Broccoli Grading Based on Improved Convolutional Neural Network Using Ensemble Deep Learning

Zaki Imaduddin[1], Yohanes Aris Purwanto[2]*, Sony Hartono Wijaya[3], Shelvie Nidya Neyman[4]
Department of Computer Science, IPB University, Bogor, Indonesia[1, 3, 4]
Department of Mechanical and Biosystem Engineering, IPB University, Bogor, Indonesia[2]
Departement of Informatics, Sekolah Tinggi Teknogi Terpadu Nurul Fikri, Depok, Indonesia[1]

***Abstract*—The demand for broccoli in Indonesia has been increasing significantly, with an annual growth of approximately 15% to 20%. However, the supply availability remains insufficient, and its quality is often inconsistent. Therefore, a grading process is needed to classify broccoli into grades A, B, and C based on color, size, and shape. Currently, the grading process is carried out solely by market intermediaries, while farmers and the general public have a limited understanding of this process. This research developed an automated grading method using a Convolutional Neural Network (CNN) based on two broccoli images' top and side views. Three main parameters, namely color, size, and shape, were identified from these images and used as grading determinants. An ensemble deep learning technique was applied by training each parameter separately using several CNN models, namely ResNet50, EfficientNetB2, VGG16, and Improved CNN. These were then combined in the testing phase using a voting technique. The test was conducted 64 times with various model combinations to achieve the best accuracy. A significant contribution of the Improved CNN lies in the shape feature, which achieved a maximum performance of 95%. This study also compared evaluation metrics such as precision, recall, F-Score, and accuracy across different model combinations.**

Keywords—Grading; convolution neural network; ensemble deep learning; voting

I. Introduction

Broccoli (Brassica oleracea L. var. italica) is a widely cultivated cruciferous vegetable valued for its high nutritional content and economic significance. It is a rich source of essential vitamins such as C, K, and A [1] and bioactive compounds such as glucosinolates, which are studied for their potential health benefits. Globally, broccoli is a key commercial vegetable, with strong export demand from regions such as the United States, China, and Europe. In Indonesia, the demand for broccoli has increased by up to 20% annually, driven by growing consumption in restaurants, hotels, modern retail markets, and exports [2]. However, domestic supply is constrained by a lack of standardized quality grading, creating inconsistencies that disadvantage farmers and traders alike. Broccoli grading is crucial in determining market value and quality, with morphological attributes such as size, shape, color, and compactness as key indicators. The morphology of the broccoli head is particularly significant, as it reflects the crop's overall quality and resilience to environmental stress [3]. Traditionally, grading has been performed manually, leading to inconsistencies due to subjective evaluation. Various approaches have been employed to assess broccoli quality, including dry geometric and weight measurements, mass spectrometry analysis, and non-contact sensor technologies [4]. Among these, image-based methods using RGB digital cameras and deep learning algorithms have emerged as promising solutions because of their non-destructive nature, cost-effectiveness, and high accuracy. Recent studies have demonstrated the potential of Convolutional Neural Networks (CNNs) in broccoli grading, utilizing algorithms such as Mask R-CNN and ResNet for tasks such as detecting harvest readiness, estimating weight, and analyzing color attributes.

Research conducted by Blok et al. employed the Mask R-CNN algorithm [5] and successfully detected 229 out of 232 harvest-ready broccoli heads across three cultivars. The study concluded that the algorithm demonstrated better generalizability across multiple broccoli cultivars. Previous studies have also contributed by combining the Viewpoint Feature Histogram (VFH) with a Support Vector Machine (SVM), enabling precise broccoli detection and facilitating automated systems for detecting and measuring broccoli heads. This approach proved effective in achieving the goal of determining the optimal harvest timing [6].

A recent study by Zhou [4] developed a dataset of 100 broccoli head images captured using a custom-designed imaging system under controlled conditions. The study employed an improved ResNet CNN algorithm to extract broccoli pixels from the background and estimate their weights. Additionally, the particle Swarm Optimization Algorithm (PSOA) and Otsu method were applied to evaluate the broccoli quality, achieving an accuracy rate of 0.896 [4]. However, Zhou et al. study was limited to color as the sole criterion, whereas grading broccoli typically requires multiple parameters to assess its overall quality.

Another study compared the performance of various models, including ResNet, DenseNet, MobileNetV2, NASNet, and EfficientNetB2, to determine the best model for grading apples and bananas [7]. EfficientNet yielded the highest accuracies of 99.2% for the training data and 98.6% for the testing data. However, this approach has not been applied to broccoli datasets.

This study introduces several contributions, and those are: 1) the development of a broccoli grading model based on three features: color, size, and shape, where the data are divided into three independent features, each assessed without reliance on the others; 2) optimization of convolutional and classification models for the broccoli grading process; and 3) comparison of grading results using multiple CNN models.

The grading criteria in this study differ from those of previous research because of the distinct physical characteristics of domestically produced broccoli compared to imported varieties. The criteria include shape (degree of roundness), color, size, and flower compactness (density). Image acquisition was conducted from two perspectives: top and side views. The research modified several deep learning models, including ResNet50, EfficientNetB2, VGG16, and Improved CNN.

In addition, the study utilized a parallel Ensemble Learning technique during the training phase. This approach allows models to be developed independently, with no interdependencies, ensuring that errors from one model are less likely to align with errors from others. Consequently, the weaknesses of one model can be mitigated by those of the other. Ensemble learning has been successfully applied across various fields and often outperforms single-model approaches [8].

The predictions from all deep learning models were aggregated and reused during the testing phase, where a voting mechanism was applied to make classification decisions. This image-based grading algorithm aims to enhance broccoli's post-harvest quality and standardization while advancing the agricultural industry in Indonesia, particularly benefiting broccoli farmers.

II. Literature Review

The utilization of Convolutional Neural Network (CNN) architectures had seen rapid growth since 2012, when Krizhevsky's breakthrough in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) demonstrated the CNN's efficiency for image classification across various domains [9][10]. Specifically in agriculture, CNN-based approaches have been employed for fruit classification and detection, highlighting CNN's pivotal role of CNNs in image processing for this sector. CNN operations typically involve six essential layers [10][12]:

- Input Layer: This layer accepts raw images as input and forwards them to the subsequent layers for feature extraction.
- Convolution Layer: Each output connects to a small region in the input using a weight matrix (filter or kernel). Multiple filters can be applied within each convolution layer to generate several 2D outputs stacked into an output volume.
- ReLU (Rectified Linear Unit): Negative values in the output of the convolution layers are replaced with zero, accelerating the training process.
- Pooling Layer: This process downsamples feature maps to achieve translation invariance. Feature map dimensions are reduced using average and max pooling techniques.
- Fully Connected Layer: This final layer integrates all filtered image data, converting it into labels and categories.
- Softmax Layer: Positioned before the output layer, this layer generates decimal probabilities for each class, ranging from 0 to 1, enabling CNNs to extract, process, and classify image data features efficiently.

TABLE I. CNN-Based Research on Broccoli Plants

Author	CNN Model	Result
[5]	Mask Region-based Convolutional Neural Network	Successfully detected 175 out of 176 test datasets
[6]	KNN & SVM	95.2%
[13]	3D information based on convolution neural network	Below 90%
[14]	improved resnet	Below 90%
[15]	Organised Edges Segmentation (OES) and Organised Region Growing Segmentation (ORG	Low generalization level
[16]	Gaussian Mixture Model	97.9%

Table I compares various CNN models and other methods in detection or classification tasks based on referenced studies. Mask R-CNN performed well, detecting 175 out of 176 test datasets, while the Gaussian Mixture Model achieved the highest accuracy at 97.9%. KNN & SVM also performed well with 95.2%, whereas 3D CNN and improved ResNet had accuracies below 90%. OES & ORG showed a low generalization level, indicating limitations in handling new data. the researchers adopted the parallel Ensemble Learning technique, which integrates multiple models or classifiers to improve the prediction accuracy compared to using a single model [8]. Each model or classifier within the ensemble is generally trained on slightly different datasets (distinct training data or varying feature selection methods).

Ensemble Learning aggregates predictions from multiple base models to produce more stable and accurate results. Two primary strategies introduce diversity among the baseline classifiers: homogeneous ensembles and heterogeneous. Homogeneous ensembles consist of baseline classifiers of the same type, in which each classifier is trained on different datasets. The feature selection method remained consistent across the datasets. By contrast, Heterogeneous ensembles comprise baseline classifiers of various types, allowing each classifier to adopt distinct methodologies in processing the training data [11]. Heterogeneous ensembles typically demonstrate superior generalization capabilities than homogeneous ensembles [17]. Several commonly used ensemble techniques include the following:

- Averaging computes the average prediction from multiple models.
- Bagging (Bootstrap Aggregating) involves training multiple models on random subsets of training data and subsequently combining their outputs. Bagging aims to stabilize models by reducing the variance [18].
- Random Forest is a bagging method that employs numerous decision trees to generate predictions.
- Stacking combines predictions from multiple base models and utilizes another model (a meta-model) to aggregate these predictions. Boosting enhances the

model performance by assigning greater importance to previously misclassified data points. A prominent boosting algorithm is AdaBoost [19], which improves the performance of the decision trees.

Among ensemble methods, bagging and boosting are the most frequently applied in classification tasks. These approaches are widely recognized for their robust theoretical foundations and exceptional empirical results [11]. Existing studies highlight that bagging and boosting are particularly effective when applied to decision tree models [20].

III. Research Method

This research involves several methodological stages to achieve broccoli grading classification: Data Acquisition, Data Pre-processing, Model Training, Model Evaluation, and Grading (Fig. 1).

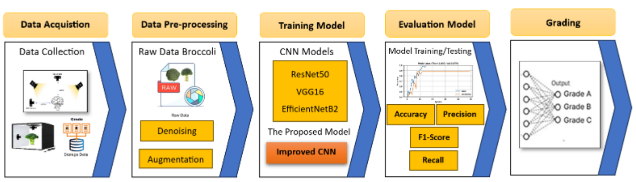

Fig. 1. Methodology of classification and grading broccoli.

A. Data Acquisition

The first stage of the research process was the dataset collection. The broccoli images were gathered and classified into grades with the assistance of experts. The dataset consisted of 450 broccoli images captured from the top and side views, which were categorized into grades A, B, and C.

B. Data Pre-Processing

After the data were collected, the images were processed through pre-processing stages.

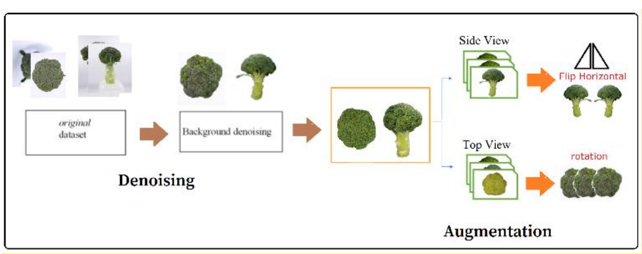

Fig. 2. The architecture of denoising and data augmentation process.

In Fig. 2, the first step is denoising, which removes unnecessary backgrounds to improve image quality. The second step is augmentation, which involves transforming the images using rotation techniques for top-view images and flipping them for side-view images. The denoising and data augmentation process aimed to increase the variation in the dataset.

C. Training Model

At this stage, the data training process uses ensemble deep learning. The ensemble learning technique enables the classification of three distinct feature subsets, color, size, and shape, by training each feature independently without relying on the others. This independent training allows the model to capture and store information from each feature more effectively, reducing the risk of interference between features. As a result, during the testing process, the model can focus on accurately reading and determining the grade of each test data point. This approach enhances the model's generalization ability across different feature types and improves overall classification performance.

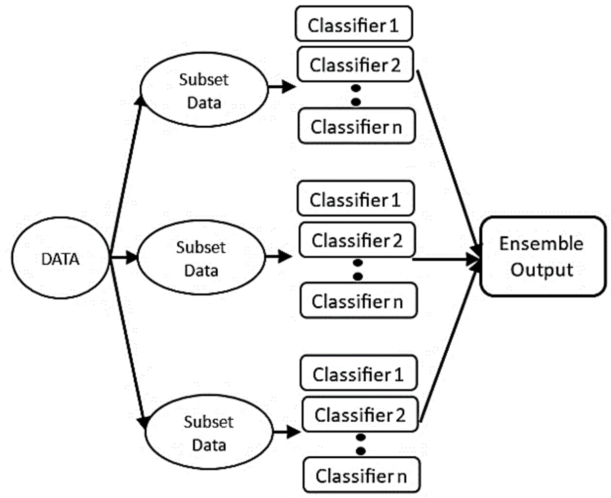

Fig. 3. Development of training architecture using ensemble learning.

As shown in Fig 3, each subset is trained using several convolutional neural network (CNN) models, which are known to be effective for image recognition. The models used included ResNet50, VGG16, EfficientNetB2, and the proposed model, which is an Improved CNN.

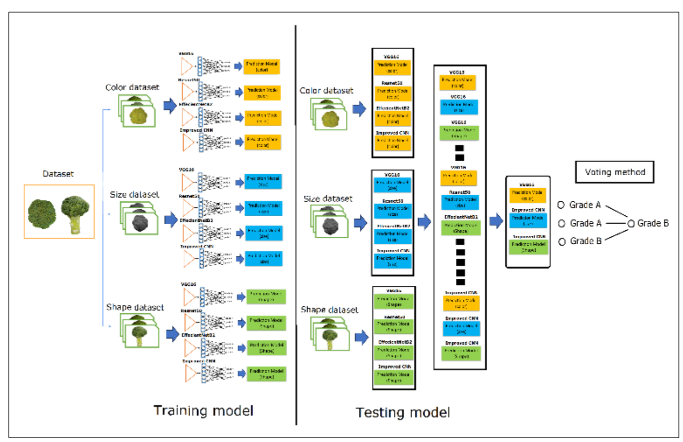

Fig. 4. The architecture training and testing data using ensemble deep learning.

In Fig. 4, each subset based on color, size, and shape is trained using several Convolutional Neural Network (CNN) models: ResNet50, VGG16, EfficientNetB2, the proposed model, and the Improved CNN. The outputs of these models were used in the testing phase.

D. Evaluation Model

The trained models were evaluated using performance metrics such as accuracy, precision, recall, and F1-Score. Below are some equations for this method.

$$\mathrm{Acc} = \frac{(\mathrm{TP}+\mathrm{TN})}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \quad (1)$$

The Accuracy (Acc) formula was used to measure the model's performance by calculating the number of True Positive and True Negative elements as the numerator and the total number of elements in the Confusion Matrix as the denominator. The True Positive and True Negative elements represent the correct predictions made by the model and are located on the matrix's main diagonal. Meanwhile, the denominator includes all the elements incorrectly classified by the model outside the main diagonal. Therefore, accuracy indicates how well the model can make correct classifications for positive and negative cases compared with the entire dataset [9].

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

Precision measures the model's accuracy in identifying positive cases, indicating how well the model avoids misclassifying negative cases as positive cases.

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

Recall or Sensitivity indicates how well the model can remember or recognize all the existing positive cases. This metric is important because it ensures the model does not miss significant positive cases. The higher the recall value, the better the model captures all positive cases.

$$F1 - score = 2.\frac{(precision \times recall)}{precision + recall} \quad (4)$$

The F1 score is a metric that combines Precision and Recall to assess the performance of a classification model. The F1-score reached its optimal value when the model had high precision and recall, indicating that it effectively identified and recognized all positive cases in the data. Conversely, the F1-score will be low if the Precision or Recall is low, signaling inadequate model performance. Therefore, the F1 score provides an overview of how well the model can balance Precision and Recall, with one being the best and zero the worst.

E. Grading

In the final stage, the model's prediction results are translated into specific categories (e.g., Grade A, B, C) based on a voting technique, where the decision is based on the lowest grade. This process was used to classify broccoli quality according to predefined standards.

IV. Result and Discussion

A. Data Acquisition

Based on Fig. 5, all broccoli data that had been labeled and collected were placed in a photo box studio and photographed individually using a 12MP SLR camera. The labeling process was conducted in collaboration with three experts experienced in grading transit locations before distribution to supermarkets. The broccoli samples were then categorized into grades A, B, and C. Subsequently, images were captured from the top and side angles, maintaining a consistent distance of 20 cm between the object and the camera. Photobox Studio was equipped with adjustable lighting settings to ensure optimal image quality and minimize noise interference from the surrounding environment. This process was designed to produce high-quality images suitable for further analysis.

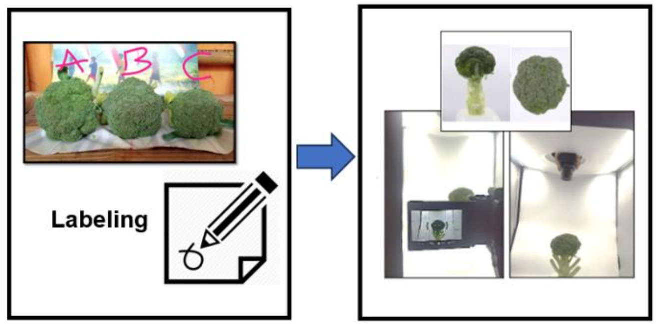

Fig. 5. Collecting data with photobox studio.

B. Data Pre-Processing

Table II shows the distribution of the dataset used for broccoli grading based on two perspectives, namely the top view and side view, as well as the impact of the data augmentation process. The table is divided into grades A, B, and C. In the top view, each grade (A, B, and C) contained 100 samples from the original dataset and 100 augmented samples, indicating that data augmentation doubles the dataset size for the top view. Meanwhile, each grade contained 50 samples from the original dataset and 50 augmented samples in the side view, demonstrating a similar augmentation process but fewer samples than in the top view. Overall, the total number of samples for the top view, including original and augmented data, was 600 (300 original + 300 augmented). In comparison, the total for the side view was 100 samples (50 original + 50 augmented), resulting in a combined dataset of 700 samples. This table emphasizes that the dataset is balanced across the three grade categories and illustrates how data augmentation is applied to both perspectives to enhance the model training performance.

TABLE II. Dataset Augmentation

Grade	Top view (original)	Augmentation	Side View (original)	Augmentation
Grade A	100	100	50	50
Grade B	100	100	50	50
Grade C	100	100	50	50
Tota Dataset	600		100	
	700			

After the data were augmented, data splitting was performed, in which 80% of the data were allocated for training and 20% for testing.

C. Training Model

At this stage, the data training process utilizes an ensemble deep learning approach with heterogeneous ensemble specifications, a technique that has shown better generalization than single models and homogeneous ensembles [21]. The broccoli dataset was divided into three subsets and processed using the four classifiers.

TABLE III. SUMMARY OF CNN AND PROPOSED METHOD

Hyperparameter	CNN Models			
	ResNet50	***EfficientNet B2***	***VGG16***	***Improved CNN***
Kernel 1	ResNet50 Architecture	EfficientNet B2 Architecture	VGG16 Architecture	32 (3x3)
Kernel 2				64 (3x3)
Kernel 3				128 (3x3)
Activation Function	ReLU	ReLU	ReLU	ReLU
Layer Pooling	Average-pooling	Average-pooling	Average-pooling	Max Pooling
Layer Dense	512	512	512	512
	256		256	256
	128		128	128
	64		64	64
	3 class	3 class	3 class	3 class
Optimizer	Adam	Adam	Adam	Adam
Fully Connected	Softmax	Softmax	Softmax	Softmax
Epoch	50	50	50	50

1) Model CNNs: Table III summarizes several top CNN models commonly used to achieve satisfactory accuracy. Hyperparameter tuning was performed to optimize the results. Each model employs pre-trained ImageNet weights to enhance the model performance. The Rectified Linear Unit (ReLU) function. Additionally, the pooling layers in these models adopt the average pooling method, which calculates the average value of the features in the pooling area. Modifications were made to the fully connected layers of all CNN models. Each dense layer was configured with two to five dense layers, with the number of neurons set to 512, 256, 128, and 64, ending with a 3-class classification. For EfficientNetB2, the number of neurons was set to 512 with a 3-class classification. These modifications aim to optimize the model during the training process. The Adam optimizer was selected because it is frequently used in deep learning classification tasks and its relatively optimal performance compared to other optimizers. The number of epochs was set as 50 to achieve the most stable results during the training phase.

2) Improved CNN: The proposed model in this research is an improvement on the CNN, designed to achieve the best performance. The model modifies the hyperparameters of the convolutional and classification layers. For those convolutional layers, 32, 64, and 128 filters were used, each with a kernel size of 3x3 pixels. The activation function is the ReLU, and the pooling layers use the max-pooling method, which selects the highest value from each feature in the pooling area. The dense layer was configured with five layers containing 512, 256, 128, and 64 neurons, ending with a 3-class classification. The Softmax function was used for the output layer to produce the probabilities required for classification. The training process was run for 50 epochs to achieve optimal and stable results.

TABLE IV. RESULT FROM COMPARISON MODELS

	Features			**ACCURACY RESULT**
	SIZE	**SHAPE**	**COLOR**	
CNN Model	VGG16	VGG16	VGG16	94.00%
	VGG16	VGG16	EfficientNetB2	94.00%
	VGG16	VGG16	RESNET50	93.00%
	VGG16	VGG16	Improved CNN	94.00%
	VGG16	EfficientNetB2	VGG16	91.00%
	VGG16	EfficientNetB2	EfficientNetB2	91.00%
	VGG16	EfficientNetB2	RESNET50	90.00%
	VGG16	EfficientNetB2	Improved CNN	91.00%
	VGG16	RESNET50	VGG16	94.00%
	VGG16	RESNET50	EfficientNetB2	94.00%
	VGG16	RESNET50	RESNET50	93.00%
	VGG16	RESNET50	Improved CNN	94.00%
	VGG16	Improved CNN	VGG16	**95.00%**
	VGG16	Improved CNN	EfficientNetB2	**95.00%**
	VGG16	Improved CNN	RESNET50	**95.00%**
	VGG16	Improved CNN	Improved CNN	**95.00%**
	EfficientNetB2	VGG16	VGG16	94.00%
	EfficientNetB2	VGG16	EfficientNetB2	94.00%
	EfficientNetB2	VGG16	RESNET50	93.00%
	EfficientNetB2	VGG16	Improved CNN	94.00%
	EfficientNetB2	EfficientNetB2	VGG16	91.00%
	EfficientNetB2	EfficientNetB2	EfficientNetB2	91.00%
	EfficientNetB2	EfficientNetB2	RESNET50	90.00%
	EfficientNetB2	EfficientNetB2	Improved CNN	91.00%
	EfficientNetB2	RESNET50	VGG16	94.00%
	EfficientNetB2	RESNET50	EfficientNetB2	94.00%
	EfficientNetB2	RESNET50	RESNET50	93.00%
	EfficientNetB2	RESNET50	Improved CNN	94.00%
	EfficientNetB2	Improved CNN	VGG16	**95.00%**

	EfficientNetB2	Improved CNN	EfficientNetB2	**95.00%**
	EfficientNetB2	Improved CNN	RESNET50	**95.00%**
	EfficientNetB2	Improved CNN	Improved CNN	**95.00%**
	RESNET50	VGG16	VGG16	94.00%
	RESNET50	VGG16	EfficientNetB2	94.00%
	RESNET50	VGG16	RESNET50	93.00%
	RESNET50	VGG16	Improved CNN	94.00%
	RESNET50	EfficientNetB2	VGG16	91.00%
	RESNET50	EfficientNetB2	EfficientNetB3	91.00%
	RESNET50	EfficientNetB2	RESNET50	90.00%
	RESNET50	EfficientNetB2	Improved CNN	91.00%
	RESNET50	RESNET50	VGG16	94.00%
	RESNET50	RESNET50	EfficientNetB2	94.00%
	RESNET50	RESNET50	RESNET50	93.00%
	RESNET50	RESNET50	Improved CNN	94.00%
	RESNET50	Improved CNN	VGG16	**95.00%**
	RESNET50	Improved CNN	EfficientNetB2	**95.00%**
	RESNET50	Improved CNN	RESNET50	**95.00%**
	RESNET50	Improved CNN	Improved CNN	**95.00%**
	Improved CNN	VGG16	VGG16	94.00%
	Improved CNN	VGG16	EfficientNetB2	94.00%
	Improved CNN	VGG16	RESNET50	93.00%
	Improved CNN	VGG16	Improved CNN	94.00%
	Improved CNN	EfficientNetB2	VGG16	91.00%
	Improved CNN	EfficientNetB2	EfficientNetB2	91.00%
	Improved CNN	EfficientNetB2	RESNET50	90.00%
	Improved CNN	EfficientNetB2	Improved CNN	91.00%
	Improved CNN	RESNET50	VGG16	94.00%
	Improved CNN	RESNET50	EfficientNetB2	94.00%
	Improved CNN	RESNET50	RESNET50	93.00%
	Improved CNN	RESNET50	Improved CNN	94.00%
	Improved CNN	Improved CNN	VGG16	**95.00%**
	Improved CNN	Improved CNN	EfficientNetB2	**95.00%**
	Improved CNN	Improved CNN	RESNET50	**95.00%**
	Improved CNN	Improved CNN	Improved CNN	**95.00%**

The combination of features size, shape, and color demonstrates significant variations in classification performance, depending on the architecture pairings used (Table IV). It is noted that the model combination with the lowest accuracy is VGG16, EfficientNetB2, and ResNet50, achieving 90.00% accuracy. This model combination indicates that this combination struggles with certain types of data and consistently performs less than other combinations. On the other hand, combinations involving the Improved CNN consistently achieve the highest accuracy of 95.00%, whether paired with VGG16, EfficientNetB2, or ResNet50. This suggests that the Improved CNN effectively addresses the weaknesses of other models and exhibits better generalization capabilities. Other models, such as the combination of VGG16 with itself (VGG16-VGG16-VGG16) or other models like EfficientNetB2 and ResNet50, show varying accuracies in the 91.00% to 94.00%. These results indicate that the base architecture of VGG16 remains fairly reliable for classification tasks, though not as optimal as the Improved CNN. Meanwhile, combinations involving EfficientNetB2, with itself or other models, yield relatively lower accuracy, ranging from 90.00% to 94.00%. This suggests that the EfficientNetB2 architecture is less optimal for specific data scenarios.

Further analysis revealed that the Improved CNN and VGG16 combination achieved consistently high and stable accuracy between 94.00% and 95.00% compared to other model combinations. This finding may indicate that these two models have strengths in handling the given features, leading to a better performance. On the other hand, combinations involving ResNet50 show consistent performance in the range of 93.00% to 95.00%, although they tend to perform slightly lower than combinations involving the Improved CNN.

TABLE V. COMPARISON ACCURACY GRADING OF BROCCOLI

CNN Models	Dense Layer	Accuracy
Resnet50	5	85.26%
VGG16	2	88.42%
GoogleNet	2	84.21%
DenseNet121	2	83.16%
EfficientNetB2	4	86.32%
Perposed Method	5	95.00%

Table V compares the accuracy between several CNN models used for broccoli grading and the proposed method utilizing ensemble learning techniques. Previous researchers have also used these CNN models to determine the grading quality of fruits and vegetables, particularly broccoli, where the models were employed to analyze the color and texture features of the objects to reach specific decisions [22]. However, in this study, the quality of broccoli objects was determined from two perspectives, the top view and the side view, based on color, size, and shape.

Based on the table above, the results show a significant difference in accuracy, reaching 95%, whereas other CNN models that do not employ ensemble techniques show results below 90%. This study also includes parameter tuning, particularly in adjusting the number of dense layers, to achieve optimal results.

It is possible to classify three feature subsets: color, size, and shape, using the ensemble learning technique, where each feature input is trained individually without relying on other features. This approach allows the model to store information from each feature more effectively, enabling it to focus on reading and determining the grade of each test data item during the testing process.

D. Evaluation Model

The trained model was evaluated using performance metrics such as accuracy, precision, recall, and F1-Score. Below are some equations related to these methods.

In this study, the grading process was conducted using an Ensemble Learning method based on Convolutional Neural Network (CNN), incorporating various feature combinations (size, shape, and color) and several CNN models selected based on the experimental results listed in Table V. The confusion matrix for each combination was generated using metrics such as precision, Recall, F1-Score, and overall accuracy. The goal was to identify the best combination that delivered optimal performance for each grade (A, B, and C). The evaluation results revealed significant performance variations across the model combinations.

For Grade A, the combination of EfficientNetB2 + VGG16 + Improved CNN achieved the best performance with an F1-Score of 1.00, reflecting the model's ability to classify grade A cases perfectly (Table VI). This combination demonstrated a good balance between Precision and Recall. In contrast, VGG16 + EfficientNetB2 + ResNet50 had a lower F1-Score of 0.94 due to less optimal Precision and Recall than other combinations. This indicates that selecting the right model combination plays a significant role in the success of the classification for specific grades.

For Grade B, the combination of ResNet50 + Improved CNN + VGG16 achieved the highest F1-Score of 0.94, indicating its capability to capture more complex data patterns for this grade. The combinations of EfficientNetB2 + VGG16 + Improved CNN and Improved CNN + VGG16 + ResNet50 also performed well, each with an F1-Score of 0.92. Conversely, the combination of VGG16 + EfficientNetB2 + ResNet50 had the lowest performance with an F1-Score of 0.90, primarily owing to a low recall value of 0.86. This suggests that this combination struggled to accurately capture the characteristics of Grade B data.

For Grade C, the combinations of ResNet50 + Improved CNN + VGG16 and Improved CNN + VGG16 + ResNet50 delivered the best results with an F1-Score of 0.94. These combinations excelled in classifying this high-complexity grade, which tends to have more diverse data distributions. The combination of EfficientNetB2 + VGG16 + Improved CNN showed stable performance with an F1-Score of 0.91, whereas the combination of VGG16 + EfficientNetB2 + ResNet50 had a lower F1-Score of 0.88, attributed to a Precision score of only 0.81.

Overall, the combination of ResNet50 + Improved CNN + VGG16 delivered the best performance with an overall accuracy of 0.95, followed by EfficientNetB2 + VGG16 + Improved CNN with an accuracy 0.94. The Improved CNN + VGG16 + ResNet50 achieved an accuracy of 0.93, whereas VGG16 + EfficientNetB2 + ResNet50 had the lowest accuracy of 0.90. When used in model combinations, these results confirm that the Improved CNN significantly enhances classification performance, particularly in capturing complex features.

E. Grading

As shown in Fig. 6, the final stage is grading, where the model's predictions are translated into specific categories (e.g., Grade A, B, C) based on a voting technique. The model predictions can be determined starting from the lowest grade as shown in Fig. 6.

TABLE VI. CONFUSION MATRIX OF SUMMARY CNN MODELS AND PROPOSED METHOD

GRADE	Result Ensemble CNN Model											
	Size Features + Shape Features + Color Features											
	VGG16 + EffecientNetB2 + ResNet50			Improved CNN + VGG16 + ResNet50			EffecientNetB2 + VGG16 + Improved CNN			ResNet50 + Improved CNN + VGG16		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Grade A	0.96	0.92	0.94	0.96	1.00	0.98	1.00	1.00	1.00	0.96	1.00	0.98
Grade B	0.95	0.86	0.90	0.92	0.90	0.91	0.92	0.92	0.92	0.95	0.93	0.94
Grade C	0.81	0.96	0.88	0.91	0.91	0.91	0.91	0.91	0.91	0.94	0.94	0.94
Accuration			**0.90**			**0.93**			**0.94**			**0.95**

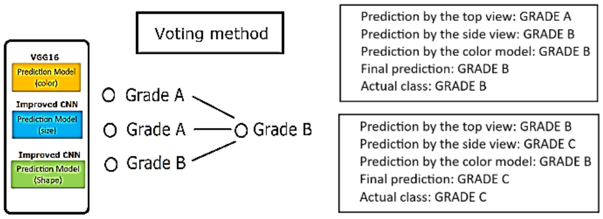

Fig. 6. The architecture of voting and grading.

Additional hyperparameter tuning is necessary for further discussion to achieve more accurate results. This research can also be expanded by developing image acquisition techniques for different lighting conditions and increasing the dataset size to enhance the algorithm's performance across a broader range of new data models.

V. Conclusion

This study on broccoli grading employs ensemble deep-learning techniques for training and testing processes. The combination of features—Size, Shape, and Color—significantly influences the prediction accuracy. Using an Improved CNN as the Shape feature substantially contributes to consistently achieving the highest performance, regardless of the models used for the Size and Color features. This indicates that the Improved CNN possesses strong generalization capabilities for various feature combinations.

Grading performance is heavily influenced by the accuracy achieved during testing, with model combinations that achieve the highest accuracy of 95% tend to produce more optimal grading results. This also proves that combining the predictive outputs from various classification models is highly effective in the grading process.

This method has significant potential for application to other data that require several parameters in the classification process. Using a combined model CNN technique, this method has proven to be capable of enhancing the performance in the classification process. The results showed a significant improvement compared to using a single CNN model alone.

This research can be further developed by enhancing image acquisition techniques using mobile phones or other devices under different lighting conditions. Additionally, it can be integrated into mobile phones and Arduino systems if, in the future, mass grading is performed using heavy machinery such as conveyor systems.

Acknowledgment

This research received financial support from the Ministry of Education, Culture, Research, and Technology of Indonesia under the 2024 Doctoral Dissertation Research Grant program (Grant No. 22179/IT3.D10/PT.01.03/P/B/2024).

References

[1] R.U. Syed, "Broccoli: a multi-faceted vegetable for health: an in-depth review of its nutritional attributes, antimicrobial abilities," Antibiotics, 2023, pp.1157-1207

[2] N.K. Raleni, "Pertumbuhan vegetatif dan produktivitas berbagai kultivar brokoli (brassica oleracea L. var. italica plenck.) Introduksi di Desa Batur, Kecamatan Kintamani, Kabupaten Bangli, Bali" Metamorfosa, 2015, pp.90-97.

[3] L. Guo, P. Wang, Z. Gu, X. Jin and R. Yang, "Proteomic analysis of broccoli sprouts by iTRAQ in response to jasmonic acid," Journal of Plant Physiology, 2017, P.16-25

[4] C. Zhou, J. Hue, Z. Xu1, J. Yue, H. Ye and G. Yang, "A monitoring system for the segmentation and grading of broccoli head based on deep learning and neural networks," Frontiers in Plant Science, 2020, pp. 402.

[5] P. M. Blok, F. K. van Evert, A. P. M. Tielen and E.J. Van Henten, "The effect of data augmentation and network simplification," J Field Robotics, 2020, pp. 1-20.

[6] K. Kusumam, T.´aˇs Krajn´ık, S. Pearson, and T. D. Grzegorz Cielniak, " 3D-Vision Based Detection, Localisation and Sizing of Broccoli Heads in the Field," Journal of Field Robotics, 2017, pp. 1505-1518.

[7] N. Ismail and O. A. Malik, "Real-time visual inspection system for grading fruits using computer vision and deep learning techniques, " Information Processing in Agriculture, 2022, pp.24–37.

[8] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: opportunities and challenges," Journal of King Saud University - Computer and Information Sciences, 2023, pp. 757-774

[9] M. Grandini, E. Bagli and G. Visani, "Metrics for multi-class classification: an overview", 2020, pp. 1–17.

[10] J. N.Torres, et all "A review of convolutional neural network applied to fruit image processing," appl science," 2020, pp. 3443

[11] S. Pardo, I. P. D´ıaz, V. B. Canedo and A. A. Betanzos, "Ensemble Feature Selection: Homogeneous and Heterogeneous Approaches," Knowledge-Based System, 2017

[12] K. Singh., D. Singh and Mishra, "Review: Convolutional neural networks and its architecture," International Journal of Health Sciences, 2022 , pp. 9183-9190

[13] J. L. Louedec, H. A. Montes, T. Duckett, G. Cielniak, "Segmentation and detection from organised 3D point clouds: a case study in broccoli head detection," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 64-65

[14] K. Zou, L. Ge, C. Zhang, T. Yuan and W. Li, "Broccoli Seedling Segmentation Based on Support Vector Machine Combined With Color Texture Features," IEEE Access, 2019, pp.168565-168574.

[15] Towards Autonomous Robotic Systems (TAROS 2019)

[16] H. A. Montes, G. Cielniak, and T. Duckett, "Model-Based 3D Point Cloud Segmentation for Automated Selective Broccoli Harvesting. In Annual Conference Towards Autonomous Robotic Systems," Springer, Cham, 2019, pp. 448-459

[17] Luzhen Ge ,ZhilunYang, ZheSun, GanZhang, MingZhang, Kaifei Zhang, Chunlong Zhang *, Yuzhi Tan * and Wei Li, "A method for broccoli seedling recognition in natural environment based on binocular stereo vision and gaussian mixture model". Sensors, 2019, pp. 1132.

[18] W. Bakasa and S. Viriri, "Stacked ensemble deep learning for pancreas cancel classification, using extreme gradient boosting," Front. Artif. Intell, 2023

[19] C. H. Chen, K. Tanaka, M. Kotera and K. Funatsu, "Comparison and improvement of the predictability and interpretability with ensemble learning models in QSPR applications," Journal of Cheminformatics, 2020.

[20] D. Opitz and R. Macli, "Popular Ensemble Methods: An Empirical Study," Journal of Artificial Intelligence Research, 1999

[21] P. Taeseung, S. Jihoon, P. Baekyung, M. Jeongsuk, C. Yoonkyung, "Generalizability evaluations of heterogeneous ensembles for river health predictions. Ecological Informatics 82," 2024.

[22] J. Wen, and J. He, "Agricultural development driven by the digital economy: improved EfficientNet vegetable quality grading," Front. Sustain. Food Syst. 8:1310042, 2024