Bird Species Classification Using CNN Models

A Somnath  
 School of Computer Science and

Engineering  
VIT-AP University

Amaravati, Andhra Pradesh

somnath.21bce9501@vitapstudent.ac.in  
V Dileep

School of Computer Science and

Engineering

VIT-AP University

Amaravati, Andhra Pradesh

dileep.21bce9650@vitapstudent.ac.in V Vishnu Prajwal

School of Computer Science and

Engineering

VIT-AP University

Amaravati, Andhra Pradesh

vishnu.21bce8538@vitapstudent.ac.in

Abstract- In recent times, bird-watching has evolved into a popular hobby among humans. People are captivated by the diverse characteristics of birds, including their colors, sizes, and the angles from which they are observed. However, due to these variations, it often becomes challenging for individuals to accurately identify and categorize different bird species. To address this issue, we present a novel approach to classifying bird species based on images. Our method involves the extraction of distinct features from bird images. Through a combination of preprocessing techniques and data augmentation, we enhance the quality of the input data. This improved dataset is then utilized to train a convolutional neural network (CNN), a type of deep learning model. The CNN's architecture is designed to process and understand the intricate patterns present in the images.

The outcomes of our approach are promising. When applied to a publicly available dataset of bird images, our CNN architecture achieves an impressive mean average precision score of 0.95. This score indicates the model's ability to accurately predict the primary species depicted in each image. Moreover, the model demonstrates a noteworthy accuracy range of 0.75 to 0.95 when classifying various bird species within the dataset. This method holds great potential for enhancing the experience of bird enthusiasts and researchers alike.

Keywords: Bird species, Machine Learning, Convolutional Neural Networks.

# **Introduction**

* Birds play a crucial role in maintaining ecological balance within our environment. The study of birds not only enhances our understanding of the natural world but also provides valuable insights into broader aspects of nature. Ornithologists commonly face the challenge of accurately identifying bird species, a task that holds immense significance. Environmental scientists leverage the sensitivity of birds to environmental changes, using them as indicators to comprehend ecosystems. Birds find applications in diverse real-world scenarios, such as the monitoring of environmental pollution. The presence and distribution of different bird species within ecosystems bear vital ecological implications. This is an arena where our classification program can prove invaluable.
* Our classification method holds the potential to contribute significantly to the field. It offers a means to track and categorize bird species from images, facilitating the understanding of their presence in different environments. Addressing the complexity of this task, we recognize the intricate variations that exist, including the distinct shapes, appearances, backgrounds, lighting conditions, and poses of birds. The central objective is to accurately identify and classify bird species.
* In the contemporary landscape of machine learning and deep learning, image classification has emerged as a prominent research area. Our work focuses on the specific challenge of identifying bird species from images, a task that poses considerable difficulties due to the wide variations even within bird subspecies. These challenges encompass issues such as shape, appearance, background, lighting, and pose discrepancies. In our pursuit to accurately classify birds and pinpoint their species, we have developed a modified CNN network tailored to our specific needs. Multiple CNN architectures were employed as models, each contributing to the process of feature extraction. By addressing these complexities and leveraging cutting-edge CNN technology, our approach strives to overcome the challenges inherent in bird species identification from images. This method holds potential not only for enhancing ornithological research but also for enabling practical applications such as ecosystem monitoring and conservation efforts.

**A. Motivation –**

Bird species classification projects are driven by a desire to better understand and conserve avian biodiversity. By leveraging advancements in deep learning and image recognition, researchers aim to create efficient tools for identifying bird species, promoting citizen science involvement, and supporting broader conservation and ecological studies. The project's ultimate goal is to contribute to the preservation of birds and their habitats while harnessing the power of technology for positive environmental impact.

**B. Contribution**

- Implement the CNN by varying the kernel size, number of kernels and optimizers.

- Implement the customized CNN to recognize the of Bird species.

- Analyzing the results of CNN model and updating the code.

# Literature Review

Specifically, ecologists utilize monitoring techniques to comprehend the reasons behind fluctuations in bird populations and to aid in the conservation and management of species that are under threat or endangered. The process of identifying birds predominantly involves visual or auditory methods. When considering visual aspects, key factors include the bird's physical shape, wings, size, posture, and coloration. However, it's crucial to account for the time of year, as a bird's wing characteristics change with growth. Acoustic components involve the recognition of bird songs and calls. Additionally, distinct markings like breast spots, wing bars (thin lines on wings), eye rings, crowns, and eyebrows are valuable for differentiation. The shape of the beak also serves as a unique identifier.

Among these characteristics, bird shape and posture are commonly relied upon for identification, especially since these attributes remain relatively stable. Silhouette recognition is a skill often mastered by experts due to its resistance to change. The tail of a bird provides further differentiation, with features such as notches, length, pointedness, or roundness being distinguishable. Even leg characteristics like length can aid in identification. However, relying solely on one parameter for identification is inadequate. Accurate results necessitate the consideration of multiple parameters. The apparent size of a bird in an image is contingent on variables such as resolution, distance from the bird, the capturing device used, and the focal length of the lens. As a result of practical observations involving numerous images, color differentiation determined by analyzing individual pixels has proven to be a dependable means of classification. It has been noted that higher image quality correlates with increased accuracy.

Convolutional Neural Networks (CNNs) are employed to determine the species to which a specific bird image belongs. John.Martinsson et al. presented the convolution neural network algorithm and deep residual neural networks to detect images based on effective feature extraction and signal classification. This approach used an experimental analysis for different images. But their approaches did not consider the images of other species found in the background. To effectively identify the background species huge volumes of training data are required.

Juha Niemi, Juha T Tanttu et al, proposed effective automatic bird image identification using Convolution Neural Network CNN along with deep Learning and machine learning classification algorithms. They also proposed a data augmentation method in which images are converted rotated, zoom and flipping in accordance with the desired color. Finally this approach takes a birds image as an input and gives the effective identity classification result of the bird as an output.

Marcelo T. Lopes, Lucas L. Gioppo et al (2011), focused mainly on the automatic identification of bird species from their audio recorded songs. Here the authors dealt with the bird species identification problem using signal processing and machine learning techniques with the MARSYAS feature set. Presented a series of experiments conducted in a database composed of bird songs from 75 species out of which problem obtained in performance with 12 species.

# Bird species classification using deep learning

* Preprocessing

preprocessing in CNN is critical for achieving better model performance, reducing overfitting, enhancing generalization, and facilitating the use of pretrained weights. Proper preprocessing contributes to a well-conditioned training process, enabling the model to learn and recognize patterns effectively.

Image Rescaling: Each pixel value is divided by 255 to bring the pixel values in the range [0, 1].

Data Augmentation: During training, the ImageDataGenerator applies various transformations:

Shear Range: Random shear transformation with a range of 0.2.

Zoom Range: Random zoom with a range of 0.2.

Horizontal Flip: Randomly flips images horizontally.

* **Proposed methods architecture**.

VGG19- The architecture used is the VGG19 model, which is a Convolutional Neural Network (CNN) with a total of 19 layers. The architecture can be summarized as follows:

The input to the network is an RGB image with dimensions 224x224 pixels (height and width) and 3 channels (Red, Green, and Blue). This is the first convolutional layer with 64 filters. Each filter has a size of 3x3 pixels and uses the Rectified Linear Unit (ReLU) activation function. The output of this layer will be 64 feature maps. This is the second convolutional layer with 64 filters of size 3x3 and ReLU activation. The output will still have 64 feature maps. This is the first MaxPooling layer that performs downsampling by taking the maximum value within a 2x2 window. It reduces the spatial dimensions of the feature maps by half, resulting in 112x112 feature maps. The third convolutional layer with 128 filters, 3x3 size, and ReLU activation. It increases the number of feature maps to 128. The fourth convolutional layer with 128 filters, 3x3 size, and ReLU activation. The output still has 128 feature maps. The second MaxPooling layer, reducing the spatial dimensions to 56x56 feature maps. The fifth convolutional layer with 256 filters, 3x3 size, and ReLU activation, resulting in 56 feature maps. The sixth convolutional layer with 256 filters, 3x3 size, and ReLU activation, maintaining 56 feature maps. The seventh convolutional layer with 256 filters, 3x3 size, and ReLU activation, keeping the output at 56 feature maps. The eighth convolutional layer with 256 filters, 3x3 size, and ReLU activation, maintaining 56 feature maps. The third MaxPooling layer, reducing the spatial dimensions to 28x28 feature maps. The ninth convolutional layer with 512 filters, 3x3 size, and ReLU activation, resulting in 28 feature maps. The tenth convolutional layer with 512 filters, 3x3 size, and ReLU activation, maintaining 28 feature maps. The eleventh convolutional layer with 512 filters, 3x3 size, and ReLU activation, keeping the output at the 28 feature maps. ReLU): The twelth convolutional layer with 512 filters, 3x3 size, and ReLU activation, maintaining 28 feature maps. The fourth MaxPooling layer, reducing the spatial dimensions to 14x14 feature maps. The thirteenth convolutional layer with 512 filters, 3x3 size, and ReLU activation, resulting in 14 feature maps. The fourteenth convolutional layer with 512 filters, 3x3 size, and ReLU activation, maintaining 14 feature maps. The fifteenth convolutional layer with 512 filters, 3x3 size, and ReLU activation, keeping the output at 14 feature maps. The sixteenth convolutional layer with 512 filters, 3x3 size, and ReLU activation, maintaining 14 feature maps. The fifth MaxPooling layer, reducing the spatial dimensions to 7x7 feature maps.

This layer flattens the 7x7 feature maps into a 1D vector of 7 \* 7 \* 512 = 25,088 elements. This is the first fully connected (dense) layer with 4096 units and ReLU activation. This is the second fully connected (dense) layer with 4096 units and ReLU activation. This is the output layer with 10 units (assuming it's a classification task with 10 classes) and uses the Softmax activation function to produce class probabilities.

And VGG16 with 13 convolutional layers and 3 fully connected layers.

**LeNet**- The architecture of the LeNet-5 model, which is a classic convolutional neural network designed for image classification.

The input to the network is a grayscale image with dimensions 32x32 pixels (height and width) and a single channel. This is the first convolutional layer with 6 filters of size 5x5 pixels. It uses the Rectified Linear Unit (ReLU) activation function. Since it is the first layer, the 'input\_shape' parameter is specified to define the input dimensions. This is the first Average Pooling layer that performs down sampling by taking the average value within a 2x2 window. It reduces the spatial dimensions of the feature maps by half, resulting in 16x16 feature maps with 6 channels. This is the second convolutional layer with 16 filters of size 5x5 pixels. It uses ReLU activation. The output will have 16 feature maps. This is the second AveragePooling layer that further down samples the feature maps. It reduces the spatial dimensions to 8x8 feature maps with 16 channels. This layer flattens the 8x8x16 feature maps into a 1D vector of 1024 elements. This is the first fully connected (dense) layer with 120 units and ReLU activation. It takes the flattened vector as input. This is the second fully connected (dense) layer with 84 units and ReLU activation. This is the output layer with 10 units (assuming it's a classification task with 10 classes) and uses the Softmax activation function to produce class probabilities.

* **Training Process**

The cnn architecture is designed to extract hierarchical features from images. The convolutional layers capture low-level features, such as edges and textures, and the fully connected layers learn high-level semantic information. By using multiple convolutional layers before pooling, cnn achieves a deeper representation, allowing it to learn complex patterns and features. The objective of using cnn is to classify bird species from input images. The model is trained to recognize patterns and features that are specific to different bird species, and the final softmax layer provides probability scores for each class.

A) Feature Extraction

**Convolution Layer:**

The output feature map of the convolution layer is computed using the convolution operation:

Where:

is the output value at position (i, j) in the k-th feature map. is the input value at position in the c-th input channel. is the weight of the convolutional filter at position (m, n) for the c-th input channel and the k-th output channel. B(k) is the bias term for the k-th output channel. S is the stride of the convolution operation.

**Max Pooling Layer:**

The output value of a max-pooling layer is the maximum value in a local region of the input:

Where:

is the output value at position in the k-th feature map. is the input value at position in the k-th input channel. S is the stride of the pooling operation

**Dense Layer:**

The output value of a dense layer is computed using the dot product of the input and weight matrix, followed by the application of the activation function:

**Softmax Function:**

The softmax function is applied in the last dense layer to obtain probability scores for each class:

Where:

is the probability score of class k.

is the output value of the k-th neuron in the last dense layer. C is the total number of classes.

**Cross-Entropy Loss:**

The cross-entropy loss is used as the loss function for multi-class classification tasks. Given the true label one-hot encoded as Y\_true and the predicted probabilities as Y\_pred, the cross-entropy loss is computed as below:

Where:

is the cross-entropy loss. is the true probability of class c (1 if the true class, 0 otherwise). is the predicted probability of class c.

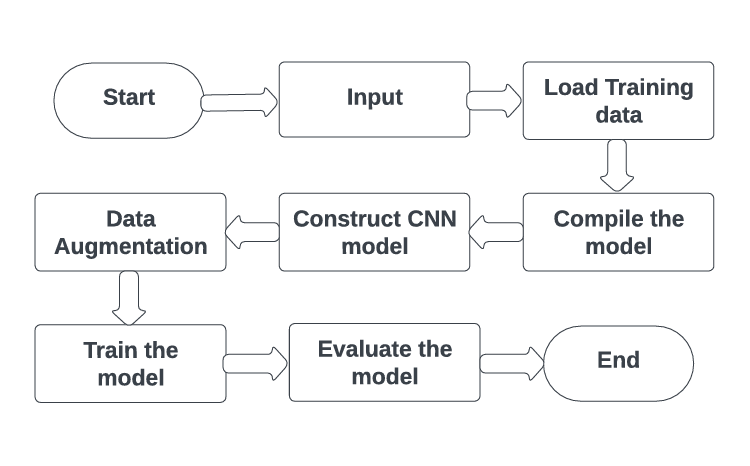


Fig- Workflow structure of the model

• Testing Process

Data Preprocessing: The images are rescaled and augmented using the ImageDataGenerator during training to increase the diversity of the training data and improve generalization. Model Compilation in the VGG19 model is compiled with the Adam optimizer, categorical cross-entropy loss function, and accuracy as the evaluation metric. Model Training: The model is trained on the training dataset using the fit method, with specified number of epochs and steps per epoch.

Data Preprocessing: The test images are rescaled using the ImageDataGenerator, but no augmentation is applied to ensure the model's performance is evaluated on the original test set.

Model Evaluation: The model's performance is evaluated on the test dataset using the evaluate method, providing the test accuracy.

# Dataset Description

* Dataset is of 10 bird species with 1601 training images, 50 test images (5 images per species) and 50 validation images (5 images per species. This is a very high quality dataset where there is only one bird species in each class and the bird typically takes up at least 50% of the pixels in the image.

All images are with the size of color images in jpg format. Data set includes a train set, test set and validation set. Each set contains 10 sub directories, one for each bird species. The data structure is convenient if you use the Keras ImageDataGenerator.flow\_from\_directory to create the train, test and valid data generators.

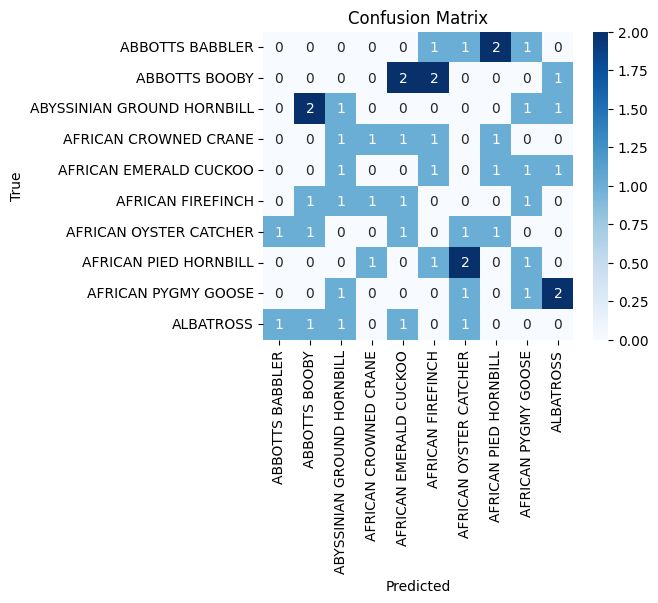
* **Dataset Description-**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Bird Species Classification | Train images | Test images | Valid images | Total Images |
| ABBOTTS BABBLER | 163 | 5 | 5 | 173 |
| ABBOTTS BOOBY | 180 | 5 | 5 | 190 |
| ABYSSINIAN GROUND HORNBILL | 179 | 5 | 5 | 189 |
| AFRICAN CROWNED CRANE | 135 | 5 | 5 | 145 |
| AFRICAN EMERALD CUCKOO | 154 | 5 | 5 | 164 |
| AFRICAN FIREFINCH | 137 | 5 | 5 | 147 |
| AFRICAN OYSTER CATCHER | 155 | 5 | 5 | 165 |
| AFRICAN PIED HORNBILL | 187 | 5 | 5 | 197 |
| AFRICAN PYGMY GOOSE | 179 | 5 | 5 | 189 |
| ALBATROSS | 132 | 5 | 5 | 142 |
| Total images | 1601 | 50 | 50 | 1701 |

# Results and Discussion

* This experiment was prepared on windows operating system in Google colab and Jupyter by using libraries like pandas, NumPy, TensorFlow, Keras, sklearn, for plotting visualizations mathplotlib and importing the dataset we uploaded the data in google drive. After training labeled dataset is ready for classifiers for image processing. The evaluation of the proposed approach for bird species classification by considering color features and parameters such as size, shape, etc. of the birds. The compilation of the model with the batch\_size of 32.

**VGG19-**



Confusion Matrix

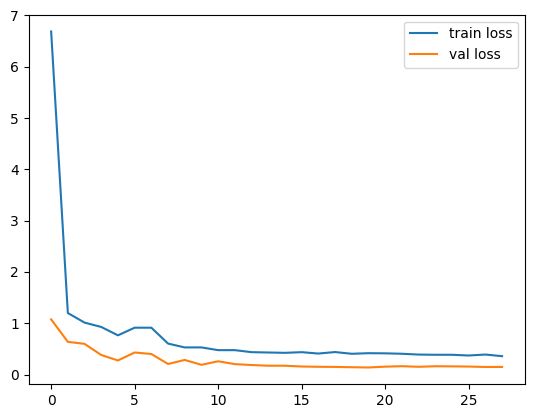
* For example, consider below Figure as input image given to the system for classification of bird which belongs to AFRICAN FIREFINCH.



AFRICAN FIREFINCH

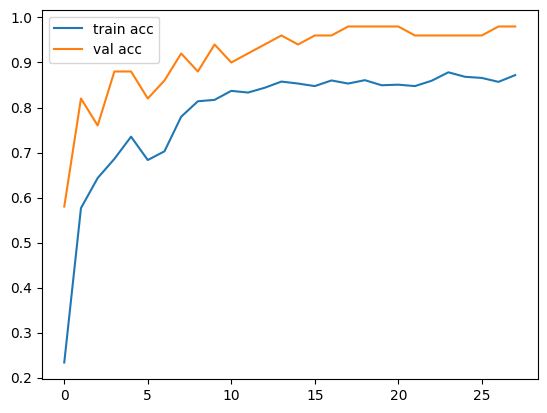
By the confusion matrix tells the true label vs predicted label of each bird image. Each image trained and tested with the size of 224,224.

shows the loss of the proposed CNN by the epochs in the range from 0 to 35. It is observed that the loss is decreased with each epoch.



Loss vs epoch

shows the training and testing accuracies of the proposed CNN by the epochs without duplicate data samples.

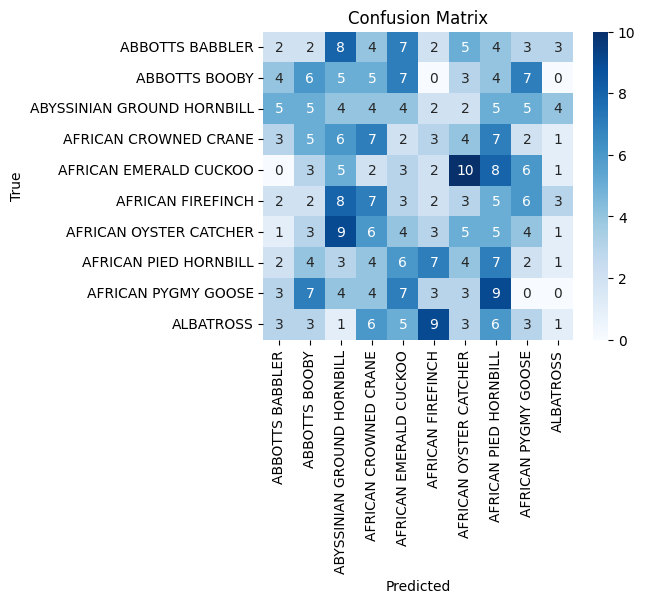


Accuracy vs epoch

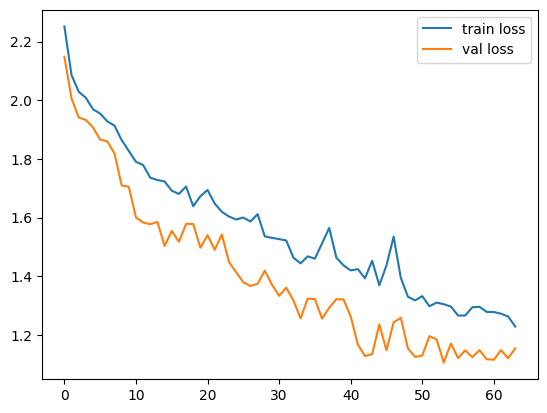
The results proved that the accuracy improves with each

epoch. The training and test accuracy curves increasing slightly for the given epochs, and consequently which is starting from the starting. Hence, algorithm can be applied with this better value epoch.

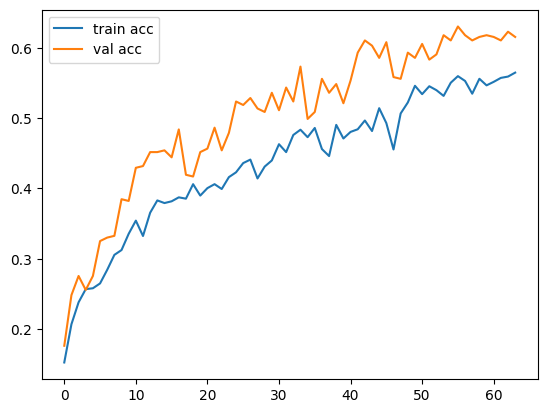
**LeNet –**

****

**Confusion Matrix**

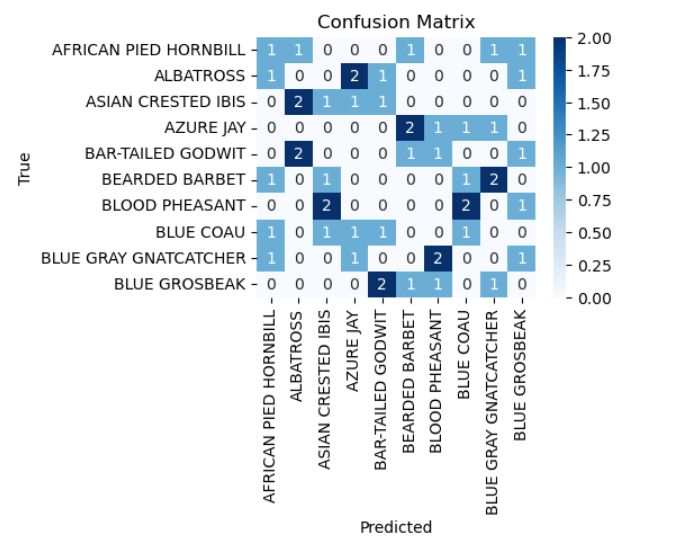
****

**Loss vs epoch**

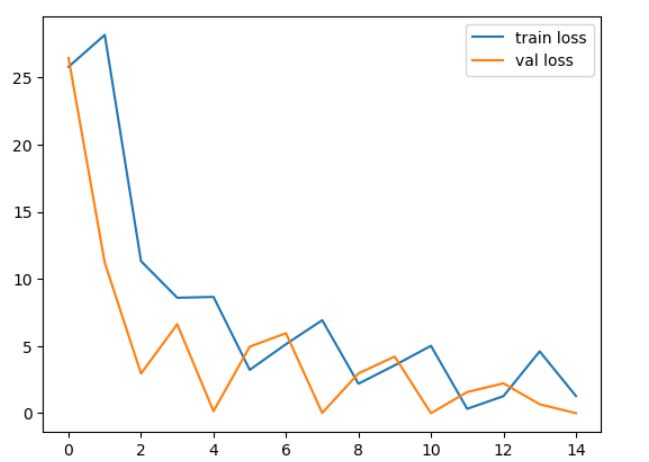
****

**Accuracy vs epoch**

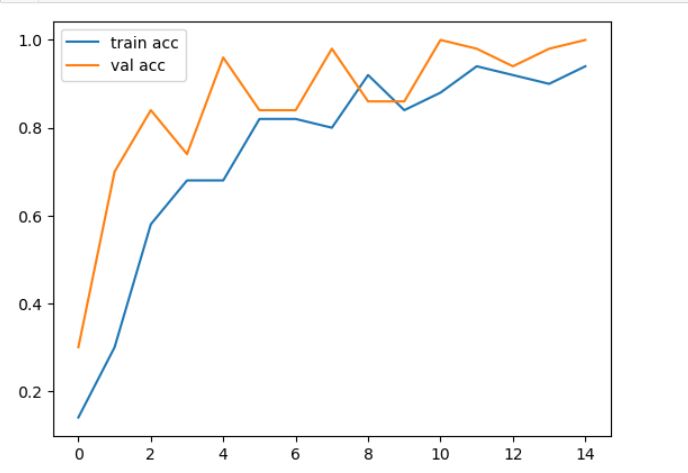
**GoogleNet –**



**Confusion Matrix**

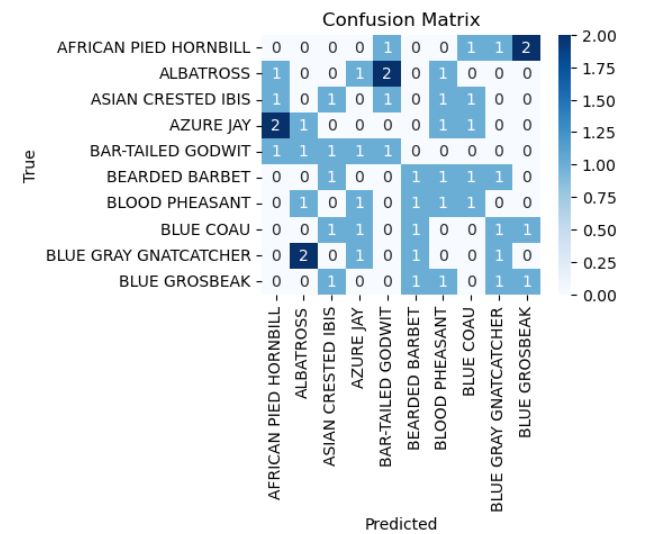


**Loss vs epoch**

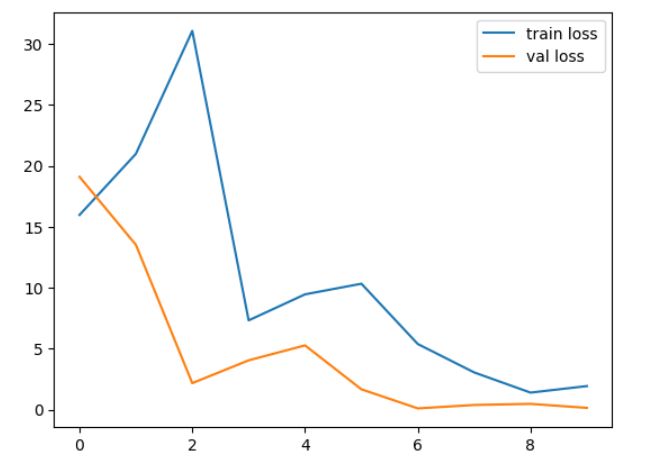


**Accuracy vs epoch**

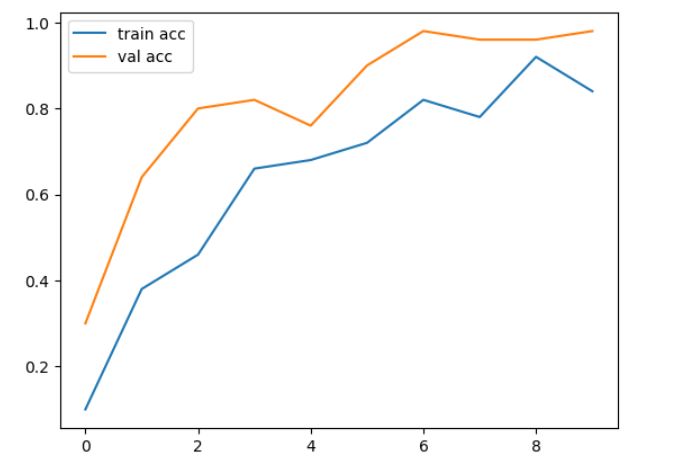
**AlexNet –**



**Confusion Matrix**

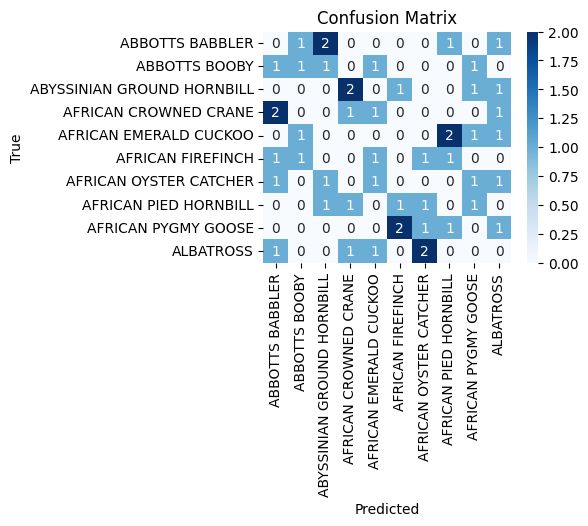


**Loss vs epoch**

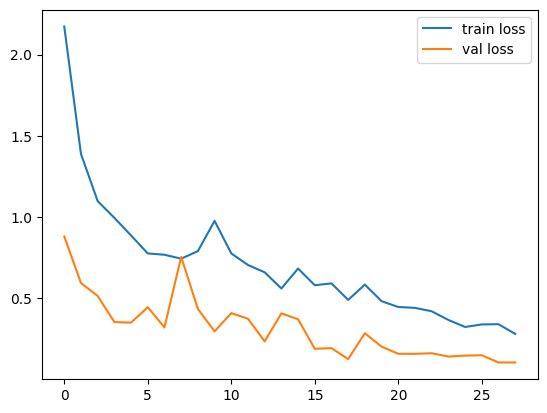


**Accuracy vs epoch**

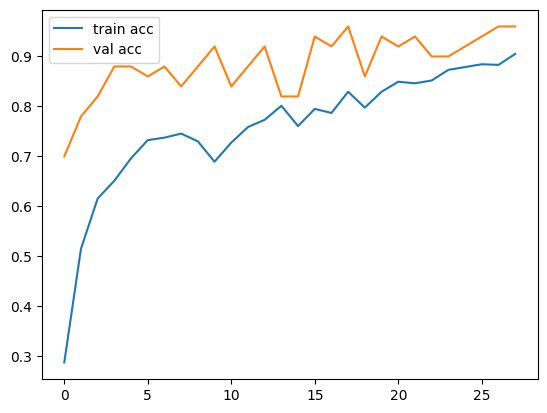
**VGG16 –**



**Confusion Matrix**

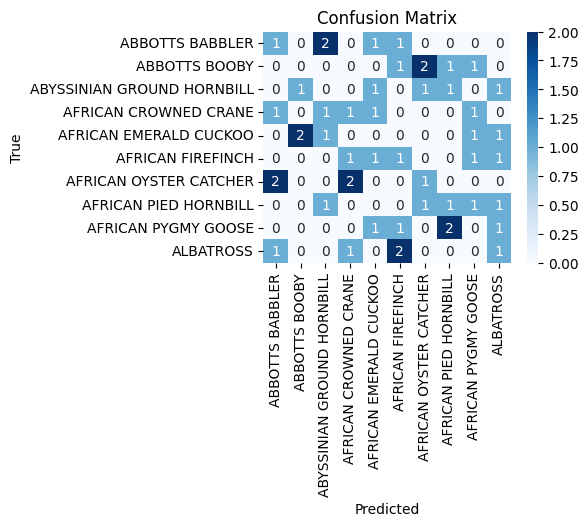


**Loss vs epoch**

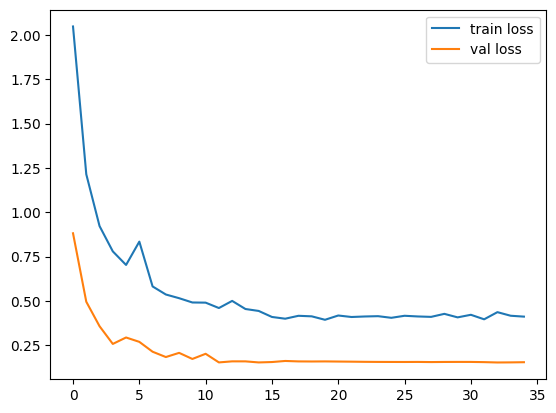


**Accuracy vs epoch**

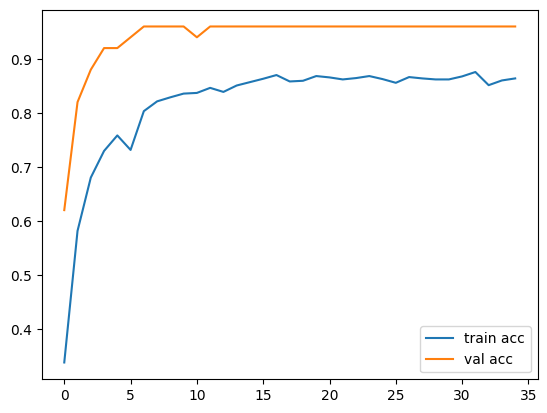
**ResNet –**

****

**Confusion Matrix**

****

**Loss vs epoch**

****

**Accuracy vs epoch**

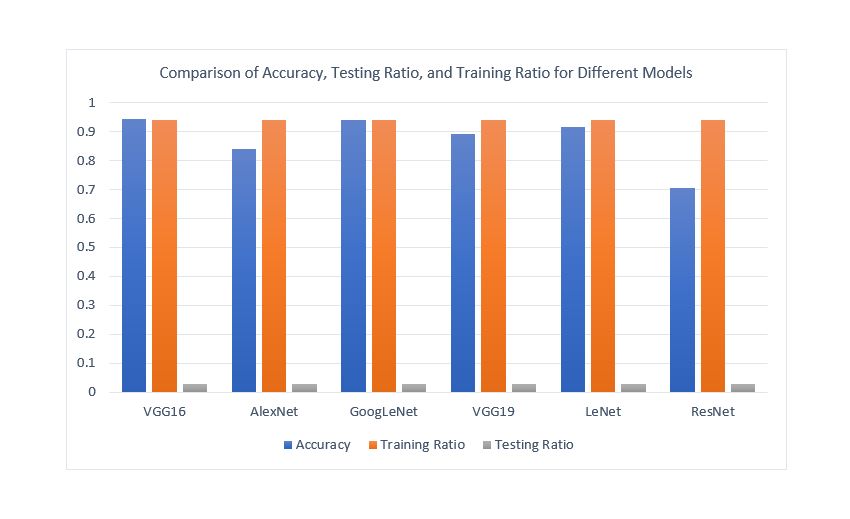
The analysis of

various deep learning architectures. The proposed model has utilized the same training, testing and valid dataset.

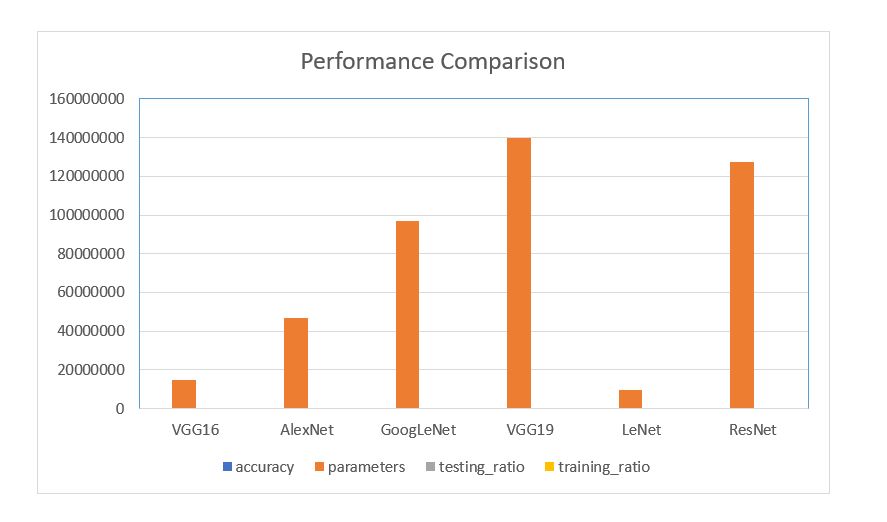
|  |  |  |  |
| --- | --- | --- | --- |
| CNN Model | Accuracy Rate | Train:Test: valid Ratio | Parameters × 10¹ million |
| VGG-19 | 0.8768 | 94:3:3 | 13.96 |
| LeNet | 0.8471 | 94:3:3 | 0.00617 |
| AlexNet | 0.8487 | 94:3:3 | 4.67 |
| GoogLeNet | 0.9130 | 94:3:3 | 9.70 |
| VGG-16 | 0.9075 | 94:3:3 | 1.49 |
| ResNet | 0.8642 | 94:3:3 | 12.74 |

COMPARISON OF VARIOUS MODELS

The proposed 6 CNN model is compared by varying the with Accuracy rate, ratios of the (test, train and valid) and parameters. There is only slight variations in Accuracy rates between CNN models.



Performance Comparison of Accuracy, testing ang training ratio.



Performance comparison of parameters

# Conclusion and Future work

##### The preparation behind developing the identification code is to create awareness regarding bird-watching, bird and their identification, notably birds found in land. The main purpose of the project is to classify the bird species from a picture given as input by the user. It additionally caters to would love of simplifying bird identification methodology then creating bird-watching easier. The technology used in the experimental setup is Convolutional Neural Networks (CNN) along with deep learning. It uses feature extraction for image recognition for classification. The method used is good enough to extract features and classify images.

##### The main purpose of our project is to identify the bird species from an image given as input by the user. We used CNN because it is suitable for implementing the advanced algorithms and gives good numerical precision accuracy. We achieved an accuracy of 75%- 95%. We believe this project extends a great deal of scope as the purpose meets. In wildlife research and monitoring, this concept can be implemented in-camera traps to maintain the record of wildlife movement in specific habitat and behavior of any species.

##### **References**

1. Dataset- -

<https://www.kaggle.com/datasets/gpiosenka/100-bird-species>

[2] Yoonchang Han and Kyogu Lee. Acoustic scene classification using convolutional neural network and multiplewidth frequency-delta data augmentation. 14(80:1-11, 2022.

1. Tóth, B.P. and Czeba, B., 2022, September. Convolutional Neural Networks for Large-Scale Bird Song Classification in Noisy Environment. In CLEF (Working Notes) (pp. 560-568)
2. Yann LeCun, Yoshua Bengio, and Hinton Geoffrey. Deep learning. Nature Methods, 13(1):35-35, 2020.
3. Sergey loffe and Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by ReducingInternal Covariate Shift. arXiv:1502.03167, pages 1-11, 2020.
4. Jaderick P. Pabico, Anne Muriel V. Gonzales, Mariann Jocel S. Villanueva, and Arlene a. Mendoza. Automatic identification of animal breeds and species using bioacoustics and artificial neural networks.arXiv preprint, pages 1- 17, 2019.
5. R. Bardeli, D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, and K.-H. Frommolt, “Detecting bird songs in a complex acoustic environment and application to bioacoustic monitoring,” Patt Recog Letters, vol. Vol.31, pp. Pp. 1524-1534, 2019.

[7] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller, “Scikit-learn,” GetMobile Mob. Comput. Commun., vol. 19, no. 1, pp. 29–33, Jun. 2018.

[8] K. Huang, S. Li, X. Kang, and L. Fang, “Spectral–Spatial Hyperspectral Image Classification Based on KNN,” Sens. Imaging, vol. 17, no. 1, p. 1, Dec. 2017.

[9] M. K. (Manoj K. Jha, B.-S. WSEAS (Organization), and S. Savarese, “Comparing image classification methods: K-nearest-neighbor and support-vector-machines,” WSEAS, 2017

[10] Zagoruyko, S. and Komodakis, N., 2016. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. arXiv preprint arXiv:1612.03928

[11] Cireşan, D., Meier, U. and Schmidhuber, J., 2016. Multi-column deep neural networks for image classification. arXiv preprint arXiv:1202.2745

[12] Thomas Berg, Jiongxin Liu, Seung Woo Lee, Michelle L. Alexander, David W. Jacobs, and Peter N. Belhumeur Birdsnap: Large-scaleFine-grainedVisualCategorizationofBirds

[13] Yuning Chai Electrical Engineering Dept. ETH Zurich,Victor Lempitsky Dept. of Engineering Science University of Oxford, Andrew Zisserman Dept. of Engineering Science University of Oxford BiCoS:A BiSegmentation Method for Image Classification.

[14] Fagerlund, S., 2015. Bird species recognition using support vector machines. EURASIP Journal on Applied Signal Processing, 2015(1), pp.64-64

[15] Inception-v4,Inception-ResNetand the Impact of Residual Connectionson Learning Christian Szegedy, Sergey Ioffe,Vincent Vanhoucke ,Alexandrr A.Alemi

[16] Cireşan, D., Meier, U. and Schmidhuber, J., 2012. Multi-column deep neural networks for image classification. arXiv preprint arXiv:1202.2745