



Automated seedling vigor estimation in cucumbers using digital image processing

 **Thanabodee Withunchettanan**¹

 **Raksak Sermsak**^{2,*}

 **Pichitra Kaewsorn**³

 **Kriengkri Kaewtrakulpong**²

¹ International School Bangkok, Nonthaburi 10900, Thailand

² Department of Farm Mechanics, Faculty of Agriculture, Kasetsart University, Bangkok 10900, Thailand

³ Department of Horticulture, Faculty of Agriculture, Kasetsart University, Bangkok 10900, Thailand

 *Corresponding author: raksak.s@ku.th

Article History

Submission: 10 May 2025

Revised: 28 June 2026

Accepted: 2 July 2026

Keywords

Cucumber seedling vigor

Digital image processing

RGB feature extraction

Non-destructive evaluation

Precision agriculture

DOI

[https://doi.org/10.65764/](https://doi.org/10.65764/tjas.2026.267412)

[tjas.2026.267412](https://doi.org/10.65764/tjas.2026.267412)

Abstract

Background and Objective: Farmers often rely on experience rather than quantitative indicators to determine seedling quality, resulting in inefficiencies in resource allocation. This study aimed to (1) compare seedling vigor and growth characteristics between open-pollinated (OP) and F1 hybrid cucumber seed types, and (2) develop a non-destructive predictive model for seedling vigor estimation based on digital image analysis.

Methodology: Seedling images were acquired under controlled LED lighting using a 5 MP digital camera positioned 30 cm above 14-day-old seedlings. A YOLOv8-based object detection model was applied to detect and isolate true leaf regions, from which pixel count and RGB color values were extracted as input features for a multiple linear regression model to predict the seedling vigor index (SVI).

Main Results: The YOLOv8 object detection model achieved a mean average precision (mAP@0.5) of 96.00%, precision of 93.40%, and recall of 94.40% in detecting true leaves. Multiple linear regression analysis was conducted using the Scikit-learn library in Python. Scikit-learn provides regression-based machine learning algorithms, including multiple linear regression. The equation was then applied to the training and test sets, using the pixel values of true leaves as the independent variable and SVI

as the dependent variable. Using multiple regression analysis, the trained model generated the equation $SVI = -2.27 + (2.84 \times 10^{-5} \times \text{pixel}) + (3.48 \times 10^{-5} \times R) + (-7.49 \times 10^{-2} \times G) + (2.06 \times 10^{-1} \times B)$, $R^2 = 0.57$ and $RMSE = 1.10$. The model achieved the test set, $r = 0.79$ and $RMSE = 1.34$. The model performance showed a correlation coefficient of 0.85 for OP data and 0.91 for F1 hybrid data.

Conclusions: The results demonstrate that digital image processing combined with object detection provides a non-destructive and effective approach to estimate seedling vigor quality. The predictive models developed from OP and F1 hybrid datasets indicate potential application in precision agriculture for automated seedling quality assessment and transplanting decision support.

How to cite : Withunchettanan, T., R. Sermsak, P. Kaewson and K. Kaewtrakulpong. 2026. Automated seedling vigor estimation in cucumbers using digital image processing. Thai J. Agric. Sci. 59(1): 50–67. <https://doi.org/10.65764/tjas.2026.267412>.

©2026

INTRODUCTION

Cucumbers (*Cucumis sativus*) are a significant crop in global agriculture, and they are known for their nutritional value and economic importance. As a widely cultivated crop, the quality of cucumber seeds directly impacts crop yield and productivity. Seed vigor is a critical determinant of seed quality and the success rate of crop production. However, cucumber seedlings are known for their sensitivity and susceptibility to various environmental stresses, pests, and diseases (Olechowska *et al.*, 2022). This inherent weakness can lead to lower success rates in seedling establishment. Suboptimal seedling transplanting has been reported to reduce marketable yield by 10–30% in vegetable crops, and the use of low-vigor seedlings increases replanting costs and delays crop establishment (Syed *et al.*, 2019; Samiei *et al.*, 2020). Specifically in cucumber production, seedling establishment failure and non-uniform growth directly translate into increased labor inputs and reduced gross returns at harvest. Therefore,

accurately estimating seed vigor is crucial for deciding whether to transplant the cucumber into a larger growing environment. The original cucumber seeds were open-pollinated (OP). The OP seeds often lead to lower yields due to their greater variability in plant characteristics, making them less ideal for agriculture (Fehr, 1987). To address these limitations, various methods have been implemented to improve seedling vigor, with F1 hybridization being one of the most effective (Fehr, 1987). However, F1 hybridization seeds are generally more expensive, which increases the cost of production for farmers.

Previous studies have demonstrated implementations made to evaluate seed vigor. However, existing research relies on destructive methods to measure the characteristics of a seedling after germination. For example, Demir and Mavi (2008) assessed the vigor of 16 commercial cucumber (*Cucumis sativus*) seed lots by treating them with different stress conditions such as low temperature, mechanical stress, and salt stress. The seed lots

underwent controlled deterioration (CD) and accelerated aging (AA) tests. This exposes seeds to high temperatures (45°C) and varying moisture contents and durations. Their germination and radicle emergence were then analyzed. Even though the methods provide a strong correlation ($P < 0.05$) between germination and the testing conditions, they are mere indicators of seedling vigor because they require damaging the seeds and are impractical for real-time seed quality monitoring (Demir and Mavi, 2008). There are a variety of methods to indicate the seedling vigor of a cucumber. Another popular method measures the electrical conductivity (EC) of solute leakage from seeds after imbibition. Seeds are soaked in deionized water, where the conductivity is measured 24 hours later. The higher the concentration, the more solute leakage there is, which translates to lower seed vigor. The EC test results in induced membrane damage and solute loss, making the seeds incapable of growing into seedlings (Matthews and Powell, 2006). The assessment of seedling vigor in cucumbers represents an aspect of agricultural production and seedling quality control. Such destructive methods are incompatible with the needs of modern agricultural systems that require real-time, nondestructive monitoring of seedling quality.

Previous studies have demonstrated the potential of combining image processing and machine learning for seedling assessment; however, these approaches have predominantly relied on either morphological features alone or deep learning architectures that require large, labeled datasets and substantial computational resources. A unified approach that integrates both spatial (pixel-based area) and spectral (RGB color) features within a lightweight regression framework, accessible to practical agricultural settings, has not been thoroughly investigated for cucumber seedlings. To address this gap, the present study applied multiple linear regression

analysis using the Scikit-learn library (Salama, 2024) to model the relationship between image-derived features and seedling vigor. One study highlights the effectiveness of linear regression in analyzing complex datasets (Alizadeh *et al.*, 2016). Similarly, research on seedling analysis has explored the role of deep learning in improving prediction accuracy. For instance, Samiei *et al.* (2020) developed a pipeline combining image processing and machine learning that utilizes deep neural networks and long short-term memory cells and achieved over 90% accuracy in classifying early seedling development stages in red clover (*Trifolium pratense*) and alfalfa (*Medicago sativa*) seedlings grown in soil. Their model classified seedlings into distinct phases during the first two weeks of sowing: first appearance of cotyledon, cotyledon opening, and appearance of the first true leaf. Another study presented an integrated approach using sensor data and image processing to quantify morphological traits of four-week-old pepper, tomato, cucumber, and lettuce seedlings, achieving a detection accuracy of 80–95% depending on variety and environmental conditions (Syed *et al.*, 2019). Image processing involves using computer algorithms to analyze and manipulate digital images to extract critical information, enhance visual quality, or detect specific features. In this case, digital image processing will be utilized to assess seedling vigor by analyzing features such as size, color, and growth patterns. This technology will be effective for farmers by providing a more accurate seedling vigor analysis while saving time.

Transplanting cucumber seedlings typically occurs 10–14 days after sowing, when the seedlings have developed their first true leaves. During the investigation, the percentage of seed germination is determined by germinating a set number of seeds under controlled conditions and recording the number of seeds that successfully germinated. The experiment focuses on cucumber seedlings, comparing the vigor

of seedlings grown from OP seeds and F1 hybrid seeds. Seedling vigor is calculated by multiplying the germination percentage by the mean seedling dry weight. Since F1 hybrid seeds typically exhibit higher vigor, they serve as the positive control in the object detection model, while OP seeds, known for their lower vigor, serve as the negative control. This comparison allows for a clear differentiation of seedling vigor levels using digital image analysis. The specific objectives of this study were: (1) to compare seedling germination, growth characteristics, and vigor index between OP and F1 hybrid cucumber cultivars; (2) to develop and evaluate a YOLOv8-based object detection model for automated true leaf identification in seedling images; and (3) to construct and validate a multiple linear regression model using pixel count and RGB color features for non-destructive prediction of the seedling vigor index.

MATERIALS AND METHODS

Seed Materials

Two types of cucumber seeds with different genetic backgrounds were used: open-pollinated (OP) and F1 hybrid cultivars. The OP seeds were 'Akkhani', supplied by the Tawan Tonkla Company in Thailand. The commercial F1 hybrid seeds were 'Micro-C 306', supplied by the East-West Seed Company.

Seed Quality, Seedling Growth and Vigor

The two types of OP and F1 hybrid cucumber seeds were sown in peat moss in a plastic tray with 100 holes in four replicates, with 100 seeds per replicate for each type. All the trays were put in the greenhouse under the same humidity dome with controlled sunlight and water. The mean light intensity inside the greenhouse was $227 \mu\text{mol m}^{-2}\text{s}^{-1}$, the highest light intensity during the day at noon was $542 \mu\text{mol m}^{-2}\text{s}^{-1}$, the temperature during the day and night was 28°C and 25°C , respectively, and the mean

relative humidity was 47.59%. The experiment was conducted in a greenhouse at Amphoe Mueang Nonthaburi, Nonthaburi, Thailand, from September 1, 2025 to September 14, 2025. A germination test in greenhouse conditions, mean germination time (MGT) was recorded daily for 11 days after sowing on four replicates with 100 seeds per replicate. Seedling growth and vigor were recorded at 14 days after sowing in four replicates with 50 plants per replicate.

Germination

Two seed lots – open-pollinated (OP) and F1 hybrid cultivars – were sown in peat moss using plastic trays with 100 holes in four replicates, with 100 seeds per replicate. All the trays were put in the same humidity dome with controlled sunlight and water. The percentage of normal seedlings was recorded for 11 days after sowing. Normal seedlings comprised a fully developed primary root, hypocotyl, and two cotyledons. Germination was calculated using equation 1 (ISTA, 2024):

Germination (%) =

$$\frac{\text{Number of normal seedlings}}{\text{Total seeds}} \times 100 \quad \text{--- (1)}$$

Mean Germination Time (MGT)

The number of normal cucumber seedlings was recorded daily from the first day after sowing until the day of the final count on day 11. The MGT was calculated using equation 2 (Ellis and Roberts, 1980):

$$\text{MGT} = \frac{\sum(n \cdot T)}{\sum n} \quad \text{--- (2)}$$

Where n is the number of normal seedlings that germinated on day T , and T is the number of days counted from the beginning of the test.

Seedling Growth and Vigor

Cucumber seedlings 14 days after sowing were randomized into four replicates with 50 plants per replicate. The following data were recorded: plant height was measured from the base level of the growing medium to the shoot apex. Stem diameter was measured at the cotyledon node. Leaf number was counted by observing the true leaves at full expansion. Fresh weights of shoot and root were measured immediately after harvesting using an analytical balance. Dry weights of shoot and root seedlings were separated into shoot and root portions at the cotyledon node immediately after fresh weight measurement. Each portion was placed in a pre-weighed paper envelope and dried in a forced-air oven at 80°C for 48 hours until a constant weight was achieved. Dry weight was recorded to the nearest 0.001 g. Shoot-to-root ratio was determined by dividing the shoot dry weight by the root dry weight. Seedling vigor index (SVI) was calculated using equation 3:

$$\text{SVI} = \text{Germination percentage} \times \text{Dry weight of seedling} \quad \text{--- (3)}$$

Independent samples t-tests were conducted to compare seedling growth characteristics between OP and F1 hybrid cucumber seedlings, using replicate means ($n = 4$ per seed type) as the statistical unit to avoid pseudo-replication. The SVI was calculated per replicate using equation 3, with the germination percentage of each replicate tray multiplied by the mean dry weight of seedlings in that replicate. For the regression analysis, the SVI value assigned to each individual seedling observation was derived from the germination percentage of its respective replicate tray multiplied by the dry weight of that individual plant.

Image Acquisition

Images were taken by placing each individual

seedling inside an inspection chamber which contained the lighting system. Each seedling was allowed a 5-second stabilization period after placement, and a single image was captured per seedling. The total processing time per seedling, from placement to removal, was approximately 15 s, using an Imaging Source DFK 72BUC02 Digital Camera and a 1/2.5-inch Onsemi CMOS MT9P006 Sensor. The resolution was $2,592 \times 1,944$ (5 MP), supporting up to six frames per second (fps). The pixel size was $2.2 \mu\text{m}$ (horizontal) $\times 2.2 \mu\text{m}$ (vertical), and the camera utilized a rolling shutter mechanism, as shown in Figure 1. The camera was placed 30 cm above the canopy of the seedlings 14 days after sowing. Qualitative observations indicate that vigorous seedlings exhibit longer shoots and a larger true leaf. To ensure color consistency across all image sessions, the camera exposure and white balance settings were locked prior to data collection, and the LED lighting array was maintained at a fixed intensity throughout the experiment. These measures were implemented to minimize inter-sample variation in illumination. A formal color checker calibration was not applied in this study; future work will incorporate a standard color reference card for per-session color calibration to further improve the accuracy and reproducibility of RGB feature extraction.

True Leaf Detection using YOLO (You Only Look Once)

The object detection model focused on isolating true leaf regions from seedling images to enable accurate vigor assessment. Since true leaves are key indicators of seedling vigor, distinguishing them from cotyledons and background elements is essential. YOLOv8 (Jocher *et al.*, 2023), an anchor-free single-stage object detection algorithm, was selected for this task due to its real-time detection speed, strong performance with small, annotated datasets, and effective boundary delineation capability suited to leaf

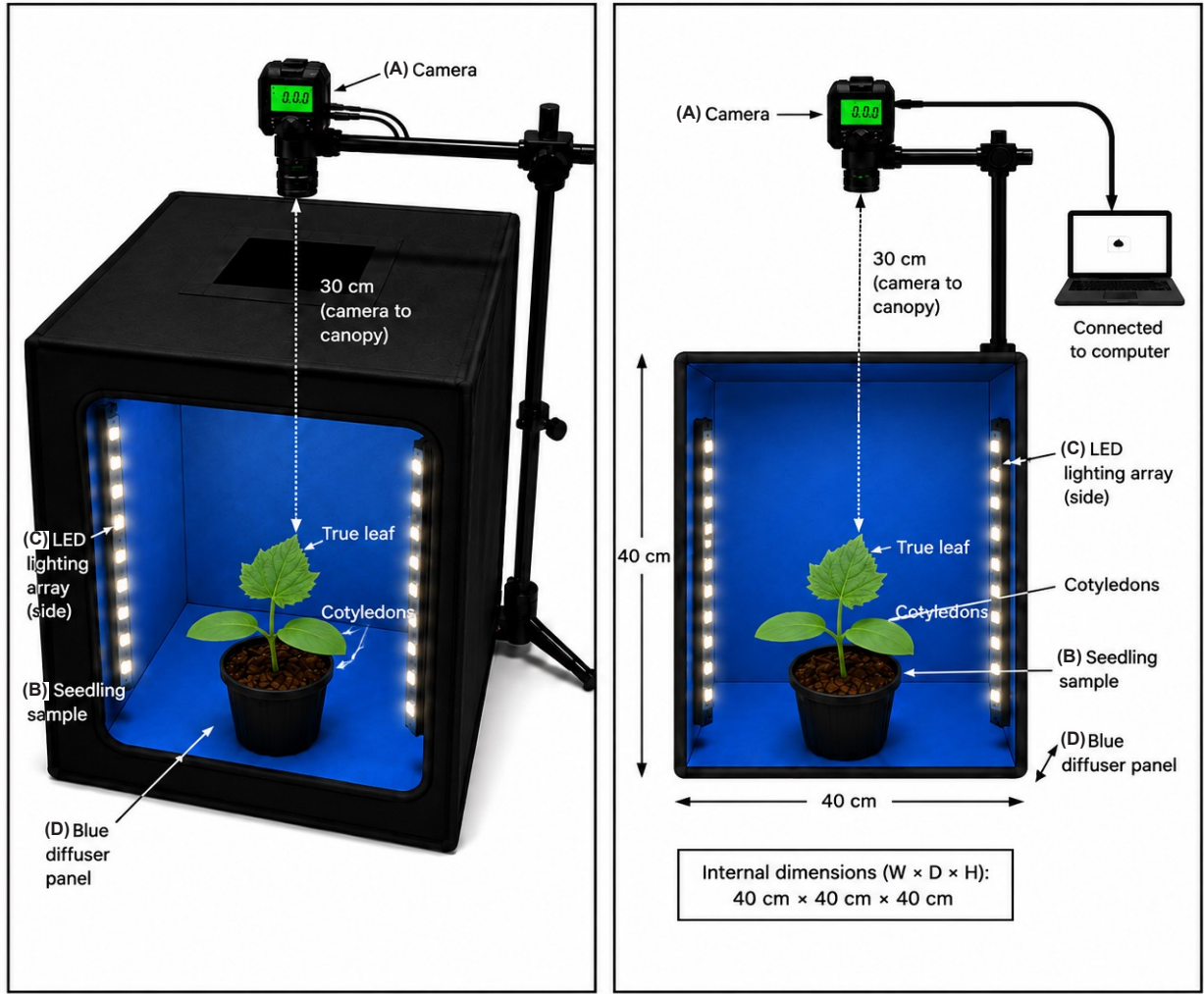


Figure 1 Image acquisition setup showing the inspection chamber (internal dimensions: 40 cm × 40 cm × 40 cm), camera position (5 MP digital camera mounted 30 cm above the seedling canopy), and LED illumination conditions. Annotated components include: (A) camera, (B) seedling sample, (C) LED lighting array, and (D) blue diffuser panel.

structure identification in agricultural applications (Shen *et al.*, 2025). The YOLOv8n (nano) architecture was used as the base model. Training was conducted for 100 epochs with a batch size of 16. The loss function combined classification loss (cross-entropy), bounding box regression loss (CIoU), and distribution focal loss. Model performance was evaluated using mean average precision at IoU threshold 0.5 (mAP@0.5), precision, and recall (Redmon *et al.*, 2016). The YOLO model's performance was evaluated using standard object detection metrics, achieving a mean average

precision (mAP), precision and recall. Training data was manually annotated using polygon tools in Roboflow, a platform optimized for YOLO-based models (Zendehdel *et al.*, 2023). The dataset consisted of 400 images of OP and F1 hybrid cucumber seedlings, showing both cotyledons and true leaves, as illustrated by the representative raw and YOLOv8-detected images in Figure 2. The dataset was split into 160 images (40%) for training and 240 images (60%) for independent testing. This split was intentionally adopted to subject the model to a more stringent

evaluation under a larger independent test set, simulating deployment conditions where generalizability is prioritized. The training and training loss curves confirmed stable convergence without evidence of overfitting. The images were captured under controlled lighting conditions with a consistent background to minimize variability. Each image was manually annotated to identify and label cotyledons and true leaves, providing ground truth data for model training. Each seedling image was labeled to differentiate true leaf regions from cotyledons, soil, and other plant structures. The annotation process incorporated diverse seedling variations, including differences in leaf size and shape, to enhance the model's generalization.

Feature Extraction

The quantification of pixel count serves as a key feature for predictive modeling as a digital proxy for leaf area measurement. The extraction of red (R), green (G), and blue (B) color values from true leaf pixels represents a technique for non-destructive assessment of plant health. The RGB color space was adopted for its simplicity, direct compatibility with the camera output format, and accessibility within the Python-based image processing pipeline. Since all images were acquired under strictly controlled and constant LED illumination within a closed inspection chamber, the influence of brightness variation on extracted RGB values was minimized by experimental design. The authors acknowledge that CIE $L^*a^*b^*$, as a perceptually uniform color space, would provide

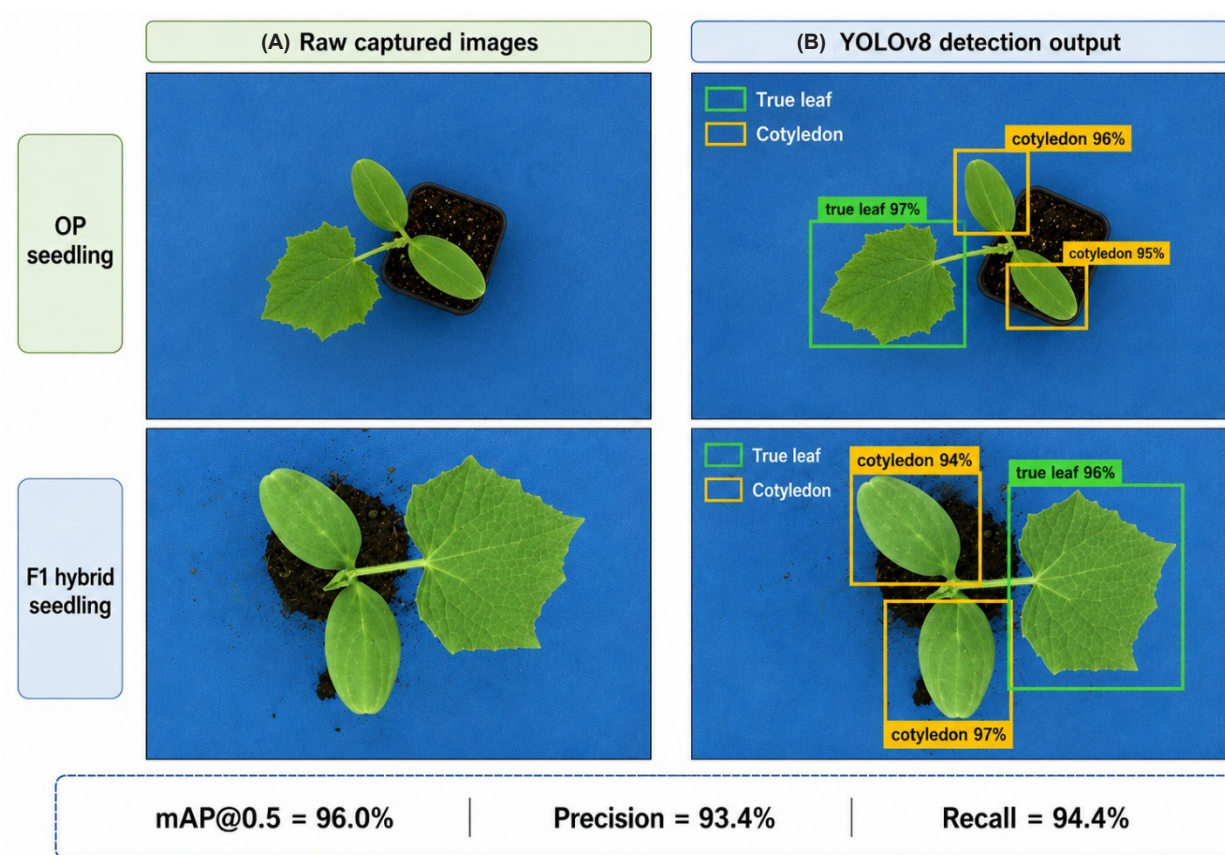


Figure 2 Representative images showing (A) raw captured images of OP and F1 hybrid seedlings; (B) YOLOv8 detection output with annotated bounding boxes distinguishing true leaf and cotyledon regions.

greater robustness to residual lighting variation by decoupling luminance from chromatic information; this is identified as a priority for future work. The pixel count and RGB color values extracted from the true leaf region were analyzed using the Python programming language. The R, G, and B channel values of pixels within the detected true leaf region were extracted, providing insight into the leaf's color characteristics. These color intensities are indicative of various physiological conditions of the seedling, such as chlorophyll content and overall plant health,

which are directly related to the vigor index. The total pixel count within the detected true leaf region was then quantified as a key feature for predictive modeling (Min *et al.*, 2022). The analytical procedure used for image-derived feature extraction is summarized in Figure 3, including YOLOv8-based true leaf detection, isolation of the detected true leaf region, generation of the segmentation mask, quantification of total leaf pixel count, and extraction of mean R, G, and B color intensity values.

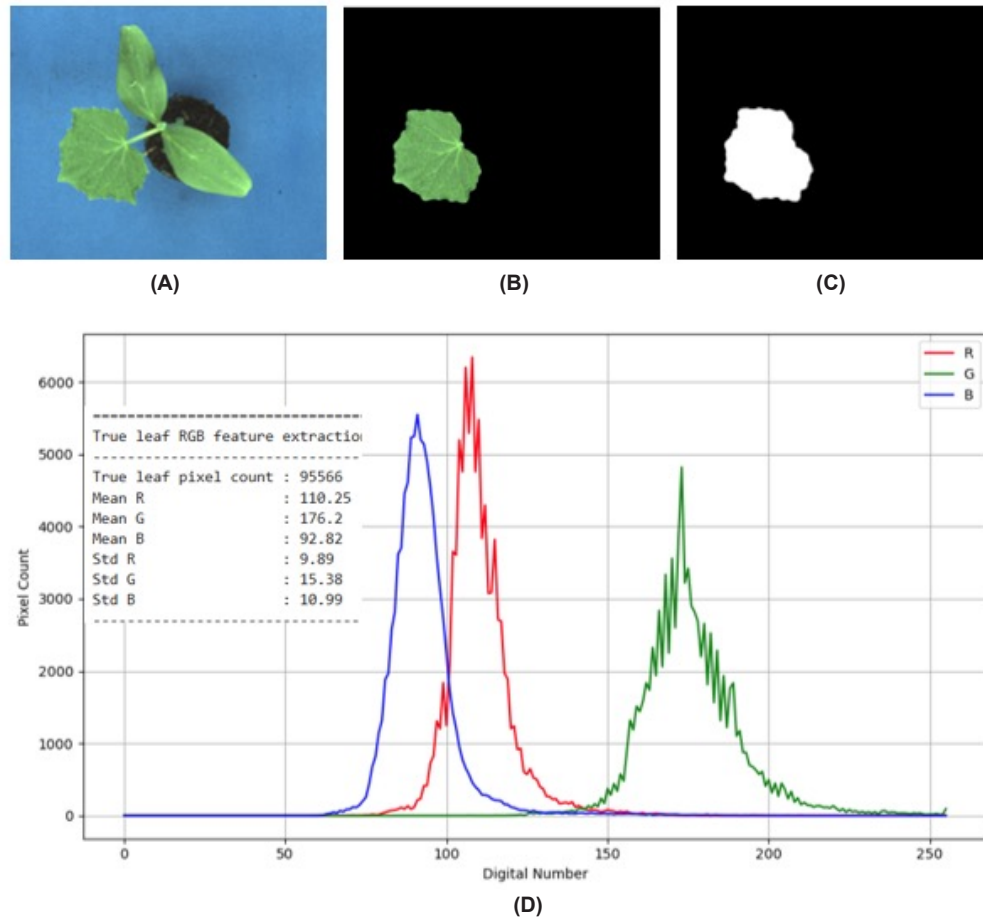


Figure 3 RGB feature extraction pipeline applied to a YOLOv8-detected true leaf region: (A) original image, (B) isolated true leaf region cropped from the original seedling image following YOLO detection, (C) binary pixel segmentation mask used to quantify total leaf pixel count, where white pixels represent the detected leaf area, and (D) mean R, G, and B color intensity values extracted from the segmented region, which served as independent variables alongside pixel count in the multiple linear regression models.

Image Model Analysis

Multiple regression analysis was employed to establish relationships between image-derived features and SVI. Following YOLOv8-based detection, the segmented true leaf region from each seedling image was processed to derive four input features: total pixel count within the detected mask (used as a proxy for leaf area) and mean R, G, and B channel values extracted from the masked region. These four features served as independent variables in the multiple linear regression model, with SVI as the dependent variable. The analysis was implemented using the Scikit-learn library (Pedregosa *et al.*, 2011), a widely adopted machine learning framework in the Python programming language. The study implemented an 80/20 train-test split ratio at the individual-seedling level, allocating 80% of the dataset (192 seedling observations) for model training and the remaining 20% (48 seedling observations) for model testing. This split was applied consistently across all three regression models (combined, OP-specific, and F1 hybrid-specific) using the `train_test_split` function in scikit-learn with a fixed random state to ensure reproducibility. Because the split was performed at the individual-seedling level rather than at the replicate-tray level, seedlings originating from the same replicate tray could potentially be assigned to both training and testing datasets. This approach was used for preliminary model development under a limited dataset size, and the potential replicate-related dependency is acknowledged as a limitation of the present study. This methodological decision was made to evaluate model performance under conditions of limited training data, simulating real-world scenarios where extensive labeled datasets may not be available in agricultural applications. The performance of the regression model was evaluated using multiple statistical metrics to provide a comprehensive

assessment of prediction accuracy. These included the coefficient of determination (R^2) and root mean squared error (RMSE) for the training model, and the Pearson correlation coefficient (r), R^2 , and RMSE for the test model, reported consistently across all three regression models (combined, OP-specific, and F1 hybrid-specific). The overall analytical workflow for automated cucumber seedling vigor estimation is summarized in Figure 4.

RESULTS AND DISCUSSION

Germination and Mean Germination Time

The germination percentage and mean germination time of OP and F1 hybrid types of cucumber seeds were non-significant difference (Table 1). Germination percentages of OP and F1 hybrid types were 92.00% and 97.00%, respectively. Similarly, the mean germination times of OP and F1 hybrid types were 3.55 and 3.31 days after sowing, respectively. This further suggests that F1 hybrid seeds exhibit superior germination rates and demonstrate excellent uniformity and consistency in germination kinetics. In contrast, OP seeds showed variability, which indicates a less uniform response due to genetic heterogeneity. This is supported by McDonald and Lubell-Brand (2024), who reported that F1 hybrid seeds produced greater crop uniformity than open-pollinated seeds.

Seedling Growth and Vigor

As seen in Table 2, cucumber seedlings of OP and F1 hybrid types showed a significant difference in stem height, stem diameter, fresh weight, dry weight, and SVI. F1 hybrid had a greater average diameter of 3.42 mm compared to 2.84 mm in OP, which is evidence of stronger seedling development. This indicates that F1 hybrid seedlings may have a more vigorous structural development at the growth stage. The data also show a noticeable difference

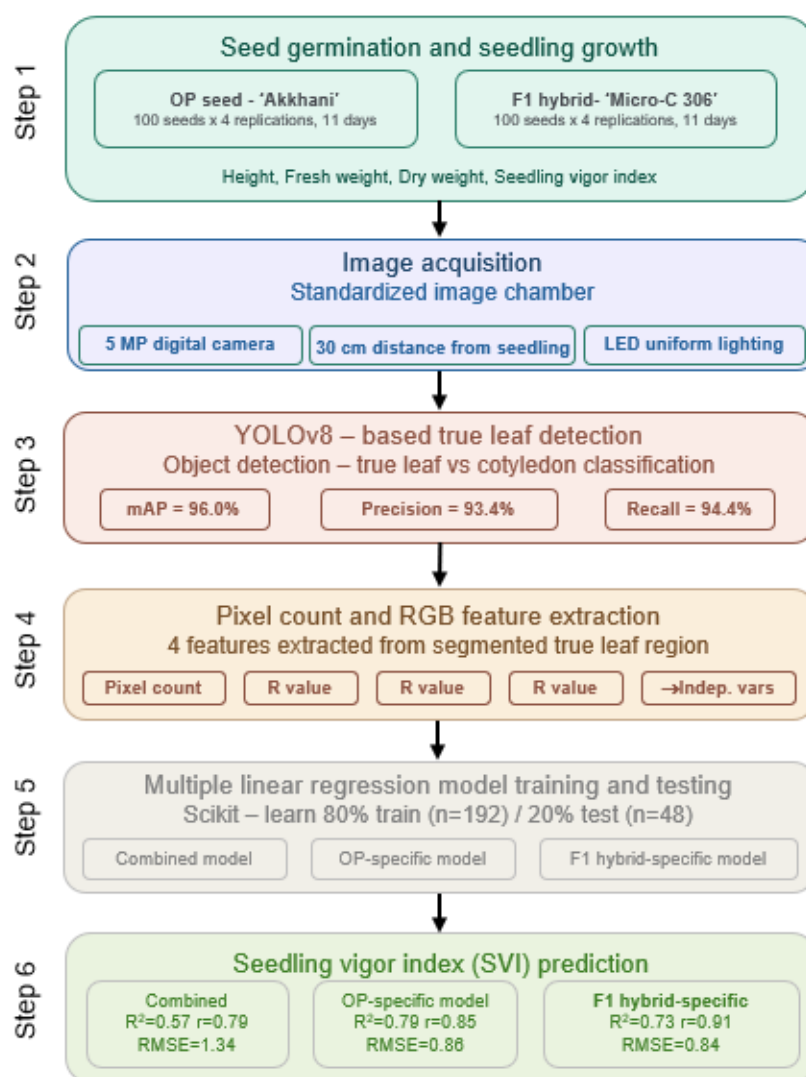


Figure 4 Analytical pipeline for automated cucumber seedling vigor estimation illustrating six sequential steps: (1) seed germination and seedling growth measurement for open-pollinated (OP) and F1 hybrid cultivars; (2) standardized image acquisition using a 5-megapixel camera under controlled LED lighting at 30 cm distance; (3) YOLOv8-based true leaf detection with classification of true leaf and cotyledon regions (4) pixel count and RGB color feature extraction from segmented true leaf regions; (5) multiple linear regression model development using Scikit-learn with an 80/20 train-test split; and (6) seedling vigor index (SVI) prediction results for the combined.

between the average plant height of the two treatments, in which OP generally had a greater height than the seedlings in the F1 hybrid. One potential interpretation is that OP seedlings allocate more resources toward vertical growth, possibly as an adaptive response to compete for light. Thus, the effect of competition for

light in OP type is shown in excessive stem height, which results in poor chlorophyll production and weak plant development. On the other hand, F1 hybrid prioritizes resource allocation for other traits, such as root development for lateral growth, which enhances overall stability and vigor. The overall superiority in

stem diameter, fresh weight, and dry weight suggests that the hybrid cucumber type had a better potential for vigorous growth. To support the hypothesis, it is evident in Table 2 that the average seedling vigor index of the F1 hybrid is higher than that of OP. The earlier hypothesis in this investigation was that the dry weight of a seedling is related to its leaf size. Since dry weight is directly related to seedling vigor

index, it is predicted that seedling vigor and leaf size are correlated. As shown in seedling analysis research by Samiei *et al.* (2020), leaf size could be translated to pixel values from a computational tool. Therefore, this presents a possibility of integrating image processing to interpret and estimate the seedling vigor index.

Table 1 Germination percentage and mean germination time of open-pollinated (OP) and F1 hybrid cucumber seeds at 11 days after sowing

Cucumber type	Germination (%)	Mean germination time (day)
Open-pollinated	92.00 ± 4.97	3.55 ± 0.33
F1 hybrid	97.00 ± 0.82	3.31 ± 0.05
P-value	0.09	0.15

Note: Values are presented as mean ± standard deviation (n = 4 replicates of 100 seeds each). P-value from independent samples t-test.

Table 2 Stem height, stem diameter, fresh weight, dry weight and seedling vigor index of open-pollinated (OP) and F1 hybrid cucumber seedlings at 14 days after sowing

Cucumber type	Stem height (mm)	Stem diameter (mm)	Fresh weight (g)	Dry weight (g)	Seedling vigor index
Open-pollinated	73.07 ± 9.75 ^a	2.84 ± 0.39 ^b	0.93 ± 0.26 ^b	0.06 ± 0.02 ^b	5.16 ± 1.66 ^b
F1 hybrid	64.63 ± 11.51 ^b	3.42 ± 2.77 ^a	1.28 ± 0.30 ^a	0.08 ± 0.02 ^a	7.42 ± 1.64 ^a
P-value	0.00	0.02	0.00	0.00	0.00

Note: Means within the same column followed by different superscript letters (a, b) are significantly different (P < 0.05). Values are presented as mean ± standard deviation (n = 4 replicates of 50 plants each).

Analysis of Seedling Leaves Structure Detection

The distinction between true leaves and cotyledons is particularly important in seedling vigor assessment. Cotyledons are the first leaves to appear after germination and differ morphologically from true leaves that develop later. The ability to accurately detect and differentiate both structures provides more comprehensive data for vigor estimation. The study examines polygon annotation tools in Roboflow for manually annotating training data, specifically optimized for YOLO-based models. In the present study, the YOLOv8 object detection model achieved

a mean average precision (mAP@0.5) of 96.00%, precision of 93.40%, and recall of 94.40%, as shown in Figure 5. The high precision indicates a low occurrence of false positives, ensuring that non-true leaf structures were rarely misclassified. The high recall suggests that nearly all true leaves were successfully detected, minimizing the risk of missing relevant plant structures. This aligns with Koirala *et al.* (2019), who emphasized the importance of high precision and recall in agricultural computer vision applications for reliable crop monitoring and management decisions.

Image Model of Seedling Vigor

The regression analysis included 30 individual seedlings randomly selected from each of the four replicate trays of OP (totaling 120 seedlings), and 30 individual seedlings from each of the four F1 hybrid trays (totaling 120 seedlings), for an overall total of 240 seedlings. Random selection was performed using a random number table to assign unique plant IDs within each tray. Images were captured from each selected seedling prior to destructive measurement,

ensuring that image-derived features and SVI values were obtained from the same individual plants at 14 days after sowing. Leaf area was quantified as the total pixel count within the YOLOv8-detected true leaf mask, which served as a non-destructive proxy for leaf size. In image analysis, the color intensity values of R, G, and B channels were extracted from each true leaf region. The relationship between SVI and the four image features (pixel count, R, G, B) is shown in Table 3.

Table 3 Pearson correlation coefficient between seedling vigor index (SVI) and image parameters

Factor	Pixel value	R	G	B
Seedling vigor index	0.64**	0.16*	0.18**	0.26**
Pixel value		0.12 ^{ns}	0.15*	-0.02 ^{ns}
R			0.86**	0.75**
G				0.86**

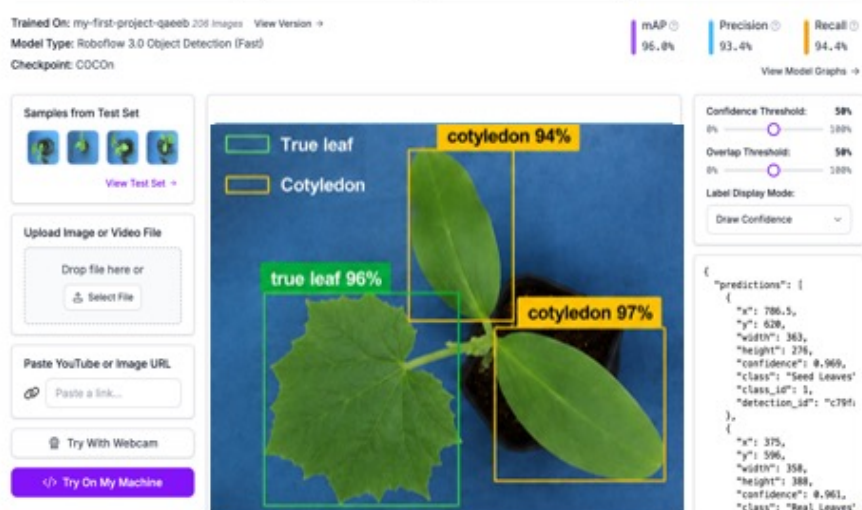
Note: ** Significantly different at $P < 0.01$. * Significantly different at $P < 0.05$. ns = not significant. R = red, G = green, B = blue.

The 240 seedling observations were randomly split into training (80%) and test (20%) datasets at the individual-seedling level using the `train_test_split` function in scikit-learn with a fixed random state for reproducibility. This split was applied consistently across the combined, OP-specific, and F1 hybrid-specific regression models. Because the random split was conducted at the individual-seedling level, possible replicate-tray dependency could not be completely excluded and should be considered when interpreting the predictive performance. Multiple regression analysis was performed to model the relationship between SVI and these image features. The derived regression equation for the training model was $SVI = -2.27 + (2.84 \times 10^{-5} \times \text{pixel}) + (3.48 \times 10^{-5} \times R) + (-7.49 \times 10^{-2} \times G) + (2.06 \times 10^{-1} \times B)$ with an R^2 of 0.57 and an RMSE of 1.10. As shown in Figure 6A, the test set yielded a Pearson correlation coefficient of 0.79 and an RMSE of 1.34, indicating a moderate positive relationship between observed and predicted SVI values for

the combined dataset. Considering that the SVI range in this study spanned approximately 2–12 units, an RMSE of 1.34 represents a relative error of approximately 11%, which is considered suitable as a general-purpose preliminary screening tool where cultivar identity is unknown. However, because regression model performance was evaluated using a single 80/20 train-test split without k-fold cross-validation or external validation, the predictive performance should be interpreted as preliminary and should be confirmed using more rigorous validation procedures in future studies. The positive correlation between total pixel number and SVI ($r = 0.64$, $P < 0.01$) demonstrates that digital quantification of seedling size via pixel count serves as an effective proxy for seedling vigor. This finding aligns with research by Dell'Aquila (2009), who established that pixel-based measurements of seedling area can accurately predict vigor-related parameters in various vegetable crops.



(A) Raw captured image



(B) YOLOv8 detection output

Figure 5 Test results of the YOLOv8-based object detection model for true leaf and cotyledon detection in cucumber seedlings: (A) raw image of a cucumber seedling captured in the inspection chamber under controlled LED illumination against the blue diffuser background, showing one true leaf and two cotyledons and (B) YOLOv8 detection output with bounding boxes distinguishing the true leaf from the cotyledons together.

The seed-type-specific models developed in the present study achieved Pearson correlation coefficients of 0.85 for OP and 0.91 for F1 hybrid, which are broadly consistent with the range reported in related image-based seedling vigor studies. It should be noted that direct numerical comparison across studies is limited by differences in crop species, vigor assessment methods, and image feature sets.

Matthews *et al.* (2012) reviewed seed quality methods and Kaewson and Sermsak (2023) found positive correlation between measured and image-estimated tomato SVI ($r = 0.94$). The results of the present study suggest that integrating both pixel-based area and RGB spectral features within a single regression framework may capture complementary physiological aspects of seedling vigor, supporting observations

by Dell'Aquila (2009) regarding the significance of seedling color as a vigor indicator.

Given that OP had a different growing pattern than the F1 hybrid, a separate analysis was made on 120 OP types to evaluate a model that fits better for a single OP cucumber type. The model evaluates an equation $SVI = 3.49 + (2.53 \times 10^{-5} \times \text{pixel}) + (-1.05 \times 10^{-3} \times R) + (-4.70 \times 10^{-2} \times G) + (9.10 \times 10^{-2} \times B)$, with an R^2 of 0.79 and an RMSE of 0.75. As shown in Figure 6B, when the model was applied to the test set comprising 20% of the 120 OP seedlings, the Pearson correlation coefficient was 0.85, and the RMSE was 0.86. An RMSE of 0.86 on an SVI range of approximately 2–9 units represents a relative error of approximately 8%, which is considered acceptable for cultivar-specific seedling screening applications in nursery operations. This demonstrates that applying the multiple regression model to a single cucumber cultivar type results in substantially higher prediction accuracy compared with the combined model. The same approach of separating analysis was made on F1 hybrid type. The model evaluates an equation $SVI = 5.81 + (3.20 \times 10^{-5} \times \text{pixel}) + (1.63 \times 10^{-3} \times R) + (-3.04 \times 10^{-2} \times G) + (3.71 \times 10^{-2} \times B)$, with an R^2 of 0.73 and an RMSE of 0.81. As shown in Figure 6C, when the model was applied to the test set comprising 20% of the 120 F1 hybrid seedlings, the Pearson correlation coefficient was 0.91, and the RMSE was 0.84. The higher correlation of the F1 hybrid-specific model ($r = 0.91$) compared with the combined model ($r = 0.79$) is attributable to the greater phenotypic uniformity of F1 hybrid seedlings arising from their homozygous genetic background. When data from both OP and F1 hybrid seed types are pooled into a single model, the regression equation must accommodate two biologically distinct growth patterns and SVI distributions, which inevitably reduces the overall fit. In the present study, type-specific models achieved higher predictive performance than the

combined model. Therefore, type-specific models are recommended when cultivar identity is known. The combined model is proposed as a general-purpose preliminary screening tool. Nevertheless, these findings were based on a single random 80/20 train-test split, and no cross-validation or independent external validation dataset was used. Therefore, the reported model performance should be considered an initial evaluation rather than a final validation of model generalizability. F1 hybrid cucumber seedlings demonstrate superior uniformity primarily due to their genetic structure (Staub *et al.*, 2008). In the present study, evaluation of the multiple regression model coefficients revealed that pixel count was positively associated with SVI, indicating that greater leaf area corresponds to higher seedling vigor. The negative coefficient for R suggested that elevated red channel values may reflect stress-related anthocyanin accumulation or reduced chlorophyll content, both associated with lower plant vigor (Dell'Aquila, 2009). The G channel also had a negative coefficient of greater magnitude than R, which may indicate that overly high green reflectance is associated with chlorophyll saturation under low-light conditions or early senescence, leading to reduced vigor. The positive coefficient for B potentially reflects higher leaf reflectance under controlled LED lighting conditions, which has been associated with younger, more vigorous leaf tissue. It should be noted that these interpretations are based on observational regression coefficients under controlled lighting; physiological validation through independent biochemical measurements would be required to confirm these relationships. Furthermore, the model was developed under controlled laboratory conditions, and performance may be affected by variability in ambient lighting, seedling orientation, or camera calibration in field deployment scenarios.

The integration of both dimensional (pixel number) and spectral (RGB) parameters into a single predictive model represents a significant advancement over single-parameter approaches. Dell'Aquila (2009) previously reported that morphological parameters alone explained 82–87% of vigor variation in various

vegetable seedlings, while color parameters independently accounted for 78–85%. Our results are consistent with Kaewson and Sermsak (2023), who reported a strong relationship ($r = 0.94$) between measured and image-estimated tomato SVI using pixelbased information.

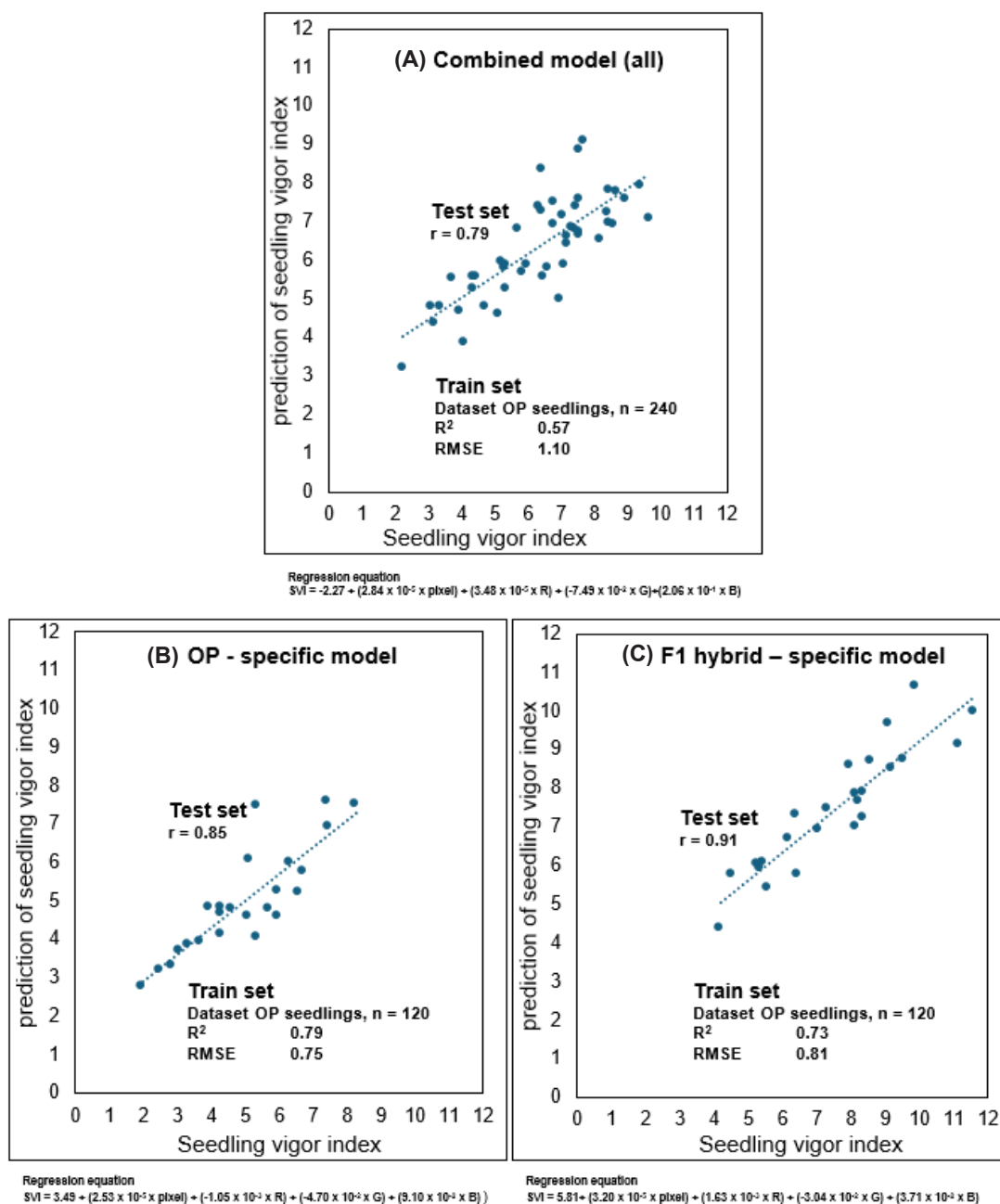


Figure 6 Relationship between observed and predicted seedling vigor index (SVI) for (A) combined model, (B) OP-specific model, and (C) F1 hybrid-specific model. The x-axis represents observed SVI, and the y-axis represents predicted SVI. The dashed line indicates the fitted trend.

CONCLUSIONS

In the present study, germination percentage and mean germination time did not differ significantly between OP and F1 hybrid seed types ($P > 0.05$), indicating comparable germination capacity under the conditions of this study. F1 hybrid cucumber seedlings demonstrated significantly higher seedling vigor index, stem diameter, fresh weight, and dry weight compared with open-pollinated seedlings at 14 days after sowing. The YOLOv8-based object detection model achieved high accuracy ($mAP@0.5 = 96.00\%$, precision = 93.40% , recall = 94.40%) in detecting and isolating true leaf regions from seedling images. Multiple linear regression models incorporating pixel count and RGB color features demonstrated the feasibility of non-destructive prediction of the seedling vigor index, with Pearson correlation coefficients of 0.85 and 0.91 for the OP-specific and F1 hybrid-specific models, respectively. These results suggest the potential of this image-based approach for integration into precision agriculture systems, including automated seedling quality assessment and transplanting decision support. A limitation of the present study is that regression model evaluation was based on a single individual-seedling-level 80/20 train-test split. Therefore,

possible replicate-tray dependency could not be completely excluded. Future studies should apply tray-level or replicate-level data splitting, repeated cross-validation, and external validation datasets to provide a stricter assessment of model generalizability. Future work should investigate the use of CIE $L^*a^*b^*$ color space to reduce the influence of lighting variation, incorporate standard color checker calibration, and evaluate model performance across additional cucumber cultivars and growing environments. In addition, because the present regression models were evaluated using only a single 80/20 train-test split without k-fold cross-validation or external validation, future studies should apply repeated cross-validation and independent validation datasets to provide a more rigorous assessment of model generalizability.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest related to this research.

ACKNOWLEDGMENTS

The authors thank the Faculty of Agriculture at Kasetsart University for providing the necessary materials and laboratory facilities for data collection and analysis.

REFERENCES

- Alizadeh, M.A., H.A. Arab, R. Tabaie, M. Nasiri and A.A. Jafari. 2016. Evaluation of seed emergence, seedling vigor enhancement of some populations from *Satureja bachtarica* with chemical, mechanical and physiological treatment. *Plant Breed. Seed Sci.* 74: 37–44. <https://doi.org/10.1515/plass-2016-0013>.
- Dell'Aquila, A. 2009. Digital imaging information technology applied to seed germination testing. A review. *Agron. Sustain. Dev.* 29: 213–221. <https://doi.org/10.1051/agro:2008039>.
- Demir, I. and K. Mavi. 2008. Seed vigour evaluation of cucumber (*Cucumis sativus* L.) seeds in relation to seedling emergence. *Res. J. Seed Sci.* 1(1): 19–25. <https://doi.org/10.3923/rjss.2008.19.25>.
- Ellis, R.H. and E.H. Roberts. 1980. Improved equations for the prediction of seed longevity. *Ann. Bot.* 45(1): 13–30. <https://doi.org/10.1093/oxfordjournals.aob.a085797>.
- Fehr, W.R. 1987. *Principles of Cultivar Development: Theory and Technique*. Macmillan Publishing Company, New York, USA.

- ISTA (International Seed Testing Association). 2024. International Rules for Seed Testing. International Seed Testing Association, Wallisellen, Switzerland.
- Jocher, G., A. Chaurasia and J. Qiu. 2023. Ultralytics YOLOv8 (Version 8.0.0) [Computer Software]. GitHub. Available Source: <https://github.com/ultralytics/ultralytics>.
- Kaewsorn, P. and R. Sermsak. 2023. Application of digital image processing for seedling vigor estimation of primed tomato seed. *Asia Pac. J. Sci. Technol.* 28(2): APST-28-02-03. <https://doi.org/10.14456/apst.2023.19>.
- Koirala, A., K.B. Walsh, Z. Wang and C. McCarthy. 2019. Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO'. *Precis. Agric.* 20: 1107–1135. <https://doi.org/10.1007/s11119-019-09642-0>.
- Matthews, S. and A. Powell. 2006. Electrical conductivity vigor test: Physiological basis and use. *Seed Testing International*. 131: 32–35.
- Matthews, S., E. Noli, I. Demir, M. Khajeh-Hosseini and M.H. Wagner. 2012. Evaluation of seed quality: From physiology to international standardization. *Seed Sci. Res.* 22(S1): S69–S73. <https://doi.org/10.1017/S0960258511000365>.
- McDonald, M.M. and J.D. Lubell-Brand. 2024. F1 hybrid seed can enhance cannabis crop uniformity and yield. *HortScience*. 59(12): 1795–1799. <https://doi.org/10.21273/HORTSCI18197-24>.
- Min, S.H., K. Kaewtrakulpong, T. Phaosang and R. Sermsak. 2022. Image processing method to check maturity index of 'SEIN TA LONE' mango in Myanmar. *Suranaree J. Sci. Technol.* 29(4): 0020017.
- Olechowska, E., R. Słomnicka, K. Kaźmińska, H. Olczak-Woltman and G. Bartoszewski. 2022. The genetic basis of cold tolerance in cucumber (*Cucumis sativus* L.)—the latest developments and perspectives. *J. Appl. Genet.* 63(4): 597–608. <https://doi.org/10.1007/s13353-022-00710-2>.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12(85): 2825–2830.
- Redmon, J., S. Divvala, R. Girshick and A. Farhadi. 2016. You Only Look Once: Unified, real-time object detection, pp. 779–788. *In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, June 27–30, 2016, Nevada, USA.
- Salama, K. 2024. Optimization of regression models using machine learning: A comprehensive study with Scikit-learn. *IUSRJ*. 5(16): 119–129. <https://doi.org/10.59271/s45500.024.0624.16>.
- Samiei, S., P. Rasti, J.L. Vu, J. Buitink and D. Rousseau. 2020. Deep learning-based detection of seedling development. *Plant Methods*. 16: 103. <https://doi.org/10.1186/s13007-020-00647-9>.
- Shen, Y., Z. Yang, Z. Khan, H. Liu, W. Chen and S. Duan. 2025. Optimization of improved YOLOv8 for precision tomato leaf disease detection in sustainable agriculture. *Sensors*. 25(5): 1398. <https://doi.org/10.3390/s25051398>.
- Staub, J.E., M.D. Robbins and T.C. Wehner. 2008. Cucumber, pp. 241–282. *In: J. Prohens and F. Nues, (Eds), Handbook of Plant Breeding*. Springer, New York, USA.

- Syed, T.N., L. Jizhan, Z. Xin, Z. Shengyi, Y. Yan, S.H.A. Mohamed and I.A. Lakhia. 2019. Seedling-lump integrated non-destructive monitoring for automatic transplanting with Intel RealSense depth camera. *Artif. Intell. Agric.* 3: 18–32. <https://doi.org/10.1016/j.aiia.2019.09.001>.
- Zendehdel, N., H. Chen and M.C. Leu. 2023. Real-time tool detection in smart manufacturing using You-Only-Look-Once (YOLO)v5. *Manuf. Lett.* 35: 1052–1059. <https://doi.org/10.1016/j.mfglet.2023.08.062>.