138	690
2	<i>Coturnix</i>	<i>Phasianidae</i>	<i>Galliformes</i>	<i>Coturnix</i>	1093	274	1367
3	<i>Phasianus</i>	<i>Phasianidae</i>	<i>Galliformes</i>	<i>Phasianus colchicus</i>	800	201	1001
4	<i>Lagopus</i>	<i>Phasianidae</i>	<i>Galliformes</i>	<i>Lagopus muta</i>	778	195	973
5	<i>Lyrurus</i>	<i>Phasianidae</i>	<i>Galliformes</i>	<i>Lyrurus tetrrix</i>	944	236	1180
6	<i>Cygnus</i>	<i>Anatidae</i>	<i>Anseriformes</i>	<i>Cygnus cygnus</i>	908	227	1135
7	<i>Asio</i>	<i>Strigidae</i>	<i>Strigiformes</i>	<i>Asio otus</i>	556	140	696
8	<i>Grus</i>	<i>Gruidae</i>	<i>Gruiformes</i>	<i>Grus grus</i>	606	152	758
9	<i>Numenius</i>	<i>Scolopacidae</i>	<i>Charadriiformes</i>	<i>Numenius phaeopus</i>	1441	361	1802
10	<i>Larus</i>	<i>Laridae</i>	<i>Charadriiformes</i>	<i>Larus canus</i>	553	139	692
11	<i>Accipiter</i>	<i>Accipitridae</i>	<i>Ciconiiformes</i>	<i>Accipiter nisus</i>	852	214	1066
12	<i>Accipiter</i>	<i>Accipitridae</i>	<i>Ciconiiformes</i>	<i>Accipiter gentilis</i>	517	130	647
13	<i>Falcons</i>	<i>Falconidae</i>	<i>Falconiformes</i>	<i>Falco tinnunculus</i>	642	161	803
14	<i>Phylloscopus</i>	<i>Sylviidae</i>	<i>Passeriformes</i>	<i>Phylloscopus trochiloides</i>	817	205	1022
15	<i>Spelaeorinis</i>	<i>Sylviidae</i>	<i>Passeriformes</i>	<i>Elachura formosa</i>	678	170	848
16	<i>Leiothrix</i>	<i>Sylviidae</i>	<i>Passeriformes</i>	<i>Leiothrix lutea (Scopoli)</i>	603	151	754

### 4. Methods

#### 4.1. Broad Learning System

The Broad Learning System (BLS) [13] is proposed based on the Random Vector Functional-link Neural Network (RVFLNN) [41]. One of the distinctive advantages of BLS is its ability to simultaneously learn and represent various levels of abstraction within data. The structure of BLS is shown in Figure 1. The input samples of the BLS undergo linear transformations to map onto the feature plane to form feature nodes. The obtained feature nodes are then transformed into enhancement nodes through activation function nonlinear transformation. The feature nodes and enhancement nodes are jointly connected as the actual input of the system, and linearly output through the connection matrix. Finally,

considering the high time cost and tendency to fall into local optima, the BLS obtains the connection matrix through the pseudoinverse.

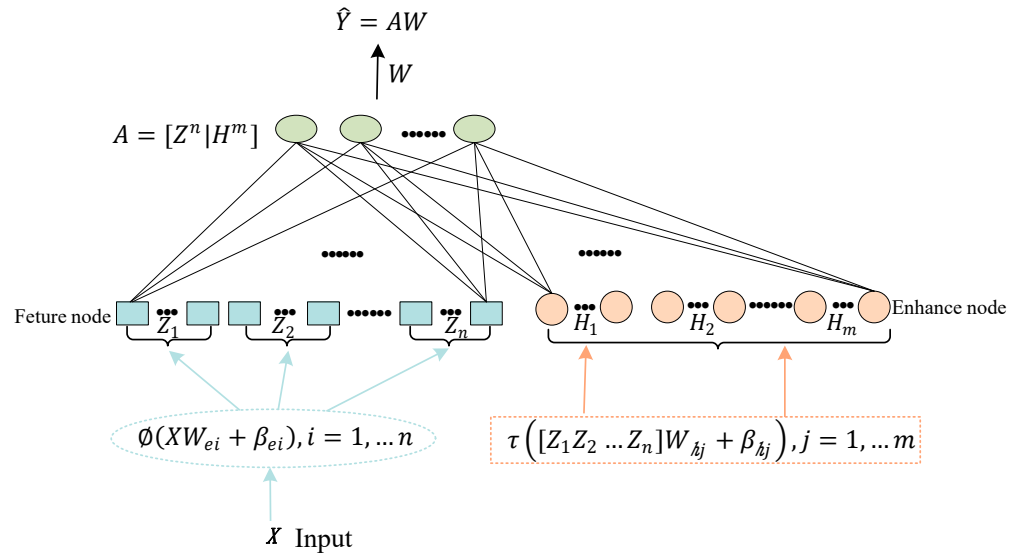


Figure 1. The structure of Broad Learning System.

First, the input samples are mapped into feature nodes  $Z$  by random linear mapping. The formula can be represented as follows:

$$Z = \varnothing(XW_{ei} + \beta_{ei}) \tag{1}$$

where  $n$  refers to the number of feature nodes.  $W_{ei}$  and  $\beta_{ei}$  represent the weights and biases of the feature nodes, respectively, which are randomly initialized.  $\varnothing(\cdot)$  stands for a linear mapping.

After obtaining the feature nodes, enhancement features are generated through non-linear mapping:

$$H = \tau(Z^n W_{hj} + \beta_{hj}) \tag{2}$$

where  $m$  denotes the number of enhancement nodes.  $W_{hj}$  represents the random matrix after orthogonal normalization,  $\beta_{hj}$  is the basis constant, and  $\tau$  represents a nonlinear mapping.

Finally, by merging the feature nodes and enhancement nodes, the input  $A$  of the ultimate network is obtained as  $A = [Z|H]$ . The output of the final network can be represented as:

$$\hat{Y} = [Z|H]W = AW \tag{3}$$

where the weight matrix  $W$  of the output is solved by pseudoinverse and ridge regression learning algorithms. The objective is to find an appropriate  $W$  to minimize the discrepancy between the output  $\hat{Y}$  and the true value of  $Y$ . Chen et al. [13] formulated the optimization problem as follows:

$$\min_w \left( \|Y - AW\|_2^2 + \gamma \|W\|_2^2 \right) \tag{4}$$

where  $\gamma$  is the regularization parameter. By taking the partial derivative of Formula (4) with respect to  $W$  and setting it to zero,  $W$  can be expressed as:

$$W = \left( A^T A + \gamma I \right)^{-1} A^T Y \tag{5}$$

We specifically have

$$A^+ = \lim_{\gamma \rightarrow 0} \left( \gamma I + A A^T \right)^{-1} A^T \tag{6}$$

where  $I$  represents the identity matrix. By solving for the parameter matrix  $W$ , the final construction of the BLS mode is achieved.

#### 4.2. Improved Broad Learning System

The birdsongs recognition model based on improved BLS mainly concludes three modules, namely, the construction of feature sequences, integration of residual block and correlation measures to BLS, and establishment of classification model, detailed as follows:

**Step1:** Construction of differential MFCC feature sequences.

First, the input sound signal is denoised, and then the effective speech segments are determined through endpoint detection. Finally the differential MFCC features of birdsong signal are extracted to obtain the feature matrix. On this basis, the first-order and second-order differential features of each frame of audio data are calculated. Finally, the feature sequences of a sample are formed by sliding window approach.

**Step2:** Introduce residual block module into the BLS and measure the correlation between nodes and class for nodes selection.

By introducing residual connections between feature nodes and enhancement nodes, the differences between nodes are eliminated and residual blocks are formed. Then, residual blocks are linearly connected with feature nodes to form the nodes layer. Finally, a mutual similarity criterion (MSC) is used to measure the similarity between nodes for nodes selection in order to obtain the most relevant and effective node subset.

**Step3:** Map the node subset to the test set to verify model performance.

The specific framework structure is shown in Figure 2.

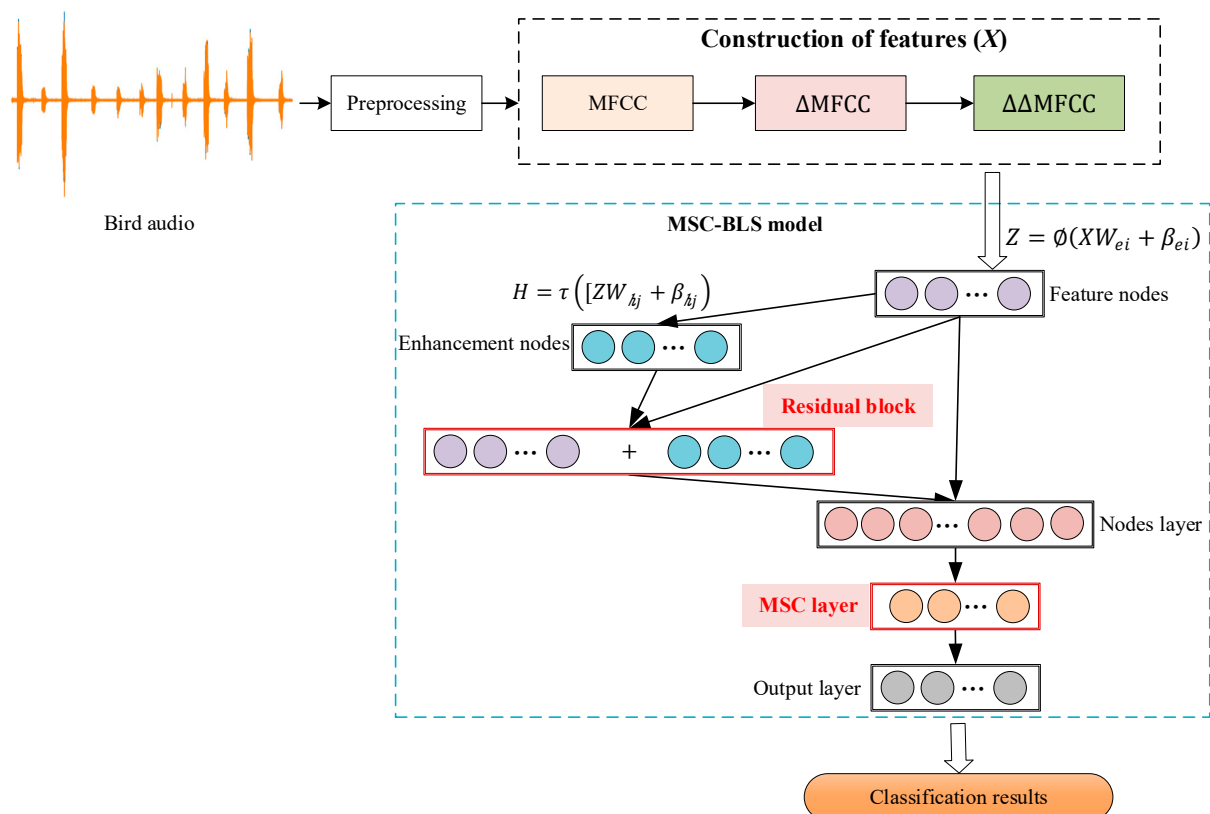


Figure 2. The Proposed framework.

##### 4.2.1. Construction of Differential MFCC Feature Sequences

Mel-Frequency Cepstral Coefficients (MFCC) is a feature parameter that closely models the auditory characteristics of the human ear and is widely employed in various audio-related tasks, including speech and sound recognition. The MFCC extraction process involves the following steps:

1. Perform the signal pre-processing: pre-emphasis, framing, and windowing.
2. Perform the Fast Fourier Transform (FFT) on each frame.
3. Pass the power spectrum through a bank of Mel filters.
4. Take the logarithm of the filter-bank energies.
5. Apply the Discrete Cosine Transform (DCT) to obtain the MFCC coefficients.

Define the original feature matrix as  $T$ , where the size of  $T$  is  $N * D$ , representing  $N$  samples, each with  $D$  dimension. After obtaining the MFCC features, the first-order ( $\Delta$ MFCC) and second-order ( $\Delta\Delta$ MFCC) differentials are calculated through differencing. Utilizing MFCC,  $\Delta$ MFCC, and  $\Delta\Delta$ MFCC, we construct a new feature,  $d_{MFCC}$ , as shown in Formula (7).

$$d_{MFCC} = Concatenate(MFCC, \Delta MFCC, \Delta\Delta MFCC) \tag{7}$$

Next, we adopt a sliding window approach with window size  $w$  and step size  $p$ . For each window of  $w$  consecutive samples, the class labels of the samples are consistent, and we concatenate  $w$  samples together. The final sample representation is denoted as  $d_{sequence}$ , and it has dimensions  $((N - w) + p) * (D * w)$ . This process effectively captures the dynamic information in birdsongs and provides richer features for subsequent recognition tasks. The overall procedure is illustrated in Figure 3.

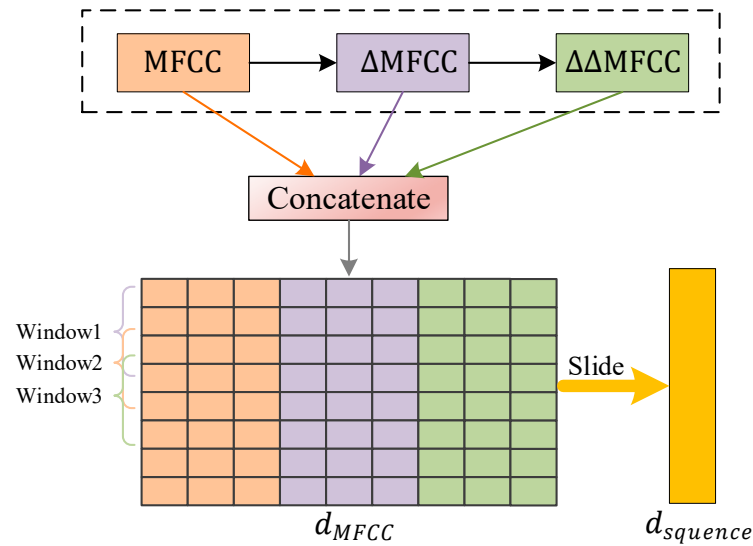


Figure 3. Construction of  $d_{sequence}$  feature.

#### 4.2.2. Improved BLS Based on Residual Block and Mutual Similarity Criterion (RMSC-BLS)

In this paper, we introduce two modules, residual block and MSC (Mutual similarity criterion) layer, into the BLS to form an improved BLS, named RMSC-BLS. The structure of the model is shown in Figure 4.

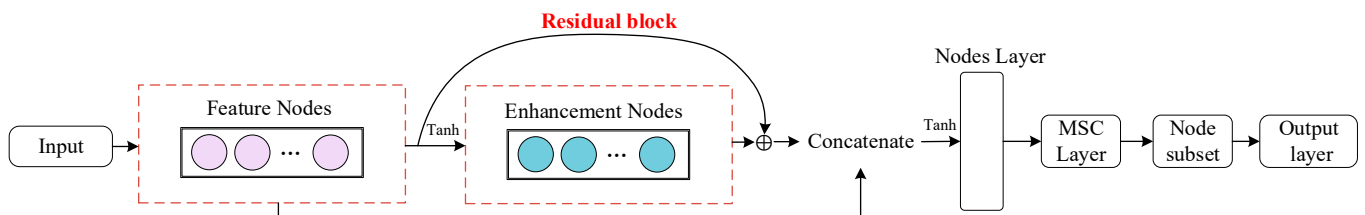


Figure 4. The improved BLS model with residual block and MSC.

- The design of residual block

The residual block allows the network to learn residual information of the differences between feature nodes and enhancement nodes during the training process by introducing

skip connections. In this way, the network can adjust the expression of the original features by learning residuals, rather than directly learning the features themselves. The design of the residual module is mainly used between feature nodes ( $Z$ ) and enhancement nodes ( $H$ ) of BLS, as shown in Formula (8).

$$NL = (Z \oplus H)|Z \quad (8)$$

where  $\oplus$  represents residual connection. After obtaining the residual block, the residual block is linearly concatenated with  $Z$  to obtain the Nodes Layer ( $NL$ ).

- Mutual similarity criterion (MSC)

Nodes in BLS have a direct impact on the network. Through the correlation between the nodes and the class, the features irrelevant to the output are removed, so that the BLS can effectively carry out feature learning. This improves the generalization ability of the model. According to the nodes and class  $c$  combined with the cosine similarity, it is called MSC (Mutual similarity criterion), and can be expressed as:

$$\cos(F^i, c) = \frac{F^i \cdot c}{\|F^i\|_2 \|c\|_2} \quad (9)$$

where  $F^i$  represents the  $i$ -th node in Nodes layer,  $c$  is the sample class.  $\cos(F^i, c)$  refers to the cosine similarity between each node and class.  $\|\cdot\|_2$  is the Frobenius norm of the matrix. Due to  $\cos(F^i, c) \in [-1, 1]$ , the closer the cosine values of  $F^i$  and  $c$  are to 1, the higher the correlation between the nodes and  $c$ . During the node selection process, the correlation between  $F^i$  and  $c$  is computed using the cosine theorem, and the correlations are ranked. Subsequently, the top  $k$  nodes are chosen from MSC layer to form the node subset ( $N_s$ ). This obtained node subset is then entered into the output layer of BLS to establish a birdsong classification model. The major steps of the proposed RMSC-BLS are summarized in Algorithm 1.

---

#### Algorithm 1: The procedure of RMSC-BLS

---

**Input:** training set  $\{X, Y\}$ , number of feature nodes  $n$ , number of enhancement nodes  $m$ , node subset  $N_s = \emptyset$ .

**Output:** Classification result  $\hat{Y}$

1. for  $i = 1; i \leq n$  do
  2. Randomly generate  $W_{ei}$  and  $\beta_{ei}$ ;
  3. Calculate  $Z_i$  by Formula (1);
  4. end // Obtained the feature nodes  $Z$ .
  5. for  $j = 1; j \leq m$  do
  6. Randomly generate  $W_{hj}$  and  $\beta_{hj}$ ;
  7. Calculate  $H_j$  by Formula (2);
  8. end // Obtained the enhancement nodes  $H$ .
  9. Obtain Nodes Layer ( $NL$ ) according to the Formula (8);
  10. Calculate correlation for each node in  $NL$  according to the Formula (9);
  11. Sort the correlation and select the top  $k$  nodes as nodes subset  $N_s$ ;
  12. Set the input  $A = [N_s|W]$  for BLS;
  13. Calculate  $A^+$  with Formula (6);
  14. Calculate the weight matrix  $W$  with Formula (5);
  15. Obtained  $\hat{Y}$  with Formula (3);
  16. Output  $\hat{Y}$ .
- 

## 5. Experiment and Result Analysis

### 5.1. Experimental Design and Environment

The hardware platform used in this experiment is a desktop computer with 128 G memory, 16 core and 32 thread CPU, 3.40 GHz frequency, and 24G GPU. The operating system is Windows 10 64-bit professional operating system. Anaconda3, PyCharm 2018.3.5,



Python 3.6, and TensorFlow 2.6 are used as deep learning platforms, and MATLAB 2021a is used as a data processing platform.

Two group experimental schemes are designed. One group compared the proposed RMSC-BLS with the original BLS (BLS) and the BLS with residual block (Res-BLS). Another group compared RMSC-BLS with other classic methods, including Random Forest (RF) [42], Support Vector Machines (SVM) [43], Extreme Learning Machine (ELM) [44], and Multilayer Perceptron (MLP) [45]. The parameters of the experiments are as follows: feature nodes per window (N1); number of windows of feature nodes (N2); number of enhancement nodes (N3); regularization parameter: C; the shrinkage scale of the enhancement nodes: s; the number of hidden layer neurons: h. The relevant parameters settings of the classifiers model are listed in Table 2.

**Table 2.** Setting of experimental parameters.

Classifier	Relevant Parameter Settings
BLS	activation function: tanh, epoch:50, N1:10, N2:100, N3:1000, C: $2^{-30}$ , s: 0.8
ELM	hidden_layer_size:1000, activation function: sigmoid
MLP	hidden_layer_size:100, activation function: ReLU
RF	alpha: 0.0001, solver: Adam, learning_rate_init: 0.001, max_iter: 200
SVM	n_estimators:100, random_state: 0, criterion: gini, max_depth: none kernel: rbf, gamma: auto, catch_size: 200, max_iter: -1, degree: 3

Each experiment is repeated 50 times independently, and the average of the experimental results is taken as the final result. In this experiment, Accuracy (*Acc*), Precision (*Pre*), Recall, and F1-score (*F1*) are used as indicators to evaluate performance of classification model. The calculation formula of *Acc* is as follows:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \quad (10)$$

Among them, *TP* refers to True Positive, which represents the number of positive samples correctly predicted as positive by the model. *TN* stands for True Negative, representing the number of negative samples correctly predicted as negative by the model. *FP* represents False Positive, indicating the number of negative samples incorrectly predicted as positive by the model. *FN* represents False Negative, indicating the number of positive samples incorrectly predicted as negative by the model.

*Pre* is used to evaluate the proportion of all predicted correct samples that contain actual correct samples. The calculation formula is as follows:

$$Pre = \frac{TP}{TP + FP} \quad (11)$$

The recall rate is the percentage of correctly predicted samples among all samples. The calculation formula is as follows:

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

*F1* score is a comprehensive consideration of *Pre* and *Recall*, defined as:

$$F1 = \frac{2 \times (Pre \times Recall)}{Pre + Recall} \quad (13)$$

## 5.2. Analysis of Experiment Results

With the sliding window approach, we set the size of *w* to 4, and the *p* size to 1. By combining the three features (MFCC,  $\Delta$ MFCC, and  $\Delta\Delta$ MFCC), we obtain  $d_{MFCC}$  features (39 dimensions). Subsequently, four frames of audio data are combined into feature sequences of a sample (156 dimensions). The final generated training and test samples are

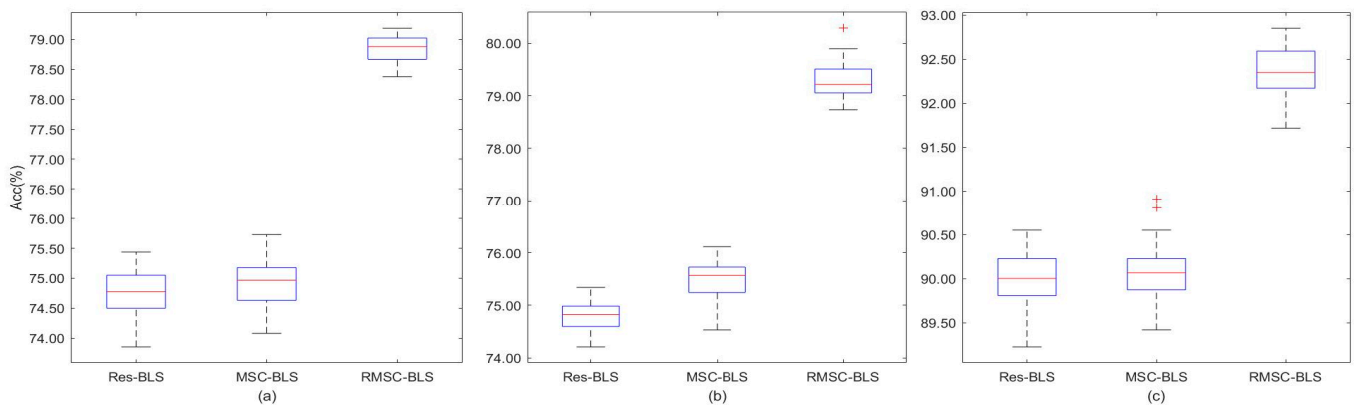
3091 and 12,337, respectively. The classification models are built using the training set and the recognition effect is verified using test set. *Acc*, *Pre*, and other evaluation indicators are calculated based on the confusion matrix.

### 5.2.1. The Results of RMSC-BLS

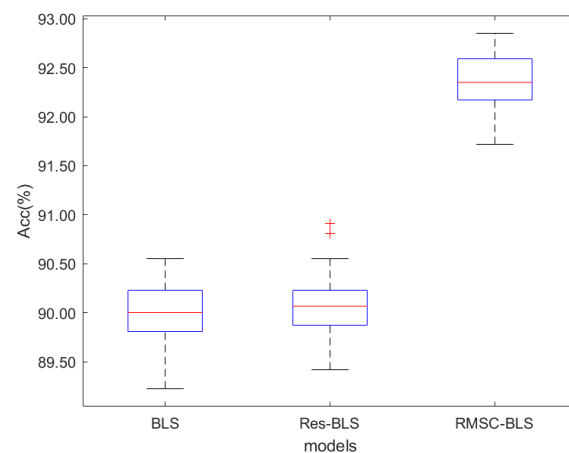
The classification performance of three models, including BLS, Res-BLS, and RMSC-BLS, on the test set is shown in Table 3. The accuracy of 50 runs of each model is shown in Figures 5 and 6.

**Table 3.** Classification results of different models.

Model	Features	Performance (%)			
		<i>Acc</i> (mean ± std)	<i>F1</i> (mean ± std)	<i>Pre</i> (mean ± std)	<i>Recall</i> (mean ± std)
BLS	MFCC	74.77 ± 0.37	74.39 ± 0.41	75.80 ± 0.39	73.84 ± 0.43
	$d_{MFCC}$	74.79 ± 0.29	74.33 ± 0.31	75.69 ± 0.30	73.81 ± 0.32
	$d_{sequence}$	89.99 ± 0.31	89.50 ± 0.32	90.39 ± 0.34	88.94 ± 0.31
Res-BLS	MFCC	74.90 ± 0.37	74.53 ± 0.40	75.97 ± 0.38	73.99 ± 0.41
	$d_{MFCC}$	75.48 ± 0.32	75.06 ± 0.36	76.45 ± 0.36	74.51 ± 0.36
	$d_{sequence}$	90.07 ± 0.29	89.58 ± 0.32	90.47 ± 0.34	89.02 ± 0.32
RMSC-BLS	MFCC	78.85 ± 0.21	78.62 ± 0.22	79.50 ± 0.23	78.45 ± 0.22
	$d_{MFCC}$	79.29 ± 0.33	79.07 ± 0.34	79.89 ± 0.34	78.85 ± 0.34
	$d_{sequence}$	92.37 ± 0.25	91.99 ± 0.27	92.61 ± 0.29	91.59 ± 0.27



**Figure 5.** Performance of different features on three models. (a) Results of MFCC on models; (b) Results of  $d_{MFCC}$  on models; (c) Results of  $d_{sequence}$  on models.



**Figure 6.** Results of  $d_{sequence}$  on RMSC-BLS.

From Table 3 and Figure 5, it shows that for each model the  $d_{sequence}$  feature is superior to the other two sets of features in all four evaluation indicators listed. The constructed feature  $d_{sequence}$  has increased by 15% in all evaluation indicators. On the aspect of standard deviation, the  $d_{sequence}$  is the lowest than the other two features, which reflects that it has a more stable performance among all features. In conclusion, the  $d_{sequence}$  feature proves to be effective for birdsong recognition.

It can be seen from Table 3 and Figure 6 that for the constructed  $d_{sequence}$  features, compared with BLS and Res-BLS, the accuracy rate of RMSC-BLS is 92.37%, which is 2.38% and 2.30% higher than that of BLS and Res-BLS, respectively. The Res-BLS has a slight improvement over the BLS. From the comparison of Table 3, Figure 5, and Figure 6, it is evident that RMSC-BLS outperforms all other models, exhibiting the highest accuracy with an improvement ranging from 2.38 to 17.60%. In addition, *F1*, *Pre*, and *Recall* are 91.99%, 92.61% and 91.59%, respectively, which increased by 17.66–2.41%, 16.92–2.14%, 17.78–2.56% compared with other models, respectively.

The comparison of the experimental results of different models is shown in Figure 7. Through comprehensive analysis and comparison of the three models of experiments, it illustrates that the proposed method has superior performance when compared with other models. The results on *Acc*, *F1*, *Pre*, and *Recall* are superior to BLS and Res-BLS models. As a whole, the RMSC-BLS has better generalization performance for the birdsong identification.

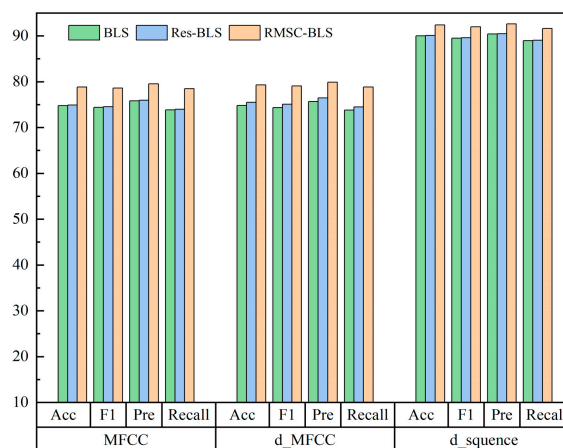
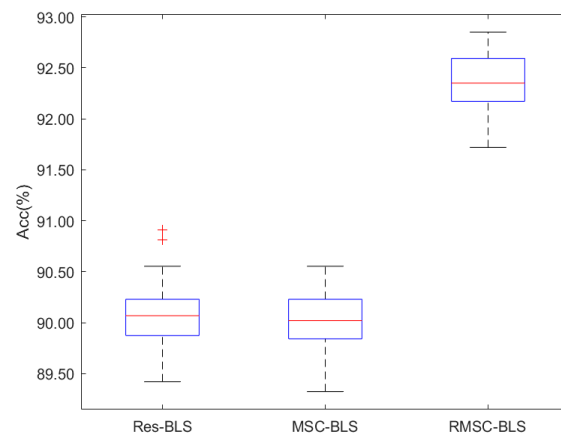


Figure 7. Comparison of different BLS models.

From Table 4 and Figure 8, it can be observed that for the  $d_{sequence}$  feature data, the performance of the Res-BLS and MSC-BLS models is similar in various indicators. However, combining Res-BLS and MSC-BLS to form the RMSC-BLS model, there is an improvement in each indicator. In the experimental schemes, the RMSC-BLS model achieves better classification results compared to other models, with a minimum and maximum improvement *Acc* of 2.3% and 2.33%, respectively. Other indicators also show some improvement. Overall, the proposed RMSC-BLS model has better classification results than other models in experimental schemes and achieves superior recognition results.

Table 4. Performance comparison of three module experiments.

Model	Performance (%)			
	<i>Acc</i> (mean ± std)	<i>F1</i> (mean ± std)	<i>Pre</i> (mean ± std)	<i>Recall</i> (mean ± std)
Res-BLS	90.07 ± 0.29	89.58 ± 0.32	90.47 ± 0.34	89.02 ± 0.32
MSC-BLS	90.04 ± 0.30	89.56 ± 0.32	90.48 ± 0.31	88.97 ± 0.33
RMSC-BLS	92.37 ± 0.25	91.99 ± 0.27	92.61 ± 0.29	91.59 ± 0.27



**Figure 8.** Comparison with Res-BLS, MSC-BLS and RMSC-BLS.

### 5.2.2. Comparison of RMSC-BLS with Other Methods

To verify the effective of the proposed model, we compared RMSC-BLS with four methods, namely RF, SVM, ELM and MLP, all of which used data of  $d_{sequence}$ .

Table 5 shows the comparative analysis of RMSC-BLS with other methods, and Figure 9 shows the confusion matrices of each model. The confusion matrix of ELM and RMSC-BLS is the experimental results of running once on  $d_{sequence}$ . From Table 5, we can conclude that the RMSC-BLS performs well in all performance indicators except for the *Pre*, with the highest *Acc* (92.37%), *F1* (91.99%), and *Recall* (91.59%). Although the *Pre* of RF is 92.75%, which is 0.14% higher than that of RMSC-BLS, RMSC-BLS outperforms the other models in most cases. This indicates that the model has achieved the best performance in handling this specific task and can effectively identify and classify samples.

**Table 5.** Comparison of RMSC-BLS with other methods.

Model	Performance (%)			
	<i>Acc</i>	<i>F1</i>	<i>Pre</i>	<i>Recall</i>
RF	91.33	91.13	92.75	90.06
SVM	85.47	84.63	86.31	83.75
MLP	90.84	90.25	90.88	90.19
ELM	86.92	86.18	87.57	85.52
RMSC-BLS	92.37	91.99	92.61	91.59

To sum up, the RMSC-BLS model demonstrates better performance in this group of experiments, and its *Acc*, *F1*, and *Recall* indicators are higher than other models. Compared with traditional machine learning models, such as RF, SVM, MLP, and ELM, RMSC-BLS has significant performance improvements, highlighting its advantages in birdsong recognition tasks.

### 5.3. Discussion

In this study, we demonstrate that the constructed  $d_{sequence}$  features exhibit superior results compared to the original MFCC feature parameters of birdsongs. The constructed  $d_{sequence}$  features enhance the expression ability of features, which is crucial for birdsong identification in complex environments. Many previous studies have also used MFCC features for birdsong recognition. For instance, Xie improved ELM by using differential evolution to classify MFCC features of nine birdsongs, with a maximum accuracy of 89.05% [46]. Wang et al. fused Mel-spectrogram and MFCC as input features and used LSTM to recognize 264 birdsongs, with an average accuracy of 77.43% [47]. Murugaiya et al. combined the improved GTCC feature with probability enhanced entropy to classify twenty bird sounds in Borneo using SVM, with an accuracy of 89.5% [48]. At the same

time, we also use ResNet18 with MFCC spectral features, achieving an accuracy of 92.11%, though this took a long time.

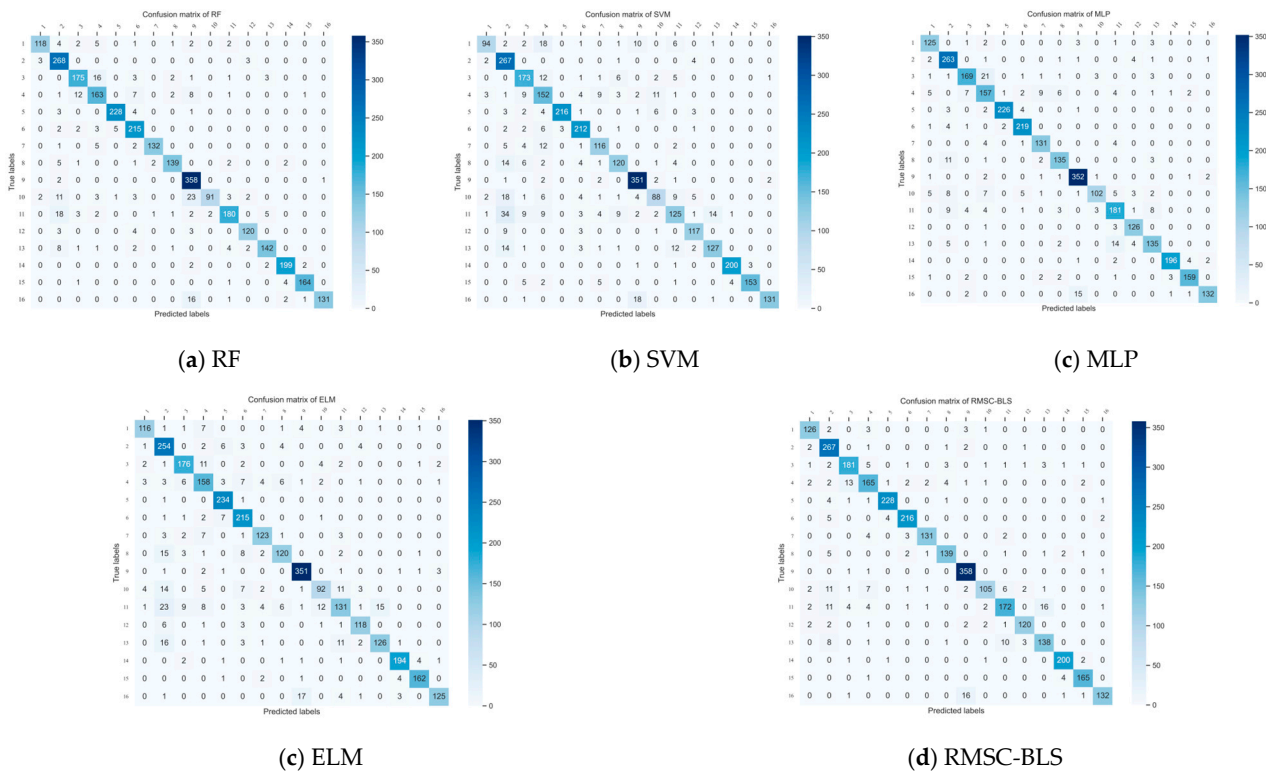


Figure 9. Confusion matrices.

For the application of deep learning in the field of birdsong recognition, SRA et al. used RNN-LSTM to explore audio feature extraction techniques (Mel Spectrograms and MFCC) for bird identification of 91 species, and ultimately achieved an accuracy of 44.26% on the validation set [49]. Mohanty et al. used spiking neural network with permutation pair frequency matrix to classify 14 bird species, with an accuracy of 92% [50]. Carvalho et al. used deep learning models to identify the Mel Spectrograms and MFCC of 91 bird species, achieving an accuracy of 44.26% [51]. Xie et al. used deep learning to combine two acoustic and visual features for late fusion and classified 14 birdsongs. The final best classification F1 score was 95.95% [52]. Liu et al. used CNN to identify the Wavelet transform (WT) [53], short-time Fourier transform (STFT) [54], and Hilbert–Huang transform (HHT) [55] spectral features of 16 bird species, achieving accuracy rates of 89.11%, 88.36%, and 81%, respectively [56]. On this basis, three spectral features were further applying CNN to obtained CNN-WT, CNN-STFT, and CNN-HHT features. The results obtained using our proposed model RMSC-BLS were 91.22%, 90.78%, and 85.20%, respectively.

Although these studies have made significant improvements, the lack of standardized benchmark datasets makes it difficult to directly compare methods even using the same indicators. From the above research, it can be concluded that the proposed RMSC-BLS model achieves a higher classification accuracy than most of these birdsong classification methods. These results clearly indicate that the improved BLS is successful in birdsong classification, as it can more fully express the sample information of birdsong.

### 6. Limitations and Future Scope

This paper constructs a differential MFCC feature sequence and proposes the RMSC-BLS- $d_{sequence}$  method based on the mutual similarity between the nodes and class to carry out research on birdsong recognition. By optimizing the nodes of BLS and proving its effectiveness on three features, it demonstrated better performance on four evaluation

indicators. There are still more problems to be solved in the feature works. Here are some of the most important points:

- (1) Extract more different features and extend the method to these features.
- (2) The categories and sample size of birdsongs need to be expanded. The RMSC-BLS model will be extended to encompass a wider range of birdsongs as well as other audio recognition applications.
- (3) The weights of the input layer and feature nodes in the BLS, as well as the weights from feature nodes to enhancement nodes, are randomly generated and not interpretable. Therefore, future research will further optimize these two parameters.
- (4) Establish the connection between bird species identification results and bird biodiversity assessment indicators to explore the relationship between diversity indices. The cross-species identification contributes to a broader understanding of bird diversity and interactions within ecosystems.

## 7. Conclusions

In this study, we focus on 16 kinds of bird species as research objects and construct the differential MFCC features. To enhance the learning process, the RMSC-BLS method is proposed by introducing residual block in BLS to learn the residuals between feature nodes and enhancement nodes and using mutual similarity to measure the correlation between nodes and class for nodes selection. Experiments show that the proposed method has improved compared to the original BLS in all three features, achieving the best recognition accuracy of 92.37%. In general, the proposed RMSC-BLS method can identify the birdsong effectively and efficiently and has a good generalization ability.

**Author Contributions:** Conceptualization, J.L.; Data curation, J.L. and S.X.; Funding acquisition, Y.Z. (Yan Zhang), D.L. (Danjv Lv), J.L.; Investigation, J.L. and Y.F.; Methodology, S.X., J.L., D.L. (Dan Lv) and Z.L.; Supervision, Y.Z. (Yan Zhang) and D.L. (Danjv Lv); Validation, S.X., D.L. (Danjv Lv) and Y.Z. (Youjie Zhao); Writing—original draft, J.L.; Writing—review and editing, Y.Z. (Yan Zhang), D.L. (Danjv Lv), Y.Z. (Youjie Zhao) and Z.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the Yunnan Provincial Science and Technology Department under Grant no: 202002AA10007, the National Natural Science Foundation of China under Grant no: 61462078 and under Grant no: 31860332, and the Yunnan Provincial Department of Education under Grant no: 2023Y0704.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data included in this study are available upon request by contact with the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Yi, Q.V.; Ming, L.U. Construction of Urban Ecological Security Pattern Based on Biodiversity Conservation. *Urban Dev. Stud.* **2017**, *24*, 134–137.
2. Hu, Y.W. Research on Feature Extraction and Classification of Audio Signals. Master's Thesis, Kunming University of Science and Technology, Kunming, China, 2018.
3. Brooker, S.A.; Stephens, P.A.; Whittingham, M.J.; Willis, S.G. Automated detection and classification of birdsong: An ensemble approach. *Ecol. Indic.* **2020**, *117*, 106609. [[CrossRef](#)]
4. Clemins, P.; Trawicki, M.; Adi, K.; Tao, J.; Johnson, M. Generalized perceptual features for vocalization analysis across multiple species. In Proceedings of the 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, IEEE, Toulouse, France, 14–19 May 2006. [[CrossRef](#)]
5. Selin, A.; Turunen, J.; Tantt, J.T. Wavelets in recognition of bird sounds. *EURASIP J. Adv. Signal Process.* **2006**, *2007*, 1–9. [[CrossRef](#)]

6. Cai, J.; Ee, D.; Pham, B.; Roe, P.; Zhang, J. Sensor network for the monitoring of ecosystem: Bird species recognition. In Proceedings of the 2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information IEEE, Melbourne, VIC, Australia, 3–6 December 2007; pp. 293–298. [\[CrossRef\]](#)
7. Fagerlund, S. Bird species recognition using support vector machines. *EURASIP J. Adv. Signal Process.* **2007**, *2007*, 1–8. [\[CrossRef\]](#)
8. Ren, Y.; Johnson, M.T.; Clemens, P.J.; Darre, M.; Glaeser, S.S.; Osiejuk, T.S.; Out-Nyarko, E. A framework for bioacoustic vocalization analysis using hidden Markov models. *Algorithms* **2009**, *2*, 1410–1428. [\[CrossRef\]](#)
9. Potamitis, I.; Ntalampiras, S.; Jahn, O.; Riede, K. Automatic bird sound detection in long real-field recordings: Applications and tools. *Appl. Acoust.* **2014**, *80*, 1–9. [\[CrossRef\]](#)
10. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *60*, 1097–1105. [\[CrossRef\]](#)
11. Issa, D.; Demirci, M.F.; Yazici, A. Speech emotion recognition with deep convolutional neural networks. *Biomed. Signal Process. Control.* **2020**, *59*, 101894. [\[CrossRef\]](#)
12. Chang, C.C.; Chang, T.Y.P.; Xu, Y.G.; Wang, M.L. Structural damage detection using an iterative neural network. *J. Intell. Mater. Syst. Struct.* **2000**, *11*, 32–42. [\[CrossRef\]](#)
13. Chen, C.L.P.; Liu, Z. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *29*, 10–24. [\[CrossRef\]](#)
14. Chen, C.L.P.; Liu, Z.; Feng, S. Universal approximation capability of broad learning system and its structural variations. *IEEE Trans Neural Netw. Learn. Syst.* **2019**, *30*, 1191–1204. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Li, G.; Xu, L. Application of local receptive field based broad learning system. *Comput. Eng. Appl.* **2019**, *56*, 162–167. [\[CrossRef\]](#)
16. Liu, Z.; Zhou, J.; Chen, C.L.P. Broad learning system: feature extraction based on K-means clustering algorithm. In Proceedings of the 4th International Conference on Information, Cybernetics and Computational Social Systems, Piscataway, NJ, USA, 24–26 July 2017; pp. 683–687. [\[CrossRef\]](#)
17. Li, W.; Yu, Z. Application of broad learning system in discrimination of mushroom toxicity. *Mod. Food Sci. Technol.* **2019**, *35*, 267–272. [\[CrossRef\]](#)
18. Fan, X.-N.; Zhang, S.-W. LPI-BLS: predicting IncRNA-protein interactions with a broad learning system-based stacked ensemble classifier. *Neurocomputing* **2019**, *370*, 88–93. [\[CrossRef\]](#)
19. Ren, C.; Yuan, C.; Sun, Y.; Liu, Z.; Chen, J. Research of broad learning system. *Appl. Res. Comput.* **2021**, *38*, 2258–2267. [\[CrossRef\]](#)
20. Logan, B. Mel frequency cepstral coefficients for music modeling. *Int. Soc. Music. Inf. Retr.* **2000**, *270*, 11.
21. Makhoul, J. Linear prediction: A tutorial review. *Proc. IEEE Inst. Electr. Electron Eng.* **1975**, *63*, 561–580. [\[CrossRef\]](#)
22. Xie, J.J.; Li, W.B.; Zhang, J.G.; Ding, C.Q. Bird species recognition method based on Chirplet spectrogram feature and deep learning. *J. Beijing For. Univ.* **2018**, *40*, 122–127. [\[CrossRef\]](#)
23. Liu, J.; Zhang, Y.; Lv, D.J.; Lu, J.; Xie, S.S.; Zi, J.L.; Yin, Y.; Xu, H.F. Birdsong classification based on ensemble multi-scale convolutional neural network. *Sci. Rep.* **2022**, *12*, 8636. [\[CrossRef\]](#)
24. Stowell, D.; Plumbley, M.D. Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. *PeerJ* **2014**, *2*, e488. [\[CrossRef\]](#)
25. Wang, E.; He, D. Bird recognition based on MFCC and dual-GMM. *Comput. Eng. Des.* **2014**, *35*, 1868–1871+1881. [\[CrossRef\]](#)
26. Xu, S.; Sun, Y.; Huangfu, L.; Fang, W.Q. Design of Synthesized Bird Sounds Classifier Based on Multi Feature Extraction Classifiers and Time-frequency Chat. *Res. Explor. Lab.* **2018**, *37*, 81–86+91.
27. Yan, N.; Chen, A.; Zhou, G.; Zhang, Z.; Liu, X.; Wang, J.; Liu, Z.; Chen, W. Birdsong classification based on multi-feature fusion. *Multimed. Tools. Appl.* **2021**, *80*, 36529–36547. [\[CrossRef\]](#)
28. Bai, J.; Chen, C.; Chen, J. Xception Based Method for Bird Sound Recognition of BirdCLEF 2020. Conference and Labs of the Evaluation Forum, 2020. Available online: <https://api.semanticscholar.org/CorpusID:225073815> (accessed on 25 September 2022).
29. Gupta, G.; Kshirsagar, M.; Zhong, M.; Gholami, S.; Ferres, J.L. Comparing recurrent convolutional neural networks for large scale bird species classification. *Sci. Rep.* **2021**, *11*, 17085. [\[CrossRef\]](#)
30. Ye, R.; Kong, Q.; Li, D.; Chen, Y.; Zhang, Y.; Liu, C. Shrimp Freshness Detection Method Based on Broad Learning System. *Spectrosc. Spectr. Anal.* **2022**, *42*, 164–169.
31. Zhu, Y.; Yan, X.; Shen, X.; Lu, Z. Fatigue driving detection based on cascade broad learning. *Comput. Eng. Des.* **2020**, *41*, 537–541. [\[CrossRef\]](#)
32. Wang, Z.; Xu, X.; Liu, H.; Sun, F. Cascade broad learning for multi-modal material recognition. *CAAI Trans. Intell. Syst.* **2020**, *15*, 787–794.
33. Huang, P.-Q. Research on Design Method of Broad Neural Network. Master's Thesis, Wuhan University of Science and Technology, Wuhan, China, 2020. [\[CrossRef\]](#)
34. Zheng, Y.; Chen, B.; Wang, S.; Wang, W. Broad learning system based on maximum correntropy criterion. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 3083–3097. [\[CrossRef\]](#)
35. Zhang, T.-L.; Chen, R.; Yang, X.; Guo, S. Rich feature combination for cost-based broad learning system. *IEEE Access* **2018**, *7*, 160–172. [\[CrossRef\]](#)
36. Ye, H.; Li, H.; Chen, C.L.P. Adaptive deep cascade broad learning system and its application in image denoising. *IEEE Trans. Cybern.* **2021**, *51*, 4450–4463. [\[CrossRef\]](#)

37. Han, M.; Feng, S.; Chen, C.L.P.; Xu, M.; Qiu, T. Structured manifold broad learning system: A manifold perspective for large-scale chaotic time series analysis and prediction. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 1809–1821. [[CrossRef](#)]
38. Ding, Z.; Chen, Y.; Li, N.; Zhao, D.; Sun, Z.; Chen, C.L.P. BNAS: Efficient neural architecture search using broad scalable architecture. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 5004–5018. [[CrossRef](#)]
39. Xie, R.; Wang, S. Downsizing and enhancing broad learning systems by feature augmentation and residuals boosting. *Complex Intell. Syst.* **2020**, *6*, 411–429. [[CrossRef](#)]
40. Ma, J.; Fan, J.; Wang, L.; Chen, C.P.; Yang, B.; Sun, F.; Zhou, J.; Zhang, X.; Gao, F.; Zhang, N. Factorization of broad expansion for broad learning system. *Inf. Sci.* **2023**, *630*, 271–285. [[CrossRef](#)]
41. Pao, Y.H.; Park, G.H.; Sobajic, D.J. Learning and generalization characteristics of the random vector functional-link net. *Neurocomputing* **1994**, *6*, 163–180. [[CrossRef](#)]
42. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
43. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28. [[CrossRef](#)]
44. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [[CrossRef](#)]
45. Gardner, M.W.; Dorling, S.R. Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmos. Environ.* **1998**, *32*, 2627–2636. [[CrossRef](#)]
46. Xie, S.; Zhang, Y.; Lv, D.; Xu, H.; Liu, J.; Yin, Y. Birdsongs recognition based on ensemble ELM with multi-strategy differential evolution. *Sci. Rep.* **2022**, *12*, 9739. [[CrossRef](#)]
47. Wang, H.L.; Xu, Y.F.; Yu, Y.; Lin, Y.C.; Ran, J.H. An efficient model for a vast number of bird species identification based on acoustic features. *Animals* **2022**, *12*, 2434. [[CrossRef](#)] [[PubMed](#)]
48. Murugaiya, R.; Abas, P.E.; De Silva, L.C. Probability enhanced entropy (PEE) novel feature for improved bird sound classification. *Mach. Intell. Res.* **2022**, *19*, 52–62. [[CrossRef](#)]
49. Chakraborty, K.; Tyagi, S.; Shridevi, S. Comparative analysis of deep learning models for bird song classification. In *AIP Conference Proceedings*; AIP Publishing: Chennai, India, 2023; Volume 2788. [[CrossRef](#)]
50. Mohanty, R.; Mallik, B.K.; Solanki, S.S. Automatic bird species recognition system using neural network based on spike. *Appl. Acoust* **2020**, *161*, 107177.1–107177.8. [[CrossRef](#)]
51. Carvalho, S.; Gomes, E.F. Automatic Classification of Bird Sounds: Using MFCC and Mel Spectrogram Features with Deep Learning. *Vietnam. J. Comput. Sci.* **2023**, *10*, 39–54. [[CrossRef](#)]
52. Xie, J.; Zhu, M.Y. Handcrafted features and late fusion with deep learning for bird sound classification. *Ecol. Inform.* **2019**, *52*, 74–81. [[CrossRef](#)]
53. Nagarajaiah, S.; Basu, B. Output only modal identification and structural damage detection using time frequency & wavelet techniques. *Earthq. Eng. Eng. Vib.* **2009**, *8*, 583–605. [[CrossRef](#)]
54. Griffin, D.; Lim, J. Signal estimation from modified short-time Fourier transform. *IEEE Trans. Acoust. Speech Signal Process.* **1974**, *32*, 236–243. [[CrossRef](#)]
55. Huang, N.E.; Wu, M.L.; Qu, W.; Long, S.R.; Shen, S.S. Applications of HilbertHuang transform to non-stationary financial time series analysis. *Appl. Stoch. Model. Bus.* **2003**, *19*, 245–268. [[CrossRef](#)]
56. Liu, J.; Zhang, Y.; Lyu, D.; Lu, J.; Xie, S.; Zi, J.; Chen, X.; Zhao, Y. Research on birdsong classification based on multi-view ensemble. *J. Nanjing For. Univ.* **2023**, *47*, 23–30. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.