

RESEARCH ARTICLE

Hardware-aware diagnostic information design: A tiny machine learning (TinyML) edge framework for late blight surveillance

Jianshu Li¹, Yuyang Xi¹, TaoSong², Chuan Yang^{1,*}

¹Sanjiang University, Nanjing, Jiangsu, China. ²Zhejiang Polytechnic University of Mechanical and Electrical Engineering, Hangzhou, Zhejiang, China.

Received: May 11, 2026; accepted: August 26, 2026.

Late blight, caused by *Phytophthora infestans*, remains one of the most destructive diseases affecting global potato production, particularly in resource-constrained regions where access to agricultural infrastructure and reliable cloud connectivity is limited. Although deep learning models, especially convolutional neural networks (CNNs), have achieved high diagnostic accuracy, their memory and computational requirements often exceed the low-cost microcontrollers used in agricultural Internet of Things (IoT) systems. This study proposed a hardware-aware Tiny Machine Learning (TinyML) framework for potato late blight detection, in which the processing pipeline was designed according to explicit edge-device constraints including flash memory capacity, static random-access memory (SRAM) availability, and over-the-air (OTA)-related storage requirements. The proposed pipeline integrated Local Binary Patterns (LBP) for texture encoding, Hu moments for geometric characterization, and a Random Forest classifier to construct a lightweight 17-dimensional feature representation suitable for predominantly integer-based computation. Using the potato subset of the PlantVillage dataset, the proposed framework achieved an overall accuracy of 83.78%, a recall of 0.96 for late blight, and a model-file size of 0.64 MB. The model file corresponded to approximately 16% of the documented 4 MB flash capacity of the ESP32-WROOM-32 module, indicating model-level storage feasibility under the reference hardware constraint. Cross-domain evaluation on a 36-image PlantDoc subset yielded a recall of 1.00 but reduced specificity, indicating that disease sensitivity was maintained in the evaluated subset at the cost of increased false-positive predictions. Comparison with the quantized MobileNetV2 baseline demonstrated a clear accuracy–model-file-size trade-off that the proposed framework reduced the model-file size from 2.66 MB to 0.64 MB with test accuracy decreasing from 100% to 83.78%. These findings indicated that interpretable feature engineering, when framed through hardware-aware information design, could provide a memory-efficient alternative for edge-based early-warning applications, in which model storage and diagnostic sensitivity were primary design considerations.

Keywords: Tiny Machine Learning (TinyML); late blight (*Phytophthora infestans*); edge computing; precision agriculture; resource-constrained IoT; information design.

*Corresponding author: Chuan Yang, Sanjiang University, Nanjing, Jiangsu 210012, China. Email: artshirui@163.com.

Introduction

Potato (*Solanum tuberosum* L.) is a major contributor to global food security and ranks as the third most important food crop worldwide in

terms of human consumption after rice and wheat. However, potato production is continuously threatened by *Phytophthora infestans*, the oomycete pathogen responsible for late blight. The disease contributed to the

Irish Potato Famine in the nineteenth century and remains a major cause of crop loss. Late blight imposes substantial annual losses through reduced yields and disease-management costs [1]. In Kenya, expert estimates attribute approximately 23% of annual potato production losses to late blight [2]. The development of accessible early-warning diagnostic tools is therefore important for protecting potato production in resource-constrained agricultural environments.

Recent advances in deep learning, particularly convolutional neural networks (CNNs), have substantially improved automated plant disease classification. Deep residual architecture has achieved high diagnostic performance. A previous study showed that a ResNet-based model achieved a test accuracy of 99.25% on PlantVillage-derived potato and tomato leaf images [3, 4]. Nevertheless, a substantial deployment gap remains between laboratory-based model development and practical field application. Many high-performance CNNs require considerable memory capacity and floating-point computation and therefore often rely on GPU-enabled, high-performance, or cloud-based computing environments. Dependence on cloud infrastructure may restrict their applicability in rural communities where internet connectivity is intermittent, characterized by high latency, or unavailable [5]. Even for architecture derived from MobileNetV2, deployment on resource-constrained agricultural equipment still requires specific optimizations regarding model size, latency, and memory consumption [6]. Tiny Machine Learning (TinyML) has emerged as an alternative approach that enables machine-learning inference to be performed directly on low-power, resource-constrained edge devices [7, 8]. Local processing can reduce communication latency, optimize bandwidth use, and decrease dependence on continuous cloud connectivity [5, 7]. Previous studies have also used deep-learning models with unmanned aerial vehicle (UAV)-based hyperspectral imagery for automated potato late blight diagnosis [9]. However, compressing deep

neural networks introduces an accuracy-resource trade-off and does not automatically guarantee suitability for microcontroller-class devices [7]. Diagnostic systems intended for low-cost agricultural use therefore require algorithms that balance diagnostic performance with storage, memory, and computational limitations of embedded hardware.

This study developed a hardware-aware TinyML framework for potato late blight surveillance in resource-constrained environments. Instead of relying on computationally intensive CNN feature extraction, the framework employed interpretable handcrafted descriptors selected according to the resource limitations of low-cost microcontrollers. Local Binary Patterns (LBP) were used to encode local grayscale texture [10], while Hu moments provided compact geometric descriptors [11]. The texture and geometric descriptors were combined into a 17-dimensional feature vector and classified using a Random Forest model. The framework was formulated according to explicit hardware constraints including flash storage capacity, static random-access memory availability, and over-the-air update-related storage considerations. By treating late blight surveillance as a diagnostic information design problem, the study investigated how pathological image information could be selected, compressed, and classified within strict edge-device limitations. The proposed approach was intended to provide a practical balance among diagnostic sensitivity, model interpretability, storage efficiency, and prospective deployability in resource-constrained agricultural settings.

Materials and methods

Dataset preparation and preprocessing

The potato images used in this study were obtained from the publicly available PlantVillage dataset (<https://github.com/spMohanty/PlantVillage-Dataset>) [12], which contained healthy, early blight, and late blight categories. The primary task was formulated as a binary

classification problem between healthy and late blight images, which was motivated by the epidemiological importance of *Phytophthora infestans* that could spread rapidly under favorable environmental conditions and cause severe crop losses. The framework was therefore designed as a high-sensitivity early-warning tool for the target disease. Early blight images were excluded from model fitting and reserved for a separate out-of-distribution evaluation to assess the model's response to a visually related but non-target potato disease. A total of 552 images were included in the primary binary experiment, comprising 152 healthy images and 400 late blight images. The healthy images accounted for 27.5% of the primary dataset. A class-adjusted hold-out procedure was used to obtain an overall training-to-test ratio of approximately 80:20 while retaining a somewhat larger proportion of healthy control images in the test set. The training set contained 441 images including 113 healthy and 328 late blight samples, whereas the held-out test set contained 111 images including 39 healthy and 72 late blight samples. The healthy images therefore accounted for 35.1% of the held-out test set. This allocation provided a larger control subset for estimating specificity while preserving an overall test proportion of 20.1%. Healthy and late blight were encoded as class 0 and class 1, respectively. All images were resized to 256×256 pixels. Red, green, and blue (RGB) images were subsequently converted to single-channel grayscale images. This conversion reduced the uncompressed input representation from $256 \times 256 \times 3$ values to $256 \times 256 \times 1$ values, corresponding to a 66.7% reduction in input data volume. Grayscale input was used because the Local Binary Patterns descriptor encoded local intensity relationships rather than color-channel information.

Reference hardware constraints

The ESP32-WROOM-32 module was used as a reference hardware profile for defining storage and memory constraints (Espressif Systems, Shanghai, China) [13]. According to the manufacturer's instructions, the module provided 4 MB of external serial peripheral

interface (SPI) flash memory and 520 KB of on-chip static random-access memory (SRAM). In a complete embedded implementation, part of the available memory would also be required for an operating system including FreeRTOS (<https://www.freertos.org/>), communication stacks, application firmware, and dynamic image buffering.

Hardware-aware feature engineering pipeline

A compact feature space was constructed to reduce the model representation while preserving texture and geometric information relevant to the binary classification task. The pipeline consisted of grayscale preprocessing, Local Binary Patterns feature extraction, Hu moments calculation, feature concatenation, and Random Forest classification. The ESP32 specification was used only to establish an analytical design budget. No physical microcontroller deployment, firmware integration, over-the-air (OTA) update test, power measurement, peak-SRAM measurement, or on-device latency experiment was conducted in this study. A nominal model-storage target of less than 0.8 MB, corresponding to less than 20% of the documented 4 MB flash capacity, was defined to reserve capacity for firmware, libraries, and application logic. Meeting this model-size target indicated storage feasibility under the reference constraint but did not by itself demonstrate full firmware compatibility or verified OTA operation. The feature-extraction and classification pipeline was therefore designed to favor comparison-based and predominantly integer-based operations. Floating-point computation was retained where required for the normalization used in Hu moments.

Texture encoding by using Local Binary Patterns

Uniform Local Binary Patterns (LBP) were used to encode local grayscale texture [10]. For a center pixel g_{cg_c} and P neighboring pixels g_{pg_p} sampled at radius R , the basic LBP code was calculated as follows.

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (1)$$

$$s(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

where g_c was the center-pixel intensity. g_p was the intensity of the p -th neighboring sample. The calculation was based mainly on local comparisons and binary encoding rather than repeated convolutional multiply–accumulate operations. The implemented uniform mapping grouped patterns with no more than two bitwise transitions and summarized the resulting codes as a 10-bin histogram that formed the 10-dimensional texture component of the feature vector.

Geometric characterization using Hu moments

Seven Hu invariant moments were calculated from normalized central moments [14] (<https://ieeexplore.ieee.org/document/1057692>) to provide a compact representation of the spatial geometry in the preprocessed image. Hu moments provided invariance to translation, scale, and in-plane rotation under their mathematical formulation. The first Hu invariant was defined as below.

$$\phi_1 = \eta_{20} + \eta_{02} \quad (2)$$

where η_{pq} was the normalized central moment of order $p + q$. The seven Hu descriptors were used as global geometric features and were concatenated with the 10-dimensional LBP histogram to form a 17-dimensional hybrid feature vector. Because Hu moment normalization involved floating-point arithmetic, the complete pipeline was predominantly integer-based rather than integer-only.

Random Forest classifier

The 17-dimensional hybrid feature vectors were classified using a Random Forest (RF) model [15]. A Support Vector Machine (SVM) was considered as an alternative classifier [16], whereas RF was adopted in the present study because its tree-based threshold decisions provided a direct,

comparison-based inference structure that could be translated into conditional logic for a prospective embedded implementation. The ensemble contained 100 decision trees with a maximum depth of 8. These parameters were fixed for the reported software experiments.

Quantized MobileNetV2 baseline

A standard MobileNetV2 architecture, originally proposed by Sandler *et al.* [17] (https://openaccess.thecvf.com/content_cvpr_2018/html/Sandler_MobileNetV2_Inverted_Residuals_CVPR_2018_paper.html), was used as a software-side deep-learning baseline. The model was trained on the same binary PlantVillage subset and converted to 8-bit integer precision using post-training quantization (PTQ). The comparison was limited to test-set classification performance and generated model-file size under the prespecified storage budget. No firmware integration or OTA compatibility test was performed.

Evaluation metrics and protocol

Diagnostic performance was evaluated using overall accuracy, precision, recall (sensitivity), and F1-score. Recall was treated as the primary diagnostic metric because a false-negative prediction would leave a diseased sample undetected, whereas a false-positive prediction could be subjected to subsequent manual verification. Precision and F1-score were also reported to characterize the effect of false-positive predictions. All classifiers were evaluated on the same held-out test set of 111 PlantVillage images. Illumination robustness was assessed by applying gamma correction over the reported range of $\gamma = 0.4 - 2.5$. The early blight category was evaluated separately as an out-of-distribution non-target disease. Cross-domain evaluation used 36 potato images from the publicly available PlantDoc dataset, comprising 18 healthy and 18 late blight images (<https://github.com/pratikkayal/PlantDoc-Dataset>). The feature ablation was conducted by evaluating Hu moments alone, LBP features alone, and the combined 17-dimensional representation while retaining the same Random Forest configuration

and held-out test set. Model storage was evaluated using the generated model file. Runtime was measured only in the host software environment used for experimentation and was treated as software-side reference, which could not be interpreted as ESP32 inference latency.

Results and discussion

Diagnostic performance evaluation

The proposed LBP-RF framework was evaluated on the held-out PlantVillage test set comprising 111 images. The model achieved an overall accuracy of 83.78%, a recall of 0.96, a precision of 0.82, and an F1-score of 0.88 for the late blight class. 69 of the 72 late blight images were correctly classified, whereas three diseased images were classified as healthy. Among the 39 healthy images, 24 were correctly classified and 15 were classified as late blight (false positives) (Figure 1).

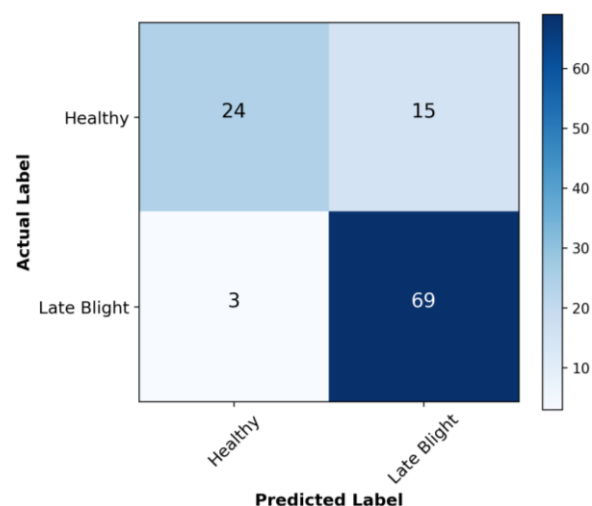


Figure 1. Confusion matrix of the LBP-RF model on the held-out test set.

The relatively high recall was consistent with the intended use of the framework as an early-warning screening tool, in which missed disease cases were considered more consequential than false alarms. However, the 15 false-positive predictions indicated that the model did not

provide sufficiently high specificity for definitive pathogen identification. Positive predictions should therefore be followed by visual inspection or additional diagnostic verification rather than being interpreted as conclusive confirmation of *Phytophthora infestans* infection.

Analysis of inter-class specificity

In the out-of-distribution evaluation of early blight images using *Alternaria solani*, 28.5% of the samples were classified as healthy and 71.5% were classified as late blight. Because early blight was excluded from the binary model-training task, these predictions were used to examine whether the feature representation responded specifically to late blight or more generally to visually similar pathological symptoms. The high proportion of early blight images classified as late blight suggested that the handcrafted representation captured general necrotic texture and geometric irregularity rather than pathogen-specific biological characteristics. Consequently, a positive prediction was more appropriately interpreted as a non-specific pathology alert requiring subsequent verification. The framework might support preliminary identification of potentially diseased leaves, but it could not reliably distinguish late blight from non-target diseases with overlapping visual symptoms.

Resource efficiency and reference-hardware feasibility

The generated LBP-RF model file had a storage size of 0.64 MB, corresponding to approximately 16% of the documented 4 MB flash capacity of the ESP32-WROOM-32 reference module [13]. The model also remained below the prespecified storage target of 0.8 MB. These results indicated model-level storage feasibility under the reference hardware constraint. Model-file size alone did not demonstrate that a complete embedded application would fit within the available memory, because firmware, libraries, communication functions, image acquisition, intermediate feature computation, and dynamic image buffering would require additional resources. In the host software environment,

model training required 0.27 s, and the software pipeline recorded an average runtime of 7.93 ms per image. This value represented software-side runtime in the experimental Python environment and should not be interpreted as ESP32 inference latency. No physical microcontroller deployment or device-side performance evaluation was conducted.

Feature ablation study

The feature ablation experiment showed that the seven-dimensional Hu moment representation achieved an accuracy of 75.68%, whereas the ten-dimensional LBP representation achieved 81.08%. Combining the two feature groups into the proposed 17-dimensional representation increased accuracy to 83.78%. The hybrid representation therefore improved accuracy by 2.70 percentage points compared with LBP alone and by 8.10 percentage points compared with Hu moments alone. These findings indicated that local texture information contributed more strongly to classification performance than global geometric information when each feature group was used independently. Nevertheless, the additional improvement obtained through feature fusion suggested that the two descriptors provided complementary information. LBP represented local grayscale contrast patterns associated with pathological texture, whereas Hu moments summarized the global spatial geometry of the preprocessed image. The ablation results demonstrated the relative contribution of the two feature groups, although they did not establish a direct biological relationship between individual feature values and specific pathological mechanisms.

Accuracy–model-file-size trade-off

The proposed LBP-RF framework achieved an accuracy of 83.78% with a generated model-file size of 0.64 MB, whereas the quantized MobileNetV2 baseline based on the architecture proposed by Sandler *et al.* [17] achieved an accuracy of 100% with a model-file size of 2.66 MB on the same binary PlantVillage test set. The quantized MobileNetV2 model file was

approximately 4.16 times larger than that of the proposed framework (Figure 2).

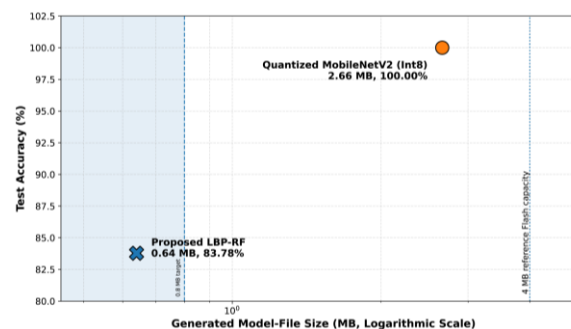


Figure 2. Accuracy–model-file-size trade-off between the proposed LBP-RF framework and the quantized MobileNetV2 baseline on the same binary PlantVillage test set.

The results demonstrated a clear accuracy–model-file-size trade-off. The proposed framework remained below the prespecified 0.8 MB model-storage target, whereas the quantized MobileNetV2 baseline exceeded this analytical target. The comparison was limited to test-set classification performance and generated model-file size and did not establish device-side processing speed, energy efficiency, firmware compatibility, or field readiness.

Illumination robustness analysis

Illumination robustness was assessed by applying simulated gamma correction using values from 0.4 to 2.5. Across the evaluated gamma-corrected images, the reported recall remained 1.00 throughout the tested range (Figure 3). The result indicated that disease sensitivity was stable under the simulated global intensity transformations used in this study. This stability was consistent with the use of local relative-intensity comparisons in the LBP descriptor. However, gamma correction represented a controlled global transformation and did not reproduce the full range of field illumination effects. Local shadows, specular reflections, uneven sunlight, changes in spectral composition, sensor noise, and motion blur were not evaluated. The result therefore demonstrated robustness to the tested gamma

transformations rather than general invariance to all outdoor lighting conditions.

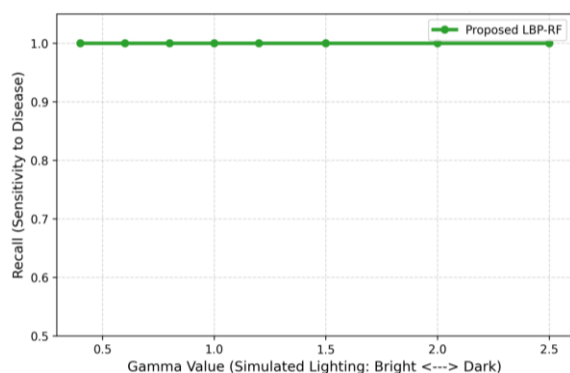


Figure 3. Recall obtained under simulated gamma correction over the range $\gamma = 0.4 - 2.5$.

Cross-domain evaluation on PlantDoc

Cross-domain performance was evaluated using a 36-image PlantDoc subset comprising 18 healthy and 18 late blight images. The proposed framework achieved an overall accuracy of 63.89%. All 18 late blight images were classified as positive, corresponding to a recall of 1.00 in this subset. However, 13 of the 18 healthy images were incorrectly classified as late blight, and only five healthy images were correctly identified (Figure 4).

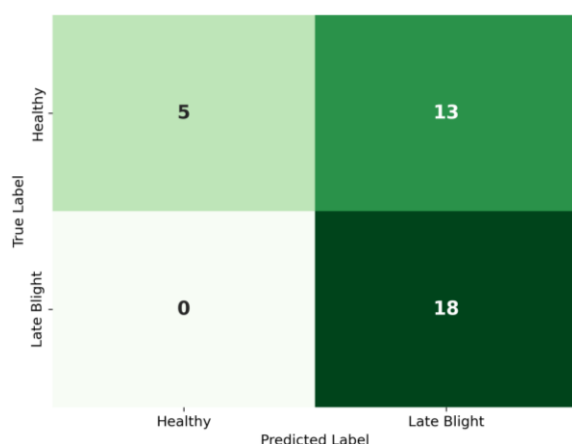


Figure 4. Confusion matrix for the cross-domain evaluation using the 36-image PlantDoc subset.

No false-negative prediction was observed among the 18 late blight images in this limited cross-domain subset. Meanwhile, the high number of false-positive predictions indicated restricted specificity under complex background conditions. These findings supported the intended high-sensitivity screening orientation of the framework but did not establish reliable field diagnosis. Repeated false alarms could increase the verification burden on users and potentially produce alert fatigue. The reduction in accuracy from 83.78% on the PlantVillage test set to 63.89% on the PlantDoc subset demonstrated a substantial domain gap between controlled leaf images and images containing more complex backgrounds and acquisition conditions. The current evidence was limited to the evaluated potato image subsets and did not establish generalization across other crops, cultivars, imaging devices, geographic regions, or concurrent infections.

Limitations and future work

A standardized image-acquisition procedure such as photographing leaves against a high-contrast neutral background may help reduce background interference during future use [18]. Further research should evaluate lightweight foreground-separation methods such as Otsu thresholding [19], complementary chromatic representations such as hue-saturation-value histograms [20], and interaction rules that combine geometric and textural evidence to reduce non-pathological alerts. Integration with environmental measurements including temperature and humidity may also help contextualize visual predictions. Future validation should include physical ESP32 deployment, firmware integration, peak-SRAM profiling, power measurement, on-device latency testing, and prospective evaluation using field-collected images. Overall, the results demonstrated that the LBP-Hu moment-Random Forest framework provided a compact software-level model with high sensitivity to late blight images and a substantially smaller model file than the evaluated quantized deep-learning baseline. The research established a hardware-

aware, storage-constrained, diagnostic information design strategy. The current evidence supported its interpretation as a software-evaluated early-warning framework rather than as a fully validated or field-deployed diagnostic device.

Acknowledgements

This work was supported by the 2025 Qinglan Project of Jiangsu Higher Education Institutions (Outstanding Young Backbone Teacher Cultivation Program).

References

1. Tiwari I, Shah KK, Tripathi S, Modi B, Subedi S, Shrestha J. 2021. Late blight of potato and its management through the application of different fungicides and organic amendments: A review. *J Agric Nat Res*. 4(1):301-320.
2. Kihui E, Ghislain M, Kibe AM, Nancy N, Gatto M, Falck-Zepeda JB. 2025. Ex ante economic impact assessment of the 3R-gene potato in Kenya. *PLoS One*. 20(3):e0309329.
3. He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. *Proc IEEE Conf Comput Vis Pattern Recognit*. 2016:770-778.
4. Anim-Ayeko AO, Schillaci C, Lipani A. 2023. Automatic blight disease detection in potato (*Solanum tuberosum* L.) and tomato (*Solanum lycopersicum*, L. 1753) plants using deep learning. *Smart Agric Technol*. 4:100178.
5. Sapna, Gauttam H, Chauhan V, Pattanaik KK, Trivedi A, Ghosh H. 2026. A comprehensive review of edge computing empowered smart agriculture: Trends, opportunities and future directions. *Comput Electron Agric*. 241:111252.
6. Duhan S, Gulia P, Gill NS, Narwal E. 2025. RTR_Lite_MobileNetV2: A lightweight and efficient model for plant disease detection and classification. *Curr Plant Biol*. 42:100459.
7. Ray PP. 2022. A review on TinyML: State-of-the-art and prospects. *J King Saud Univ Comput Inf Sci*. 34(4):1595-1623.
8. Warden P, Situnayake D. 2019. TinyML: Machine learning with TensorFlow Lite on Arduino and ultra-low-power microcontrollers. Sebastopol: O'Reilly Media. Available at: https://tinymlbook.com/wp-content/uploads/2020/01/tflite_micro_preview.pdf.
9. Shi Y, Han L, Kleerekoper A, Chang S, Hu T. 2022. Novel CropdocNet model for automated potato late blight disease detection from unmanned aerial vehicle-based hyperspectral imagery. *Remote Sens*. 14(2):396.
10. Ojala T, Pietikäinen M, Mäenpää T. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell*. 24(7):971-987.
11. Basavaiah J, Arlene Anthony A. 2020. Tomato leaf disease classification using multiple feature extraction techniques. *Wirel Pers Commun*. 115(1):633-651.
12. Mohanty SP, Hughes DP, Salathé M. 2016. Using deep learning for image-based plant disease detection. *Front Plant Sci*. 7:1419.
13. Espressif Systems. 2025. ESP32-WROOM-32 datasheet. Version 3.6. Available at: https://www.espressif.com/sites/default/files/documentation/esp32-wroom-32_datasheet_en.pdf (accessed January 4, 2026).
14. Hu MK. 1962. Visual pattern recognition by moment invariants. *IRE Trans Inf Theory*. 8(2):179-187.
15. Breiman L. 2001. Random forests. *Mach Learn*. 45(1):5-32.
16. Yang C, Tang R, Feng H. 2023. Application of SVM classifiers in information design: Case study of a new method of classifying exons and introns based on 10-dimensional vectors. *J Biotech Res*. 15:244-252.
17. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. 2018. MobileNetV2: Inverted residuals and linear bottlenecks. *Proc IEEE Conf Comput Vis Pattern Recognit*. 2018:4510-4520.
18. Yang C. 2026. Design and implementation of a computationally efficient framework for tomato disease diagnosis. *J Biotech Res*. 25:65-76.
19. Otsu N. 1979. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern*. 9(1):62-66.
20. Smith AR. 1978. Color gamut transform pairs. *ACM SIGGRAPH Comput Graph*. 12(3):12-19.