Interpretable Color–Texture–Shape Feature Fusion for RGB-Based Citrus Canker and Melanose Classification on Orange Fruit

Hoa-Cuc. Nguyen
Thu Dau Mot University
Ngoc-Duong. Nguyen
Thu Dau Mot University
Viet-Hung. Pham
Thu Dau Mot University
Bich-Ngoc. Mach
Thu Dau Mot University
Trinh Thi Nhu Quynh
Thu Dau Mot University
Duong Thi Kim Chi
chidtk@tdmu.edu.vn
Thu Dau Mot University

Research Article

Keywords: Citrus canker, Melanose, Orange fruit disease, RGB image classification, HSV color histogram, Local Binary Pattern, Histogram of Oriented Gradients, Logistic Regression, Interpretable machine learning

Posted Date: May 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9594358/v1

Additional Declarations: No competing interests reported.

Interpretable Color–Texture–Shape Feature Fusion for RGB-Based Citrus Canker and Melanose Classification on Orange Fruit

Hoa-Cuc. Nguyen[1], Ngoc-Duong. Nguyen[1], Viet-Hung. Pham[1],
Bich-Ngoc. Mach[1], Trinh Thi Nhu Quynh[2], Duong Thi Kim Chi[3,✉]

[1]Faculty of Engineering and Technology, Thu Dau Mot University, No. 06 Tran Van On, Phu Loi Ward, Ho Chi Minh City, Viet Nam

[2]School of Law and Management, Thu Dau Mot University, No. 06 Tran Van On, Phu Loi Ward, Ho Chi Minh City, Viet Nam

[3]Institute of Digital Technology, Thu Dau Mot University, No. 06 Tran Van On, Phu Loi Ward, Ho Chi Minh City, Viet Nam

✉Email: chidtk@tdmu.edu.vn

Abstract

Citrus canker and melanose substantially reduce the visual quality and commercial value of orange fruit, yet routine diagnosis remains largely dependent on subjective visual inspection. This study presents an interpretable color–texture–shape feature fusion framework for RGB-based classification of healthy, canker-infected, and melanose-affected orange fruit. Each image was represented by a compact descriptor integrating HSV color histograms, Local Binary Pattern micro-texture features, and Histogram of Oriented Gradients edge–shape information, followed by normalized feature concatenation and one-vs-rest Logistic Regression classification. On a balanced held-out test set, the proposed pipeline achieved 93.93% overall accuracy and a macro-F1 score of approximately 0.94, with class-wise F1-scores of 0.927 for citrus canker, 0.933 for healthy fruit, and 0.958 for melanose. Error analysis showed that residual misclassifications were concentrated mainly along the canker–healthy boundary. These findings demonstrate that well-designed handcrafted descriptors can provide accurate, transparent, and diagnostically meaningful citrus fruit disease recognition.

Keywords: *Citrus canker; Melanose; Orange fruit disease; RGB image classification; HSV color histogram; Local Binary Pattern; Histogram of Oriented Gradients; Logistic Regression; Interpretable machine learning*.

1. Introduction

Citrus fruit production plays an important role in global horticulture, yet its commercial value is highly vulnerable to surface diseases that degrade visual quality, shorten marketability, and increase post-harvest rejection [1]. Among these diseases, citrus canker and melanose are particularly important because their symptoms are expressed through visible rind changes, including dark speckling, roughened lesions, discoloration, pitting, and irregular lesion boundaries. In routine orchard and post-harvest practice, disease recognition still depends largely on manual visual inspection by growers, technicians, or quality-control staff [2]. Although human assessment remains valuable, it is inherently subjective, labor-intensive, and inconsistent under variable illumination, fruit maturity, background conditions, and early-stage symptom expression [3]. These limitations have strengthened the need for image-based diagnostic tools that are accurate, transparent, and practical for repeated use in citrus monitoring workflows.

Recent advances in plant disease recognition have been dominated by deep convolutional neural networks and transfer-learning architectures [4]. These models have demonstrated impressive classification performance across many leaf and fruit disease datasets, especially when

large annotated datasets and adequate computational resources are available [5]. However, their use in routine agricultural environments remains constrained by several practical factors. High-capacity neural models often require extensive training data, specialized hardware, careful model maintenance, and opaque decision mechanisms that are difficult for non-specialist users to interpret [6]. In resource-limited settings, where low-cost cameras, commodity computers, and simple deployment pipelines remain common, the best-performing model is not necessarily the most complex one [7]. For diseases whose symptoms are strongly expressed through color, texture, and lesion morphology, carefully engineered visual descriptors can still provide a strong and scientifically meaningful alternative.

This study developed an interpretable RGB-based classification framework for distinguishing healthy orange fruit, citrus canker, and melanose using a compact color–texture–shape feature representation. The framework combines HSV color histograms to capture chromatic changes, Local Binary Pattern descriptors to encode rind micro-texture, and Histogram of Oriented Gradients features to describe lesion edge and shape structure [8]. These complementary descriptors were concatenated into a normalized feature vector and classified using a one-vs-rest Logistic Regression model [9]. This design preserves a direct link between visual disease phenotypes and model inputs, allowing the resulting decisions to be interpreted in relation to observable fruit-surface characteristics rather than relying on opaque latent embeddings [10].

The experimental results confirmed that this lightweight handcrafted-feature pipeline achieved strong and balanced diagnostic performance on a three-class orange fruit dataset. The model reached 93.93% overall accuracy and a macro-F1 score of approximately 0.94, with particularly strong recognition of melanose and a clearly identifiable residual boundary between early canker symptoms and visually healthy rind texture. Beyond the numerical performance, the confusion structure and qualitative error patterns showed that the remaining errors were concentrated in phenotypically ambiguous cases, such as low-contrast lesions, glare-affected surfaces, and fine pitting patterns. These findings demonstrate that interpretable color–texture–shape fusion remains a robust and practically relevant strategy for RGB-based citrus fruit disease classification, especially when diagnostic transparency and implementation simplicity are valued alongside accuracy.

2. Materials and Methods

2.1. Dataset and classification task

The study was conducted on a balanced RGB image dataset of orange fruit comprising three diagnostic categories: healthy fruit, citrus canker, and melanose. The final evaluation set contained 1,170 images, with 390 images per class, ensuring that the reported metrics reflected intrinsic class separability rather than prior class imbalance. Each sample was denoted as $(\mathbf{I}_i, y_i)$, where $\mathbf{I}_i \in \mathbb{R}^{H\times W\times 3}$ represents the RGB image and $y_i \in \mathcal{Y}$, with $\mathcal{Y} = \{\text{healthy,canker,melanose}\}$, represents the corresponding class label. The classification task was formulated as a three-class supervised learning problem in which the model assigned each fruit image to one of the three mutually exclusive categories.

The dataset was arranged under a stratified protocol so that each class contributed equally to training, model selection, and held-out testing. This design was important because citrus canker and melanose differ not only in lesion color but also in surface texture and lesion geometry. Healthy fruit provided the reference class for normal rind appearance, while canker and melanose represented two disease phenotypes with partially overlapping but visually distinguishable color–

texture–shape signatures. The balanced design therefore allowed the model to be evaluated under a fair diagnostic setting and made macro-averaged metrics directly interpretable.

2.2. Pre-processing

All input images were standardized before feature extraction to reduce variability unrelated to disease phenotype. Each RGB image was resized to a fixed spatial resolution, ensuring that the extracted descriptors had consistent dimensionality across all samples. The RGB representation was retained for visualization, converted to HSV space for color analysis, and transformed into grayscale intensity for texture and gradient-based descriptors. The grayscale image was computed as

$$I_g = 0.299R + 0.587G + 0.114B,$$

where R, G, and B denote the red, green, and blue channels. No aggressive geometric augmentation was applied in the core evaluation, as the objective was to measure the discriminative capacity of the handcrafted descriptors under a controlled and interpretable setting. Feature scaling was estimated only from the training partition and then applied consistently to validation and test data, preventing information leakage.

2.3. Color–texture–shape feature fusion

The proposed representation integrated three complementary groups of handcrafted descriptors: HSV color histograms, Local Binary Pattern features, and Histogram of Oriented Gradients features. **Figure 1** summarizes the full feature-extraction and classification workflow, showing how each RGB image was decomposed into parallel color, texture, and edge–shape branches before feature concatenation and Logistic Regression classification. This design maintained a direct connection between model inputs and observable disease phenotypes: chromatic discoloration, rind micro-texture, and lesion-boundary structure.

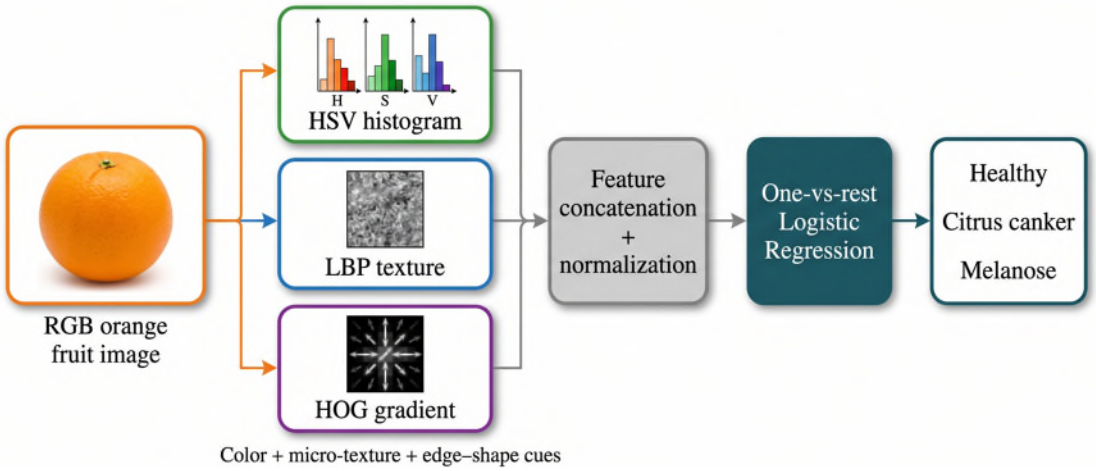

Figure 1. *Overview of the proposed interpretable color–texture–shape feature fusion pipeline*

For the color branch, each image was converted from RGB to HSV space, $\mathbf{I}_{HSV} = (H, S, V)$. For each channel $C \in \{H, S, V\}$, a histogram h_C was computed over predefined bins $\mathcal{B}_b$:

$$h_C[b] = \sum_{p \in \Omega} \mathbf{1}(C(p) \in \mathcal{B}_b), \tag{1}$$

where $p$denotes a pixel location and Ω is the image lattice. Each channel histogram was ℓ_1-normalized to reduce sensitivity to image size and global intensity variation:

$$\tilde{h}_C = \frac{h_C}{\| h_C \|_1}. \tag{2}$$

The final HSV descriptor was obtained by concatenating the normalized channel histograms.

For the texture branch, Local Binary Pattern descriptors were extracted from the grayscale image to capture rind roughness, small pits, and disease-related micro-patterns. At each pixel p, the LBP code was computed as

$$LBP_{P,R}(p) = \sum_{k=0}^{P-1} s\left(I_g(p_k) - I_g(p)\right)2^k, s(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases} \tag{3}$$

The resulting LBP codes were summarized into a normalized histogram, providing a compact description of local surface texture.

For the shape branch, HOG features were computed from grayscale gradients:

$$m = \sqrt{G_x^2 + G_y^2}, \theta = \tan^{-1}\left(\frac{G_y}{G_x}\right), \tag{4}$$

where m and θ denote gradient magnitude and orientation. Orientation histograms were accumulated within local cells and normalized across blocks, allowing the descriptor to represent lesion edges and irregular boundaries. The final image descriptor was constructed as

$$\mathbf{x} = \text{norm}\left(\mathbf{f}_{HSV} \oplus \mathbf{f}_{LBP} \oplus \mathbf{f}_{HOG}\right), \tag{5}$$

where $\oplus$ denotes concatenation and $\text{norm}(\cdot)$ denotes final feature normalization.

2.4. One-vs-rest Logistic Regression classifier

The normalized feature vector $\mathbf{x}$ was classified using a one-vs-rest Logistic Regression model. Three binary classifiers were trained, each separating one target class from the remaining two classes. For class c, the decision score and posterior probability were computed as

$$z_c = \mathbf{w}_c^\top \mathbf{x} + b_c, p_c = \sigma(z_c) = \frac{1}{1 + e^{-z_c}}, \tag{6}$$

where $\mathbf{w}_c$ and b_c are the class-specific weight vector and bias term. Model parameters were estimated by minimizing the regularized logistic loss:

$$\mathcal{L} = -\sum_i \left[y_{i,c} \log\left(p_{i,c}\right) + (1 - y_{i,c}) \log\left(1 - p_{i,c}\right)\right] + \frac{\lambda}{2} \| \mathbf{w}_c \|_2^2. \tag{7}$$

The predicted class was assigned by the maximum posterior rule:

$$\hat{y} = \arg\max_{c \in \mathcal{Y}} p_c. \tag{8}$$

This classifier was selected because it preserved probabilistic outputs, provided transparent linear decision boundaries, and allowed the contribution of each handcrafted feature group to remain interpretable.

2.5. Evaluation protocol and metrics

Model performance was assessed on a stratified held-out test set that was not used during training or model selection. Accuracy, precision, recall, F1-score, macro-F1, and the confusion

matrix were reported to provide both global and class-specific views of diagnostic performance. For each class c, precision, recall, and F1-score were defined as

$$Precision_c = \frac{TP_c}{TP_c + FP_c}, Recall_c = \frac{TP_c}{TP_c + FN_c},$$
$$F1_c = \frac{2Precision_c Recall_c}{Precision_c + Recall_c}. \quad (9)$$

Overall accuracy was computed as

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(\hat{y}_i = y_i). \quad (10$$

Macro-F1 was used as the primary summary metric because it assigns equal importance to healthy, canker, and melanose classes. The confusion matrix was further analyzed to identify structured residual errors, particularly the clinically meaningful boundary between early citrus canker and visually healthy rind texture. This evaluation protocol allowed the study to report not only whether the model was accurate, but also where its remaining diagnostic uncertainty was concentrated.

3. Results

3.1. Overall classification performance

The proposed HSV–LBP–HOG feature fusion pipeline coupled with one-vs-rest Logistic Regression achieved strong and balanced performance on the held-out test set. As summarized in Table 3, the model reached an overall accuracy of 93.93% and an overall F1-score of 0.939 across 1,170 test images, with equal class support of 390 images per category. The macro-F1 score of approximately 0.94 confirms that the performance was not dominated by a single class but was distributed consistently across healthy, citrus canker, and melanose samples.

Figure 2 provides a compact visual comparison of precision, recall, and F1-score across the three categories. The clustered bar pattern shows that all class-level F1-scores remained above 0.92, indicating that the handcrafted descriptor set preserved sufficient discriminative information for reliable three-class separation. Melanose produced the highest F1-score, reflecting the strong separability of its dark speckled lesions and characteristic rind texture. Citrus canker showed a slightly lower but still robust F1-score, mainly due to reduced recall in visually subtle cases. Healthy fruit exhibited high recall, confirming that normal rind appearance was well represented by the fused color–texture–shape descriptor.

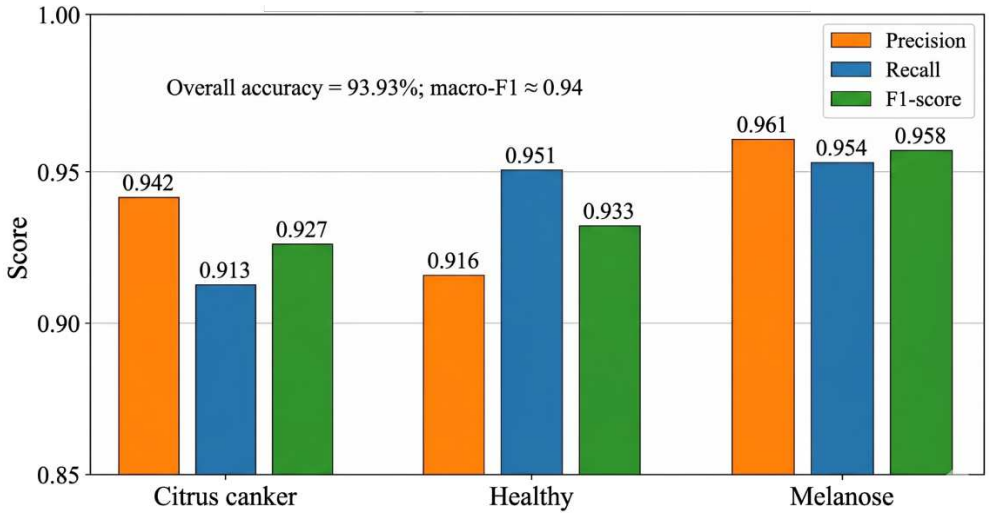

Figure 2. *Class-wise precision, recall, and F1-score of the proposed classifier*

Table 3. Overall and class-wise classification performance on the held-out test set

Class	Precision	Recall	F1-score	Support
Citrus canker	0.942	0.913	0.927	390
Healthy	0.916	0.951	0.933	390
Melanose	0.961	0.954	0.958	390
Overall	—	—	**0.939**	**1170**
Accuracy	—	—	**0.9393**	**1170**

3.2. Class-specific diagnostic behavior

The class-specific results reveal distinct diagnostic behavior for each disease category. Citrus canker obtained high precision (0.942) but lower recall (0.913), indicating that when the classifier predicted canker, the decision was generally reliable, although some canker samples were missed. This pattern is diagnostically meaningful because early or low-contrast canker lesions may resemble normal rind pores and weak surface discoloration. In contrast, healthy fruit achieved the highest recall among the non-melanose categories (0.951), showing that the model effectively captured the baseline appearance of disease-free orange rind.

Melanose showed the strongest overall behavior, with precision of 0.961, recall of 0.954, and F1-score of 0.958. As illustrated in **Figure 3**, melanose occupied the most stable region of the diagnostic profile, with a small precision–recall gap and the highest class-level F1-score. This result indicates that the integrated descriptor captured the dark speckling, surface roughness, and localized edge patterns that characterize melanose more distinctly than the early-stage visual signatures of canker. **Table 4** further interprets these class-wise patterns by linking each metric profile to its dominant visual cues.

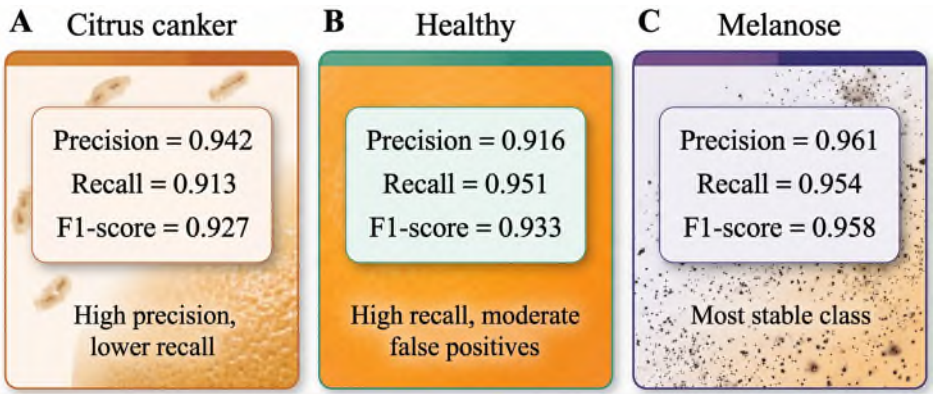

Figure 3. *Class-specific diagnostic profile showing metric balance and phenotype-level interpretation*

Table 4. Diagnostic interpretation of class-specific model behavior

Class	Dominant visual cue	Precision–recall pattern	F1-score	Diagnostic interpretation
Citrus canker	Low-contrast lesions, pitting, roughened spots	P > R	0.927	Reliable positive decisions, but some subtle canker cases were missed
Healthy	Normal rind color and texture	R > P	0.933	Healthy samples were recovered well, with some disease-like texture confusion
Melanose	Dark speckling and localized rind roughness	Balanced P/R	0.958	Most separable class due to distinctive color–texture structure

3.3. Confusion matrix and residual error structure

The confusion matrix in **Figure 4** provides a more detailed account of the remaining diagnostic uncertainty. Correct classifications were concentrated along the diagonal, with 356 citrus canker, 371 healthy, and 372 melanose images correctly assigned. This distribution confirms that the high global accuracy was supported by strong class-wise recognition rather than by uneven performance across categories. The row-normalized structure also shows that melanose and healthy fruit reached recall values above 95%, whereas canker recall remained slightly lower.

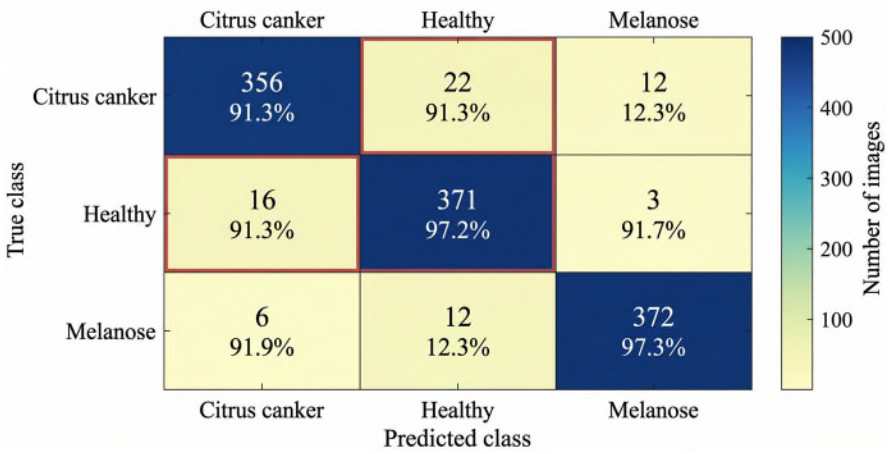

Figure 4. *Confusion matrix of the HSV–LBP–HOG + Logistic Regression classifier*

The residual errors were not randomly distributed. As shown in **Table 5**, the largest off-diagonal error was canker predicted as healthy with 22 cases, equivalent to 5.64% of all canker samples. The second major confusion was healthy predicted as canker with 16 cases, equivalent to 4.10% of healthy samples. Errors involving melanose were comparatively limited, particularly for healthy-to-melanose confusion, which occurred in only three cases. This structure indicates that the dominant decision boundary was located between early canker and visually healthy rind, while melanose remained comparatively well separated. The confusion matrix therefore strengthened the main interpretation of the study: the proposed model was accurate, but its most meaningful limitation occurred where disease phenotype and normal rind texture naturally overlap.

Table 5. Residual error structure derived from the confusion matrix

True class	**Predicted as citrus canker**	**Predicted as healthy**	**Predicted as melanose**	**Main residual pattern**
Citrus canker	356	22	12	Canker missed as healthy
Healthy	16	371	3	Healthy rind misread as canker
Melanose	6	12	372	Limited cross-class leakage
Largest off-diagonal error	—	**22**	—	**Canker → Healthy**

3.4. Qualitative interpretation of misclassified samples

The qualitative error analysis in **Figure 5** confirmed that the numerical errors reflected visually interpretable borderline cases rather than unstable model behavior. Representative false negatives showed canker lesions with weak contrast, small lesion areas, or glare-affected surfaces, causing the color and gradient descriptors to shift toward the healthy class. Conversely, some healthy samples contained natural rind pores, fine pitting, or rough local texture that partially resembled early canker, producing false-positive canker predictions.

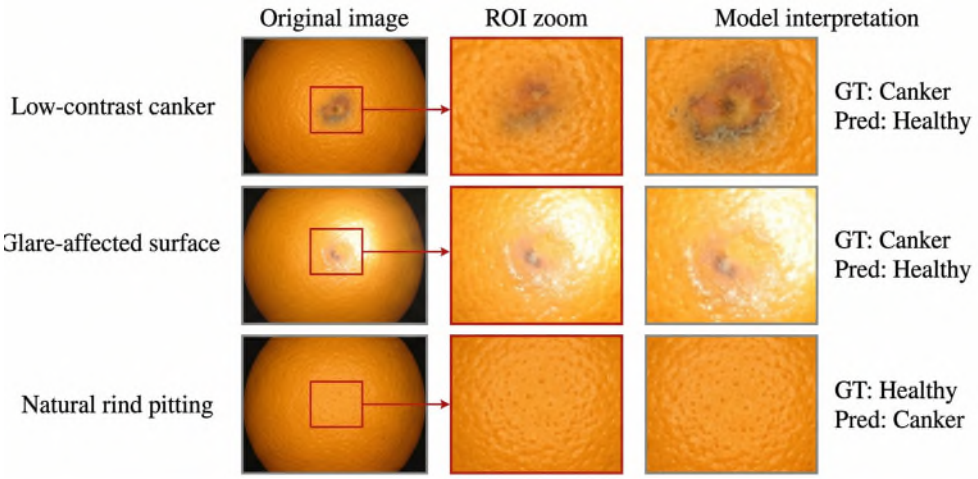

Figure 5. *Representative visual error modes for canker–healthy ambiguity*

Table 6 summarizes the principal visual regimes associated with these errors. Low-contrast lesions mainly explained canker-to-healthy confusion, while naturally rough healthy rind explained healthy-to-canker confusion. Glare and over-exposure reduced the separability of HSV and texture cues by compressing local color differences. These findings show that the classifier's residual errors were concentrated in biologically and optically plausible ambiguity zones. This error profile is important because it identifies clear improvement pathways, including illumination normalization, multi-scale texture extraction, and lesion-focused region-of-interest analysis.

Table 6. Phenotypic interpretation of representative misclassification regimes

Error regime	**Main affected direction**	**Visual cause**	**Descriptor limitation**	**Suggested refinement**
Low-contrast early lesion	Canker → Healthy	Weak discoloration and faint lesion edge	HSV and HOG cues become less distinctive	Illumination normalization
Glare or over-exposure	Canker → Healthy	Saturated surface highlights	Color and texture differences are compressed	Specular highlight control
Natural rind pitting	Healthy → Canker	Fine pores mimic lesion micro-texture	LBP patterns overlap with disease texture	Multi-radius texture features
Small localized lesion	Canker → Healthy/Melanose	Lesion occupies limited image area	Global descriptors dilute local signals	ROI-based feature extraction

4. Conclusion

This study demonstrated that an interpretable color–texture–shape feature fusion pipeline can provide accurate and diagnostically meaningful classification of citrus canker, melanose, and healthy orange fruit from RGB images. By integrating HSV color histograms, Local Binary Pattern micro-texture descriptors, and Histogram of Oriented Gradients edge–shape features with a one-vs-rest Logistic Regression classifier, the proposed framework achieved strong and balanced performance, reaching 93.93% accuracy and a macro-F1 score of approximately 0.94. The main

contribution of this work lies in showing that a carefully engineered handcrafted-feature model can approach the practical accuracy range of more complex vision systems while preserving transparency, simplicity, and phenotype-level interpretability. The residual errors were concentrated mainly along the canker–healthy boundary, reflecting genuine visual ambiguity rather than unstable model behavior. These findings support interpretable RGB-based feature fusion as a reliable and accessible strategy for citrus fruit disease recognition.

Ethics Statement

This study uses an open, publicly available orange-fruit image dataset; no human/animal data are involved. All usage follows the dataset's terms.

Data Availability

All experiments reference the open dataset described in the source papers; class balancing to 2,600 images per class is documented therein. Feature code and training scripts can be packaged for release upon acceptance.

Acknowledgements

We thank the creators and maintainers of the open orange-fruit dataset used in both source papers.

Competing Interest Declaration:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P. K. Yadav, T. Burks, Q. Frederick, J. Qin, M. Kim, and M. A. Ritenour, "Citrus disease detection using convolution neural network generated features and Softmax classifier on hyperspectral image data," *Frontiers in Plant Science,* vol. 13, p. 1043712, 2022, doi: 10.3389/fpls.2022.1043712.

[2] C. Zhang, J. Valente, S. Blackmore, Y. Chen, and M. Li, "Automated identification of citrus diseases in orchards using deep learning," *Biosystems Engineering,* vol. 223, pp. 249-258, 2022, doi: 10.1016/j.biosystemseng.2022.09.006.

[3] Y. Wang, Z. Xiao, H. Wang, F. Li, and J. Tian, "A study on an efficient citrus Huanglong disease detection algorithm based on three-channel aggregated attention," *PeerJ Computer Science,* vol. 11, p. e2943, 2025, doi: 10.7717/peerj-cs.2943.

[4] M. Islam, M. Habib, and M. K. Hossain, "Lemon and Orange Disease Classification using CNN-Extracted Features and Machine Learning Classifier," *arXiv,* 2024, doi: 10.48550/arXiv.2408.14206.

[5] Z. Iqbal, M. A. Khan, M. Sharif, J. H. Shah, M. H. ur Rehman, and K. Javed, "An automated detection and classification of citrus plant diseases using image processing techniques: A

review," *Computers and Electronics in Agriculture,* vol. 153, pp. 12-32, 2018, doi: 10.1016/j.compag.2018.07.032.

[6] B. Kaur, S. K. Gupta, M. Janarthan, D. M. Alsekait, and D. S. AbdElminaam, "Precision diagnosis of citrus leaf diseases using image enhancement and nonlinear fuzzy ranking ensemble approach NLFuRBe," *Scientific Reports,* vol. 15, p. 32296, 2025, doi: 10.1038/s41598-025-16923-4.

[7] A. Khattak, M. U. Asghar, U. Batool, and M. Asghar, "Automatic detection of citrus fruit and leaves diseases using deep neural network model," *IEEE Access,* vol. 9, pp. 112942-112954, 2021, doi: 10.1109/ACCESS.2021.3096895.

[8] N. Butt, M. M. Iqbal, S. Ramzan, A. Raza, L. Abualigah, N. L. Fitriyani, and I. T. R. Yanto, "Citrus diseases detection using innovative deep learning approach and Hybrid Meta-Heuristic," *PLOS ONE,* vol. 20, no. 1, p. e0316081, 2025, doi: 10.1371/journal.pone.0316081.

[9] S. K. Behera, A. K. Rath, A. Mahapatra, and P. K. Sethy, "Disease classification and grading of orange using machine learning and fuzzy logic," in *2018 International Conference on Communication and Signal Processing (ICCSP)*, 2018: IEEE, pp. 0678-0682, doi: 10.1109/ICCSP.2018.8524415.

[10] K. N. Arifin, S. A. Rupa, M. M. Anwar, and I. Jahan, "Lemon and orange disease classification using cnn-extracted features and machine learning classifier," *arXiv preprint arXiv:2408.14206,* 2024.