



# The Evaluation of Deep Learning Using Convolutional Neural Network (CNN) Approach for Identifying Arabica and Robusta Coffee Plants

Bayu Taruna Widjaja Putra<sup>1,2</sup> · Rizki Amirudin<sup>1</sup> · Bambang Marhaenanto<sup>1</sup>

Received: 12 November 2021 / Revised: 8 March 2022 / Accepted: 10 March 2022 / Published online: 22 March 2022  
© The Author(s), under exclusive licence to The Korean Society for Agricultural Machinery 2022

## Abstract

**Purpose** Arabica and Robusta coffee plants are physically distinctive as manifested in their leaves, leaf shape, color, and size. However, for ordinary people or those who have just begun their business in coffee cultivation, identifying the type of coffee plant can be challenging. In this study, we incorporated and evaluated deep learning technology to identify the types of coffee based on leaf image identification.

**Methods** In this study, we designed a deep learning architecture and compared it with the well-known approaches, including LeNet, AlexNet, ResNet-50, and GoogleNet. A total of 19,980 image datasets were split into training and testing data, consisting of 15,984 images and 3,996 images, respectively.

**Results** The hyperparameters were taken into account where the use of 100 epoch and 0.0001 learning rate provided the highest accuracy. In addition, 10-fold cross-validation and ROC were used for evaluating the proposed architectures. The results show that the developed convolutional neural network (CNN) generated the highest accuracy of 97.67% compared to LeNet, AlexNet, ResNet-50, and GoogleNet with an accuracy rate of 97.20%, 95.10%, 72.35%, and 82.16%, respectively.

**Conclusions** The modified-CNN algorithm had satisfactory accuracy in identifying different types of coffee. The underlying principles of such classification draw specific attention to the leaf shape, size, and color of Arabica and Robusta coffee. For future works, it is a potential method that can be used to rapidly identify diverse varieties of Robusta and Arabica coffee plants based on leaf tissue and above canopy characteristics.

**Keywords** Leaf classification · Robusta coffee · Arabica coffee · Convolutional neural network · Coffee species · Precision agriculture

## Introduction

Indonesia is one of the tropical countries of which climate affects the crops cultivation on a regular basis; one of those is the coffee plant. Arabica and Robusta are the two most widely cultivated coffees by smallholders. According to the Directorate General of Plantations (2019), the plantations of Arabica and Robusta coffee

in Indonesia in 2019 were 342,393 hectares and 396,676 hectares, respectively. The coffee plantation is projected to increase to meet consumer needs. The increase is also marked by the increase of new workers starting to work in the coffee cultivation business. This may become a problem as the new workers often find it challenging to distinguish the Arabica from the Robusta coffee. According to Ferreira et al. (2019), Arabica and Robusta coffee can be classified according to their shapes as they are physically different. Arabica has smaller leaves with a darker and glossy surface, while the leaves of Robusta coffee are lighter, less shiny, more extensive, and slightly wavy. Arabica coffee leaves are approximately 12–15 cm × 6 cm, and Robusta coffee's are more than 20 cm × 10 cm (Aak, 1988).

In order to overcome these challenges, training can usually be carried out to introduce the knowledge to the community and workers, despite the time and costs required. As widely acknowledged, training/workshops are very unlikely to be

---

✉ Bayu Taruna Widjaja Putra  
bayu@unej.ac.id

<sup>1</sup> Laboratory of Precision Agriculture and Geo-informatics, Faculty of Agricultural Technology, Jember University, Jl. Kalimantan 37 Jember 68121, East Java, Indonesia

<sup>2</sup> Center of Excellence on Artificial Intelligence for Industrial Agriculture, Jember University, Jember, Indonesia

carried out face-to-face in this pandemic era. Therefore, measures to better identify these types of coffee can be done by utilizing advanced technology, such as deep learning. Deep learning is a part of machine learning closely related to artificial neural network algorithms with multiple layers in the network that are immediately deduced and optimized for expected outcomes.

Over the years, several techniques and methods have been developed to support agriculture, including on-farm and off-farm. Several studies successfully used traditional statistics to support their experimental analysis and help to take a decision. However, not all problems can be solved by using traditional statistics, other approaches and techniques are called upon. Some of those approaches are machine learning and deep learning. Bolívar-Santamaría and Reu (2021) incorporate remote sensing and machine learning such as random forest for classifying agroforestry systems. Uyeh et al. (2021) use machine learning to optimize the sensor's placement in a greenhouse.

According to Bzdok et al. (2018) and Lewis (2000), traditional statistics, such as discriminant analysis and logistic regression, can be used for classification purposes. It requires several assumptions to follow a particular distribution, both for variable responses and predictors and data normality assurance. These properties ultimately make these traditional statistical methods challenging to use. In comparison, deep learning produces predictions that aim to predict future results so that it can be used to identify the most action without understanding the underlying mechanism. In addition, deep learning only requires input in the form of data and generates the desired output, followed by independent computerization to study the existing input.

Deep learning offers a wide array of advantages in agriculture, such as crop management, yield prediction (Bhanumathi et al., 2019), disease detection (Lu et al., 2021), weed detection (Hasan et al., 2021; Osorio et al., 2020), and water and land management (Liakos et al., 2018). However, the most widely observed phenomenon in the agricultural sector relates to visual/optical data. Thus, in its implementation, the convolutional neural network (CNN) part of deep learning is mostly used in computer vision. CNN is an automatic image detector (Jia et al., 2020) and classification (Barré et al., 2017; Koklu et al., 2021; Sihalath et al., 2021) utilizing supervised learning using computer vision technology. In addition, many studies working with signal processing and natural language processing (NLP) incorporate CNN to solve their problems (Kvam & Kongsro, 2017; Li et al., 2021). In other sectors such as robotics, researcher uses CNN to process the biological signal of the human to control the robot (Ak et al., 2022).

Some existing deep learning architectures have been successfully trained to identify objects such as fruit detection, pest detection, diseases detection, and high accuracy. However, specific warrants need to be addressed to the output, which

may vary when implemented to different objects, making it difficult to maintain its generalization. A specific application and implementation need to be conducted to perform the principle of precision agriculture. This study is projected to improve the CNN model for fast and accurate identification of coffee types based on leaves characteristics. Several existing deep learning techniques for object identification include LeNet, AlexNet, ResNet-50, and GoogleNet, each of which has been examined to identify diverse coffee types.

## Material and Methods

This section describes the research procedure, from image acquisition, labeling, preprocessing, image processing to evaluation. The CNN deep learning architecture modification was carried out to develop an identification model to determine the coffee type. The resultant model was then analyzed by looking at the resulting validation and calculating accuracy using the confusion matrix. In addition, the predicted accuracy was then compared against other architectures, namely LeNet, GoogleNet, AlexNet, and ResNet-50.

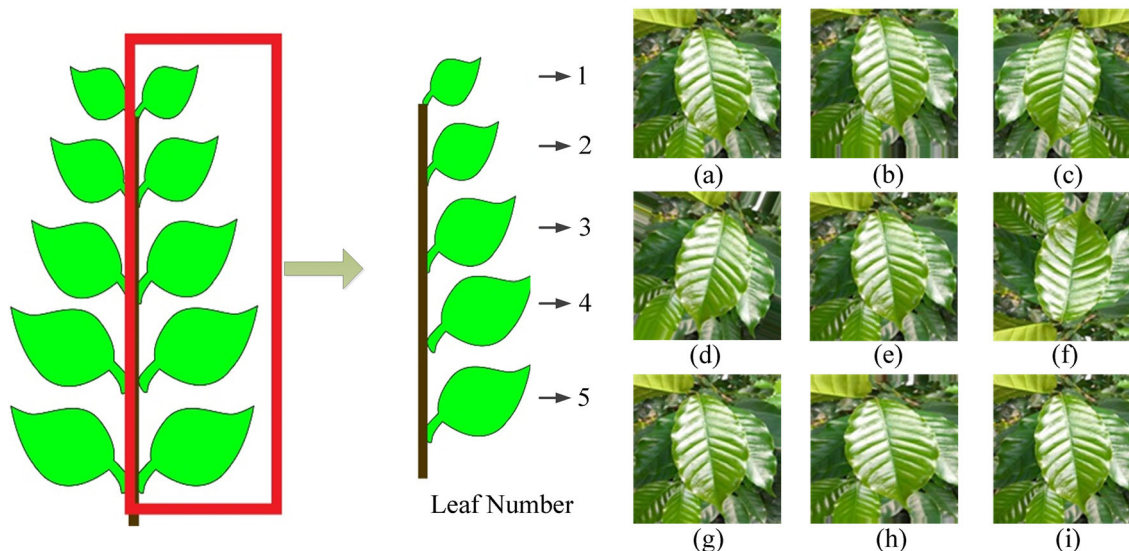
### Image Acquisition

This study used two research objects, namely Robusta and Arabica coffees, spread over several plantations in Jember and Bondowoso regencies, Indonesia. Information on the type of coffee was collected through interviews with farmers or smallholder farmers. Leaves samples were collected from Robusta and Arabica coffee plants without picking or separating the leaf from the branch. This mechanism is essential to avoid photosynthetic metabolic disorders in the coffee plants. Data were acquired using a smartphone with a 13 MP camera to capture the leaves on each coffee type. All of the images were converted to JPEG format and resized with a size ratio of 1:1.

Each coffee plant had multiple branches, and the leaves in each branch were captured using a smartphone camera. At least five leaves per branch were sequentially captured, starting from the topmost. The procedure of collecting leaves dataset is described in Fig. 1—left.

### Data Preparation

The datasets of Robusta and Arabica coffee leaves were collected from the coffee plantations belonging to smallholders. The collected images were renamed based on the type of coffee plant and sequentially given a specific number at the end. Then, each image resolution was reduced from  $3024 \times 3024$  pixels to  $224 \times 224$  pixels. Previous studies (Hao et al., 2020; Paymode & Malode, 2022; Sihalath et al., 2021) show that resizing the image to  $224 \times 224$  pixels adequately represents the essence of the original image. The reduction was



**Fig. 1.** Data acquisition and augmentation; (left) the procedure for collecting dataset of five coffee leaves in a branch; (right) the comparison of the augmentation process and its result; **a** no-augmentation; **b** height shift;

**c** horizontal alignment; **d** rotation; **e** shear; **f** vertical alignment; **g** width shift; **h** magnification; **i** fill mode

imperative because high-resolution images would be lengthy for training purposes (Öztürk & Akdemir, 2018). An image with a large size (dimension) has a complex structure and a lot of color information about the image texture, affecting the duration required for deep learning (Shi et al., 2017).

## Augmentation

Data augmentation is a technique of manipulating an image without losing the essence of those image data. Some of the augmentation techniques used include flipping, color space, and its transformation, cropping, and rotation (Shorten & Khoshgoftaar, 2019). A study conducted by Suhajito et al. (2021) used various data augmentation for increased accuracy of palm oil ripeness. Other studies used data augmentation for enhancing the identification of pests and diseases on coffee leaves (Tassis et al., 2021).

In this study, each image was augmented with different treatments using the tensorflow/keras preprocessing layers. The augmentation was used to enhance the classification results. This augmentation was done by changing or modifying the images so that the computer was able to detect different images. This process was conducted using the resized image, on which modifications were made using rotation, magnification, width shift, height shift, shear, horizontal alignment, and fill modes (Fig. 1—right).

The initial data included 2,000 leaf images (1,000 leaves for each coffee type) and were augmented, with total resultant data of 19,653 images. Adequate data can resolve overfitting problems and increase the classification result, so augmentation is the solution to this issue (Adrian, 2017; Shin et al., 2021; Thenmozhi & Srinivasulu Reddy, 2019). Another study

involving data augmentation on datasets contains less than 500 images (Su et al., 2021).

## Data Training and Validation

Data generated from the augmentation were divided by a ratio of 80:20. The number of images for training and validation was 15,984 and 2,396, respectively. Training data aimed to generate a model relevant to test data validation. Training data was taken randomly by including 80% of the total data, and the remaining 20% was entered into data testing. Data testing aimed to examine the model accuracy from the training results. The training model was devoted to generating the most accurate identification.

## Architectures

Convolutional neural networks are neural networks employing convolution operations as a layer. This neural network is devoted to processing two-dimensional data such as images and sound. According to Traore et al. (2018), convolution is a linear algebraic operation process that multiplies the matrix of image filter prior to further process. Convolution involves a special feature containing multiple layers. Each layer is interconnected with one another. There are three main layers at work: the convolution layer, pool layer, and fully connected layer. The input of CNN architecture was the image of coffee plants, and the output was the identified coffee type. This study used two-dimensional modified-CNN architecture consisting of four convolution layers, four pooling layers, and a fully connected layer (Fig. 2).

The convolution layers were done using (1) where the input image =  $r$ , the kernel =  $s$ . The indexes of rows and columns resulting from the matrix were denoted by  $p$  and  $q$ , respectively.

$$G[p, q] = (r \times s)[p, q] = \sum_t \sum_u s[t, u] r[p-t, q-u] \quad (1)$$

In the convolution layers, rectified linear unit (ReLU) was used to convert the summed weighted input from the node into the node's activation or output. ReLU function denoted by (2).

$$f(x) = \max(0, x) \quad (2)$$

## CNNs Modification

This study modified the CNN architecture using four convolution layers, four pooling layers, and a fully connected layer. The convolution result was a linear transformation of the input image according to the spatial information in the collected data. The data input was  $224 \times 224$  pixels with three image channels (red, green, blue). A series of receptive fields resulted from repeated application of the filter. Parameters changed to modify the properties of each layer were kernel size, stride, and padding.

The first convolution process used a  $3 \times 3$  kernel with 16 filters. Afterward, the ReLU function was activated, aiming to change the negative value to zero (removing negative values in a convoluted matrix). The convolution result had a size of  $224 \times 224$  because it used zero padding. The convolution output was then entered into the pooling process, and this process was used to reduce the matrix size. This study used max pooling to obtain the new matrix value from the pooling results. The result was a new matrix sized  $112 \times 112$  using  $2 \times 2$  kernel pooling. The max pooling was operated by considering the maximum value, which was 2, based on the kernel shift as much as the stride value.

The second convolution process dealing with the results of the first convolution with an input matrix of  $112 \times 112$  involved 32 filters and a kernel size of  $3 \times 3$ . The second convolution used the ReLU activation. Furthermore, a new matrix

was obtained in the pooling process using max pooling, which was  $56 \times 56$  in size.

The third convolution process used a  $56 \times 56$  input matrix with 64 filters and a  $3 \times 3$  kernel. Similar to the previous process, deploying the ReLU function and max pooling, the subsequent output was a new matrix measuring  $28 \times 28$ .

In the fourth convolution process, the input matrix was  $28 \times 28$ . It used 128 filters with a kernel size of  $3 \times 3$ . Next, ReLU activation function was operative, which proceeded to the max pooling process, with an output matrix of  $14 \times 14$ . The following figure displays the architecture of CNN modified (Fig. 3).

The model on CNN was modified to obtain the best classification accuracy and prediction. These modifications included the number of layers, epoch parameters, optimizer, dropout, and learning rate. The modified-CNN model had four layers with  $224 \times 224$  dimension input data. The epoch parameters or the number of repetitions during training used were 50, 75, and 100. The optimizer used in this study was the Adam optimizer. The optimizer denotes an algorithm for updating weights and biases in deep learning. A number of optimizer algorithms have been developed, including Adam (Kingma & Ba, 2014), RMSProp, NAG, AdaGrad, Adamax, and Nadam (Kandel et al., 2020). However, not all problems can be solved by relying on just one optimizer. Many studies (Pathan et al., 2021; Waheed et al., 2020) show that Adam optimizer performs better than other optimizers. Several parameters used for the modified-CNN were involved in this study. The architecture differentiation between CNN and modified-CNN is shown in Table 1.

## LeNet

LeNet is a simple architecture developed by Lecun et al. (1998). The LeNet architecture has three convolution layers, and two fully connected layers. The input image size for the LeNet architecture is  $28 \times 28$  pixels. The batch size used in this architecture is 64. We used the Adam optimizer for stochastic gradient descent (SGD), three convolution layers, two

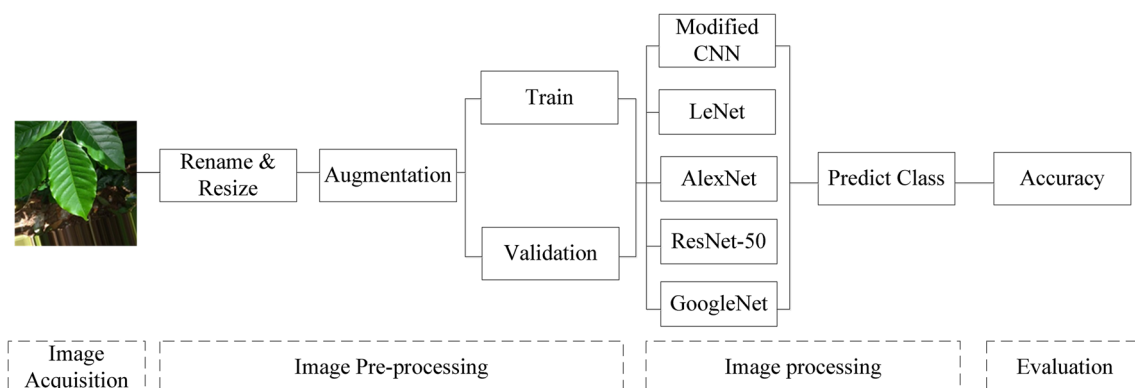
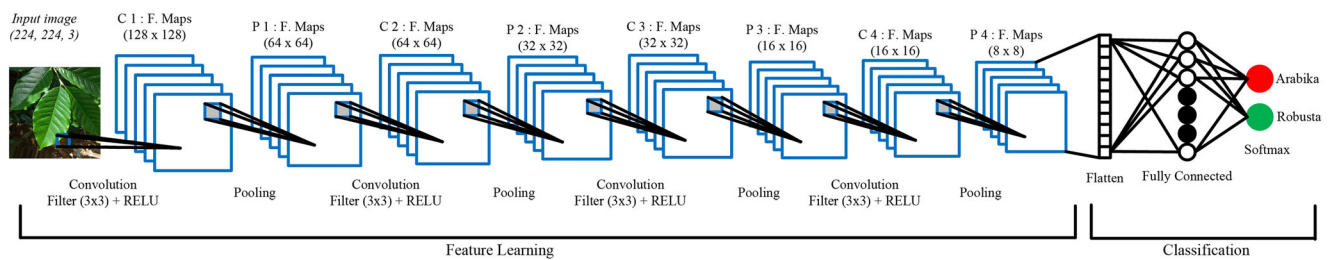


Fig. 2. Workflow of coffee identification using several deep learning architectures



**Fig. 3.** The architecture of the modified-CNN

pooling layers, and two fully connected layers during the deep learning training process.

## AlexNet

AlexNet is an extension of the LeNet architecture and published by Krizhevsky et al. (2012). This architecture has five convolution layers, three max pooling layers, and three fully connected layers. In this study, the input images size for AlexNet architecture was  $224 \times 224$  pixels. Several hyperparameters, namely regularization, batch size, optimizer, dropout, and Adam, were also incorporated in this architecture. Lastly, the fully connected layer was applied for performing binary classification.

## GoogleNet

This architecture is built from a network of modules or blocks. Since its emergence was emphasized by Szegedy et al. (2015), GoogleNet has become a differentiator in deep learning. The concept of GoogleNet is to use stacked modules or blocks. This study used a 22-layer deep convolutional network and a wider inceptions network. The input images for the GoogleNet architecture were  $224 \times 224$  pixels.

## ResNet-50

The ResNet-50 architecture has become popular with the concept of skip connections. Although not the first to use skip connections, it remains a popular architecture. In the study conducted by He et al. (2016), the architecture uses 152 layers without compromising its capability. In this study, the input

images for the ResNet-50 architecture were  $224 \times 224$  pixels with the batch size of 8.

## Parameters

The accuracy in identifying coffee plant species was examined using the confusion matrix method. This method accurately counts and correctly classifies images and incorrect images. According to Lopes et al. (2020), the accuracy performance is obtained by using (3).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (3)$$

$TP$  = true positives

$TN$  = true negatives

$FP$  = false positives

$FN$  = false negatives

On the other hand, other studies (Hasnain et al., 2020; Thenmozhi & Srinivasulu Reddy, 2019) describe a shorter accuracy equation, namely (4). In this study, this equation was used to evaluate the model accuracy.

$$Accuracy = \frac{Total\ True\ Prediction}{Total\ data} \quad (4)$$

In addition, another parameter, namely the receiver operating characteristic (ROC), was used to evaluate the model. ROC is a graphical plot consisting of true positive rate (TPR) and false-positive rate (TFR). These parameters are calculated using (5) and (6), respectively.

$$TPR = \frac{TP}{TP + FN} \quad (5)$$

**Table 1.** Architecture differentiation between CNN and modified-CNN

Parameter	CNN	Modified-CNN
Input image	$32 \times 32$	$224 \times 224$
Convolution layer	1	4
Regularization	Dropout	Dropout and data augmentation
Optimizer	Adam	Adam, learning rate 0.001, 0.0001, and 0.00001
Epoch	Any	Any (50, 75, and 100)

$$TFR = \frac{FP}{FP + TN} \quad (6)$$

## Results and Discussion

### Metrics and Environment Setup

The learning rate and epoch parameters determine the accuracy and tune during CNN training in deep learning. Using a higher epoch indicates that the data used has better accuracy to be implemented in data validation. Epoch 100 showed the best results in this study compared to Epoch 50 and Epoch 75 (Fig. 4).

Also, different learning rates can affect the duration of the training process. A smaller value of the learning rate is negatively correlated to the time required for the training process.

A Supercomputer NVIDIA Station DGX A100 with Ubuntu 20.04 and Docker container was used in this study. Hardware environments: CPU Single AMD 7742, 64 cores, 512G RAM; GPU 160 GB. Software environments: Python3.8, TensorFlow-GPU. 2.5, CUDA 11.2.

### The Training Accuracy of Modified-CNN Model

These hyperparameters are configurable and important use for the training of deep learning. The learning rate is important to evaluate how fast the model can identify and classify the coffee species based on the image data, while epoch indicates the time used to proceed whole training dataset of coffee leaf images using deep learning. There is no limit in the number of epochs used. The number of epochs can qualify for the training process if the accuracy and loss are stagnant. Studies conducted by Cruz Ulloa et al. (2022), Gan et al. (2021), and Wang et al. (2021) use 30, 60, and 100 epochs for training to produce the model, respectively.

The model resulting from the training process in each epoch produces a graph indicative of training accuracy and training loss. A good model possesses a high accuracy value and

minimum loss value. The training results presented in Fig. 5 show that the 50th, 75th, and 100th epochs increase training accuracy and decrease training loss. However, the learning rate of 0.0001 provides better accuracy and decreases the loss than other learning rates of 0.001 and 0.00001. In the learning rate of 0.0001, the training accuracy graph is relatively constant, and the training error appears consistently low. Thus, the learning rate of 0.0001 is used in the following evaluation.

### Performance of the Modified-CNN Model

The training model needs to be tested using data tests to evaluate the model's performance. The numbers of leaf images of different coffee types were assessed using (4). The correct identification (T) and incorrect identification (F) were used to determine the model accuracy. The higher T values in the identification imply the higher accuracy.

The model generated from several hyperparameters (learning rate and epoch) marks the best accuracy at a learning rate and epoch of 0.0001 and 100. Under these conditions, the model obtained from modified-CNN can correctly identify 1,987 Arabica coffee and 1,916 Robusta coffee with an exemplary accuracy of 97.67%. By contrast, a study conducted by Tassis et al. (2021) has successfully identified coffee pests and diseases using CNN based on leaves dataset.

### Comparison with the Well-Known Architectures

In this study, some developed architectures were incorporated to render a comprehensive comparison related to the classification of coffee types using deep learning. Both epoch and learning rate were also considered for evaluating the performance of the model using different architectures. The epochs and learning rate were 100 and 0.0001, respectively.

The results of modified-CNN were then compared against several well-known architectures such as LeNet, AlexNet, GoogleNet, and ResNet. The detailed performance of deep learning architectures is shown in Table 2. The test dataset used for each type of Arabica and Robusta coffee was 1998.

**Fig. 4.** **a** Accuracy deviation of each learning rate based on different epoch; **b** time performance of the learning rate based on different epoch

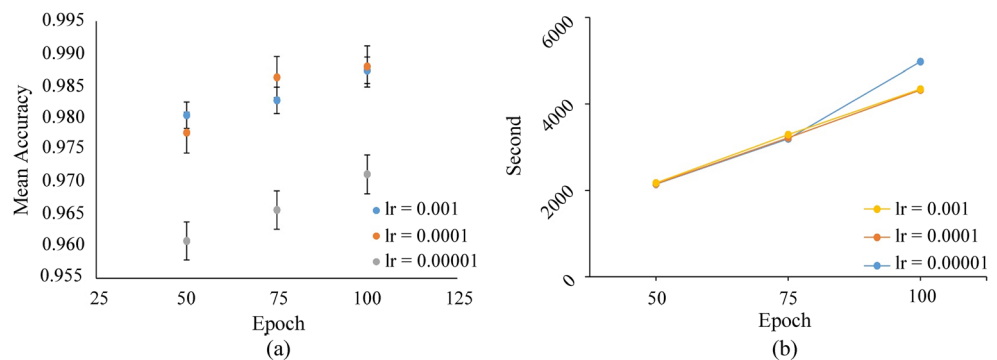
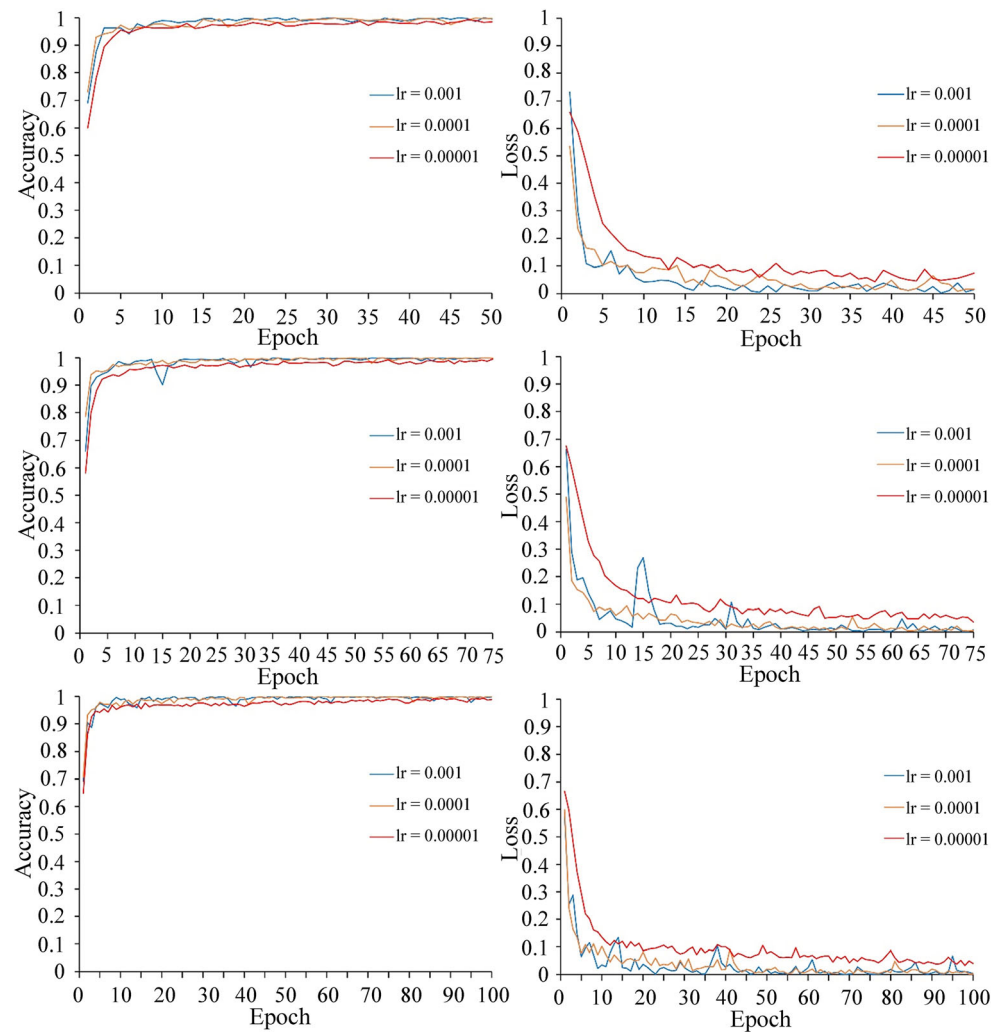


Fig. 5. Training model evaluation



From the existing model, modified-CNN had the least error in predicting Arabica and Robusta coffee. This can occur due to several factors, namely too homogeneous images, inappropriate augmentation results, and the inappropriate position of the leaves (e.g., too hidden position) or the unsuitable angle of the leaves (e.g., leaves are in wilting condition) when captured in the field making it difficult for the model to capture the relevant features.

**Table 2.** The performance of well-known deep learning architectures (learning rate 0.0001 and epoch 100)

Architecture	Arabica		Robusta		Accuracy (%)
	True	False	True	False	
Modified-CNN	1987	11	1916	82	97.67
AlexNet	1976	22	1824	174	95.10
LeNet	1881	117	1923	75	95.20
ResNet-50	1918	80	973	1025	72.35
GoogleNet	1960	38	1323	675	82.16

In addition, the modified-CNN, AlexNet, LeNet, ResNet-50, and GoogleNet architectures were compared based on their respective accuracy level. The comparison of the well-known architectures to the modified-CNN architecture highlights different performances in identifying Arabica and Robusta coffee. The accuracies of modified-CNN, AlexNet, LeNet, ResNet-50, and GoogleNet are 97.67%, 95.10%, 97.20%, 72.35%, and 82.16%, respectively. However, although the accuracies of AlexNet and GoogLeNet provide good accuracy, the graphs show fluctuation, while the ResNet-50 shows poor accuracy and high loss (Fig. 6).

The architecture that approximates the performance of a modified-CNN is the LeNet architecture. Several studies used the CNN modification and several well-known deep learning architectures for categorizing such agricultural objects such as pests (Thenmozhi & Srinivasulu Reddy, 2019), diseases (Ayan et al., 2020), and weeds (Jiang et al., 2020). A study conducted by Rauf et al. (2019) also compares the well-known deep learning architectures, namely GoogleNet, LeNet, AlexNet, and ResNet-50, with their CNN modification.

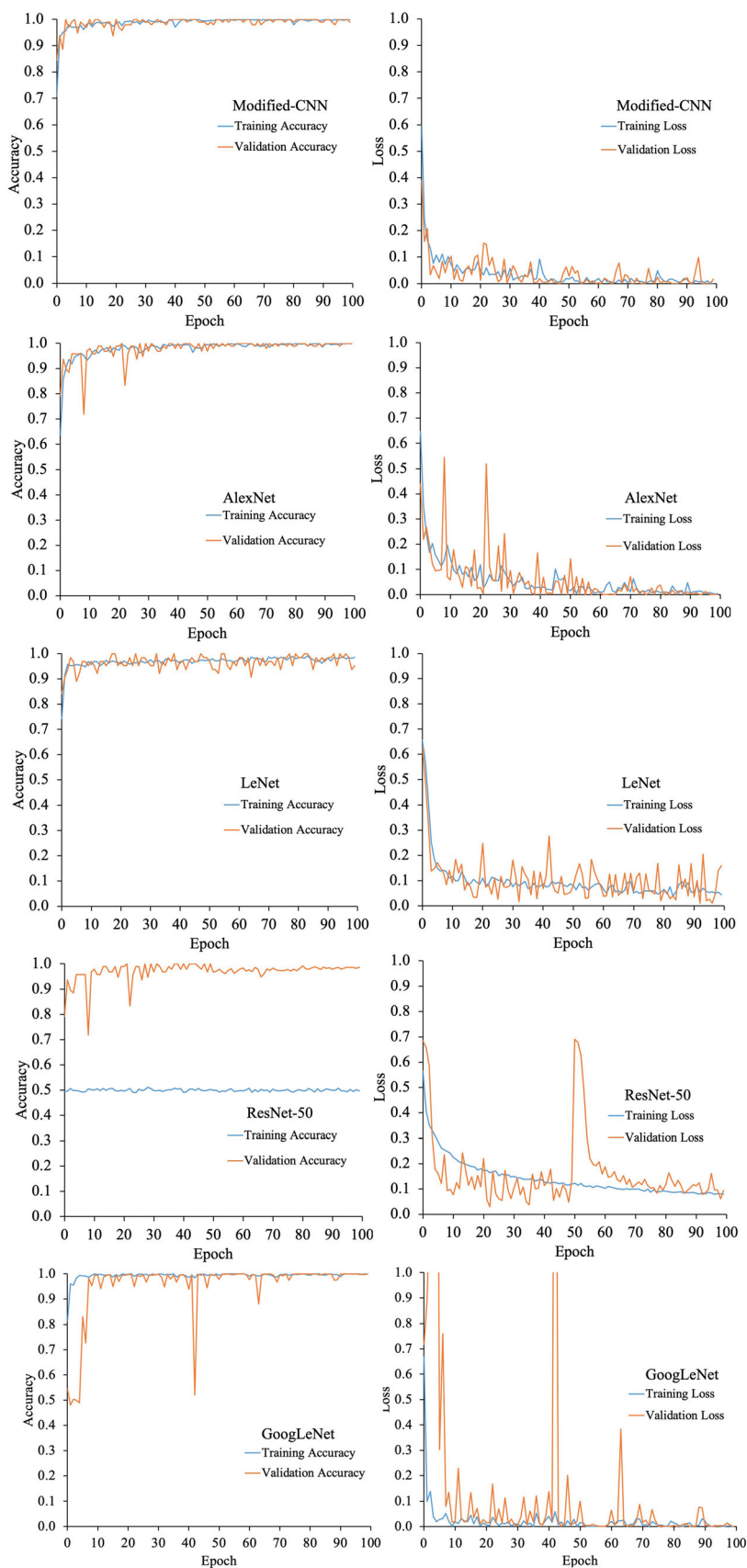
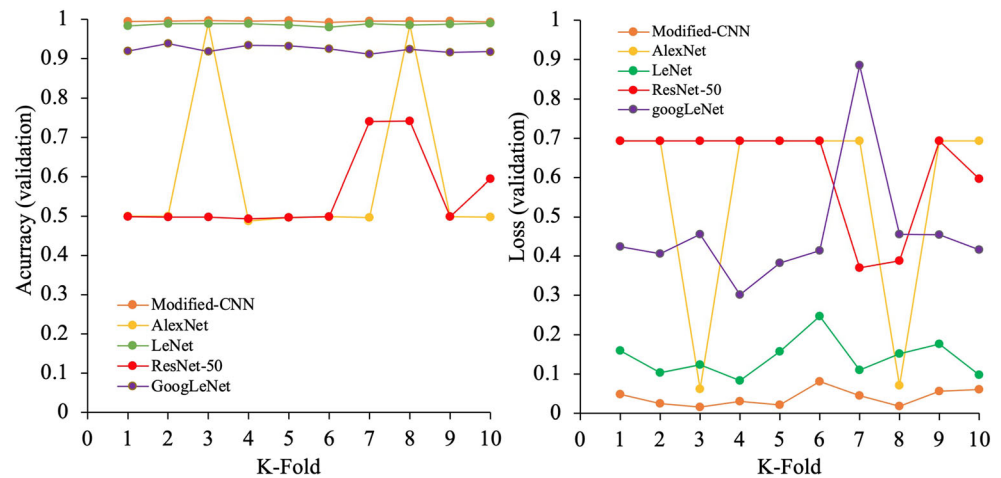


Fig. 6. Model performance of different architectures



**Fig. 7.** The accuracy and loss validation comparisons of architectures

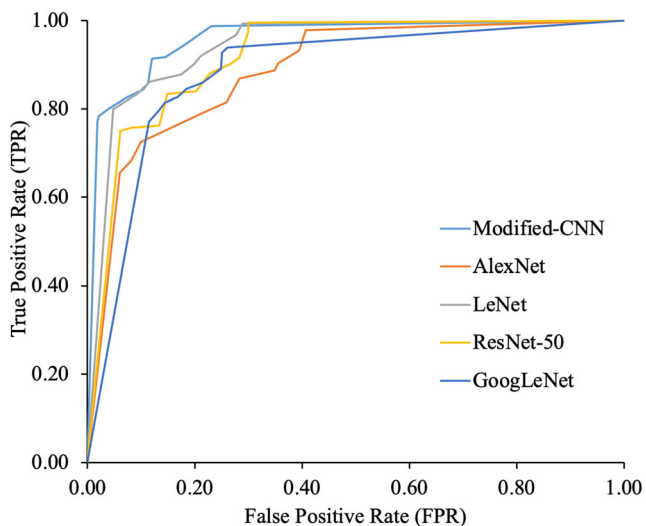


Although the results show that modified-CNN, AlexNet, and LeNet architectures perform better than ResNet-50 and GoogleNet architectures, further evaluation needs to be examined using K-fold cross-validation and ROC.

### K-Fold Cross-validation and Models Evaluation Using ROC

To corroborate the results obtained from the evaluation of performance architectures, analysis using K-fold cross-validation (CV) is required. K-fold cross-validation is used to evaluate the performance of architectures where the data are separated into two subsets, namely training data and validation data. In this study, a 10-fold CV was used to estimate the accuracy and choose the appropriate model to classify the Arabica and Robusta coffee species.

In a 10-fold CV, the data is split into tenfolds at the same size. This splitting process is done randomly by the system. Thus, the ten datasets were prepared to evaluate the



**Fig. 8.** The receiver characteristic curve (ROC) of different deep learning architectures

performance of the architectures. During CV process, each of the 10 datasets was split into 9-fold and 1-fold used for training and validation, respectively. The results show that the modified-CNN provides the highest validation accuracy and lowest loss, followed by LeNet. The accuracy of GoogLeNet produces good accuracy, but the validation loss is high, followed by AlexNet and ResNet-50 (Fig. 7).

To evaluate the model's performance using ROC, TPR and FPR of each model were plotted into a graph (Fig. 8). This ROC provides the performance of the proposed architectures model for classifying the Arabica and Robusta coffee leaves at all classification thresholds. The modified-CNN provides the highest TPR of all the other architecture models.

The modified-CNN architecture is better than other well-known architectures based on a series evaluation using different methods. In the modified-CNN, 4 convolution layers were used to classify Arabica and Robusta leaves. A study conducted by Geetharamani and J. (2019) used three convolution layers and provided satisfactory results in classifying plant leaf diseases. AlexNet used a deeper network using 5 convolution layers but provided low accuracy while tested using k-fold cross-validation and ROC. According to Chakraborty et al. (2018), CNN shows better object classification and recognition while used in deeper networks or more layers. The features can be further extracted using the deeper network.

### Limitation of Study

Although the modified-CNN provides the highest accuracy in this study, the model has a limited use to identify Arabica and Robusta coffee leaves only in the natural condition of Indonesia. The images were collected from coffee plants in natural conditions without picking or separating the leaf from the branch. Since the Arabica and Robusta coffee types are the most consumed and produced by most of the world's coffee-producing countries, another type of coffee, namely Liberica coffee, was not

included. This study is also limited in that the proposed architectures (modified-CNN, LeNet, AlexNet, ResNet-50, and GoogLeNet) and the particular hyperparameters (epoch and learning rate) were tested to the limited data.

## Conclusion and Direction for Future Works

Image processing architecture using deep learning will continue to develop along with the performance of computer technology. Deep learning can be utilized in the agricultural sector, particularly for classifying plant diseases and detecting weeds for counting fruit. This study proposes an identification mechanism for Arabica and Robusta coffee based on leaf images using modified-CNN. The underlying principles of such classification draw specific attention to the leaf shape, size, and color of Arabica and Robusta coffee. Several hyperparameters such as epoch and learning rate were taken into account where the use of 100 epoch and 0.0001 learning rate provided the highest accuracy. The test results showed an identification accuracy of 97.67% by using modified-CNN, higher than LeNet, AlexNet, ResNet-50, and GoogLeNet architectures. Also, 10-fold cross-validation and ROC were used for evaluating the proposed architectures. The results showed that modified-CNN provided the strongest accuracy followed by LeNet architectures.

By implication, the modified-CNN used in this study consisting of four convolution layers, four pooling layers, and a fully connected layer successfully identified coffee plant types based on leaf characteristics. Deeper convolution layers provided the benefit to extracting more features rather than the shallow ones. However, the proposed architecture has a limited use only for classifying Arabica and Robusta coffee types in the coffee plantations in Indonesia. For future works, this architecture with deeper layers is highly potential to be implemented on a micro-scale, such as classifying the coffee varieties in each Robusta and Arabica coffee type by capturing the image from above canopy measurement. Several varieties in Robusta and Arabica coffee types available in Indonesia are BP 42, BP 436, BP 409, BP 936, BP 939, S 795, Gayo 1, Andungsari 1, Andungsari 2K, Komasti, Sigararutung, and Gayo 2.

**Acknowledgements** We would like to extend our uttermost gratitude for the IT Center of the University of Jember for providing Supercomputer NVIDIA Station DGX A 100 for this project.

## Declarations

**Conflict of Interest** The authors declare no competing interests.

## References

- Aak. (1988). *Budidaya Tanaman Kopi*. Kanisius.
- Adrian, R. (2017). Deep learning for computer vision with python. PyImageSearch.
- Ak, A., Topuz, V., & Midi, I. (2022). Motor imagery EEG signal classification using image processing technique over GoogLeNet deep learning algorithm for controlling the robot manipulator. *Biomedical Signal Processing and Control*, 72, 1–10. <https://doi.org/10.1016/J.BSPC.2021.103295>
- Ayan, E., Erbay, H., & Varçın, F. (2020). Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks. *Computers and Electronics in Agriculture*, 179, 105809. <https://doi.org/10.1016/J.COMPAG.2020.105809>
- Barré, P., Stöver, B. C., Müller, K. F., & Steinhage, V. (2017). LeafNet: A computer vision system for automatic plant species identification. *Ecological Informatics*, 40, 50–56. <https://doi.org/10.1016/J.ECOINF.2017.05.005>
- Bhanumathi, S., Vineeth, M., & Rohit, N. (2019). Crop yield prediction and efficient use of fertilizers. In: *2019 International Conference on Communication and Signal Processing (ICCSPP)*, pp. 769–773, India. <https://doi.org/10.1109/ICCSPP.2019.8698087>
- Bolívar-Santamaría, S., & Reu, B. (2021). Detection and characterization of agroforestry systems in the Colombian Andes using sentinel-2 imagery. *Agroforestry Systems* 2021., 95(3), 499–514. <https://doi.org/10.1007/S10457-021-00597-8>
- Bzdok, D., Altman, N., & Krzywinski, M. (2018). Points of significance: Statistics versus machine learning. *Nature Methods* 2018, 15(4), 1–7. <https://doi.org/10.1038/nmeth.4642>
- Chakraborty, B., Shaw, B., Aich, J., Bhattacharya, U., & Parui, S. K. (2018). Does deeper network lead to better accuracy: A case study on handwritten devanagari characters. In *Proceedings-13th IAPR International Workshop on Document Analysis Systems-DAS 2018* (pp. 411–416). Vienna. <https://doi.org/10.1109/DAS.2018.72>
- Cruz Ulloa, C., Krus, A., Barrientos, A., del Cerro, J., & Valero, C. (2022). Robotic fertilization in strip cropping using a CNN vegetables detection-characterization method. *Computers and Electronics in Agriculture*, 193, 106684. <https://doi.org/10.1016/J.COMPAG.2022.106684>
- Ferreira, T., Shuler, J., Guimarães, R., & Farah, A. (2019). CHAPTER 1 introduction to coffee plant and genetics. In *Coffee: Production, Quality and Chemistry* (pp. 1–25). The Royal Society of Chemistry. <https://doi.org/10.1039/9781782622437-00001>
- Gan, H., Li, S., Ou, M., Yang, X., Huang, B., Liu, K., & Xue, Y. (2021). Fast and accurate detection of lactating sow nursing behavior with CNN-based optical flow and features. *Computers and Electronics in Agriculture*, 189, 106384. <https://doi.org/10.1016/J.COMPAG.2021.106384>
- Geetharamani, G., & J, A. P. (2019). Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Computers and Electrical Engineering*, 76, 323–338. <https://doi.org/10.1016/J.COMPELECENG.2019.04.011>
- Hao, X., Jia, J., Mateen Khattak, A., Zhang, L., Guo, X., Gao, W., & Wang, M. (2020). Growing period classification of Gynura bicolor DC using GL-CNN. *Computers and Electronics in Agriculture*, 174, 105497. <https://doi.org/10.1016/J.COMPAG.2020.105497>
- Hasan, A. S. M. M., Sohel, F., Diepeveen, D., Laga, H., & Jones, M. G. K. (2021). A survey of deep learning techniques for weed detection from images. *Computers and Electronics in Agriculture*, 184, 106067. <https://doi.org/10.1016/j.compag.2021.106067>
- Hasnain, M., Pasha, M. F., Ghani, I., Imran, M., Alzahrani, M. Y., & Budiarto, R. (2020). Evaluating trust prediction and confusion matrix measures for web services ranking. *IEEE Access*, 8, 90847–90861. <https://doi.org/10.1109/ACCESS.2020.2994222>

- 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. pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Jia, W., Tian, Y., Luo, R., Zhang, Z., Lian, J., & Zheng, Y. (2020). Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot. *Computers and Electronics in Agriculture*, 172, 105380. <https://doi.org/10.1016/J.COMPAG.2020.105380>
- Jiang, H., Zhang, C., Qiao, Y., Zhang, Z., Zhang, W., & Song, C. (2020). CNN feature based graph convolutional network for weed and crop recognition in smart farming. *Computers and Electronics in Agriculture*, 174, 105450. <https://doi.org/10.1016/J.COMPAG.2020.105450>
- Kandel, I., Castelli, M., & Popović, A. (2020). Comparative study of first order optimizers for image classification using convolutional neural networks on histopathology images. *Journal of Imaging*, 6(9), 1–17. <https://doi.org/10.3390/jimaging6090092>
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. . Cornell University. <https://arxiv.org/pdf/1412.6980>. Accessed 16 February 2021.
- Koklu, M., Cinar, I., & Taspinar, Y. S. (2021). Classification of rice varieties with deep learning methods. *Computers and Electronics in Agriculture*, 187, 106285. <https://doi.org/10.1016/J.COMPAG.2021.106285>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
- Kvam, J., & Kongsro, J. (2017). In vivo prediction of intramuscular fat using ultrasound and deep learning. *Computers and Electronics in Agriculture*, 142, 521–523. <https://doi.org/10.1016/J.COMPAG.2017.11.020>
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>
- Lewis, R. J. (2000). An introduction to classification and regression tree (CART) analysis. Resource document. Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.4103&rep=rep1&type=pdf>. Accessed 8 April 2021.
- Li, X., Ma, D., & Yin, B. (2021). Advance research in agricultural text-to-speech: The word segmentation of analytic language and the deep learning-based end-to-end system. *Computers and Electronics in Agriculture*, 180, 105908. <https://doi.org/10.1016/J.COMPAG.2020.105908>
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18, 3–29. <https://doi.org/10.3390/s18082674>
- Lopes, F., Agnelo, J., Teixeira, C. A., Laranjeiro, N., & Bernardino, J. (2020). Automating orthogonal defect classification using machine learning algorithms. *Future Generation Computer Systems*, 102, 932–947. <https://doi.org/10.1016/j.future.2019.09.009>
- Lu, J., Tan, L., & Jiang, H. (2021). Review on convolutional neural network (CNN) applied to plant leaf disease classification. *Agriculture*, 11(8), 2–18. <https://doi.org/10.3390/agriculture11080707>
- Osorio, K., Puerto, A., Pedraza, C., Jamaica, D., & Rodríguez, L. (2020). A deep learning approach for weed detection in lettuce crops using multispectral images. *AgriEngineering*, 2(3), 471–488. <https://doi.org/10.3390/agriengineering2030032>
- Öztürk, Ş., & Akdemir, B. (2018). Effects of histopathological image pre-processing on convolutional neural networks. *Procedia Computer Science*, 132, 396–403. <https://doi.org/10.1016/j.procs.2018.05.166>
- Pathan, S., Siddalingaswamy, P. C., Kumar, P., Pai M M, M., Ali, T., & Acharya, U. R. (2021). Novel ensemble of optimized CNN and dynamic selection techniques for accurate COVID-19 screening using chest CT images. *Computers in Biology and Medicine*, 137, 1–14. <https://doi.org/10.1016/J.COMPBIOMED.2021.104835>
- Paymode, A. S., & Malode, V. B. (2022). Transfer learning for multi-crop leaf disease image classification using convolutional neural networks VGG. *Artificial Intelligence in Agriculture*, 6, 1–11. <https://doi.org/10.1016/J.AIIA.2021.12.002>
- Rauf, H. T., Lali, M. I. U., Zahoor, S., Shah, S. Z. H., Rehman, A. U., & Bukhari, S. A. C. (2019). Visual features based automated identification of fish species using deep convolutional neural networks. *Computers and Electronics in Agriculture*, 167, 105075. <https://doi.org/10.1016/J.COMPAG.2019.105075>
- Shi, J., Wu, J., Li, Y., Zhang, Q., & Ying, S. (2017). Histopathological image classification with color pattern random binary hashing-based PCANet and matrix-form classifier. *IEEE Journal of Biomedical and Health Informatics*, 21(5), 1327–1337. <https://doi.org/10.1109/JBHI.2016.2602823>
- Shin, J., Chang, Y. K., Heung, B., Nguyen-Quang, T., Price, G. W., & Al-Mallahi, A. (2021). A deep learning approach for RGB image-based powdery mildew disease detection on strawberry leaves. *Computers and Electronics in Agriculture*, 183, 106042. <https://doi.org/10.1016/J.COMPAG.2021.106042>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1–48. <https://doi.org/10.1186/S40537-019-0197-0>
- Sihalath, T., Basak, J. K., Bhujel, A., Arulmozhi, E., Moon, B. E., & Kim, H. T. (2021). Pig identification using deep convolutional neural network based on different age range. *Journal of Biosystems Engineering*, 46(2), 182–195. <https://doi.org/10.1007/S42853-021-00098-7>
- Su, D., Kong, H., Qiao, Y., & Sukkariyah, S. (2021). Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. *Computers and Electronics in Agriculture*, 190, 106418. <https://doi.org/10.1016/J.COMPAG.2021.106418>
- Suharjito, Elwirehardja, G. N., & Prayoga, J. S. (2021). Oil palm fresh fruit bunch ripeness classification on mobile devices using deep learning approaches. *Computers and Electronics in Agriculture*, 188, 106359. <https://doi.org/10.1016/J.COMPAG.2021.106359>
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA: June 2015. <https://doi.org/10.1109/CVPR.2015.7298594>
- Tassis, L. M., Tozzi de Souza, J. E., & Krohling, R. A. (2021). A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images. *Computers and Electronics in Agriculture*, 186, 106191. <https://doi.org/10.1016/J.COMPAG.2021.106191>
- Thenmozhi, K., & Srinivasulu Reddy, U. (2019). Crop pest classification based on deep convolutional neural network and transfer learning. *Computers and Electronics in Agriculture*, 164, 104906. <https://doi.org/10.1016/J.COMPAG.2019.104906>
- Traore, B. B., Kamsu-Foguem, B., & Tangara, F. (2018). Deep convolutional neural network for image recognition. *Ecological Informatics*, 48, 257–268. <https://doi.org/10.1016/J.ECOINF.2018.10.002>

- Uyeh, D. D., Bassey, B. I., Mallipeddi, R., Asem-Hiablie, S., Amaizu, M., Woo, S., Ha, Y., & Park, T. (2021). A reinforcement learning approach for optimal placement of sensors in protected cultivation systems. *IEEE Access*, *9*, 100781–100800. <https://doi.org/10.1109/ACCESS.2021.3096828>
- Waheed, A., Goyal, M., Gupta, D., Khanna, A., Hassanien, A. E., & Pandey, H. M. (2020). An optimized dense convolutional neural network model for disease recognition and classification in corn leaf. *Computers and Electronics in Agriculture*, *175*, 105456. <https://doi.org/10.1016/J.COMPAG.2020.105456>
- Wang, D., Wang, J., Li, W., & Guan, P. (2021). T-CNN: Trilinear convolutional neural networks model for visual detection of plant diseases. *Computers and Electronics in Agriculture*, *190*, 106468. <https://doi.org/10.1016/J.COMPAG.2021.106468>