SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I10P101

Volume 12 Issue 10, 1-8, October 2025

Original Article

Dual Encoder Ensemble with EfficientNet and VGG for Bird Species Identification

Sireesha Abotula[1,3*], Srinivas Gorla[2], Prasad Reddy PVGD[4], Dasari Siva Krishna[5]

[1]*Department of Information and Technology, Andhra University, Visakhapatnam, India.*
[2]*Department of Computer Science & Engineering, Anil Neerukonda Institute of Technology & Sciences, Visakhapatnam, India.*
[3]*Department of AI and Data Science, Gandhi Institute of Technology and Management, Visakhapatnam, India.*
[4]*Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam, India.*
[5]*Department of Computer Science & Engineering, Gandhi Institute of Technology and Management, Visakhapatnam, India.*

[*]*Corresponding Author : sabotula85@gmail.com*

Received: 01 August 2025 Revised: 03 September 2025 Accepted: 02 October 2025 Published: 31 October 2025

Abstract *- The identification of bird species is crucial in various fields such as wildlife conservation, ecological research, and biodiversity monitoring. Identifying bird species from images is traditionally labor-intensive and prone to errors, particularly due to the global diversity of avian species. In this work, we propose an Image-Based Bird Species Identification (IMGBSI-NET) system using Deep Learning, which is both highly accurate and innovative. The system integrates two encoders, such as VGG-19 and EfficientNet-B0 pre-trained models. In this approach, a Dual Encoder-based architecture is employed, where various stages of the VGG-19 and EfficientNet models are used to extract features for prediction in bird species identification. In contrast to previous studies, this novel approach enhances the model performance of bird species recognition across a wide range of species. The most appropriate in this framework is EfficientNet, which extracts potent features that can capture fine-grained details that are required to distinguish between species of birds. The suggested method also proves to be compatible with other trained models and various mechanisms. The bird species recognition ability introduced in this study sets a new standard, and its performance is much higher in comparison to state-of-the-art models. The solution is a strong and effective one, and provides the perspectives of future research and practice in computer vision.*

Keywords *- EfficientNet, Encoder, Decoder, Bird Species, Features, Deep Learning.*

1. Introduction

The classification of bird species is a sophisticated problem in computer vision, particularly when dealing with fine-grained categorization. This area of research has attracted significant interest due to its potential applications in biodiversity monitoring, ecological studies, and conservation efforts. The task is complicated by the slight visual differences between species and the variability introduced by various factors such as lighting conditions, background clutter, and bird pose. These challenges have motivated researchers to develop various methods aimed at improving classification accuracy under real-world conditions.

Initial studies in bird classification were mainly based on classical methods of machine learning approaches, such as regression, Support Vector Machines (SVMs), and handcrafted feature extraction, followed by classifiers. Other studies include Histograms of Oriented Gradients (HOG) and RGB color descriptors. As illustrated by Alter et al. [1], the use of pre-trained Convolutional Neural Networks (CNNs) to extract features, when compared to the use of traditional classifiers, was much more effective, especially in situations where there is a change in the lighting and the presence of complex backgrounds. As deep learning came into view, the direction changed to ones that were more effective in fine-grained image classification. Atanbori et al. [2] demonstrated that combining motion and appearance features demonstrated a significant jump in classification accuracy when birds were in motion, rather than the classifier, but this approach was limited to static images only. Their work mentions the possibility of utilizing the temporal data of video collections in order to determine species.

In recent times, transfer learning has become popular in this area. Research conducted by Tayal et al. [3] and Cai et al. [4] indicated that using pre-trained deep learning models that extract task-specific features enhanced the accuracy of the classification. Kumar et al. [5] took this study a step further in a multistage training process with advanced networks like the Mask-RCNN and Inception networks to solve the issues of image resolution and different environments. Despite the fact that deep learning techniques are dominating the scene,

classical techniques like SVMs and decision trees are also used in bird classification. As an illustration, Fagerlund et al. [14] and Li Jian et al. [15] examined methods such as feature selection, image similarity measure, and color histogram application to improve classification accuracy such investigations highlight the fact that although classical methods tend to be less accurate compared to deep learning models, they still have some worth as far as computational efficiency is concerned.

The work is summarized as follows:

- In this proposed dual encoder ensemble framework, EfficientNet-B0 and VGG-19 are used for bird species identification.
- It integrates Squeeze-and-Excitation (SE) blocks to enhance channel-wise feature representation and improve model performance.
- Compare the effectiveness of EfficientNet-B0 and VGG-19 in different encoder placements to analyze their contributions to feature extraction.
- Finally, the hybrid approach using both encoders with GAP and FC layers outperforms individual models in bird species classification.

2. Related Work

The problem of bird classification is a fine-grained classification problem, a difficult but significant step to the overall problem of computer vision algorithm development for image recognition. Many of the studies have tried to deal with this issue and concentrate on making the classification more accurate in a real-life situation. As an illustration, Alter et al. [1] tried a multi-class SVM on HOG and RGB features together with softmax regression and transfer learning with a pre-trained CNN. Their findings indicated that pre-trained CNNs were more effective in feature extraction to deal with the problems of lighting variability, cluttered background, and subtle interspecies differences. An automatic model of birds was also explored by Atanbori et al. [2], who noted that to overcome the manual data collection process, which was labor- and error-driven, there was a necessity to develop an automated model of the problem. In order to address the further complexity of categorizing flying birds, where issues of worse image quality, fast pose change, and inter-species similarity occur, they generated a video dataset of sequences of thirteen bird classes. They used a combination of motion and appearance features, and with their approach, their classification accuracy was improved by 7% over single-image classifiers with a 90% correct classification rate.

Other researchers have investigated various ways and data to enhance the identification of bird species. Tayal et al. [3] employed MATLAB to extract image features and employed SVMs and transfer learning, and the data was obtained through Google Images. In the same way, Cai et al. [4] applied neural networks to bird species recognition and used neural networks that could learn dynamic characteristics of bird songs, which had an accuracy of 95% on the Caltech dataset, and this is with the help of a noise reduction algorithm. Kumar et al. [5] formulated a multistage training framework utilizing a transfer study alongside Mask-RCNN and an ensemble of Inception networks (InceptionV3 and InceptionResNetV2) on an Indian bird dataset, with an F1 score of 55.67% on a challenging set of solutions like different perspectives. A K-Nearest Neighbors (KNN) algorithm proposed by Hassanat et al. [6] uses a Binary Search Tree (BST) to classify big data effectively and showed significant progress in both speed and accuracy on both benchmark datasets.

The most current research still expands to the classification of bird species using machine learning and deep learning. Qiao et al. [7] used a hybrid of SVM and decision tree methods to increase the classification rates, whereas Huang et al. [8] used the decision tree to identify species on the Internet of Birds (IOB) dataset with the help of deep learning. Omkarini et al. [10] created a deep neural net-based method that has really promising results on a dataset of 253 species. Dhamodaran et al. [11] proposed the use of photo recognition software in bird watchers using CNNs to attain high accuracy. On the same note, Shyamkrishna et al. [12] adopted a CNN-based algorithm to boost accuracy, which resulted in a straightforward but valuable tool for observers. Pureti Anusha et al. [13] examined the learning methods in an unsupervised manner used to classify birds, and they showed an adequate performance with regard to diversity in the input images.

Traditional machine learning approaches remain relevant. Fagerlund et al. [14] employed statistical learning theory, structural risk minimization, and SVMs with kernel methods, achieving higher accuracy on Kaggle datasets. Li Jian et al. [15] developed a classification system based on color histograms and image similarity analysis, enabling first-level bird identification with detailed feature analysis. These feature-based methods, though less accurate than deep learning, remain valuable due to their computational efficiency and versatility when integrated with classical image processing. Abishal et al. [16] showed that the CUB-200-2011 dataset was classified at 80% accuracy with the help of the Random Forest models. Nevertheless, the previous models also had the disadvantage of working with different angles, and so researchers attempted to utilize more advanced deep learning approaches to enhance accuracy. They have also been utilized in a practical manner. In addition, Rodriguez-Juan et al. [19] introduced the Visual WetlandBirds dataset for species identification and behavior recognition using video-based ecological monitoring. Gavali and Banu [20] proposed a visual-acoustic fusion model, enhancing the accuracy of Indian bird species classification by combining visual and auditory cues.

3. Materials and Methods

The proposed dual encoder architecture (IMGBSI_NET) has encoders Encoder-1 and Encoder-2, which process the input data using different feature extraction backbones and attention mechanisms, as shown in Figure 1. The outputs of both encoders are aggregated using Global Average Pooling (GAP) and concatenated to pass through a Fully Connected (FC) layer to generate the final output. Using this mechanism allows the network to tune to the most informative channels for better performance and efficiency.

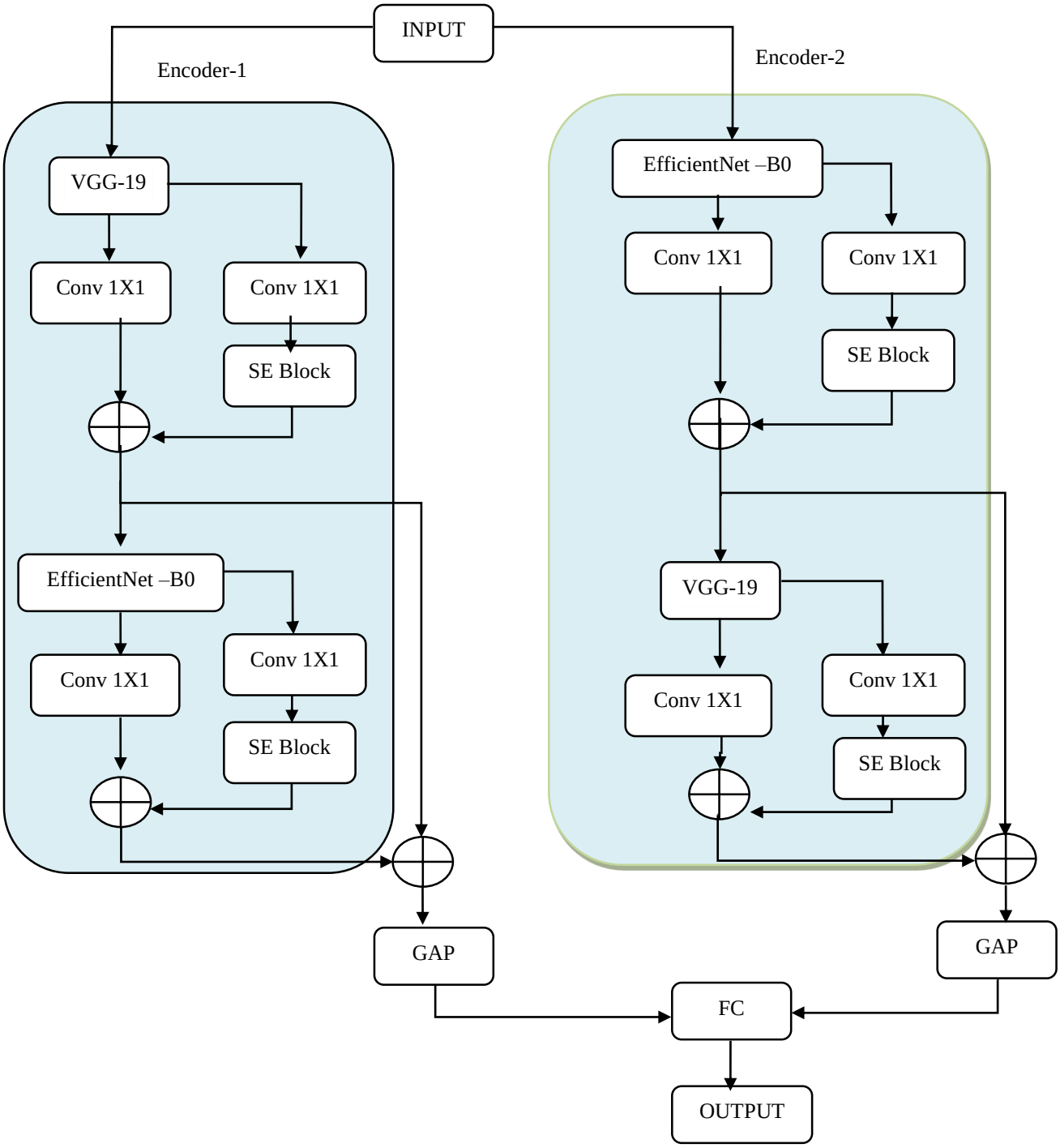

Fig. 1 Dual encoder architecture

Notations: GAP- Global Max polling, FC-Fully Connected Layers, SE Block-Squeeze-and-Excitation blocks

3.1. Encoder-1 Architecture

Firstly, IMGBSI-NET feeds the VGG-19 network with the input and extracts feature maps. VGG-19 is a well-known deep convolutional neural network used to extract hierarchical features. There are two subsequent 1x1 convolutional layers after VGG-19. They are used to reduce the dimensionality of feature maps and, hence, make the network more expressive. The 1 x 1 convolution then feeds the feature maps to a Squeeze and Excitation (SE) Block to adaptively recalibrate channel-wise responses of features by modeling interactions within channels. The input is fed into an efficient net B0 network in parallel. The network width, depth, and resolution are then optimized using a compound scaling method for efficient feature extraction. Two 1x1 convolution layers followed by an SE block are similarly processed over the EfficientNet output. At different depths, skip connections are used to fuse features from different depths and keep the spatial information.

3.2. Encoder-2 Architecture

Similar to Encoder-1, except that the backbones are flipped. First, the input is obtained through the EfficientNet-B0 model. For dimensionality reduction and in-channel recalibration, first perform the same 1x1 convolutions and SE block on the output of EfficientNet-B0. Parallel, it then processes the input through a VGG-19 network. VGG 19's output will also follow that of two 1x1 convolutions and an SE block pattern. Skip connections are used to merge feature maps between stages in Encoder-2 as well.

3.3. Squeeze-and-Excitation (SE) Block

The SE block is a form of channel attention mechanism [21] that improves the representational power of a network by recalibrating channel-wise feature responses.

- *Squeeze*: A Global Average Pooling (GAP) layer is applied to each channel of the feature map to generate a channel descriptor. This step squeezes spatial dimensions (H, W) into a single scalar value for each channel, thus creating a vector of size equal to the number of channels.
- *Excitation:* The squeezed vector is passed through two Fully Connected (FC) layers with a ReLU activation. The first FC layer reduces the dimensionality (using a reduction ratio), and the second FC layer restores it to the original size. A sigmoid activation function is applied to produce a set of channel-wise weights in the range [0, 1].
- *Recalibration:* The original feature map is multiplied (element-wise) by the recalibrated channel weights, allowing the network to emphasize or suppress certain channels based on their importance.

3.4. VGG-19

VGG-19 is a convolutional neural network architecture [22] that is widely used for image classification and feature extraction. VGG-19 consists of 16 convolutional layers, 3 fully connected layers, and a softmax for classification. In this architecture, only the convolutional layers are used. The network uses small (3x3) convolutional filters but stacks multiple layers to increase the depth and extract more complex features. After a few convolutional layers, max pooling is used to reduce spatial dimensions while retaining the most important information. Rectified Linear Units (ReLU) are applied after each convolution operation to introduce non-linearity. VGG-19 is known for its simplicity, ease of implementation, and strong performance in extracting hierarchical features.

3.5. EfficientNet-B0

At present, a family of convolutional neural networks, EfficientNet [23], is optimized for efficient training and inference. Compound scaling is the method used to scale depth, width, and resolution, keeping them balanced. This leads to a high-efficiency model with better accuracy via fewer parameters. EfficientNet B0 is based on MBConv blocks or light and efficient versions of the inverted residuals employed in MobileNetV2. In this design, each MBConv block is followed by an SE block, which boosts channel-wise feature representation. Depthwise separable convolutions enable efficiency in decreasing parameters and computation costs. EfficientNet B0 is the smallest model in the EfficientNet family and is optimized for reasonable computational resources.

4. Results and Discussion

The architecture consists of two parallel encoders (Encoder-1 and Encoder-2), each utilizing different feature extraction backbones (VGG-19 and EfficientNet-B0) with 1x1 convolutions and Squeeze-and-Excitation (SE) blocks. The outputs from both encoders are aggregated using Global Average Pooling (GAP), followed by a Fully Connected (FC) layer for the final output.

4.1. Encoder -1 Analysis

VGG-19 expects an input of shape (3, 224, 224). The network performs a series of convolutions (with 3x3 filters), followed by max-pooling. For simplicity, let us summarize VGG-19's output: After the final pooling layer of VGG-19, let the feature map size be reduced to (512, 7, 7). The feature map (512, 7, 7) is passed through two successive. These convolutions primarily adjust the channel dimensions while preserving the spatial dimensions. Suppose these reduce the channels to 256, resulting in a feature map of shape (256, 7, 7).

The SE block recalibrates the channel-wise feature responses, and Global Average Pooling (GAP) reduces each channel to a scalar, creating a (256,) vector. Fully connected layers adjust and rescale these values, which are multiplied back to the original feature map. The output shape remains (256, 7, 7).

The input image is also passed through an EfficientNet-B0 backbone.EfficientNet-B0 uses a combination of MBConv blocks and SE blocks. The final output of EfficientNet-B0 has

a feature map of (1280, 7, 7). Additional 1x1 Convolution Layers and SE Block. The output (1280, 7, 7) undergoes two 1x1 convolutions, reducing the channel size to 512. After SE recalibration, the output shape is (512, 7, 7).

The outputs from VGG-19 and EfficientNet paths are concatenated, and the merged output shape is (768, 7, 7). Global Average Pooling (GAP) reduces this to (768,).

4.2. Encoder-2 Analysis

EfficientNet-B0 Backbone Input goes first through EfficientNet-B0, resulting in (1280, 7, 7). 1x1 Convolution Layers and SE Block Feature map (1280, 7, 7) is processed by two 1x1 convolutions and an SE block, resulting in (512, 7, 7).VGG-19 Backbone. In parallel, the input is passed through VGG-19: Final feature map after VGG-19 processing is (512, 7, 7). The VGG-19 output undergoes two additional 1x1 convolutions and an SE block, resulting in (256, 7, 7) aggregation (Encoder-2 Output).

The outputs from the EfficientNet and VGG-19 paths are merged, with the merged feature map size being (768, 7, 7). Global Average Pooling (GAP) reduces this to (768,). The outputs from Encoder-1 ((768,)) and Encoder-2 ((768,)) are concatenated, resulting in a combined feature vector of (1536,). This vector is passed through a Fully Connected (FC) layer. The FC layer generates the final output, which classifies the bird species.

4.3. DataSet Description

The Data Set consists of 510 species, and each species contains 50 different kinds of images that capture the same species. The dataset is divided into training, testing, and validation. The validation set consists of 2550 images, each of which consists of 5 images. Similarly, the testing set consists of 2550 images, each of which consists of 5 images. The Data Set consists of a total of 25,500 training images, 2550 validating images, and 2550 testing images, and the sample images shown in Figure 2.

The data set is collected from the kaggle.com repository, which contains only bird species images, and the non-bird images set is prepared from search engine results. The non-bird species training set, validation, and testing consist of 25550, 2550, and 2550, respectively. The complete dataset consists of 51,100, 5,100, and 5,100, totaling 61,300. The splitting ratio of the dataset is 80:10:10, as shown in Table 1.

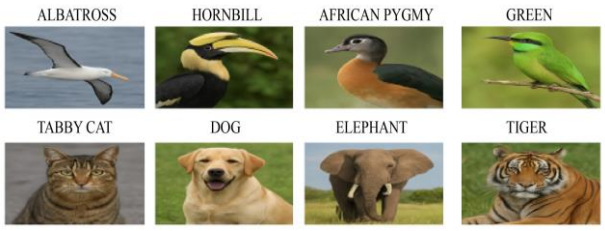

Fig. 2 Sample dataset

Data Set Availability:
https://www.kaggle.com/code/gpiosenka/explore-510-bird-species-dataset

Table 1. Dataset description

Species Type	Training	Testing	Validation
Bird Species	25,550	2550	2550
Non-Bird Species	25,550	2550	2550
Total	51,100	5100	5100

4.4. System Specifications

To effectively handle the comprehensive dataset consisting of 61,300 images (with 51,100 training images, 5,100 validation images, and 5,100 testing images) for optimal performance, which includes a powerful CPU such as an Intel Core i9-13900K with 12 or more cores for efficient data preprocessing. A high-performance GPU like the NVIDIA RTX 4090 with 24GB of VRAM is essential to manage large batch sizes and complex neural network models. The system has 128GB DDR5 RAM to handle the significant memory. For storage, it consists of a 2TB NVMe SSD for the operating system and software, along with a 4TB SSD for data storage. To sustain prolonged training sessions without overheating, a liquid cooling system is advisable, supported by a 1000W power supply unit.

4.5. Experimental Results

In testing the data set with various deep learning models such as VGG-16, VGG-19, ResNet-18, ResNet-34, and ResNet-50, Efficient B0, B1, B2, B3, B3, B4, B5, and B6. Tables 2-4 show the experimental results of the deep learning models. In these models, each model exhibited a maximum of 50 epochs, and the best trained model with validated trained parameters is stored with the explicit path. The model was tested with the best trainable parameter and the sample code available in the link below
https://www.kaggle.com/code/siva007dasari/imgbsi-net/edit.

4.5.1. Performance Metrics

The performance of a model can be evaluated using various metrics such as Accuracy (Acc), Precision (Pre), Recall, and the F1-Score.

4.5.2. Deep Learning Model Performances

The utility of VGG-19 over VGG-16 on all 3 metrics shows the highest accuracy (92.40%), recall (93.48%), and f1-score (94.26%). The performance of VGG 19 is well balanced, with high precision and recall resulting in an excellent f1-score, thus indicating good balance between true positives and true negatives. VGG-16 slightly outperforms VGG-19, but the recall (88.60%) is the same as the test set, which suggests that the model would miss some positive predictions. This has resulted in its additional layers assisting the deeper network (VGG-19) in extracting better features and generalizing them. The problem comes with simple but strong performing VGG models, which are unseated by the more modern architectures like EfficientNet.

Table 2. VGGNet performance

Model	Acc	Pre	Recall	F1
VGG-16	90.35	90.22	88.60	89.20
VGG-19	92.40	91.48	93.48	94.26

ResNet-50 offers the best balance between precision and recall among ResNet models, resulting in a competitive F1 score (84.86%). ResNet-18 shows similar performance to ResNet-50 in terms of accuracy and precision, but falls slightly behind in recall and F1-score. ResNet-34 has the lowest performance across all metrics, indicating that increasing depth from ResNet-18 to ResNet-34 does not necessarily guarantee better results in this case. While deeper models like ResNet-50 generally perform better due to their ability to learn more complex patterns, they may also face challenges such as overfitting or vanishing gradients if not properly optimized. The use of skip connections in ResNet models helps alleviate issues like vanishing gradients, but performance gains diminish beyond a certain depth (as seen in ResNet-34).

Table 3. ResNet performance

Model	Acc	Pre	Recall	F1-Score
ResNet-18	86.48	88.56	84.72	88.20
Resnet-34	82.40	84.52	84.52	83.40
Resnet-50	86.54	88.24	86.26	84.86

EfficientNet-B0 achieves the highest accuracy (92.48%) among all EfficientNet variants, suggesting it can be a strong starting model with efficient parameter usage. EfficientNet-B1 and EfficientNet-B2 show competitive performance, particularly excelling in recall, with B2 achieving the highest recall (92.52%) among all models in this category. As the models scale up to EfficientNet-B6, the performance does not always increase linearly. EfficientNet-B5 and B6 show a dip in f1-score, indicating potential overfitting in training deeper models with limited data. EfficientNet models leverage compound scaling, adjusting width, depth, and resolution simultaneously, which generally improves performance and efficiency. However, increasing the model size (from B0 to B6) does not always guarantee better results, as seen in the decline of the F1-score for B5 and B6, possibly due to overfitting or data inefficiencies.

Table 4. EfficientNet performance

Model	Acc	Pre	Recall	F1-score
EfficientNet-B0	92.48	88.56	84.72	88.20
EfficientNet-B1	90.46	90.48	91.40	90.22
EfficientNet-B2	90.40	90.32	92.52	90.40
EfficientNet-B3	84.48	86.36	88.26	86.36
EfficientNet-B4	88.40	86.56	84.24	83.40
EfficientNet-B5	88.56	84.72	78.20	80.20
EfficientNet-B6	90.24	88.50	84.42	88.22

4.5.3. Comparison Study

The proposed model (IMGBSI-NET) achieves the highest accuracy at 96.24%, as shown in Table 5, EfficientNet-B0 (92.48%), and VGG-19 (92.40%).ResNet-50 shows a significantly lower accuracy of 86.54%, indicating that it is outperformed by both the classical VGG and the EfficientNet architecture. IMGBSI-NET leads with a Precision of 94.24%, which suggests it is highly effective in minimizing false positives.

VGG-19 follows closely with a precision of 91.48%, while EfficientNet-B0 and ResNet-50 lag slightly behind at 88.56% and 88.24%, respectively. The Proposed Model excels with a Recall of 96.28%, meaning it is quite good at finding true positives and minimizing false negatives, and VGG-19 achieves a strong 93.48% recall; however, EfficientNetB0 falls back to 84.72%. The most balanced model in terms of both precision and recall: It achieves an F1-score of 94.80% with IMGBSI-NET. Notably, the F1-score of 94.26% is also very close to the VGG-19; however, ResNet shows a lower F1-score of 84.86, suggesting a trade-off between precision and recall score.

Table 5. Comparative study

Model	Acc	Pre	Recall	F1-Score
EfficientNet-B0	92.48	88.56	84.72	88.20
VGG-19	92.40	91.48	93.48	94.26
Resnet-50	86.54	88.24	86.26	84.86
Sachin Kumar et al. [17]	89.50	88.40	85.60	88.90
Jashanpreet Kaur [18]	95.80	95.60	95.40	95.50
Proposed Model (IMGBSI-NET)	96.24	94.24	96.28	94.80

Overall, the architecture present in this work outperforms all others in every metric it tested, and across all common datasets. The fact that the new model is able to accurately identify true positives with very few false predictions suggests that there is not only great potential to further increase its accuracy, but also incredible potential to decrease the cost (in terms of the number of false predictions required) for which a model can distinguish between these two classes of datapoints. Although VGG-19 is a dependable performer - particularly in precision and recall - it is still an outperformer, which is appropriate for tasks needing a balanced performance. It, however, does fall behind IMGBSI NET. Compared to VGG-19 and the proposed model, both ResNet-50 and EfficientNet-B0 are more efficient models but are outperformed by the VGG-19 and proposed model, especially in recall and F1-Score, which is critical for applications where false negatives incur a high cost.

5. Conclusion

In this work, we introduce IMGBSI-NET, a new dual encoder-based deep learning architecture to identify bird

species. The method takes the capabilities of two models, VGG-19 and EfficientNet-B0, to extract the features at different levels of the network, and therefore, enhances the classification accuracy for bird species identification. The dual encoder architecture performs better than the single encoder, which extracts features to discriminate against closely related bird species.

The experimental findings give substantial evidence that IMGBSI-NET is better than the traditional models, and the accuracy, precision, recall, and f1-score of IMGBSI-NET are significantly higher with a variety of bird species datasets. This proves that the proposed method can overcome the difficulties of the enormous diversity and the minute variations of species of birds. The effectiveness of IMGBSI-NET shows its possible use in the field of wildlife protection, ecological observation, and biodiversity studies, as it can provide a scalable and automated solution to identify bird species. To apply this architecture to larger datasets, to real-time bird recognition in videos, and to fit the model to deploy it on edge computing systems, we will work on it in the future. The research is a new standard of the bird identification task, and its potential future is bright with the help of computer vision in the hands of ornithologists and other fields.

References

[1] Anne L. Alter, and Karen M. Wang, "An Exploration of Computer Vision Techniques for Bird Species Classification," *stanford.edu*, pp. 1-7, 2017. [Google Scholar] [Publisher Link]

[2] John Atanbori et al., "Classification of Bird Species from Video Using Appearance and Motion Features," *Ecological Informatics*, vol. 48, pp. 12-23, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[3] Madhuri Tayal, "Bird Identification by Image Recognition," *Helix*, vol. 8, no. 6, pp. 4349-4352, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[4] Jinhai Cai et al., "Sensor Network for the Monitoring of Ecosystem: Bird Species Recognition," *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, Melbourne, VIC, Australia, pp. 293-298, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[5] Akash Kumar, and Sourya Dipta Das, "Bird Species Classification using Transfer Learning with Multistage Training," *Workshop on Computer Vision Applications*, Hyderabad, India, pp. 28-38, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Ahmad B.A. Hassanat, "Furthest-Pair-Based Binary Search Tree for Speeding Big Data Classification Using K-Nearest Neighbors," *Big Data*, vol. 6, no. 3, pp. 171-235, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[7] Baowen Qiao et al., "Bird Species Recognition Based on SVM Classifier and Decision Tree," *2017 First International Conference on Electronics Instrumentation & Information Systems (EIIS)*, Harbin, China, pp. 1-4, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[8] Yo-Ping Huang, and Haobijam Basanta, "Bird Image Retrieval and Recognition Using a Deep Learning Platform," *IEEE Access*, vol. 7, pp. 66980-66989, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[9] Satyam Raj et al., "Image-Based Bird Species Identification using Convolutional Neural Network," *International Journal of Engineering Research & Technology*, vol. 9, no. 6, pp. 1-6, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Vemula Omkarini, and G. Krishna Mohan, "Automated Bird Species Identification Using Neural Networks," *Annals of the Romanian Society for Cell Biology*, vol. 25, no. 6, pp. 5402-5407, 2021. [Google Scholar] [Publisher Link]

[11] S. Dhamodaran et al., "Bird Species Identification from Image," *Mathematical Statistician and Engineering Application*, vol. 71, pp. 495-501, 2022. [Google Scholar]

[12] Aleena Varghese, K. Shyamkrishna, and M. Rajeswari, "Utilization of Deep Learning Technology in Recognizing Bird Species," *AIP Conference Proceedings*, vol. 2463, no. 1, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Pureti Anusha, and Kundurthi ManiSai, "Bird Species Classification Using Deep Learning," *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP)*, Hyderabad, India, pp. 1-5, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Seppo Fagerlund, "Bird Species Recognition Using Support Vector Machines," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1-8, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[15] Li Jian, Zhang Lei, and Yan Baoping, "Research and Application of Bird Species Identification Algorithm Based on Image Features," *2014 International Symposium on Computer, Consumer and Control*, Taichung, Taiwan, pp. 139-142, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[16] B. Persia Abishal, and Sujitha Juliet, "Image-Based Bird Species Identification Using Machine Learning," *2023 9th International Conference on Advanced Computing and Communication Systems FIGICACCS)*, Coimbatore, India, pp. 1963-1968, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] Sachin Kumar et al., "Enhancing Bird Species Identification Using a Custom CNN Architecture," *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Tashkent, Uzbekistan, pp. 1333-1338, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Jashanpreet Kaur et al., "Enhanced Bird Species Identification using ResNet-50: A Deep Learning Framework for High-Performance Classification," *2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, Pudukkottai, India, pp. 821-826, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Javier Rodriguez-Juan et al., "Visual Wetlandbirds Dataset: Bird Species Identification and Behavior Recognition in Videos," *Scientific Data*, vol. 12, pp. 1-13, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[20] Pralhad Gavali, and J. Saira Banu, "A Novel Approach to Indian Bird Species Identification: Employing Visual-Acoustic Fusion Techniques for Improved Classification Accuracy," *Frontiers in Artificial Intelligence*, vol. 8, pp. 1-15, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[21] Jie Hu, Li Shen, and Gang Sun, "Squeeze-and-Excitation Networks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 7132-7141, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[22] Karen Simonyan, and Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv Preprint*, pp. 1-14, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[23] Mingxing Tan, and Quoc Le, "Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks," *International Conference on Machine Learning*, pp. 6105-6114, 2019. [Google Scholar] [Publisher Link]