Industria: Jurnal Teknologi dan Manajemen Agroindustri
Volume 12 No 1: 73-88 (2023)

Industria: Jurnal Teknologi dan Manajemen Agroindustri
http://www.industria.ub.ac.id
ISSN 2252-7877 (Print) ISSN 2548-3582 (Online)
https://doi.org/10.21776/ub.industria.2023.012.01.7

Image-based Quality Identification of Black Soybean (*Glycine soja*) Using Convolutional Neural Network

Identifikasi Mutu Kedelai Hitam (Glycine soja) Berbasis Citra Digital Menggunakan Metode Jaringan Saraf Konvolusional

Mas'ud Effendi*, Naufal Hilmi Ramadhan, Arif Hidayat
Department of Agro-Industrial Technology, Faculty of Agricultural Technology, Universitas Brawijaya
Jl. Veteran, Malang 65145, Indonesia
*mas.ud@ub.ac.id
Received: 14th January, 2022; 1st Revision: 14th July, 2022; 2nd Revision: 25th December, 2022; Accepted: 27th July, 2023

Abstract

The problem faced in identifying the quality of black soybeans is that the quality of the assessment is inconsistent and it takes a relatively long time. This study aims to determine the best convolutional neural network architecture by comparing the performance of Custom CNN, MobileNetV2, and ResNet-34 architectures in identifying the quality level (grade) of black soybeans. The quality of black soybean is split into 4 different classes based on physical characteristics (split, damaged, other colors, wrinkles, dirt) and moisture content test. The number of images used is 1300 images, with the ratio of training data, validation data, and testing data are 50:25:25, 60:25:15, and 70:20:10. The best model for identifying the quality based on the physical characteristics is the MobileNetV2 architecture with a ratio of 50:25:25 which produces an accuracy of 90.18%. Morover, the best model for identifying the quality based on the moisture content is the ResNet-34 architecture with a ratio of 70:20:10, which produces an accuracy of 78.12%. The best overall accuracy in identifying the quality based on both physical characteristics and moisture content is the ResNet-34 architecture, with a ratio of 70:20:10, with an average accuracy of testing data of 79.21%.
Keywords: black soybean, Convolutional Neural Network, image, MobileNetV2, ResNet-34

Abstrak

Permasalahan yang dihadapi dalam identifikasi mutu kedelai hitam adalah kualitas penilaian yang tidak konsisten dan membutuhkan waktu yang relatif lama. Penelitian ini bertujuan untuk menentukan arsitektur jaringan saraf konvolusional terbaik dengan membandingkan kinerja antara arsitektur Custom CNN, MobileNetV2, dan ResNet-34 dalam identifikasi tingkat mutu kedelai hitam. Mutu kedelai hitam terdiri dari 4 kelas dengan parameter uji fisik (belah, rusak, warna lain, keriput, kotoran) dan uji kadar air. Peneliti ini menggunakan1300 citra dengan rasio data latih, data validasi, dan data uji yang digunakan adalah 50:25:25, 60:25:15, dan 70:20:10. Hasil terbaik untuk identifikasi mutu parameter uji fisik adalah pada arsitektur MobileNetV2 dengan rasio 50:25:25 dan akurasi 90,18%. Hasil terbaik untuk identifikasi mutu parameter uji kadar air adalah arsitektur ResNet-34 dengan rasio 70:20:10 dan akurasi 78,12%. Hasil akurasi terbaik secara keseluruhan dengan identifikasi parameter fisik dan kadar air adalah arsitektur ResNet-34 dengan rasio 70:20:10 yang memiliki rata-rata akurasi data uji 79,21%.
Kata kunci: *citra, Jaringan Saraf Konvolusional, kedelai hitam, MobileNetV2, ResNet-34*

INTRODUCTION

One of Indonesia's most essential commodities from the food crops sub-sector is soybeans. Soybean is a raw material that is quite popular in making various processed foods such as tempeh, tofu, soy sauce, flour, milk, tauco, and so on. The level of consumption of soybeans in 2019 was 2.26 million tons of dry soybeans, with around 96% of the total consumption used as processed food raw materials (indirect consumption) (Pusat Data dan Sistem Informasi Pertanian, 2020).

Black soybean (*Glycine soja*) is one of the often used soybean varieties. Black soybeans in Indonesia are usually used as raw materials for making soy sauce. Several studies have shown that black soybeans also have the potential to be processed into other products, such as making functional soy (Wardani &

Wardani, 2014), yogurt (Riyanto & Rahayuningsih, 2015), ice cream (Sanjaya et al., 2019), tofu (Zakaria et al., 2016) and tempeh (Nurrahman & Nurhidajah, 2015). The potential utilization of black soybeans is extensive, so the quality of black soybeans must be considered following the purpose of their consumption.

Quality is the material properties' tolerance limit that affects the consumer acceptance degree. The tolerance limit is arranged in a standard that can function as a determinant of the selling price to follow the material quality level. Black soybean quality standards are regulated in SNI 01-3922-1995, consisting of four quality levels. Typical black soybean test parameters are water content, split grains, damaged grains, different colored grains, wrinkled grains, and dirt.

Problems often experienced in manually identifying the quality of grain and legume species include inconsistent quality assessments because they are only based on the subjectivity of each operator (Wallelign et al., 2019). Neethirajan et al. (2007) also stated that identifying grain quality by visual inspection by skilled workers requires a relatively long time and a high degree of subjectivity. Therefore, technological adaptation is needed for quality assessment which is still carried out traditionally. One technology that can be applied is digital image analysis, such as corn (Effendi et al., 2019) and coffee (Effendi et al., 2017). Research on the identification of the yellow soybean quality (Glycine max) was conducted by Muthmainnah et al. (2016) based on physical appearance using digital image processing and artificial neural network methods. The research only conducted physical tests without testing the water content. Digital processing is also used to identify the quality of this black soybean. This research also considers the water content test. The development of digital image analysis at this time uses a lot of deep learning algorithms.

Deep learning is a popular algorithm for retrieving information from a digital image. The advantage of deep learning is that it can identify various positions in an image (translation invariance) well. This is very difficult to do when only using machine learning because it requires much larger parameters, so the computational process takes longer. The advantage of deep learning in terms of computing is that it can take advantage of the Central Processing Unit (CPU), Random Access Memory (RAM), and Graphics Processing Unit (GPU) capabilities so that the computing process will be slightly faster (Ilahiyah & Nilogiri, 2018). One of the deep learning methods that can be used is a Convolutional Neural Network (CNN). This method is designed to process two-dimensional data using convolution layers which can extract information spontaneously in studying a feature/property of an image. So, this method results in a better accuracy theoretically than using ordinary artificial neural networks, which are only designed to process one-dimensional data. One-dimensional data design in image classification assumes that each pixel is an independent feature (Putra et al., 2016). One of the disadvantages of a simple convolutional neural network is the relatively longer computation time when the number of layers is very deep. Good accuracy may not be achieved if the layer is not too deep. The many layers may also cause a vanishing gradient (Praveenkumar & Dharmalingam, 2019). A vanishing gradient is a condition when the network in the initial layer does not train because the gradient is already or is close to (Small & Briscoe, 2020). Based on this, the architecture of transfer learning can be a solution because it uses architectural models that have been researched and weights that have been previously trained. Transfer learning is a CNN approach to solving similar problems by updating parameters to obtain compatibility with the new dataset (Naufal & Kusuma, 2021).

This study aims to determine the best convolutional neural network architecture by comparing the performance of the Custom CNN, MobileNetV2 and ResNet-34 architectures in identifying the quality level of black soybeans. The use of MobileNetV2 and ResNet-34 architectures as a comparison of Custom CNN because these architectures are good enough to identify complex soybean images. Skip connection in the ResNet-34 architectural concept can overcome the vanishing gradient problem when the network architecture is made very deep (He et al., 2016). The concept of MobileNetV2 that makes it very computationally efficient and allows it to be applied to mobile phones is due to bottlenecks and skip connections. Operations on the inner layer of the bottleneck allow converting low-level inputs, such as pixels, to high-level descriptors, such as image categories. This makes the training process faster with fairly good accuracy (Sandler et al., 2018). This study results are expected to be the basis for developing a black soybean quality identification system using digital imagery and an alternative method of quality identification that is faster with good and consistent quality assessments.

METHODS

Sample Preparation

The sample used was black soybean variety detam-1 obtained from research center for various legumes and tubers which is called “Balai Penelitian Tanaman Aneka Kacang dan Umbi (Balitkabi)”, Malang Regency, East Java Province. Samples need to be classified manually into four quality levels with a weight of 20 grams each. This size was obtained from the results of several initial experiments with 14 cm distance from the camera to the object. If the sample weight is more than 20 grams, the details of each parameter during the image acquisition process are not very visible, which makes it difficult for the training process to get good accuracy. The samples taken consisted of 25 samples of quality I, 30 samples of quality II, 35 samples of quality III, and 40 samples of quality IV.

Quality Labeling

Quality labeling is necessary because the approach type used in building the system is supervised learning. Supervised learning is an approach that requires the system to study datasets that have labeled data (Santoso et al., 2021). Quality labeling refers to SNI 01-3922-1995 with parameters that can determine the quality class, called water content, split grain, broken grain, other color grains, wrinkled grain, dirt. Each maximum levels of these parameters are presented in Table 1.

Table 1. Quality standards

Parameter	Quality Classification (%)			
	I	II	III	IV
Water content	13	14	14	16
Split grain	1	2	3	5
Broken grain	1	2	3	5
Other color grains	1	3	5	10
Wrinkled grain	0	1	3	5
Dirt	0	1	2	3

(Sourcer: SNI 01-3922-1995)

Physical parameter sorting was carried out on 20 grams of sample. Each physical parameter is weighed and the percentage is calculated using Equation (1).

$$\text{Physical parameter (\%)} = \frac{\text{each parameter weight}}{\text{sample weight}} \times 100\% \qquad (1)$$

Samples for each parameter were then put together again and put into a plastic standing pouch to be tested for water content using an AR991 moisture tester. The sample is then labeled according to quality based on the measurement of each parameter. Variations are made on quality labeling by providing a combination of various parameter values (Alzubaidi et al., 2021; Kim et al., 2019), such as physical parameters being omitted at quality I or only 1% of split grain parameters, damaged grain parameters and other colors being made 0%. Variation was carried out to test and evaluate the performance of the architecture in identifying the quality of black soybeans in various scenarios and conditions. This variation is also applied to other quality levels so that the model can learn to identify various cases and improve its performance on unseen data, such as a new image with a combination of features that are not the same as the combination in the image used in the training data.

Image Acquisition

Image acquisition was carried out in a box measuring 30 cm x 30 cm x 30 cm using the Xiaomi 11T Pro smartphone camera (108 MP, 0.7µm pixel size, 9-in-1 2.1µm superpixel, f/1.75). Illumination in this process uses a 12-watt/120-volt white Stripe LED lamp. Image acquisition was performed ten times for one sample to obtain 1300 images from 130 samples. Parameter position changes are made for each image taken. This is done to test the model to understand each image's different parameter positions. The image acquisition process arrangement is presented in Figure 1, with the distance between the sample and the camera lens being 14 cm.

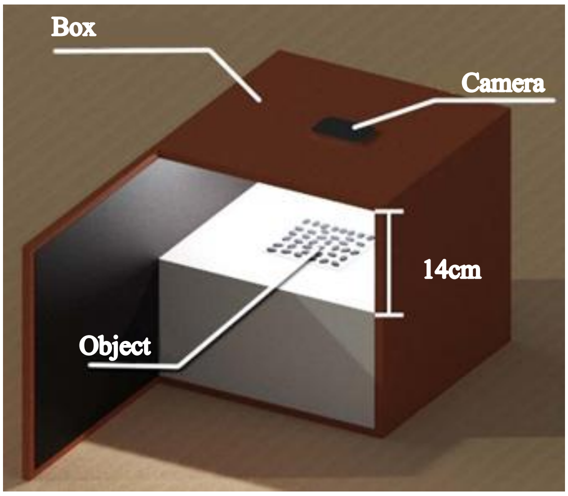


Figure 1. Image Acquisition

The camera for image standardization uses manual mode with ISO 400, a shutter speed of 1/160", and a resolution of 300 x 300 pixels. Reducing image resolution is very common in digital image processing to make computation faster. Reducing image accuracy does not significantly affect accuracy, and the main factor is the network architecture (Effendi et al., 2019; Gao et al., 2021; He et al., 2016). An image example of each quality level is presented in Figure 2.

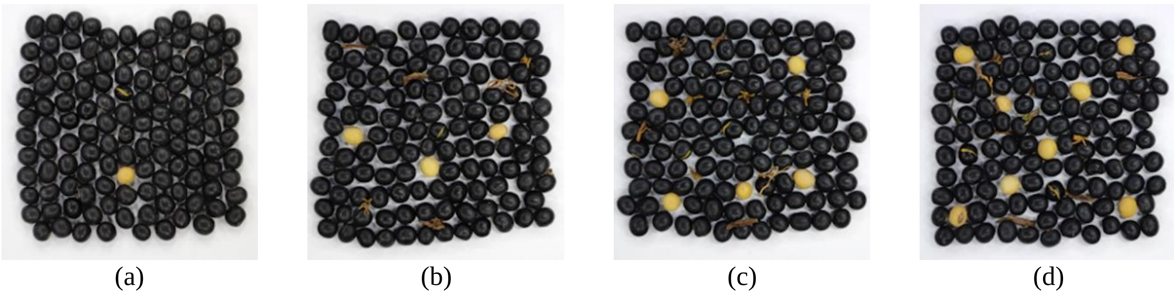


Figure 2. Image of Black Soybean; (a) Quality I, (b) Quality II, (c) Quality III, (d) Quality IV

Classification Using Convolutional Neural Networks

Compiling the program code was completed on jupyter notebook version 6.4.3 with the miniconda3 package management. Jupyter notebooks have a faster auto-complete feature and a snippet feature (J.Cop Snippets v1.0) to simplify the process of compiling program code. The library used is PyTorch because it is more flexible for deep learning programming. PyTorch is also highly recommended for research because it optimizes the training process (Hendri et al., 2021). The architecture is trained and tested at Google Colaboratory by importing the code compiled in the jupyter notebook. The training process in Google Colaboratory aims to speed up training time because there is a free Compute Unified Device Architecture (CUDA) Graphics Processing Unit (GPU).

Data Splitting

Data splitting is the process of dividing the image dataset into three parts, training data, validation data, and test data for training, validation, and testing (Gao et al., 2021). The proportions of training data, validation data, and test data are selected by trial and error to find the best architectural performance. The proportion of data used in this study is 50:25:25, 60:25:15, and 70:20:10 based on research conducted by Gao et al. (2021). and Géron (2019). Details of the proportion of data at each quality level can be shown in Table 2.

Table 2. Details of the images number in each proportion

Proportions	Label	Training Data	Validation Data	Test Data	
				Physical	Water Content
50:25:25	M1	125	63	31	31
	M2	150	75	38	37
	M3	175	88	44	43
	M4	200	100	50	50
60:25:15	M1	150	63	19	18
	M2	180	75	23	22
	M3	210	88	26	26
	M4	240	100	30	30
70:20:10	M1	175	50	13	12
	M2	210	60	15	15
	M3	245	70	18	17
	M4	280	80	20	20

Augmentation and Normalization

Augmentation is used to increase the variety of images to increase the architecture's ability in classification. The augmentations used in the training data are rotation of 5°, horizontal-vertical flip, and random resized crop. Image augmentation is not used in data validation because the goal is to test the architecture with the original image. The images on the training data, validation data, and test data are then normalized by changing the image's resolution to 224 x 224 pixels and changing the image data form to tensor format because PyTorch can only process data in that format. The transfer learning architecture needs to normalize the RGB channel by adding mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225] (Islam et al., 2022). The batch size initialization performed at this stage is 64.

Architecture Design

This study uses the Custom CNN architecture and two transfer learning architectures. The convblock layer in the Custom CNN architecture contains convolution operations, ReLu activation functions, and MaxPooling. The direct classifier is at the fully-connected layer with the LogSoftmax activation function, according to Dhomne et al. (2018) and Gao et al. (2021). The CNN Custom Architecture is presented in Figure 3.

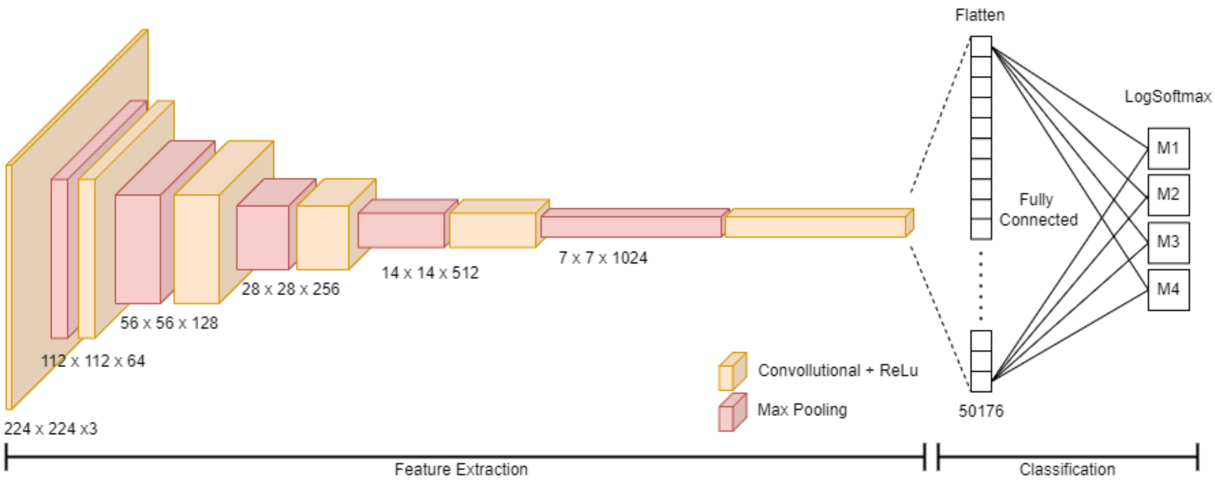


Figure 3. CNN Custom Architecture

The transfer learning architectures used are MobileNetV2 and ResNet-34. Both architectures are used because the skip connection process in these architectures can overcome the gradient vanishing problem if the architecture is made with very deep layers. Another advantage of the skip connection is that the data processed in each layer group will retain the characteristics of the input data studied without losing input data information by data extinction and gradient vanishing (Ahn & Yim, 2020). The MobileNetV2 architecture is superior in terms of computational speed and size because operations on the inner layers of the bottleneck allow changing low-level inputs to high-level descriptors, such as pixels to image categories, so that the training process is faster (Sandler et al., 2018). The architectures of MobileNetV2 and ResNet-34 are presented in Figure 4 and Figure 5.

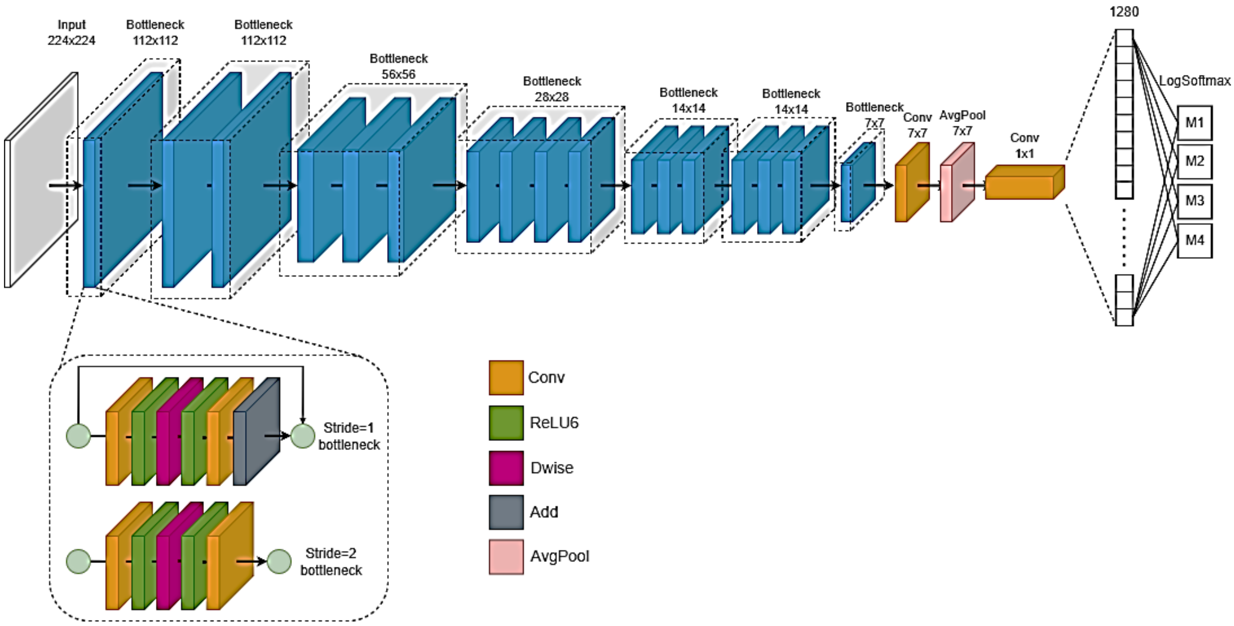


Figure 4. MobileNetV2 Architecture

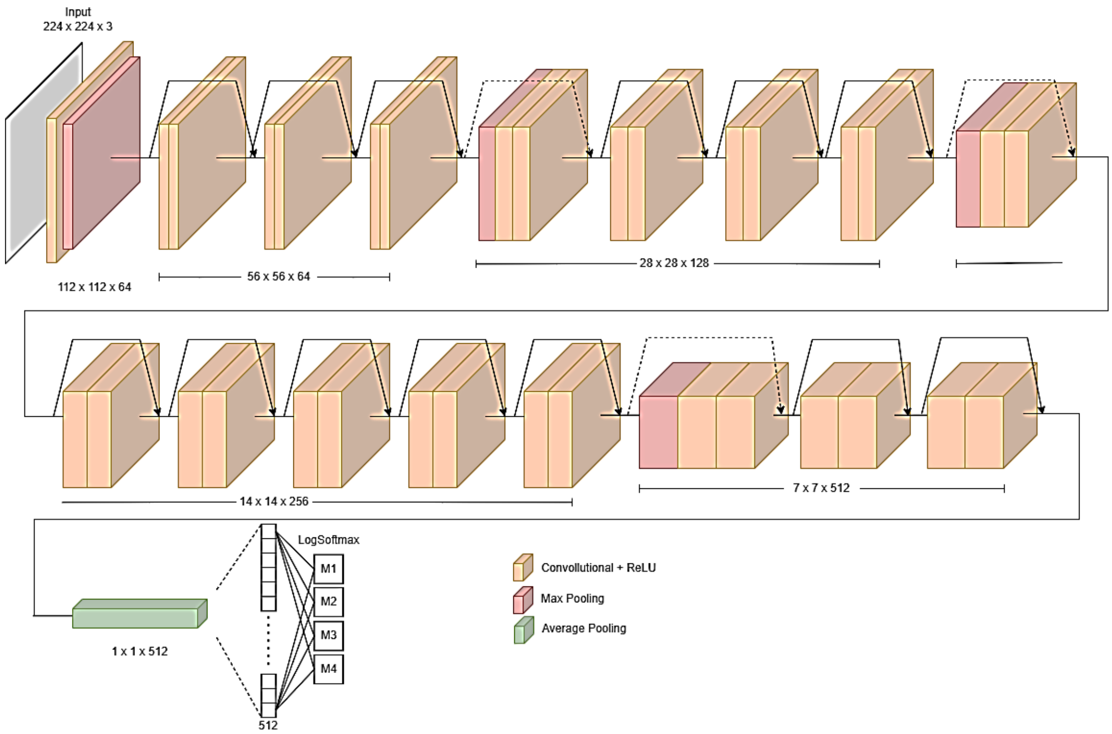


Figure 5. ResNet-34 Architecture

<u>Training Preparation</u>

The variables initialized in this process are model, criterion, optimizer, and callback (Chollet, 2018; Géron, 2019). The model variable wraps the architecture compiled in the previous process. The criterion variable contains a loss function that measures the model's performance based on the prediction of error (loss) on the target. The Loss Function used is Negative Log Likelihood Loss (NLLLoss) because it adapts to LogSoftmax as the last layer activation function. The optimizer variable contains the optimization algorithm for updating the weights in the model. The optimizer used is the adaptive moment with weight decay (AdamW) which can generalize the model better than Adam, which was the previous version (Landro et al., 2021). The callback variable contains an algorithm to control training and testing so that it stops when accuracy has not increased in several early-stop patience (epoch).

<u>Training and Testing</u>

The purpose of training and testing data on machine learning (including on CNN) is to build models that can recognize data patterns and provide accurate predictions on new data that have never been seen before. The custom CNN architecture differs from the architecture derived from transfer learning in this process. The training and testing process on the custom CNN architecture is carried out in one phase with a standard learning rate of 0.001 and an early stop patience of 5 to help prevent overfitting and allow the model to generalize better to invisible data. A learning rate of 0.001 is used in this phase because it is considered not too big or too small. The training process is faster if the learning rate is high, but the results are less accurate. The training process takes longer if the learning rate is low, but the results are more accurate (Géron, 2019).

Training and testing the architecture from transfer learning is performed in two phases: adaptation and fine-tuning. The weight for feature extraction during the adaptation phase will be frozen, so the training process for updating the weight only occurs in fully connected. The learning rate in this phase uses the standard learning rate, but the early stop patience is reduced to 2. The fine-tuning phase is a retraining of the entire architecture. Unfreeze at feature extraction is carried out in this phase, so it must reduce the learning rate not to damage the weight trained by the architecture maker (Prasetyo et al., 2021). The learning rate used in this phase is 10^{-5}, and the early stop patience used is 5 to prevent overfitting and ensure the model is not trained for too long, which can cause a decrease in generalization performance.

<u>Architecture Performance Measurement</u>

Architecture performance measurement uses test data that does not participate in the training and testing process. This aims to test accuracy with data that the model has never seen before. Performance measurement was done on physical parameters (splits, damage, wrinkles, other colors, dirt) and water content parameters. The performance of the predicted results is compared with the actual results, which can be seen using the confusion matrix. The confusion matrix is used to tell the prediction results that are True and False so that accuracy calculations are needed to determine architectural performance. The equation for this calculation can be seen in Equation (2).

$$\text{Accuracy (\%)} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} = \frac{\text{Total amount of true predictions}}{\text{The total amount of data}} \quad (2)$$

True Positive (TP): the amount of data that is positive and correctly predicted as positive
True Negative (TN): the amount of data that is negative and correctly predicted as negative
False Positive (FP): the amount of data that is negative but predicted as positive
False Negative (FN): the amount of data that is positive but predicted as negative.

RESULTS AND DISCUSSION

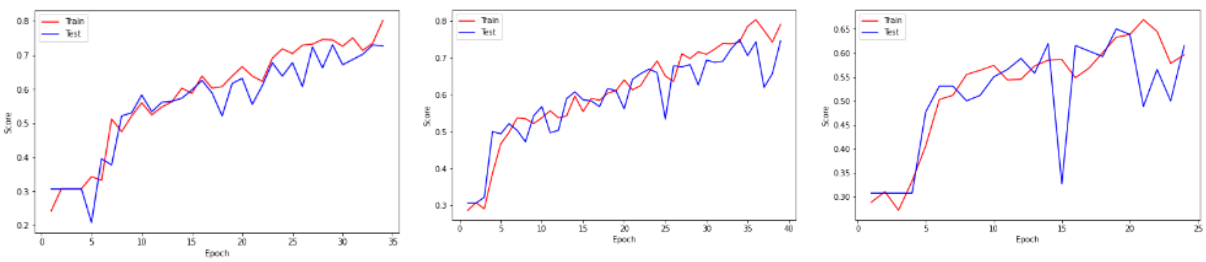
Custom CNN Architecture

The Custom CNN architecture training and testing phase resulted in training data and validation data accuracy from three different data proportions, 50:25, 60:25, and 70:20, according to Géron (2019). The result of Custom CNN architecture training data and validation data accuracy is shown in Table 3 and Figure 6. The best validation data accuracy from the Custom CNN architecture is at 60:25 proportion.

The proportion of data with the best validation data accuracy will not always produce the best training data accuracy. This condition can be influenced by overfitting, dataset size, and model complexity (Alzubaidi et al., 2021; Ying, 2019).

Table 3. Accuracy of training data and validation data on the Custom CNN architecture

Training Data : Validation Data Proportion	Accuracy (%)		Epoch
	Training Data	Validation Data	
50:25	74.46	73.01	29
60:25	74.23	74.85	34
70:20	63.25	65.00	19



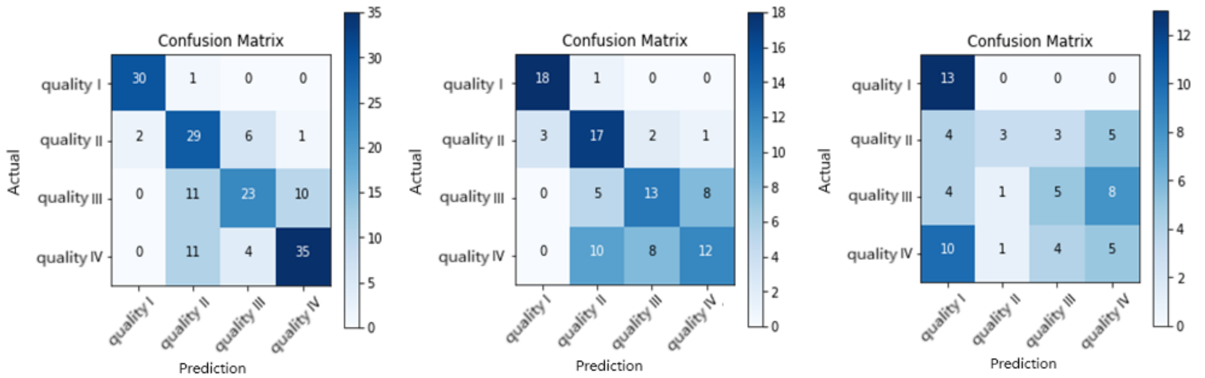
(a) 50:25 Proportion (b) 60:25 Proportion (c) 70:20 Proportion

Figure 6. Custom CNN Accuracy Graph

The architecture accuracy measurement for predicting physical parameters and water content parameters was performed using test data with a proportion of 25%, 15%, and 10% of the total dataset. This proportion adjusts to the previously selected training data and validation data. The predicting accuracy of physical parameters for the Custom CNN architecture is presented in Table 4, and the confusion matrix is presented in Figure 7. The best accuracy for predicting physical parameters on the Custom CNN architecture is in the 25% data proportion. This result is likely due to the sufficient size of the dataset (Alzubaidi et al., 2021).

Table 4. Predicting accuracy of physical parameters on the Custom CNN architecture

Test Data Proportion (%)	Predictions		Accuracy (%)
	True	False	
25	117	46	71.77
15	60	38	61.22
10	26	40	39.39



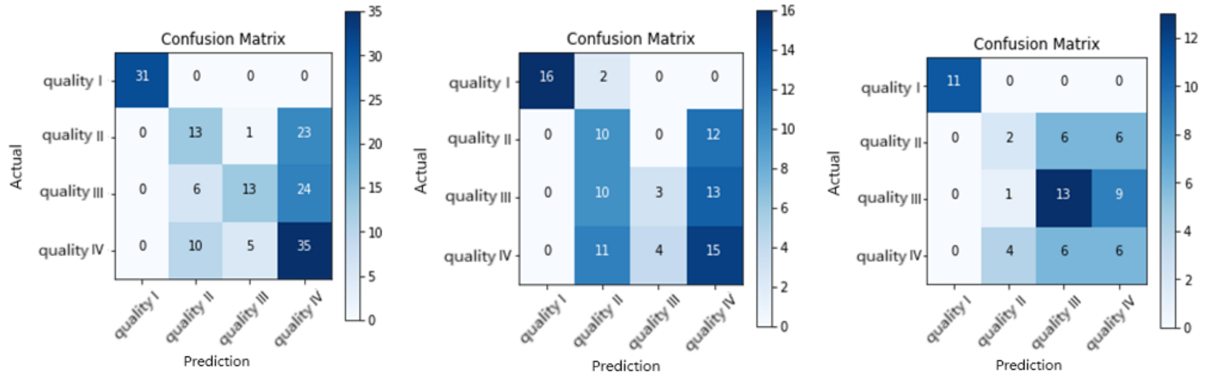
(a) 25% Proportion (b) 15% Proportion (c) 10% Proportion

Figure 7. Custom CNN Confusion Matrix of Physical Parameters

The predicting accuracy of water content parameters on the Custom CNN architecture is presented in Table 5, and the confusion matrix is presented in Figure 8. The best accuracy for predicting the water content parameter on the Custom CNN architecture is in the 25% data proportion. This result is likely due to the sufficient size of the dataset (Alzubaidi et al., 2021).

Table 5. Predicting accuracy of water content parameters on the Custom CNN architecture

Test Data Proportion (%)	Predictions		Accuracy (%)
	True	False	
25	92	69	57.14
15	44	52	45.83
10	32	32	50.00



(a) 25% Proportion (b) 15% Proportion (c) 10% Proportion

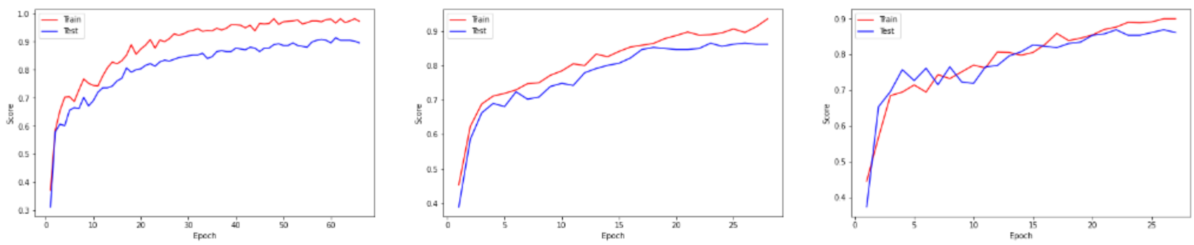
Figure 8. Custom CNN Confusion Matrix of Water Content Parameters

MobileNetV2 Architecture

The MobileNetV2 architecture training and testing phase resulted in training data and validation data accuracy from three different data proportions, 50:25, 60:25, and 70:20. The result of MobileNetV2 architecture training data and validation data accuracy is shown in Table 6 and Figure 9. The best validation data accuracy from the MobileNetV2 architecture is at 50:25 proportion. This result is likely due to the sufficient size of the dataset (Alzubaidi et al., 2021).

Table 6. Accuracy of training data and validation data on the MobileNetV2 architecture

Training Data : Validation Data Proportion	Accuracy (%)		Epoch
	Training Data	Validation Data	
50:25	96.62	91.41	61
60:25	88.97	86.50	23
70:20	87.68	86.92	22



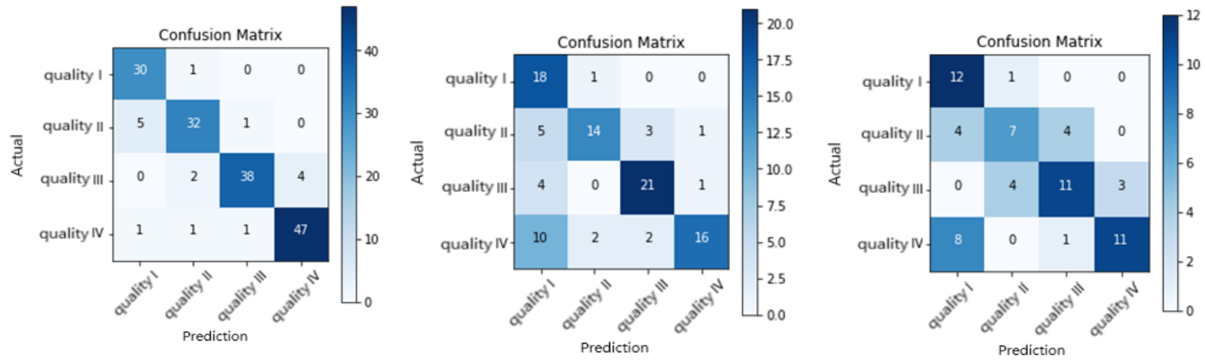
(a) 50:25 Proportion (b) 60:25 Proportion (c) 70:20 Proportion

Figure 9. MobileNetV2 Accuracy Graph

The predicting accuracy of physical parameters for the MobileNetV2 architecture is presented in Table 7, and the confusion matrix is presented in Figure 10. The best accuracy for predicting physical parameters on the MobileNetV2 architecture is in the 25% data proportion. This result is likely due to the sufficient size of the dataset (Alzubaidi et al., 2021).

Table 7. Predicting accuracy of physical parameters on the MobileNetV2 architecture

Test Data Proportion (%)	Predictions		Accuracy (%)
	True	False	
25	147	16	90.18
15	69	29	70.40
10	41	25	62.10



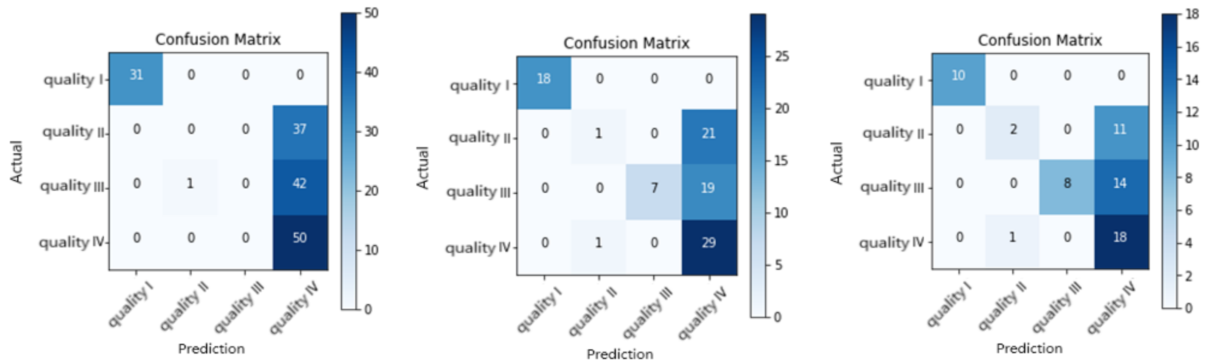
(a) 25% Proportion (b) 15% Proportion (c) 10% Proportion

Figure 10. MobileNetV2 Confusion Matrix of Physical Parameters

The predicting accuracy of water content parameters on the MobileNetV2 architecture is presented in Table 8, and the confusion matrix is presented in Figure 11. The best accuracy for predicting the water content parameter on the MobileNetV2 architecture is in the 10% data proportion. This result is likely due to the sufficient size of the dataset (Alzubaidi et al., 2021).

Table 8. Predicting accuracy of water content parameters on the MobileNetV2 architecture

Test Data Proportion (%)	Predictions		Accuracy (%)
	True	False	
25	81	80	50.31
15	55	41	57.29
10	38	26	59.37



(a) 25% Proportion (b) 15% Proportion (c) 10% Proportion

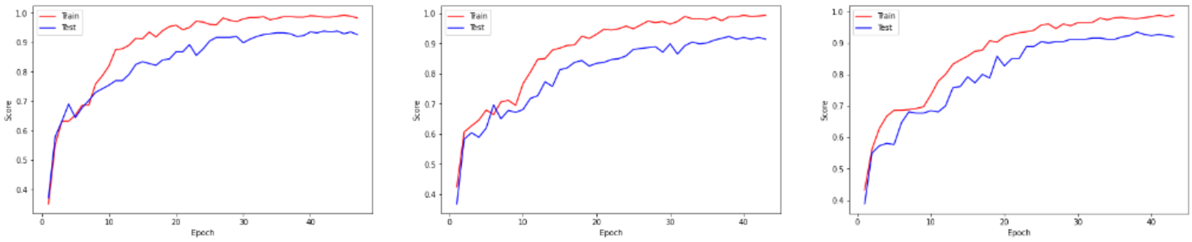
Figure 11. MobileNetV2 Confusion Matrix of Water Content Parameters

ResNet-34 Architecture

The ResNet-34 architecture training and testing phase resulted in training data and validation data accuracy from three different data proportions, 50:25, 60:25, and 70:20. The result of ResNet-34 architecture training data and validation data accuracy is shown in Table 9 and Figure 12. The best validation data accuracy from the ResNet-34 architecture is at 50:25 proportion. This result is likely due to the sufficient size of the dataset (Alzubaidi et al., 2021).

Table 9. Accuracy of training data and validation data on the ResNet-34 architecture

Training Data : Validation Data Proportion	Accuracy (%)		Epoch
	Training Data	Validation Data	
50 : 25	98.62	93.87	42
60 : 25	98.85	92.33	38
70 : 20	97.67	93.46	38



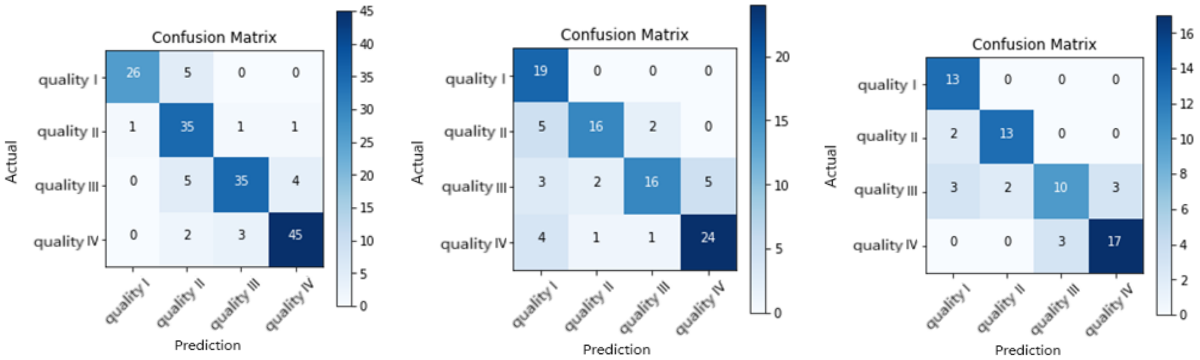
(a) 50:25 Proportion (b) 60:25 Proportion (c) 70:20 Proportion

Figure 12. ResNet-34 Accuracy Graph

The predicting accuracy of physical parameters for the ResNet-34 architecture is presented in Table 10, and the confusion matrix is presented in Figure 13. The best accuracy for predicting physical parameters on the ResNet-34 architecture is in the 25% data proportion. This result is likely due to the sufficient size of the dataset (Alzubaidi et al., 2021).

Table 10. Predicting accuracy of physical parameters on the ResNet-34 architecture

Test Data Proportion (%)	Predictions		Accuracy (%)
	True	False	
25	141	22	86.50
15	75	23	76.53
10	53	13	80.30



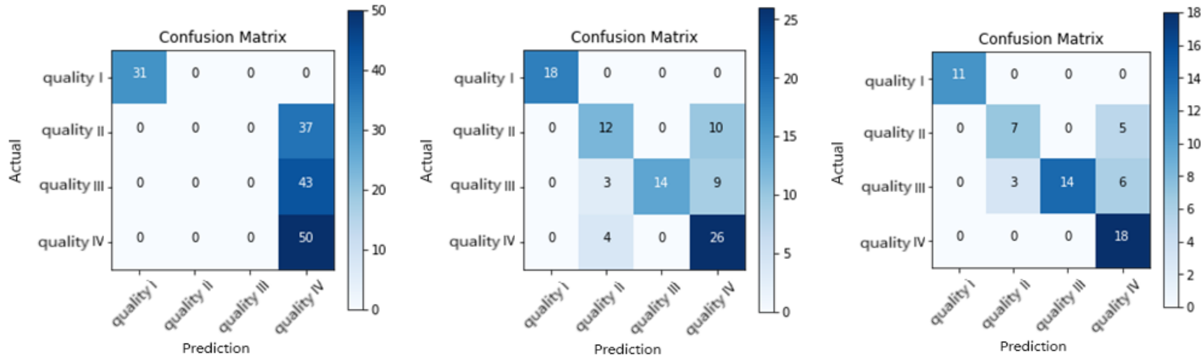
(a) 25% Proportion (b) 15% Proportion (c) 10% Proportion

Figure 13. ResNet-34 Confusion Matrix of Physical Parameters

The predicting accuracy of water content parameters on the ResNet-34 architecture is presented in Table 11, and the confusion matrix is presented in Figure 14. The best accuracy for predicting the water content parameter on the ResNet-34 architecture is in the 10% data proportion. This result is likely due to the model's complexity (Alzubaidi et al., 2021).

Table 11. Predicting accuracy of water content parameters on the ResNet-34 architecture

Test Data Proportion (%)	Predictions		Accuracy (%)
	True	False	
25	81	80	50.31
15	70	26	72.91
10	50	14	78.12



(a) 25% Proportion (b) 15% Proportion (c) 10% Proportion

Figure 14. ResNet-34 Confusion Matrix of Moisture Content Parameters

Best Architecture Determination

The best architecture determination is based on the overall accuracy results shown in Table 12. The accuracy resulting from the three data proportions on the Custom CNN architecture is still lower than the other architectures because it has an average test data accuracy below 70%. Adding the training data proportion in this architecture decreases the validation and test data accuracy. The decrease in level accuracy after adding a variety of training data is because this architecture only uses five layers in feature extraction and one fully-connected layer, so the model cannot study the image's features/properties properly. This result is likely due to the low model complexity and lack of feature representation (Alzubaidi et al., 2021). The model can learn the image's features/properties well if there are more layers and fully-connected layers, but this requires more extended training and testing time. The accuracy score is also related to the object complexity in the classified image. Pratiwi et al. (2021) research related to malaria parasites detection, which only uses three layers in feature extraction and one fully-connected layer, can produce up to 95.83% accuracy. The layer in that study is lower than in this black soybean quality detection study, but the accuracy is higher. This condition shows that the identification of black soybean quality is categorized as complex enough objects to be trained by machines, so this simple Custom CNN architecture is highly not recommended.

Adding the training data proportion on the MobileNetV2 architecture reduces the physical parameter test data accuracy but increases the moisture content parameter accuracy. This trade-off between physical parameters and moisture content accuracy indicates that the MobileNetV2 architecture, in this case, is maybe overfitting (Alzubaidi et al., 2021). In this study, the decrease in physical parameter accuracy was not proportional to the increase in moisture content parameter accuracy. The model tries to generalize the physical and water content parameters to identify the quality when the training data proportion increases. This result is indicated by the decreased difference in the accuracy between training data and validation data and also between physical parameter test data and water content parameter. Various factors can affect model performance, including dataset, model architecture, and parameter settings (Alzubaidi et al., 2021; Enkvetchakul & Surinta, 2022). This generalization is still not optimal because the average test data accuracy is still lower than ResNet-34. Reducing the test data proportion makes the accuracy even lower. In contrast, Budiman et al. (2021) stated that mask-wearing detection using the MobileNetV2 architecture shows that reducing the test data proportion increases accuracy. The MobileNetV2 architecture in this study can be recommended more for identifying physical parameters because of its high accuracy. The data proportion recommended for identifying physical parameters is 50:25:25, which produces the highest accuracy, 90.18%.

The ResNet-34 architecture produces better accuracy in all data proportions in this study compared to the Custom CNN and MobileNetV2 architectures. The model cannot generalize well between physical parameters and water content for quality identification in the 50:25:25 data proportion. This is shown by the physical prediction accuracy, which reached 86.50%, but the prediction accuracy of the water content was only 50.31%. The trade-off between physical parameter accuracy and water content for the ResNet-34 architecture is quite good at the 60:25:15 data proportion because the accuracy decrease of physical parameter test data is offset by a significant increase in the accuracy of the water content parameter. The accuracy increase in the physical parameters and water content occurs in the 70:20:10 data proportion. This result also happened in Li & Rai (2020) research on detecting apple leaf disease, which showed that using more training data could increase the model's ability to classify so that accuracy also increases. The model in the 70:20:10 data proportion was able to generalize quite well, as indicated by the difference in the physical parameter test data and the water content accuracy, which was quite close, 80.30% and 78.12%. Therefore, the best recommendation for identifying black soybean quality is the ResNet-34 architecture, with 70:20:10 data proportion, with an average test data accuracy of 79.21%.

Table 12. Architecture accuracy

Architecture	**Data Proportion**	**Training Data Accuracy (%)**	**Validation Data Accuracy (%)**	**Test Data Accuracy (%)**		**Average Test Data Accuracy (%)**
				Physical	**Water Content**	
Custom CNN	50:25:25	74.46	73.01	71.77	57.14	64.45
	60:25:15	74.23	74.85	61.22	45.83	53.52
	70:20:10	63.26	65.00	39.39	50.00	44.69
MobileNetV2	50:25:25	96.62	91.41	90.18	50.31	70.24
	60:25:15	88.97	86.50	70.40	57.29	63.84
	70:20:10	87.68	86.92	62.10	59.37	60.73
ResNet-34	50:25:25	98.62	93.87	86.50	50.31	68.40
	60:25:15	98.85	92.33	76.53	72.91	74.72
	70:20:10	97.67	93.46	80.30	78.12	79.21

Accuracy improvement can be achieved through data and models. Improvement of accuracy through data can be achieved by increasing the quantity and quality of the training dataset (Ying, 2019). Accuracy improvements through the model can be made by implementing a transfer learning architecture and applying a newer activation function. The latest architecture that can be applied, among others, is Contrastive Captioners (CoCa). This architecture can produce 91% accuracy on the ImageNet (Yu et al., 2022), while the MobileNetV2 and ResNet architectures can only produce 74.7% and 76.35% accuracy on those datasets. The newest activation functions that can be implemented include Swish, Mish, and Logish. Using the old version of the activation function (Sigmoid, Tanh, and ReLU) on the CIFAR-10 dataset by implementing the ResNet architecture resulted in an accuracy of 88.8%, 87.5%, and 89.6%. Using the latest version of the activation function (Swish, Mish, and Logish) for the dataset with the same architecture, each can produce 93.6%, 92.4%, and 94.8% accuracy, respectively (Zhu et al., 2021).

CONCLUSIONS

The study to get the best convolutional neural network architecture for identifying black soybean quality based on digital images was conducted by comparing the performance of Custom CNN, MobileNetV2, and ResNet-34 architectures. The proportions of training data and validation data used in this study were 50:25, 60:25, and 70:20. The best accuracy for identifying black soybean quality based on physical parameters is 90.18% which is produced by the MobileNetV2 architecture with a data proportion of 50:25. The best accuracy for identifying the water content parameter is 78.12% which is produced by the ResNet-34 architecture with a data proportion of 70:20. The best accuracy for identifying soybean quality based on physical parameters and moisture content is 79.21% which is produced by the ResNet-34 architecture with a data proportion of 70:20. The accuracy of the identification can be increased by adding variations in the data amount. The Accuracy improvements can also be achieved by implementing the latest transfer learning architectures, such as CoCa, or implementing new activation functions, such as Mish, Swish or Logish.

Architectural integration with hardware, such as Raspberry Pi, and camera sensors that enable real-time detection of black soybean quality can be considered for further research. The Raspberry Pi has general-purpose input/output pins that allow direct interaction with the camera. The Python programming language used in this study can also be run by Raspberry Pi.

References

Ahn, H., & Yim, C. (2020). Convolutional neural networks using skip connections with layer groups for super-resolution image reconstruction based on deep learning. *Applied Sciences*, *10*(6), 1959. https://doi.org/10.3390/app10061959

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8

Budiman, B., Lubis, C., & Perdana, N. J. (2021). Pendeteksian penggunaan masker wajah dengan metode convolutional neural network. *JIKSI (Jurnal Ilmu Komputer Dan Sistem Informasi)*, *9*(1), 40–47. https://doi.org/10.24912/jiksi.v9i1.11556

Chollet, F. (2018). *Deep Learning with Python*. Manning Publication.

Dhomne, A., Kumar, R., & Bhan, V. (2018). Gender recognition through face using deep learning. *Procedia Computer Science*, *132*, 2–10. https://doi.org/10.1016/j.procs.2018.05.053

Effendi, M., Fatasya, U., & Effendi, U. (2017). Identifikasi jenis dan mutu kopi menggunakan pengolahan citra digital dengan metode jaringan syaraf tiruan. *Jurnal Ilmiah Teknologi Pertanian Agrotechno*, *2*(1), 140–146.

Effendi, M., Jannah, M., & Effendi, U. (2019). Corn quality identification using image processing with k-nearest neighbor classifier based on color and texture features. *IOP Conference Series: Earth and Environmental Science*, *230*, 012066. https://doi.org/10.1088/1755-1315/230/1/012066

Enkvetchakul, P., & Surinta, O. (2022). Effective data augmentation and training techniques for improving deep learning in plant leaf disease recognition. *Applied Science and Engineering Progress*, *15*(3), 1–12. https://doi.org/10.14416/j.asep.2021.01.003

Gao, M., Song, P., Wang, F., Liu, J., Mandelis, A., & Qi, D. (2021). A novel deep convolutional neural network based on resnet-18 and transfer learning for detection of wood knot defects. *Journal of Sensors*, *2021*, 1–16. https://doi.org/10.1155/2021/4428964

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Hendri, Hoki, L., Augusman, V., & Aryanto, D. (2021). Penerapan machine learning untuk mengategorikan sampah plastik rumah tangga. *Jurnal TIMES*, *10*(1), 1–5.

Ilahiyah, S., & Nilogiri, A. (2018). Implementasi deep learning pada identifikasi jenis tumbuhan berdasarkan citra daun menggunakan convolutional neural network. *Justindo (Jurnal Sistem Dan Teknologi Informasi Indonesia)*, *3*(2), 49–56.

Islam, M. M., Hossain, M. B., Akhtar, M. N., Moni, M. A., & Hasan, K. F. (2022). CNN based on transfer learning models using data augmentation and transformation for detection of concrete crack. *Algorithms*, *15*(8), 287. https://doi.org/10.3390/a15080287

Kim, S.-W., Cho, S.-J., Uhm, K.-H., Ji, S.-W., Lee, S.-W., & Ko, S.-J. (2019). Evaluating parameterization methods for Convolutional Neural Network (CNN)-based image operators. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1862–1870. https://doi.org/10.1109/CVPRW.2019.00237

Landro, N., Gallo, I., & La Grassa, R. (2021). Combining optimization methods using an adaptive meta optimizer. *Algorithms*, *14*(6), 186. https://doi.org/10.3390/a14060186

Li, X., & Rai, L. (2020). Apple leaf disease identification and classification using ResNet models. *2020 IEEE 3rd International Conference on Electronic Information and Communication Technology (ICEICT)*, 738–742. https://doi.org/10.1109/ICEICT51264.2020.9334214

Muthmainnah, Z. L., Suyantohadi, A., & Khuriyati, N. (2016). *Identifikasi tingkat mutu kedelai (Glycine max) berdasarkan kenampakan fisik menggunakan pengolahan citra digital*. Departemen Teknologi Industri Pertanian. Fakultas Teknologi Pertanian. Universitas Gadjah Mada. Yogyakarta.

Naufal, M. F., & Kusuma, S. F. (2021). Pendeteksi citra masker wajah menggunakan CNN dan transfer learning. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *8*(6), 1293–11300. https://doi.org/10.25126/jtiik.2021865201

Neethirajan, S., Karunakaran, C., Jayas, D. S., & White, N. D. G. (2007). Detection techniques for stored-product insects in grain. *Food Control*, *18*(2), 157–162. https://doi.org/10.1016/j.foodcont.2005.09.008

Nurrahman, & Nurhidajah. (2015). Pengaruh konsumsi tempe kedelai hitam terhadap aktivitas makrofag dan kadar interleukin 1(IL-1) pada tikus secara in vivo. *Jurnal Agritech*, *35*(3), 294–299. https://doi.org/10.22146/agritech.9340

Prasetyo, E., Purbaningtyas, R., Adityo, R. D., Prabowo, E. T., & Ferdiansyah, A. (2021). Perbandingan convolution neural network untuk klasifikasi kesegaran ikan bandeng pada citramata. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *8*(3), 601–608. https://doi.org/10.25126/jtiik.2021834369

Pratiwi, N. K. C., Ibrahim, N., Fu'adah, Y. N., & Rizal, S. (2021). Deteksi parasit plasmodium pada citra mikroskopis hapusan darah dengan metode deep learning. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, *9*(2), 306–317. https://doi.org/10.26760/elkomika.v9i2.306

Praveenkumar, G. D., & Dharmalingam, M. (2019). Recurrent cascade neural network for image classification. *International Journal of Scientific & Technology Research*, *8*(10), 1009–1012.

Pusat Data dan Sistem Informasi Pertanian. (2020). *Outlook Kedelai: Komoditas Pertanian Subsektor Tanaman Pangan* (A. A. Susanti & A. Supriyatna (eds.)). Pusat Data dan Sistem Informasi Pertanian. Sekretariat Jenderal Kementerian Pertanian.

Putra, I. W. S. E., Wijaya, A. Y., & Soelaiman, R. (2016). Klasifikasi citra menggunakan convolutional neural network (cnn) pada Caltech 10. *Jurnal Teknik ITS*, *5*(1), A65–A69. https://doi.org/10.12962/j23373539.v5i1.15696

Riyanto, S., & Rahayuningsih, H. M. (2015). Yoghurt kedelai hitam (black soyghurt) dapat menurunkan kadar LDL tikus hiperkolesterolemia. *Jurnal Gizi Dan Dietetik Indonesia (Indonesian Journal of Nutrition and Dietetics)*, *3*(1), 1–9. https://doi.org/10.21927/ijnd.2015.3(1).1-9

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4510–4520.

https://doi.org/10.1109/CVPR.2018.00474

Sanjaya, R., Suhartatik, N., & Wulandari, Y. W. (2019). Es krim kedelai hitam (Glycine soja L) dengan penambahan lidah buaya (aloe vera). *JITIPARI (Jurnal Ilmiah Teknologi Dan Industri Pangan UNISRI)*, *4*(2), 74–81.

Santoso, P., Abijono, H., & Anggreini, N. L. (2021). Algoritma supervised learning dan unsupervised learning dalam pengolahan data. *G-Tech: Jurnal Teknologi Terapan*, *4*(2), 315–318.

Small, S., & Briscoe, J. (2020). *Gradients and Tissue Patterning*. Academic Press.

Wallelign, S., Polceanu, M., Jemal, T., & Buche, C. (2019). Coffee grading with convolutional neural networks using small datasets with high variance. *Journal of WSCG*, *27*(2), 113–120. https://doi.org/10.24132/JWSCG.2019.27.2.4

Wardani, A. K., & Wardani, I. R. (2014). Eksplorasi potensi kedelai hitam untuk produksi minuman fungsional sebagai upaya meningkatkan kesehatan masyarakat. *Jurnal Pangan Dan Agroindustri*, *2*(4), 58–67.

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, *1168*, 022022. https://doi.org/10.1088/1742-6596/1168/2/022022

Yu, J., Wang, Z., Vasudevan, V., Yeung, L., Seyedhosseini, M., & Wu, Y. (2022). CoCa: Contrastive captioners are image-text foundation models. *Transactions on Machine Learning Research*, 1–20. https://doi.org/10.48550/arXiv.2205.01917

Zakaria, F. R., Firdaus, D. P. R., & Yuliana, N. D. (2016). Konsumsi tahu kedelai hitam untuk memperbaiki nilai SGOT/ SGPT dan aktivitas antioksidan plasma penderita diabetes tipe 2. *Pangan*, *25*(2), 95–104.

Zhu, H., Zeng, H., Liu, J., & Zhang, X. (2021). Logish: A new nonlinear nonmonotonic activation function for convolutional neural network. *Neurocomputing*, *458*, 490–499. https://doi.org/10.1016/j.neucom.2021.06.067