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

Indonesian Traditional Cake Classification Using Convolutional Neural Networks

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

Bimo Prasetyo Afif
Informatics Engineering
Electronic Engineering
Polythechnic Institute of Surabaya
Surabaya, Indonesia
bimoprasetyoafif@it.student.pens.ac.id

Ira Prasetyaningrum
Informatics Engineering
Electronic Engineering
Polythechnic Institute of Surabaya
Surabaya, Indonesia
ira@pens.ac.id

***Abstract* — In Indonesia, there are many types of cakes that are categorized as traditional snacks. Refer to the Kamus Besar Bahasa Indonesia; snacks are defined as foods that are peddled or mean bites. Snacks are classified based on how they are made, and some are based on the taste of the snacks. Traditional snacks are a part of Nusantara culture that is mandatory for those born and live in Indonesia to preserve them. But in reality, many people tend to consume and know more about modern snacks than traditional ones. In fact, not many people have even tried traditional snacks or even made their own at home. This application was developed to help people distinguish and recognize the various kinds of cakes on the market. With convolutional neural networks technology in machine learning, people can use image classification presented through mobile applications with accuracy above 90% for Indonesian traditional cake recognition.**

Keywords—convolutional neural networks, transfer learning, mobilenetv2,Indonesian traditional cake.

I. INTRODUCTION

Food is a basic need that is very important for human life, both physiologically, psychologically, socially, and anthropologically. Food is always related to human efforts to maintain their survival and health on earth. Food can also be the identity of a nation or civilization. There are many varieties and types of food, from traditional to modern, from dry to wet types, some require a complicated process to make, and some can be considered fast food. From the Big Indonesian Dictionary (KBBI), food is defined as anything that can be eaten, such as snacks, side dishes, or cakes.

In Indonesia, there are many types of cakes that are included in traditional snacks. From the KBBI, snacks are defined as snacks that are peddled or mean bites. Snacks are classified based on the type of processing method, and some are based on the taste of the snacks. From the type of snacks, they are divided into 10 types, namely kue abuk, kue basah, kue dadar, kue kering, kue koci, kue ku, kue lapis, kue mangkok, kue sengkulun, and kue tart. Meanwhile, from the taste, some snacks taste sweet, salty, savory, spicy, and the rest can be a combination of flavors such as savory, salty, or sweet and sour. Based on the period, there are traditional snacks and modern snacks. The difference between these snacks is influenced by the region, culture, and available natural resources.

Traditional cakes are part of the culture of the archipelago that is obligatory for those of us who were born and live in Indonesia to continue to preserve them. But in reality, many people tend to consume and know more about the modern cake than traditional ones [1]. In fact, not many have even tried traditional cake or even made their own at home.

Several researchers proposed image textures [2][3]. Farinella et al. classify food images using their texture features [2]. A bag of visual words model (BoW) is employed to extract food image's texture features. Before feeding the images into the system, they were processed with a bank of rotation and scale-invariant filters, and then a small codebook of Textons was built for each food class. Finally, the learned class-based Textons were obtained for every single visual dictionary existing in the 61 classes of the Pittsburgh Fast-Food Image Dataset. The Support Vector Machine (SSVM) is used for the classification stage. Average accuracy of 67.9% was obtained for all food image classes. Food texture is also used by Nakamoto et al. to classify food [3]. The bulk of magnetoresistance elements and an inductor as sensing elements were extracted from images. Then, the food textures were clustered using principal component analysis. In the end, an SVM classified the foods. The

component scores analyzed by the principal component analysis were used as data inputs for the SVM. A score of 95% accuracy was obtained using this method.

Although the food texture profiles have been widely used as the features to classify food images [1][4][5][6]. Some researchers utilize convolutional neural networks to extract image features automatically. Since Indonesia is famous for its traditional food that is popular both domestically and abroad, Dian et al. classify several Indonesian traditional cakes using Convolutional Neural Networks [1]. There are kue dadar gulung, kastangel, klepon, lapis, lumpur, putri salju, risoles and serabi. They used 1676 images for data training and testing with an accuracy of 65%. David et al. made an effort to classify food images for further diet monitoring applications using Convolutional neural networks [4]. They extract image features automatically using convolutional layers for the task of food classification. The standard Food-101 dataset has been selected as the dataset for this works. They classified food images using convolutional neural networks. A 2D convolution layer that creates a convolution kernel convolved with the layer input to produce a tensor of outputs was utilized. There are multiple such layers, and the outputs are concatenated at parts to form the final tensor of outputs. The Max-Pooling function for the data was adopted, and the features extracted from this function are used to train the network. In their study, an accuracy of 86.97% for the classes of the FOOD-101 dataset is classified using the proposed implementation. Rajayogi et al. used the transfer learning method to classify the Indian food dataset of 20 classes with a total of 500 images [5]. The images were divided into training and validating. The models used were InccceptionV3, VGG16, VGG19, and ResNet. The experimentation showed that the Google InceptionV3 outperformed other models with an accuracy of 87.9% and loss rate of 0.59. Other models such as the VGG19 produced 78.9% of accuracy. The VGG16 model and the ResNet model were able to produce accuracy of 78.2% and 69.91% respectively. A prediction model for classifying Thai fast food images was developed by Hnoohom et al [6]. The model uses a deep learning process that has been trained on natural images taken from GoogLeNet dataset. It then was fine-tuned to be able to recognize Thai fast food images. The authors created the Thai Fast Food dataset, which contains 3,960 images. The dataset has eleven groups of food images comprised of omelet on rice, rice topped with stir-fried chicken and basil, barbecued red pork in sauce with rice, stewed pork leg on rice, Thai fried noodle, rice with curried chicken, steamed chicken with rice, shrimp-paste fried rice, fried noodle with pork in soy sauce and vegetables, wide rice noodles with vegetables and meat. The last group comprises dishes that are not included in the other ten groups listed but exist among Thai fast food. The classification average accuracy on separate experiments showed that Thai fast food classification accuracy at 88.33%.

In this paper, we made an effort to develop deep learning neural networks which use convolutional layers to estimate and extract features of Indonesian traditional cake images automatically. We used MobileNetV2 networks as a benchmark in extracting features and then followed with fully connected layers to classify Indonesian traditional cake images. The remaining of this paper is organized as follows. Section II presents the proposed methods, including the image dataset used in this research. It also describes the CNN model. Section III discusses the experimental results and the analysis. Finally, section IV concludes the work with some future directions.

II. MATERIALS AND METHODS

Convolutional Neural Networks (CNN) is a type of neural network commonly used in image data. CNN can be used to detect and recognize objects in an image. CNN is a technique inspired by the neural network that resides in the human body [11]. Broadly speaking, CNN utilizes the convolution process by moving a convolution kernel (filter) of a specific size to an image. The computer gets new representative information from the result of multiplying that part of the image with the filter used.

The architecture of CNN is divided into two major parts, the Feature Extraction Layer and the Fully-Connected Layer, which is Multi-Layer Perceptron. Feature Extraction is the process of encoding from an image into features in the form of numbers that represent the image. Fully-Connected is usually used in the Multi-layer Perceptron method and aims to process data so that it can be classified.

2.1 CNN Layer

In this CNN method, we can apply transfer learning techniques or methods. Transfer learning utilizes a deep neural network model that has been trained and then applies it to a large dataset for classification [10]. The trained model can be named as a base model. With this method, we do not need to create a neural network model with many layers and train for days to get optimal results because we can transfer the weight of the network's weight into our CNN model. Our proposed CNN architecture is shown in Figure 1. The MobileNetV2 is used to extract image features automatically. In the classification layer, we used Average Pooling, 0.5% Dropout, 1024 neurons with RELU, then 0.5% of Dropout again, followed with eight output using softmax.

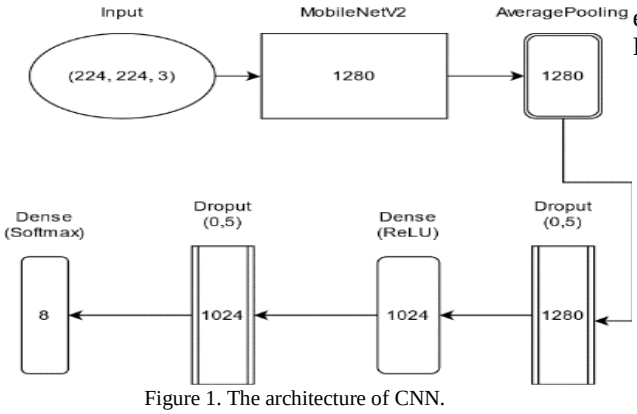

Figure 1. The architecture of CNN.

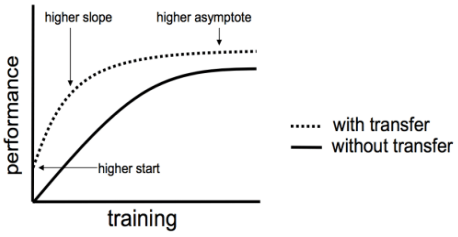

Figure 2 Difference Using Transfer Learning

In this paper, we propose a CNN layer that uses the base model from MobileNetV2. The transfer learning approach using MobileNetV2 is selected since it is well-known that it has better performance than without the transfer learning technique. The comparison of the two approaches can be seen in Figure 2. This base model is designed to have a small size and is very reliable in optimizing Neural networks in mobile applications [7].

The output of the base model is fed into the pooling layer in order to get the average value and convert it to flatten or vector. Then the dropout in the fully-connected layer is 0.5 to avoid overfitting [8]. There are two activation functions in the hidden layer, namely ReLU and Softmax, where a dropout of 0.5 is given between the two layers. The activation function is used to determine whether the neuron should be active or not based on its sum threshold.

The output that will be generated in the classification process will have one of the eight traditional cake labels, namely: kue_klepon, kue_mendut, kue_dadar_gulung, kue_clorot, kue_cenil, kue_lumpur, kue_layer, and kue_serabi, along with the accuracy values obtained.

2.2 Dataset

The traditional cake images dataset used in this research is a combination of snack images from self-collection and snack images from the Kaggle [14] and search engine results. Self-collection data is obtained from photographing traditional cakes from different sides circulating in the market directly. Some examples of the traditional cake images are shown in Figure 3.

Figure 3 Example of Cake's Picture

There are eight categories of cakes. Among them: Kue Klepon, Kue Lumpur, Kue Serabi, Kue Dadar Gulung, Kue Lapis, Kue Cenil, Kue Mendut dan Kue Clorot. The eight types of cakes were chosen because they are difficult to find in the market and cause unpopularity. The total number distribution of the cakes can be seen in Table I.

TABLE I. THE DISTRIBUTION OF INDONESIAN TRADITIONAL CAKES IN THE DATASET

Class Name	Number Of Images
kue_cenil	110
kue_clorot	82
kue_dadar_gulung	131
kue_klepon	189
kue_lapis	106
kue_lumpur	83
kue_mendut	53
kue_serabi	91

The total dataset of all classes is 845 images, where each class has a different number of images of traditional cakes. From the whole images in the dataset, we divided it into the training dataset and the validation dataset, where the ratio of 80% and 20%.

III. EXPERIMENTAL RESULT

3.1 Experimental Parameters

We use the accuracy and loss parameters from the training results to recognize its performance. We also use a confusion matrix in analyzing the optimization of the model in classifying the images. The model is good if the accuracy reaches above 90% with a loss below 0.2%. And can also detect at least 17 images in 20 images per class.

3.2 Dataset

In this experiment, we use a test dataset. This dataset is used in every trial outside the training process. The test dataset has 160 images divided into

eight classes, where each class has 20 different images. The collection of test datasets is done manually. This collection aims to ensure that no dataset has the same characteristics or characteristics as the dataset used during the training process.

3.3 Specification Requirements

The experiments run in the personal computer which is equipped with GPU. The complete specification of the hardware and software is shown in the Table II.

TABLE II. EXPERIMENTAL SPCIFICATION REQUIREMENTS

Hardware	
Motherboard	Gigabyte H110M Gaming3
CPU	Intel i3-6100 @ 3.70 GHz
Memory	HyperX Fury DDR4 4GB x 2
VGA	Gigabyte GTX 1050 OC 2GB
Storage	Seagate Barracuda SATA SSD 500GB
Software	
OS	Ubuntu 20.04.2 LTS
GPU Driver	NVIDIA 465.19.01
CUDA	11.3
Python	3.8.10
Tensorflow	2.5
Training Platform	Jupyter Lab 3.0.16

3.4 Results

1) Training Process

In this experiment, we conducted repeated training with the same CNN architecture and model. We use various variations of epochs and learning as differentiating values for each case. Epoch and learning rate are adjusted to the results of previous experiments where the model reaches the highest value in the training process. Figure 4 shows the illustration of the training and testing process.

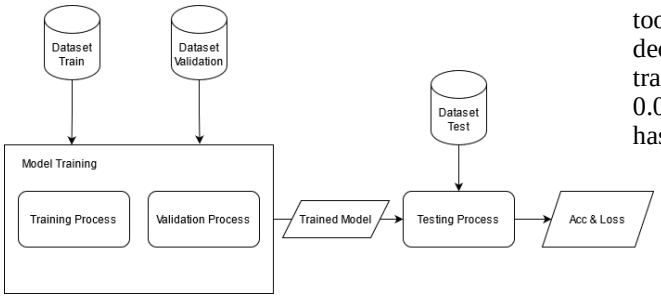

Figure 4. Training and Testing Process

The accuracy and loss values obtained in the model come from model validation on the test dataset after the training is complete. Therefore, there are two validations: validation during training using dataset validation and validation after training using dataset test.

The training and testing case results are shown in Table III. Each case has a different time according to the device's performance to carry out the training and validation process. From the results that are shown in Table III, four cases have good accuracy and loss values, namely numbers 3, 4, 5, and 6. Accuracy has a value above 0.90 and a loss value below 0.20.

It can be seen in case 1 that the model has an accuracy of more than 0.90 and a loss of less than 0.20. However, we try to add epochs and reduce the learning rate to find optimal results where the 100th epoch (4th case) with a learning rate of 0.00001. The results obtained in this experiment are the accuracy value of 0.96 and the loss value of 0.14. The best testing case results is shown in Table IV.

TABLE III. TRAINING AND TESTING CASE RESULT

No	Epoch	Optimizer	Learning Rate	Accuracy	Loss	Time
1	10	Adam	Default (0.001)	0.9250	0.2792	4m 48s
2	20	Adam	Default (0.001)	0.9375	0.3486	11m
3	10	Adam	0.0001	0.9375	0.1998	5m 12s
4	13	Adam	0.0001	0.9563	0.1758	7m 4s
5	20	Adam	0.0001	0.9500	0.1519	14m 0s
6	50	Adam	0.00001	0.9375	0.2359	26m 20s
7	100	Adam	0.00001	0.9625	0.1453	41m 37s

TABLE IV. BEST RESULT TESTING CASE

Case	Epoch	Learning Rate	Accuracy	Loss	Time
1	10	0.0001	0.9375	0.1998	5m 12s
2	13	0.0001	0.9563	0.1758	7m 4s
3	20	0.0001	0.9500	0.1519	14m 0s
4	100	0.00001	0.9625	0.1453	41m 37s

2) Training Accuracy And Loss Comparison

From the graph that is shown in Figure 5, case 1 to case 2, in epochs 1 to 2, the model has a significant gap in the train accuracy value of 0.70 to 0.50 on the validation accuracy line. This causes the model to be too complex to learn something and can result in decreased accuracy. In the 4th case, simultaneously train accuracy and validation accuracy score between 0.0 to 0.1 in the first epoch. This shows that the model has the capability to learn better at each epoch.

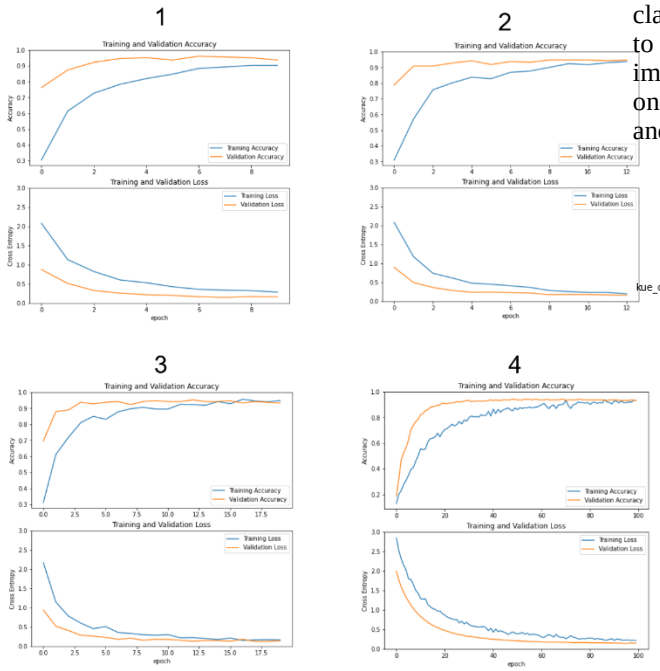

Figure 5. Accuracy and Loss Comparison

The training evaluation and validation results in the 4th case get an accuracy value of 0.96 (96%) and a loss of 0.15 (0.15%). Because the number of classification classes is more than 2, the accuracy value is obtained by calculating the average of the accuracy values contained in each class.

	precision	recall	f1-score	support
kue_cenil	1.00	0.80	0.89	20
kue_clorot	1.00	0.90	0.95	20
kue_dadar_gulung	0.91	1.00	0.95	20
kue_klepon	0.95	1.00	0.98	20
kue_lapis	0.95	1.00	0.98	20
kue_lumpur	1.00	1.00	1.00	20
kue_mendut	1.00	0.85	0.92	20
kue_serabi	0.87	1.00	0.93	20
micro avg	0.96	0.94	0.95	160
macro avg	0.96	0.94	0.95	160
weighted avg	0.96	0.94	0.95	160
samples avg	0.94	0.94	0.94	160

Figure 6. Classification Report

The complete classification report can be seen in Figure 6. It displays the performance of the model system precision, recall, f1-score, which is usually used to conclude how precise each class in the model is in classifying [7]. The results of the classification report obtained show that each class in the model has a precision value above 0.85 which is quite good in recognizing its class.

3) Confusion Matrix

Figure 7 is the confusion matrix results of the experiment. It is used to measure the performance of the model in classifying the images [9]. The confusion matrix results show the model's optimality in the classification indicated by constant numbers above 18 to 20 images even though there are 1 to 2 incorrect images. These errors can be caused by several things, one of which is the lack of datasets during the training and validation process.

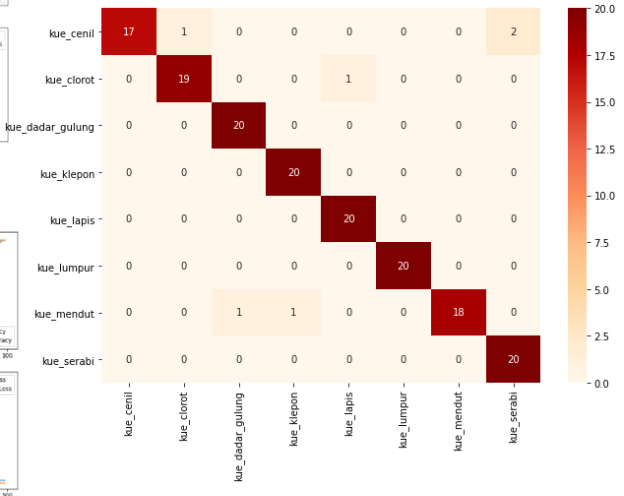

Figure 7. Confusion Matrix Result

4) Random Image Experiments

From 15 random images, the average accuracy value in object classification obtained is 90% to 100%. However, there may be deficiencies in the classification. Of the 15 images, one image has an accuracy below 50% with the correct classification results and one image with an accuracy above 85% with an incorrect classification result. The result of random image experiments is shown in Figure 8.

Figure 8. Random Image Result

IV. Conclusion

Convolutional Neural Networks and the transfer learning technique were successful in automatically extracting Indonesian food images. It can be this can be proven by the results of high classification accuracy. In this work, we used the MobileNetV2 as a base model in the neural network architecture. The proposed classification networks obtained an accuracy value of 0.96 and a loss of 0.14 in its validation. The best classification accuracy value obtained from the model in this research reached 95%, with an average of 90% in each class.

In future works, we will try to use smaller convolutional networks since the Convolutional Neural Networks are consuming high computational time. We will also add more other types of Indonesian traditional cakes to the dataset.

ACKNOWLEDGMENT

We would like to thank to Electronic Engineering Polytechnic Institute of Surabaya who is support this research.

REFERENCES

[1] Dian Ade Kurnia, Andi Setiawan, Dita Rizki Amalia, Rita Wahyuni Arifin, Didik Setiyadi, "Image Processing Identifacation for Indonesian Cake Cuisine using CNN Classification Technique", Journal of Physics Conference Series 1783(1):012047, February 2021.

[2] G. M. Farinella, M. Moltisanti, S. Battiato, "Classifying food images represented as Bag of Textons, IEEE International Conference on Image Processing (ICIP)", 2014.

[3] H. Nakamoto, D. Nishikubo, S. Okada, F. Kobayashi, & F. Kojima. "Food Texture Classification Using Magnetic Sensor and Principal Component Analysis", Third International Conference on Computing Measurement Control and Sensor Network (CMCSN), 2016.

[4] David Joseph Attokaren, Ian G. Fernandes, A. Sriram, Y.V. Srinivasa Murthy, Shashidhar G. Koolagudi, "Food classification from images using convolutional neural networks", Proc. of the 2017 IEEE Region 10 Conference November 2017.

[5] J. R. Rajayogi, G. Manjunath, G. Shobha, "Indian Food Image Classification with Transfer Learning". The 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), 2019.

[6] N. Hnoohom, S. Yuenyong, "Thai fast food image classification using deep learning". The International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON), 2018.

[7] Malina Jiang, "Food Image Classification with Convolutional Neural Networks", CS230: Deep Learning, Fall 2019, Stanford University, CA.

[8] Alsing, O. "Mobile Object Detection using TensorFlow Lite and Transfer Learning", 2018.

[9] Goutte, C., & Gaussier, E. "A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. Lecture Notes in Computer Science", 3408, 345–359, June 2014.

[10] Hu, F., Xia, G. S., Hu, J., & Zhang, L." Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sensing", 7(11), 14680–14707, 2015.

[11] Novakovic, J., Veljovi, A., Iiic, S., Papic, Z., & Tomovic, M. "Evaluation of Classification Models in Machine Learning. Theory and Applications of Mathematics & Computer Science", 7(1), 39–46. 2017

[12] Saha, R. (2018). "Transfer Learning – A Comparative Analysis". December 2018.

[13] Sokolova, M., & Lapalme, G. (2009). "A systematic analysis of performance measures for classification tasks". Information Processing and Management, 45(4), 427–437.

[14] Ilham Firdausi, "Kue Indonesia, Image of various Indonesian traditional cake", 2018, https://www.kaggle.com/ilhamfp31/kue-indonesia.