EasyChair Preprint

№ 15712

Towards Classifying Bird Sounds Using a Deep Transfer Learning Model

Saptarshi Dey, Soumi Ghosh, Soumapriyo Mondal, Akash Harh, Spandan Bandhu, Bidhan Barai and Pawan Kumar Singh

January 14, 2025

Towards Classifying Bird Sounds Using a Deep Transfer Learning Model

Saptarshi Dey[1], Soumi Ghosh[1], Soumapriyo Mondal[1], Akash Harh[2], Spandan Bandhu[2], Bidhan Barai[3,*], Pawan Kumar Singh [4, 5, [0000-0002-9598-7981]]

[1]Department of Computer Science and Engineering (Data Science), Techno Main Salt Lake, EM-4/1, Sector-V, Salt Lake, Kolkata-700091, West Bengal, India

[2]Department of Computer Science and Engineering (Data Science), Techno Main Salt Lake, EM-4/1, Sector-V, Salt Lake, Kolkata-700091, West Bengal, India

[3]Department of Computer Science and Engineering (AI-ML), Techno Main Salt Lake, EM-4/1, Sector-V, Salt Lake, Kolkata-700091, West Bengal, India

[4]Department of Information Technology, Jadavpur University, Jadavpur University Second Campus, Plot No. 8, Salt Lake Bypass, LB Block, Sector III, Salt Lake City, Kolkata, Pin: 700106, West Bengal, India

[5]Shinawatra University, 99, Moo 10, Bang Toei, Sam Khok, Pathum Thani, Thailand, 12160

{saptarshidey2120@gmail.com,ghoshsoumi562@gmail.com,soumapriyomondal1@gmail.com,akashharh2002@gmail.com,bandhuspandan@gmail.com, bidhan.ju.cse@gmail.com, pawansingh.ju@gmail.com }

*Corresponding author

Abstract. The conservation of bird biodiversity relies on accurately identifying and classifying species, which is often time-consuming and requires specialized knowledge. Recent advances in deep learning, particularly in convolutional neural networks (CNNs), have made it possible to detect species passively from acoustic signals, even in challenging environments. This paper presents a high-performance deep convolutional neural network (CNN) model using the VGG-16 architecture for the passive classification of bird sounds, using a remarkably accurate model of Short-Time Fourier Transform (STFT) that accounts for 97.31% of the BirdCLEF 2022 data set and 98.41% for the Cornell Birdcall Identification dataset. The model discriminates between species, even in complex soundscapes with overlapping records. The framework also uses a tool-based consensus framework to enhance the focus on relevant features, improving classification accuracy for rare and endangered species. This method is highly effective in various phonological and language processing tasks and enhances the model's robustness, making it suitable for real-world applications.

.

Keywords: Bird species classification**,** Mel - Spectrogram, Short-time Fourier transform, Bird Clef 2022 dataset, VGG-16, Feature Extraction, Cornell Birdcall Identification dataset.

1 Introduction

Avian biodiversity is crucial for preserving ecological stability, and correct identification and monitoring of bird species are important for information on biodiversity traits and implementing conservation techniques [1], as the motivation to choose this project relies precisely on those grounds. Bird species identification through their vocalizations draws crucial inferences about ecosystem health. These are cumbersome to do manually and prone to errors, especially across large datasets and varied environments; hence, it creates the need for automated solutions.

Traditional methods, along with visual statements or guide evaluations, are exertion-intensive, prone to human errors, and time-ingesting. As environmental data becomes more abundant and diverse, the manual process becomes less scalable, particularly for large datasets collected over long periods or across wide geographic areas. With many bird species facing threats from habitat loss, climate change, and other environmental factors so the global efforts to protect endangered species have intensified, there may be a developing call for automated systems that could efficiently classify bird vocalizations in various and complex soundscapes [2] with robust and scalable approach. Recent advances in machine learning, mainly deep learning, have significantly improved the ability to automate sound classification. We select deep neural networks because deep neural networks, CNNs, have extraordinary capability in capturing intricate spectral and temporal features in audio signals. They can handle the spectral nature of spectrograms and are hence best suited for that purpose. Finally, VGG-16 layers have been added to some extent due to their pre-training skills and feature extraction ability so as to do better in identifying complex patterns. Thus, with this combination, strong and effective bird sound classification can be achieved, outperforming traditional classifiers, while at the same time minimizing the hassle of training such models. This study aims to develop an automated bird sound classification system that not only enhances classification accuracy but also leverages impact on avian biodiversity conservation by providing a faster and scalable monitoring solution.

The research presents a hybrid deep mastering model that uses a Conv2D layer and VGG-16 architecture for automatic bird sound classification. It uses Short-Time Fourier Transform (STFT) to transform audio indicators into Mel spectrograms and the Viridis shade map for feature extraction. The BirdCLEF 2022 and the Cornell Birdcall Identification dataset, comprising 15,000 and 4,733 audio samples, serve as the premise for training and assessment. Please refer to Fig.1 for the overall block diagram of our proposed deep CNN framework for solving automated bird sound classification problems using a deep transfer learning model.

The highlighting features of our proposed approach are as follows:

1. The research introduces a hybrid deep learning model combining Conv2D layers and VGG-16 architecture for automatic bird sound classification.

2. The Short-Time Fourier Transform (STFT) converts audio signals into Mel spectrograms, with the Viridis color map applied for feature extraction.
3. We have used two diverse datasets ‘BirdCLEF 2022’ and ‘Cornell Birdcall Identification’ datasets for broadening the model's learning scope, which has 25 and 50 bird classes respectively.
4. The proposed deep transfer learning model architecture performs better than other state-of-the-art methods.

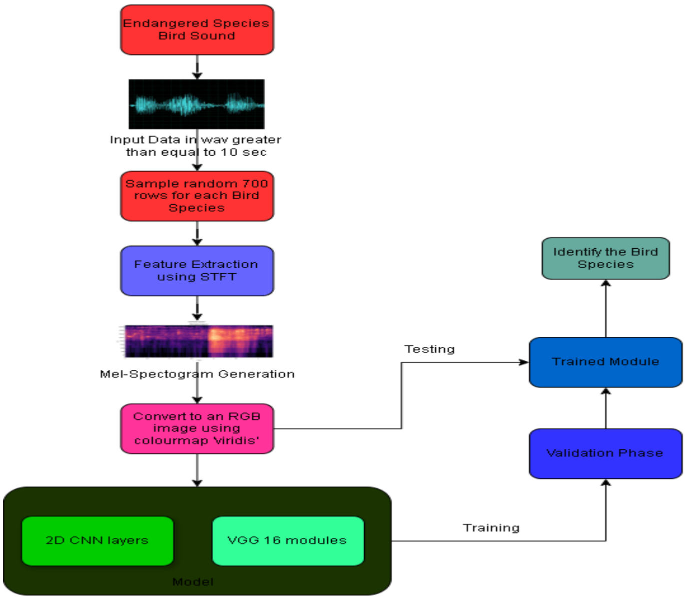

Endangered Species Bird Sound
Input Data in wav greater than equal to 10 sec
Sample random 700 rows for each Bird Species
Feature Extraction using STFT
Mel-Spectogram Generation
Convert to an RGB image using colourmap Viridis'
2D CNN layers
VGG 16 modules
Model
Training
Validation Phase
Trained Module
Testing
Identify the Bird Species

Fig. 1. Architectural diagram of our proposed deep CNN framework for classification of endangered bird’s sounds.

2 Related Study

This section discusses the advancements in automated bird species classification based on acoustic analysis. Noumida et al. [3] used an Attention-BiGRU version for actual-time bird species classification using the Xeno-canto database, achieving an F1-score of 0.84. Hong [4] used CNNs for bird species classification using the OpenVINO tool, with a dataset of 119,000 audios covering 834 bird species. Yang et al. [5] developed a lightweight bird sound recognition model using MobileNetV3, a multiscale feature fusion structure, and a Pyramid Split Attention module, achieving a Top-1 accuracy of 95.12% and Top-5 accuracy of 100% on a dataset of 264 bird species. Sun et al. [6] presented a lightweight model with frequency dynamic convolution for bird species identification, reaching 95.21% accuracy. Huang et al. [7] employed a transfer learning approach with Inception-ResNet-v2 to classify bird species endemic to Taiwan, achieving 98.39% accuracy. Lucio et al. [8] utilized texture capabilities extracted from spectrogram photographs, achieving 77.65% accuracy on a 46-class bird species dataset. Liu et al. [9] suggested a Bi-LSTM-DenseNet model for bird song categorization, outperforming existing neural networks in detecting diverse bird species. Ragib et al. [10] proposed a deep learning model utilizing a pre-trained ResNet network to identify individual birds from images, enhancing image classification accuracy in complex scenarios. Gupta et al. [11] utilized a hybrid CNN-RNN model, achieving an average accuracy of 67% across 100 species, with a peak accuracy of 90% for Red Crossbill. Mehyadin et al.[12]used Mel-frequency cepstral coefficients (MFCC) to analyze bird calls and employ machine learning techniques for species identification. Among the tested algorithms, J4.8 achieved the highest accuracy at 78.40%, proving the most effective for classifying bird species. Non-desirable noise is filtered using noise suppression techniques. Koh et al.[13] employed Inception and ResNet models to classify 659 bird species from 50,000 audio recordings in the BirdCLEF 2019 competition. Despite challenges like signal-to-noise ratio mismatch, the Inception model achieved a classification mean average precision (c-mAP) of 0.16. Heinrich et al.[14] introduced ProtoPNet with a ConvNeXt backbone for bird sound classification, focussing on interpretability. The model uses spectrograms to extract features and classify species by comparing new data with learned prototypical patterns. Achieving an AUROC of 0.82 and cmAP of 0.37, it rivals state-of-the-art black-box models. Incze et al.[15] refined the pre-trained MobileNet CNN model for bird sound classification using spectrograms from Xeno-Canto recordings. Experiments compare various configurations, showing that aligning the color map with pre-trained image data enhances performance. The system is effective for a limited number of bird species.

The research introduces a hybrid model that integrates Conv2D layers with the VGG-16 architecture for bird sound classification, aiming to deliver high accuracy while ensuring robustness across diverse environmental conditions and dataset sizes. Utilizing Short-Time Fourier Transform (STFT) to convert audio signals into mel spectrograms, the model processes these with the Viridis color map to ensure precise classifications. Data augmentation techniques are employed to address class imbalance, ensuring consistent model performance throughout training. Utilizing both the 'BirdCLEF 2022' and 'Cornell Birdcall Identification datasets offers a unique opportunity to train a model with diverse species and environmental conditions, fostering robustness and accuracy in bird species classification across varied contexts and regions.

The motivation for this work arises from the growing need for efficient, automated bird sound classification, essential for ecological monitoring and conservation. Manual identification of bird species through their vocalizations is time-consuming and prone to inaccuracies, whereas an automated system can significantly enhance the ability to track species populations and understand ecosystem health. This approach offers a scalable solution to monitor avian biodiversity more effectively.

3 Datasets Used and Dataset Pre-processing

Bird calls are classified in this study using the BirdCLEF 2022 data set of 15,000 audio recordings from 25 species and the Cornell Birdcall Identification dataset of 4733 audio recordings from 50 species. Metadata in the data set is a combination of species labels and recording status, which can be important for type challenges. The data set has been preprocessed with the aid of transforming the raw audio signals into Mel spectrograms using the transient Fourier transform (STFT), which preserves the temporal and spectral features and the subsequent Mel spectrograms, which is preferable to using Viridis color maps. The dataset size has been changed to 128x431 pixels with 3 channels (RGB format) to ensure compatibility with the VGG-hexadecimal size used in the hybrid version. The preprocessing pipeline additionally deals with class imbalances related to information enhancement techniques such as random cropping, pitch shifting, and time-related extensions. Fig.2. and Fig.3 show the Mel-Spectrogram images of bird species from the 'Bird Clef 2022' and 'Cornell Birdcall Identification ' datasets respectively.

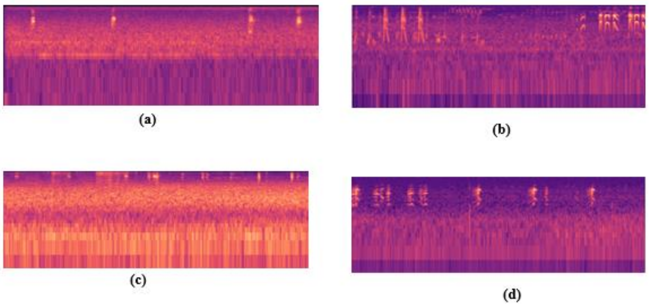


Fig. 2 Sample Mel - Spectrogram images of (a) Black-Crowned Night Heron (b) California Quail (c) Common Waxbill (d) Canada Goose bird species from the 'Bird Clef 2022' dataset.

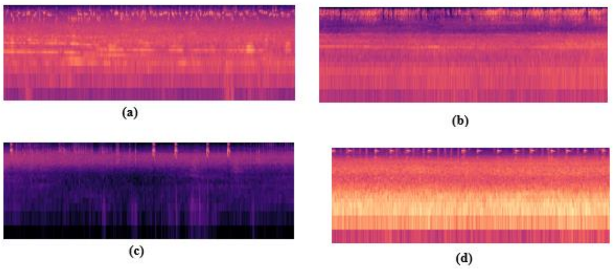
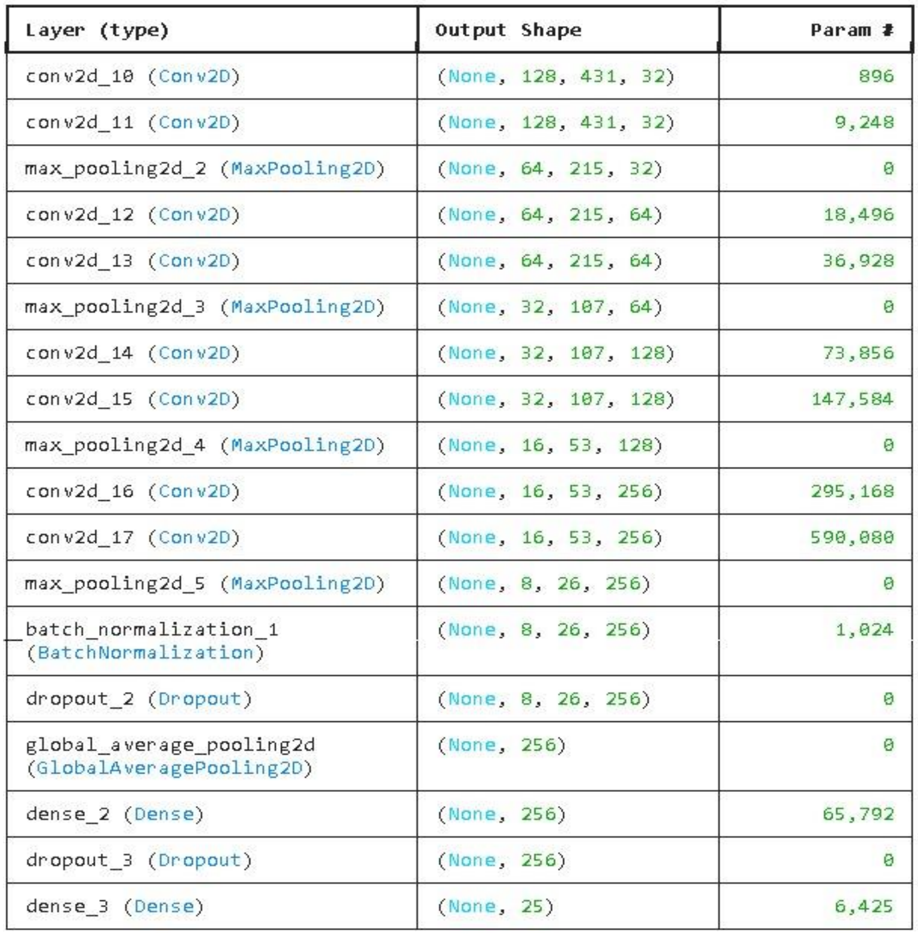


Fig. 3. Sample Mel - Spectrogram images of (a) Brown Thrasher (b) Barn Swallow (c) Northern Flicker (d) Loggerhead Shrike bird species from the 'Cornell Birdcall Identification ' dataset.

4 Proposed Deep CNN Framework

The proposed deep CNN framework is designed to classify bird species based on their vocalizations using a model inspired by the VGG-16 structure. The network uses mel-spectrograms from bird audio recordings, which are fed into the network with a length of 128x431 pixels and 3 color channels. Fig. 4 illustrates that the framework starts with a chain of Conv2D layers, each configured to capture time-frequency functions from the Mel-spectrograms. The layers gradually evolve to 64 and 128 filters, enhancing the community's ability to analyze complex features. The Global Average Pooling (GAP) layer condenses each feature map to an unmarried price, reducing the model's parameters and computational value. The final dense layer consists of 25 and 50 classes, representing the 25 and 50 bird species in the dataset, with a Softmax activation characteristic that outputs the classification chances. This structure presents a strong and green answer for automatic bird sound classification, achieving high accuracy and overall performance metrices. The Table 1 outlines the structure of a Convolutional Neural Network, listing each layer's type, output shape, and parameter count. It includes Conv2D, MaxPooling2D, BatchNormalization, Dropout, and Dense layers, highlighting the network's hierarchical design and parameter efficiency.

Table 1. Layer, Output Shape, and Number of Parameters in the Convolutional Neural Network Architecture.



Layer (type)	Output Shape	Param #
conv2d_10 (Conv2D)	(None, 128, 431, 32)	896
conv2d_11 (Conv2D)	(None, 128, 431, 32)	9,248
max_pooling2d_2 (MaxPooling2D)	(None, 64, 215, 32)	0
conv2d_12 (Conv2D)	(None, 64, 215, 64)	18,496
conv2d_13 (Conv2D)	(None, 64, 215, 64)	36,928
max_pooling2d_3 (MaxPooling2D)	(None, 32, 107, 64)	0
conv2d_14 (Conv2D)	(None, 32, 107, 128)	73,856
conv2d_15 (Conv2D)	(None, 32, 107, 128)	147,584
max_pooling2d_4 (MaxPooling2D)	(None, 16, 53, 128)	0
conv2d_16 (Conv2D)	(None, 16, 53, 256)	295,168
conv2d_17 (Conv2D)	(None, 16, 53, 256)	590,080
max_pooling2d_5 (MaxPooling2D)	(None, 8, 26, 256)	0
batch_normalization_1 (BatchNormalization)	(None, 8, 26, 256)	1,024
dropout_2 (Dropout)	(None, 8, 26, 256)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 256)	0
dense_2 (Dense)	(None, 256)	65,792
dropout_3 (Dropout)	(None, 256)	0
dense_3 (Dense)	(None, 25)	6,425

Algorithm 1: Proposed_Deep Transfer_Learning_Model

```
  Input: Bird sound dataset (BirdCLEF 2022, Cornell Birdcall)
 Output: Trained hybrid model and performance metrics

 # Step 1: Load and Pre-process Data
  Procedure Preprocess_Data
     Load bird audio files and metadata
     For each audio file in the dataset do
              Apply STFT to audio signal
              Generate Mel-Spectrogram from STFT
              Normalize Mel-Spectrogram pixel values to range [0, 1]

End For
Perform
Data Augmentation:
```

```
            Apply random shifts to Mel-Spectrograms
            Apply pitch changes to Mel-Spectrograms
            Apply time stretching to Mel-Spectrograms
        End Procedure
    # Step 2: Design Hybrid CNN Model
        Procedure Design_Model
            Initialize Sequential Hybrid CNN Model with VGG-16 Layers
            Add Additional Convolutional Layers to the model
            Add Regularization Layers:
            Add BatchNormalization layer
            Add Dropout layer with rate = 0.5
        End Procedure
    # Step 3: Classification Layer
        Procedure Add_Classification_Layer
            Add Dense Layer with units = 256 and activation = 'ReLU'
            Add Dropout layer with rate = 0.5
            Add Output Layer with activation = 'softmax'
        End Procedure
    # Step 4: Compile and Train the Model
        Procedure Compile_Train_Model
            Compile Model:
                Set Loss Function to Categorical Crossentropy
                Set Optimizer to Adam
                Set Metrics to Accuracy
            Train the Model
        End Procedure
    # Step 5: Evaluate Model Performance
        Procedure Evaluate_Performance
            Input: Trained model, test data
            Evaluate model on the test dataset
            Calculate accuracy
            Calculate precision
            Calculate recall
            Calculate F1-score
            Generate a confusion matrix to analyze misclassifications
        End Procedure
    # Step 6: Post-training Fine-tuning
        Procedure Fine_Tune_Model
            Perform hyperparameter tuning:
                Tune learning rate
                Tune batch size
                Tune dropout rate
            Optimize hyperparameters
        End Procedure
        # Execute Procedures
         Call Prepro cess_Data
         Call Design_Model
         Call Add_Classification_Layer
         Call Compile_Train_Model
         Call Evaluate_Performance
         Call Fine_Tune_Model
End Algorithm
```

The pseudocode (shown in Algorithm 1) demonstrates the architecture and implementation details of the proposed hybrid deep convolutional neural network model used for automated bird sound classification. It highlights the key layers and operations performed during training and evaluation.

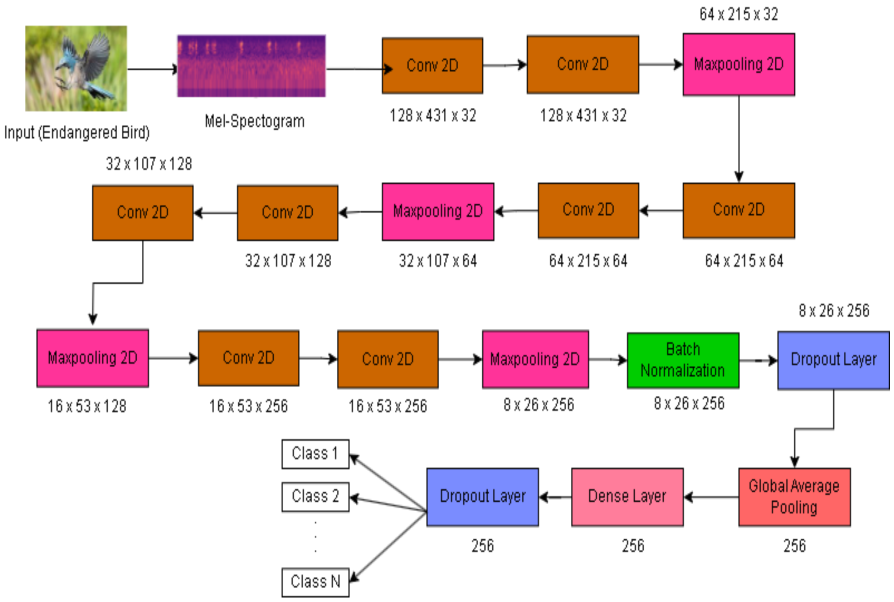


Fig. 4. Block diagram of our proposed deep transfer learning framework which consists of 2D-CNN incorporated with VGG-16 layers.

5 Experimental Results and Analysis

We evaluated the performance of our proposed deep CNN framework using well-known classification metrics: accuracy, precision, recall, and F1 score. Accuracy assesses the version's general correctness in species classification, even as precision displays the proportion of real tremendous predictions. Recall measures the version's effectiveness in detecting rare or subtle vocalizations, and the F1 score balances precision and don't forget, offering a comprehensive evaluation. These metrics spotlight the model's robustness, generalization functionality, and proficiency in shooting quality-grained acoustic capabilities crucial for species discrimination. The following section gives the experimental results of our custom model at the BirdCLEF 2022 dataset after 40 epochs. Table 2 shows the overall classification accuracy of our proposed custom CNN model for both datasets. The model has been trained on a dataset of 'BirdClef 2022' and achieved a recall of 96.47%, a precision of 98.07%, an F1 score of 97.02%, and an accuracy of 97.31%. The model was then tested on a separate dataset of 'The Cornell Birdcall Identification dataset', achieving a recall of 91.12%, a precision of 96.30%, an F1 score of 93.52%, and an accuracy of 98.41%. Table 3 represents the classification accuracy under different training/testing proportions on both Bird Clef 2022 and Cornell Birdcall Identification datasets. Table 4 and Table 5 show the classification report of our proposed hybrid deep CNN model in the 'Bird Clef 2022' and 'Cornell Birdcall Identification' datasets respectively.

Table 2. Experimental findings regarding evaluation metrics of our proposed hybrid deep CNN model on our used datasets.

Dataset Used	Accuracy (%)	Precision (%)	Recall(%)	F1 Score(%)
Bird Clef 2022	97.31	98.07	96.47	97.02
Cornell Birdcall Identification dataset	98.41	98.58	97.53	98.2

Table 3. Experimental results of our proposed hybrid CNN model under different training/testing proportions.

Dataset Used	Classification accuracy (%) for Various Training-Testing Ratios		
	75:25	**80:20**	**90:10**
Bird Clef 2022	95.86	97.31	98.50
Cornell Birdcall Identification dataset	97.23	98.41	98.64

Table 4. Classification report of our proposed hybrid deep CNN model of top 5 highly classified Bird classes in the 'Bird Clef 2022' dataset.

Class Name	Precision	Recall	F1 Score
California Quail	0.78	0.66	0.72
Green-Winged Teal	0.88	0.45	0.6
House Finch	0.91	0.4	0.55

Northern Cardinal	0.9	0.53	0.67
Western Meadowlark	0.7	0.61	0.65

Table 5. Classification report of our proposed hybrid deep CNN model of top 5 highly classified Bird classes in the ‘Cornell Birdcall Identification’ dataset.

Class Name	**Precision**	**Recall**	**F1 Score**
Common Raven	0.71	0.60	0.65
Marsh Wren	0.54	0.70	0.61
Blackpoll Warbler	0.83	0.74	0.78
House Wren	0.89	0.29	0.43
Brewer’s Sparrow	0.42	0.61	0.50

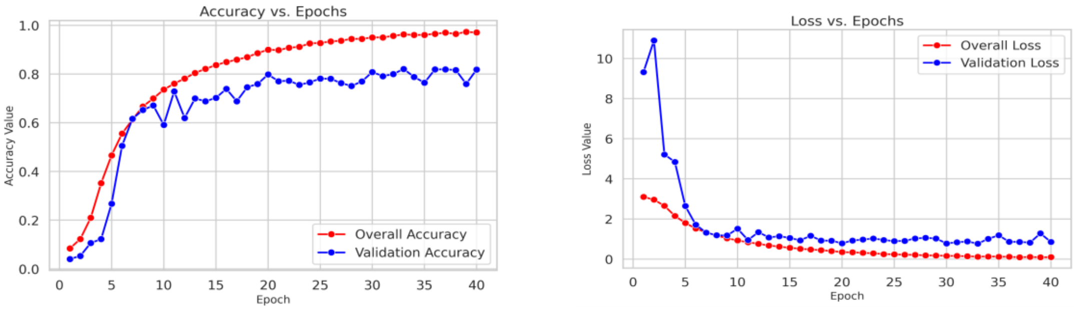


Fig.5. Learning rate curves illustrating the recognition accuracy and loss values attained per epoch on the ‘Bird Clef 2022’ dataset.

The development of model accuracy and loss over epochs during training on the dataset is shown by the learning curves in Fig.5. While the right subplot displays loss values, which demonstrate the difference between real labels and predictions, the left subplot displays accuracy trends, which represent the percentage of properly categorized occurrences. Here in the left subplot, we can see after running the 1st epoch accuracy starts at a low value, approximately 10%, indicating the model is just beginning to learn, after running 10 epochs accuracy improves significantly, reaching 75%, showing notable progress as the model adjusts its weights, after 20 epochs it stabilized near 90%, after 30 epochs it gained 95% accuracy and after running 40 epochs our final accuracy is steady at almost 97.31%, showing the model has converged effectively. These curves shed light on the model's convergence, showing that overfitting occurred after strong initial learning, as seen by the gradual divergence of training and validation measures.

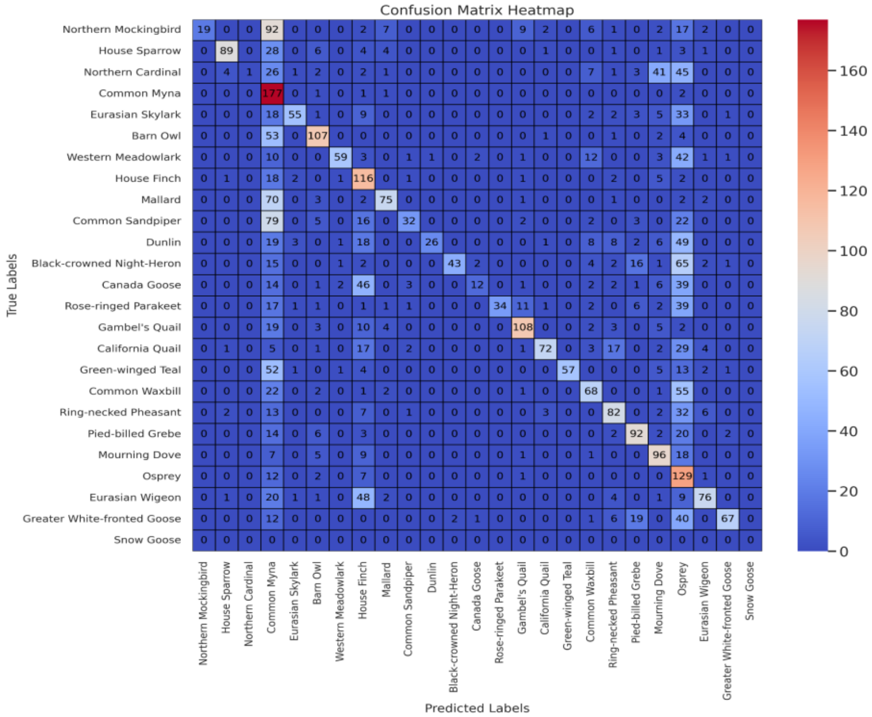


Fig.6. Confusion matrix of our proposed hybrid deep CNN framework on a test dataset of 'Bird Clef 2022' Dataset.

Fig.6. illustrates the classification effectiveness of the 'BirdClef 2022' Dataset, showcasing the classification hierarchy of the 25 classes. From this figure, we can easily understand the two most classified bird classes are 'Common Myna' and 'Osprey'. In the similar way, by analyzing the diagonal values of this matrix, we can depict two highly misclassified bird classes are 'Snow Goose' and 'Northern Cardinal'.

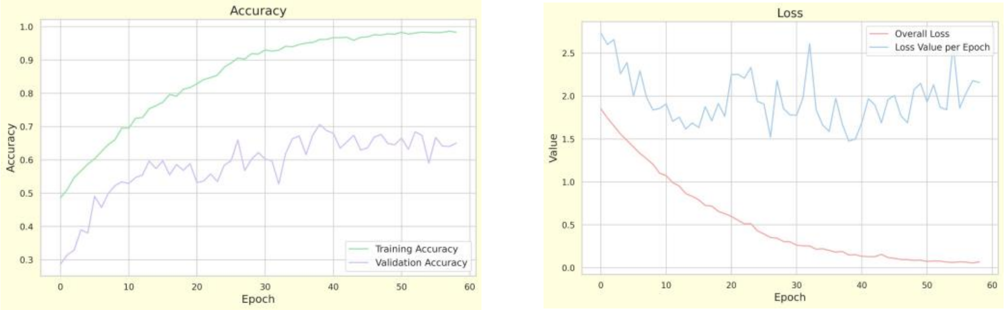


Fig.7. Learning rate curves illustrating the recognition accuracy and loss values attained per epoch on the 'Cornell Bird Identification' dataset

The learning rate curve (highlighted in Fig.7.) shows our model's training progress on the 'Cornell Bird Identification' dataset, highlighting the relationship between the learning rate of accuracy and loss values per epoch. Here the left subplot illustrates that after running the 1st epoch accuracy starts at 50%, indicating the model is beginning to learn, after running 10 epochs accuracy improves to around 70%, showing significant progress, then after 20 epochs the accuracy reached 83%, with steady improvement, after 30 epochs it approached 92%, showing continued learning., after running 40 epochs the model stabilizes further, with accuracy nearing 96%, after 50 epochs it achieved 98% accuracy and after running 60 epochs our final accuracy is steady at almost 98.41%, indicating the model has effectively learned the patterns.

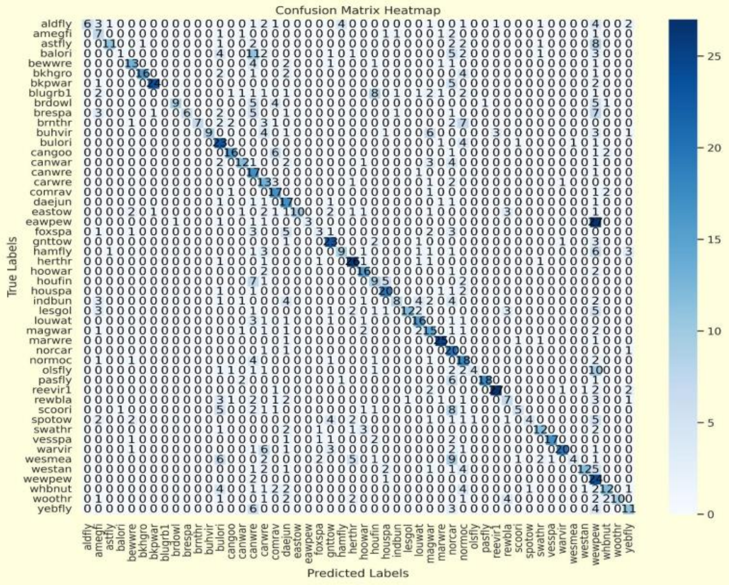


Fig.8. Confusion matrix of our proposed deep CNN framework on a test dataset of the ‘Cornell Bird Identification’ Dataset.

Fig.8. illustrates the classification effectiveness of the Cornell Bird Identification Dataset, showcasing the classification hierarchy of the 50 classes. From the figure, it is evident that the two most accurately classified bird species are 'Reevirl' and 'Herthr'. Similarly, analyzing the diagonal values of the matrix reveals that 'Balori' and 'Blugrbl' are the two most misclassified bird species.

5.1 Comparison with state-of-the-art works

The study evaluated the performance of a hybrid deep CNN version on the BirdCLEF and Cornell Bird Identification datasets. The Xception version achieved an accuracy of 80.66%[2], demonstrating the effectiveness of depthwise separable convolutions for characteristic extraction. However, the proposed version outperformed this with a 97.31% accuracy, indicating the benefits of incorporating both Conv2D and VGG-16 layers. The lightweight version with frequency dynamic convolution achieved a respectable accuracy of 95.21% [6]. In contrast, The custom CNN deep residual network achieved 80% accuracy[16] but fell quickly compared to the proposed method. Similarly, the custom CNN model finished with 90% accuracy[17], and CNN with fusion acoustic features recorded 95.25% accuracy[18]. Both models highlight the importance of customized architecture and characteristic fusion. For the Cornell Bird Identification dataset, the hybrid CNN-RNN model reached 67% accuracy[11], and the faster R-CNN model varied between 75% to 92.3% accuracy[1]. Overall, our proposed version sets a new benchmark in automatic bird sound type, validating the effectiveness of the hybrid technique. The comparison of our proposed efficient deep CNN model with state-of-the-art models for both datasets is shown in Table 6.

Table 6. State-of-art comparison of our proposed efficient deep transfer learning model.

Datasets Used	**Author Name**	**Publishing Year**	**Used Model**	**Accuracy**	**Accuracy of our proposed model**
‘Bird Clef’ dataset	Revadekar et al. [2]	2023	Xception	80.66%	**97.31%**
	Sun et al.[6]	2023	Lightweight Model with Frequency Dynamic Convolution	95.21%	
	Madhavi et al. [16]	2018	Custom CNN deep residual network	80%	
	Patil et al. [17]	2022	Custom CNN model	90%	
	Xie et al.[18]	2016	CNN with Fusion Acoustic Features	95.25%	
‘Cornell Bird Identification’ dataset	Gupta et al.[11]	2021	Hybrid CNN+RNN model	67%	**98.41%**
	Mirugwe et al. [1]	2022	Faster R CNN	92.3%	

6 Conclusion & Future Works

The study presents a hybrid deep CNN framework for identifying bird species based on their vocalizations, using the Conv2D structure and selected layers from the VGG-16 model. The model uses Short-Time Fourier Transform (STFT) for characteristic extraction and Mel spectrograms with a Viridis color map to capture complex spectral and temporal functions in chicken calls. Experiments on the BirdCLEF 2022 and Cornell Birdcall Identification datasets show the model's ability to achieve classification accuracy of 97.31% and 98.41%, surpassing modern methods. The model's design includes multiple Conv2D layers for extracting difficult capabilities, MaxPooling, Batch Normalisation, and Global Average Pooling layers for discriminative power and generalization skills. It addresses brightness imbalances through centered strategies and fine-tuning hyperparameters, ensuring high precision, recall, and F1 scores across various bird species. The model's balanced simplicity and robustness make it ideal for deployment in small-scale environments, consisting of mobile gadgets, wherein green bird species identity is vital. Its overall performance across various situations highlights its capacity for real-time tracking in ecological settings.

While the outcomes are promising, Destiny Work could focus on similarly addressing elegance imbalances, improving the model's adaptability to varying recording qualities, and enhancing its real-time processing abilities [19-20]. Additionally, exploring superior sign processing strategies and record augmentation techniques ought to similarly refine the model's accuracy.

References

1. Mirugwe, A., Nyirenda, J., & Dufourq, E. (2022, July). Automating bird detection based on webcam captured images using deep learning. In *Proceedings of 43rd Conference of the South African Insti* (Vol. 85, pp. 62-76).
2. Revadekar, S. ., Panchal, V. ., Kanani, P. ., Shah, K. ., Vasoya, A. ., & Pandey, R. . (2023). Bird Sound Classification using Deep Neural Networks: A Comparative Analysis of State-of-the-Art Models and Custom Architectures. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(4), 614–622. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/3596
3. Noumida, A., & Rajan, R. (2022). Multi-label bird species classification from audio recordings using attention framework. *Applied Acoustics*, *197*, 108901.
4. Hong, L. (2023). Acoustic Bird Species Recognition at BirdCLEF 2023: Training Strategies for Convolutional Neural Network and Inference Acceleration using OpenVINO. In *CLEF (Working Notes)* (pp. 2043-2050).
5. Yang, F., Jiang, Y., & Xu, Y. (2022). Design of bird sound recognition model based on lightweight. *Ieee Access*, *10*, 85189-85198.
6. Sun, Y. P., Jiang, Y., Wang, Z., Zhang, Y., & Zhang, L. L. (2023). Wild Bird Species Identification Based on a Lightweight Model With Frequency Dynamic Convolution. *IEEE Access*, *11*, 54352-54362.
7. Huang, Y. P., & Basanta, H. (2021). Recognition of endemic bird species using deep learning models. *Ieee Access*, *9*, 102975-102984.

8. Lucio, D. R., Maldonado, Y., & da Costa, G. (2015, October). Bird species classification using spectrograms. In *2015 Latin American Computing Conference (CLEI)* (pp. 1-11). IEEE.
9. Liu, H., Liu, C., Zhao, T., & Liu, Y. (2021, November). Bird song classification based on improved Bi-LSTM-DenseNet network. In *2021 4th International Conference on Robotics, Control and Automation Engineering (RCAE)* (pp. 152-155). IEEE.
10. Ragib, K. M., Shithi, R. T., Haq, S. A., Hasan, M., Sakib, K. M., & Farah, T. (2020, July). Pakhichini: Automatic bird species identification using deep learning. In *2020 Fourth world conference on smart trends in systems, security and sustainability (WorldS4)* (pp. 1-6). IEEE.
11. Gupta, G., Kshirsagar, M., Zhong, M., Gholami, S., & Ferres, J. L. (2021). Comparing recurrent convolutional neural networks for large scale bird species classification. *Scientific reports*, *11*(1), 17085.
12. Mehyadin, A. E., Abdulazeez, A. M., Hasan, D. A., & Saeed, J. N. (2021). Birds sound classification based on machine learning algorithms. *Asian Journal of Research in Computer Science*, *9*(4), 1-11.
13. Koh, C. Y., Chang, J. Y., Tai, C. L., Huang, D. Y., Hsieh, H. H., & Liu, Y. W. (2019, September). Bird Sound Classification Using Convolutional Neural Networks. In *Clef (working notes)*.
14. Heinrich, R., Sick, B., & Scholz, C. (2024). AudioProtoPNet: An interpretable deep learning model for bird sound classification. *arXiv preprint arXiv:2404.10420*.
15. Incze, A., Jancsó, H. B., Szilágyi, Z., Farkas, A., & Sulyok, C. (2018, September). Bird sound recognition using a convolutional neural network. In *2018 IEEE 16th international symposium on intelligent systems and informatics (SISY)* (pp. 000295-000300). IEEE.
16. Madhavi, A., & Pamnani, R. (2018). Deep learning based audio classifier for bird species. *Int. J. Sci. Res*, *3*, 228-233.
17. Patil, D., Bodhe, R., Pawar, R., Doshi, T., & Vasekar, V. (2022). Visual and acoustic identification of bird species. *Int. Res. J. Engg. Technol*, *9*(5), 3504-3507.
18. Xie, J., & Zhu, M. (2019). Handcrafted features and late fusion with deep learning for bird sound classification. *Ecological Informatics*, *52*, 74-81.
19. Harh, A., Bandhu, S., Barai, B., Das, N., & Singh, P. K. (2024). An efficient deep convolutional neural network for automated bird sound classification. In Proceedings of 2nd International Conference on Data Analytics and Insights (ICDAI-2024).
20. Harh, A., Bandhu, S., Barai, B., & Singh, P. K. (2024). A hybrid deep learning framework for text-independent automatic speaker recognition system. In Proceedings of 3rd International Conference on Advanced Computing and Applications (ICACA-2024).