International Journal on Advanced Science Engineering Information Technology

Vol.15 (2025) No. 2
ISSN: 2088-5334

Volume Prediction of Axisymmetric Fruits Using Optimized Convolutional Neural Networks with Transfer Learning Strategy

Vincentius Christian Wariky [a], Joko Siswantoro [a,*], Njoto Benarkah [a]

[a] *Department of Informatics Engineering, University of Surabaya, Surabaya, Indonesia*
Corresponding author: [*]joko_siswantoro@staff.ubaya.ac.id

***Abstract*—Accurate volume measurement of fruits is key in the agricultural and food industries, supporting better logistics, quality assessment, and processing decisions. However, traditional methods for measuring volume are often manual and labor-intensive, creating bottlenecks in high-scale operations. This study presents a novel approach that utilizes convolutional neural networks (CNNs) combined with a transfer learning strategy to predict the volume of axisymmetric fruit from images, offering a more automated and efficient solution. To achieve this, five pretrained CNN architectures, including MobileNetV2, VGG-16, DenseNet201, ResNet50, and EfficientNetV2B0, were employed by modifying the fully connected layers and optimized through a random search process, allowing for optimal hyperparameter selection. Only the fully connected layers were fine-tuned, while the pretrained convolutional layers retained their original weights, enabling the models to focus on relevant image features without extensive retraining. The methodology encompassed dataset creation, image preprocessing, and segmentation, with training supported by the Adam optimizer and evaluated using mean squared error and mean absolute error. The performance of CNNs was assessed through metrics like mean absolute relative error (ARE) and the coefficient of determination (R^2). Experimental results demonstrate that ResNet50 achieved the highest prediction accuracy with a mean ARE of 3.76% and an R^2 of 0.9721, outperforming other models and several existing methods from previous research. This study's findings highlight the potential of CNN-based models, especially ResNet50, for precise axisymmetric fruit volume estimation. Future research may extend this method to encompass diverse fruit types and real-time applications, advancing automated processing technologies in the industry.**

***Keywords*— Convolutional neural networks; fruit volume prediction; hyperparameter optimization; ResNet50.**

Manuscript received 15 Nov. 2024; revised 20 Jan. 2025; accepted 9 Apr. 2025. Date of publication 30 Apr. 2025.

I. INTRODUCTION

The accurate measurement of fruit volume is essential in various industries, including agriculture, food processing, and retail, as it plays a crucial role in determining product quality grading, pricing, transporting, and packaging requirements [1]–[6]. Traditional methods, such as water displacement, are often time-consuming, prone to error for porous objects, and impractical for large-scale operations [2], [7], [8]. With the advancement of image processing and machine learning technologies, automated volume prediction using image-based methods has emerged as a promising alternative, offering higher accuracy, efficiency, and scalability [9] .

Previous studies on fruit volume prediction based on image can be categorized into two main approaches: 2D and 3D. The 2D approach typically utilized a single image taken from one viewpoint, where geometric features such as length, width, or perimeter of the object were measured to estimate the volume [2]. This approach was commonly used for axisymmetric objects. Axisymmetric object is an object that maintains symmetry along its axis, such as apples or oranges, where measurements from one perspective can sufficiently represent the whole object [2]. In contrast, the 3D approach involved using multiple images captured from various angles [4], [10] or employed RGB-Depth images [7]–[9], [11], [12] to generate a three-dimensional model. This method allowed for more flexible applications, as it could accommodate objects with irregular or non-symmetric shapes, though it required more complex computations and resulted in longer processing times [2].

The methods used for volume prediction can be broadly categorized into four types: geometric-based, interpolation-based, machine learning-based, and deep learning-based. Geometric-based methods rely on the assumption that axisymmetric objects are formed by rotating a cross-sectional profile around their axis of symmetry. Consequently, their volume can be calculated using the volume of revolution in

calculus [13]. Geometric-based approaches approximate this integral using specific geometric shapes such as discs [14], the elliptic cylinder method [15], the cone and conical frustum method [16], as well as employing Pappus' theorem [17] for volume estimation based on object cross-section captured in the image. Despite its simplicity, this method requires careful input of scale factors, which are highly sensitive to the object's size and can affect the accuracy of the volume measurement.

Interpolation-based techniques are used to approximate the boundary curves of an object's cross-section. This involves fitting a mathematical function to the boundary points obtained from the object's cross-section in an image. The function is then applied to estimate the volume using the integral for the volume of revolution. Jana et al. [18] applied 10^{th} order polynomial interpolation to estimate the volume of potatoes, citrus, and tomatoes, with error rates between 7.46% and 10.98%. Similarly, Siswantoro et al. [2] used cubic spline interpolation for axisymmetric food products volume estimation, achieving a very low mean absolute relative error of 1.03%. However, users need to carefully determine the number of points used in the interpolation based on the object's shape. Additionally, scaling factors must be included to ensure the accuracy of the volume estimates.

Several machine learning models have been used to predict the volume of various fruits and vegetables. Rahman et al. , [19] used an artificial neural network (ANN) for potato volume prediction with a coefficient of determination (R^2) of 0.882. Nyalala et al. [3] applied a radial basis function SVM to estimate the volume of tomatoes, achieving an R^2 of 0.982. Mansuri et al. [20] developed an SVM model to estimate the volume of Thai Apple Ber, achieving an R^2 of 0.965. Saikumar et al. [21] applied linear and non-linear models for elephant apple volume prediction. A length-based rational model achieved the best performance with R^2 of 0.924. Xu et al. [8] employed multiple linear regression (MLR), shallow neural networks (SNN), and deep neural network (DNN) for sweet potato volume estimation and achieved the highest R^2 of 0.993 with DNN. Xie et al. [20] also used MLR to estimate the volume of Rosa roxburghii fruits and obtained the best R^2 of 0.898. Although these models can predict volume with high R^2Each is only used to predict the volume of a single type of object. Moreover, some features need to be extracted to predict the volume of axisymmetric objects using machine learning. This process inevitably adds to the computation time required for volume prediction.

The use of deep learning in fruit volume prediction remains relatively limited in prior studies. Dalai et al. [23] employed VGG-ResNet framework for point cloud generation using edge features and SIFT, followed by volume estimation using a hybrid 3D U-Net and graph neural network. The model achieved an error rate of 6.1% and R^2 of 0.982 in volume prediction. Nevertheless, with a computational time of 3.2 seconds per image, this method may be challenging to implement in industrial settings where faster processing is essential.

A Convolutional Neural Network (CNN) is one of the most popular deep learning models used to address classification and regression problems related to digital images [24]. Unlike traditional machine learning methods, where feature extraction must be performed as a preliminary step, CNNs streamline this process by directly processing the raw input images. This allows CNNs to predict the volume of an object without requiring separate feature extraction steps. As a result, CNNs offer greater efficiency compared to traditional machine learning models regarding volume prediction. To adapt a CNN for volume prediction, the final layer of the model must be replaced with a regression layer, enabling the network to output continuous values rather than discrete class labels [25].

This study aims to develop a CNN model to predict the volume of axisymmetric fruits. Using transfer learning techniques, several pretrained CNN models will be employed as base models. Transfer learning allows for the rapid and effective development of a CNN model without requiring a large dataset by leveraging a model previously trained on a large-scale dataset [26]. Additionally, the hyperparameters of the CNN model will be optimized using random search optimization [27] to achieve the best possible performance.

II. Materials and Methods

This section outlines the research process, which includes creating an image dataset, preprocessing images, and segmenting them. It then proceeds to develop a CNN model using transfer learning, hyperparameter tuning for performance optimization, model training, and model evaluation. Fig. 1 illustrates the detailed process flow.

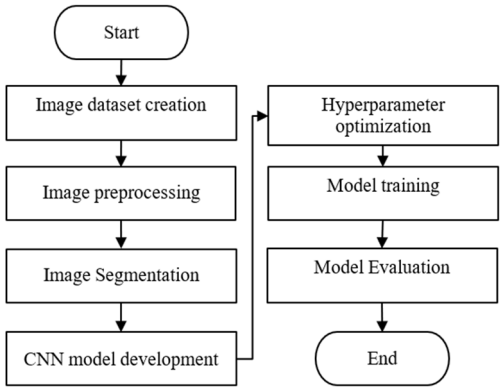


Fig. 1 The flow of the research process.

A. Image Dataset

To create a comprehensive image dataset for training and evaluating a CNN model aimed at predicting the volume of axisymmetric fruits, a specific data collection process was conducted in this study. The dataset comprised four types of fruits: tangerines, Sunkist oranges, green apples, and lemons. These fruits were chosen due to their axisymmetric shapes and smooth, even surfaces, which make them ideal candidates for this study. All samples were sourced from local fruit markets in the Surabaya region of Indonesia. Each fruit type included 30 individual samples.

Each sample was captured four times in different positions, resulting in a total dataset of 480 images. During image capture, each sample was positioned so that its rotation axis was parallel to the horizontal axis of the image coordinate system. All images in the dataset were captured using a Logitech HD Webcam C270h with an HD 720p resolution. The camera was mounted 50 cm above the objects to maintain a consistent height. Each object was placed against a white background to ensure clarity and contrast in the images. To

maintain scale consistency, the fruits were positioned within a red square measuring 15×15 cm, which was aligned with a 600×600 pixel bounding box displayed on the camera screen. This setup ensured that all images had the same size and aspect ratio, crucial for the CNN model, as it relies on standardized input to learn and make accurate predictions. The lighting source used during the image acquisition was overhead fluorescent lamps (TL lamps) provided on the room's ceiling, ensuring consistent illumination throughout the image-capturing process. All images were in RGB color format, maintaining a 1:1 aspect ratio and a uniform size of 600×600 pixels. The captured images were then saved in JPEG file format. Examples of fruit images in the dataset can be seen in Fig. 2.

After capturing the images, the actual volume of each sample was measured using the water displacement method. This process involved submerging the sample in water and measuring the volume of water displaced. Each measurement was performed three times for accuracy, and the average value was recorded as the exact volume of the sample. This precise volume measurement provided the ground truth data necessary for training and evaluating the CNN model. Using this well-structured and standardized dataset, the model can be effectively employed to predict the volume of axisymmetric food products.

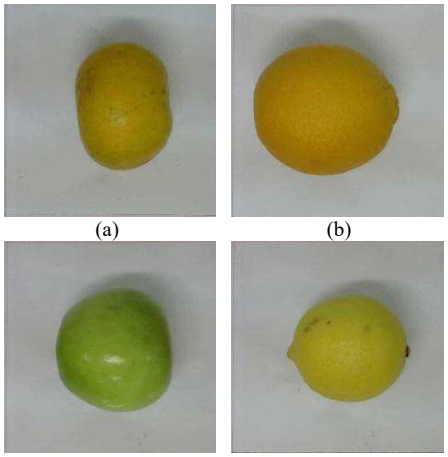


Fig. 2 Examples of fruit image: (a) tangerine, (b) Sunkist orange, (c) green apples, and (d) lemon.

B. Image Preprocessing

Image preprocessing is a crucial step to facilitate subsequent tasks in the workflow. In this study, image preprocessing involved several steps, including resizing, transformation to grayscale image, and filtering. The capture image was resized from 600×600 pixels to 224×224 pixels. This resizing is performed to adapt the image size for input into a pre-trained CNN model for volume prediction. Resizing allows the images to be compatible with the input size required by the pre-trained model.

According to Siswantoro et al. [2], natural objects are easier to distinguish from their background in the HSV color space. Therefore, the resized image was first transformed from RGB into HSV color space to construct a grayscale image. After that, the image was decomposed into single H, S, and V channels, as shown in Fig. 3. The fruit object is easily segmented on the S and V channels. Hence, the grayscale image was constructed from a weighted mean of the images in the S and V channels as in Eq. (1). An example of the grayscale image can be seen in Fig. 4 (a).

$$grayscale = 0.7S + 0.3V \quad (1)$$

The last step for image processing was Gaussian filtering [28]. Gaussian filtering was used to reduce noise and improve the quality of grayscale images. This study employed a Gaussian filter with a kernel size of 15×15. This specific kernel size was chosen to effectively suppress high-frequency noise without significantly blurring important structural details of the food products. The filter works by convolving the image with the kernel, which assigns more weight to the central pixels and gradually decreases the weight for pixels further away, thus ensuring a smooth transition and preserving the essential features necessary for accurate volume estimation. An example of the filtered image can be seen in Fig. 4 (b).

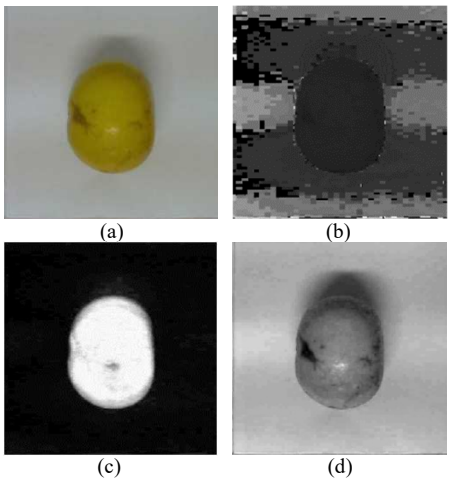


Fig. 3 Examples of (a) RGB image, (b) H channel image, (c) S channel image, and (d) V channel image

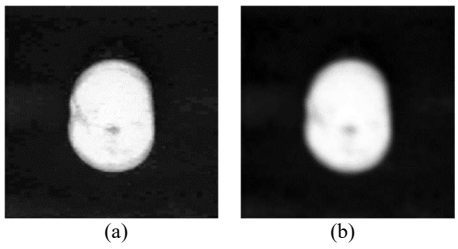


Fig. 4 Examples of (a) grayscale image, and (b) filtered image

C. Image Segmentation

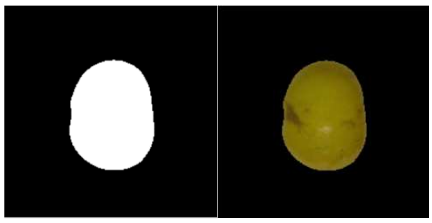
Image segmentation is a crucial step in the volume prediction from image pipeline, particularly for separating the object of interest from the background. In this step, the pre-processed images underwent segmentation to separate the background and the fruit. The thresholding method was used in this step. Otsu's method [28] was employed to determine the threshold value T automatically by minimizing the intra-class variance, as formulated in Eq. (2) and Eq. (3). Where $\sigma_w^2(t)$ is the intra-class variance at the threshold value t. $\sigma_0^2(t)$ and $\sigma_1^2(t)$ are the variance of background pixels and object pixels at the threshold value t, respectively. $P_0(t)$ and $P_1(t)$

are the proportion of background pixels and object pixels at the threshold value *t*, respectively.

$$\sigma_w^2(t) = P_0(t)\sigma_0^2(t) + P_1(t)\sigma_1^2(t) \quad (2)$$

$$T = \underset{t}{\operatorname{argmin}}\, \sigma_w^2(t) \quad (3)$$

After determining the threshold value, the image was binarized by assigning pixel values based on this threshold. A pixel with intensity value greater than *T* was categorized as object pixel with binary value 1 (white). Otherwise, the pixel was categorized as background with binary value 0 (black). This binary image was then multiplied elementwise with the original image to black out the background. This step is crucial as it allows the CNN model to focus solely on the pixels representing the fruit, thus facilitating the extraction of important features necessary for accurate volume prediction. The example of binary image and segmented image can be seen in Fig. 5.



(a) (b)

Fig. 5 Examples of (a) binary image and (b) segmented image

D. CNN Model Architecture

Five pretrained models were employed as the base architecture for the CNN model used to predict the volume of axisymmetric fruits from images, by leveraging transfer learning approach [26]. Transfer learning offers several advantages, including reducing training time and computing resources, while improving model performance through effective feature extraction by using pretrained models trained on large dataset. The pretrained models utilized in this study included MobileNetV2 [29], VGG-16 [30], DenseNet201 [31], ResNet50 [32], and EfficientNetV2B0 [33]. These pretrained models were used for features extraction. The convolutional layers of these models retained their original weights and were not retrained. Before the fully connected layers, a global average pooling layer was added to the network. The fully connected layers were modified specifically to enable accurate volume prediction.

MobileNetV2 is a CNN model designed for mobile and IoT use. It improves on the original MobileNet by introducing inverted residual blocks, where the input and output are narrow layers, and intermediate layers use lightweight depthwise convolutions. The network starts with a convolutional input layer, followed by inverted residual blocks with expansion, depthwise, and projection layers using linear activations. ReLU6 is applied to most layers to limit activations, and batch normalization ensures stable training. At the end, average pooling reduces dimensions before a fully connected layer outputs the class predictions.

VGG-16 is a deep convolutional neural network architecture known for its simplicity and effectiveness in image classification tasks. It consists of 16 weight layers, including 13 convolutional layers and three fully connected layers. The network starts with a series of convolutional layers, using 3×3 filters, followed by max-pooling layers to reduce the spatial dimensions. This design enables the network to capture intricate features while controlling the number of parameters. The convolutional layers are divided into five blocks, each followed by a max-pooling layer. After the convolutional blocks, the network includes three fully connected layers, with the final layer producing the classification output. ReLU activation functions are applied after each convolutional and fully connected layer to introduce non-linearity, and dropout is used in the fully connected layers to prevent overfitting.

DenseNet-201 is a deep convolutional neural network architecture that addresses the vanishing gradient problem and promotes feature reuse by connecting each layer to every other layer in a feed-forward fashion. This architecture consists of 201 layers, including multiple dense blocks and transition layers. The network begins with a convolution and pooling layer to process the input image. It is followed by four dense blocks, where each dense block contains multiple layers, and each layer receives input from all previous layers within the same block. This dense connectivity pattern results in shorter paths between the layers, improving gradient flow and encouraging feature reuse, which makes the network more efficient and effective. Between dense blocks, transition layers, consisting of a batch normalization layer, a 1×1 convolution, and a 2×2 average pooling layer, are used to reduce the number of feature maps and spatial dimensions. DenseNet-201 uses ReLU activation functions and batch normalization throughout the network to improve convergence and stability. The final layers include a global average pooling layer and a fully connected.

ResNet50 is a widely used deep convolutional neural network architecture that introduces the concept of residual learning to address the problem of vanishing gradients in very deep networks. It consists of 50 layers, including convolutional layers, batch normalization layers, activation layers, and fully connected layers, structured into residual blocks. The network begins with an initial convolutional layer followed by a max-pooling layer to reduce the spatial dimensions. It is organized into four stages, each containing several residual blocks, where each block includes two or three convolutional layers with shortcut connections that bypass one or more layers. These shortcut connections add the input of the block directly to the output of the convolutional layers, enabling the network to learn residual functions with reference to the layer inputs, rather than learning unreferenced functions. Each convolutional layer is followed by batch normalization and ReLU activation to stabilize and accelerate training. The network concludes with a global average pooling layer, followed by a fully connected layer that produces the final output.

EfficientNetV2B0 is a highly optimized convolutional neural network architecture designed to balance high performance with computational efficiency. It builds upon the original EfficientNet by incorporating a more efficient scaling method and advanced training techniques, using compound scaling to adjust depth, width, and resolution uniformly. The network starts with mobile inverted bottleneck convolution (MBConv) blocks, known for their parameter efficiency and

low computational cost, including squeeze-and-excitation layers to enhance representational power. It also employs fused MBConv blocks in initial layers for faster training. Each block features a 1x1 convolution for expansion, a depthwise convolution, and a 1x1 convolution for projection, with batch normalization and Swish activation functions for improved stability and performance. The architecture concludes with a global average pooling layer and a fully connected layer for the final output.

E. Hyperparameter Optimization

To achieve an optimal architecture for predicting the volume of axisymmetric fruits using the CNN model, hyperparameter optimization was performed on the fully connected layers of each architecture using the random search optimizer. Random search works by randomly selecting combinations of hyperparameters from predefined ranges and evaluating the performance of the model with each combination. Unlike grid search, which exhaustively evaluates all possible combinations within the specified ranges, random search samples a specified number of hyperparameter combinations, making it a more efficient approach when dealing with a large hyperparameter space. This method can often find optimal hyperparameter values with fewer iterations, reducing computational time and resources while still exploring a diverse set of configurations.

The optimized hyperparameters included the number of dense layers, the number of neurons per dense layer, the presence of dropout layer, and the dropout rate. Furthermore, the learning rate of the Adam optimizer used during the training process was also optimized. The optimized hyperparameters of fully connected layer and their respective search domains are tabulated in Table I. In this study, 20 random combinations of hyperparameter were selected using random search. The objective function used by random search optimizer was the mean absolute error (MAE), which measures the average magnitude of the errors in the predictions without considering their direction, as in Eq. (4). Where V_i and $\widehat{V_i}$ are the exact volume and predicted volume of i^{th} sample, respectively, and n is the number of samples.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|V_i - \widehat{V_i}\right| \tag{4}$$

F. Model Training

The training process for the CNN model involved optimizing the weights of fully connected layers for accurate volume prediction using optimal hyperparameters. The dataset involved in the training process was 360 images, which represented 75% of the total dataset. This subset was further split into 288 images (80%) for training and 72 images (20%) for validation. The Adam optimizer [34] was utilized to minimize the mean squared error (MSE) loss function as in Eq. (5), while MAE as in Eq. (4) was tracked as a performance metric.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(V_i - \widehat{V_i}\right)^2 \tag{5}$$

The training process was set to run for up to 1,000 epochs with a batch size of 32. However, the process was monitored using early stopping to prevent overfitting, halting the training when the validation performance based on MAE ceased to improve. The chosen checkpointing strategy ensured that only the best-performing model on the validation set was saved, providing a model with high predictive accuracy for the task of fruit volume prediction.

TABLE I
OPTIMIZED HYPERPARAMETERS OF FULLY CONNECTED

Hyperparameter	Search domain
Number of dense layers	{1, 2, 3, …, 16}
Number of neurons per dense layer	{8, 16, 32, 64, 128, 256}
Presence of dropout layer	{True, False}
Dropout rate	{0.2, 0.3, 0.4, 0.5}
Learning rate	$\{10^{-3}, 10^{-4}, 10^{-5}, \ldots, 10^{-9}\}$

G. Model Evaluation

The proposed volume prediction method was evaluated by comparing the predicted volume of each fruit sample ($\widehat{V_i}$) with the actual volume (V_i) measured through water displacement for the remaining 25% dataset (120 images). This comparison was quantified using the mean absolute relative error (MARE) of all samples, as in Eq. (6). MARE provided an indication of the model's accuracy in predicting the fruit volumes. Lower MARE values would demonstrate the model's effectiveness in minimizing prediction errors and its reliability in volume estimation tasks.

$$\text{MARE}(\%) = \frac{1}{n}\sum_{i=1}^{n}\frac{|V_i - \widehat{V_i}|}{V_i} \times 100 \tag{6}$$

Additionally, the linear relationship between the predicted volumes and the actual volumes measured by water displacement was assessed using the coefficient of determination (R^2). This statistical measure helped quantify how well the predicted volumes matched the measured volumes, providing a clear indicator of the model's predictive strength. A high R^2 value would suggest that the proposed method accurately captures the relationship between image features and fruit volume. The coefficient of determination was calculated using Eq. (7), where $\bar{V}$ is the mean of actual volume.

$$R^2 = 1 - \sum_{i=1}^{n}\frac{(V_i - \widehat{V_i})^2}{(V_i - \bar{V})^2} \tag{7}$$

III. RESULTS AND DISCUSSION

The summary of predicted fruit volumes using five different CNN models: MobileNetV2, VGG-16, DenseNet201, ResNet50, and EfficientNetV2B, are presented alongside the summary of actual measured volumes and their corresponding absolute relative error (ARE) in Tables II through VI. Each table details the performance of a specific model in predicting the volumes of four fruit types: tangerine, Sunkist orange, green apple, and lemon. The evaluation focuses on the accuracy of the predictions as indicated by the ARE for each fruit type, as well as the overall performance of the models across all samples.

As can be seen in Table II, MobileNetV2 exhibits strong predictive performance with an overall mean ARE of 4.02% and a standard deviation (SD) of 3.22%, indicating consistent prediction accuracy across all samples. The model shows particularly high accuracy for Sunkist oranges, with mean ARE of 2.84% and a low SD of 2.12%, reflecting minimal error variation. For tangerines, the mean ARE is 4.02% with

an SD of 2.83%, demonstrating stable and reliable predictions. However, the model shows higher prediction errors for green apples and lemons, with MAREs of 4.95% and 4.25%, respectively, and higher SDs of 3.44% and 3.97%, suggesting more variability in these predictions.

TABLE II
SUMMARY OF ACTUAL AND PREDICTED VOLUMES USING MOBILENETV2

Fruit type	n	Volume (cm^3)				ARE (%)	
		Actual		Predicted			
		Mean	SD	Mean	SD	Mean	SD
Tangerine	30	181.54	27.99	183.07	27.53	4.02	2.83
Sunkist orange	30	326.80	33.03	323.69	31.34	2.84	2.12
Green apple	30	230.69	31.99	230.26	29.54	4.95	3.44
Lemon	30	179.67	11.87	185.43	14.46	4.25	3.97
All	120	229.68	65.83	230.62	62.90	4.02	3.22

Table III illustrates the performance of VGG-16, which demonstrates a higher overall mean ARE of 5.66% compared to MobileNetV2, indicating less accurate predictions across the sample set. The SD of the ARE for all samples is 3.93%, showing some variability in prediction errors. For tangerines, the mean ARE is 5.26% with an SD of 4.08%, revealing that the model tends to overestimate the volume with moderate error fluctuation. Sunkist oranges exhibit a similar trend, with MARE of 5.37% and SD of 3.43%, reflecting less precise predictions and increased variability. The model's accuracy further diminishes for green apples and lemons, with MAREs of 5.38% and 6.64%, respectively, and SDs of 4.69% and 3.40%, showing greater inaccuracies and variability.

TABLE III
SUMMARY OF ACTUAL AND PREDICTED VOLUMES USING VGG-16

Fruit type	n	Volume (cm^3)				ARE (%)	
		Actual		Predicted			
		Mean	S. D.	Mean	SD	Mean	SD
Tangerine	30	181.54	27.99	186.59	31.73	5.26	4.08
Sunkist orange	30	326.80	33.03	337.85	28.89	5.37	3.43
Green apple	30	230.69	31.99	235.25	33.60	5.38	4.69
Lemon	30	179.67	11.87	191.55	13.53	6.64	3.40
All	120	229.68	65.83	237.81	67.04	5.66	3.93

TABLE IV
SUMMARY OF ACTUAL AND PREDICTED VOLUMES USING DENSENET201

Fruit type	n	Volume (cm^3)				ARE (%)	
		Actual		Predicted			
		Mean	SD	Mean	SD	Mean	SD
Tangerine	30	181.54	27.99	180.51	25.18	4.75	3.46
Sunkist orange	30	326.80	33.03	330.83	29.30	3.20	2.13
Green apple	30	230.69	31.99	217.02	25.10	6.57	4.97
Lemon	30	179.67	11.87	190.43	12.05	6.27	4.22
All	120	229.68	65.83	229.70	64.58	5.20	4.03

Table IV displays the performance of DenseNet201, which achieves an overall mean ARE of 5.20% with SD of 4.03%. This indicates a moderate level of accuracy in predicting fruit volumes and some variability in prediction errors. For tangerines, DenseNet201 provides a mean ARE of 4.75% with SD of 3.46%, suggesting relatively stable predictions with moderate errors. The model performs best for Sunkist oranges, with a mean ARE of 3.20% and a low SD of 2.13%, indicating both accurate and consistent predictions. However, for green apples and lemons, DenseNet201 produces higher MAREs of 6.57% and 6.27%, respectively, along with higher SDs of 4.97% and 4.22%, revealing significant prediction inaccuracies and variability.

Table V demonstrates the performance of ResNet50, which stands out with the lowest overall mean ARE of 3.76%, indicating the highest level of accuracy among the models evaluated. The SD of the ARE across all samples is 3.30%, also reflecting some variability in the prediction errors. For tangerines, ResNet50 achieves a mean ARE of 3.36% with an SD of 2.37%, showing precise and consistent predictions. The model performs similarly well with Sunkist oranges, exhibiting a mean ARE of 2.95% and a low SD of 2.10%, indicating very accurate predictions with minimal error variation. The performance is also strong for lemons, with a mean ARE of 3.25% and an SD of 2.48%, showing low error rates and consistency. However, for green apples, ResNet50 has a higher mean ARE of 5.47% and a higher SD of 4.94%, revealing increased prediction errors and variability for this fruit.

TABLE V
SUMMARY OF ACTUAL AND PREDICTED VOLUMES USING RESNET50

Fruit type	n	Volume (cm^3)				ARE (%)	
		Actual		Predicted			
		Mean	SD	Mean	SD	Mean	SD
Tangerine	30	181.54	27.99	182.49	27.34	3.36	2.37
Sunkist orange	30	326.80	33.03	327.44	30.27	2.95	2.10
Green apple	30	230.69	31.99	231.60	26.57	5.47	4.94
Lemon	30	179.67	11.87	183.80	12.66	3.25	2.48
All	120	229.68	65.83	231.34	64.16	3.76	3.30

As can be observed from Table VI, EfficientNetV2B0 produces an overall mean ARE of 4.23% with the SD of 3.28%. These results indicate that EfficientNetV2B0 has moderate level of accuracy in volume predictions. The SD of the ARE for all samples reflects some variability in the prediction errors. For tangerines, the model achieves a mean ARE of 4.16% with SD of 3.88%, suggesting relatively accurate predictions with moderate variability.

TABLE VI
SUMMARY OF ACTUAL AND PREDICTED VOLUMES USING EFFICIENTNETV2B0

Fruit type	n	Volume (cm^3)				ARE (%)	
		Actual		Predicted			
		Mean	SD	Mean	SD	Mean	SD
Tangerine	30	181.54	27.99	182.81	26.42	4.16	3.88
Sunkist orange	30	326.80	33.03	330.89	24.03	4.02	2.94
Green apple	30	230.69	31.99	222.59	28.83	4.65	3.39
Lemon	30	179.67	11.87	185.97	12.13	4.08	2.96
All	120	229.68	65.83	230.56	64.64	4.23	3.28

The performance for Sunkist oranges is better, with a mean ARE of 4.02% and a lower SD of 2.94%, indicating good accuracy and consistency. However, for green apples, the model exhibits a higher mean ARE of 4.65% and SD of 3.39%, revealing increased errors and variability in predictions. In contrast, EfficientNetV2B0 performs better for lemons, with a mean ARE of 4.08% and a lower SD of 2.96%, reflecting relatively accurate and stable predictions.

The coefficient of determination (R^2) values in Table VII further support the performance rankings of the models. R^2 measure the linear relationship between actual and predicted volumes. As can be seen in Table VII, ResNet50 has the highest R^2 value of 0.9721, indicating that it explains 97.21% of the variance in predicted volumes. Fig. 6 presents the linear

relationship between actual and predicted volumes using ResNet50, demonstrating its ability to provide accurate volume estimations across all fruit types.

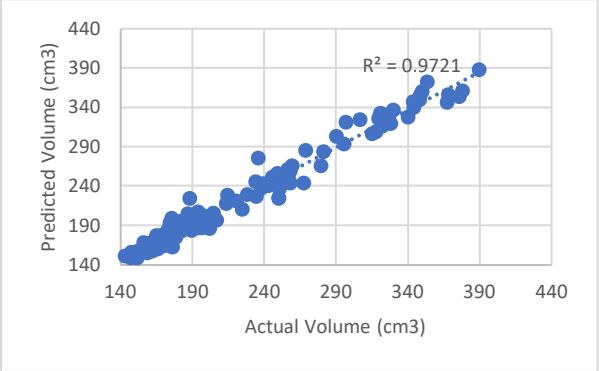


Fig. 6 The linear relationship between actual and predicted volumes using ResNet50

This result aligns with its superior performance noted in Table V. MobileNetV2 follows closely with R^2 of 0.9712, supporting its strong accuracy as shown in Table II. EfficientNetV2B0 has R^2 values of 0.9648, which indicates substantial but lower explanatory power compared to ResNet50 and MobileNetV2. This is consistent with the moderate prediction accuracy observed in Table VI. In comparison, VGG-16 and DenseNet201 had lower R^2 values of 0.9591 and 0.9512, respectively, indicating weaker performance. Overall, the R^2 values confirm that ResNet50 provide the most accurate and consistent prediction, while DenseNet201 and VGG-16 exhibit comparatively weaker performance.

TABLE VII
COEFFICIENT OF DETERMINATION BETWEEN ACTUAL AND PREDICTED VOLUMES FOR ALL MODELS

Model	R^2
MobileNetV2	0.9712
VGG-16	0.9591
DenseNet201	0.9512
ResNet50	0.9721
EfficientNetV2B0	0.9648

To provide a comprehensive evaluation of ResNet50's performance in fruit volume prediction, the model's accuracy is benchmarked against several existing volume measurement methods using the same set of fruit image samples. Table VIII outlines the mean ARE and SD for various methods, including geometric, traditional machine learning, and interpolation-based methods. ResNet50 achieves the lowest mean ARE of 3.76% with an SD of 3.30%, reflecting its superior accuracy and consistency. Geometric approaches like the disk method (mean ARE 5.34%, SD 3.65%), Pappus's theorem (mean ARE 5.64%, SD 5.74%), and conical frustum method (mean ARE 24.96%, SD 17.09%) have significantly higher errors and variabilities, indicating less reliable predictions, especially in conical frustum method. Traditional machine learning models, such as ANN (mean ARE 4.57%, SD 3.63%) and RBF SVR (mean ARE 4.88%, SD 4.09%), also exhibit greater prediction errors compared to ResNet50, though they offer improved accuracy over geometric methods. Similarly, interpolation techniques, including 10th-order polynomial interpolation (mean ARE 5.27%, SD 3.64%) and cubic spline interpolation (mean ARE 4.79%, SD 3.68%), also demonstrate larger errors and variability. These results show that ResNet50 outperforms both geometric and traditional machine learning methods by offering the most precise and consistent volume predictions.

TABLE VIII
COMPARISON WITH EXISTING VOLUME MEASUREMENT METHODS

Method	ARE(%)	
	Mean	SD
Disk [14]	5.34	3.65
Pappus's theorem [17]	5.64	5.74
Cone and conical frustum [16]	24.96	17.09
ANN [19]	4.57	3.63
10th order polynomial interpolation [18]	5.27	3.64
RBF SVR [3]	4.88	4.09
Cubic spline interpolation [2]	4.79	3.68
ResNet50 (this study)	3.76	3.30

IV. CONCLUSION

This study demonstrates the effectiveness of convolutional neural networks (CNNs) with transfer learning in predicting the volume of axisymmetric fruits based on image data. The methodology involved creating an image dataset, preprocessing the images, segmenting the fruit areas, and applying CNNs to extract features and predict volumes. Five pretrained models, including MobileNetV2, VGG-16, DenseNet201, ResNet50, and EfficientNetV2B0, were fine-tuned for this task, with the fully connected layers modified for volume prediction. Among these models, ResNet50 outperformed the others, achieving the highest predictive accuracy with a mean absolute relative error (ARE) of 3.76% and a coefficient of determination (R^2) of 0.9721. These results highlight the model's superior accuracy and consistency compared to geometric methods, traditional machine learning models, and interpolation techniques. The findings confirm the potential of CNNs for precise and reliable volume estimation, particularly for fruits with complex shapes. However, some models, such as DenseNet201 and VGG-16, exhibited lower accuracy and higher error variability, indicating the need for further optimization.

Future study could explore expanding the dataset to include a wider variety of fruits, including non-axisymmetric ones, to improve the generalization of the models. The application of these models in real-time systems for agriculture, retail, or industrial purposes could significantly enhance operational efficiency and reduce human error.

ACKNOWLEDGMENT

This study was supported by the University of Surabaya under PPB Grant No. 151/SP-Lit/LPPM-01/FT/XI/2023.

REFERENCES

[1] S. M. H. Mousavi and S. M. H. Mosavi, "Automatic Infrared-Based Volume and Mass Estimation System for Agricultural Products: Along with Major Geometrical Properties," in *2021 11th International Conference on Computer Engineering and Knowledge (ICCKE)*, 2021, pp. 140–149, doi:10.1109/iccke54056.2021.9721526.

[2] J. Siswantoro, E. Asmawati, and M. Z. F. N. Siswantoro, "A rapid and accurate computer vision system for measuring the volume of axi-symmetric natural products based on cubic spline interpolation," *J. Food Eng.*, vol. 333, p. 111139, 2022, doi:10.1016/j.jfoodeng.2022.111139.

[3] I. Nyalala *et al.*, "Weight and volume estimation of single and occluded tomatoes using machine vision," *Int. J. Food Prop.*, vol. 24, no. 1, pp. 818–832, 2021, doi: 10.1080/10942912.2021.1933024.

[4] Y. Han, S. Xu, Q. Zhang, H. Lu, X. Liang, and C. Fan, "Non-destructive detection method and experiment of pomelo volume and flesh content based on image fusion," *Postharvest Biol. Technol.*, vol. 213, p. 112953, 2024, doi: 10.1016/j.postharvbio.2024.112953.

[5] I. Nyalala, C. Okinda, C. Kunjie, T. Korohou, L. Nyalala, and Q. Chao, "Weight and volume estimation of poultry and products based on computer vision systems: a review," *Poult. Sci.*, vol. 100, no. 5, p. 101072, 2021, doi: 10.1016/j.psj.2021.101072.

[6] H. Fitriyah, "Accuracy of Various Methods to Estimate Volume and Weight of Symmetrical and Non-Symmetrical Fruits using Computer Vision.," *J. ICT Res. Appl.*, vol. 16, no. 3, 2022, doi:10.5614/itbj.ict.res.appl.2022.16.3.2.

[7] Y. Zhu, S. Cao, T. Song, Z. Xu, and Q. Jiang, "3D reconstruction and volume measurement of irregular objects based on RGB-D camera," *Meas. Sci. Technol.*, vol. 35, no. 12, p. 125010, 2024, doi:10.1088/1361-6501/ad7621.

[8] J. Xu, Y. Lu, E. Olaniyi, and L. Harvey, "Online volume measurement of sweetpotatoes by A LiDAR-based machine vision system," *J. Food Eng.*, vol. 361, p. 111725, 2024, doi:10.1016/j.jfoodeng.2023.111725.

[9] S. Luo, J. Tang, J. Peng, and H. Yin, "A novel approach for measuring the volume of Pleurotus eryngii based on depth camera and improved circular disk method," *Sci. Hortic. (Amsterdam).*, vol. 336, p. 113382, 2024, doi: 10.1016/j.scienta.2024.113382.

[10] Y. S. Gan, L. Wei, Y. Han, C. Zhang, Y.-C. Huang, and S.-T. Liong, "A statistical approach in enhancing the volume prediction of ellipsoidal ham," *J. Food Eng.*, vol. 290, p. 110186, 2021, doi:10.1016/j.jfoodeng.2020.110186.

[11] Y. A. Sari and A. Gofuku, "Measuring food volume from RGB-Depth image with point cloud conversion method using geometrical approach and robust ellipsoid fitting algorithm," *J. Food Eng.*, vol. 358, p. 111656, 2023, doi: 10.1016/j.jfoodeng.2023.111656.

[12] J. Li, M. Wu, and H. Li, "3D reconstruction and volume estimation of jujube using consumer-grade RGB-depth sensor," *IEEE Access*, vol. 11, pp. 61502–61512, 2023, doi:10.1109/access.2023.3285713.

[13] M. T. Nair, "Definite Integral BT - Calculus of One Variable," M. T. Nair, Ed. Cham: Springer International Publishing, 2021, pp. 173–249, doi: .10.1007/978-3-030-88637-0

[14] H. M. Tran, K. T. Pham, T. M. Vo, T.-H. Le, T. T. M. Huynh, and S. V. T. Dao, "A new approach for estimation of physical properties of irregular shape fruit," *IEEE Access*, vol. 11, pp. 46550–46560, 2023, doi: 10.1109/access.2023.3273777.

[15] T. Mon and N. ZarAung, "Vision based volume estimation method for automatic mango grading system," *Biosyst. Eng.*, vol. 198, pp. 338–349, 2020, doi: 10.1016/j.biosystemseng.2020.08.021.

[16] T. Huynh, L. Tran, and S. Dao, "Real-Time Size and Mass Estimation of Slender Axi-Symmetric Fruit/Vegetable Using a Single Top View Image," *Sensors*, vol. 20, no. 18, p. 5406, Sep. 2020, doi:10.3390/s20185406.

[17] M. Soltani, M. Omid, and R. Alimardani, "Egg volume prediction using machine vision technique based on pappus theorem and artificial neural network," *J. Food Sci. Technol.*, vol. 52, no. 5, pp. 3065–3071, 2015, doi: 10.1007/s13197-014-1350-6.

[18] S. Jana, R. Parekh, and B. Sarkar, "A De novo approach for automatic volume and mass estimation of fruits and vegetables," *Optik (Stuttg).*, vol. 200, p. 163443, Jan. 2020, doi:10.1016/j.ijleo.2019.163443.

[19] F. Rahman *et al.*, "Prediction of Potato Volume by Neural Network Regression Model," *J. Agric. Mach. Bioresour. Eng.*, vol. 8, no. 2, pp. 40–52, 2024, doi: 10.61361/jambe.v8i2.124.

[20] S. M. Mansuri, P. V. Gautam, D. Jain, and C. Nickhil, "Computer vision model for estimating the mass and volume of freshly harvested Thai apple ber (Ziziphus mauritiana L.) and its variation with storage days," *Sci. Hortic. (Amsterdam).*, vol. 305, p. 111436, 2022, doi:10.1016/j.scienta.2022.111436.

[21] A. Saikumar, C. Nickhil, and L. S. Badwaik, "Physicochemical characterization of elephant apple (Dillenia indica L.) fruit and its mass and volume modeling using computer vision," *Sci. Hortic. (Amsterdam).*, vol. 314, p. 111947, 2023, doi:10.1016/j.scienta.2023.111947.

[22] Z. Xie, J. Wang, Y. Yang, P. Mao, J. Guo, and M. Sun, "Image processing based modeling for Rosa roxburghii fruits mass and volume estimation," *Sci. Rep.*, vol. 14, no. 1, p. 15507, 2024, doi:10.1038/s41598-024-65321-9.

[23] R. Dalai, N. Dalai, and K. K. Senapati, "An accurate volume estimation on single view object images by deep learning based depth map analysis and 3D reconstruction," *Multimed. Tools Appl.*, vol. 82, no. 18, pp. 28235–28258, 2023, doi: 10.1007/s11042-023-14615-7.

[24] M. M. Taye, "Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions," *Computation*, vol. 11, no. 3, p. 52, 2023, doi:10.3390/computation11030052.

[25] S. Ajala, H. Muraleedharan Jalajamony, M. Nair, P. Marimuthu, and R. E. Fernandez, "Comparing machine learning and deep learning regression frameworks for accurate prediction of dielectrophoretic force," *Sci. Rep.*, vol. 12, no. 1, pp. 1–17, 2022, doi: 10.1038/s41598-022-16114-5.

[26] T. Lu, B. Han, L. Chen, F. Yu, and C. Xue, "A generic intelligent tomato classification system for practical applications using DenseNet-201 with transfer learning," *Sci. Rep.*, vol. 11, no. 1, p. 15824, 2021, doi: 10.1038/s41598-021-95218-w.

[27] E. A. Tsvetkov and R. A. Krymov, "Pure random search with virtual extension of feasible region," *J. Optim. Theory Appl.*, vol. 195, no. 2, pp. 575–595, 2022, doi: 10.1007/s10957-022-02097-w.

[28] W. Burger and M. J. Burge, *Digital Image Processing An Algorithmic Introduction*. Cham, Switzerland: Springer Nature Switzerland AG, 2022.

[29] Y. Gulzar, "Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique," *Sustainability*, vol. 15, no. 3. 2023, doi: 10.3390/su15031906.

[30] X. Li, Z. Liu, T. Zhai, and X. Yang, "Fruit Classification Identification Research Based on VGG16 Network," in *2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA)*, 2024, pp. 387–391, doi:10.1109/icipca61593.2024.10709233.

[31] M. S. Morshed, S. Ahmed, T. Ahmed, M. U. Islam, and A. B. M. A. Rahman, "Fruit Quality Assessment with Densely Connected Convolutional Neural Network," in *2022 12th International Conference on Electrical and Computer Engineering (ICECE)*, 2022, pp. 1–4, doi: 10.1109/icece57408.2022.10088873.

[32] A. Doshi, P. Khatri, V. Dodiya, S. Muni, and M. Narvekar, "PikFresh - Fruit Quality Detection using ResNet50," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2023, pp. 1–5, doi:10.1109/icccnt56998.2023.10308324.

[33] G. Singh, K. Guleria, and S. Sharma, "A Pre-trained EfficientNetV2B0 Model for the Accurate classification of Fake and Real Images," in *2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2024, pp. 1082–1086, doi:10.1109/iceca63461.2024.10801011.

[34] H. Salem, A. E. Kabeel, E. M. S. El-Said, and O. M. Elzeki, "Predictive modelling for solar power-driven hybrid desalination system using artificial neural network regression with Adam optimization," *Desalination*, vol. 522, p. 115411, 2022, doi:10.1016/j.desal.2021.115411.