

Convolutional Neural Network for Ground Coffee Particle Size Classification

Convolutional Neural Network untuk Klasifikasi Ukuran Partikel Serbuk Kopi Giling

Dimas Zaki Alkani Putra, Achmad Pratama Rifai*

Department of Mechanical and Industrial Engineering, Universitas Gadjah Mada, Yogyakarta

Abstract. Indonesia is the fourth largest coffee-producing country in the world. The popularity of coffee is increasing due to people's curiosity about the origin of coffee, from harvest to the hot cup of coffee on their table. This coffee culture drives innovators to develop coffee processing technology. Currently, there are tens of different coffee brewing methods available, each with their own unique flavor characteristics. The particle size of coffee beans is the basis for brewing coffee using specific methods. Identifying the particle size and calibrating tools to grind coffee requires special skills, expertise, experience, and a time-consuming process. Therefore, this study aims to develop a tool to classify the particle size of ground coffee based on computer vision. The object of this research is ground coffee with various particle sizes, which are acquired through imagery and will be classified using Convolutional Neural Network to provide recommendations for brewing coffee according to the particle size of the ground coffee. To build the classification model, the architectures were trained by full learning and transfer learning using VGG-19, MobileNet, and InceptionV3. The results showed that the classification model using the Convolutional Neural Network using the cellphone camera dataset achieved an accuracy value of 0.80. Meanwhile, with the microscope dataset, the model's accuracy only reached 0.58. Therefore, the classification model using the cellphone dataset is feasible to be implemented to determine the particle size.

Keywords: computer vision, deep learning, grinder calibration, ground coffee particle size, multiclass classification

Abstrak. Indonesia adalah negara penghasil kopi terbesar keempat di dunia. popularitas kopi semakin meningkat karena minat orang terhadap asal-usul kopi, mulai dari panen hingga menjadi secangkir kopi panas di meja. Budaya kopi ini mendorong para inovator untuk mengembangkan teknologi pengolahan kopi. Saat ini, terdapat puluhan metode penyeduhan kopi yang berbeda, masing-masing dengan karakteristik rasa yang unik. Ukuran partikel serbuk kopi giling menjadi dasar dalam menyeduh kopi menggunakan alat-alat tertentu. Identifikasi ukuran partikel serbuk kopi dan kalibrasi alat untuk menggiling kopi memerlukan keterampilan khusus, keahlian, pengalaman, dan proses yang memakan waktu. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan metode untuk mengklasifikasikan ukuran partikel serbuk kopi giling berdasarkan visi komputer. Objek dari penelitian ini adalah serbuk kopi giling dengan berbagai ukuran partikel, yang divisualisasi melalui citra dan diklasifikasikan menggunakan convolutional neural network untuk memberikan rekomendasi penyeduhan kopi yang sesuai dengan ukuran partikel serbuk kopi. Untuk membangun model klasifikasi ini, arsitektur jaringan dilatih dengan pendekatan full learning dan transfer learning menggunakan VGG-19, MobileNet, dan InceptionV3. Hasil penelitian menunjukkan bahwa model klasifikasi menggunakan convolutional neural network dengan menggunakan dataset kamera ponsel mencapai nilai akurasi sebesar 0.80. Sementara itu, dengan dataset mikroskop, akurasi model hanya mencapai 0.58. Oleh karena itu, model klasifikasi menggunakan dataset ponsel layak dan sesuai untuk diimplementasikan.

Kata Kunci: computer vision, deep learning, kalibrasi gilingan, klasifikasi multi-kelas, ukuran partikel serbuk kopi giling

Practical Application: The study's classification method offers guidance for choosing coffee brewing methods based on the particle size of ground coffee powder. This benefits consumers who are unsure about grind size and streamlines cafe operations by providing a tool to recalibrate coffee grinders to meet coffee standards. Using computer vision eliminates the need for coffee grounds extraction during calibration, saving time, and reducing costs associated with wasted materials.

*Corresponding: achmad.p.rifai@ugm.ac.id

INTRODUCTION

Indonesia is the fourth-largest coffee-producing country in the world, with 145.9 thousand tons of coffee commodities exported between January and July 2018, valued at USD 442,387,100 (BPS 2017). According to the Ministry of Agriculture, coffee consumption in Indonesia increased from 249,800 tons in 2016 to 314,400 tons in 2018 (Kementan 2018), indicating a growing coffee culture in the country. This has resulted in innovators and businesspeople investing in technology to improve the coffee production process. There are now over ten coffee brewing methods with different coffee extraction characteristics.

The character and taste of coffee depend on the particles size of the ground coffee. The particle size of ground coffee affects the coffee extraction process when water is added. Ground coffee particles are categorized into nine sizes based on the brewing methods, from extra super coarse for direct brewing, to extra super fine for Turkish brewing methods. Selecting the right coffee ground size is essential to bring out the desired coffee flavor character according to the desired brewing method. The basic principles for extracting the flavor and character of coffee are contact time, extraction rate, and flow rate (Musika 2017).

Due to the many types of brewing methods, the particle size of the ground coffee, the coffee grinder for the coffee extraction process, and special skills are needed in determining the particle size of the ground coffee that is suitable for each brewing method. This requires training and experience of brewing coffee. Therefore, it is difficult to identify the ground coffee fineness category by untrained eyes. As such, it is essential to develop a tool for classifying the ground coffee fineness level.

Technological developments in the fourth industrial revolution are also influencing coffee processing technology. Tools to simplify the production process are being developed, such as quality control of defect seeds, quality control of maturity levels, automated roasting machines, and automated coffee extraction using artificial intelligence (AI). One of its applications is in the field of computer vision, which also saw wide implementation for coffee identification problems.

Computer vision is a field of study that aims to develop algorithms and techniques that enable machines to interpret and understand visual data, much like human intelligence and vision. It integrates various subfields such as image acquisition, image processing, image recognition, and decision-making through automated processes (Wijaya and Prayudi 2010). The process of computer vision mimics the way humans see objects through their eyes. The image of the object is then forwarded to the brain for processing, allowing humans to recognize what they see. The results of this recognition are then used to make decisions.

In the context of image classification using computer vision, two scientific approaches can be employed: classi-

cal machine learning and deep learning neural networks. Classical machine learning is suitable for solving simple classification problems with clear feature extraction, while deep learning is more appropriate for classifying large amounts of data, with the feature extraction process occurring simultaneously with the classification.

Khamitova *et al.* (2020) studied the distribution of subtlety coffee grinder for brewing espresso with the aim of processing coffee grounds with the same quality. From the result analysis, they noted that homogenizing the powder particles could reduce the use of raw material, thus cutting production costs.

Other researchers are using AI to improve coffee quality control. Pinto *et al.* (2017) used the Convolutional Neural Network (CNN) to classify defects in raw coffee beans, achieving accuracies ranging from 72.4 to 98.7% based on the types of defects. Santos *et al.* (2020) and García *et al.* (2019) discussed quality control of ground coffee beans by detecting defects. Santos *et al.* (2020) used Support Vector Machine (SVM), Random Forest (RF), and Deep Neural Network (DNN) to classify coffee beans defects according to the shape and color features, achieving high accuracy in their results.

Another study by Leme *et al.* (2019) focuses on examining the degree of roasting of coffee beans using RGB and CIE Lab* methods. The first stage is the selection of bean colors based on Specialty Coffee Association of America (SCAA). Then, the CIE Lab* values are measured. The next stage involves acquiring a dataset of 300 samples, which are then converted to a 16-bit TIFF format. The final stage involves using Matlab software for Neural Network Color space transformation. The results showed that the model can predict the degree of roasting accurately and objectively for whole and ground beans based on CIE Lab*. Arboleda *et al.* (2018) discusses the classification of different coffee bean species, namely Arabica, Liberica, and Excelsa, using the K-Nearest Neighbor method. The aim of this study is to classify different species of coffee beans. Other studies which focused on coffee beans type detection include Rodríguez *et al.* (2020) and Unal *et al.* (2022) who developed deep learning detection model using CNN.

Apart from the development of coffee processing technology, the research interest in the application of AI in coffee processing technology is still in its infancy, particularly in the determination of the ground coffee particle size category. To fill these gaps, this research focuses on developing a computer vision-based tool to analyze the fineness level of ground coffee particle for certain brewing methods and for calibrating coffee grinders. The models are developed based on the convolutional neural network (CNN), which is a deep learning tool widely used for image classification. The data is captured using a microscope and mobile phone camera. The use of a mobile phone camera with a CNN-based model to identify the category of coffee ground particle can facilitate and speed up the process. It can be used by everyone, while the manual identification method requires expert knowledge and a longer processing time.

Furthermore, the developed computer vision method can replace the coffee bean grinding machine calibration process by eliminating one step, namely the machine extraction calibration. In practical use, computer vision can work as an automatic quality control to speed up the calibration process and save raw materials, ensuring that the quality of the coffee produced remains at a taste that is in accordance with quality standards. The implementation of computer vision is expected to help calibrate coffee grinders both in new conditions that need to be calibrated and after a long period of use in which the burr and wear cause the quality of the ground coffee to no longer be up to standard. As such, the purpose of this study is to develop a classification model based on vision that can identify the ground coffee particle size using CNN. These classification models with high accuracy are useful to assist the calibration process of grinding machines.

MATERIALS AND METHODS

Materials

The object of this research is the ground coffee product with nine different categories derived from various types of coffee beans, presented in subsequent subsection. The coffee beans used in experiments come from several variants in Indonesia, both Arabica and Robusta, such as Toraja, Gayo, Flores, Kintamani, etc. Noted that, this study aims to develop a classification model for identifying the ground coffee particle sizes, regardless of the coffee origin. Since the developed method relies on vision-based techniques, the data employed consists of images depicting the ground coffee samples for each category. These images were captured using two devices: a microscope and a mobile phone camera. Table 1 presents the categories of ground coffee particle size which becomes the image classes on the classification model.

Coffee brewing method

To brew the perfect coffee, it's essential to consider measurements, proportions, serving methods, and tools that complement each other. Modern brewing tools on the market have developed around four basic methods. The immersing method, which involves soaking the coffee grinds for a specific time, is the most straightforward and commonly used. While some prefer to brew directly, others use tools like coffee presses or French presses for this method. The dripping method, on the other hand, involves dripping hot or cold water over coffee, and popular brewing tools include V60, Vietnam drip, flat bottom, and Kalitawave. The pressing method, which uses high pressure to extract coffee essence, produces the well-known espresso and involves automatic espresso machines, rockpresso, and aeropress. Finally, the boiling method boils water to brew coffee and uses siphon and Turkish coffee as the brewing method (Greene 2019).

Ground coffee particle size

The size of coffee grounds used for making pure coffee drinks or coffee with other ingredients is a crucial factor to consider. The grind size has a significant impact on the resulting coffee flavor after brewing. The taste of coffee can be influenced by the duration of coffee-water contact, extraction level, and water flow rate (Delta Coffee Roaster 2020). Generally, a larger surface area of coffee grounds results in a higher extraction rate. Thus, a finer grind size is preferable for a larger surface area. A higher extraction rate reduces the contact time between coffee and water. However, finer coffee beans can make it harder for water to penetrate, resulting in longer contact time between coffee and water. Ground coffee particle size can be categorized into nine levels, each with a specific brewing method. These nine classes of ground coffee article sizes are the classification targets of the developed model. The descriptions of each level in decreasing order from the roughest to finest particle sizes are explained as follows. Extra Super Coarse: The largest and coarsest grind size, used for the Tubruk brewing method. Super Coarse: Slightly smaller than Extra Super Coarse, with some finer particles. Suitable for French press brewing. Coarse: Slightly finer than Super Coarse, but still only felt when touched. Recommended for immersing brewing methods. Medium Coarse: The same size as Coarse but can be clearly felt by hand. Used for Vietnamese drip brewing. Medium: The grind size is still difficult to differentiate visually but can be felt clearly by hand. Suitable for pourover brewing methods like Kalitawave and V60. Medium Fine: The distribution of grind size has a smaller deviation than the previous levels. Used for Aeropress brewing. Fine: Finer than Medium Fine, but still distinguishable by hand. Suitable for brewing espresso with low-spec machines or manual espresso machines. Superfine: The second finest level, looking and feeling like flour. Used for espresso brewing with high-spec machines. Extra Super Fine: The finest grind level, resembling powdered sugar. Recommended for Turkish brewing.

Research flow

Figure 1 presents the research flow of this study. The stages of research to be conducted for this study are outlined as follows:

Image acquisition of the dataset

The study focuses on ground coffee beans, distributed randomly on paper for image acquisition. The dataset comprises two sets of images captured using a cell phone camera and a microscope, respectively. Each set was used separately to evaluate its performance in the same model.

Before starting the processing of the image, the images were captured using the cellphone camera. The images were taken by a sensor that can digitize the captured image (Wandell *et al.* 2002). If the results from the sensor are not digital, a switch from analog to digital sensor was required.

Table 1. Image classes of the datasets

Class	Image from Microscope	Image from Mobile Phone	Ground Coffee Particle Size Category	Extraction Method
1			Extra Super Fine	Turkish, Ibrik
2			Super Fine	Espresso Machine
3			Fine	Manual Espresso Machine
4			Medium Fine	Moka Pot, Aeropress
5			Medium	Kalitawave, V60, pourover
6			Medium Coarse	Vietnamese Drip
7			Coarse	Immersing
8			Super Coarse	French Press
9			Extra Super Coarse	Tubruk

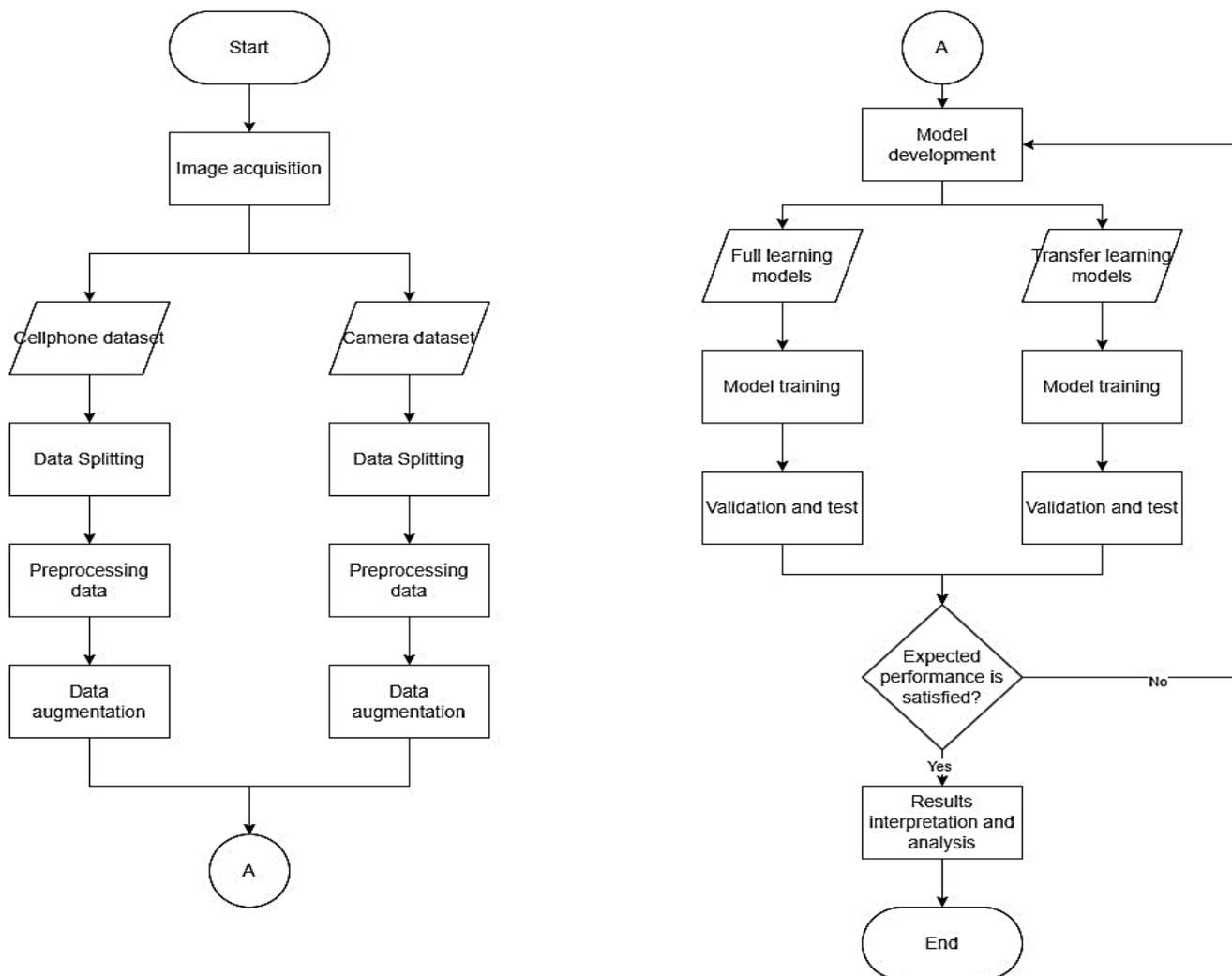


Figure 1. Research flow of this study

Digital images will be represented in a two-dimensional matrix with real numbers. The acquisition process was performed with a 50x magnification microscope and a cellphone camera without magnification. Image acquisition was carried out in a well-lit room during the day to minimize noise. The cell phone camera images were taken at a resolution of 12 mega-pixels and include a total of 1,350 images belonging to nine different categories. Each category contains 150 images with a size of 4032 x 3024 pixels.

In contrast, the microscope images were taken at a lower resolution of 1.3 megapixels and had a size of 1280 x 1024 pixels. An LED lamp with a brightness of 28 lumens and a temperature of 7000K reduced noise during image capture. A computer was used to take screenshots of images from the microscope. The object was taken perpendicularly and was 10 cm from the sensor.

Data preparation

The datasets were then split into training, validation, and testing sets with an 80:10:10 ratio for training and evaluating the deep learning model (Pandey *et al.* 2022). The first part was Training Data consisting of 80% of the total images with nine categories. The second and third parts were Validation and Test Data, each consisting of

10% of the total images. The splitting process was executed randomly and evenly in each class.

Pre-processing and augmentation

Pre-processing improves the image data quality before proceeding to the next step. Operations such as smoothing, color balance, denoise, deblur, normalization, and exposure are used to process the image (Choi *et al.* 2011). Data Augmentation changes data in a way that the computer perceives the data differently while preserving its original content. It increases the model's accuracy by introducing new data for better generalizations (Mahmud *et al.* 2019).

The pre-processing process for the cellphone camera dataset included three stages. During pre-coding, two stages of pre-processing were carried out using the Roboflow application. In the first stage, the images were cropped and resized with fill (within center crop) to a resolution of 1000x1000 pixels. In the second stage, auto-adjust contrast was applied using the contrast stretching command. The final stage of pre-processing was done during coding, which involved re-scaling the pixels to 1/255 pixels to ensure that the scale value ranged from 0.1 to 0.255.

For the microscope dataset, the pre-processing treatment involved two stages. The first stage included cropping and resizing the images to fit within 512x512 pixels using the Roboflow application. In the second stage, the RGB values were re-scaled to 1/255 to ensure consistency.

For the cellphone camera dataset, data augmentation involved a single stage performed during pre-coding using the Roboflow application. The tiling process was used with a 3x3 command, resulting in nine parts for each processed image. This process generated a total of 12,150 image data. Similarly, the microscope dataset was subjected to data augmentation using the tiling method, which was carried out during pre-coding with the Roboflow application. The 5x5 command was used to produce 25 parts for each input image, resulting in a dataset of 8,100 images.

Model development

The model is constructed using Jupyter Notebook software with tensorflow library written in python, employing the convolutional neural network (CNN) algorithm for computer vision. Model development encompasses several distinct architectures, namely full learning, VGG-19, MobileNet, and Inception V3.

Convolutional Neural Network

Convolutional Neural Network (CNN) is a deep learning algorithm commonly used for object recognition, detection, and segmentation. It automatically extracts features directly from images and consists of image input, convolution, pooling, fully connected, and dense layers for feature extraction (Bjerrum *et al.* 2017). CNN does not perform feature extraction manually (Lecun *et al.* 2015) and can learn from data with high-dimensional characteristics (Chakravartula *et al.* 2022). Figure 2 illustrates the different components of the CNN.

Convolution layer

The Convolution layer in the Convolutional Neural Network (CNN) algorithm involves a convolution pro-

cess that operates on the output of the previous layer (Bjerrum *et al.* 2017). Convolution refers to the repetitive application of a function to the output. In the case of image processing, convolution involves applying a kernel to the image at all possible offsets. This operation results in a feature map that represents the output function applied to the input. The convolution operation can be represented by Equation 1.

$$s(t) = (x * t)(t) = \sum_{\alpha=-\infty}^{\infty} x(\alpha) * w(t - \alpha) \dots \dots \dots (1)$$

Note: S(t) = Function operation convolution; X= Input
W= weight from kernel

In the equation above, *x* is input, and *w* is the kernel, or also called as filter. For inputs that have more than one dimension, the equations are as follows (Equation 2 and 3).

$$s(i,j) = (K * I)(i,j) = \sum_{\infty} \sum_{\infty} I(i - m, j - n) K(m, n) \dots (2)$$

$$s(i,j) = (K * I)(i,j) = \sum_{\infty} \sum_{\infty} I(i + m, j + n) K(m, n) \dots (3)$$

In the equation 2 dan 3, *i* and *j* represent the pixels of the image, while *k* can be the kernel and *I* can be the input. The operation is cumulative and can be applied at each position of the kernel as it moves over the image. The convolution layer is composed of neurons that form a filter with length and height (Karpathy 2016). The layers are represented as follows, for example there is a convolution layer with a size of 5x5x3. Then what is meant is 5 pixels in length, 5 pixels in height and 3 dimensions in thickness. The convolution filter will move from left to right using the dot operation between the input and the constant of the filter itself. The result of the dot operation is a feature map or activation map.

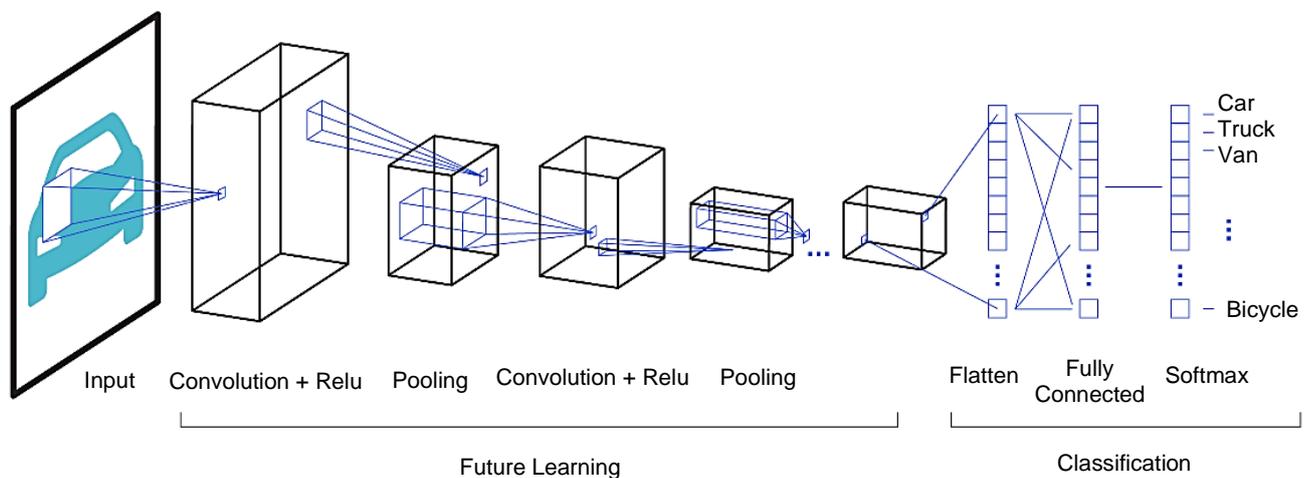


Figure 2. CNN Architecture (Saha 2018)

Subsampling layer

The concept of the subsampling layer simplifies the matrix that is formed to be simpler. Another purpose of the subsampling layer is to increase the positional invariance of features. The method used in subsampling is max pooling (Karpthy 2016). In max pooling, it will divide the layer into several segments of smaller boxes, then the maximum value will be taken from each of the boxes that have been divided. From the max pooling process, the features will look the same and invariant if the matrix experiences translation. An illustration of the pooling operation can be seen in Figure 3.

Flattening

The next stage is the Flattening layer which is a process for converting three-dimensional arrays into one dimension. The matrix flattening layer gets input from the pooling layer and then processes it to the fully connected layer for image classification (Karpthy 2016).

Fully connected layer

The Fully Connected Layer is a layer that is also used in the Multi-Layer Perceptron (MLP) which has a function to perform transformations so that data can be classified linearly. For each neuron in the convolution layer, it is necessary to transform into one dimension, then enter into the fully connected layer (Karpthy 2016). This process causes loss of spatial information and only applies in one direction, so that the fully connected layer will only be performed in the last stage.

The difference between the convolution layer and the fully connected layer is in the connected neurons. The convolutional layer neurons are only connected to certain areas, in contrast to the fully connected layer which is connected to all regions. However, there are similarities between the two, namely operating the dot product so that the results are not too different.

Activation function

The activation function is used after the output layer stages. The Activation Function is used to improve the non-linear characterization of the network by calculating the total weight of the input (Chi *et al.* 2020). Rectified Linear Unit (ReLU), tanh function and sigmoid are

commonly used activation functions. In the case of multi-class classification, the softmax activation function is used.

Loss function

The loss function is an error calculation for the model during the model optimization process. The loss function defines how close the distribution of the predicted labels created by the model corresponds to the true labels (Karpthy 2016). The loss function used is cross entropy for binary classification and categorical cross entropy for multi classification.

Model training

The training phase involves testing the constructed CNN model, utilizing 80% of the total dataset. Here, we use two approaches for model training, which are transfer learning and full learning.

Transfer learning

Transfer learning is a machine learning and deep learning method that reuses previously trained models. The previously trained model will be used as a starting point for building a new model with a new task. The use of previously trained models can easily provide optimization for modeling new tasks. The advantage of transfer learning is that it provides good performance even with a small dataset. In addition, transfer learning can cut time in making models.

Inception V3

Inception V3 is a convolution-based neural network architecture used to classify images. This network was previously introduced by the development team from Google with the name GoogLeNet or Inception V1 (Jahandad *et al.* 2019). It is 42 layers thick and pre-trained using one million images from the ImageNet subset. The Inception V3 architecture can classify 1000 object categories, such as pencils, mice, keyboards, and plants. In the initial version of the Inception model, there are four layers running in parallel, including convolution layers with sizes of 1x1, 3x3, and 5x5, and max pooling with a size of 3x3. Figure 4 presents an illustration of inception architecture.

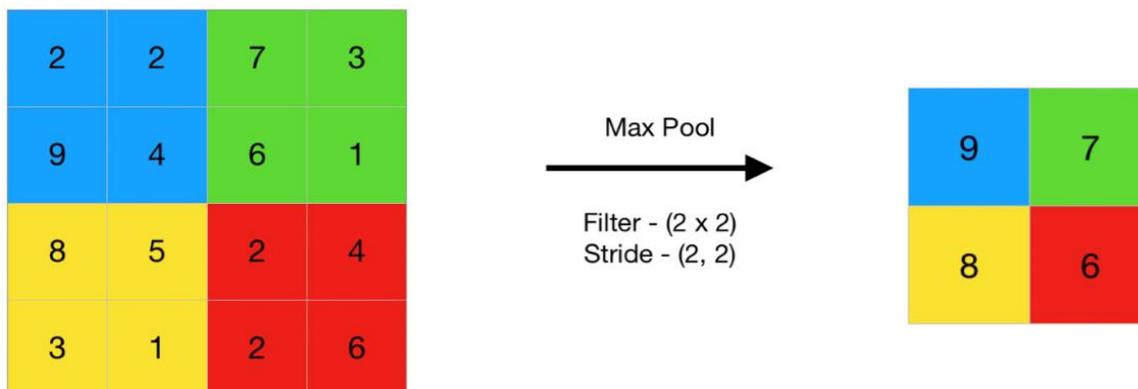


Figure 3. Max pooling operation

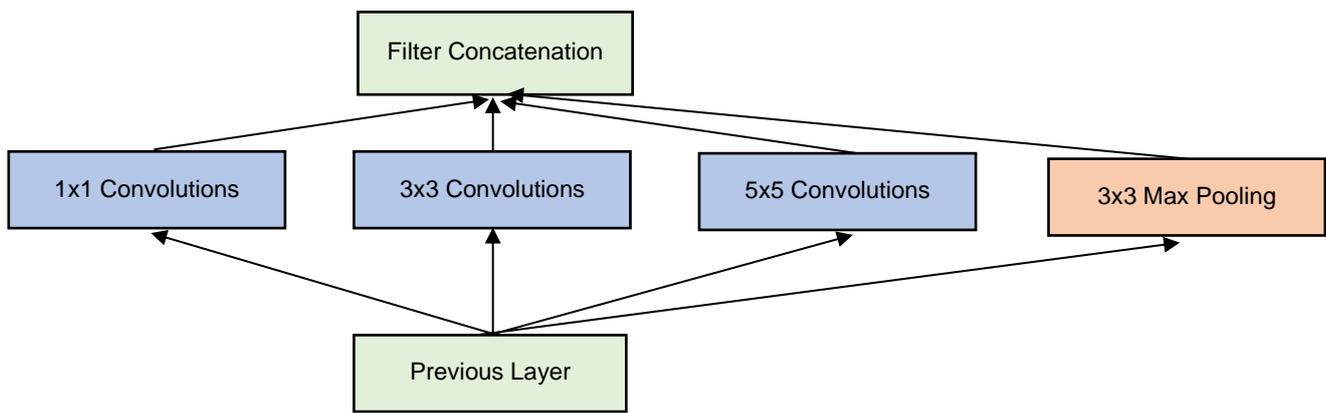


Figure 4. Early versions of inception architecture

Improvements to the Inception v3 model were introduced in 2015, this model made modifications to several parameters. The main modifications are changing the convolution factorization so that it becomes smaller, performing asymmetric spatial factorization, additional utility classification, and grid sizing that is more efficient than before. Modification of the inception model provides advantages from several sides, namely higher model efficiency, deeper models so that it can cut computation time, and computation becomes cheaper. The final model of Inception V3 has 42 layers.

VGG-19

VGG-19 is an artificial neural network with a convolution basis to perform image classification tasks. VGG is a development of Alexnet which was introduced in 2012. The development of VGG is carried out by a different group from Alexnet but uses the basic ideas of its predecessor. Visual Geometry Group is the developer of the VGG model introduced in 2014 (Ikechukwu *et al.* 2021). VGG uses deeper layers to improve model accuracy. The VGG-19 architecture has been trained using 1.2 million training image datasets, 50 thousand validation datasets and 150 thousand testing datasets with 1000 classes using the imagenet subset.

The VGG architecture takes an input to the network in the form of a matrix with a dimension of (244, 244, 3), which is pre-processed by subtracting the average value of RGB from each pixel. The architecture uses a 3x3 kernel with a stride size of 1 and spatial padding to maintain the spatial resolution of the image. Next, a 2x2 max pooling is applied with a stride size of 2 and Rectified Linear Units (ReLU) are used to enhance classification and reduce computation time. The fully connected layer consists of two layers, each with a size of 4096, and a third layer with a size of 1000, followed by the use of the softmax function in the last layer. The architectural diagram of VGG 19 can be found in Table 2.

MobileNet

MobileNet is an artificial neural network renowned for its portable architecture and impressive efficiency. What sets the MobileNet architecture apart from others is its unique approach to convolution, utilizing a depthwise

convolution filter (Howard *et al.* 2017). Comprising 28 layers, the MobileNet architecture is composed of depthwise convolution and pointwise convolution. The default settings result in 4.2 million parameters, but they can be adjusted based on modifications to certain hyperparameters. A more in-depth view of MobileNet's architecture is shown in Table 3.

Full learning

Full learning is an approach for constructing machine learning or deep learning models from scratch, without relying on pre-trained architectures. These models can be highly customized and offer great flexibility in terms of layer type and depth. However, the complexity and computation time of a full learning model can vary greatly depending on the layers built. By optimizing computation time and accuracy results, full learning aims to improve upon existing architectures. Although full learning does not use pre-trained architectures, it may reference layer arrangements from previous architectures for guidance.

Model Validation and Evaluation

Model validation is conducted to test the CNN-generated model, employing 10% of the dataset. Testing the model aims to verify if the constructed model has achieved the expected performance in classifying images of ground coffee particle size.

RESULTS AND DISCUSSION

Datasets after acquisition, preprocessing and augmentation

This dataset included 324 obtained images belonging to nine categories, with 36 images per category. Table 4 presents the distribution of images across the categories and datasets. Afterward, the data is split into three sets, which are training, validation, and test. The results of data splitting are presented in Table 5. At the last stage of data preparation, data augmentation is executed to multiply the number of images to enlarge the dataset, as well as to provide further variability of the dataset. Table 6 provides the number of images in the dataset after data augmentation.

Table 2. VGG-19 layer (Simonyan and Zisserman 2014)

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
Input (224x224 RGB image)					
Conv3-64	Conv3-64 LRN	Conv3-64 Conv3-64	Conv3-64 Conv3-64	Conv3-64 Conv3-64	Conv3-64 Conv3-64
maxpool					
Conv3-128	Conv3-128	Conv3-128 Conv3-128	Conv3-128 Conv3-128	Conv3-128 Conv3-128	Conv3-128 Conv3-128
maxpool					
Conv3-256 Conv3-256	Conv3-256 Conv3-256	Conv3-256 Conv3-256	Conv3-256 Conv3-256 Conv1-256	Conv3-256 Conv3-256 Conv3-256	Conv3-256 Conv3-256 Conv3-256 Conv3-256
maxpool					
Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512 Conv1-512	Conv3-512 Conv3-512 Conv3-512	Conv3-512 Conv3-512 Conv3-512 Conv3-512
maxpool					
Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512 Conv1-512	Conv3-512 Conv3-512 Conv3-512	Conv3-512 Conv3-512 Conv3-512 Conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
Soft-max					

Table 3. MobileNet architecture of the developed model

Type	Stride	Kernel Shapes	Input Size
Conv	2	3x3x3x32	224x224x3
Conv dw	1	3x3x32	112x112x32
Conv	1	1x1x32x64	112x112x32
Conv dw	2	3x3x64	112x112x64
Conv	1	1x1x64x128	56x56x64
Conv dw	1	3x3x128	56x56x128
Conv	1	1x1x128x128	56x56x128
Conv dw	2	3x3x128	56x56x128
Conv	1	1x1x128x256	56x56x128
Conv dw	1	3x3x256	28x28x256
Conv	1	1x1x256x256	28x28x256
Conv dw	2	3x3x256	28x28x256
Conv	1	1x1x256x512	14x14x256
Conv dw	1	3x3x512	14x14x512
Conv	1	1x1x512x512	14x14x512
Conv dw	1	3x3x512	14x14x512
Conv	1	1x1x512x512	14x14x512
Conv dw	1	3x3x512	14x14x512
Conv	1	1x1x512x512	14x14x512
Conv dw	2	3x3x512	14x14x512
Conv	1	1x1x512x1024	7x7x512
Conv dw	2	3x3x1024	7x7x1024
Conv	1	1x1x1024x1024	7x7x1024
Avg Pool	1	Pool 7 x 7	7x7x1024
Fully Connected	1	1024 x 1000	1x1x1024
Softmax	1	Classifier	1x1x1000

Table 4. Image acquisition of the dataset

Category of Ground Coffee	Number of Images	
	Microscope Dataset	Cellphone Camera Dataset
Size 1	36	150
Size 2	36	150
Size 3	36	150
Size 4	36	150
Size 5	36	150
Size 6	36	150
Size 7	36	150
Size 8	36	150
Size 9	36	150
Total	324	1350

Model performance

The experiment was performed to evaluate the performance of the four developed classification models (full learning, VGG19, MobileNet, and Inception) based on their accuracy in classifying ground coffee particle size on the obtained datasets. Based on the results of the model observations, Table 7 shows the best training and testing computation times for all the architectures used in the study. In the cellphone camera dataset, the full learning model performed the best in terms of training and validation speeds, with a value of 2253 seconds. Additionally, the testing time for each image was only 0.375 seconds.

Based on the observations of the microscope dataset, the MobileNet architecture achieved the best training speed performance, with a total training and validation time of 2732 seconds. Additionally, the testing time for each image was only 1 second, indicating that the Mobile-

Net architecture had a significantly faster testing speed than the other architectures.

The VGG-19 architecture had a longer training and validation time for both datasets compared to the other three architectures, which can be attributed to its thicker layers and larger number of parameters. However, since training and validation are one-time runs, these longer times may not be a significant concern. On the other hand, testing time is important, as each image must pass through the image classification stage. The VGG-19 architecture had the longest testing time, taking 21

seconds per image, indicating that it may not be the most efficient option for real-time image classification tasks.

Table 8 presents the results of performance testing on the four architectures, with the VGG-19 model demonstrating the best performance on both datasets. Specifically, the cellphone camera dataset achieved a training accuracy score of 0.990, validation of 0.901, and testing of 0.800. Meanwhile, the microscope dataset achieved a training accuracy score of 0.974, but lower validation and testing scores of 0.502 and 0.576, respectively, suggesting that the model may be overfitting to the training data.

Table 5. Data splitting of the dataset

Category of Ground Coffee	Microscope Dataset			Cellphone camera dataset		
	Training	Validation	Dataset	Training	Validation	Dataset
Size 1	120	15	15	33	2	1
Size 2	120	15	15	27	4	5
Size 3	120	15	15	26	5	5
Size 4	120	15	15	26	5	5
Size 5	120	15	15	30	3	3
Size 6	120	15	15	28	4	4
Size 7	120	15	15	31	1	4
Size 8	120	15	15	31	4	1
Size 9	120	15	15	28	4	4
Total	1080	135	135	260	32	32

Table 6. The number of images after preprocessing and augmentation of the dataset

Category of Ground Coffee	Microscope Dataset			Cellphone camera dataset		
	Training	Validation	Dataset	Training	Validation	Dataset
Size 1	1080	135	135	825	50	25
Size 2	1080	135	135	675	100	125
Size 3	1080	135	135	650	125	125
Size 4	1080	135	135	650	125	125
Size 5	1080	135	135	750	75	75
Size 6	1080	135	135	700	100	100
Size 7	1080	135	135	775	25	100
Size 8	1080	135	135	775	100	25
Size 9	1080	135	135	700	100	100
Total	9720	1215	1215	6500	800	800

Table 7. Time performance of all models

Model	Dataset	Epoch	Training and Validation	Testing
			Duration (s)	Duration per Image (s)
Full Learning	Cellphone Camera	5	2253	0.375
	Microscope	22	6447	0.379
VGG-19	Cellphone Camera	23	161392	21
	Microscope	22	106548	21
MobileNet	Cellphone Camera	22	8933	1
	Microscope	10	2732	1
Inception V3	Cellphone Camera	30	54962	6
	Microscope	5	6115	6

Table 8. Model performance of all models

Model	Dataset	Epoch	Training	Validation	Testing
			Accuracy	Accuracy	Accuracy
Full Learning	Cellphone Camera	5	0.970	0.754	0.638
	Microscope	22	0.982	0.353	0.414
VGG-19	Cellphone Camera	23	0.990	0.901	0.800
	Microscope	22	0.974	0.502	0.576
MobileNet	Cellphone Camera	22	0.978	0.829	0.703
	Microscope	10	0.884	0.40	0.444
Inception V3	Cellphone Camera	30	0.970	0.808	0.690
	Microscope	5	0.569	0.357	0.402

The VGG-19 model outperformed the other three architectures, namely transfer learning, Inception V3, and MobileNet for both the cellphone camera and microscope datasets, owing to its larger training parameters. The VGG-19 architecture had a layer depth of 19 and 143,667,240 training parameters (Chollet 2015), which results in excellent accuracy as compared to the other architectures.

On the other hand, the MobileNet architecture had only 4,300,000 training parameters and 55-layer depths (Chollet 2015), but it still produced good performance with a short training time of 8933 seconds for the cellphone camera dataset and 2732 seconds for the microscope dataset. The MobileNet architecture achieved an accuracy score of 0.703 for the cellphone camera dataset and 0.444 for the microscope dataset.

Meanwhile, the Inception V3 architecture had the deepest layer among the three architectures with a total of 24,000,000 training parameters and 189 layer depths (Chollet 2015). However, the performance of this architecture was worse than the simpler MobileNet architecture. The number of layer depths does not determine the performance of the model; the performance is more affected by the type of filter used and the number of training parameters. The more training parameters a model has, the better its accuracy.

The full learning model performed the worst among the other three models. This can be attributed to the fact

that the VGG-19, Inception V3, and MobileNet transfer learning architectures were trained on a vast collection of diverse images, consisting of over 1.3 million data points and 1000 classes, called the ImageNet dataset. On the other hand, the full learning model only used 8 sequential layers without any functional layers, unlike the other three architectures. However, the full learning model had its advantages over the other architectures. With its straightforward layer arrangement, model training was quicker and more straightforward.

Best model

The accuracy of the developed model in classifying each class is presented in the confusion matrix. Based on the presented confusion matrix in Figure 5, the VGG 19 model achieved an accuracy of 0.8 on the cellphone camera dataset. The number of correct predictions for each class were as follows: 119 images in class 1, 121 images in class 2, 114 images in class 3, 124 images in class 4, 104 images in class 5, 88 images in class 6, 104 images in class 7, 83 images in class 8, and 116 images in class 9. However, the confusion matrix showed that categories 6 and 8 had accuracies below the average accuracy, and the precisions of class 6 and 8 were reported to be 0.65 and 0.61 respectively in Figure 6. Therefore, these classes had lower performance than the other classes.

Confusion Matrix VGG-19 cellphone Camera Dataset

Actual Value	1	119 9.79%	14 1.15%	0 0.00%	2 0.16%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	
	2	4 0.33%	121 9.96%	1 0.08%	8 0.66%	0 0.00%	0 0.00%	1 0.08%	0 0.00%	
	3	3 0.25%	5 0.41%	114 9.38%	10 0.82%	3 0.25%	0 0.00%	0 0.00%	0 0.00%	
	4	0 0.00%	1 0.08%	1 0.08%	124 10.21%	4 0.33%	2 0.16%	0 0.00%	0 0.00%	3 0.25%
	5	0 0.00%	3 0.25%	4 0.33%	9 0.74%	104 8.56%	6 0.49%	4 0.33%	0 0.00%	5 0.41%
	6	0 0.00%	3 0.25%	0 0.00%	10 0.82%	9 0.74%	88 7.24%	7 0.58%	1 0.08%	17 1.40%
	7	0 0.00%	0 0.00%	0 0.00%	1 0.08%	3 0.25%	0 0.00%	104 8.56%	5 0.41%	22 1.81%
	8	0 0.00%	1 0.08%	1 0.08%	3 0.25%	6 0.49%	6 0.49%	3 0.25%	83 6.83%	32 2.63%
	9	0 0.00%	0 0.00%	0 0.00%	1 0.08%	1 0.08%	4 0.33%	5 0.41%	8 0.66%	116 9.55%
			1	2	3	4	5	6	7	8
		Predicted Value								

Figure 5. Confusion matrix of VGG19 for cellphone camera dataset

	precision	recall	f1-score	support
0	0.88	0.94	0.91	126
1	0.90	0.82	0.86	148
2	0.84	0.94	0.89	121
3	0.92	0.74	0.82	168
4	0.77	0.80	0.78	130
5	0.65	0.83	0.73	106
6	0.77	0.84	0.80	124
7	0.61	0.86	0.72	97
8	0.86	0.59	0.70	195
accuracy			0.80	1215
macro avg	0.80	0.82	0.80	1215
weighted avg	0.82	0.80	0.80	1215

Figure 6. Classification report model of VGG19 for cellphone camera dataset

Upon observation of the acquired cellphone camera images, it was found that the precision score of classes 6 and 8 was below that of the other classes due to the characteristics of the ground coffee particle size of the coffee beans. Larger particle sizes of coffee beans resulted in more inhomogeneous particle sizes, while smaller particle sizes produced more homogeneous particle sizes. This inhomogeneous particle size caused the precision of the grade 6 and 8 coffee bean particle sizes to be below the average precision of the other classes. Classes 1 to 4, on the other hand, had a precision score above the average macro accuracy, which was consistent with the particle homogeneity character possessed by coffee mills with smaller sizes.

Figure 7 showcases an instance of correct prediction made by the VGG-19 model on the testing data. The images below were accurately predicted as class 0 or mill category 1, with the original images also belonging to class 0 or mill category 1.

Predicted Class 0, Actual Class 0 Predicted Class 0, Actual Class 0



Predicted Class 0, Actual Class 0 Predicted Class 0, Actual Class 0



Figure 7. Correct Predictions of the VGG-19 Model

Figure 8 illustrates some instances of prediction errors made by the VGG-19 model on the cellphone camera dataset. The actual images belonged to class 0 or mill category 1. However, the three images below were predicted as class 1 or mill category 2, and one image was predicted as class 3 or mill category 4. Upon closer inspection, it was found that the image predicted as class 3, which was originally from class 0, had a predominantly white background. This was one of the factors that may hinder the model's ability to generalize the test image.

Predicted Class 1, Actual Class 0 Predicted Class 1, Actual Class 0



Predicted Class 3, Actual Class 0 Predicted Class 1, Actual Class 0



Figure 8. Wrong predictions of the VGG-19 model

Observing the confusion matrix in Figure 9, the VGG 19 model with a microscope dataset achieved an accuracy value of 0.58. This value was based on the correct predictions of 3 images in class 1, 87 images in class 2, 76 images in class 3, 79 images in class 4, 45 images in class 5, 33 images in class 6, 46 images in class 7, 18 images in class 8, and 74 images in class 9. A closer look at each class revealed contrasting results, with some classes having a high number of true classes (e.g., class

2) while others have low accuracy (e.g., class 1). This could be due to the uneven distribution of the test dataset, but it was generally not a concern as the training process introduces and generalizes each class.

The classification report in Figure 10 shows that some classes had lower precision scores than others. For example, class 1 had a precision score of only 0.12, while class 6 had a score of 0.33. These results can be attributed to the fact that the testing data for class 1 only consists of one image prior to the augmentation process. If this image had features that cannot be generalized by the model, precision may suffer. Additionally, the homogeneity of coffee particle sizes may have contributed to the lower precision in class 6 compared to other classes.

Two factors contributed to the difference in performance between cellphone camera and microscope datasets. First, the pixel size of the microscope sensor was only 1.2 megapixels, while the cellphone camera had a 12 megapixel sensor. Since small particles from coffee grounds required high-quality images, a smartphone camera was better suited for this task. Second, after the augmentation process, the cellphone camera dataset contained 12,150 images while the microscope dataset only had 8,100 images. Deep learning required a large amount of data for optimal accuracy, so the cellphone camera dataset outperformed the microscope dataset across different models.

While VGG-19 was a model with good performance, there was a significant gap between training accuracy and

testing accuracy in its overall architecture, as depicted in Figure 11. The difference between validation and testing accuracy ranged from 0.19 at best to 0.5 at worst. Additionally, the loss value experienced a similar gap, with a high difference between training and validation as shown in Figure 12. Moreover, the validation loss value increased with each higher epoch. This condition of a large gap between training and validation accuracy and loss values was referred to as overfitting. The cause of this issue across the entire model can be attributed to the dataset, which failed to meet the required criteria in terms of both the amount of data and the quality of images.

Based on the experimental results and information provided, it was evident that the VGG-19 model using a microscope dataset was not suitable for classifying coffee ground particle sizes due to its low accuracy of 0.58. Implementing such a model can lead to inaccurate brewing recommendations, affecting the taste of the brewed coffee and causing wastage of coffee material in industrial settings.

On the other hand, using the VGG-19 model with a cellphone camera dataset yielded an accuracy score of 0.8, making it useful for assisting in grinder calibration. Notably, precision values for class 1 are important in grinder calibration, as this class corresponds to the particle size used in the calibration process. However, precision values for classes 6 and 8 were below average, indicating a potential for errors in the brewing method when using this coffee particle size classification application.

Confusion Matrix VGG-19 Microscope Dataset

Actual Value	1	3 0.38%	5 0.63%	9 1.13%	8 1.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%
	2	20 2.50%	87 10.88%	10 1.25%	6 0.75%	2 0.25%	0 0.00%	0 0.00%	0 0.00%	0 0.00%
	3	5 0.63%	8 1.00%	76 9.50%	21 2.63%	9 1.13%	0 0.00%	5 0.63%	0 0.00%	1 0.13%
	4	0 0.00%	4 0.50%	12 1.50%	79 9.88%	21 2.63%	1 0.13%	6 0.75%	2 0.25%	0 0.00%
	5	0 0.00%	2 0.25%	2 0.25%	15 1.88%	45 5.63%	0 0.00%	8 1.00%	0 0.00%	3 0.38%
	6	0 0.00%	2 0.25%	5 0.63%	11 1.38%	19 2.38%	33 4.13%	15 1.88%	2 0.25%	13 1.63%
	7	0 0.00%	0 0.00%	3 0.38%	12 1.50%	8 1.00%	2 0.25%	46 5.75%	5 0.63%	24 3.00%
	8	0 0.00%	0 0.00%	1 0.13%	3 0.38%	0 0.00%	1 0.13%	0 0.00%	18 2.25%	2 0.25%
	9	0 0.00%	0 0.00%	0 0.00%	5 0.63%	1 0.13%	1 0.13%	8 1.00%	11 1.38%	74 9.25%
			1	2	3	4	5	6	7	8
		Predicted Value								

Figure 9. Confusion matrix of VGG19 model for microscope dataset

	precision	recall	f1-score	support
0	0.12	0.11	0.11	28
1	0.70	0.81	0.75	108
2	0.61	0.64	0.63	118
3	0.63	0.49	0.55	160
4	0.60	0.43	0.50	105
5	0.33	0.87	0.48	38
6	0.46	0.52	0.49	88
7	0.72	0.47	0.57	38
8	0.74	0.63	0.68	117
accuracy			0.58	800
macro avg	0.55	0.55	0.53	800
weighted avg	0.60	0.58	0.58	800

Figure 10. Classification report of VGG19 model for microscope dataset

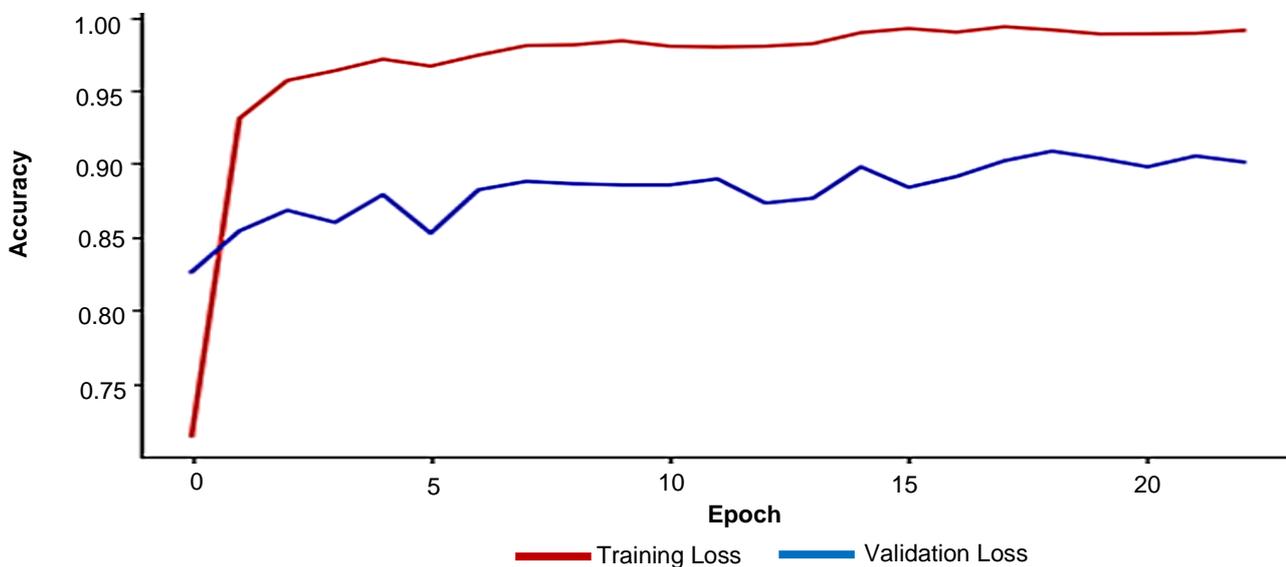


Figure 11. Training and validation accuracy of the VGG-19 model for cellphone camera dataset

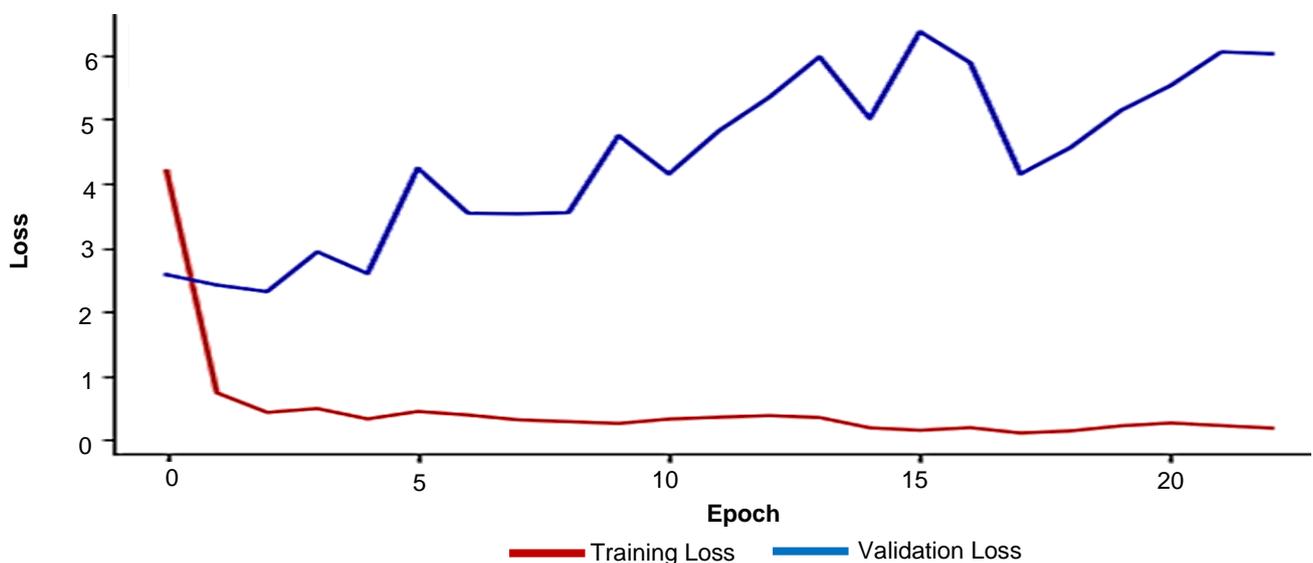


Figure 12. Training and validation loss of the VGG-19 for cellphone cameras dataset

It is important to note that the VGG-19 model tended to fail in predicting grades 6, 7, and 8, instead predicting them as class 9. This phenomenon occurred in both cellphone cameras and microscope datasets and may be attributed to the grinder machine's decreasing homogeneity in particle size in higher classes. As a result, the model may misclassify ground coffee with size 9 in classes 6, 7, and 8, leading to inaccurate predictions.

The classification results can determine coffee brewing recommendations according to the size of the ground coffee powder. This can make it easier for consumers to find ground coffee when they do not know the size of the particle and help to determine the suitable method for brewing coffee. Furthermore, this classification can simplify the production process at the café as a tool for recalibrating coffee grinder machines so that they comply with standards for making perfect coffee without having to discard the raw coffee powder material to recalibrate the grinder machine.

The calibration process using computer vision cuts out one process, namely the extraction of coffee grounds to look for flavors according to standards. So, practically, its use is for the adjustment of the grinder machine, in which the coffee particle categories can be inspected by using the developed model, based on the image of the powder. If it is not appropriate, the grinder machine is reset. The conventional calibration stage requires one necessary step, which is the coffee powder extraction process. By using the proposed prediction model, this step can be omitted, thus reducing setup time and costs due to wasted raw materials.

In addition, this study enriches the literature regarding artificial intelligence applications in agriculture especially for inspection in coffee industry, which is then helpful to promote the growth of coffee consumption among enthusiast and casual consumers.

CONCLUSION

Based on the results of research that has been conducted using the full learning models VGG-19, Inception V3, and MobileNet, the following conclusions can be drawn. Firstly, experiments on coffee ground particle fineness classification using camera phone datasets and microscopes resulted in a model capable of classifying coffee bean fineness into nine different classes. Secondly, the VGG-19 transfer learning architecture was found to be the best model for coffee ground size classification, achieving an accuracy of 0.8 for the cellphone camera dataset, while the model that used the microscope dataset only achieved an accuracy of 0.58. The full learning architecture model for the camera phone dataset achieved an accuracy value of 0.638, while for the microscope dataset, the accuracy value achieved was only 0.414. The model with the best architecture was generated from an experiment using the VGG-19 architecture transfer learning with an accuracy value of 0.8. Thirdly, the classification model performance with the VGG-19 architec-

ture for the whole class had an accuracy of 0.8. However, in classes 6 and 8, it had a precision value of 0.65 and 0.61, respectively. This was due to the homogeneity of the mill size, which worsens with the larger class size. Therefore, precision in classes 6 to 9 tended to be lower than classes 1 to 5. The results confirmed that the proposed classification model can be implemented in assisting the calibration process of grinding machine. Additionally, the proposed model can be easily used by coffee enthusiasts to correctly identify ground coffee particle types for specific brewing methods since it uses phone camera images. By correctly using suitable particle size, the brewing process can deliver coffee beverages according to the standard quality and characteristics of each brewing method.

ACKNOWLEDGEMENTS

This research was funded by Riset dan Inovasi untuk Indonesia Maju (RIIM) Gelombang IV research grant with grant numbers: 172/IV/KS/11/2023 and 6815/UN1/DITLIT/Dit-Lit/KP.01.03/2023.

REFERENCES

- Arboleda ER, Fajardo AC, Medina RP. 2018. Classification of coffee bean species using image processing, artificial neural network and K nearest neighbors. In 2018 IEEE International Conference on Innovative Research and Development (ICIRD), pp. 1-5. IEEE. DOI: 10.1109/ICIRD.2018.8376326
- [BPS] Badan Pusat Statistik. 2017. Statistik Kopi Indonesia. <https://www.bps.go.id/publication/2018/12/20/71e7ada61b13e59be0b1bf80/statistik-kopi-indonesia-2017.html> [Accessed on 17 September 2022].
- Bjerrum EJ, Glahder M, Skov T. 2017. Data augmentation of spectral data for convolutional neural network (CNN) based deep chemometrics. arXiv: 1710.01927. DOI: 10.48550/arXiv.1710.01927
- Chakravartula SSN, Moschetti R, Bedini G, Nardella M, Massantini R. 2022. Use of convolutional neural network (CNN) combined with FT-NIR spectroscopy to predict food adulteration: A case study on coffee. *Food Control* 135: 108816. DOI: 10.1016/j.foodcont.2022.108816.
- Chi P, Zhang Z, Liang R, Cheng C, Chen S. 2020. A CNN recognition method for early stage of 10 kV single core cable based on sheath current. *Electr Power Syst Res* 184: 106292. DOI: 10.1016/j.epsr.2020.106292.
- Choi JY, Ro YM, Plataniotis KN. 2011. A comparative study of preprocessing mismatch effects in color image based face recognition. *Pattern Recognit* 44(2): 412-430. DOI: 10.1016/j.patcog.2010.08.020.

- Chollet F. 2015. Keras open-Source Neural-Network Library Written in Python. <https://github.com/fchollet/keras> [Accessed on 9 August 2022].
- Delta Coffee Roaster. 2020. Ukuran Gilingan Kopi. <https://tamandelta.com/ukuran-gilingan-kopi/> [Accessed on 26 July 2022].
- García M, Candelo-Becerra JE, Hoyos FE. 2019. Quality and defect inspection of green coffee beans using a computer vision system. *Appl Sci* 9(19): 4195. DOI: 10.3390/app9194195.
- Greene M. 2019. A Beginners Guide to Immersion Coffee Brewing. <https://prima-coffee.com/learn/article/coffee-basics/a-beginners-guide-immersion-coffee-brewing/29121> [Accessed on 15 September 2022].
- Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861.
- Ikechukwu AV, Murali S, Deepu R, Shivamurthy RC. 2021. ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images. *Glob Transit Proc* 2(2): 375-381. DOI: 10.1016/j.gltpr.2021.08.027.
- Jahandad, Sam SM, Kamardin K, Sjarif NNA, Mohamed N. 2019. Offline signature verification using deep learning convolutional neural network (CNN) architectures GoogLeNet inception-v1 and inception-v3. *Procedia Comput Sci* 161(2019): 475-483. DOI: 10.1016/j.procs.2019.11.147.
- Karpathy A. 2016. Cs231n convolutional neural networks for visual recognition. <https://cs231n.github.io/> [Accessed on 20 September 2022].
- [Kementan] Kementerian Pertanian RI. 2018. Konsumsi Kopi Nasional. <https://databoks.katadata.co.id/datapublish/2018/07/31/2021-konsumsi-kopi-indonesia-diprediksi-mencapai-370-ribu-ton> [Accessed on 17 September 2022].
- Khamitova G, Angeloni S, Borsetta G, Xiao J, Maggi F, Sagratini G, Vittori S, Caprioli G. 2020. Optimization of espresso coffee extraction through variation of particle sizes, perforated disk height and filter basket aimed at lowering the amount of ground coffee used. *Food Chem* 314: 126220. DOI: 10.1016/j.foodchem.2020.126220.
- LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature* 521(7553): 436-444. DOI: 10.1038/nature14539.
- Leme DS, da Silva SA, Barbosa BHG, Borém FM, Pereira RGFA. 2019. Recognition of coffee roasting degree using a computer vision system. *Comput Electron Agric* 156: 312-317. DOI: 10.1016/j.compag.2018.11.029.
- Mahmud KH, Adiwijaya, Al Faraby S. 2019. Klasifikasi Citra Multi-Kelas Menggunakan Convolutional Neural Network. 2127-2136. eProceedings of Engineering, Telkom University, Bandung.
- Musika YA. 2017. Jenis Coffee Grind Size dan Metode Seduhannya, Otten Coffee. <https://ottencoffee.co.id/majalah/jenis-coffee-grind-size-dan-metode-seduhannya> [Accessed on 25 July 2022].
- Pandey RK, Kumar A, Mandal A. 2022. A robust deep structured prediction model for petroleum reservoir characterization using pressure transient test data. *Pet Res* 7(2): 204-219. DOI: 10.1016/j.ptlrs.2021.09.003.
- Pinto C, Furukawa J, Fukai H, Tamura S. 2017. Classification of Green coffee bean images basec on defect types using convolutional neural network (CNN). In 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA), Bali, Indonesia, pp. 1-5. IEEE. DOI: 10.1109/ICAICTA.2017.8090980.
- Rodríguez JP, Corrales DC, Aubertot JN, Corrales JC. 2020. A computer vision system for automatic cherry beans detection on coffee trees. *Pattern Recognit Lett* 136(2020): 142-153. DOI: 10.1016/j.patrec.2020.05.034.
- Saha S. 2018. A comprehensive guide to convolutional neural network. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53> [Accessed on 20 September 2022].
- Santos FFLD, Rosas JTF, Martins RN, Araújo GDM, Viana LDA, Gonçalves JDP. 2020. Quality assessment of coffee beans through computer vision and machine learning algorithms. *Coffee Sci* 15(2020): e151752. DOI: 10.25186/v15i.1752.
- Simonyan K, Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.
- Unal Y, Taspınar YS, Cinar I, Kursun R, Koklu M. 2022. Application of pre-trained deep convolutional neural networks for coffee beans species detection. *Food Anal Methods* 15(12): 3232-3243. DOI: 10.1007/s12161-022-02362-8.
- Wandell BA, El Gamal A, Girod B. 2002. Common principles of image acquisition systems and biological vision. *Proceedings of the IEEE* 90(1): 5-17. DOI: 10.1109/5.982401
- Wijaya AA, Prayudi Y. 2010. Implementasi Visi Komputer dan Segmentasi Citra untuk Klasifikasi Bobot Telur Ayam Ras. Seminar Nasional Aplikasi Teknologi Informasi (SNATI). G1-G5. Yogyakarta, 19 Juni 2010.