Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 9, Number 3, 2025
DOI: 10.30829/zero.v9i3.27199
Page: 1109 - 1122

ResNet-18 Citrus Classification with Augmentation: Peel Color Associations with pH and Vitamin C

[1] **Titin Afrina Priani Dewi**
Department of Mathematics, Ahmad Dahlan University, Yogyakarta 55191, Indonesia

[2] **Sugiyarto Surono**
Department of Mathematics, Ahmad Dahlan University, Yogyakarta 55191, Indonesia

[3] **Aris Thobirin**
Department of Mathematics, Ahmad Dahlan University, Yogyakarta 55191, Indonesia

Article Info

Article history:

Accepted 26 December 2025

Keywords:

CIELab;
Data Augmentation;
Deep Learning;
ResNet-18;
RGB.

ABSTRACT

Limes and lemons are widely consumed, but field quality assessment is often subjective. This study built a ResNet-18 citrus classifier using transfer learning and data augmentation, and evaluated it with stratified 5-fold cross-validation on 80 images (40 limes, 40 lemons). All analyses were conducted per fold to reduce optimistic bias. The model reached 98.75% mean validation accuracy, misclassifying one lime image while correctly recognizing all lemons. For interpretation, peel regions were quantified using NDYI and CIELab (L*, a*, b*), and related to pH and vitamin C using Spearman correlation. Uncertainty was quantified with bootstrap-based 95% confidence intervals for each correlation. Peel color features were more consistently associated with pH (especially in limes), whereas correlations with vitamin C were weak and non-significant for both fruits. Results indicate strong performance under controlled imaging, but using peel color as a vitamin C proxy requires broader data and external validation across cameras and lighting.

Corresponding Author:

Titin Afrina Priani Dewi,
Department of Mathematics
Ahmad Dahlan University, Yogyakarta, Indonesia
Email: 2200015026@webmail.uad.ac.id

1. INTRODUCTION

Increased focus on artificial intelligence has increased the introduction and use of deep learning in various fields, including agriculture and food [1][2]. In precision agriculture, image-based vision supports monitoring and decision-making in cultivation processes, making management more efficient [3]. In the post-harvest stage, CNN is widely used for quality inspection and fruit grading based on external characteristics such as color, shape, size, ripeness level, and surface defects, making assessment results more consistent than manual inspections [4][5].

Citrus fruits such as lime (Citrus aurantifolia) and lemon (Citrus limon) are important commodities that are widely used as beverage ingredients, culinary spices, as well as a source of vitamin C, phenolic compounds, and other bioactive components that act as antioxidants, both in the flesh and peel of the fruit [6]. However, the practice of quality assessment at the level of farmers, traders, and consumers is still dominated by subjective visual

observations. This condition can cause inconsistencies in assessments, so an image-based quantitative approach is needed to assess the quality of lime and lemon more objectively [7].

The quality and consumer acceptance of limes and lemons are strongly influenced by peel appearance, especially the green or yellow color which reflects the level of ripeness and quality. Supply chains in agriculture increasingly require quality control that runs automatically and quickly, without tampering with samples, and remains stable despite limited labeled data [8]. In this case, the relationship between peel color and chemical properties can be studied by extracting color features from digital imagery and analyzing patterns of their relationship with chemical parameters, such as acidity level (pH) and vitamin C levels [9]. This approach is important because it provides a more interpretive path of analysis, i.e. visual features are not only used for classification, but are also used to see the tendency of visual associations with chemical content in a measurable way.

In terms of modeling and statistical validation, the main challenge in image-based classification for fruit quality is the limitation of labeled data, especially if each sample must be accompanied by chemical measurements such as pH and vitamin C, which of course requires time and cost. Small data sizes can increase the risk of overfitting and make performance estimates more uncertain [10][11]. Therefore, evaluation designs such as stratified k-fold cross-validation are required to maintain the proportion of classes on each fold, and results are reported as inter-fold summaries and confidence intervals to reflect the uncertainty of evaluation metrics in small data [12][13].

Several previous studies have applied deep learning to classification, detection of fruit damage or disease, or determination of ripeness in citrus fruits based on imagery [14][15]. However, a number of studies have shown that the application of computer vision to agriculture and quality assessment, small datasets or limited data variations, plus less comprehensive evaluation procedures, can result in performance metrics that look very high but do not necessarily reflect generalization capabilities under real conditions. Therefore, uncertainty reporting such as through confidence intervals on evaluation metrics needs to be included so that interpretation is more careful and transparent [16][17]. For example, the classification study of lime and jungga lime resulted in high validation, but the confusion matrix showed low overall accuracy of 52.9%, indicating significant misclassification and the need for generalization strategies [18].

On the other hand, lightweight architectures such as ResNet-18 are reported to be effective on a wide range of fruit classification tasks. As in the detection of the ripeness stage of mulberries, the ResNet-18 model was able to achieve an accuracy of 98.65% [19]. However, success in non-citrus commodities does not automatically guarantee similar performance in limes and lemons. In addition, the application of ResNet-18 for the recognition of citrus, such as lime and lemon is still rare, especially in the context of correlation analysis between visual characteristics and the chemical content of fruits, such as vitamin C and acidity levels [20]. This shows that there is a research gap that can be further explored using a deep learning-based approach. These limitations indicate that there is a gap in applied research to develop a more objective image-based quality research framework on the two types of images.

In addition to being lightweight, ResNet-18 is also based on the concept of residual learning through skip connections, which is a shortcut that connects inputs and outputs in a block. With this mechanism, the network focuses more on learning minor corrections to features from previous stages, rather than building feature transformations from scratch. Skip connections help keep gradient flow smooth, making optimization more stable and reducing the risk of issues such as vanishing gradients during training [21]. This becomes especially relevant when transfer learning is applied to small datasets, particularry when combined with robust regularization and evaluation [22].

As an applied study to strengthen image-based citrus quality assessment under limited data conditions, this study applies a data augmentation strategy to improve the generalization ability of the model on a relatively small dataset. The data augmentation approach has been reported to be effective in improving image classification performance by enriching the variety of training data. The main contribution of this study lies in the application of transfer learning in the ResNet-18 architecture combined with data augmentation applied to the training data, as well as evaluating the stratified k-fold scheme for the classification of lime and lemon. In addition to producing a classification model, this study also analyzed the relationship between fruit skin color features to pH and vitamin C using Spearman correlation, as this approach is suitable for small data and relationships that are not always linear.

Thus, this research is directed by the following research questions:

RQ1: How stable is the performance of ResNetNet-18 based lime and lemon classification on limited data when augmented and evaluated with stratified k-fold cross-validation?

RQ2: To what extent does data augmentation help improve the generalization capabilities of the model?

RQ3: Do fruit skin color features show a meaningful relationship with pH and vitamin C based on Spearman correlations?

2. RESEARCH METHOD

This study adopts a deep learning-based computational approach, which focuses on the classification of citrus images using the ResNet-18 architecture and describes the pattern of the relationship between fruit peel color and pH and vitamin C levels. The entire image processing process, ResNet-18 model training, and model performance evaluation were performed on CPU-based Google Colab using Python with the PyTorch library.

The series of research stages is shown in Figure 1. The workflow starts with dataset collection, followed by image pre-processing. Next, the data are partitioned using stratified k-fold cross-validation into training and validation sets. Data augmentation is applied only to the train set, and the augmented data are used to trin the ResNet-18 model. The trained model is then assessed through model evaluation on the validation set. Finally, the study continues with peel color feature and citrus chemistry analysis, and the relationship between color features, pH, and vitamin C is examined using Spearman correlation with a 95% confidence interval (CI).

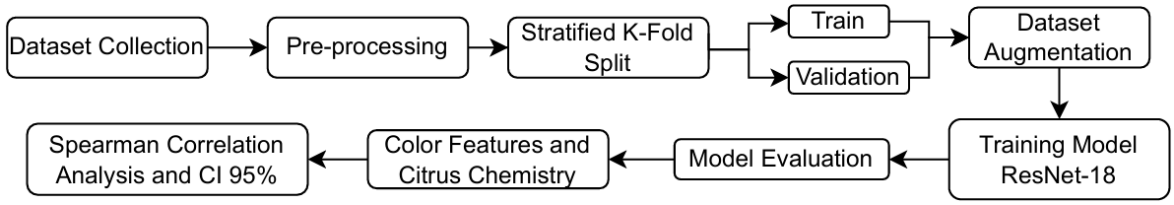


Figure 1. Research Flow

2.1 Dataset Collection

The dataset used in this study consisted of two classes of citrus fruits, namely lime and lemon, with a total of 80 images, with 40 images per class. Fruit samples were purchased at Mayong Market and roadside shophouses in Jepara and were randomly selected from the fruits available at the time of purchase with inclusion criteria including ro not, no excessive abrasion, and varying fruit sizes. Lime cultivar information was not recorded, whereas the lemon used was a California lemon with grade A. Therefore, the generalization coverage mainly represented the lime fruit in circulation at the location and time of data collection.

Each sample is represented by one digital image and two chemical parameters, namely pH and vitamin C levels. The image collection process was carried out in one data collection session using the iPhone 15 camera in Portrait mode, with natural lighting and exposure compensation setting of −1 to maintain relatively uniform lighting conditions. To minimize acquisition variations that may affect color values and classification performance, shooting was done with a consistent black background and the fruit position and camera angles are made uniform.

After the shooting process, fruit water extract was taken from each fruit for pH measurement and vitamin C analysis. pH values were measured using a pH meter that was calibrated before measurement with a standard buffer solution through a two-three point calibration, namely a pH buffer of 7.00 as the starting point, followed by a pH buffer of 4.00 or pH 10.00 until the reading was stable and in accordance with the buffer value. Each sample is measured 2-3 times and the recorded values were stable results, i.e. when repeated measurements produce the same values. Vitamin C levels were calculated using the iodometric titration method with duplicate measurements, then the titrant volume was averaged for the calculation of vitamin C levels. Therefore, the interpretation of the relationship was carried out with caution and accompanied by a summary of the basic variations in pH and vitamin C in each class shown in Table 1.

Table 1. Descriptive Statistics of pH and Vitamin C per Class

Class	Chemical Parameters	Mean	SD	Min	Max
Lime	pH	2.067500	0.840143	1.200000	4.400000
	Vitamin C (%)	0.884326	0.011553	0.868442	0.925000
Lemon	pH	1.782500	0.448294	1.400000	3.700000
	Vitamin C (%)	1.320678	0.007539	1.307822	1.3369943

2.2 Pre-processing

The image is processed deterministically through resizing to 256×256 pixels, then center-cropping to 224×224 pixels. The image is then normalized using ImageNet normalization parameters and converted into a PyTorch tensor using ToTensorV2, so that all data has a consistent input scale and formatbefore entering the model training and evaluation stage.

2.3 Stratified K-Fold Split

The performance evaluation of ResNet-18 was carried out by utilizing a stratified k-fold scheme with k=5, so that the number of lime and lemon images remained balanced in each fold. All datasets containing N images

are divided into k parts of almost equal size, and at each fold the proportions of the two classes are maintained to be close to the initial distribution. In general, the amount of training data and validation data in a round of k-fold can be expressed as follows [23].

$$N_{train} = N - \frac{N}{k} \quad (1)$$

$$N_{val} = \frac{N}{k} \quad (2)$$

From the total data used N = 80 images, each fold obtained 64 training data and 16 validation data, with a balanced proportion between the lime and lemon classes in each fold.

The selection of k=5 was used to maintain a compromise of bias and variance in the condition of a small dataset. If k is too large, then the validation size per fold becomes very small, so the evaluation metric can easily fluctuate because a few prediction errors alone can change the accuracy value significantly. This approach prevents class imbalance and allows each image to play a role alternately as training data and validation data [24].

2.4 Dataset Augmentation

Data augmentation is used as a structured process to reproduce and enrich training images by forming new variations of the same image through a series of transformations, without changing their meaning or label [25][26]. In this study, a data augmentation scheme was used, namely the application of random crop transformation, horizontal flip, rotation, color jitter, and Gaussian noise that were randomly selected on the image only on the training data. Data augmentation is only applied to the training data in each fold that is selected for classification using the Albumentations library in Python. Data augmentation is not applied to images extracted using skin tone features to keep the scale representative of the original image condition.

A summary of the transformation types and key parameters is presented in Table 2. This table contains the crop size, the angle range of rotation, and the probability of applying each transform used to the training data.

Table 2. Data Augmentation Parameters

Transformation	Parameters
Random crop	224 × 224 pixels
Horizontal flip	0.5
Rotation	±20° (ρ = 0.5)
Color jitter	Brightness = 0.3, contrast = 0.3, saturation = 0.2, hue = 0.05
Gaussian noise	Var_limit = (10, 50), ρ = 0.3

2.5 ResNet-18 Architecture

In this study, the citrus image classification model was built using the ResNet-18 architecture. The model is initialized with the ImageNet pretrained weights, which are IMAGENET1K_V1. ResNet-18 applies the concept of residual learning with skip connections, so that the network does not directly learn the mapping of $H(x)$, but learns the residual function of $F(x)$ against input x which is defined as follows [27].

$$F(x) = H(x) - x \quad (3)$$

The output of the residual block is obtained by adding the residual function to the initial input, i.e.

$$y = F(x) + x \quad (4)$$

The ResNet-18 backbone structure is composed of an initial layer of convolution and max pooling, followed by four stages of residual blocks, then global average pooling to produce a 512-dimensional feature vector. In the implementation of this study, the fully connected layer of ResNet-18 was changed to Identity so that the backbone functions as a feature extractor. The classification was carried out using a classifier head measuring 512→256→2 equipped with a $p = 0.5$ dropout to suppress overfitting in small datasets.

The training process was optimized using AdamW with weight decay $= 1 \times 10^{-4}$. and batch size 16. The loss function used was cross-entropy loss with label smoothing of 0.15. In addition, to increase regularization under limited data, training is conducted up to a maximum of 15 epochs on each fold, with the application of early stopping based on validation performance to stop training when no meaningful improvements occur. Training stability is also maintained using gradient clipping with a limit of max_norm=1.0. The best model on each fold is stored based on the minimum validation loss. The final output of the model is in the form of logits for two classes, while the probabilistic activation function is not explicitly defined because it has been accommodated in the cross-entropy scheme during training.

Figure 2 shows the standard structure of ResNet-18 as an illustration. The main adjustment in the implementation of this study is the replacement of the fully connected layer with Identity and the use of classifier heads with dropout to produce two classes.

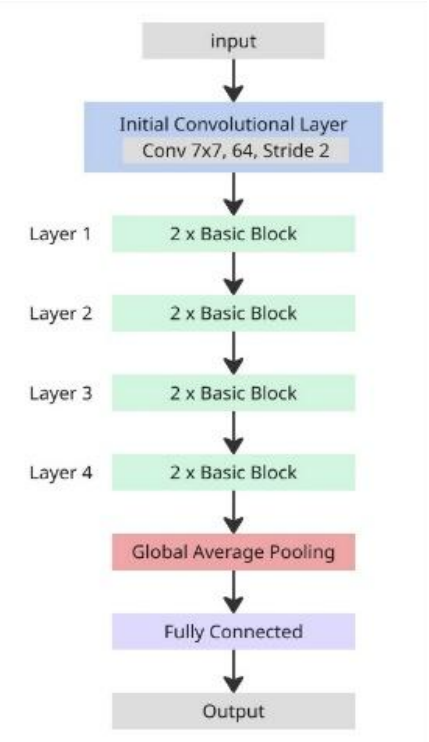


Figure 2. Structure of ResNet-18

2.6 Model Evaluation

The performance of the model was evaluated in several metrics derived from the confusion matrix, namely accuracy, precision, recall, and F1-score [28][29]. The confusion matrix presents four main components:

1. True Positive (TP) : The actual number of images includes a positive class and is predicted as a positive class by the model.
2. True Negative (TN): The actual number of images includes a negative class and is predicted as a negative class by the model.
3. False Positive (FP) : The number of images is actually a negative class but is predicted to be a positive class.
4. False Negative (FN) : The number of images that are actually a positive class but predicted to be a negative class.

Accuracy is calculated as the percentage of the number of correctly classified images against all images in the fold, namely:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad (5)$$

Precision is used to measure how much a model's positive predictions are correct, and is defined as follows.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

Recall is also called sensitivity or true positive rate which functions to measure how many positive samples are successfully recognized by the model, defined as follows.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

To balance the trade-off between precision and recall, F1-score is used, which is the harmonic average of the two.

$$\text{F1} - \text{Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

2.7 The Relationship of Fruit Peel Color Features to Fruit Chemical Content

Analysis of the pattern of relationship between fruit color features and citrus chemical content was carried out by combining information from digital images and laboratory measurement results. For each sample, the image is processed so that the color features are calculated only from the area of the fruit peel. The skin area in the image is cropped first and segmented to separate the skin covering from the background. Then the pixel values on the fruit skin area are extracted to obtain average color features.

a. RGB

In the RGB color space, each pixel in the fruit peel area is represented by three intensity channels, namely red (R), green (G), and blue (B). To illustrate the tendency of yellow hue in citrus peel, the Normalized Difference Yellowness Index (NDYI) is used, which is a yellowness index based on the normalized difference between channels G and B, which is expressed in the equation (9) [30].

$$I_{NDYI} = \frac{G - B}{G + B} \tag{9}$$

The greater value of I_{NDYI} indicates the dominance of the green-yellow component compared to blue, thus indicating a yellowish-looking fruit peel.

b. CIELab

The fruit peel image was also converted to the CIELab color space to obtain a more perceptually uniform color description. The conversion process is carried out in two stages, namely the pixel value in the RGB image (sRGB) is first transformed to CIEXYZ using a standard transformation with D65 illuminant, then the CIEXYZ value is converted to CIELab.

The pixel value of the image is converted to the CIELab coordinates using the standard transformation from CIEXYZ to CIELab. For each pixel, three components are obtained, namely L*, a*, and b*, which represent brightness, green-red axis, and blue-yellow axis, respectively [31]. The value of the feature used is the average value in the area of the fruit peel, i.e.

$$L^* = 116f\left(\frac{Y}{Y_n}\right) - 16 \tag{10}$$

$$a^* = 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \tag{11}$$

$$b^* = 200\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \tag{12}$$

with X, Y, Z expressing the CIEXYZ tristimulus value of the pixel and Xn, Yn, and Zn is the reference white dot tristimulus value (illuminant D65). The nonlinear function f(t) follows the CIELab standard definition, i.e.

$$f(t) = \begin{cases} t^{1/3}, & t > 0.008856 \\ 7{,}787t + \dfrac{16}{116}, & t \le 0.008856 \end{cases} \tag{13}$$

Next, the values of L*, a*, and b* are averaged on the pixels of the fruit peel area to obtain the CIELab color features.

2.8 Spearman Correlation Analysis and 95% Confidence Interval

To assess the relationship between chemical parameters, i.e. pH with vitamin C and skin tone features, such as NDYI and CIELab, Spearman correlation is used because this model is rank-based so it is safer to use when the relationship does not have to be linear and the distribution of data does not have to be normal.

Suppose there are data pairs (x_i, y_i) for $i = 1, 2, \ldots, n$. value ,x-i. and ,y-i. first changed to $R(x_i)$ dan $R(y_i)$ ratings. Spearman coefficient, ρ-s. calculated as a Pearson correlation on the ranking data [32]:

$$\rho_s = corr(R_x, R_y) = \frac{\sum_{i=1}^{n}(R_{x,i} - \bar{R}_x)(R_{y,i} - \bar{R}_y)}{\sqrt{\sum_{i=1}^{n}(R_{x,i} - \bar{R}_x)^2}\sqrt{\sum_{i=1}^{n}(R_{y,i} - \bar{R}_y)^2}} \tag{14}$$

Furthermore, in addition to reporting the value of ρ_s and p-value, the study also reported a 95% CI for ρ_s using a bootstrap approach because the sample size is relatively small and CI analytics for rank-based correlation is not always simple.

The step is to resample in pairs (x_i, y_i) B times. On each bootstrap sample b, it is calculated $\rho_s^{(b)}$. The CI of 95% is then taken from the 2.5% quartile and 97.5% from the distribution $\{\rho_s^{(1)}, \ldots, \rho_s^{(B)}\}$ [33]:

$$CI_{95\%}(\rho_s) = [Q_{0.025}(\rho_s^*), Q_{0.975}(\rho_s^*)] \tag{15}$$

3. RESULT AND ANALYSIS

In this study, the results discussed were the performance of ResNet-18 in classifying citrus images as well as the pattern of the relationship between the color of the fruit skin and the pH and levels of vitamin C. The model performance was summarized through the accuracy of each fold, the combined confusion matrix, as well as the loss and accuracy graphs during training and validation. Visualization of the relationship between RGB and CIELab-based color features is presented to illustrate the pattern trends in the dataset.

Given the small size of the dataset and relatively controlled image acquisition conditions, very high metrics need to be interpreted carefully as they have the potential to contain optimistic bias. Therefore, in addition to the average value, stability is evaluated through inter-fold variations in the 5-fold stratified scheme, as well as the combined validation predictions across all folds.

3.1. Model Evaluation

This subsection presents the results of the evaluation applied to ResNet-18 which includes the accuracy of each fold, the combined confusion matrix, as well as the loss and accuracy graphs during training and validation. The 5-fold division is made at the image level, based on img_path and labels. Because each fruit is represented with one image, the split level of the image and the level of the fruit are the same in this study.

3.1.1 Accuracy of each Fold, Classification Report, and Confusion Matrix

A summary of the validation accuracy for each fold is shown in the Table 3.

Table 3. Model Accuracy of each fold

Fold	Accuracy
1	0.9375
2	1.0000
3	1.0000
4	1.0000
5	1.0000

Based on Table 3, an average accuracy of 0.9875 or 98.75% was obtained. The value between the folds is quite stable as most folds reach 1.0000, but there is still one fold lower, which is 0.9375. This indicates that there are still samples that are more difficult to process so that the potential for prediction errors remains, in small datasets the results can still change depending on the data division.

In addition to the average per-fold accuracy, the quality of the model predictions in each class was also evaluated in the classification report. A summary of precision, recall, and f1-score values for the lime and lemon classes is presented in Table 4.

Table 4. Classification Report

Class	Precision	Recall	F1-Score
Lime	1.00	0.97	0.99
Lemon	0.98	1.00	0.99

Table 4 shows that the ResNet-18 model performed very well in both citrus classes, with precision and recall close to 1.00, and an F1-Score of 0.99 for lime and lemon. This value indicates that the prediction error is relatively small and that the model is able to distinguish between lime and lemon consistently. To see the number of true and false predictions clearly, the validation results of all the folds are combined into the combined confusion matrix shown in Figure 3. Predictions are taken from the best models on each fold.

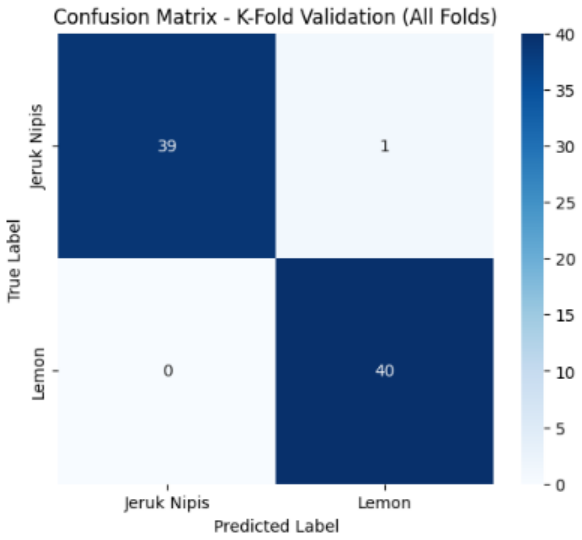


Figure 3. Confusion Matrix

Aa shown in Figure 3, out of 40 images of limes, 39 images were correctly classified as lime and only 1 image was incorrectly identified as lemon. Meanwhile, 40 lemon images were successfully classified as lemons. These results show that the model provides very consistent predictions for the lemon class, whereas in the lime class there is still a slight overlap of visual features that can trigger misclassification in a small portion of the data.

To assess the stability of the evaluation results, the metrics were summarized with a 95% confidence interval (CI) based on inter-fold variation. This 95% CI is calculated from 5 fold values using the t-interval method on the average of the metric between folds, so this CI is approximate and mainly accounts for variations due to fold splits. The average accuracy of 98.75% has a 95% CI of 95.28% to 100%. A weighted precision of 98.89% has a 95% confidence interval of 95.80% to 100%, a weighted recall of 98.75% has a 95% confidence interval of 95.28% to 100%, and a weighted F1-score of 98.75% has a 95% confidence interval of 95.26% to 100%. In general, this summary indicates that the results tend to be consistent across different data divisions, although the amount of data used is still limited.

3.1.2 Loss and Accuracy Charts

The development of loss and accuracy during training and validation is visible in the Figure 4.

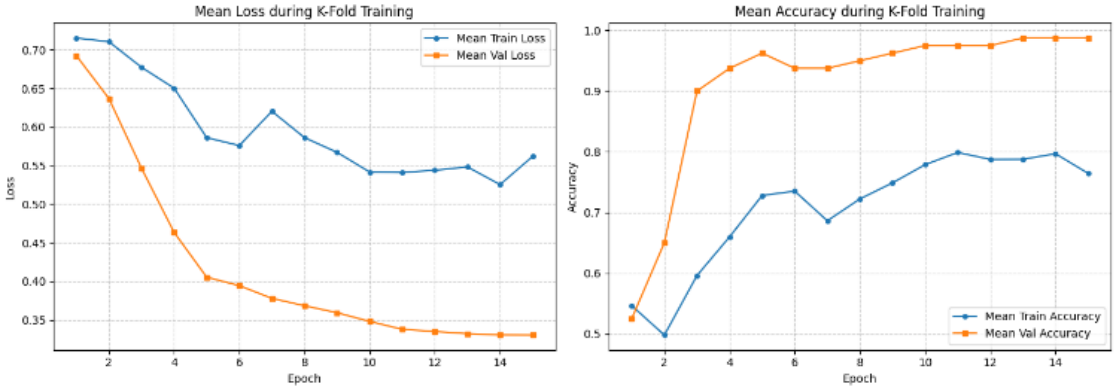


Figure 4. Loss and Accuracy Graph in Training and Validation

This shows that after a few initial iterations the model is already able to separate classes quite well. The accuracy graph shows the accuracy of the training rising gradually, while the accuracy of the validation jumps sharply to very high values and then forms an almost horizontal pattern. The combination of reduced losses and increased accuracy, especially on the validation curve, showed that on average, across all k-folds, the model managed to learn the high classification performance on the validation data. The training is controlled using early stopping based on validation loss, and the best model per fold is stored at the minimum validation loss to be used in the evaluation of the fold. This mechanism can curtail the risk of training continuing past the optimum point that often occurs in small datasets.

Although the validation accuracy looks very good, the size of the dataset in this study is relatively small compared to the capacity of the CNN model used. Therefore, there is a risk that the model partially adapts to the specific characteristics of the dataset, such as uniform shooting conditions, camera settings, background, and shooting angles. As a result, the high performance of the dataset does not necessarily fully persist when applied to new citrus images taken under different conditions.

The 5-fold stratified scheme and augmentation on the training data help to increase resilience in the dataset, but they have not been able to replace the need for external testing using independent data from different sessions, devices, and lighting.

3.2. Data Augmentation on Model Performance

This study applied data augmentation to training data in each fold in the form of random crop, horizontal flip, rotation, color jitter, and Gaussian noise. This transformation enriches the variance of the training data so that the training model deals with changes in angle, scale, and visual variation that may arise when the size of the dataset is limited. Thus, augmentation acts as a data-driven regularization strategy to help reduce overfitting and increase model resilience in small sample regimes.

Previous studies have also reported high performance on fruit classification using ResNet-18, including on different objects and scenarios. However, due to differences in fruit types, acquisition conditions, and evaluation protocols, the results between studies cannot be directly compared and are not intended to assert superiority. In this study, high internal validation results showed that augmentation can support learning on small datasets, but external tests using independent data are still needed to ensure performance stability under different imaging conditions.

3.3. Visual Relationship Pattern of Fruit Peel Color with Fruit Chemical Content

This section analyzes the relationship between pH and vitamin C levels as a percentage (%) with fruit peel color features extracted from RGB and CIELab to elucidate the relationship between visual signals and the chemical properties of fruits.

Since all images were captured under natural light, ambient illumination changes such as cloud cover and time-of-day differences may shift the recorded peel color and affect the extracted NDYI and CIELab features. Therefore, the interpretation of color–chemistry trends is limited to the current acquisition setting, and future work should include explicit illumination or color normalization and additional data under more diverse lighting conditions.

3.3.1 Visual Pattern of the Relationship of Lime Peel Color to pH and Vitamin C

The visual pattern of the relationship between the color of the lime peel and the pH and vitamin C levels can be seen in Figure 5 and Figure 6.

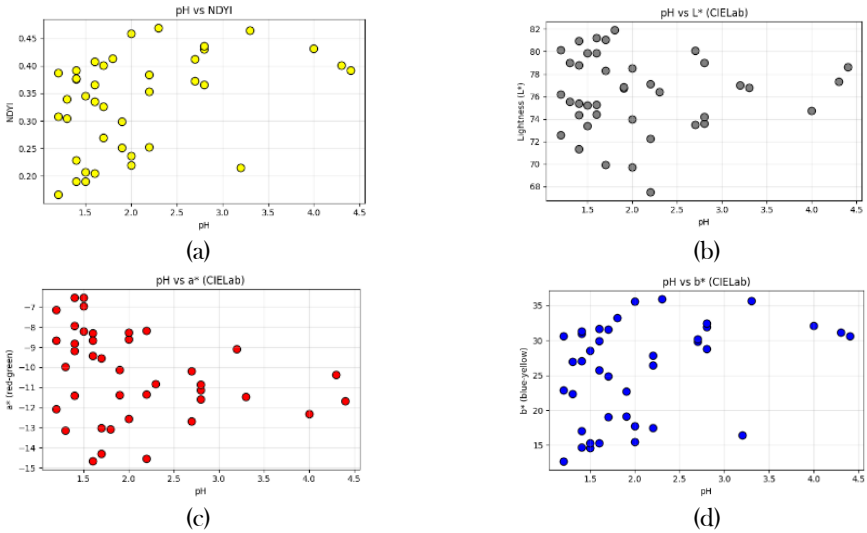


Figure 5. Visual Pattern of the Relationship Between Lime Peel Color Features and pH

In general, the scattering of points at pH vs NDYI and pH vs b* shows a tendency to increase as the pH increases, so that the higher the pH of the peel, the yellower it appears, although the distribution of data is still wide enough that the pattern formed is not completely linear. At pH vs L*, the dots appear to spread out without a consistent direction of change. Meanwhile, in pH vs a*, there is a tendency to change to more negative values in some pH ranges, but in this pattern it is also not strong and still shows large variations between samples.

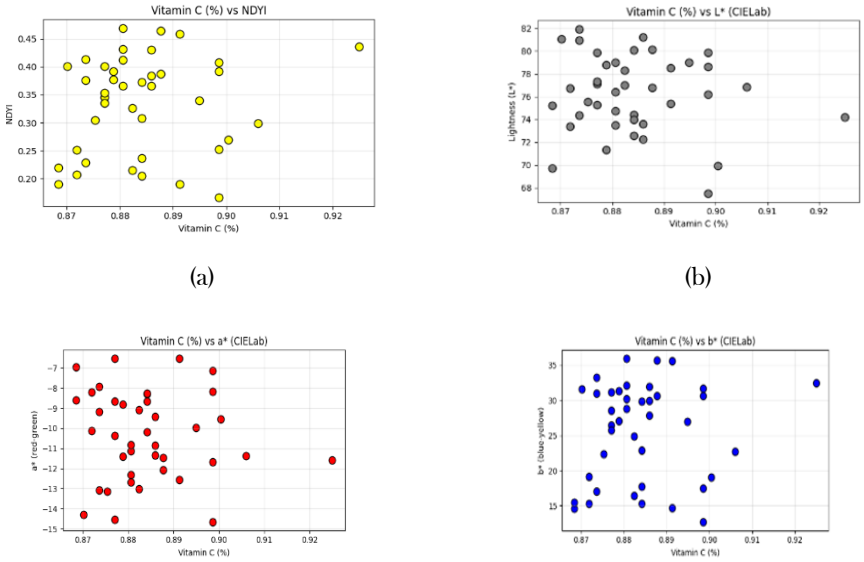
To strengthen the visual interpretation, the relationship between pH and color features was also quantitatively tested using the Spearman correlation test, and the uncertainty of the correlation coefficient was reported with a CI of 95%. The results of the correlation test are presented in Table 5.

Table 5. Spearman pH Correlation to Lime Peel Color Features

Variable Pairs	ρ Spearman	p-value	CI 95%
pH vs NDYI	0.4522	0.0034	(0.1537, 0.6607)
pH vs L*	-0.0676	0.6784	(-0.3491, 0.2252)
pH vs a*	-0.3332	0.0357	(-0.5863, -0.0321)
pH vs b*	0.4044	0.0097	(0.1105, 0.6251)

Based on Table 5, pH has a significant positive relationship with NDYI with (ρ=0.4522, p=0.0034) and b* (ρ=0.4044, p=0.0097). In magnitude, these two indicate a medium relationship, a CI of 95% which is entirely positive reinforcing that the direction of the relationship is consistent. That is, in limes, the increase in pH tends to be related to an increase in the yellowness index and the yellow component of the CIELab, although the pattern in the scatter plot is still spreading.

In contrast, the pH relationship with L* was not significant (p = 0.6784) and the CI was 95% across zero, so the brightness of the lime peel did not show a consistent affinity with pH in this dataset. For component a*, a significant negative relationship (ρ = -0.3332, p = 0.0357) with small to medium magnitude and a 95% CI was entirely negative (-0.5863 to -0.0321). This shows that in some samples, an increase in pH tends to be followed by a decrease in a* or more negative values, i.e. a shift in hue that is more towards green in some samples.



(a) (b)

(c) (d)

Figure 6. Visual Pattern of the Relationship Between Lime Peel Color Features and Vitamin C (%)

Figure 6 shows that higher vitamin C levels tend to appear in skin with relatively larger NDYI and b*, so samples with slightly higher vitamin C levels are generally in areas of a stronger yellow hue, although the pattern is still weak due to the narrow range of vitamin C levels. However, the dot distribution on the graph of vitamin C vs L* levels and vitamin C vs a* levels does not form a consistent direction of change, making it less informative to distinguish small variations in vitamin C levels.

To reinforce such visual interpretations, the relationship between vitamin C levels and color features was also quantitatively tested using the Spearman correlation test, and the uncertainty of the correlation coefficient was reported with a CI of 95%. The results of the correlation test are presented in Table 6.

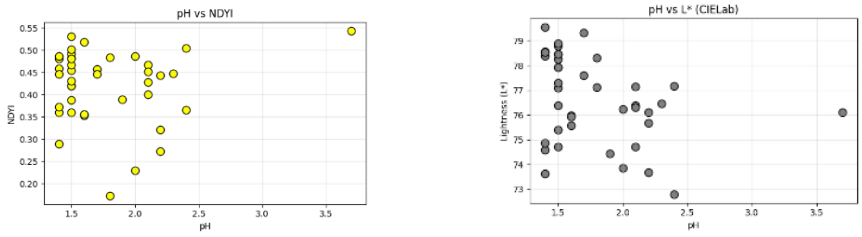
Table 6. Spearman's Correlation of Vitamin C Levels to Lime Peel Color Features

Variable Pairs	ρ Spearman	p-value	CI 95%
Vitamin C vs NDYI	0.1355	0.4044	(-0.2039, 0.4667)
Vitamin C vs L*	-0.0520	0.7499	(-0.3841, 0.2907)
Vitamin C vs a*	-0.0968	0.5525	(-0.4329, 0.2600)
Vitamin C vs b*	0.1103	0.4980	(-0.2359, 0.4494)

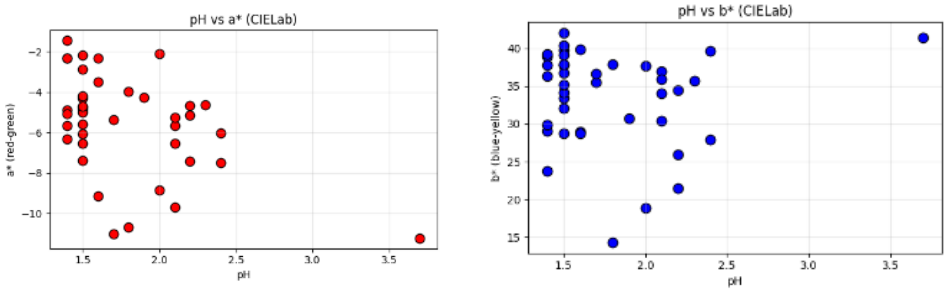
Based on Table 6, the entire correlation between vitamin C levels and color features in lime had a small coefficient with a ρ value of < 0.20 and insignificant at a p-value of > 0.05. The CI value of 95% on all variable pairs also crosses zero, so the direction of the relationship cannot be said to be consistent. However, the relationship between vitamin C and NDYI and the CIELab component is very weak, and the apparent tendency to scatter plots has not been supported by strong statistical evidence. This condition is likely affected by a narrow range of vitamin C and a wide variety of samples.

3.3.2 Visual Pattern of the Relationship of Lemon Peel Color to pH and Vitamin C Levels

A visual pattern of the relationship between lemon peel color and pH and vitamin C levels can be seen in Figure 7 and Figure 8.



(a) (b)



(c) (d)

Figure 7. Visual Pattern of the Relationship Between Lemon Peel Color Features and pH

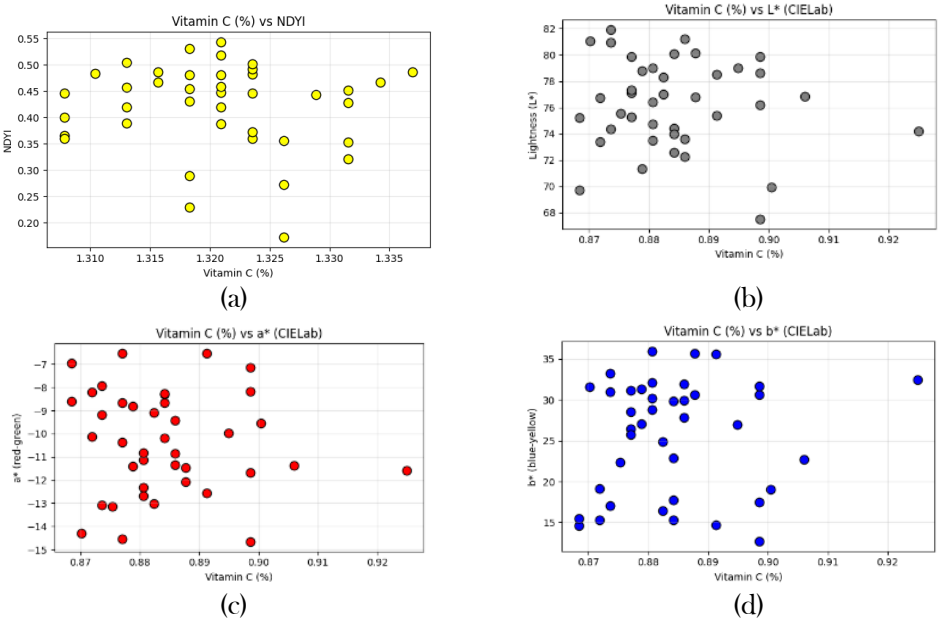
In Figure 7, the dot distribution on the pH vs NDYI and pH vs b* graphs shows that most of the points are at relatively high values in almost the entire pH range, so the color of the lemon peel tends to be predominantly yellow and no obvious change is seen when the pH value changes. At pH vs L*, it can be seen that the tendency of L* decreases slightly at higher pH so that the fruit appears somewhat darker, although the dot distribution is still quite wide. For pH vs a*, the a* value generally remained negative in some samples, indicating a predominance of green hue and only shifting slightly even when pH increased.

To reinforce such visual interpretations, the relationship between pH and color features was also quantitatively tested using the Spearman correlation test, and the uncertainty of the correlation coefficient was reported with a CI of 95%. The results of the correlation test are presented in Table 7.

Table 7. Spearman pH Correlation to Lemon Peel Color Feature

Variable Pairs	ρ Spearman	p-value	CI 95%
pH vs NDYI	-0.0544	0.7388	(-0.3831, 0.2758)
pH vs L*	-0.3494	0.0271	(-0.6518, -0.0012)
pH vs a*	-0.3689	0.0191	(-0.6082, -0.0702)
pH vs b*	-0.1489	0.3591	(-0.4716, 0.1887)

Based on Table 7, pH was not significantly related to NDYI or b* because the p-value > 0.05 and CI 95% both crossed zero, so the pH change in lemons did not show a consistent shift in the yellowish index. In contrast, pH was significantly negatively correlated with L* (ρ = -0.3494, p = 0.0271, CI 95% = -0.6518 to -0.0012) and a* (ρ = -0.3689, p = 0.0191, CI 95% = -0.6082 to -0.0702). In magnitude, these two relationships are small to moderate, indicating that an increase in pH values tends to be followed by a decrease in L* and a decrease in a* value in some samples. However, because the distribution of points on the graph is still quite wide and there are some deviating points, the interpretation of this relationship is still done carefully.



(a) (b)

(c) (d)

Figure 8. Visual Pattern of the Relationship Between Lemon Peel Color Features and Vitamin C (%)

Figure 8 shows that the graph of vitamin C vs NDYI and vitamin C vs b* is spread over a fairly wide range of values with no clear direction of change. Some samples had high NDYI and b* values, but their appearance was inconsistent with increased vitamin C, so the resulting visual patterns tended to be weak. On the graph, vitamin C vs L* and vitamin C vs a* are in the medium to high brightness range without a consistent trend, so they do not provide strong information about the variation of vitamin C in lemons.

To reinforce such visual interpretations, the relationship between vitamin C levels and color features was also quantitatively tested using the Spearman correlation test, and the uncertainty of the correlation coefficient was reported with a CI of 95%. The results of the correlation test are presented in Table 8.

Table 8. Spearman's Correlation of Vitamin C Levels to Lemon Peel Color Features

Variable Pairs	ρ Spearman	p-value	CI 95%
Vitamin C vs NDYI	-0.1008	0.5361	(-0.3973, 0.2037)
Vitamin C vs L*	0.0440	0.7877	(-0.2782, 0.3559)
Vitamin C vs a*	0.1037	0.5243	(-0.2291, 0.4289)
Vitamin C vs b*	-0.0791	0.6275	(-0.3667, 0.2221)

Based on Table 8, it shows that the entire relationship between vitamin C levels and color features in lemons has a small coefficient of $\rho < 0.20$ and is insignificant due to p-value > 0.05. The 95% CI value on all variable pairs also crosses zero, so the direction of the relationship cannot be said to be consistent. Thus, the relationship between vitamin C and NDYI and the CIELab component in lemons is very weak, and the pattern in Figure 8 does not show a statistically consistent relationship between vitamin C and color features in lemons.

4. CONCLUSION

This study aims to classify lime and lemon images using ResNetNet-18 based CNN under limited data conditions and examine the relationship pattern between fruit skin color features and chemical properties, namely pH and vitamin C levels.

The results obtained showed that ResNet-18 provides very high validation performance under uniform data capture conditions, with an average accuracy of 98.75%. These results indicate that the difference in visual patterns between lime and lemon is quite pronounced in a controlled scenario, so this deep learning approach can be used as an initial prototype for citrus classification. In the analysis of color features with pH and vitamin C, color features tended to be more consistent with pH, especially in limes, whereas the relationship of vitamin C to color features was not significant in both limes and lemons. Thus, no strong and stable relationship between vitamin C and peel color metrics was found in this dataset, so color metrics cannot be used as a reliable indicator for vitamin C screening. Despite the high classification performance, this study is still limited by the small data size and relatively homogeneous one-sided acquisition, so there is a possibility of optimistic bias and generalization limitations when applied to the conditions of the lighting devices, and different backgrounds. Further research is suggested to be strengthened by the addition of data from several sessions and lighting variations, the capture of more than one angle per piece, and the use of external test data. In addition, from the applied mathematics/statistical side, the research can be developed by adding predictive uncertainty analysis to find out how confident the model is, utilizing many additional unlabeled imagery through semi-unsupervised learning, and applying methods to keep the model stable when used on photos from different cameras or lighting, for example with color adjustment or domain adaptation/generalization.

5. REFERENCES

[1] P. Changtor, W. Ratiphaphongthon, M. Saengthong, K. Buddhachat, and N. Yimtragool, "Hybrid Cassava Identification from Morphometric Analysis to Deep Convolutional Neural Networks and Confirmation Strategies," *Trends Sci.*, vol. 22, no. 5, pp. 1–20, 2025. doi: https://doi.org/10.48048/tis.2025.9475

[2] D. Mirwansyah and A. Wibowo, "Fruit Image Classification Using Deep Learning Algorithm: Systematic Literature Review (SLR)," *MULTICA Sci. Technol.*, vol. 2, no. 2, pp. 38–41, 2022. doi: https://doi.org/10.47002/mst.v2i2.356

[3] S. Ghazal, A. Munir, and W. S. Qureshi, "Computer Vision in Smart Agriculture and Precision Farming: Techniques and Applications," *Artif. Intell. Agric.*, vol. 13, pp. 64–83, 2024. doi: https://doi.org/10.1016/j.aiia.2024.06.004

[4] L. E. Chuquimarca, B. X. Vintimilla, and S. A. Velastin, "A Review of External Quality Inspection for Fruit Grading Using CNN Models," *Artif. Intell. Agric.*, vol. 14, pp. 1–20, 2024. doi: https://doi.org/10.1016/j.aiia.2024.10.002

[5] S. Magnussen, "Calibration of a Confidence Interval for a Classification Accuracy," *Sci. Res. Publ.*, vol. 11, pp. 14–36, 2021. doi: https://doi.org/10.4236/ojf.2021.111002

[6] H. Guo *et al.*, "Quality Evaluation of Citrus Varieties Based on Phytochemical Profiles and Nutritional Properties," *Front. Nutr.*, vol. 10, pp. 1–11, 2023. doi: https://doi.org/10.3389/fnut.2023.1165841

[7] I. R. Santelices, S. Cano, F. Moreira, and Á. P. Frintz, "Artificial Vision Systems for Fruit Inspection and Classification: Systematic Literature Review," *sensors*, vol. 25, no. 5, pp. 1–27, 2025. doi: https://doi.org/10.3390/s25051524

[8] K. Yu *et al.*, "Advances in Computer Vision and Spectroscopy Techniques for Non-Destructive Quality Assessment of Citrus Fruits: A Comprehensive Review," *Foods*, vol. 14. China, pp. 1-35, 2025. doi: https://doi.org/10.3390/foods14030386

[9] S. Ghanghas, V. K. Singh, S. Kumar, N. Kumar, and M. K. Gang, "Prediction of Fruit Quality Parameters Using Peel Color in Citrus Reticulata L . Fruit by Multiple Linear Regression and Artificial Neural Network Approach," *Res. Sq.*, pp. 1–13, 2022. doi: https://doi.org/10.21203/rs.3.rs-2332668/v1

[10] M. İ. Çatal, "Predicting Macroelement Content in Legumes with Machine Learning," *Sci. Rep.*, vol. 4, p. 34656, 2025. doi: https://doi.org/10.1038/s41598-025-22371-x

[11] A. Dolatabadian, T. X. Neik, M. F. Danilevicz, S. R. Upadhyaya, J. Batley, and D. Edwards, "Image-Based Crop Disease Detection Using Machine Learning," *plant Pathol.*, vol. 74, no. September 2024, pp. 18–38, 2025. doi: https://doi.org/10.1111/ppa.14006

[12] S. Prusty, S. Patnaik, and S. K. Dash, "Stratified K-Fold Cross-Validation on ML Classifiers for Predicting Cervical Cancer," *Front. Nanotechnol.*, vol. 15, pp. 1–12, 2022. doi: https://doi.org/10.3389/fnano.2022.972421v

[13] O. Rainio, J. Teuho, and R. Klén, "Evaluation Metrics and Statistical Tests for Machine Learning," *Sci. Rep.*, vol. 14, pp. 1–14, 2024. doi: https://doi.org/10.1038/s41598-024-56706-x

[14] Z. Jrondi, A. Moussaid, and M. Y. Hadi, "Interpretable Citrus Fruit Quality Assessment Using Vision Transformers and Lightweight Large Language Models," *AgriEngineering*, vol. 7, pp. 1–18, 2025. https://doi.org/10.3390/agriengineering7090286

[15] Q. Wang *et al.*, "In Situ Nondestructive Identification of Citrus Fruit Ripeness Via Hyperspectral Imaging Technology," *Plant Methods*, vol. 21, pp. 1–13, 2025. doi: https://doi.org/10.1186/s13007-025-01354-z

[16] R. Singh, C. Nickhil, K. Upendar, P. Mishra, and S. C. Deka, "Deep Learning-Based Approach for Classifying Mandarin Orange Maturity," *Sustain. Food Technol.*, vol. 3, pp. 2215–2225, 2025. doi: https://doi.org/10.1039/d5fb00408j

[17] A. Goyal and K. Lakhwani, "Integrating Advanced Deep Learning Techniques for Enhanced Detection and Classification of Citrus Leaf and Fruit Diseases," *Sci. Rep.*, vol. 15, p. 2659, 2025. doi: https://doi.org/10.1038/s41598-025-97159-0

[18] A. Hafiz, A. A. Tumanggor, M. D. Ibo, M. Y. Noor, K. P. Sitepu, and H. Syahputra, "Deteksi Jeruk Nipis dan Jeruk Jungga Menggunakan Algoritma CNN," *J. Mhs. Tek. Inform.*, vol. 9, no. 4, pp. 6245–6251, 2025. doi: https://doi.org/10.36040/jati.v9i4.14040

[19] S. H. Miraei Ashtiani, S. Javanmardi, M. Jahanbanifard, A. Martynenko, and F. J. Verbeek, "Detection of Mulberry Ripeness Stages Using Deep Learning Models," *IEEE Access*, vol. 9, no. 12, pp. 100380–100394, 2021. doi: https://doi.org/10.1109/ACCESS.2021.3096550

[20] O. J. Adelaja *et al.*, "Insecticidal Activities and Comparative Efficacy of Essential Oils of Three Citrus Species (Sapindales: Rutaceae) Against Anopheles Gambiae Mosquitoes," *Sri Lankan J. Biol.*, vol. 10, no. 1, pp. 26–35, 2025. doi: https://doi.org/10.1038/10.4038/sljb.v10i1.166

[21] W. Wu, L. Huo, G. Yang, X. Liu, and H. Li, "Research into the Application of ResNet in Soil: A Review," *agricultur*, vol. 15, pp. 1–29, 2025. doi: https://doi.org/10.3390/agriculture15060661

[22] J. Plested, M. Phiri, and T. Gedeon, "Deep Transfer Learning for Image Classification: A Survey," *ArXiv*, pp. 1–45, 2025. doi: https://doi.org/10.48550/arXiv.2205.09904

[23] M. Rafało, "Cross Validation Methods: Analysis Based on Diagnostics of Thyroid Cancer Metastasis," *Korean Ist. Commun. Inf. Sci.*, vol. 8, no. 2, pp. 183–188, 2022. doi: https://doi.org/10.11113/mjfas.v19n1.2660

[24] S. F. Mokhtar, Z. M. Yusof, and H. Sapiri, "Confidence Intervals by Bootstrapping Approach: A Significance Review," *Malaysian J. Fundam. Appl. Sci.*, vol. 19, pp. 30–42, 2023. doi: https://doi.org/10.11113/mjfas.v19n1.2660

[25] A. Mumuni and F. Mumuni, "Data Augmentation: A Comprehensive Survey of Modern Approaches," *Array*, vol. 16, no. August, pp. 1–27, 2022. doi: https://doi.org/10.1016/j.array.2022.100258

[26] Z. Wang *et al.*, "A Comprehensive Survey on Data Augmentation," *IEEE TKDE*, vol. 1, no. 1, pp. 1–36, 2024. doi: https://doi.org/10.48550/arXiv.2405.09591

[27] Y. Ni, S. Li, and P. Guo, "Discrete Wavelet Integrated Convolutional Residual Network for Bearing Fault Diagnosis under Noise and Variable Operating Conditions," *Sci. Rep.*, vol. 15, pp. 1–26, 2025. doi: https://doi.org/10.1038/s41598-025-99346-5

[28] Y. R. Fauzan and S. Uyun, "Land Cover Classification in Mountainous Regions Using Multi-Scale Fusion and Convolutional Neural Networks : A Case Study on Mount Slamet," *JOIN (Jurnal Online Inform.*, vol. 10, no. 2, pp. 286–298, 2025. doi: https://doi.org/10.15575/join.v10i2.1612

[29] A. Khairan, A. Hendri, N. Aesha Durratul, and H. Abd Mujahid, "An Applied Computer Mathematics Approach to Transliteration : YOLOv8-Based Detection of Harah Jawoe Script," *Zero J. Sains, Mat. dan Terap.*, vol. 9, no. 1, pp. 66–76, 2025. doi: https://doi.org/10.30829/zero.v9i1.24064

[30] H. Zhou, J. Yang, W. Lou, and L. Sheng, "Improving Grain Yield Prediction Through Fusion of Multi-Temporal Spectral Features and Agronomic Trait Parameters Derived from UAV Imagery," *Front. Plant Sci.*, vol. 14, no. October, pp. 1–17, 2023. doi: https://doi.org/10.3389/fpls.2023.1217448

[31] T. Azetsu and N. Suetake, "Chroma Enhancement in CIELAB Color Space Using a Lookup Table," *design*, vol. 5, no. 32, pp. 1–14, 2021. doi: https://doi.org/10.3390/designs5020032

[32] L. Zhang and L. Whang, "Optimization of Site Investigation Program for Reliability Assessment of Undrained Slope Using Spearman Rank Correlation Coefficient," *Comput. Geotech.*, vol. 155, p. 105208, 2023. doi: https://doi.org/10.1016/j.compgeo.2022.105208

[33] G. A. Rousselet, C. R. Pernet, and R. R. Wilcox, "An introduction to the Bootstrap: A Versatile Method to Make Inferences by Using Data-Driven Simulations.," *Meta-Psychology*, vol. 7, pp. 1–24, 2023. doi: https://doi.org/10.17605/OSF.IO/JGDW8