Journal
IJCSM

Iraqi Journal for Computer Science and Mathematics

Journal Homepage: http://journal.esj.edu.iq/index.php/IJCM
e-ISSN: 2788-7421 p-ISSN: 2958-0544

A Novel Deep Learning Approach for Classification of Bird Sound Using Mel Frequency Cepstral Coefficients

Aymen Saad[1,2] *, Muhammad Mun'im Ahmad Zabidi[1], Israa S. Kamil[3] and Usman Ullah Sheikh[1]

[1] School of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia.
[2]Department of Information Technology, Management Technical College, Al-Furat Al-Awsat Technical University, Kufa, 54003, Iraq.
[3]University of Babylon, College of Information Technology, Department of Information Security, Babil, 51007, Iraq

*Corresponding Author: Aymen Saad

DOI: https://doi.org/10.52866/ijcsm.2024.05.03.040
Received April 2024; Accepted June 2024; Available online August 2024

ABSTRACT: Monitoring animal populations is one important matter to better understand changes in their population, behavior, and biodiversity. Bird sounds are the main tool to classify bird species acoustically. The sounds of birds are an indicator for ecologists as it responds to changes in their environment. The recognition among a variety of bird species to get important features is computationally expensive. With the unbalanced classes and scarcity of training data, the performance accuracy is degrading. This paper aims to classify species of birds using lightweight convolutional neural networks (LWCNNs) basis on using a spectrogram image of Brazilian bird sounds as a dataset. For extracting spectrogram images, Mel Frequency Cepstral Coefficient (MFCC) algorithm is used. To prove the high performance of the classifier, ten species of birds with 10,000 spectrogram images are provided to the classifier. Our LWCNN model achieved a training and testing accuracy of 99.68 % and 92.80 % respectively in 10.54 min with 5 epochs.

Keywords: Bird sound, Deep Learning, Light Wight Convolutional Neural Networks, Classification, MFCC.

1. INTRODUCTION

Exploration of the interaction between organisms and their environment is a significant problem in ecology, especially with the threat of climate change. Acoustically monitoring and classifying animals in their physical environment has sparked a lot of interest recently [1].

Classification and segmentation methods that operate automatically and can generalize while dealing with various species and noisy environments are required to perform higher-level work, such as the detection of migration patterns. The structure of bird songs could be understood while also improving reproducibility and reducing human bias by employing new theories related to bird songs that rely on extracting features from large-scale sounds [2]. The collection and interpretation of bird data require a significant amount of human labour [3]. Making this process work automatically is challenging given the acoustic diversity of bird songs, sound quality in recordings, noise environments, and the presence of multiple species in the same recording. These factors motivate the development of a new automatic system for classifying bird songs [4].

The use of a group of experts can improve classification reliability, but this is time-consuming, expensive, and unsuitable for real-time classification. Nonetheless, the fact that manual monitoring of sound spectrographs produces correct judgments has prompted researchers to develop a classifier based on standards derived objectively from expert opinions [5]. In recent times, the utilization of deep learning methods has brought about a significant transformation in the potential applications of machine learning, particularly in the fields of speech analysis, visual data analysis, and text analysis. Deep architectures, particularly deep convolutional neural networks (CNNs), have been used to improve several classification tasks [6].

The classification scheme in deep learning is designed in such a way that the classifier itself learns the best features for describing patterns during training. CNNs, in particular, describe patterns as images at the start of their classification process [7]. Deep learning has strong capabilities for feature extraction and self-learning, which can automatically acquire characteristic information [8]. In this paper, 10,000 sound samples of Brazilian birds were fed into neural networks. These birds were chosen due to the abundance of samples in the Xeno-Canto repository. Each audio clip is resampled and segmented into 1-second sample data at 16KHz. Each sample containing the bird call signal is then divided into three samples. The spectrogram representation of the samples is obtained using the MFCC algorithm, and all the images are resized to 224*224 using MATLAB 2019b.

2. Related Work

To gain a comprehensive understanding of how the environment has changed, the study of ecology must analyze the birds present in each area. Manual classification is typically performed by ornithologists using either visual or audio data collected directly or indirectly. A summary of related works is provided, with some works coming from the acoustic scene classification (ASC) task. Combining acoustic or sound-based bird species classification with visual or imagery-based classifiers and deep learning is demonstrated in [9]. In this study, CNN was used to analyze general features, while reducing dimensions and fully connected layers will be used to classify bird species. There were 60 votes for each type and voting scheme, with 60 votes for normal and 60 votes for threatened. Fire torches and smoke from simulated fires were used to simulate threatening conditions in the scheme.

Bird sounds can be pre-processed by passing them through a filter to remove unwanted noise and disturbance, segmenting them, and extracting important features from each segment [10]. Once a classification model has been trained, selected bird sound features can be fed directly into the model to predict the species of the unidentified bird sound. The classification model is initially trained using a database of labelled bird sounds. Intuitively, some factors such as the used classification model, feature selection, and training data quality influence the trained classification model's performance in predicting unknown bird sounds. The labelled bird sound is a difficult step in bird classification because it is rich and reliable, so it was used to train the classification.

T-MSNet, based on ShuffleNet and MobileNet, is proposed by Kek et al [11] for the task of acoustic scene classification (ASC) by controlling the number of parameters and input size. The proposed method employs an improved wavelet scattering representation and a shuffling module that shuffles the feature maps in a frequency-dependent manner. The paper demonstrates that this method effectively improves the model's generalization.

Cakir et al [12] employ two stages to represent sound and map acoustic features. The first stage of sound representation is the process of extracting spectro-temporal features from audio recordings. CRNN is used in the second stage to map acoustic features to a binary estimate of bird song presence. It used supervised learning to extract acoustic features from annotations of bird song activity and a training database. In the context of academic literature, CRNN serves as the classification component, comprising a solitary feedforward layer featuring a single neuron with sigmoid activation. It also includes a temporal max-pooling layer, alongside convolutional layers that employ rectified linear unit (ReLU) activations and non-overlapping pooling along the frequency axis.

Yang et al [13] created a lightweight bird sound recognition model to improve recognition accuracy. Two modules are used to improve the network's ability to scale channel information and extract spatial information: multi-scale feature fusion structure and PSA (pyramid split attention). Three modules are introduced to improve the model's refinement ability to global information and the simplicity of feature extraction by extracting the Mel spectrogram and passing it as a three-dimensional feature into the recognition model: ordinary convolution, channel attention mechanism, and Back block. Non-overlapping 4s segments are obtained from long unlabeled recordings using segmentation, which are then classified as positive or negative, with seven positive classes and two negative classes. The network assigns a probability to each class after the classification process. The advantage of these probabilities and predictions is that they allow you to choose a subset with many examples of the target species [14].

In [15] layers of integration are included in CNN properties. Image separation function is used for these CNN properties. The sequence starts with the image, and network input, ending with conversion and merging. Thereafter, one or more fully connected layers are provided with these activities. Finally, the class label is removed by a fully integrated layer. First, the collection and localization of raw input data for most bird semantic components are performed. Second, some characteristics like colour, size, and shape are used to identify and sort common components. Third, the training of CNN to extract a feature with predefined features is accomplished by using images in a graphics processing unit. A capture using a camera is performed on the obtained information from the image uploaded by the user then this captured image could be scanned from information and prediction from a trained model.

Developing a convolutional neural network with five steps is introduced in [16]. These steps are represented by convolution, application of the Rectifier Function, Pooling, Flattening, and Full Connection. To increase the linearity of the image, a corrective function in Relu (Rectified Linear Unit) is adopted. Converting a non-linear image into a linear one is important in passing the image to the convolution operation. Spatial variance competence is provided to the convolutional neural network by pooling. In addition, the process of minimizing the size of the image and reducing the number of parameters is the services of pooling that leads to reducing the computational power required in data processing and that in turn helps in over-fitting prevention. A re-organization of the resulting feature map from pooling into a column of values is executed. The detection of a certain feature in a fully connected layer and retaining its value by the receptor neuron is the main function of the full connection process. The neuron communicates this value to all the classes. Finally, controlling and determining if the value of the feature is acceptable to them is performed by the classes.

The framework presented in reference [17] is structured into two main phases. Firstly, it focuses on the reconstruction of high-resolution (HR) images from low-resolution (LR) images using super-resolution (SISR) techniques, such as DRRN, VDSR, and FSRCNN. In the second phase, the HR images are subjected to refinement and optimization for bird detection using R-CNN and YOLO. To establish the LR-to-HR mapping function (F), which takes the LR image Y as input and produces the HR image X as output, a model with adjustable parameters is trained through SISR methods. Specifically, VDSR introduces the concept of residual learning between LR and HR images. Notably, FSRCNN stands out as a real-time SISR method, achieving a remarkable speed of 24 frames per second (fps) and superior super-resolution quality, outperforming other contemporaneous SISR methods by a significant margin (approximately 17.36 times faster). The core components of FSRCNN encompass feature extraction, mapping, shrinking, deconvolution, and expanding.

ResNet-50 is presented in [18], the original network architecture used for training was modified to be able for 46 bird species classification. This modification is accomplished by using a fully connected layer with 46 neurons instead of a densely connected layer with 1000 neurons. Specifically, let c denote the number of columns and r denotes the number of rows in a spectrogram. There are (r x c x 3) input images accepted by the network, whereas the ResNet CNN accepts the three-colour channels loaded by spectrogram image. Choosing the subsets of species for each run of classification of n species is not careful. Rather a generator of a random number is used to obtain 20 bags of species. Using this method, a bag of species drawn randomly without repetition from 659 complete species could be obtained. To get a good estimate of the performance of the classification using a specified number of species, the computations should be repeated 20 times [19].

To highlight the main points of the above-related works, Table 1. presents criteria to show the most important points that are used in each research.

Table 1. Comparison between related works.

Related Works	Methodologies	Sample Characteristics	Accuracy of Results	Limitations
[9]	CNN, Deep Learning	14 species of birds	94.36 %	The system does not give a special mark on the predicted value parallel to the available label choices.
[10]	Nearest Centroid (NC) and Artificial Neural Network (ANN)	10 endemic bird species	96.70%	The complexity of the ANN classifier is always higher than the NC classifier, irrespective of using LDA as feature reduction since its complexity depends on the number of neurons in both the hidden and output layers and the activation function used.
[11]	T-MSNet	ESC-50 (2000) U8K (8732) DCASE 2021 (23040)	70.60 %	Hyperparameters of T-MSNet
[12]	CRNN	40 ms audio bird sound frames	88.50%	none
[13]	depthwise separable convolution (DSC)	229164, 183690 (training set) and 45924 (test set).	95.12 %	none
[14]	Resnet34	a square image of size 224 _ 224 pixels	74.00%	Data Mismatch and Bias
[15]	CNN	1000 images	93.00%	Some species are difficult to extinguish.
[16]	CNN with Rectified Linear Unit or the ReLU	8218 images for 60 bird species	93.19%	none
[17]	Faster R-CNN and YOLO	2751 images (about 55 images per category)	96.20%	none
[18]	ResNet-50	2814 spectrogram of bird calls	94.50%	The large number of syllables
[19]	shallow ANN	659 bird species	94.00%	none

3. Methodology

The CNN-based bird call classifier presented in this study was developed within the MATLAB 2021a environment, utilizing a computer setup comprising 16 GB of DDR4 RAM, an AMD Ryzen 5 3550H CPU with Radeon Vega Mobile GFX clocked at 2.10 GHz, and the Windows 10 operating system.

3.1 Data Collection

The main challenge in the adoption of CNNs is the availability of training data. Sound files are readily available for the more common species. Synthetic samples are created as a process of increasing the size of data in rare species. To handle this issue, many sounds of birds are used from the Xeno-canto repository that are obtained from different areas around the world. Table 2 lists the details of alpha codes associated with these birds [20].

Table 2. Birds' alpha codes

No	English Name	Scientific Name	Six-letter Alpha Codes
1	Eurasian Wigeon	Maraca Penelope	MARPEN
2	American Robin	Turdus migratorius	TURMIG
3	Eurasian Skylark	Alauda arvensis	ALAARV
4	Purple Honeycreeper	Cyanerpes caeruleus	CYACAE
5	White-winged Dove	Zenaida asiatica	ZENASI
6	Brown Shrike	Lanius cristatus	LANCRI
7	Northern Lapwing	Vanellus vanellus	VANVAN
8	Long-billed Hermit	Phaethornis longirostris	PHALON
9	Pigeon Guillemot	Cepphus columba	CEPCOL
10	Swallow Tanager	Tersina viridis	TERVIR

3.2 Data Processing

The training dataset audio files are downloaded from the Xeno-canto repository for 10 different bird species and then processed to produce 16 kHz samples. Because the number of available files is limited, several 2-second sound clips are generated from audio files using voice activity detection (VAD), as shown in Fig. 1 [21].

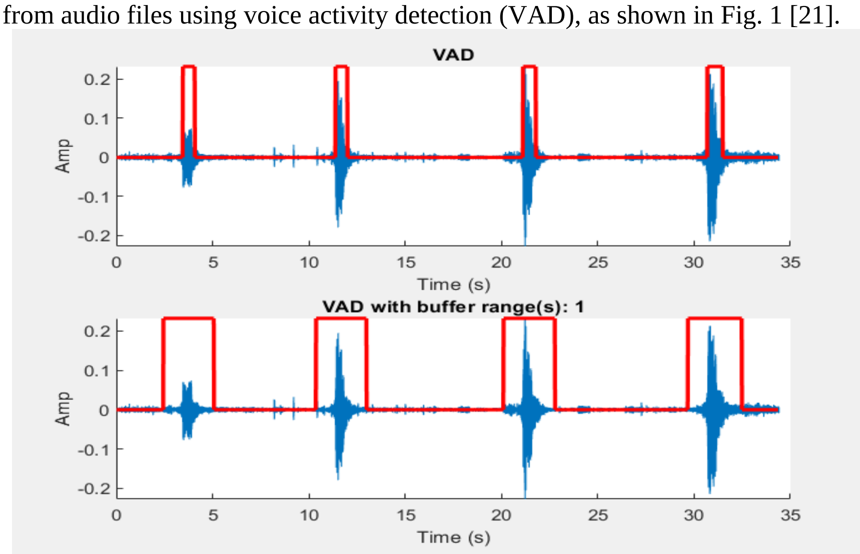

Figure 1. Framing bird vocalizations by Voice Activity Detection (VAD).

Spectrogram images that represent the coloured frequency values of the audio are obtained by converting each sound clip after segmentation. Mel Frequency Cepstral Coefficients (MFCC) algorithm [22,23] is used to convert sound to image, as shown in Fig. 2.

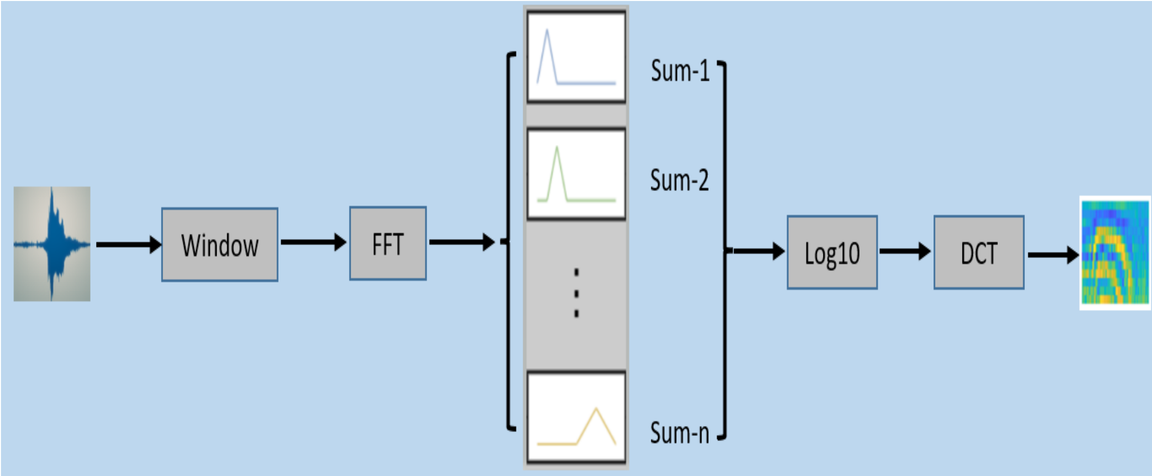

Figure 2. MFCC image generating process.

MFCC is applied as in Equation (1), to extract features from the signal, using MATLAB, the parameters of this function are as follows: (S: Signal, Fs: Frequency, n: Number of FFT points, Tf: Frame duration in seconds, N: Number of samples per frame as in Equation (2), Fn: Number of mel filters, L: total number of samples in speech, Ts: Frame step in seconds, Frame-Step: Frame step in samples as in Equation (3), noFrames: Maximum no of frames in speech sample as in Equation (3), FMatrix: Matrix to hold cepstral coefficients as in Equation (5)).

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

$$N = Fs * Tf \tag{2}$$

$$FrameStep = Fs * Ts \tag{3}$$

$$noFrames = floor\left(\frac{1}{FrameStep}\right) \tag{4}$$

$$FMatrix = zeros\,(noFrames - 2\,, num) \tag{5}$$

The Mel scale of the normal frequency scale denoted as 'm', will be used. The spectrograms will be resized to 224x224x3 to match the dimensions shown in Figure 3. As a result, the classifier will be trained using 1,000 samples for 10 different bird species.

3.3 Model Architecture

In this model, five hidden layers are used. Each layer contains additional layers (batch normalization, leaky Relu, max-pooling layers, to address weight scattering and take important features from images. The Convolution Neural Network involving these parts of layers is shown in Fig. 3 and its details are presented in Tab 3.

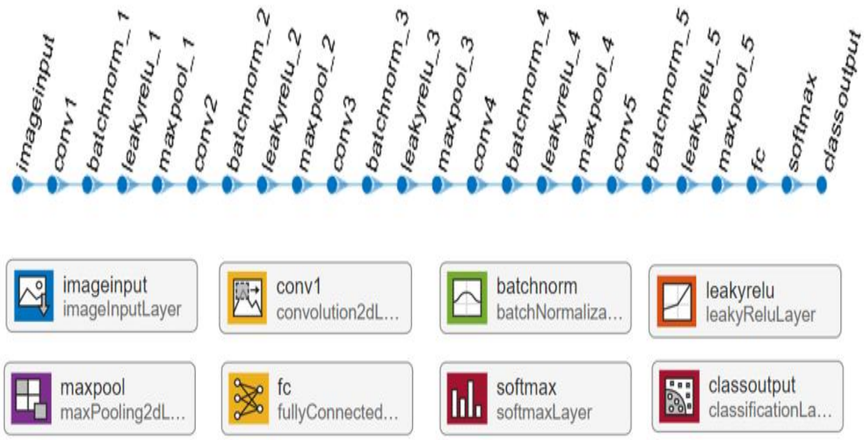

Figure 3. LWCNN Architecture.

Table 3. Details of the proposed LWCNN model.

Name of layer	Decimation	# Of Filter	Padding	Stride
Input	224 224 3			
Conv1	3 3	8	same	
batch normalization				
leaky Relu	0.01	1		
max-pooling	2 2			2 2
Conv2	3 3	16	same	
batch normalization				
leaky Relu	0.01			
max-pooling	2 2			2 2
Conv3	3 3	32	same	
batch normalization				
leaky Relu	0.01			
max-pooling	2 2			2 2
Conv4	3 3	16	same	
batch normalization				
leaky Relu	0.01			
max-pooling	2 2			2 2
Conv5	3 3	8	same	
batch normalization				
leaky Relu	0.01			
max-pooling	2 2			2 2
Fully Connect		10		
Softmax		To classification.		
classification		1, 2, …,10.		

3.4 Model Training

MATLAB is used to implement the proposed classifier. The stages of the training of 10,000 samples are explained as follows:

- Data Calling: For loading and managing datasets, Image Datastore is used.
- Checking the data: For checking and summarizing the number of images in species, “CountEachLabel” is used.
- Convolutional Neural Network (CNN): Design and implementation of LWNET model.
- Data Preprocessing: During this phase, the dataset will undergo division into training and validation sets. Specifically, 80% of the samples within each category will be allocated for the training dataset, while the remaining 20% will form the validation dataset. To ensure an unbiased selection, a randomization process is employed.
- Output Layer: 10 deep CNN layers are used as output layers.

4. Results

This section discusses the training and validation results of the LWCNN model, where the accuracy is benchmarked with the previous study.

4.1 Training Results

In the proposed classifier, there are 8,000 images extracted from the training dataset, where each bird species derives 800 spectrogram images. These images were processed in the training stage of a new model (LWNET):

In the LWNET model, five epochs are applied, which means that each type of dataset is processed in the training stage five times. Thus, there are 1250 iterations and 250 iterations per epoch. The details of the results are shown in Table 4.

Table 4. Comparison between the results of the previously proposed model [24] and the new proposed model.

C	Previous study		New study
	(ResNet-50) + MFCC	(MobilNet) + MFCC	LWCNN + MFCC
Training accuracy	**90.56%**	**85.73%**	**99.68%**
Time	**523 min**	**187**	**1.54 min**
epoch	**5**	**5**	**5**

Figure 4 illustrates the evolution of accuracy and loss during the training process. As accuracy steadily rises with each epoch until it reaches a saturation point, where it remains stable with minimal fluctuations. Conversely, the loss decreases as the number of epochs decreases, eventually reaching a saturation level where it stabilizes.

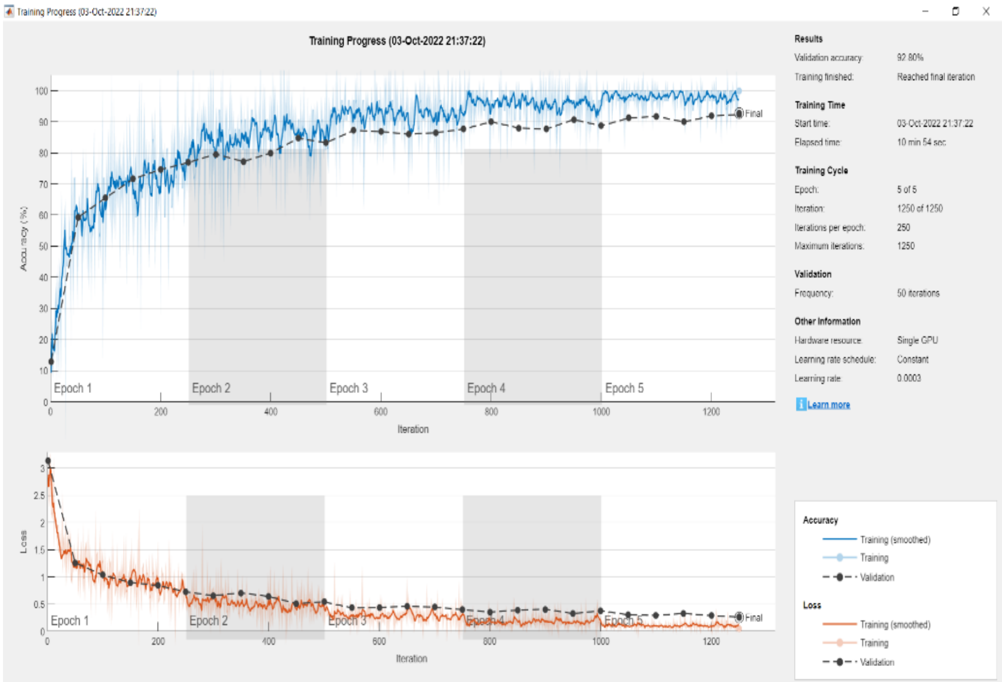

Figure 4. Illustrate the accuracy of the training datasets.

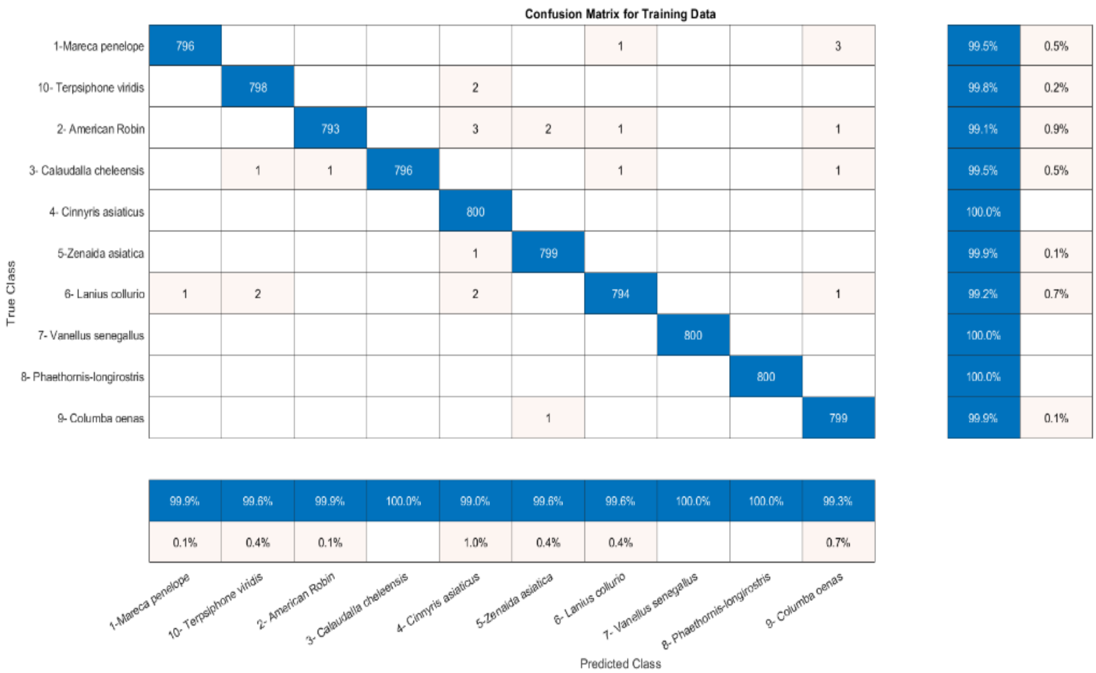

Figure 5. Illustrate training confusion matrix for WLCNN with MFCC.

The model's ability to anticipate the class that will be utilized in supervised learning is displayed in a confusion matrix. The proposed model demonstrates precise predictions for the ten classes within the validation test. The matrix displays the count of predictions for each class in its respective column, while the occurrences within the actual classes are depicted in the rows. Both the previous study [24] and the present study's results are documented in the confusion matrix presented in Figure 5.

4.2. Validation Results

The second partition of the dataset is the validation dataset probability distributed in the same way that training data are distributed. There are 2000 images used in the proposed classifier, where from 10 species there are 200 images from each species. Table 1 presents alpha codes for all ten birds [20]. A confusion matrix is a method of validating the proposed model's ability to accurately predict the class. The proposed model predicts the classes of the ten species during the validation test. In the confusion matrix, rows represent instances of the real class, while columns represent the number of predictions. The outputs that were correctly predicted are highlighted. It displays the prediction using the colour hue. When the accuracy is high, the block is dark, and when the accuracy is low, the block is light.

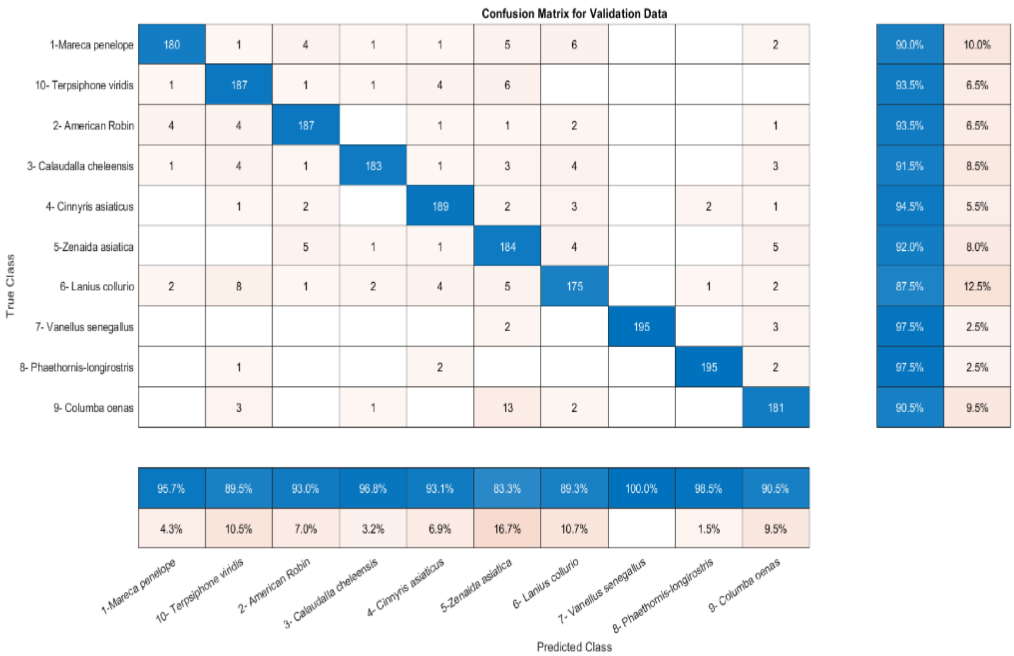

Figure 6. Validation confusion matrix for WLCNN with MFCC.

4.3 Discussion

Accuracy serves as a metric to evaluate the effectiveness of the proposed model. It gauges the model's ability to make accurate predictions when presented with new, unseen data. On the other hand, time complexity quantifies the duration needed to complete the testing process. The following formula is used to calculate accuracy:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ predictions} \quad (6)$$

The performance measures used in this study are the most widely used metrics such as sensitivity, specificity, and precision for each class [25,26] as shown in Tab 5.

Table 4: Results of (Specificity, Sensitivity/Recall, Precision and F1-Score).

True Positive *TP*	*False Positive* *FP*	$Precision = \frac{TP}{TP + FP}$
False Negative *FN*	*True Negative* *TN*	$Negative\ Predictve\ Value(NPV) = \frac{TN}{TN + FN}$
$Sensitivity\ /Recall = \frac{TP}{TP + FP}$	$Specificity = \frac{TN}{TN + FP}$	$F1 - Score = \frac{Precision\ + Recall}{Precision\ + Recall}$

Where is:

The number of correctly labelled positive samples is called True Positive (TP), the number of negative samples incorrectly labelled as positive is called False Positive (FP), the number of correctly labelled negative samples is called True Negative (TN) and the number of positive samples incorrectly labelled as negative is called False Negative (FN).

Table 5. Evaluation metrics results by using WLCNN + (MFCC)

NAME OF CLASS	Ac.%	Pre. %	Sen.%	Spe. %	F-measure%
MARPEN	99	96	90	100	93
TURMIG	98	91	91	99	91
ALAARV	99	93	94	99	93
CYACAE	99	97	94	100	96
ZENASI	98	94	92	99	93
LANCRI	97	83	88	98	86
VANVAN	99	89	99	99	94
PHALON	100	98	98	100	98
COLUOE	100	98	98	100	98
TERPVI	98	89	94	99	91

CONCLUSION

The WLCNN (Low-Complexity Convolutional Neural Network) model is employed to develop a bird-call classifier. This classifier is designed to accurately identify bird species based on their calls. To train the model, the Xeno-Canto repository is utilized as a source of bird sounds, specifically focusing on ten different bird species. To prepare the data for training, short 2-second clips are extracted from the sound samples. To ensure the quality of the training data, a Voice Activity Detection (VAD) technique is employed to filter out background noise. Additionally, to enhance the diversity of the dataset, more short clips are generated, resulting in a total of 1,000 sound clips for each bird species. To extract meaningful features from the audio clips, they are converted into spectrograms using the Mel-Frequency Cepstral Coefficients (MFCC) technique. This transformation allows the model to capture the frequency content of the bird calls effectively. Furthermore, the size of the spectrograms is adjusted to 244x224 pixels to match the input requirements of the CNN architecture. Finally, the performance of the WLCNN model is compared to two other CNN models: ResNet-50, which is a high-complexity CNN model, and MobileNet, which is a low-complexity CNN model. The purpose of this comparison is to evaluate the effectiveness of the proposed classifier while considering factors such as model size and testing time. The results of the evaluation indicate that the WLCNN model achieves satisfactory accuracy in bird species classification without compromising the size of the model or the time required for testing. This suggests that the WLCNN model strikes a balance between accuracy and computational efficiency, making it a practical choice for bird-call classification tasks.

FUNDING

None

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

“The authors affirm that they have no financial or other conflicts of interest to disclose about the current work”.

REFERENCES

[1] Martinson, J., ‘Bird Species Identification using Convolutional Neural Networks ‘, Master’s Thesis, 2017, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden.
[2] Renter´ıa, S., E. Vallejo, E., and E. Taylor, C., ‘ Birdsong Phrase Verification and Classification Using Siamese Neural Networks’, bioRXiv, Preprint, https://doi.org/10.1101/2021.03.16.435625
[3] Omkarini, V., Mohan, G., K., 2021, ‘Automated Bird Species Identification Using Neural Networks ‘, Annals of R.S.C.B., Vol. 25, Issue 6, Pages. 5402 – 5407.
[4] Figueiredo, N., Felix, F., Medeiros, C., B., and Queiroz, M., ‘A comparative study on filtering and classification of bird songs’, 15th Sound and Music Computing Conference (SMC2018), Limassol, Cyprus, 4-7 July, 2018.
[5] Tachibana, R., O., Oosugi N., Okanoya, K., ‘Semi-Automatic Classification of Birdsong Elements Using a Linear Support Vector Machine’, 2014, PLoS ONE, volume (9), no. (3): e92584. https://doi.org/10.1371/journal.pone.0092584.
[6] Himawan, I., Towsey, M., & Roe, P., ‘ 3D convolution recurrent neural networks for bird sound detection’. 2018, In Wood, M, Glotin, H, Stowell, D, & Stylianou, Y (Eds.) Proceedings of the 3rd Workshop on Detection and Classification of Acoustic Scenes and Events. Detection and Classification of Acoustic Scenes and Events, pp. 1-4.
[7] Nanni, L., Costa, Y.M.G., Aguiar, R.L. et al., ‘Ensemble of convolutional neural networks to improve animal audio classification’. EURASIP Journal on Audio, Speech, and Music Processing, 2020.
[8] Zhang, F., Zhang, L., Chen, H., and Xie J., ‘Bird Species Identification Using Spectrogram Based on Multi-Channel Fusion of DCNNs’. Entropy (Basel). 2021 Nov 13;23(11):1507.
[9] Permana, S. D. H, Saputra, G., Arifitama, B., et al., ‘Classification of bird sounds as an early warning method of forest fires using Convolutional Neural Network (CNN) algorithm’, Journal of King Saud University - Computer and Information Sciences, 2022, Volume 34, Issue 7, pp 4345–4357
[10] Ramashini, M., Abas, P. E., Silva, L., C., D., ‘A Novel Approach of Audio Based Feature Optimisation for Bird Classification, Pertanika Journal of Science & Technology (Malaysia), 29 (4), p. 2383-2407.
[11] Kek, X. Y., Chin, C. S., and Li, Y., "An Intelligent Low-Complexity Computing Interleaving Wavelet Scattering Based Mobile Shuffling Network for Acoustic Scene Classification," in IEEE Access, 2022, vol. 10, pp. 82185-82201.
[12] Cakir, E., Adavanne, S., Parascandolo, G., et al., ‘Convolutional Recurrent Neural Networks for Bird Audio Detection’, 2017, 25th European Signal Processing Conference (EUSIPCO), pp. 1744-1748.
[13] Yang, F., Jiang, Y., and Xu, Y., "Design of Bird Sound Recognition Model Based on Lightweight," 2022, in IEEE Access, vol. 10, pp. 85189-85198.
[14] Philip, E., Callan, A., Paul, R., etal., ‘A Convolutional Neural Network Bird Species Recognizer Built from Little Data by Iteratively Training, Detecting, and Labeling’, Frontiers in Ecology and Evolution, 2022, v. (10).
[15] Tamilalagan, S., ‘Image-based Bird Species Identification Using Convolutional Neural Network’, International Journal of Innovative Research in Technology, 2022, volume (8), Issue (11).
[16] Raj, S., Garyali, S., Kumar, S., et al., ‘Image-based Bird Species Identification using Convolutional Neural Network’, International Journal of Engineering Research & Technology (IJERT), 2020, Vol. 9 Issue 06.
[17] Li, C., Zhang, B., Hu, H., and et al., ‘ Enhanced Bird Detection from Low-Resolution Aerial Image Using Deep Neural Networks’, Neural Processing Letters, Volume (49), Issue (3), pp 1021–1039.
[18] Sankupellay, M., Konovalov, D., ‘Bird Call Recognition using Deep Convolutional Neural Network, ResNet-50’, Proceedings of ACOUSTICS, Adelaide, Australia, 2018.
[19] Ghani, B., Hallerberg, S., ‘A Randomized Bag-of-Birds Approach to Study Robustness of Automated Audio Based Bird Species Classification’, Appl. Sci., 2021, 11, 9226. https://doi.org/10.3390/app11199226.
[20] A. Fritzler, S. Koitka, and C. M. Friedrich, (2017). “Recognizing bird species in audio files using transfer learning.” in CLEF Working Notes.
[21] M. H. Moattar and M. M. Homayounpour, (2009). “A simple but efficient real-time voice activity detection algorithm,” in 2009 17th European Signal Processing Conference. IEEE, pp. 2549–2553.
[22] C.-H. Lee, C.-C. Han, and C.-C. Chuang, (2008). “Automatic classification of bird species from their sounds using two-dimensional cepstral coefficients,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 8, pp. 1541–1550.

[23] S. ahmed, Zainab Ali Abbood, hameed Mutlag Farhan, Baraa Taha Yasen, Mohammed Rashid Ahmed, and Adil Deniz Duru, “Speaker Identification Model Based on Deep Neural Networks ”, Iraqi Journal For Computer Science and Mathematics, vol. 3, no. 1, pp. 108–114, Jan. 2022.
[24] Saad, A., Ahmed, J., Elaraby, A. (2022). Classification of bird sound using high-and low-complexity convolutional neural networks. Traitement du Signal, Vol. 39, No. 1, pp. 187-193. https://doi.org/10.18280/ts.390119.
[25] Osama abdulaziz and O. A. Saltykova, “CNN-PS: Electroencephalogram Classification of Brain States Using Hybrid Machine - Deep Learning Approach”, Iraqi Journal For Computer Science and Mathematics, vol. 4, no. 4, pp. 63–75, Oct. 2023.
[26] Hadjouni, M., Elmannai, H., Saad, A., Altaher, A., Elaraby, A. (2023). A novel deep learning approach for brain tumors classification using MRI images. Traitement du Signal, Vol. 40, No. 3, pp. 1137-1144. https://doi.org/10.18280/ts.400329.