Bird Species Classification from an Image Using VGG-16 Network

Shazzadul Islam
North South University
Dhaka, Bangladesh
shazzadul.islam@northsouth.edu

Sabit Ibn Ali Khan
North South University
Dhaka, Bangladesh
sabit.khan@northsouth.edu

Md. Minhazul Abedin
North South University
Dhaka, bangladesh
abedin.minhazul@northsouth.edu

Khan Mohammad Habibullah
North South University
Dhaka, Bangladesh
h.rana10@gmail.com

Amit Kumar Das
East West University
Dhaka, Bangladesh
amit.csedu@gmail.com

ABSTRACT
Birds are an integral part of any environment and they are of the utmost importance to nature. Considering this, it is clear how necessary it is to be able to identify birds in the wilderness. This paper proposes a Machine Learning approach to identify Bangladeshi birds according to their species. We used VGG-16 network as our model to extract the features from bird images. In order to perform the classification, we used a data set that contains pictures of different bird species of Bangladesh which were used as they are, without any annotation. We then used various classification methods, where each method gave us different results. However, compared to other classification methods such as Random Forest and K-Nearest Neighbor (KNN), Support Vector Machine (SVM) gave us the maximum accuracy of 89%.

CCS Concepts
• **Computing methodologies→Supervised learning by classification.**

Keywords
VGG-16; Bird's species classification; Support Vector Machine; K-Nearest Neighbors; Random Forest.

1. INTRODUCTION
Birds are vital in regards to maintaining the balance in the ecosystem. The study of birds can help to better understand the world around us and to grasp significant information regarding nature. Bird identification is a well-known problem for ornithologists. Environmental scientists often use birds to understand ecosystems because they are sensitive to environmental changes. Various real-world applications rely on birds, such as monitoring of environmental pollution [1]. The presence of birds of different species in an ecosystem is also important for many environmental reasons. This is another area where our classification program can be of use. With the help of our classification method, the authorities can also keep track of bird hunting in an area by monitoring the population rise and fall of each species of birds.

ICCCM 2019, July 27–29, 2019, Bangkok, Thailand

ACM ISBN 978-1-4503-7195-7/19/07…$15.00
https://doi.org/10.1145/3348445.3348480

Nowadays, image classification has become one of the major research fields of machine learning and deep learning [22]. Identification of bird species from an image is a challenging task due to issues like different subspecies of birds varying dramatically in shape and appearance, the background of images, lighting conditions in photos and extreme variation in the pose. In this paper, an attempt to correctly classify a bird and state its species will be described. We used a VGG-16 network that was modified to suit our needs. VGG-16 was used as our model because it was found to be the best choice regarding feature extraction [3]. VGG-16 also gives us very rich feature representative vectors, which is very helpful when implementing the classification methods. Classification methods such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Random Forest (RF) were then implemented on the model.

Since birds are extremely important to Bangladesh's ecosystem, this paper will help understand the ecology, population distribution of birds and other such factors in Bangladesh. Initially, we had to create the dataset from collected data and then the data had to be processed so that it become suitable to use. The features of the birds were then extracted from each image using the VGG-16 model. Based on the extracted features, we implemented various methods in order to perform classification of the bird species. Finally, we used certain metrics to calculate the quality of classification done by our method.

Throughout the paper, in section 2 we have talked about other related works, section 3 described about dataset and features extraction using VGG-16, section 4 described classifiers that are used for classification. In section 5 we have discussed about performance and result summary. At the end in section 6, we talked about conclusion and future work.

2. RELATED WORK
There have already been several types of research work done which have focused on transfer learning techniques [16, 22] for the purpose of classification. The current state-of-the-art CNNs utilize pre-trained networks and are able to categorized bird poses

from different angles and bird positions. These methods have improved accuracies to about 85.4% [16] . Another work based on transfer learning of the pre-trained AlexNet Convolutional Neural Networks (CNN) with fixed feature extraction has achieved the accuracy of 46% [22]. Their second implementation is a multi-class SVM and utilizes computer vision methods to extract Histogram Oriented Gradients and RGB histogram values to classify the birds which boosted their accuracy 5% to 9% than HOG alone.

Recent similar works that have been done to classify trees used images acquired by the WorldView-3 satellite [17]. They initially segmented the images to delineate the subject and then used a VGG-16 network to classify the subject. The results were compared to results from Random Forest and Gradient Boosting and it is observed that the VGG-16 network reaches an accuracy of 92.13%, outperforming both the RF and GB. Overall, we can state that CNN's show the most attractive identifying bird features and thus improving bird classification accuracy.

In [6], D. Rathi, et al recommends a method to automatically classify underwater fish species. The first step of the methodology is to remove all the noise from the dataset. This is achieved by using Image Processing before training to remove dirt, underwater obstructions and non-fish bodies from the dataset. The second step makes use of Deep Learning by implementing CNN in order to make the classifications. When using ReLU as an activation function, the accuracy achieved was 96.29%, whereas it was 72.62% for tanh and 61.91% for Softmax. The computation time was 0.00183 seconds per frame which makes it feasible to use in real-time applications.

In [18], the author proposed VGG-16 for image captioning model. In that case, they remove the last classification layer. After removal of the last layer VGG-16 returns 4096-dimensional feature representative vectors which are later conditioned with word vectors. Finally, they achieved state-of-art performance regarding Bengali image caption. One thing to take note of regarding [22] [16] and [19] is that they all worked on the same data sets, which are the CUB-200-2011 and CUB-200 data sets[20]. This means that the results found by these methods are heavily biased towards these data sets. In order to counter this fact, we have decided to work on a data set based on birds that can be found in Bangladesh.

3. METHODOLOGY

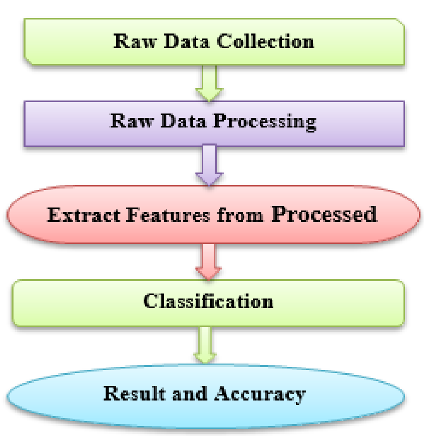

Figure 1. Workflow diagram of bird species classification procedure.

The method used in this paper is a transfer learning technique where a pre-trained CNN is loaded with the last two layers modified. We propose a new model using VGG-16 networks as transfer learning techniques. The reason behind VGG-16 is VGG release 16 layers and 19 layers CNN model. VGG-16 are no longer State-of-art by only few points. However, they are very powerful and useful for image classification and basement for new model which take image as input [3].

Since TensorFlow also runs in the background of VGG-16 we take advantage of TensorFlow library for bird detection. VGG-16 gives us a very high number of dimensional features, and thus we use SVM to perform the classification. KNN, Decision tree, and Multinomial Naïve Bayes are also used because they are all renowned algorithms for classification problems.

3.1. Feature Extraction

For our feature extraction, we used a VGG-16 network that consists of 16 layers [3]. During training, the input images are converted to images of size 224 x 224 as is required by VGG-16. A full stack of convolutional layers which has three fully connected (FC) layers [3]. First two layers have 4096 channels each and third layer contains 1000 channels. The final layer is a soft-max layer.

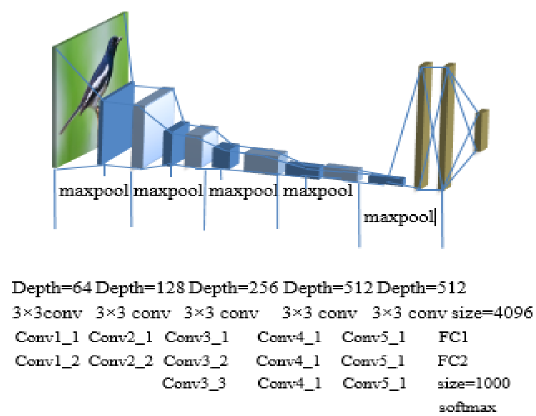

Figure 2. VGG-16 Model.

K. Simonyan et al [3], evaluated the ConvNet configurations by their names (A-E) which is shown in Figure 2. All configurations differ in depth, starting from network A which has 11 weight layers (8 convolutional and 3 FC layers) to 19 weight layers in network E (16 convolutional and 3 FC layers). The width of convolutional layers containing the number of channels start from 64 in first layer. After each max-pooling layer the number is increased by a factor of 2 until the number reaches 512.

However, we remove the last two layers, which are classification layer FC-1000 and Softmax. After removal of the last layer, VGG-16 returns 4096-dimensional feature representative vectors. After getting the features vector we make our own classification layer.

3.2. Dataset Overview

The dataset used for the classification is one that we created manually and it consists of bird species that are native to the country of Bangladesh. Scraping the internet, we collected 1600 images of 27 different species of birds. The images were handpicked from Flickr, Pinterest and Google Images. The images used are high-res pictures of birds in various stages of motion and with varying backgrounds and surroundings. This was done to make sure that the network will be able to identify a bird, irrespective of the surroundings and setting. The average dimensions of the images were 550 x 600. The images did vary in dimension but this was not an issue as the VGG-16 layer converts all images to 224 x 224 before feeding it as input. We used the images without annotation because real life image data does not have annotation in any form. We used the pictures without annotation so that we can absorb the genuine features from the image.

Figure **3**. **A batch of samples from the dataset.**

3.3. Algorithm Implementation

We realized that Scikit learns library [10] can help us to tweak different aspects of an algorithm. As a result, the Scikit learn library is used to call the classification methods. In order to perform the classification, SVM, KNN with n=5 and n=10, Decision Tree and Naive Bayes were implemented for the classification purposes. In order to acquire the training set and test set, 4-fold cross validation was used where the dataset is randomly partitioned to make 4 equal sized subsamples. Out of the 4 subsamples, 1 is used to test the mode, and the other 3 sub-samples are used to train the model. This cross-validation process is repeated 4 times where all the subsamples are taken as the validation data exactly once. The results from all the processes are combined in some way, such as averaged, in order to give a single estimation. Using this method ensures that all the samples are used in both training and testing and also all the samples are used as validation one time.

4. ALGORITHM OVERVIEW

The complete mathematical formula of related algorithms SVM, RF, KNN can be found in [12], [13], [14], [15] and [4]. In this section we are going to discuss about algorithm reliability and process of classification.

4.1. Support Vector Machine (SVM)

Support Vector Machine usually deals with pattern recognition or classification [11]. There are two types of patterns, linear and non-linear. The main theme behind SVM is finding the optimal separation between classes from training samples of the side of class. Those samples of classes are called Support Vectors [12]. If training data set is not linearly separable then kernel method is used for non-linear projection data in high dimension space.

Since our sample is non-linear, we use kernel method so that non-linear data fit in high dimension space that the data can easily separate. If a ϵ^n is projected in a high dimensional space and a_(i) ϵ^n features vector with n components then linear classifier can be expressed as,

$$K(a, a_i) = a^t . a_i$$

4.2. K-Nearest Neighbor (KNN)

K-nearest neighbor is a classifying object method based on the closest training example in the feature [4]. The training process of the algorithm depends on a strong feature vector and labels of the training images.

The K-NN algorithm focused on a distance metric and a voting function in k-Nearest Neighbors, the metric employed is the Euclidean distance [5]. The KNN usually used pixel-based method which rely on the optimal distance matric and feature space [7-9] and [21]. Each pixel is classified in the similar class with respect to the closest intensity of the training dataset. Since we converted each image to a vector of fixed-length with real numbers from VGG-16, we used the Euclidean distance function.

$$d = (a, B) = \|a - B\| = \sqrt{(a - B).(a - B)} \quad = \quad (\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}((\boldsymbol{a_i} - \boldsymbol{B_i})^{\boldsymbol{2}})^{\boldsymbol{1}/\boldsymbol{2}})$$

4.3. Random Forest

The Random Forests [2] is an ensemble classifier that uses decision trees. The random forests compute a response variable by creating many (usually more than hundred) different decision trees and then placing each object to be modeled down each of the decision trees [23], in our sample object is a multi-layered pixel which is ascertained by evaluating the responses from all of the decision trees. For image classification the class that is predicted most is the class that is imposed for that multi-layered pixel object.

In our case image is classified by sending it down every tree and accumulating the leaf ordination. Randomness can be inserted at two points during training. Subsampling the training data so that every individual tree is grown using a different subset; and in picking the node tests.

The trees are all binary trees and they are designed using a top-down approach. At each node, the binary test can be chosen by either of the two methods: Randomly or greedy algorithm.

$$\Delta E = -\sum_i \frac{|R_i|}{|R|} E(R_i)$$

5. RESULT ANALYSIS

The performance of our proposed approach was evaluated on 1600 images of 27 bird species available on our dataset which is described in section 3.2.

We have used four different machine learning algorithms. All of the algorithms gives different types of performance with respect different parameters. For SVM we have used linear kernel. Without any kind of parameter, SVM gives 88% accuracy but after using linear kernel, our performance boosts up for 1%.The performance accuracies of different method are reported in Table 1.

Table 1. Accuracy table

METHOD	ACCURACY RATE	STANDARD DEVIATION
SVM	0.89	0.03
KNN(N=5)	0.82	0.02

KNN(N=10)	0.85	0.01
RANDOM FOREST	0.87	0.02

From table.1, it is observed that the implementation of SVM with linear kernel gives the highest accuracy of 89% followed closely by Random Forest at 87%. When implementing KNN, it can be noted that having N=10 gives a better result than having N=5 giving accuracies of 85% and 82% respectively.

Receiver Operating Characteristic (ROC) is a metric used to evaluate how well our classification method works. The Y-axis of the curves show true positives and X axis shows false positive. We used 5 out of the 25 classes to plot the ROC curves for each individual classifier and calculate the corresponding macro-averaged and micro-averaged accuracy.

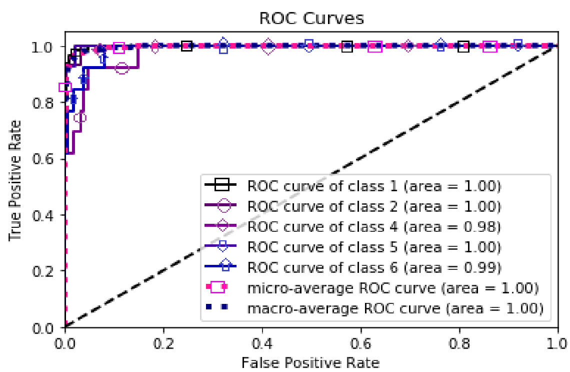

Figure 4. ROC Curve for SVM.

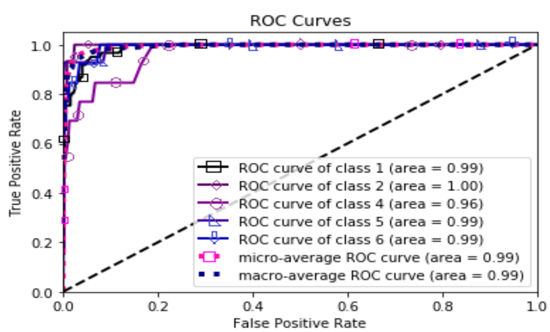

Figure 5. ROC Curve for RF.

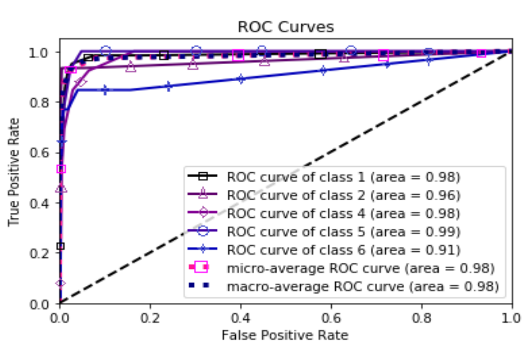

Figure 6. ROC Curve for KNN(N=10).

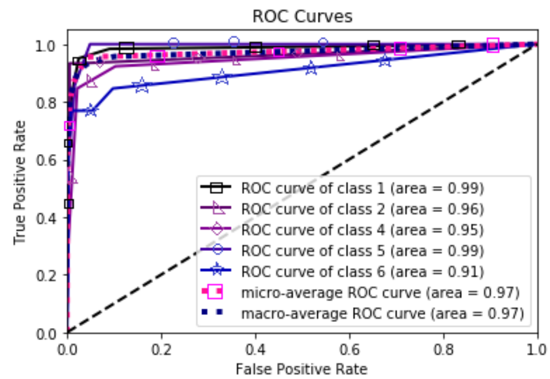

Figure 7. ROC Curve for KNN(N=5).

6. CONCLUSION AND FUTURE WORK

In this paper, we have classified different species of birds in Bangladesh using their images. Initially, we have collected 1600 images of 27 species of birds and created our dataset. VGG-16 was used for creating the model and to extract features from the images. Using these features, we have implemented different machine learning algorithms to classify the species of the birds and achieved 89% accuracy using SVM and kernel method. In the future, we will try to improve the accuracy of the classification methods possibly by using a larger dataset that consists of more species of birds. We will also try to extract features from real-time video in order to perform classification and develop an android application which will be able to classify bird species from real-time photo and video.

7. REFERENCES

[1] "Birds as Environmental Indicators | Environmental Science.org." [Online]. Available: https://www.environmentalscience.org/birds-environmental-indicators. [Accessed: 04-Jun-2019].

[2] W. Man, Y. Ji, and Z. Zhang, "Image classification based on improved random forest algorithm," in 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 2018, pp. 346–350.

[3] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," ICLR,2015.

[4] G. Amato and F. Falchi, "kNN based image classification relying on local feature similarity," in Proceedings of the Third International Conference on SImilarity Search and APplications - SISAP '10, 2010, p. 101.

[5] S. Dave, "Image Classification Algorithm based on Multi-Feature Extraction and KNN Classifier," Int. J. Adv. Eng. Res. Dev., vol. 4, no. 6, 2017.

[6] D. Rathi, S. Jain, and S. Indu, "Underwater Fish Species Classification using Convolutional Neural Network and Deep Learning," in 2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR), 2017, pp. 1–6.

[7] Y. Chen, Z. Lin, and X. Zhao, "Riemannian manifold learning based k-nearest-neighbor for hyperspectral image classification," in 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS, 2013, pp. 1975–1978.

[8] E. Blanzieri and F. Melgani, "Nearest Neighbor Classification of Remote Sensing Images with the Maximal Margin

Principle," IEEE Trans. Geosci. Remote Sens., vol. 46, no. 6, pp. 1804–1811, Jun. 2008.

[9] X. Jia and J.a. Richards. 2005. Fast K-NN Classification Using the Cluster-Space Approach. IEEE Geoscience and Remote Sensing Letters2, 2 (2005), 225–228.

[10] 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. 2015.

[11] G. Mercier and M. Lennon, "Support vector machines for hyperspectral image classification with spectral-based kernels," in IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477), vol. 1, pp. 288–290.

[12] C. J. C. Burges and C. J. C., "A Tutorial on Support Vector Machines for Pattern Recognition," Data Min. Knowl. Discov., vol. 2, no. 2, pp. 121–167, 1998.

[13] M. K. (Manoj K. Jha, B.-S. WSEAS (Organization), and S. Savarese, "Comparing image classification methods: K-nearest-neighbor and support-vector-machines," WSEAS, 2012.

[14] L. Breiman and Leo, "Random Forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[15] A. Bosch, A. Zisserman, and X. Munoz, "Image Classification using Random Forests and Ferns," in 2007 IEEE 11th International Conference on Computer Vision, 2007, pp. 1–8.

[16] S. Branson, G. Van Horn, S. Belongie, and P. Perona, "Bird Species Categorization Using Pose Normalized Deep Convolutional Nets," Jun. 2014.

[17] M. Wang, X. Zhang, X. Niu, F. Wang, Z. Xu, and S. Chen, 2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS 2018).

[18] T. Deb, A. Firoze, and R. Rahman M. Oboyob: A sequential-Semantic Bengali Image Captioning Engine (JIFS).

[19] A. Bhandari, A. Joshi, and R. Patki, "Bird Species Identification from an Image."

[20] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "The Caltech-UCSD Birds-200-2011 Dataset," 2011.

[21] 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. 2016.

[22] A. L. Alter and K. M. Wang, "An Exploration of Computer Vision Techniques for Bird Species Classification," 2017.

[23] N. Horning, "Random Forests : An algorithm for image classification and generation of continuous field data sets."