

Fruit and Vegetable Classification using Convolutional Neural Network with MobileNetV2

Muhammad Khoiruddin¹, Silvester Tena²(✉)

¹ Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia

² Department of Electrical Engineering, Faculty of Science and Engineering, Universitas Nusa Cendana, Kupang, Indonesia
siltena@staf.undana.ac.id, muhammadkhoiruddin@mail.ugm.ac.id

Article Info

Article history:

Received November 21, 2024

Revised December 28, 2024

Accepted December 29, 2024

Keywords:

Fruit

Vegetable

Classification

CNN

MobileNetV2

ABSTRACT

Fruits are parts of plants that originate from the plant's pistils and usually contain seeds. Meanwhile, vegetables are leaves, legumes, or seeds that can be cooked. Fruits and vegetables have many variations that can be distinguished based on color, shape, and texture. However, the development of Artificial Intelligence (AI) technology has become pervasive in everyday life, one aspect of which is demonstrated through deep learning, a method of AI learning. Therefore, developing deep learning for tasks such as automatically detecting surrounding objects is necessary. This study aims to classify types of fruits and vegetables by applying a Convolutional Neural Network (CNN) with the MobileNetV2 architecture. In this study, fruits and vegetables encompassing 36 categories, including significant types in daily life, were considered. The results show that the classification system achieved an excellent accuracy rate of 97.31%, demonstrating the effectiveness of using deep learning techniques for this application.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Silvester Tena

Department of Electrical Engineering, Faculty of Science and Engineering, Universitas Nusa Cendana, Kupang, Indonesia

Email: siltena@staf.undana.ac.id

1. INTRODUCTION

Classifying fruits and vegetables remains a challenge in image recognition within computer vision. Fruits and vegetables often have similar colors, shapes, and textures. Much time is spent labelling or counting various fruits and vegetables in production and sales, yet these processes need more automation. Moreover, there is a pursuit of higher quality and efficiency. Therefore, researchers are designing a high-accuracy classification system aimed at fruits and vegetables that minimizes time consumption, enabling quick, precise, and accurate identification of various types of fruits and vegetables.

Deep Learning is a class of machine learning algorithms that uses multiple layers of algorithms containing nonlinear processing units [1]. Deep Learning also enables computers to learn tasks that mimic human capabilities. This technology supports the development of artificial intelligence (AI), and its progress has led to advancements in various fields [2]. One application of deep learning is in image processing or digital image processing. The purpose of an image processing system is to assist humans in recognizing or classifying objects efficiently, quickly, accurately, and capable of simultaneously processing large amounts of data [3].

Convolutional Neural Network (CNN) is a type of Deep Learning algorithm that can take images as input, identify various aspects or objects within the images that a machine can "learn" to recognize, and distinguish one image from another. The architecture of CNN is similar to the pattern of connections between neurons or nerve cells in the human brain. CNN is inspired by the visual cortex, the part of the brain responsible

for processing visual information. In CNN, the data propagated through the network is two-dimensional, making the linear operations and weight parameters different. In CNN, linear operations use convolution, and the weights are not one-dimensional but four-dimensional, consisting of convolution kernels. In this research, Deep Learning using the CNN method will be implemented to accurately and precisely detect different types of fruits and vegetables [3]. The research was conducted to identify and classify local fruits using deep CNN. Several deep learning models were used, and MobileNet achieved the best results in feature extraction compared to the other models [4][5].

This research aims to classify fruits and vegetables, identifying similarities and differences between the two. The user requirement in this research is to provide a method that enables users to utilize the research results to distinguish between different types of fruits and vegetables. One of the methods used in this study is the Convolutional Neural Network (CNN). The deep learning model used in this study is MobileNetV2, as it has achieved high accuracy based on previous research [6]. However, in this study, the maximum number of epochs is 10, so the accuracy has not reached its full potential.

The remaining parts of this paper are divided into various parts, including Section 2, which explains the materials and research methods. Section 3 describes the research results and discussion. Finally, Section 4 concludes the paper.

2. MATERIAL AND RESEARCH METHOD

The methodology used for this task includes, followed by the specification and implementation of the system. Figure 1 shows the research framework for fruit and vegetable classification using the CNN method. This paper uses a MobileNetV2 pre-trained model for image feature extraction and classification. The classification process begins with a training dataset. The next stage is to use the prepared test dataset to test the trained model.

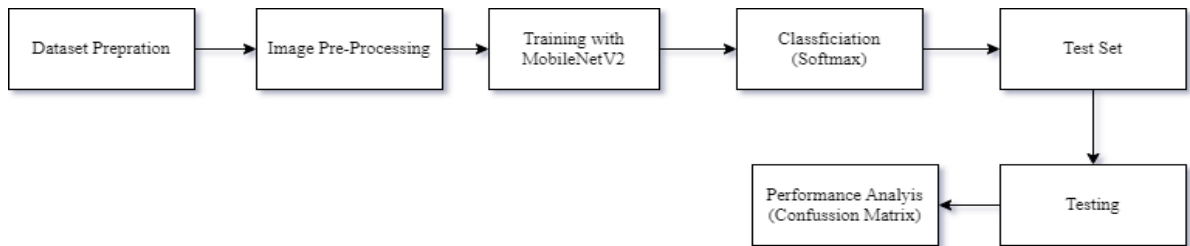


Figure 1. Research Framework

2.1 Dataset

The dataset used in this research consists of fruits and vegetables obtained from the Kaggle dataset "fruits_classification" by SHIVAGOWRI_19. All images in the fruits and vegetables dataset are 224x224 pixels, totalling 3,825 images. The fruit and vegetable dataset contains a total of 36 classes. The dataset is divided into 3,115 training images, 351 testing images, and 359 validation images, as shown in table 1.

Table 1. Dataset of fruits and vegetables

Types of Fruits	Training Data	Validation Data	Test Data
Apple	68	10	10
Banana	75	9	9
Beetroot	88	10	10
Bell pepper	90	9	10
Cabbage	92	10	10
Capsicum	89	10	10
Carrot	82	9	10
Cauliflower	79	10	10
Chili Pepper	87	9	10
Corn	87	10	10
Cucumber	94	10	10
Eggplant	84	10	10
Garlic	92	10	10
Ginger	68	10	10
Grapes	100	9	10
Jelepeno	88	9	10

Types of Fruits	Training Data	Validation Data	Test Data
Kiwi	88	10	10
Lemon	82	10	10
Lettuce	97	9	10
Mango	86	10	10
Onion	94	10	10
Orange	69	9	10
Paprika	83	10	10
Pear	89	10	10
Peas	100	10	10
Pineapple	99	10	10
Pomegranate	79	10	10
Potato	77	10	10
Raddish	81	9	10
Soy beans	97	10	10
Spinach	97	10	10
Sweetcorn	91	10	10
Sweetpotato	69	10	10
Tomato	92	10	10
Turnip	98	10	10
Watermelon	84	10	10
Total of Data	3,115	351	359

2.2 Preprocessing

The preprocessing process is intended to prepare the data [7]. The preprocessing steps include adjusting the input data size, color mode, rotation range, and zoom range. Another goal of preprocessing is to demonstrate the impact of the data on accuracy and to use techniques such as resizing the images to 224x224 pixels with a channel depth of 3, indicating that the input images are in RGB (Red, Green, Blue) color.

2.3 Learning Process

The learning process, also known as training, is intended to train the designed CNN model to understand and differentiate between images of fruits and vegetables that have been indexed according to their classes [2][8][9][10]. The learning process is conducted using several epoch categories, namely 5, 10, 15, and 20 epochs. The CNN model that produces a high accuracy index for classifying fruit and vegetable images will be selected. The target variables used in this learning process aim for an accuracy of > 0.999 and a loss of < 0.05 [8][11].

2.4 MobileNetV2 Architecture

The Convolutional Neural Network(CNN) in this research is intended to classify images of fruits and vegetables. Unlike traditional classification algorithms, which typically perform feature extraction and image classification as separate processes, the CNN algorithm integrates feature extraction into the learning process. In other words, the feature extraction in the CNN algorithm is also part of the learning process [7]. The CNN model is divided into two sections: the convolution layer (CL) at the front and the fully connected layer (FCL) at the back [12].

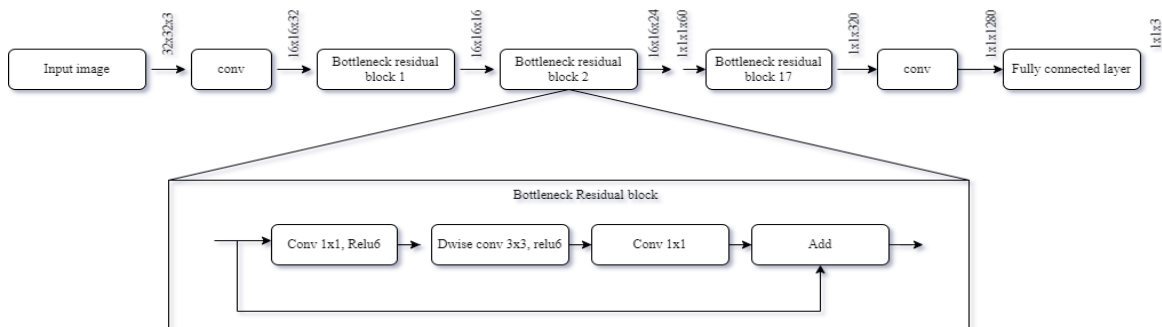


Figure 2. MobileNetV2 architecture

The general trend in creating CNN models is that an increasing number of layers tends to result in higher accuracy. This trend is supported by analysis results from the ImageNet Large Scale Visual Recognition Challenges (ILSVRC) from 2010 to 2015 [13][14]. The number of layers has almost always increased, with models such as ResNet having more than 100 layers. However, increasing the number of layers does not necessarily make the model more efficient in size and speed, which is a significant consideration for real-world applications such as robotics. In applications like self-driving cars and augmented reality, efficiency is crucial due to the limited computational capabilities of the devices. MobileNet is an architecture designed for efficiency, utilizing two sets of hyper-parameters to build a tiny model with low latency, making it suitable for mobile and embedded applications. MobileNet is based on depthwise separable convolutions to reduce computation in the early layers [15][16]. The research conducted by [13] demonstrated that the application of the CNN method with the MobileNet architecture to classify seven types of diseases using an Android smartphone achieved an accuracy of 84.28%. This was accomplished by utilizing the built-in under-sampling and preprocessing methods of MobileNet. The highest accuracy of 96.4% was achieved by applying oversampling techniques and augmenting the input data [17][18][19].

2.5 Accuracy Calculation

The accuracy calculation process in this research represents a variable used to assess the performance and measure the success of the CNN model in classifying images of fruits and vegetables. The equation used to calculate accuracy is shown in Equation (1)[20].

$$Accuracy = \frac{\text{number of correct predictions}}{\text{Number of test Data}} \times 100\% \quad (1)$$

2.6 System Design

The CNN method requires preprocessing steps to prepare the data before training, including reshaping the data into a three-dimensional array of 224x224x3. This means that each dimension of the array contains the RGB color values of the image, with a range of 0-255. These values are then rescaled to a range of 0-1. The CNN algorithm model is implemented using Python and libraries such as Numpy, Pandas, and Os. Path, Matplotlib, TensorFlow, and Keras. The types of devices used to perform the algorithm training are as follows: MSI GF63 Thin SCR Laptop, Processor Intel Core i5 10200H – 10 Th Gen, Random Access Memory (RAM) 16 GB DDR4, Video Graphics Adapter (VGA) NVIDIA Geforce GTX 1650 Ti 4 GB DDR6. The architecture used in this research is MobileNet V2, which has a depth of 28 layers and utilizes Average Pooling. The input consists of images in RGB mode, which are resized to a dimension of 224x224 pixels, resulting in a three-dimensional array, with each dimension representing the image's red, green, and blue color values. The purpose of the training process is to develop an efficient and reliable model for the task of fruit and vegetable recognition.



Figure 3. Training process flow

3. RESULTS AND DISCUSSION

3.1 Training Result

The data collection and research using Convolutional Neural Network (CNN) aim to apply the MobileNet architecture model to digital image data of fruits and vegetables [21]. The images have dimensions of 224x224 pixels, and each dimension has a different impact on the classification of the fruits and vegetables. During the training process, four trials were conducted with adjustments to the epoch and learning rate parameters. The resulting image data is shown in Figure 4.



Figure 4. Dataset image after testing

The accuracy training and validation results for each epoch is shown in Figure 5. The best training performance for the MobileNetV2 architecture was achieved at epochs 15; after prediction, many results were accurate, such as carrots being predicted as carrots, pineapples as pineapples, garlic as garlic, grapes as grapes, watermelons as watermelons, soybeans as soybeans, tomatoes as tomatoes, corn as corn, cabbage leaves as cabbage leaves, bell peppers as bell peppers, pomegranates as pomegranates, and cabbage as cabbage. A snapshot of the accuracy achieved by the CNN model during the learning process is shown in Figure 5. The accuracy shown in Figure 5 is for a maximum epoch of 10. In this study, the maximum epoch is 20. Loss is a function that describes the error associated with the probabilities generated by the CNN model.

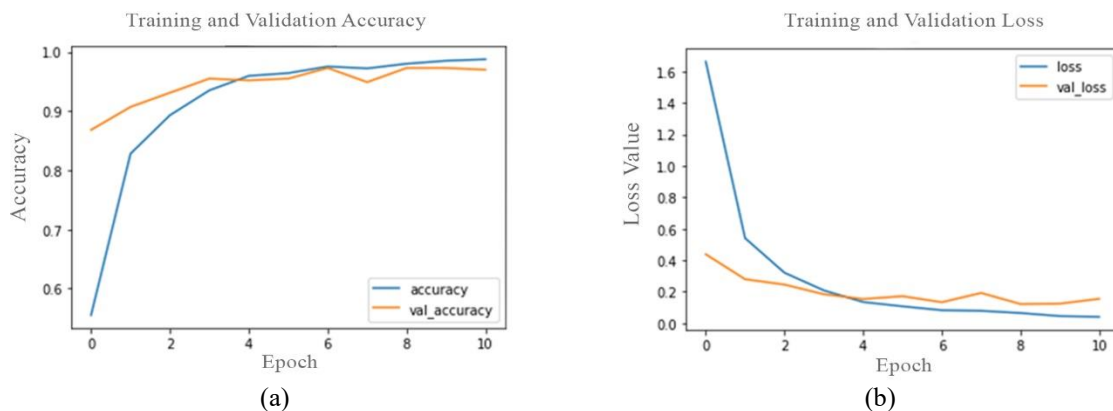


Figure 5. (a) Accuracy Training and validation, (b) Loss Training and validation

Figure 6 presents the classification results of the model for the test data. Each type of fruit and vegetable was used for testing. Based on the testing process, the MobileNetV2 model successfully classified the fruits and vegetables. According to the calculations from the equation (1), the accuracy achieved by the CNN model with the MobileNet architecture is 97.31% . The highest classification accuracy value is at epoch 15. This means that at epoch 15, the model provided good results. The number of epochs increases performance. However, at certain epochs, it will experience saturation. Therefore, even if the epoch value is increased, it will not affect accuracy, and the accuracy value will decrease.

The results are quite promising, showing six images with strong correlations. The system predicts the apple as chilli, the banana as chilli, the capsicum as paprika, the corn as sweet corn, the potato as pear, and the sweet corn as corn. One of the causes of misclassification is class imbalance in the dataset. For example, the apple class may be predicted as chilli because the training data for apples is more abundant than for chilli. As a result, the model tends to predict the more dominant and similar class. However, in this study, only a few classes were misclassified due to the advantage of the CNN method in extracting object features. The advantage of CNNs in feature extraction from fruits lies in their ability to manage variations in data. Images of fruits can exhibit significant variability in size, shape, color, and texture. The CNN method has the advantage of image feature extraction, which results in optimal model performance [24][25]. CNNs effectively handle this variability through deep learning processes, enabling them to recognize fruits even when substantial visual differences exist among individual samples. Moreover, CNNs can accurately identify and classify fruits with different variations from the training dataset, enhancing overall prediction accuracy.

4. CONCLUSION

Based on the testing results, the Convolutional Neural Network (CNN) application with the MobileNetV2 architecture for identifying fruits and vegetables achieved an accuracy of 97.31%. There were classification correlations where the system predicted the apple as chilli, the banana as chilli, the capsicum as paprika, the corn as sweet corn, the potato as pear, and the sweet corn as corn. Using features such as color, shape, and size of fruits and vegetables has proven optimal. The fruit and vegetable recognition system can also handle a wider variety of fruits and vegetables, which can be used to recognize, identify, classify, and count the number of fruits and vegetables in a digital image.

REFERENCES

- [1] H. Mureşan and M. Oltean, "Fruit recognition from images using deep learning," *Acta Univ. Sapientiae, Inform.*, vol. 10, no. 1, pp. 26–42, 2018, doi: 10.2478/ausi-2018-0002.
- [2] D. Mirwansyah and Arief Wibowo, "Fruit Image Classification Using Deep Learning Algorithm: Systematic Literature Review (Slr)," *Multica Sci. Technol. J.*, vol. 2, no. 2, pp. 120–123, 2022, doi: 10.47002/mst.v2i2.356.
- [3] F. F. Maulana and N. Rochmawati, "Klasifikasi Citra Buah Menggunakan Convolutional Neural Network," *J. Informatics Comput. Sci.*, vol. 1, no. 02, pp. 104–108, 2020, doi: 10.26740/jinacs.v1n02.p104-108.
- [4] M. M. Rahman, M. A. Basar, T. S. Shinti, M. S. I. Khan, H. M. H. Babu, and K. M. M. Uddin, "A deep CNN approach to detect and classify local fruits through a web interface," *Smart Agric. Technol.*, vol. 5, no. July, p. 100321, 2023, doi: 10.1016/j.atech.2023.100321.
- [5] M. F. Nadhif and S. Dwiasnati, "Classification of Date Fruit Types Using CNN Algorithm Based on Type," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 36–42, 2023, doi: 10.57152/malcom.v3i1.724.
- [6] G. Kaur, N. Sharma, R. Chauhan, H. S. Pokhariya, and R. Gupta, "Fruit and Vegetable Classification Using MobileNet V2 Transfer Learning Model," *2023 3rd Int. Conf. Smart Gener. Comput. Commun. Networking, SMART GENCON 2023*, pp. 1–6, 2023, doi: 10.1109/SMARTGENCON60755.2023.10442618.
- [7] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, 2018, doi: 10.1016/j.patcog.2017.10.013.
- [8] F. A. Azis, H. Suhaimi, and E. Abas, "Waste Classification using Convolutional Neural Network," *ACM Int. Conf. Proceeding Ser.*, no. July, pp. 9–13, 2020, doi: 10.1145/3417473.3417474.
- [9] K. Hameed, D. Chai, and A. Rassau, "A sample weight and adaboost CNN-based coarse to fine classification of fruit and vegetables at a supermarket self-checkout," *Appl. Sci.*, vol. 10, no. 23, pp. 1–18, 2020, doi: 10.3390/app10238667.
- [10] M. T. Vasumathi and M. Kamarasan, "An lstm based cnn model for pomegranate fruit classification with weight optimization using dragonfly technique," *Indian J. Comput. Sci. Eng.*, vol. 12, no. 2, pp. 371–384, 2021, doi: 10.21817/indjcs/2021/v12i2/211202051.
- [11] S. S. S. Palakodati, V. R. R. Chirra, Y. Dasari, and S. Bulla, "Fresh and rotten fruits classification using CNN and transfer learning," *Rev. d'Intelligence Artif.*, vol. 34, no. 5, pp. 617–622, 2020, doi: 10.18280/ria.340512.
- [12] S. Tena, R. Hartanto, and I. Ardiyanto, "Content-Based Image Retrieval for Traditional Indonesian Woven Fabric Images Using a Modified Convolutional Neural Network Method," *J. Imaging*, vol. 9, no. 8, 2023, doi: 10.3390/jimaging9080165.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.
- [14] C. Szegedy *et al.*, "Going deeper with convolutions," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 1–9, 2015, doi: 10.1109/CVPR.2015.7298594.
- [15] T. Turap, T. B. Merupakan, T. B. Lebih, and T. D. Turap, "object recognition with gradient based learning," vol. 1, pp. 1–17.
- [16] Dwi Fitriana Sari and D. Swanjaya, "Implementasi Convolutional Neural Network Untuk Identifikasi Penyakit Daun Gambas," *Semin. Nas. Inov. Teknol.*, vol. 04, no. 03, pp. 137–142, 2020.
- [17] A. G. Howard *et al.*, "MobileNets: efficient convolutional neural networks for mobile vision applications," Apr.

Fruit and Vegetable Classification using Convolutional Neural Network with MobileNetV2 ... (Muhammad Khoiruddin)

- 2017, doi: 10.48550/arXiv.1704.04861.
- [18] J. Velasco *et al.*, "A smartphone-based skin disease classification using mobilenet CNN," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 5, pp. 2632–2637, 2019, doi: 10.30534/ijatcse/2019/116852019.
- [19] U. Seidaliyeva, D. Akhmetov, L. Ilipbayeva, and E. T. Matson, "Real-time and accurate drone detection in a video with a static background," *Sensors (Switzerland)*, vol. 20, no. 14, pp. 1–18, 2020, doi: 10.3390/s20143856.
- [20] S. Khan, H. Rahmani, S. A. A. Shah, and M. Bennamoun, "A Guide to Convolutional Neural Networks for Computer Vision," *Synth. Lect. Comput. Vis.*, vol. 8, no. 1, pp. 1–207, Feb. 2018, doi: 10.2200/S00822ED1V01Y201712COV015.
- [21] S. Sari, I. Soesanti, and N. A. Setiawan, "Best Performance Comparative Analysis of Architecture Deep Learning on CT Images for Lung Nodules Classification," *Proc. - 2021 IEEE 5th Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. Appl. Data Sci. Artif. Intell. Technol. Glob. Challenges Dur. Pandemic Era, ICITISEE 2021*, pp. 138–143, 2021, doi: 10.1109/ICITISEE53823.2021.9655872.
- [22] S. Farhadpour, T. A. Warner, and A. E. Maxwell, "Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices," *Remote Sens.*, vol. 16, no. 3, pp. 1–22, 2024, doi: 10.3390/rs16030533.
- [23] S. Tena and B. Y. Dwiandiyanta, "Transforming Woven Ikat Fabric : Advanced Classification Via Transfer Learning and Convolutional Neural Networks," vol. XII, no. 2, pp. 73–82, 2023, doi: 10.35508/jme.v12i2.12579.
- [24] H. Prasetyo and B. A. Putra Akardihas, "Batik image retrieval using convolutional neural network," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 17, no. 6, p. 3010, Dec. 2019, doi: 10.12928/telkomnika.v17i6.12701.
- [25] S. Tena, R. Hartanto, and I. Ardiyanto, "Content-based image retrieval for fabric images: A survey," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 23, no. 3, p. 1861, Sep. 2021, doi: 10.11591/ijeecs.v23.i3.pp1861-1872.