

Recognition of Local Birds of Bangladesh using MobileNet and Inception-v3

Md. Mahbubur Rahman¹, Al Amin Biswas², Aditya Rajbongshi³, Anup Majumder⁴

Crowd Realty, Tokyo, Japan¹

Dept. of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh²

Dept. of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh^{3,4}

Abstract—Recognition of bird species can be a challenging task due to various complex factors. The purpose of this work is to distinguish various local bird species of Bangladesh from the image data. The MobileNet and Inception-v3 model which is mainly an image classification model used here to accomplish this work. Here, we have used a total of four approaches namely Inception-v3 without transfer learning, Inception-v3 with transfer learning, MobileNet without transfer learning, and MobileNet with transfer learning to accomplish the task. To evaluate our experimental results, we have calculated F1 Score besides the model's accuracy and also presented the ROC curve to evaluate the model's output quality. Then we have done a comparison among the applied four approaches. The experimental result has proved the working capability of the applied four approaches. Among these four approaches, MobileNet with transfer learning outperforms the others and obtained a test accuracy of 91.00%. For each of the classes, MobileNet with transfer learning obtained the highest F1 Score than other approaches.

Keywords—Recognition; MobileNet; Inception-v3; transfer learning; computer vision; Bangladeshi bird

I. INTRODUCTION

The bird is one of the amazing creations of God. Besides adding charm and beauty to nature, they also help to sustain a balance of the fresh environment in the world. Bangladesh has a friendly environment to live for the birds. It is very difficult to recognize the local birds because of the bird's nature, color, structure, and voice.

Images can be classified in several ways. Image classification aims to category a distinct picture according to a set of probable categories. From a deep learning aspect [1], the image classification problem can be solved by transfer learning. This learning technique is becoming more popular day by day in computer vision because of its ability to construct accurate models within a very short time [2]. The main advantages of transfer learning techniques [3], there is no need for a huge training dataset and also no need for much more computational power. In computer vision, Transfer learning is normally exposed through the practice of pre-trained models. A pre-trained model [1] is a technique that was trained on a huge dataset to resolve a problem similar to the one problem that individual desires to solve.

Here, we have mainly performed the task of recognition of the seven birds. Seven local birds of Bangladesh namely Bulbuli(বুলবুলি), Chorui(চড়ই), Kaak(কাক), Doel(দোয়েল), Machranga(মাছরাঙ্গা), Shalik(শালিক), and Kokil(কোকিল) including the national birds which are shown in the Fig. 1. In this paper, we have implemented two models (two different approaches for each of the models) to accomplish this work of local bird recognition in Bangladesh. Firstly, we have performed Inception-v3 and MobileNet without using transfer technology and then we have again performed Inception-v3 and MobileNet using transfer technology. Then we have compared our observations on the basis of performance evaluation metrics namely accuracy and F1 Score.

This paper is divided into five sections and presented as follows: Literature Review is presented in Section II. Section III describes the Methodology of this work. Experimental Result and Analysis are presented in Section IV. Conclusion is presented in Section V. Lastly; Future Work is mentioned in Section VI, respectively.



Fig. 1. An Introduction of Seven Local Birds.

II. LITERATURE REVIEW

Much research has been worked on transfer learning by many researchers but few have been applied this learning technique on local birds of Bangladesh. Many researchers have worked on bird recognition with many traditional diverse solutions.

A method of bird recognition based on Inception-v3 is presented by J. Bai et al. [4]. To improve the robustness and generalization of the model, they have applied some of the data augmentation techniques. Lastly, they evaluated their system in BirdCLEF test data and obtained 0.055 of classification mean average precision.

A deep learning neural network technique is presented to identify the bird species by S.K. Pillai et al. [5]. To implement this work, they used the Tensor Flow framework and twelve bird species were taken from the Caltech dataset to test their system. From their result, the true prediction (training dataset) as well as true prediction (unknown image) is always more than 0.5 and in some cases, both the values are more than 0.9 which proves the accuracy of the algorithm. For this dataset, train accuracy starts from 29% (at step 0) and goes till 100% (at step 1500). For their work, it is observable that the relationship between validation accuracy and the cross-entropy is inversely proportional.

M. Lasseck [6] presented deep learning methods for the detection of the acoustic bird. Deep Convolutional Neural Networks originally outlined for the classification of the image, are adjusted and fine-tuned to recognize the bird's presence in audio recordings. To improve model performance, various data augmentation techniques were implemented here.

N. R. Gavai et al. [7] have presented the experimental performance of the MobileNets model to retrain the flower kind datasets, which can exceedingly reduce the time and space for flower classification compromising the accuracy insignificantly.

J. Bankar and N. R. Gavai [8] used a machine learning technique to classify the animal into particularized classes. They used the transfer learning technology to retrain the animal kind datasets based on the Inception-v3 model in the TensorFlow platform, which can exceedingly enhance the accuracy of classification in this circumstance.

Detection of pest birds in the field of agriculture is addressed by S. Lee et al. [9]. They proposed adapting deep learning techniques to certain cropped small areas of a frame where there is a huge chance of bird presence, based on the image processing's result. By conducting the background subtraction based on Gaussian Mixture Model, the moving objects are extracted. After that, color extraction and median filter eliminate the undesirable things in the agricultural environment. Lastly, the neural network object classifier was used to classify and minimize moving objects. The test results showed that employing neural networks to image's precise areas causes greater accuracy than the entire original frame.

S. S. Londhe and S. S. Kanade [10] studied the way of automatic bird's species identification by their vocalization.

Bird sounds are classed by their function into songs and calls which are moreover subdivided into hierarchical levels of phrases, syllables, and elements. It is pointed out here that the syllable is a suitable unit to recognize bird species. Variety within diverse types of syllables birds are capable of producing is large. To identify species, the Support Vector Machine is chosen here.

A real-time detection system of birds in flight based on background subtraction and tracking through point correspondence is described by Moein Shakeri and Hong Zhang [11]. They have used a single fixed camera for their bird detection system and used Zivkovic's background subtraction technique. They appended a correspondence component based on point-tracking to the background subtraction technique to reach reliable bird detection. Investigations were carried to analyze the detection performance accepting objects of varied sizes, colors, and velocity. The obtained results showed the accuracy and efficiency of their system.

M. R. Islam et al. [12] developed a MobileNet model for local bird classification using CNN. They have used a small dataset. They utilized 500 images in 5 classes of local birds. Their obtained accuracy is about 100%.

X. Xia et al. [13] utilized the Inception-v3 model for the classification of flowers. For retraining the datasets, transfer learning technology was practiced. Two types of datasets namely Oxford-17 and Oxford -102 flowers were employed and the classification process was conducted using four steps. 95% percent accuracy is found for the dataset Oxford-17 and 94% for the dataset Oxford-102.

Howard et al. [14] presented a model which is the application of a mobile vision named ModelNet. The MobileNet architecture is based on streamlined. For choosing the right sized model based on the problems, they mentioned two hyperparameters which are global. MobileNet effectiveness is also demonstrated by focusing on various classification problems.

III. METHODOLOGY

This section represents the way we have used to implement this work. This section basically consists of five subsections namely Dataset, Data Preprocessing, Model Description, Training, and lastly Testing. The implementation procedure of our work is presented in Fig. 2.

The detailed description of the abovementioned five subsections is presented below.

A. Dataset

In this work, we have collected seven local birds' images. Details of the seven local birds with mentioning the local name, English name, and also scientific name is presented in Table I. Primarily, we have collected 100 images for each bird species. As we are working with seven types of local birds. So, our total collected data is 700 image data.

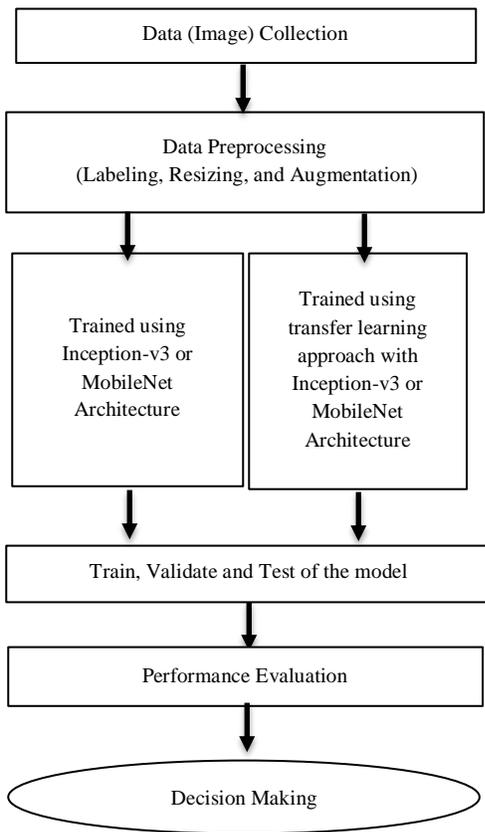


Fig. 2. The Implementation Procedure of Recognition of Local Birds of Bangladesh using MobileNet and Inception-v3.

TABLE I. REPRESENTATION OF THE SEVEN LOCAL BIRDS

Bird's Local Bangla Name	Bird's English Name	Bird's Scientific Name
বুলবুলি পাখি (Bulbuli)	Red-vented bulbul	<i>Pycnonotus cafer</i>
চডুই পাখি (Chorui)	House Sparrow	<i>Passer domesticus</i>
কাক পাখি (Kaak)	Crow	<i>Corvus corax</i>
দোয়েল পাখি (Doel)	Magpie	<i>Copsychus saularis</i>
মাছরাঙ্গা পাখি (Machranga)	Common kingfisher	<i>Alcedo atthis</i>
শালিক পাখি (Shalik)	Myna	<i>Acridotheres tristis</i>
কোকিল পাখি (Kokil)	Cuckoo	<i>Cuculus canorus</i>

B. Data Preprocessing

The data preprocessing step is included with data resizing, labeling, and augmentation. Since, in deep learning, we need a sufficient amount of data but in some cases, it is not feasible to collect. So, in this type of case, data augmentation techniques can help individuals to rescue. So, to increase the

size of the dataset, we have performed the data augmentation technique on the original dataset. Data augmentation can be done by several operations. The most commonly used operations are rotation, shearing, zooming, cropping, flipping, and contrast adjustment using histogram equalization, adaptive histogram equalization technique. We got 500 images for each class after augmentation. So, after the augmentation of data, we have a total of 3500 image data. We have resized this augmented label dataset for training and testing purposes. Then, we divided the data randomly into two sets (train-tests) with an 80% to 20% ratio.

C. Model Description

Here, we worked on Inception-v3 and MobileNet architecture in two different ways. One, we have used Inception-v3 and MobileNet architectures without transfer learning. And another way is using the transfer learning approach and for this, we used the pre-trained weight of these architectures on the imagenet dataset of 1000 classes. The details of MobileNet and Inception-v3 architectures for the imagenet dataset are presented in Table II. For our training purpose, we discarded the output layer of 1000 neurons and added an output layer of 7 neurons as in our dataset we have only worked on the 7 classes. During transfer learning with Inception-v3, we make the top 2 inception blocks freeze and the rest of the layer trainable. We made all the layers trainable when we did transfer learning with MobileNet architecture. As we were getting high overfitting with transfer learning, we used a dropout layer with a rate of 0.5 after the final fully connected layer to reduce the overfitting.

D. Training

To train the models, we used an Nvidia Geforce 2080 GPU with 8GB Memory. We trained the data in 100 epochs with a batch size of 8. For transfer learning, we used the pre-trained weight on the imagenet dataset. During training, we used the test set as the validation set. The accuracy with respect to epochs for all the models is presented in the Fig. 3. Blue line is used here to present the training accuracy and the other one is used here to present the validation accuracy.

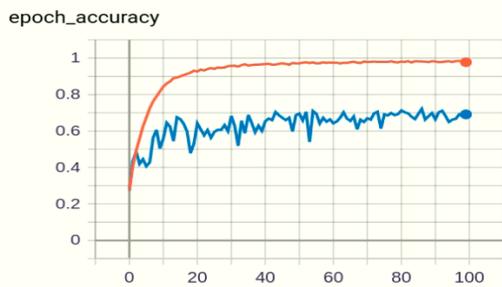
E. Testing

After training, we tested the newly trained models with the test data. We got the following 7*7 confusion matrix for the class set {'Bulbuli', 'Chorui', 'Kaak', 'Doel', 'Machranga', 'Shalik', 'Kokil'} which is shown in Table III.

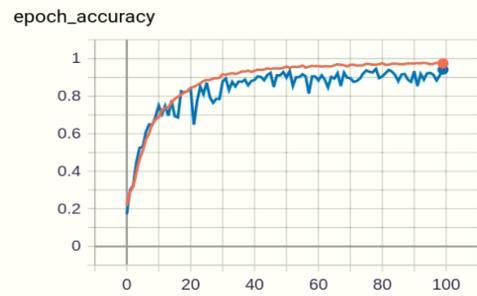
The binary representation of the 7*7 confusion matrix shown in Table III is presented in Table IV. A Binary matrix is easy to interpret and understand. As 700 images are used to testing so, the summation of TP, FP, FN, and TN is equal to 700 for each of the bird classes.

TABLE II. MOBILENET AND INCEPTION-V3 ARCHITECTURE

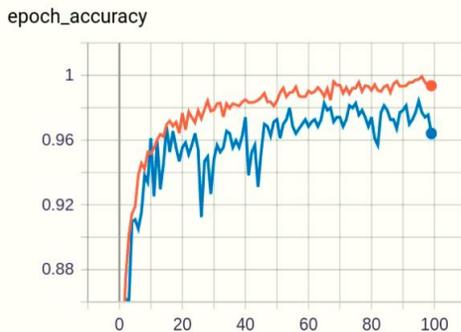
MobileNet			Inception-v3		
Type/Stride	Filter Shape	Input Size	Type/ Stride	Patch size	Input size
Conv / s2	3×3×3×32	224×224×3	Conv /s2	3×3	299×299×3
Conv dw / s1	3×3×32 dw	112×112×32	Conv /s1	3×3	149×149×32
Conv / s1	1×1×32×64	112×112×32	Conv padded /s1	3×3	147×147×32
Conv dw / s2	3×3×64 dw	112×112×64	Pool /s2	3×3	147×147×64
Conv / s1	1×1×64×128	56×56×64	Conv / s1	3×3	73×73×64
Conv dw / s1	3×3×128 dw	56×56×128	Conv / s2	3×3	71×71×80
Conv / s1	1×1×128×128	56×56×128	Conv /s1	3×3	35×35×192
Conv dw /s2	3×3×128 dw	56×56×128	3×Inception	Inception Module	35×35×288
Conv / s1	1×1×128×256	28×28×128	5×Inception	Inception Module	17×17×768
Conv dw /s1	3×3×256 dw	28×28×256	2×Inception	Inception Module	8×8×1280
Conv / s1	1×1×256×256	28×28×256	Avg Pool	Pool 8×8	8×8×2048
Conv dw / s2	3×3×256 dw	28×28×256	FC	2048 × 1000	1×1×2048
Conv / s1	1×1×256×512	14×14×256	Softmax	classifier	1×1×1000
Conv dw/s1 5× Conv / s1.	3×3×512 dw 1×1×512×512	14×14×512 14×14×512			
Conv dw /s2	3×3×512 dw	14×14×512			
Conv / s1	1×1×512×1024	7×7×512			
Conv dw /s2	3×3×1024 dw	7×7×1024			
Conv / s1	1×1×1024×1024	7×7×1024			
Avg Pool	Pool 7×7	7×7×1024			
FC	1024×1000	1×1×1024			
Softmax	Classifier	1×1×1000			



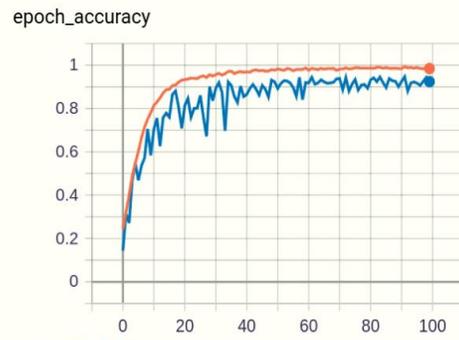
(a) Inception-v3 with Transfer Learning.



(b) Inception-v3 without Transfer Learning.



(c) MobileNet with Transfer Learning.



(d) MobileNet without Transfer Learning.

Fig. 3. Accuracy over the Epochs of different Models.

TABLE III. CONFUSION MATRIX BASED ON THE TEST SET

Method: Inception-v3 without transfer learning		Predicted						
		Bulbuli	Chorui	Kaak	Doel	Machranga	Shalik	Kokil
Actual	Bulbuli	87	8	3	0	0	1	1
	Chorui	1	82	10	0	0	3	4
	Kaak	0	0	87	6	0	1	6
	Doel	0	0	0	91	8	0	1
	Machranga	0	0	3	0	89	5	3
	Shalik	2	1	1	1	2	82	11
	Kokil	8	2	7	2	0	0	81
Method: MobileNet without transfer learning		Predicted						
		Bulbuli	Chorui	Kaak	Doel	Machranga	Shalik	Kokil
Actual	Bulbuli	86	8	2	1	0	0	3
	Chorui	1	88	9	1	0	0	1
	Kaak	0	0	84	9	0	0	7
	Doel	0	0	2	90	8	0	0
	Machranga	1	0	2	2	88	5	2
	Shalik	1	1	6	1	0	81	10
	Kokil	8	0	8	0	0	0	84
Method: Inception-v3 with transfer learning		Predicted						
		Bulbuli	Chorui	Kaak	Doel	Machranga	Shalik	Kokil
Actual	Bulbuli	53	8	8	4	5	11	11
	Chorui	1	64	15	4	3	12	1
	Kaak	6	0	76	5	0	6	7
	Doel	1	2	3	71	14	8	1
	Machranga	6	10	1	4	64	12	3
	Shalik	4	5	7	3	6	64	11
	Kokil	9	1	7	2	4	7	70
Method: MobileNet with transfer learning		Predicted						
		Bulbuli	Chorui	Kaak	Doel	Machranga	Shalik	Kokil
Actual	Bulbuli	90	7	2	0	0	0	1
	Chorui	0	91	8	0	0	0	1
	Kaak	0	0	92	7	0	0	1
	Doel	0	0	1	90	8	0	1
	Machranga	0	0	1	0	92	7	0
	Shalik	0	0	1	0	1	90	8
	Kokil	8	0	0	0	0	0	92

TABLE IV. CONFUSION MATRIX AS A BINARY FORMAT

Class Name	Method	True Positive	False Positive	False Negative	True Negative
বুলবুলি (Bulbuli)	Inception-v3 without transfer learning	87	11	13	589
চডুই (Chorui)		82	11	18	589
কাক (Kaak)		87	24	13	576
দোয়েল (Doel)		91	9	9	591
মাছরাঙ্গা (Machranga)		89	10	11	590
শালিক (Shalik)		82	10	18	590
কোকিল (Kokil)		81	26	19	574
বুলবুলি (Bulbuli)	MobileNet without transfer learning	86	11	14	589
চডুই (Chorui)		88	9	12	591
কাক (Kaak)		84	29	16	571
দোয়েল (Doel)		90	14	10	586
মাছরাঙ্গা (Machranga)		88	8	12	592
শালিক (Shalik)		81	5	19	595
কোকিল (Kokil)		84	23	16	577
বুলবুলি (Bulbuli)	Inception-v3 with transfer learning	53	27	47	573
চডুই (Chorui)		64	26	36	574
কাক (Kaak)		76	41	24	559
দোয়েল (Doel)		71	22	29	578
মাছরাঙ্গা (Machranga)		64	32	36	568
শালিক (Shalik)		64	56	36	544
কোকিল (Kokil)		70	34	30	566
বুলবুলি (Bulbuli)	MobileNet with transfer learning	90	8	10	592
চডুই (Chorui)		91	7	9	593
কাক (Kaak)		92	13	8	587
দোয়েল (Doel)		90	7	10	593
মাছরাঙ্গা (Machranga)		92	9	8	591
শালিক (Shalik)		90	7	10	593
কোকিল (Kokil)		92	12	8	588

IV. EXPERIMENTAL RESULT AND ANALYSIS

The following formula ((1) to (4)) is used to calculate the performance evaluation metrics.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (4)$$

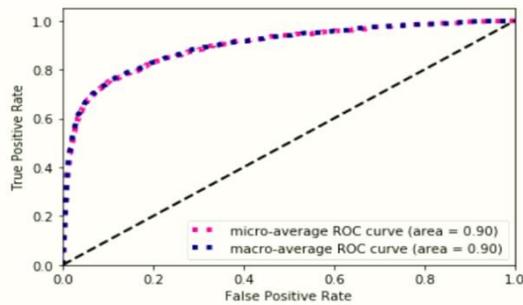
In Table V, we can see that MobileNet with Transfer Learning among the four approaches achieved the highest accuracy of 91.00%. We have also calculated another performance evaluation metrics F1 Score of the applied four approaches for each of the seven classes. From Table V, we found that F1 Score of MobileNet with the transfer learning

approach for the class বুলবুলি পাখি (Bulbuli), চডুই পাখি (Chorui), কাক পাখি (Kaak), দোয়েল পাখি (Doel), মাছরাঙ্গা পাখি (Machranga), শালিক পাখি (Shalik), and কোকিল পাখি (Kokil) is 90.91%, 91.92%, 89.76%, 91.37%, 91.54%, 91.37%, and 90.20%, respectively. Here, MobileNet with the transfer learning approach has outperformed than other approaches for each of the classes.

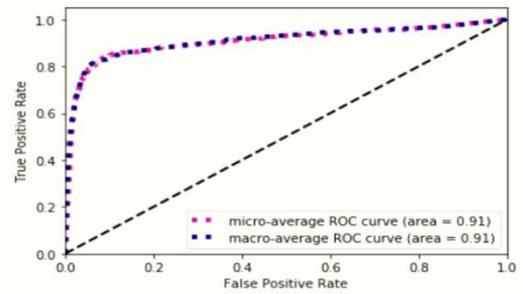
Receiver Operating Characteristic (ROC) metric is usually used to evaluate the quality of classifier's output. In the ROC curve, true positive rate places on the Y-axis and on the other hand false positive rate places on the X-axis. This denotes that the top left corner of the plot is the ideal point because here FPR is lowest, and TPR is highest. This is not very practical, but it does indicate that a larger AUC is usually better. From Fig. 4, we can see that the ROC of MobileNet with transfer learning approaches has larger AUC than others which indicates that this approach is working better here.

TABLE V. RESULT ANALYSIS BASED ON PERFORMANCE EVALUATION METRICS

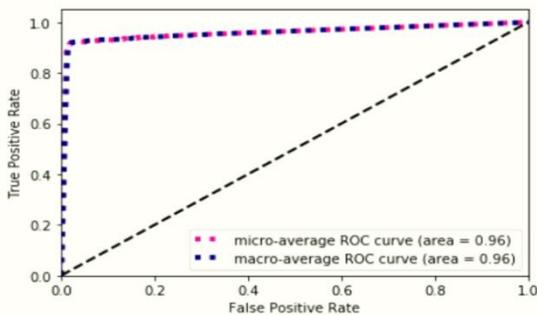
	Without Transfer Learning		With Transfer Learning	
	Inception-v3	MobileNet	Inception-v3	MobileNet
Accuracy (Test)	85.57%	85.86%	66.00%	91.00%
F1 Score [বুলবুলি (Bulbuli)]	87.88%	87.31%	58.89%	90.91%
F1 Score [চডুই (Chorui)]	84.97%	89.34%	67.37%	91.92%
F1 Score [কাক (Kak)]	82.46%	78.87%	70.05%	89.76%
F1 Score [দোয়েল (Doel)]	91.00%	88.24%	73.58%	91.37%
F1 Score [মাছরাঙ্গা (Machranga)]	89.45%	89.80%	65.31%	91.54%
F1 Score [শালিক(Shalik)]	85.42%	87.10%	58.18%	91.37%
F1 Score [কুকিল (Kokil)]	78.26%	81.16%	68.63%	90.20%



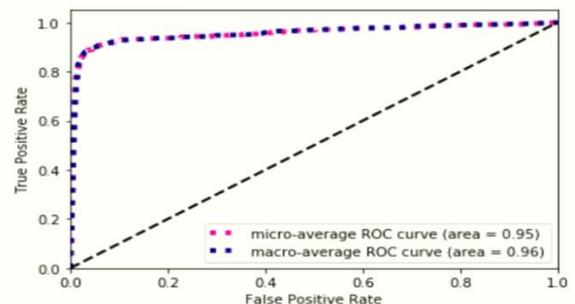
(a) Roc of Inception-v3 with Transfer Learning.



(b) Roc of Inception-v3 without Transfer Learning.



(c) ROC of MobileNet with Transfer Learning.



(d) ROC of MobileNet without Transfer Learning.

Fig. 4. ROC Curve of Four different Approaches.

ROC curves are usually used in binary classification to analyze the output of a classifier. In order to extend the ROC curve and ROC area to multi-label classification, it is essential to binarize the output. The two evaluation measures for multi-label classification are micro-averaging and macro-averaging. This is possible to draw a ROC curve per label, but it is also possible to draw a ROC curve by considering each element of the label indicator matrix as a binary prediction (micro-averaging). Macro-averaging, basically known as another evaluation measure for multi-label classification which gives equal weight to the classification of each label.

V. CONCLUSION

This research work is mainly done on the seven local bird's species of Bangladesh, namely, বুলবুলি পাখি (Bulbuli), চডুই পাখি (Chorui), কাক পাখি (Kaak), দোয়েল পাখি (Doel), মাছরাঙ্গা পাখি (Machranga), শালিক পাখি (Shalik), and

কোকিল পাখি (Kokil). Here, we have applied four approaches. To accomplish the work, we have considered 2800 images data as a training set and 700 images data as a test set. Among the four approaches, MobileNet with transfer learning performed better here in terms of performance evaluation metrics which is shown in Table V.

VI. FUTURE WORK

There are so many local bird species in Bangladesh. So, in the future, we will work with more local bird species with many other transfer learning techniques. Also, we will work on the recognition of the local bird species from the video data in the future.

REFERENCES

- [1] "Transfer learning from pre-trained models," Available Online: <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751> [Last Access: 26 December 2019].

- [2] "Transfer Learning for Image Classification in Keras ," Available Online: <https://towardsdatascience.com/transfer-learning-for-image-classification-in-keras-5585d3ddf54e> [Last Access: 26 December 2019].
- [3] "Deep Learning For Beginners Using Transfer Learning In Keras," Available Online: <https://towardsdatascience.com/keras-transfer-learning-for-beginners-6c9b8b7143e> [Last Access: 26 December 2019].
- [4] J. Bai, B. Wang, C. Chen, Z. H. Fu, and J. Chen, "Inception-v3 Based Method of LifeCLEF 2019 Bird Recognition," CLEF working notes, 2019.
- [5] S. K. Pillai, M. M. Raghuvanshi, and U. Shrawankar, "Deep Learning Neural Network for Identification of Bird Species," In Computing and Network Sustainability, vol. 75, pp. 291-298, 2019.
- [6] M. Lasseck, "Acoustic Bird Detection with Deep Convolutional Neural Networks," in Proceedings of the Detection and Classification of Acoustic Scenes and Events 2018 Workshop (DCASE2018), pp. 143-147, 2018.
- [7] N. R. Gavai, Y. A. Jakhade, S. A. Tribhuvan, and R. Bhattad, "MobileNets for Flower Classification Using TensorFlow," International Conference on Big Data, IoT and Data Science (BIG), pp. 154-158, 2017.
- [8] J. Bankar, and N. R. Gavai, "Convolutional Neural Network-based Inception v3 Model for Animal Classification," International Journal of Advanced Research in Computer and Communication Engineering, vol. 7, no. 5, pp. 142-146, 2018.
- [9] S. Lee, M. Lee, H. Jeon, and A. Smith, "Bird Detection in Agriculture Environment using Image Processing and Neural Network," In 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), pp. 1658-1663, 2019.
- [10] S. S. Londhe, and S. S. Kanade, "Bird Species Identification Using Support Vector Machine," International Journal of Computer Science and Mobile Computing, vol. 4, no. 9, pp. 125-135, 2015.
- [11] M. Shakeri, and H. Zhang, "Real-Time Bird Detection Based on Background Subtraction," In Proceedings of the 10th World Congress on Intelligent Control and Automation, pp. 4507-4510, 2012.
- [12] M. R. Islam, N. Tasnim, and S. B. Shuvo, "MobileNet Model for Classifying Local Birds of Bangladesh from Image Content Using Convolutional Neural Network," In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-4. IEEE, 2019.
- [13] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," In 2017 2nd International Conference on Image, Vision and Computing (ICIVC), pp. 783-787. IEEE, 2017.
- [14] A. G. Howard., M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.