Advances in Social Science, Education and Humanities Research, volume 647
International Conference on Applied Science and Technology on Social Science 2021 (iCAST-SS 2021)

Cat Breeds Classification Using Compound Model Scaling Convolutional Neural Networks.

*Tita Karlita
Informatics Engineering
Electronic Engineering Polythechnic Institute of Surabaya
Surabaya, Indonesia
tita@pens.ac.id

Rengga Asmara
Informatics Engineering
Electronic Engineering Polythechnic Institute of Surabaya
Surabaya, Indonesia
rengga@pens.ac.id

Nadia Azahro Choirunisa
Informatics Engineering
Electronic Engineering Polythechnic Institute of Surabaya
Surabaya, Indonesia
nadiaazahrochoirunisa@gmail.com

Fitri Setyorini
Informatics Engineering
Electronic Engineering Polythechnic Institute of Surabaya
Surabaya, Indonesia
fitri@pens.ac.id

***Abstract*— Cats are one of the most popular animals in the world. Many cat breeds in the world are only about 1%. Therefore, most are dominated by mixed cats or domestic cats. Nevertheless, there are so many different types of cat breeds in the world that it is sometimes difficult to identify them. Therefore, we need a system that can recognize and classify the types of cat breeds automatically. In this study, we used one of the deep learning methods that can recognize and classify an object, a Convolutional Neural Networks (CNN). The EfficientNet-B0 architecture was used as a model to extract image features automatically. The collection of nine different cat breeds containing 2700 images was used as a working dataset fed into the EfficientNet-B0 architecture. Based on the experiments, the system succeeds in classify cat breeds images, and the best model has achieved classification accuracy of 95%.**

Keywords—Deep Learning; Convolutional Neural Networks; Cat Breeds; EfficientNet-B0.

I. INTRODUCTION

Cat (Felis catus) is a type of carnivorous mammal from the Felidae family. In the past 6000 years BC, cats are known to have mingled with humans and spread in various parts of the world. The cat who is usually referred to is a cat that has been tamed. The number of cat breeds worldwide is only about 1%, so most are dominated by mixed cats or domestic cats. This minimal number of cat breeds makes the price of purebred cats much more expensive. Every cat breed has special characteristics, but because of the large number of interbreeding, the determination of cat breeds becomes more difficult [1][2].

An attractive appearance with various types makes cats become one of the most popular pets in the world. Many people have cats because of the various advantages like being able to repel mice, practical because it is not noisy, feeds a little, and does not require such a large room. In addition, keeping a cat can also positively impact the owner, one of which is to reduce stress.

The improvement of the artificial intelligence era to recognize images is growing very rapidly. Often, the collection of images used at some stage in the category system is imperfect, as the numerous distractions, the shape of shadows, low contrast, and blurred images. So, we need a method that is able to process these images. One approach that might be used is the deep learning technique that may apprehend and hit upon an item in a virtual photograph [3]. One technique of deep learning that may recognize and classify an image's category is Convolutional Neural Networks (CNN).

Convolutional neural networks (CNNs) are a type of deep learning neural network. CNN represents a major breakthrough in image recognition. They are most often used to analyze visual images and often

work behind the scenes in image classification. CNN has an excellent technique in phrases of item detection and item recognition [3]. The CNN methods have been similar to those of neurons in the human brain, equipped with the weight, bias, and activation functions. In the CNN, several convolutional layers are added to obtain image features automatically. Nevertheless, the CNN method has a weak point: it needs tremendous computational time to accomplish its task.

Some studies about animal classification have been conducted by researchers using deep learning as one of the latest technologies in computer science [4][5][6]. Zhang et al. present an Android application used to predict the location and breed of a given cat using a mobile phone camera [4]. The SSD Mobilenet_v1 FPN was used as the model to classify the images. The app has been trained to recognize 14 types of cats with an average accuracy of the finalized model of 81.74%. Images of dogs and cats were classified using CNN by Mahardi et al [5]. They tried to build an image classifier to recognize various breeds of dogs and cats using retrained VGG models. Two standard models, VGG16 and VGG19, were used to construct the classifier. Both models have a training accuracy of around 98%. The resulting model from VGG16 has a validation accuracy of 98.56% and a testing accuracy of 83.68%. The model from VGG19 has a validation accuracy of 98.56% and a testing accuracy of 84.07%. Lee et al. compared and analyzed the classification performance from different machine learning and deep learning [6]. They implemented support vector machine and convolutional neural networks to solve the classical Cats vs Dogs problem, and compared how different parameters affect CNN.

In this paper, we present a system to classify cat breeds automatically. The EfficientNet-B0 architecture is utilized as a model that was used to detect and extract image features. Then a fully connected neuron that have some layers was used to label the images as one category of nine cat categories. The rest of the paper is organized as follows. Section II describes the proposed methods, including the working image dataset. It also describes CNN. Section III discusses the experiments and results. Finally, section IV concludes the work with some future directions.

II. Literature Reviews

In this study, an experimental research method was carried out with several stages.

2.1 Dataset

In this study, collection of cat images was collected from Kaggle [13], which originally had 67 different categories of cat breeds. This images has been collected by the usage of petpy, a PetFinder API wrapper [14]. We selected eight-category of cat breeds from the original dataset. Moreover, we also collected cat images by directly capture image using a smartphone for the local cat (Kucing Lokal) type. After combining all of the images, there are nine types of cats in this dataset, including Abyssinian, Bombay, British Shorthair, Kucing Kampung, Persian, Ragdoll, and Savannah Sphynx, and Turkish Angora. Before the dataset is used, we grouped them in the corresponding folder.

2.2 Proposed Method

Convolutional Neural Networks (CNN) is the development of Multilayer Perceptron (MLP), designed to process two-dimensional data. CNN is included in the type of deep neural networks because of its high network depth and is widely used in image data. MLP is unsuitable for image classification because it does not store spatial information from image data and assumes that each pixel is an independent feature, so the results are poor [7] [8]. CNN has two parts. The first part is feature learning. It contains some layers and some convolutional layers. The second part is the learning stage using the backpropagation method. Overall, CNN consists of several hidden layers, namely the convolution layer, activation function (ReLU), pooling, and fully connected layers, which have some neurons. The illustration of the CNN architecture is shown in Figure 1.

1) EfficientNet

EfficientNet is one type of the many CNN architectures developed by the Google Brain team, Mingxing Tan, and Quoc V. Le. They used the idea of model scaling, which is about scaling the existing model in terms of model depth, model width, and less popular input image resolution to improve the model's performance and scaling all three of them together to deliver better results. This method is called compound scaling [9][10]. The smallest version of EfficientNet is EfficientNet-B0, that is mobile sized architecture having 11M trainable parameters. The B0 architecture is shown in Table I.

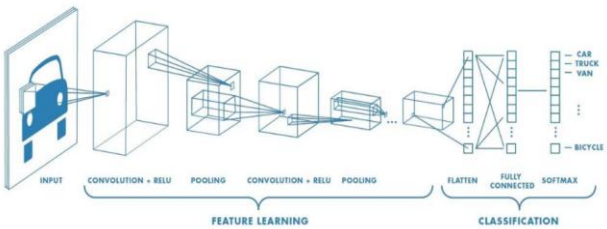


Fig. 1. Convolutional Neural Networks Architecture

The seven MBConv blocks are inverted residual blocks that were introduced in MobileNetV2 architecture, with different settings. These blocks also use squeeze and excitation block along with swish activation. An equal weightage was given to each of

feature map channels produced by a convolutional layer. Squeeze and excitation (SE) block is a method to give weightage to each channel instead of treating them all equally. Swish is a multiplication of a linear and a sigmoid activation. They apply the compound scaling method to scale it up from this baseline and obtain EfficientNet-B1 to B7 [11]. Inverted residual blocks uses depthwise separable convolution inside the residual block which uses depthwise convolution first and then pointwise convolution. This approach decreases a large number trainable parameters.

TABLE I. TABLE I. EFFICIENTNETB0 BASELINE NETWORKS

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	28×28	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

2) *Architecture Approach*

Considering the result of the research above, we decide to use EfficientNet-B0 as the base model of this study. The proposed architecture approach can be seen in the Figure 2.

The input data is 224x224x3 with three as an RGB channel (because the input data is an image). Next EfficientNet-B0 architecture is used for the training process and given some extra layers. The first is pooling, whose function is to reduce the number of parameters. Next, dropout is done as much as 0.7 to reduce the possibility of overfitting [12]. Overfitting occurs when the model is too focused on specific training datasets not to generalize other similar datasets. This can be characterized by a much higher accuracy value than the validation accuracy value in the training process. Continued dense layer using ReLU activation function followed dropout. This is done twice with the different values of each. And the last is the dense layer using the ReLU activation function, followed by the dense with softmax activation function as an output layer. Then the output will be obtained in the form of the classification results of nine types of cat breeds.

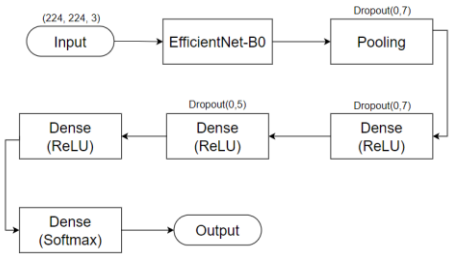


Fig. 2. The proposed CNN architecture design.

III. EXPERIMENTAL RESULT

3.1 *Data Preparation*

There are nine types of cat breeds used in this study. This image data is then called a dataset. The dataset collected was 2700 images for nine types of cat breeds. The dataset was divided into 80% for training and 20% for validation. Before the dataset is used in the training process, the dataset undergoes an augmentation process including random flip, random zoom, random rotation, increasing brightness, and random shift range. Some examples of images from the dataset are shown in Figure 3.

Fig. 3. Image example of the dataset.

3.2 *Training Result*

The input data is 224x224x3 with three as an RGB channel. Dataset training is an early stage that aims to process available datasets. In this training process, image data input will go through a training process using the Convolutional Neural Networks that will form a model which will be tested for performance (in this case, the base model used is EfficientNet-B0).

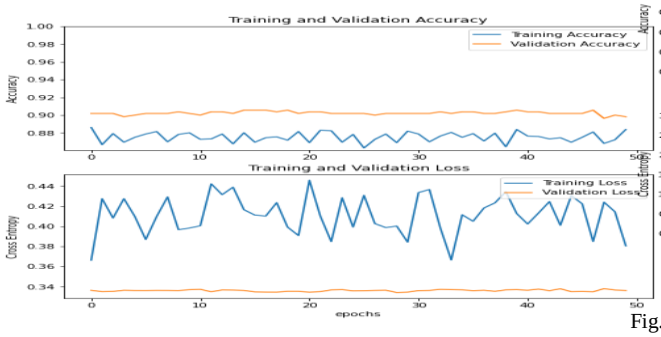
The training process is done with three scenarios; The first scenario compares the model with pre-processing data and without pre-processing data using Gaussian Blur from scikit-image. The next step is to compare the model with the Adam optimizer and RMSprop with a learning rate (lr) of 0.001. The third is the same as before, using the Adam optimizer and RMSprop, but with a different learning rate, 0.0001.

1) *Scenario 1*

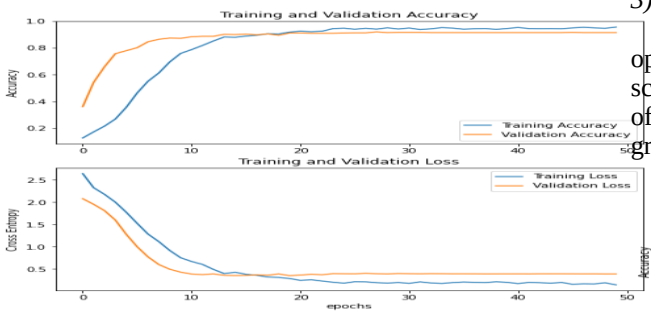
This test is carried out by comparing the model using data pre-processing with models without using data pre-processing. The pre-processing used in this experiment is the Gaussian Blur filter from the scikit-image module. The results can be seen in the training graph in Figure 4.

From Figure 5, it can be seen that the model using pre-processing data has a better accuracy value than the model without pre-processing data. The accuracy value of the model without pre-processing data has fluctuated around 88%. At the same time, the model with pre-processing data showed a higher value, namely 95%. This indicates that the model's

performance using pre-processed data is better than the model without pre-processing data.



(a)



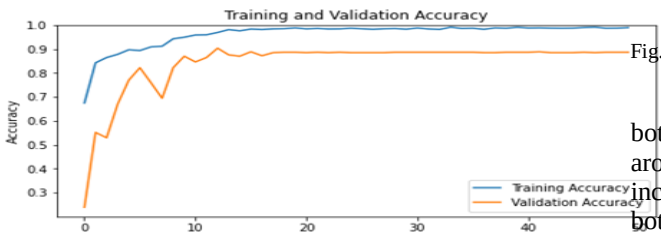
(b)

Fig. 4. Training graph of scenario 1, (a) without pre-processing, (b) with pre-processing

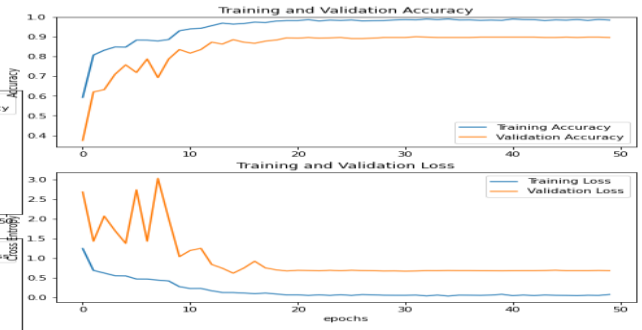
2) Scenario 2

This test was conducted by comparing the Adam and RMSprop optimizers with a learning rate of 0.001 and an epoch of 50 epochs. The results can be seen in the training graph in Figure 5.

The graph above shows that the accuracy value in both experiments rose to near 1.0 at around the 15th epoch. In contrast, the validation value fluctuated up to around the 15th epoch. In the next epoch, both accuracy and validation values are constant. In the Adam optimizer experiment, the accuracy value obtained was 98%, and the validation value was 88%. While in the experiment using the RMSprop optimizer, the accuracy value obtained was 98%, and the validation value was 89%.



(a)

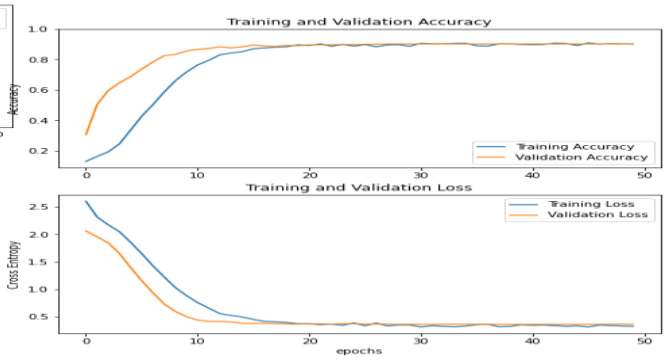


(b)

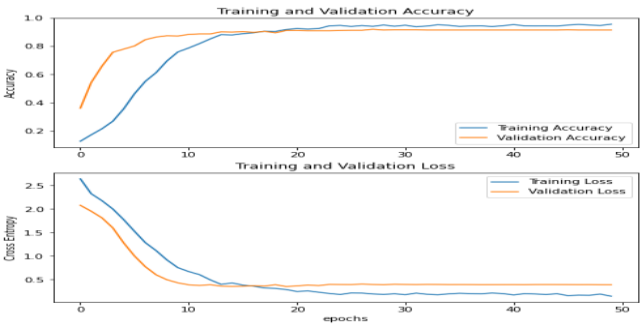
Fig. 5. Training graph of scenario 2, (a) Adam (b) RMSprop.

3) Scenario 3

This test was conducted by comparing the Adam optimizer and RMSprop, the same as the previous scenario but with a learning rate of 0.0001 and epochs of 50 epochs. The results can be seen in the training graph in Figure 6.



(a)



(b)

Fig. 6. Training graph of scenario 3, (a) Adam (b) RMSprop.

The graph above shows that the accuracy value in both experiments increased significantly to around 0.9 around the 15th epoch. While the validation value also increased around the 15th epoch. In the next epoch, both accuracy and validation values are constant. In the Adam optimizer experiment, the accuracy value obtained was 90%, and the validation value was 90%. While in the experiment using the RMSprop optimizer, the accuracy value obtained was 95%, and the validation value was 91%. Compared to scenario 1, the accuracy value in this scenario is slightly lower,

but there is no overfitting. Table II shows precision and recall for each scenario.

TABLE II. TABLE II. CLASSIFICATION METRICS COMPARISON TABLE

	Scenarios	Precision	Recall
Scenario 1	Without pre-processing	0.98	0.97
	With pre-processing	0.99	0.99
Scenario 2	Adam (lr=0,001)	0.99	0.98
	RMSprop (lr=0,001)	0.98	0.98
Scenario 3	Adam (lr=0,0001)	0.98	0.96
	RMSprop (lr=0,0001)	0.99	0.99

The best precision value is obtained by the model using the Adam optimizer (lr=0.001) and RMSprop (lr=0.0001) were using pre-processing, which means that the model is very good at recognizing each class. Meanwhile, the experiment conducted the largest recall value using the RMSprop optimizer (lr=0.0001) and using pre-processing, which was 0.99. A high recall value indicates the model is working well in recognizing true positives and reducing the number of false negatives. Table III is a summary to make it easier to compare the accuracy of the three scenarios above.

TABLE III. TABLE III. EXPERIMENT ACCURACY COMPARISON TABLE

Optimizer	Learning rate	Preprocessing	Accuracy	Validation accuracy
RMSprop	0.0001	No	88%	89%
Adam	0.001	Yes	98%	88%
RMSprop	0.001	Yes	98%	89%
Adam	0.0001	Yes	90%	90%
RMSprop	0.0001	Yes	95%	91%

Based on Table III, the experiment with the highest accuracy value uses the Adam optimizer and RMSprop with a learning rate of 0.001. However, these two experiments were overfitting, so they were still unable to classify images well. The experiment using the RMSprop optimizer with a learning rate of 0.0001 got an accuracy of 95%, lower than the previous experiment. However, the results of this experiment do not experience overfitting so that the model can classify images better. At the same time, the experiments with the lowest accuracy values were found in the two experiments without pre-processing data. So it can be concluded that pre-processing data is essential to improve the model's performance.

3.3 Confusion Matrix

Then testing is done with new test data with 180 images on the model with the best results using the RMSprop optimizer with lr = 0.0001 and pre-processing. The results of this process are presented in the confusion matrix as as shown in Figure 7. Based on the figure, generally, it can be seen that all classes are well grouped at their classes.

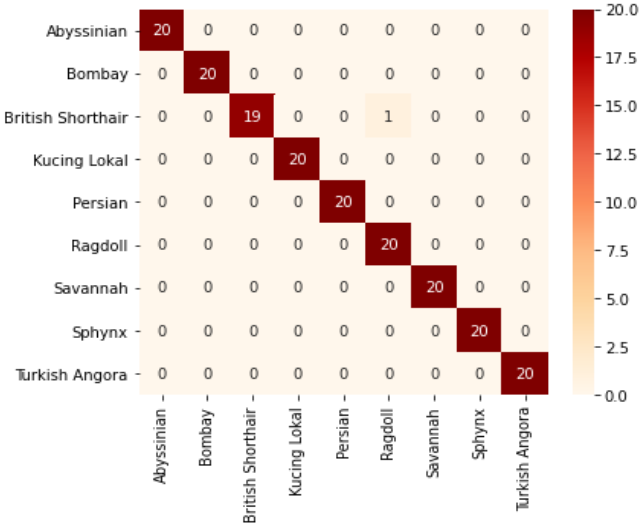


Fig. 7. Confusion matrix from the best model

From the confusion matrix, it can be concluded that in an experiment using the RMSprop optimizer, the number of images that were successfully predicted was 179, with the error only in the British Shorthair class.

IV. CONCLUSIONS

This paper focused on the use of deep learning for animal identification and classification. One of the deep learning methods, Convolutional Neural Networks (CNN), is used in this study using one of the models, namely EfficientNet-B0. This work was conducted with a dataset of 2700 images of 9 different cat breeds, where 80% of images are used for training, and 20% images are used for validation. After the training process was conducted with three scenarios, the most optimal result obtained the value of the accuracy of 95% with the validation accuracy of 91%. Using test data of 180 images, this model managed to classify 179 cats with the correct class.

For further research, we suggest improving the existing model. Besides that, add a class to the dataset for cat breeds whose characteristics are pretty similar and add variations such as coat color, pattern, or mixed breed. And the last, try to use another data pre-processing method to get better results. Also, we suggest adding more cat breeds images for the training stage and adding more cat breeds categories to the system.

ACKNOWLEDGMENT

Authors would like to thank to Electronic Engineering Polythechnic Institute of Surabaya that is support this research.

REFERENCES

[1] L. Mariani, “Aplikasi Pendeteksian Ras Kucing Dengan Mendeteksi Wajah Kucing Dengan Metode Viola Jones Berbasis Android,” Pelayanan Kesehat., no. 2015, pp. 3–13, 2020.

[2] F. Paola and R. Giancarlo, “Purebred dogs and cats: a proposal for a better protection,” J. Vet. Behav., Jul. 2021.

[3] I. N. Purnama and S. Primakara, “Herbal Plant Detection Based on Leaves Image Using Convolutional Neural Network With Mobile Net Architecture,” vol. 6, no. 1, pp. 27–32, 2020.

[4] X. Zhang, L. Yang, & R. Sinnott. "A Mobile Application for Cat Detection and Breed Recognition Based on Deep Learning", IEEE 1st International Workshop on Artificial Intelligence for Mobile (AI4Mobile), 2019.

[5] Mahardi, I.-H. Wang, K.-C. Lee, & S.-L. Chang, "Images Classification of Dogs and Cats using Fine-Tuned VGG Models", IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), 2020.

[6] Y. Lee, “Image Classification with Artificial Intelligence: Cats vs Dogs”, 2nd International Conference on Computing and Data Science (CDS), 2021.

[7] W. S. Eka Putra, “Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101,” J. Tek. ITS, vol. 5, no. 1, 2016.

[8] E. N. Arrofiqoh and H. Harintaka, “Implementasi Metode Convolutional Neural Network Untuk Klasifikasi Tanaman Pada Citra Resolusi Tinggi,” Geomatika, vol. 24, no. 2, p. 61, 2018.

[9] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” 36th Int. Conf. Mach. Learn. ICML 2019, vol. 2019-June, pp. 10691–10700, 2019.

[10] R. Widodo, T. Badriyah, and I. Syarif, “Detection of Lung Cancer Cell Based on Cytological Examination Using Convolutional Neural Network,” Klik - Kumpul. J. Ilmu Komput., vol. 7, no. 3, p. 296, 2020.

[11] S. Rizal, N. Ibrahim, N. K. C. Pratiwi, S. Saidah, and R. Y. N. Fu’adah, “Deep Learning untuk Klasifikasi Diabetic Retinopathy menggunakan Model EfficientNet,” TELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron., vol. 8, no. 3, p. 693, 2020.

[12] W. Anggraini, “Deep Learning Untuk Deteksi Wajah Yang Berhijab Menggunakan Algoritma Convolutional Neural Network (CNN) Dengan Tensorflow,” J. Phys. A Math. Theor., vol. 44, no. 8, pp. 1–9, 2011.

[13] Cat Breeds Dataset, Images for 67 different cat breeds as labeled by advertisers for adoption, Version 2, 2019, https://www.kaggle.com/ma7555/cat-breeds-dataset/metadata.

[14] A. Schlegel, petpy - Python Wrapper of the Petfinder API, 2017