How to Cite:
Sitepu, A. C., Liu, C.-M., Sigiro, M., Panjaitan, J., & Copa, V. (2022). A convolutional neural network bird's classification using north American bird images. *International Journal of Health Sciences*, *6*(S2), 15067–15080.
https://doi.org/10.53730/ijhs.v6nS2.8988

A convolutional neural network bird's classification using north American bird images

Ade Clinton Sitepu
International Electrical Engineering and Computer Science National Taipei University of Technology
Corresponding author email: t110999406@ntut.org.tw

Chuan-Ming Liu
Department of Computer Science and Information Engineering National Taipei University of Technology
Email: cmliu@ntut.edu.tw

Mula Sigiro
Department of Physics Education University of HKBP Nommensen
Email: mulasigiro@gmail.com

Joel Panjaitan
Department of Electrical Engineering Academy of Deli Serdang Engineering
Email: joel.pandjaitan@gmail.com

Venkatesh Copa
International Electrical Engineering and Computer Science National Taipei University of Technology
Email: t110998406@ntut.org.tw

***Abstract*---**In general, image classification is more like to classify objects with large categories, where these objects have a low level of similarity that is relatively rare. Birds image classification is a tough image dataset annotated with many bird species. It may be a challenging issue as numerous of the species of birds have degree of visual closeness. Bird species recognition can be challenging for people, let alone computer vision calculations. To analyze the image, this paper resizes the image into 224×224 pixels. We use the Convolutional Neural network (CNN) approach and add the structure of MobileNetV2, EfficientNetB0, EfficientNetB3 and the weight of the network that has previously trained using ImageNet. This paper attempts to analyze the comparative results from using MobileNetV2, EfficientNetB0 and EfficientNetB3 architecture. The F1 weighted average score MobileNetV2 is 75%, EfficientNet B0 is 81% and

Manuscript submitted: 27 March 2022, Manuscript revised: 18 May 2022, Accepted for publication: 9 June 2022

EfficientNet B3 is 81%. This result shows that some models are completely unable to predict some classes of birds, but other models can recognize them. This can happen because of differences of the operation in each CNN architecture parameter. These models complement each other.

***Keywords*---**convolutional neural network, EfficientNetB0, EfficientNetB3, MobileNetV2.

Introduction

Beside the advancement of data innovation in this digital era, making information accessible and easy to use. The ease of use of this information makes the data grow very rapidly. With so many images uploaded every day, better models and algorithms for image classification can be made by utilizing these images [1]. The computer receives input as an image which will then be processed, identified and given output in the form of classification results. The image classification is a relatively simple task for humans [2], unlike the case for computers. Image classification is one of the challenging problems in Computer Vision which has a complex calculation to get the solution. Machine learning-based methods are commonly used techniques to solve these problems. One of the most widely used methods for image data is the Convolutional Neural Network, also known as ConvNet or CNN [3]. CNN is the advancement from MLP (multi-layer perceptron) as one of the Deep Learning Algorithm. The CNN application has the most significant results in computer vision, this is often since CNN tries to imitate the image recognition framework within the human visual cortex, so that it has the capacity to handle image data [4].

In general, image classification that has been carried out is more likely to classify objects with larger categories, such as the classification of mammals, birds and so on which have a low level of similarity in physical form. Classification of images with objects that have similar physical forms is relatively rare. Research by Achmad Yusuf, Randy Cahya and Randy Chaya Wilhandika in 2019 regarding the Classification of Emotions Based on Facial Characteristics Using the Convolutional Neural Network. This paper discusses how to build an optimal CNN model to be able to classify emotions, the determination of the parameter values on the network is very influential in the training process with learning rate 0.000850981267220764 to reduce the cost value in the model very significantly and get an accuracy with 80.75% for face emotions classification emotions [5]. Isna Wulandari, Hasbi Yasin and Tatik Widiharih (2020) on the Classification of Digital Imagery of Spices and Spices Using the CNN and from their paper discusses building a CNN model to classify the image of spices and herbs, which results in an accuracy of 88.89% [6]. Research by Sarirotul Divineyah and Agung Nilogiri in 2018, they try to identify the plant types based on the leaf image and also using deep learning approaches. This paper builds a CNN model to identify plant genera with an accuracy value of 90.8% [7].

The dataset Caltech-UCSD Birds 200-2011 also known as CUB-200-2011 is a dataset pointed at subordinate category classification. Dataset CUB-200-2011 has 200 species birds (mostly North American) images with 11,788 annotated images [8]. The huge number of classes ought to make it a curiously dataset for subordinate categorization. As we mention, the differentiate precisely classifying more than a modest bunch of birds is something only a small proportion of people can do without get to the field directly. In addition, since few people do well on subordinate categorization case, it is arguably a region where a computer vision system would be valuable indeed in case it was not idealize.

The dataset tries to facilitate research on applications where image recognition that can help people to classify objects that are obscure to them. For example, if an accurate bird classifier were created, a client might submit a photo of a recently spotted bird to query an information database, such as Wikipedia [9]. Such classifiers may moreover offer assistance to computerize other areas of science. Based on the background that has been initiated previously, this research focuses on building a Convolutional Neural Network model on objects with many classes. The dataset used as the object in this paper is poultry, especially North American birds, which have similar physical shapes and significant differences are only found in motifs and colors.

Materials and Methods

Materials and General Method

This research begins by identifying a problem that is being experienced by the researcher and the formulation of the problem will be investigated. The problem that has been identified is how to make a model to classify objects that tend to have a high similarity. To solve this problem, it was formulated to create a model to classify 200 North American bird species using the CNN method. Next, a literature study phase related to image classification and CNN methods was carried out which was obtained in articles, research journals and papers. Then, downloading data in the form of digital images is done through the Kaggle website according to the object of research.

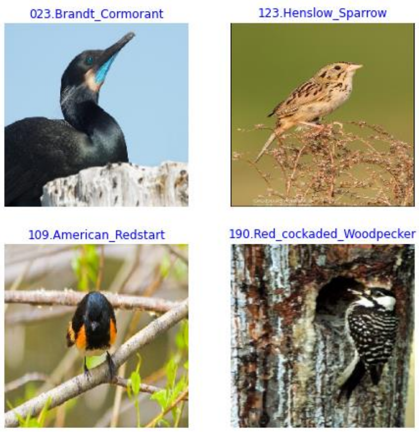


Fig. 1. Samples of Several North American Birds Species

Fig. 2. Rotating

Next, the pre-processing stage was carried out on the data that had been collected by rotating the image so that the amount of training data for each class was balanced to be used for the next stage, namely the design and selection of the model. At this stage, the CNN model design process and the selection of the best model are carried out. After that, an evaluation is carried out to test the accuracy of the model that has been formed using the Confusion Matrix. **Error! Reference source not found.** is a flowchart of the research stages.

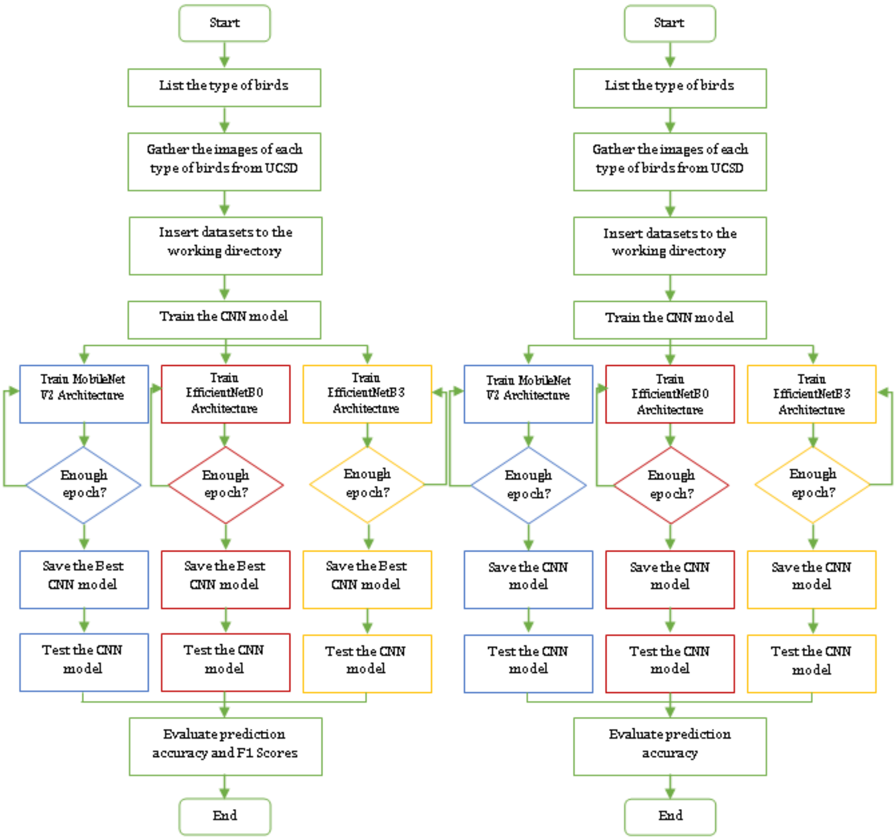


Fig. 3. General Research Flowchart

The research tools and materials used for this research can be seen in **Error! Reference source not found.**.

Table I. Research Tools and Materials

No.	Tools	Notes
1	Computer	Intel Core i7-7700 CPU @ 3.60GHz with Memory 32768MB RAM
2	Windows 10	Operation System
3	Visual Code Studio	Package Management
4	Python	Program Language
5	Tensorflow Keras	High-level application programming interface for Machine Learning.
6	CNN Architecture	- MobileNetV2 - EfficientNetB0 - EfficientNetB3

Dataset

The data used in this paper is a color image that downloaded through the Caltech website with the main object being the image of the North American Bird which is used to classify 200 species of birds. Fig. 1 is a sample of data from images of North American bird species that have been taken and will be processed as a dataset for use in this paper. The dataset will be divided and split into 80% for training data, 10% for validation data and 10% for testing data [10]. There are 11,788 total images data before pre-processing (training data = 9414, validation data = 1139 and testing data = 1235).

Preprocessing

Rotating

At this stage each starling image data will be rotated. We want to create balance dataset in each class for training data, so it will become a balanced data analysis. Rotating the image is also an effort to reduce the model to be overfitting [11]. The result of a rotated image can be seen in Fig. 2. After rotating, the number of training data = 12000 images.

Resizing

At this stage, each starling image data will be resized in pixels with the same length and width, which is 224×224 pixels. So, each data has the same size evenly. An example of an image that has been resized can be seen in Fig. 3.


Before Resize 309×500 pixels


After Resize 224× 224 pixels

Fig. 3. Resizing

Normalization

The dataset that has been divided, then the data will be normalized using a hard library, ImageDataGenerator, besides that it also gives a label to each data later. This stage makes all pixel values normalized to 0 to 1, with the aim of getting better learning results [12].

CNN Model Architecture (Keras Application)

Keras Applications are deep learning models that are made available alongside pre-trained weights. These models can be used for prediction, feature extraction, and fine-tuning. The size and number of parameters, and Time (ms) per inference step (GPU) refers to the model MobileNetV2 [13], EfficientNetB0 and EfficientNetB3 [14] performance on the ImageNet validation dataset in Table II.

Table II. CNN Model Architecture

No.	Tools	Size (MB)	Parameters	Time (ms) per inference step (GPU)
1	MobileNet V2	14	3,538,984	3.83
2	EfficientNetB0	29	5,330,571	4.91
3	EfficientNetB3	48	12,320,535	8.77

Time per inference step based on the average of 30 batches and 10 repetitions using AMD EPYC Processor (with IBPB), RAM 1.7 TB, Tesla A100 GPU with Batch size is equal to 32 [15].

Evaluation Technique

To measure the performance of the CNN model in classifying bird images, it will calculate how many bird species were correctly predicted from the overall image data to be tested [16]. There are several parameters used that refer to the process of measuring the performance of a model using the confusion matrix method [17].

True Positive

True Positive (TP) is a condition where the CNN model predicts that the information is within the positive class and the real information is also in the positive class [18].

True Negative

True Negative (TN) is a condition where the CNN model predicts that the information is within the negative class, but the real information is in the positive class [19].

False Positive

False Positive (FP) is a situation where the CNN model predicts that the information is within the positive class, but the real information is in the negative class .

False Negative

False Negative (FN) is a situation where the CNN model predicts that the information is within the negative class, but the real information is in the positive class [20].

Accuracy

Accuracy is calculated as the proportion of the number of correct classifications to the whole number of classifications [12]. The correct classification is True Positive for each type or class. The equation to calculate the accuracy can be seen in (1).

$$\text{accuracy} = \frac{\text{TP}}{\text{Total number of dataset}} \quad (1)$$

Precision

Precision is an accuracy value that is relative to a certain class prediction. The equation to calculate the precision can be seen in (2).

$$\text{precission} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

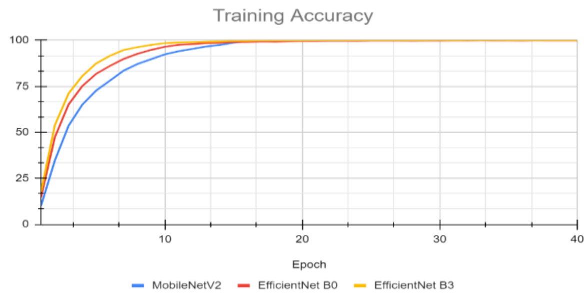
Recall

Recall describes the success of the model in retrieving an information. Thus, recall is the proportion prediction of true positive compared to the whole true positive data. The equation to calculate the recall can be seen in (3).

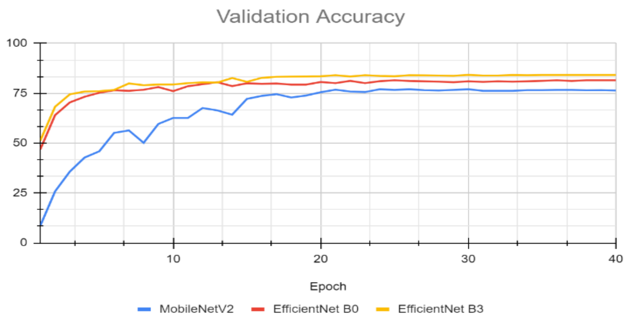
$$\text{recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

Result and Discussion

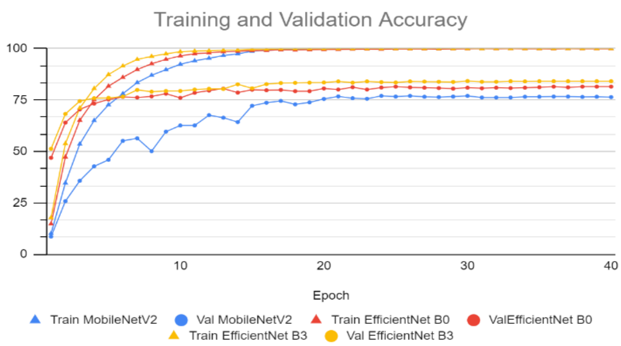
Prior to the classification stage, the Convolutional Neural Network or CNN model must first be obtained by training the datasets that have been created previously. The training process aims to be a learning material for the model so that the model can perform a better classification. The dataset has been normalized, then the dataset will be used to carry out the model design process. The training data will be used as input in the first convolution layer. In the Convolutional Neural Network training process, the Keras Application is used.



(a)



(b)

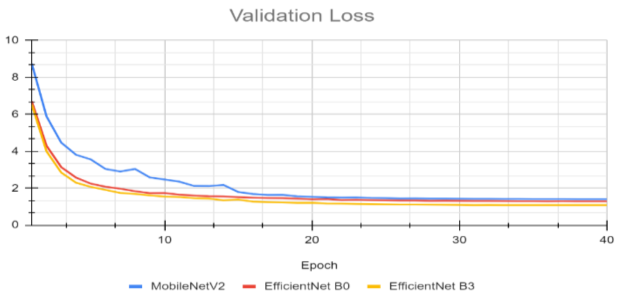


(c)

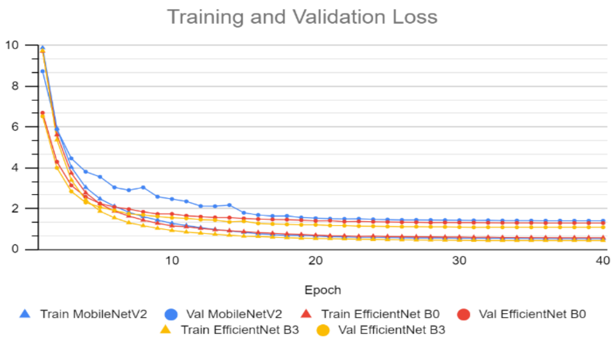
Fig. 4. Accuracy Comparison of MobileNetV2, EfficientNet B0, EfficientNet B3 to classify the types of birds: (a) training dataset; (b) validation dataset; (c) training and validation

Fig. 4 shows a comparison of the learning accuracy of training data and validation data for each epoch for each CNN architectural model. Fig. 5 shows the comparison of learning loss training data and validation data for each epoch for each CNN architecture model. The model that has been trained is then tested on the test data. The results can be seen in Table III. To measure the accuracy of the CNN model in classifying the image of 200 classes of birds, it will be calculated how many species of starlings were correctly predicted from the overall image data to be tested. Furthermore, the CNN model performance test was carried out to classify the bird species category by evaluating and predicting using testing data to determine its accuracy.

(a)



(b)



(c)

Fig. 5. Loss Comparison of MobileNetV2, EfficientNet B0, EfficientNet B3 to classify the types of birds: (a) training dataset; (b) validation dataset; (c) training and validation

Table III. The Accuracy of Validation Data and Elapsed Time of Training Data

No.	CNN Architecture	Accuracy	Training Elapsed Time (s)
1	MobileNet V2	76.734	44,875.65
2	EfficientNetB0	80.421	59,865.49
3	EfficientNetB3	80.948	99,831.89

The process of evaluating the accuracy of the CNN model will be described in the form of a recall and precision comparison as can be seen in Fig. 7 and Fig. 8.

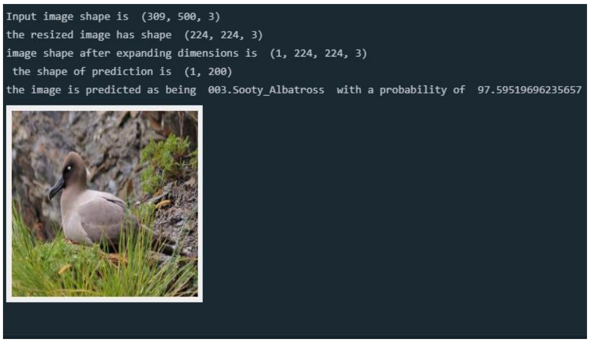
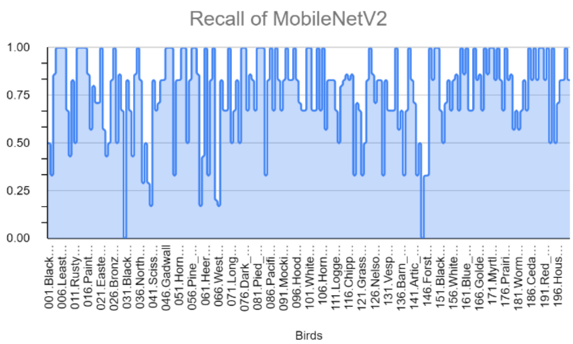
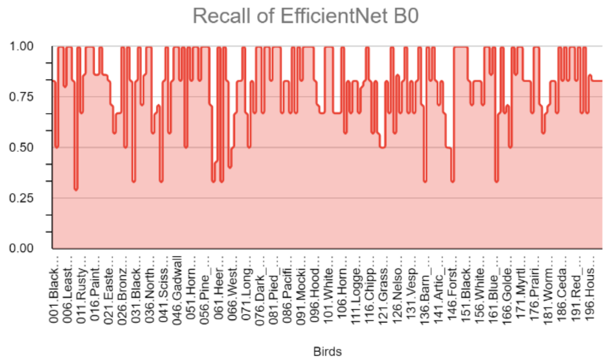


Fig. 6. Predict the image with probability

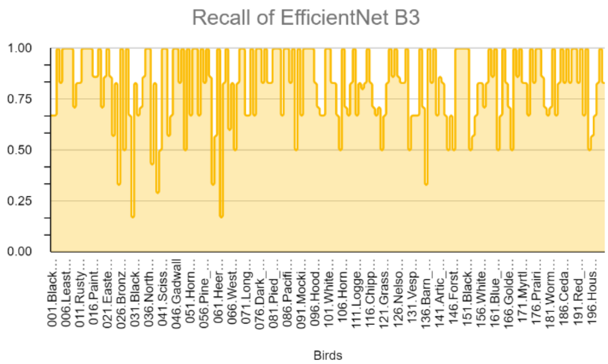
Fig. 7 and Fig. 8 show that some models are completely unable to predict some classes of birds, but other models can recognize them. This can happen because of differences in the operation of each CNN architecture parameter used. These models complement each other.



(a)



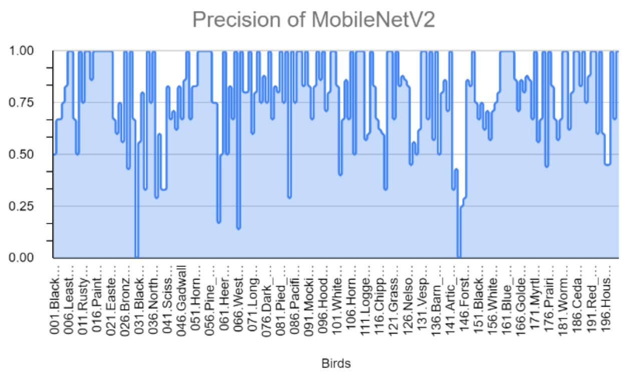
(b)



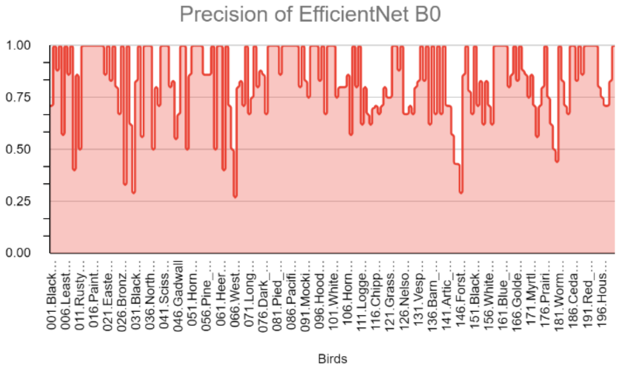
(c)

Fig. 7. Recall comparison each model for every testing data: (a) MobileNetV2; (b) EfficientNet B0; (c) EfficientNet B3

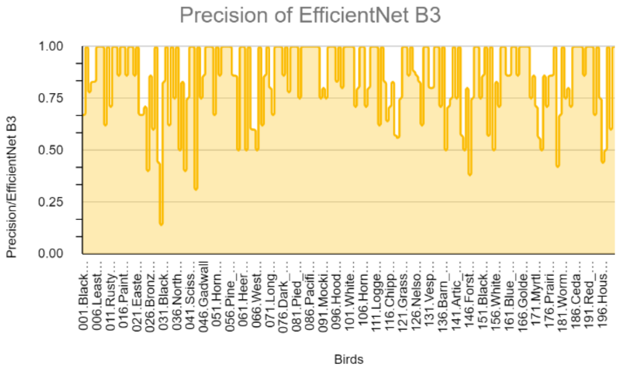
A class in the testing data has Recall = 0 means that the model is completely unable to predict the class. This model is also able to predict the probability that the bird is correct according to its class. Can be shown in Fig. 6.



(a)



(b)



(c)

Fig. 8. Precision comparison each model for every testing data: (a) MobileNetV2; (b) EfficientNet B0; (c) EfficientNet B3

In common, in case we are working with an imbalanced dataset where all classes are similarly vital, the value of the macro average would be a good choice because it treats all classes similarly. It implies that in this paper for bird's image classification, we would use the F1 macro score. To relegate more great contribution to classes with more cases within th dataset in imbalanced dataset, the weighted average is preferred. Since in weighted averaging, the contribution of each class F1 average is weighted by its size. The F1 weighted average score for each model: MobileNetV2 is 75%, EfficientNet B0 is 81% and EfficientNet B3 is 81%.

Conclusion

Based on the papers that have been carried out, it is concluded that the Convolutional Neural Network method can classify objects that have high altitudes. In this study, the learning rate value was 0.0001 in the training process, the comparison of the three models obtained accuracy in the validation data of = MobileNetV2 : EfficientNetB0 : EfficientNetB3 = 76.383% : 81.475% : 84.021%. Data augmentation can also improve the accuracy of the CNN model. Comparison of the accuracy of the model to classify images of 200 bird species on the test data are MobileNetV2 : EfficientNetB0 : EfficientNetB3 = 75.30% : 80.73% : 81.30%. The F1 weighted average score MobileNetV2 is 75%, EfficientNet B0 is 81% and EfficientNet B3 is 81%.

For advanced works, it is possible to use images with different objects as datasets in the development of research related to classification using Convolutional Neural Networks or other CNN Architecture that might help other computer vision problems. We need big dataset if we are working on the large feature. Data accuracy affects the background of the image. For further analysis, the edge analysis is carried out first and then the image cropping augmentation is performed. This is expected to give better results. We are interest too about how to optimize the model by use Transfer Learning approaches or using Yolo Algorithm for better idea.

References

1. I. Kandel, M. Castelli, and A. Popovič, “Comparative study of first order optimizers for image classification using convolutional neural networks on histopathology images,” *J. imaging*, vol. 6, no. 9, p. 92, 2020.
2. F. Chollet, *Deep learning with Python*. Simon and Schuster, 2021.
3. S. Ibrahim, N. Hasan, N. Sabri, K. A. F. Abu Samah, and M. Rahimi Rusland, “Palm leaf nutrient deficiency detection using convolutional neural network (CNN),” *Int. J. Nonlinear Anal. Appl.*, vol. 13, no. 1, pp. 1949–1956, 2022.
4. T. Purwaningsih, I. A. Anjani, and P. B. Utami, “Convolutional Neural Networks Implementation for Chili Classification,” in *2018 International Symposium on Advanced Intelligent Informatics (SAIN)*, 2018, pp. 190–194.
5. E. Christianto, A. M. Sambul, and F. D. Kambey, “Implementation of Convolutional Neural Network on Images for Starlings Classification,” 2021.
6. I. Wulandari, H. Yasin, and T. Widiharih, “Klasifikasi CItra Digital Bumbu dan Rempah dengan Algoritma Convolutional Neural Network,” *J. Gaussian*, vol. 9, no. 3, pp. 273–282, 2020.
7. S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” *Front. Plant Sci.*, vol. 7, p. 1419, 2016.
8. C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, “The caltech-ucsd birds-200-2011 dataset,” 2011.
9. P. Welinder *et al.*, “Caltech-UCSD birds 200,” 2010.
10. F. A. Chalik and W. F. Al Maki, “Classification of Dried Clove Flower Quality using Convolutional Neural Network,” in *2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA)*, Nov. 2021, pp. 40–45. doi: 10.1109/DATABIA53375.2021.9650199.
11. C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *J. Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
12. A. C. Sitepu and M. Sigiro, “Analisis Fungsi Aktivasi Relu Dan Sigmoid Menggunakan Optimizer Sgd Dengan Representasi Mse Pada Model Backpropagation,” *JUTISAL J. Tek. Inform. Univers.*, vol. 1, no. 1, pp. 12–25, 2021.
13. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
14. M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional

neural networks," in *International conference on machine learning*, 2019, pp. 6105–6114.
15. F. Chollet, "Keras documentation," *keras. io*, vol. 33, 2015.
16. S. C. Satapathy, V. Bhateja, K. S. Raju, and B. Janakiramaiah, Eds., *Data Engineering and Intelligent Computing*, vol. 542. Singapore: Springer Singapore, 2018. doi: 10.1007/978-981-10-3223-3.
17. T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS One*, vol. 10, no. 3, p. e0118432, 2015.
18. S. Park, S. Kim, and Y. Ha, "Highway traffic accident prediction using VDS big data analysis," *J. Supercomput.*, vol. 72, no. 7, pp. 2815–2831, Jul. 2016, doi: 10.1007/s11227-016-1624-z.
19. N. Salmi and Z. Rustam, "Naïve Bayes Classifier Models for Predicting the Colon Cancer," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 546, no. 5, p. 52068, 2019, doi: 10.1088/1757-899x/546/5/052068.
20. M. Chandra Wijaya, "Automatic Short Answer Grading System in Indonesian Language Using BERT Machine Learning", [Online]. Available: https://doi.org/10.18280/ria.350609
21. Widana, I.K., Sumetri, N.W., Sutapa, I.K., Suryasa, W. (2021). Anthropometric measures for better cardiovascular and musculoskeletal health. *Computer Applications in Engineering Education*, *29*(3), 550–561. https://doi.org/10.1002/cae.22202