Field-Reliability Analysis of Lightweight Orange Disease Screening: Image-Quality Effects, Failure Taxonomy, and Practical Triage for Edge Deployment

Nguyen Khac Hoang
Lac Hong University
Nguyen Tuan Kiet
Lac Hong University
Nguyen Hoang Vu
Lac Hong University
Thanh Q. Nguyen
nqthanh@ntt.edu.vn
Nguyen Tat Thanh University

Research Article

Keywords: Orange disease screening, citrus canker, melanose, image quality assessment, failure taxonomy, confidence-aware triage, edge artificial intelligence

Posted Date: May 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9586741/v1

Additional Declarations: No competing interests reported.

Field-Reliability Analysis of Lightweight Orange Disease Screening: Image-Quality Effects, Failure Taxonomy, and Practical Triage for Edge Deployment

Nguyen Khac Hoang[1], Nguyen Tuan Kiet[1], Nguyen Hoang Vu[1], Thanh Q. Nguyen[2]

[1]Faculty of Information Technology, Lac Hong University, Bien Hoa, Dong Nai, 810000, Vietnam

[2]Institute of Interdisciplinary Sciences, Nguyen Tat Thanh University, Ho Chi Minh City 700000, Vietnam

✉Corresponding author: [2]nqthanh@ntt.edu.vn (Thanh Q. Nguyen)

Abstract

Reliable edge-based citrus disease screening requires more than high benchmark accuracy; it must identify when an image and its prediction are trustworthy under field conditions. This study analyzed a balanced three-class orange image corpus comprising healthy fruit, citrus canker, and melanose using a lightweight HSV–LBP–HOG representation and calibrated confidence profiling. Rather than treating all classifications as equally actionable, the analysis linked prediction confidence with image-quality limitations and residual error patterns. The results showed that difficult cases were concentrated around the early canker–healthy boundary, where weak chromatic changes, subtle rind texture, shadowing, blur, and non-disease surface defects reduced diagnostic reliability. A quality-aware triage scheme was therefore established to separate direct acceptance, image recapture, and expert or local refinement. The findings support a practical field-reliability framework for low-cost orange disease screening on mobile and edge devices.

Keywords: *Orange disease screening; citrus canker; melanose; image quality assessment; failure taxonomy; confidence-aware triage; edge artificial intelligence.*

1. Introduction

Image-based plant disease recognition has progressed rapidly from laboratory image analysis to mobile and edge-oriented screening systems. In citrus production, this transition is particularly important because early detection of fruit surface diseases can support timely intervention, reduce unnecessary chemical use, and limit the spread of infection across orchards [1]. However, high classification accuracy on curated datasets does not automatically translate into reliable field operation. Once deployed outside controlled acquisition conditions, disease images are affected by uneven illumination, hard shadows, specular glare, hand-held camera motion, defocus, background clutter, and non-disease rind defects [2]. These factors can distort the visual evidence used by a classifier and may turn a numerically accurate benchmark model into an uncertain decision tool under real orchard conditions.

For citrus canker and melanose, the reliability challenge is not merely the presence of occasional random errors. The most consequential failure occurs when early canker is mistaken for healthy fruit because symptoms are still visually weak [3]. At this stage, chromatic changes may be faint, lesion edges may be poorly defined, and surface roughness may be subtle. Such cases are operationally important because a false healthy decision can delay intervention, whereas a false disease alert can still be corrected through inspection [4]. Therefore, citrus disease screening should not be evaluated solely by overall accuracy; it should also identify which predictions are trustworthy, which images are unsuitable for automated interpretation, and which cases require additional review or local refinement.

Edge deployment further changes the evaluation logic [5]. A field-ready system must operate under constrained computation, limited memory, unstable connectivity, and variable user capture behavior. More importantly, it must provide actionable outputs rather than forced labels for every input image [6]. In practical orchard screening, a high-confidence prediction from a clear image can be accepted directly; a blurred or shadowed image should trigger recapture; and a visually adequate but low-confidence image should be routed for expert review or secondary refinement [7]. This accept–recapture–review structure is central to field reliability, yet it is rarely treated as a formal component of citrus disease classification pipelines.

Handcrafted visual descriptors offer a useful foundation for such reliability analysis because they correspond to interpretable image evidence. HSV color histograms reflect hue, saturation, and brightness changes associated with diseased peel regions, but they are vulnerable to illumination distortion [8, 9]. Local Binary Patterns capture micro-texture and rind roughness, making them relevant to early canker where color evidence is insufficient. Histogram of Oriented Gradients encodes lesion boundaries and local shape structure, but its reliability decreases under blur or defocus [10]. Thus, HSV, LBP, and HOG do not only support lightweight classification; they also provide a transparent lens for understanding why and when misclassification occurs.

Despite growing interest in lightweight agricultural artificial intelligence, limited attention has been given to image quality and failure taxonomy as explicit parts of citrus disease screening [11]. Most studies report predictive performance but provide less guidance on whether a field image should be trusted, reacquired, or escalated [10]. This study addresses that gap by reframing lightweight orange disease recognition as a field-reliability problem. Using a balanced three-class corpus of healthy, citrus canker, and melanose images, the analysis links confidence behavior, image-quality degradation, descriptor-level evidence, and residual error patterns into a practical triage framework for low-cost mobile and edge-based orchard screening.

2. Materials and Methods

2.1. Dataset and image preparation

This study used a balanced orange surface image corpus containing three diagnostic categories: healthy fruit, citrus canker, and melanose. The original dataset was harmonized to support reliability-oriented analysis rather than only classifier benchmarking. Each class contained 2,600 images, yielding a final dataset of 7,800 samples. All images were converted to a standardized 8-bit RGB format, resized to a fixed spatial resolution, and center-cropped where necessary to preserve the fruit surface while reducing background dominance. The dataset was represented as

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}, y_i \in \{\text{healthy,canker,melanose}\}, N = 7800. \quad (1)$$

Stratified image-level splitting was applied to construct training, validation, and test partitions with preserved class proportions. Fixed random seeds and stored split indices were used throughout the experiments to prevent leakage and to ensure that image-quality assessment, classifier confidence analysis, and error auditing were performed under the same reproducible evaluation protocol.

2.2. Lightweight visual descriptors as interpretable disease cues

Each image was encoded using a lightweight descriptor stack composed of HSV color histograms, Local Binary Patterns, and Histogram of Oriented Gradients. The final representation was defined as

$$\mathbf{z}_i = [\mathbf{h}_{HSV,i}; \mathbf{f}_{LBP,i}; \mathbf{f}_{HOG,i}], \hat{\mathbf{z}}_i = \frac{\mathbf{z}_i}{\| \mathbf{z}_i \|_2}. \tag{2}$$

Unlike purely accuracy-oriented feature extraction, these descriptors were retained because each component corresponds to an interpretable disease cue. HSV captured hue, saturation, and brightness changes associated with diseased rind regions, but also reflected the influence of shadow and glare. LBP summarized local micro-texture and was therefore informative for early canker, where lesion roughness may appear before strong discoloration. HOG represented boundary orientation and local shape structure, which helped characterize visible lesion edges but was sensitive to blur and defocus. Figure 1 summarizes how these descriptors were used not only for classification, but also as evidence channels for reliability analysis. This workflow clarifies the central methodological shift of the study: the image was not forced directly into a disease label; instead, its visual quality, confidence behavior, and failure risk were assessed before assigning an operational action.

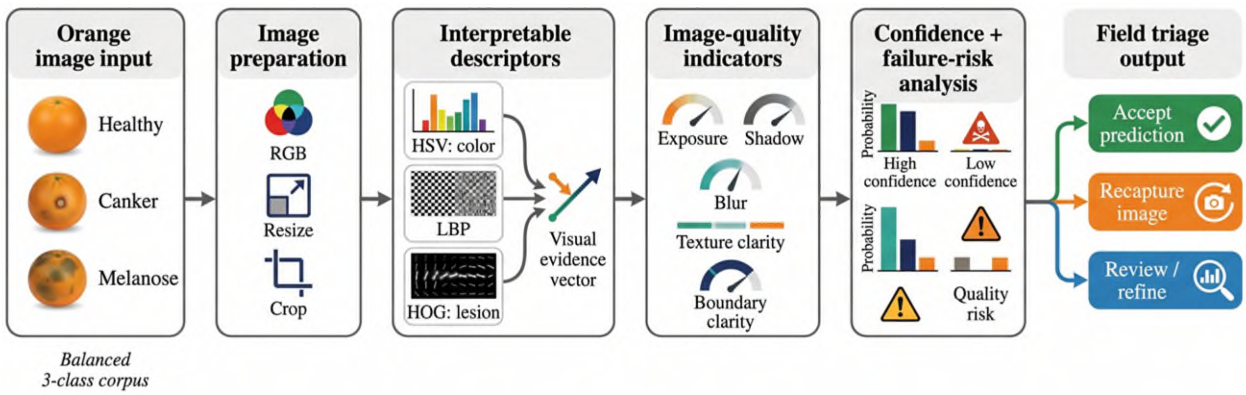

Figure 1. *Quality-aware reliability workflow for lightweight orange disease screening on edge devices*

2.3. Image-quality indicators

To quantify whether an image was suitable for automated interpretation, a compact set of edge-friendly image-quality indicators was computed from the same visual evidence used for classification. Illumination reliability was described using the mean and standard deviation of the HSV value channel, denoted as μ_V and σ_V. A shadow-sensitive ratio was calculated as the proportion of pixels with low value intensity:

$$r_{shadow} = \frac{1}{|\Omega|} \sum_{p \in \Omega} \mathbf{1}\{V(p) < \theta_V\}, \tag{3}$$

where Ω denotes the image domain and θ_V is a validation-defined low-brightness threshold. Color reliability was further characterized through saturation dispersion, since extremely compressed saturation distributions reduced the separability between healthy and diseased surfaces. Blur and defocus were quantified using the variance of the Laplacian response:

$$Q_{blur} = \mathrm{Var}(\nabla^2 I), \tag{4}$$

where lower values indicated weaker high-frequency structure. Texture reliability was summarized by LBP entropy, reflecting whether local rind micro-patterns remained sufficiently informative. Boundary clarity was represented by the average HOG magnitude, which decreased when lesion edges became indistinct. These indicators were collected into a quality vector

$$\mathbf{q}_i = [\mu_V, \sigma_V, r_{shadow}, \sigma_S, Q_{blur}, H_{LBP}, E_{HOG}], \tag{5}$$

allowing each image to be interpreted jointly through exposure, color stability, focus, texture, and lesion-boundary evidence.

2.4. Quality-aware triage protocol and evaluation

The final screening output was structured as an operational triage decision rather than a forced disease label. For each image, the classifier produced a predicted class and confidence score

$$c_i = \max_k P(y_i = k \mid \hat{\mathbf{z}}_i). \tag{6}$$

This confidence score was interpreted together with the image-quality vector. Images with acceptable quality and high confidence were assigned to the accept pathway, meaning that the predicted class was considered directly actionable. Images with poor exposure, strong shadow, blur, or inadequate boundary clarity were assigned to the recapture pathway, because the visual evidence was insufficient for reliable automated diagnosis. Images with acceptable quality but low or intermediate confidence were assigned to the review/refine pathway, reflecting the likelihood of early canker, melanose ambiguity, or hard-negative rind defects. Performance was evaluated using classification accuracy, class-wise recall, canker recall, confidence distribution, error composition by failure mode, recapture rate, review rate, and accepted-sample reliability. This evaluation design measured not only whether the model was correct, but whether the system could identify when its decision was safe for field use.

3. Results

3.1. Overall field-reliability profile

The field-reliability analysis showed that lightweight orange disease screening was not limited by global classification performance alone, but by the concentration of uncertainty in a small number of visually consequential cases. On the balanced test set of 900 images, the optimized Logistic Regression model achieved 94.00% accuracy, while the confidence-gated Hybrid model reached 95.00%, confirming that a compact HSV–LBP–HOG representation retained strong diagnostic value under an edge-oriented setting. However, **Figure 2** shows that the most important result was the structure of the remaining errors. Healthy fruit and clearly expressed disease cases formed stable high-confidence regions, whereas early canker occupied a less reliable transition zone because color evidence was weak and rind texture changes were subtle. This pattern is critical for orchard screening: a false healthy decision for early canker carries greater operational risk than many other misclassifications because it may delay field intervention. **Table 1** summarizes this reliability profile by class. Healthy fruit showed the highest recall under LR, but its main risk was false disease triggering under shadow or surface defects. Canker showed the most important reliability limitation, with residual errors concentrated at the healthy boundary. Melanose remained generally stable, although boundary ambiguity increased under glare or flattened chromatic contrast. These results justified reframing the task from pure label prediction to reliability-aware decision support.

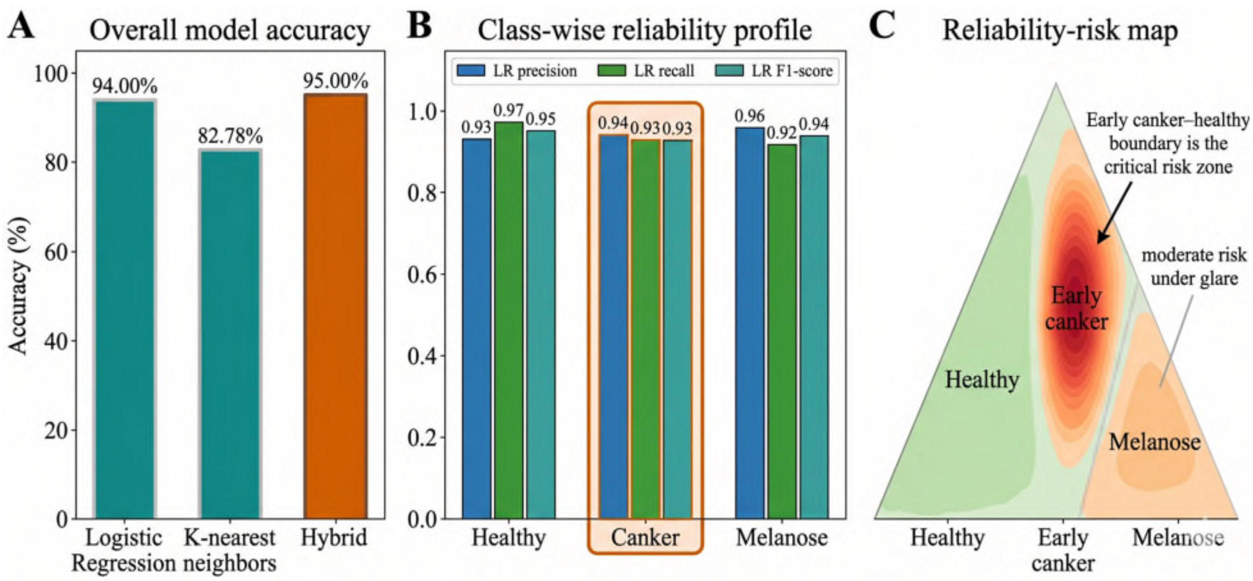

Figure 2. *Overall field-reliability map of lightweight orange disease screening*

Table 1. Class-wise field-reliability profile on the balanced test set

Class	LR precision	LR recall	LR F1-score	Dominant reliability risk	Field interpretation
Healthy	0.93	0.97	0.95	False disease under shadow or rind scars	Reliable when image quality is clear
Canker	0.94	0.93	0.93	Early canker misread as healthy	Highest-risk class for field triage
Melanose	0.96	0.92	0.94	Confusion with canker under glare	Stable when lesion boundaries are visible
Macro-average	0.94	0.94	0.94	Boundary ambiguity	Suitable for reliability-guided screening

3.2. Image-quality effects on classification errors

The residual-error audit demonstrated that most remaining mistakes were strongly associated with identifiable image-quality limitations rather than random classifier behavior. As shown in **Figure 3**, illumination and shadow effects accounted for the largest share of audited failures, representing 20 of 45 residual errors. These cases typically compressed or distorted HSV evidence, causing healthy peel, early canker, and melanose regions to appear more visually similar than they were under clear acquisition. Motion blur and defocus formed the second largest group, with 15 residual errors. This degradation weakened micro-texture and boundary information, reducing the reliability of both LBP and HOG cues. Non-disease rind defects accounted for the remaining 10 errors and represented a different kind of reliability challenge: the image quality was often acceptable, but scars, scratches, or mechanical marks produced texture patterns that resembled pathological lesions. **Table 2** links each failure mode to its visual cause, descriptor-level disruption, typical confusion, and practical mitigation. The result is operationally important because the first two categories can often be addressed by recapture or acquisition guidance, while the third category requires hard-negative enrichment and review. This separation converted the error analysis from a retrospective performance explanation into a field-actionable reliability taxonomy. It also showed that improving real-world performance does not necessarily require heavier models; in many cases, better capture control and quality gating directly target the dominant sources of failure.

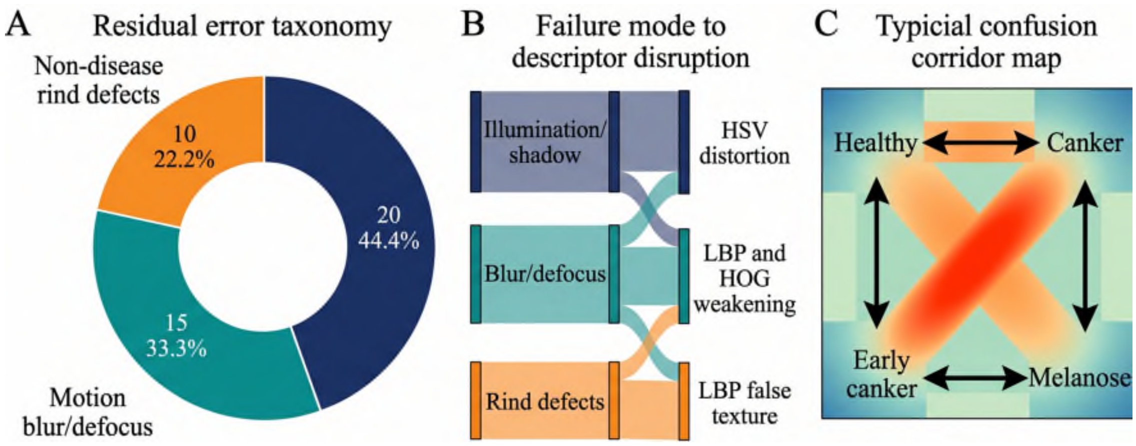

Figure 3. *Image-quality drivers of residual classification failures*

Table 2. Failure taxonomy and descriptor-level interpretation of residual errors

Failure mode	**Count**	**Share of residual errors (%)**	**Main visual cause**	**Descriptor most affected**	**Typical confusion**	**Practical mitigation**
Illumination / shadow	20	44.4	Uneven exposure, hard shadows, glare	HSV	Healthy ↔ Canker; Canker ↔ Melanose	Exposure normalization; recapture
Motion blur / defocus	15	33.3	Hand motion, poor focus, low sharpness	LBP / HOG	Early canker ↔ Healthy	Blur gate; sharper acquisition
Non-disease rind defects	10	22.2	Scars, scratches, mechanical blemishes	LBP	Healthy ↔ Disease	Hard-negative mining; review
Total audited residual errors	45	100.0	—	—	—	—

3.3. Quality-aware triage performance

The quality-aware triage analysis translated confidence and image-quality evidence into three operational pathways: accept, recapture, and review/refine. At the selected confidence operating point, 73.2% of test images followed the direct-accept pathway, while 26.8% were assigned to secondary scrutiny through review or local refinement. **Figure 4** illustrates this behavior as a field workflow rather than a conventional classifier output. The direct-accept pathway contained images with clear acquisition and high model confidence, where automated prediction was sufficiently reliable for routine screening. The review/refine pathway concentrated low- and mid-confidence images, including early canker and lesion-boundary ambiguity. The recapture pathway was informed by the residual-error audit: illumination/shadow and blur/defocus together represented 35 of 45 residual errors, indicating that many unsafe decisions were not intrinsic disease ambiguities but acquisition-quality failures. **Table 3** summarizes the operational triage evidence. The most relevant result is that the model was not forced to issue the same type of decision for every input. Instead, the system separated reliable predictions from images that required better visual evidence or additional scrutiny. This structure is particularly suitable for mobile orchard screening because a grower-facing system can immediately accept clear images, request another capture when the image is unreliable, and reserve expert or local refinement for biologically ambiguous cases.

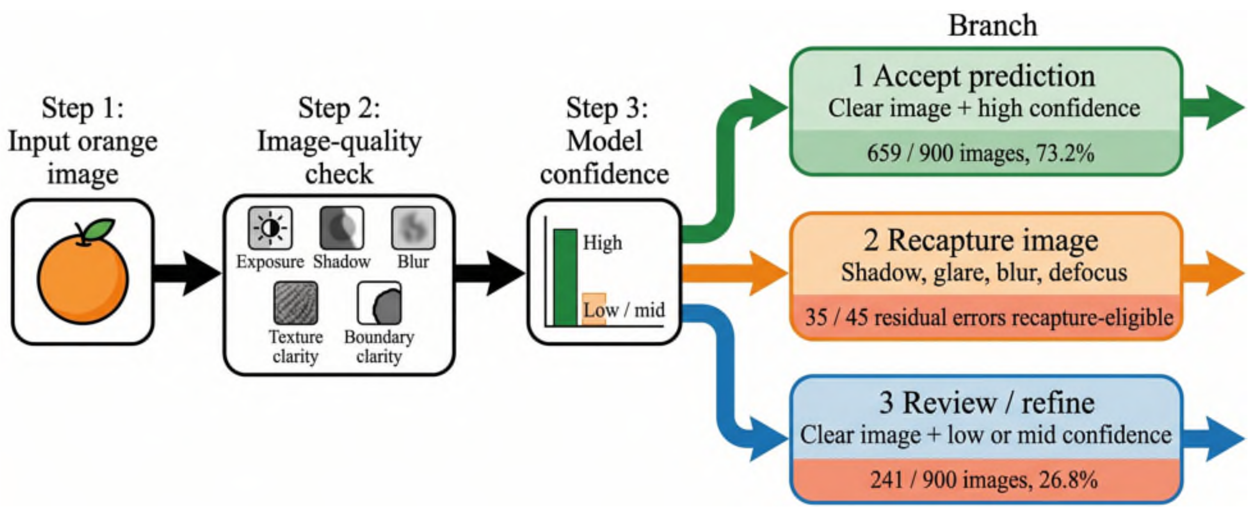

Figure 4. *Quality-aware accept–recapture–review triage behavior for edge-based screening*

Table 3. Operational triage summary at the selected confidence operating point

Triage pathway	**Main trigger**	**Number of images / cases**	**Share (%)**	**Operational meaning**
Accept prediction	Clear image and high confidence	659 of 900 test images	73.2	Fast automated decision
Review / refine	Low or mid confidence	241 of 900 test images	26.8	Secondary scrutiny for ambiguous cases
Recapture-eligible residual cases	Shadow, glare, blur, or defocus among errors	35 of 45 residual errors	77.8	Image quality should be improved before decision
Hard-negative review cases	Non-disease rind defects among errors	10 of 45 residual errors	22.2	Requires human review or hard-negative training

3.4. Practical implications for edge-based orchard screening

The results indicate that field-ready citrus disease screening should be judged by decision reliability rather than by test accuracy alone. **Figure 5** presents the practical deployment concept that emerged from the analysis. A mobile or edge device first captures the orange surface, then applies a lightweight quality check, computes interpretable HSV–LBP–HOG evidence, and produces either an accepted label, a recapture request, or a review/refine recommendation. This output structure is more useful for orchard practice than a forced class label because it reduces unsafe decisions under poor visual evidence. **Table 4** contrasts the conventional accuracy-oriented workflow with the proposed field-reliability-oriented workflow. The difference is particularly relevant for early canker. When a clear image receives high confidence, automated screening is efficient and defensible. When the image is blurred or shadowed, a recapture instruction is more reliable than accepting a potentially misleading label. When confidence remains low despite acceptable image quality, the case is likely to represent genuine biological ambiguity, hard-negative rind defects, or early-stage disease expression, and should therefore be reviewed or refined. This interpretation also clarifies why lightweight descriptors remain valuable. HSV, LBP, and HOG do not only support low-cost classification; they identify the visual channel through which reliability breaks down. The practical next step is therefore not simply to add model complexity, but to strengthen acquisition guidance, illumination adaptation, blur warnings, and hard-negative enrichment for mobile orchard deployment.

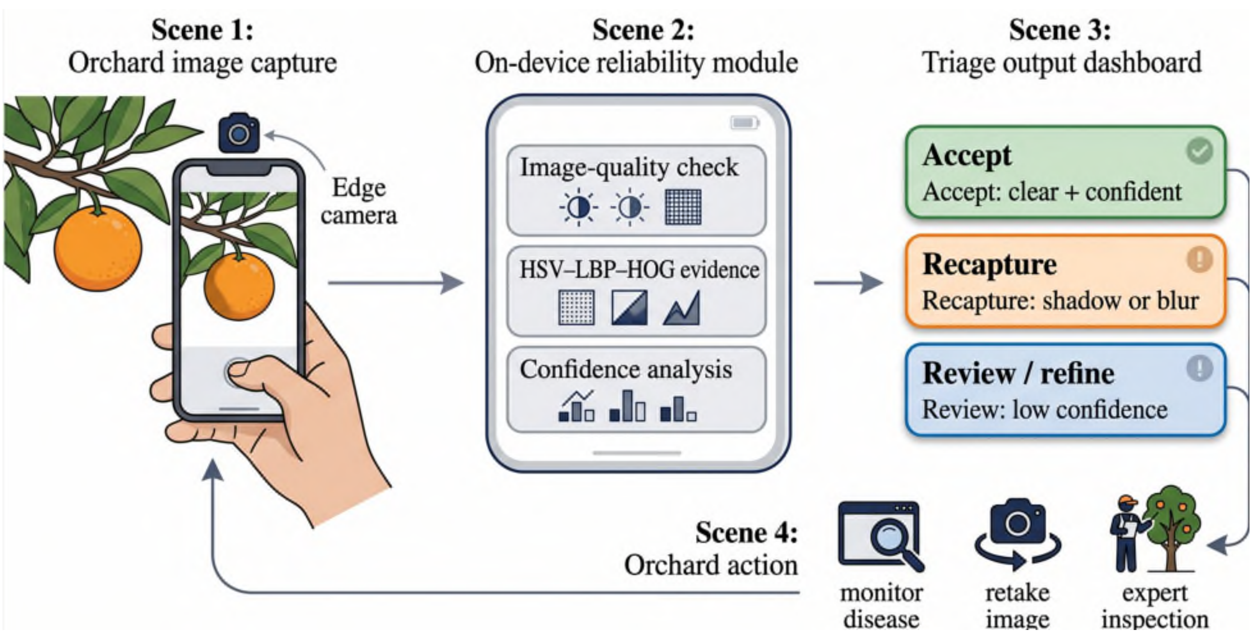

Figure 5. *Practical deployment concept for quality-aware citrus disease screening on edge devices*

Table 4. Accuracy-oriented classification versus field-reliability-oriented screening

Dimension	**Accuracy-oriented classifier**	**Field-reliability-oriented screening**	**Evidence from this study**
Primary goal	Maximize test accuracy	Avoid unsafe field decisions	Errors concentrated in specific visual-risk zones
Low-quality image	Still receives a forced label	Recapture is requested	77.8% of residual errors were recapture-eligible

Low-confidence prediction	Treated as a weak label	Routed to review/refinement	26.8% of test images required secondary scrutiny
Output type	Disease class only	Class label plus operational action	Accept, recapture, or review/refine
Deployment value	Benchmark performance	Orchard usability and risk control	Supports mobile and edge screening

4. Conclusion

This study reframed lightweight orange disease recognition from a conventional accuracy-oriented classification task into a field-reliability problem for mobile and edge-based orchard screening. Using interpretable HSV, LBP, and HOG visual evidence, the analysis showed that residual errors were not randomly distributed but were concentrated in operationally meaningful risk zones, especially the early canker–healthy boundary and images affected by shadow, glare, blur, defocus, or non-disease rind defects. The main contribution of the study is a quality-aware screening framework that links prediction confidence, image-quality indicators, descriptor-level failure modes, and practical triage actions. Instead of forcing every image into a disease label, the proposed workflow separates cases into accept, recapture, and review/refine pathways, thereby reducing unsafe decisions under unreliable visual evidence. These findings support a more deployable form of citrus disease screening, where model output is coupled with field-action guidance. Future work should validate the framework across orchards, devices, lighting conditions, and larger hard-negative datasets.

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] N. Butt, M. M. Iqbal, S. Ramzan, A. Raza, and L. Abualigah, "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.

[2] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. Nanehkaran, "Using deep transfer learning for image-based plant disease identification," *Computers and Electronics in Agriculture,* vol. 173, p. 105393, 2020, doi: 10.1016/j.compag.2020.105393.

[3] A. Abade, P. A. Ferreira, and F. B. Vidal, "Plant diseases recognition on images using convolutional neural networks: A systematic review," *Computers and Electronics in Agriculture,* vol. 185, pp. 106125-106125, 2021, doi: 10.1016/j.compag.2021.106125.

[4] J. Eunice, K. M. Priya, and C. N. Lakshmi, "Deep Learning-Based Leaf Disease Detection in Crops," *Agronomy,* vol. 12, no. 10, p. 2395, 2022, doi: 10.3390/agronomy12102395.

[5] A. Foroughi, J. Lloret, J. M. Jimenez, and S. Sendra, "An edge computing wireless sensor network for diagnosing orange fruit disease," *Cluster Computing,* vol. 28, no. 5, pp. 1-26, 2025.

[6] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture,* vol. 145, pp. 311-318, 2018, doi: 10.1016/j.compag.2018.01.009.

[7] 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 (NL-FuRBe)," *Scientific Reports,* vol. 15, p. 32296, 2025, doi: 10.1038/s41598-025-16923-4.

[8] J. C. F. da Silva, M. C. Silva, E. J. S. Luz, S. Delabrida, and R. A. R. Oliveira, "Using Mobile Edge AI to Detect and Map Diseases in Citrus Orchards," *Sensors,* vol. 23, no. 4, p. 2165, 2023, doi: 10.3390/s23042165.

[9] S. Dananjayan, Y. Tang, J. Zhuang, C. Hou, and S. Luo, "Assessment of state-of-the-art deep learning based citrus disease detection techniques using annotated optical leaf images," *Computers and Electronics in Agriculture,* vol. 193, p. 106658, 2022, doi: 10.1016/j.compag.2021.106658.

[10] M. Sharif, M. A. Khan, Z. Iqbal, M. F. Azam, M. I. U. Lali, and M. Y. Javed, "Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection," *Computers and Electronics in Agriculture,* vol. 150, pp. 220-234, 2018, doi: 10.1016/j.compag.2018.04.023.

[11] P. K. Yadav, T. Burks, J. Qin, M. Kim, Q. Frederick, M. M. Dewdney, and M. A. Ritenour, "Automated classification of citrus disease on fruits and leaves using convolutional neural network generated features from hyperspectral images and machine learning classifiers," *Journal of Applied Remote Sensing,* vol. 18, no. 1, pp. 014512-014512, 2024.