Journal of Applied Horticulture, 28(2): 524-532, 2026

Journal of Applied Horticulture
ISSN: 0972-1045

Artificial intelligence-based mango leaf disease detection: current progress, emerging technologies, challenges, and future perspectives

Vaibhav Srivastava[1], M. Srivastava[1]* and Shailendra Rajan[2]*

*[1]Amity Institute of Information Technology, Amity University Uttar Pradesh, Lucknow, India. [2]ICAR-Central Institute for Subtropical Horticulture, Rehmankhera, Lucknow, India. *E-mail: msrivastava@lko.amity.edu, srajanlko@gmail.com*

Abstract

Mango (*Mangifera indica* L.) production is constrained by fungal, bacterial and physiological disorders that reduce yield, fruit quality and market value. Conventional diagnosis depends largely on visual inspection, which is labour-intensive, subjective and farmers are often unable to recognise subtle or early symptoms. Artificial intelligence (AI), machine learning (ML) and deep learning (DL) offer non-destructive alternatives by learning disease-related patterns from images and other sensor data. This review consolidates current work on mango leaf disease detection, tracing the progression from handcrafted colour, texture and shape descriptors to convolutional neural networks (CNNs), transfer learning, hybrid classifiers, vision transformers and explainable AI. Reported studies show strong experimental performance for CNN-SVM systems, ResNet and MobileNet variants, severity classifiers and CNN-transformer models. However, headline accuracies frequently reflect curated datasets and do not necessarily demonstrate reliability in orchards, where illumination, background clutter, leaf occlusion, cultivar differences, mixed disorders and disease stage can alter symptom appearance. The review therefore emphasises dataset quality, independent validation, transparent reporting and deployment efficiency rather than accuracy alone. It also examines mobile and edge AI, thermal and multimodal imaging, severity assessment and integration with precision agriculture. Future research should prioritise geographically diverse datasets, expert-verified annotations, temporal and multimodal learning, self-supervised representation learning, continual model updating, explainable predictions and farmer-centred decision support. Accurate, interpretable and accessible AI tools could support earlier intervention, more targeted agrochemical use and sustainable mango production, but practical impact will depend on interdisciplinary validation and integration with plant pathology and extension services.

Key words: Artificial intelligence, mango disease detection, deep learning, convolutional neural network, transfer learning, vision transformer, explainable AI, precision agriculture, smart farming

Introduction

Mango is one of the most important tropical and subtropical fruit crops and is widely valued for its nutritional, cultural and commercial importance (Rajan, 2021). Its productivity and marketability are nevertheless affected by anthracnose, powdery mildew, bacterial canker, sooty mould, gall midge, leaf blight, scab, dieback and several leaf-spotting disorders. Symptoms can reduce photosynthetic area, weaken plant growth and damage fruit quality, while late recognition increases the likelihood of disease spread and costly intervention. The role of digital decision-support systems (DSS), including web-based phenological data management systems (for climate-sensitive fruit crop mango), has also been emphasized recently (Srivastava *et al*., 2024). Earlier, crop disease research mainly considered automated image analysis as a problem in lab classification but it is now emerging as a feasible precision agriculture tool (Nalawade *et al*., 2023; Pratap and Kumar, 2024; Ramkumar, 2025).

Conventional diagnosis generally involves inspection of colour changes, lesions, fungal growth, deformation and tissue necrosis. Experienced observers can recognise characteristic symptoms, but manual assessment remains subjective and may vary with expertise, fatigue, symptom stage and environmental conditions. Nutrient deficiency, insect damage, dust, senescence and environmental stress can also resemble disease. Delayed or incorrect diagnosis can lead either to untreated infection or to unnecessary chemical use. Automated systems are attractive because they can provide rapid, repeatable screening and can potentially extend diagnostic support to locations where specialists are not immediately available (Anand, 2025; Giri *et al*., 2023).

The technological trajectory of mango disease detection reflects the broader development of computer vision. Early methods segmented suspected lesions and described them through handcrafted colour, texture, edge and shape features. Conventional classifiers such as support vector machines, random forests, gradient boosting and artificial neural networks then mapped those descriptors to disease classes. Deep learning changed this pipeline by allowing models to learn hierarchical representations directly from images. CNNs detect simple structures in shallow layers and progressively combine them into more complex patterns such as lesions, fungal colonies and symptom distributions. Transfer learning further reduced the need to train large networks from the beginning by adapting models pretrained on large image collections to agricultural datasets (Saravanan *et al*., 2023; Thaseentaj and Ilango, 2023).

Current research extends beyond standard CNN classification.

Lightweight MobileNet and quantised EfficientNet models address mobile and edge deployment; CNN-SVM and CNN-random forest systems combine learned representations with classical decision boundaries; thermal imaging seeks physiological evidence that may precede conspicuous visual symptoms; and transformers use attention to model global relationships within an image. Explainable AI techniques such as class-activation mapping are being introduced to show which leaf regions influenced a prediction. Multimodal frameworks also link visual diagnosis with textual explanations or environmental information. These developments indicate a shift from isolated recognition models toward integrated decision-support systems (Zenodo, 2025; Mane *et al*., 2025; Ortiz *et al*., 2025; Srivastava *et al*., 2025).

Despite high reported performance, the central research gap is translation. Many studies use limited or highly curated datasets, random image-level splits and controlled backgrounds. Such conditions can produce strong internal accuracy while failing to test whether a model transfers across orchards, seasons, devices, cultivars or regions. A reliable system must recognise disease under variable light, clutter, occlusion and mixed symptoms, estimate uncertainty, explain its focus and provide management-relevant output. This review therefore consolidates overlapping strands of the literature into four linked questions: how methods have evolved; how datasets and validation shape reported results; which applications are closest to field use; and what technical and institutional changes are required for dependable deployment.

Review methodology

We performed a systematic narrative search in Scopus, Web of Science, Google Scholar, ScienceDirect and SpringerLink for studies published until April, 2026. The search terms included mango leaf disease, artificial intelligence, machine learning, deep learning, CNN, transfer learning, MobileNet, EfficientNet, vision transformer, explainable AI, severity assessment, thermal imaging and edge AI and were combined using AND and OR . Relevant studies on mango leaf disease detection, model evaluation and field deployment were screened and synthesised narratively due to differences in datasets, disease classes and validation methods.

Review scope and evolution of AI approaches: This narrative review synthesises published studies on AI-based mango leaf disease detection and organises them by methodological contribution rather than repeating each paper in multiple sections. The evidence includes conventional ML, CNNs, transfer-learning models, hybrid classifiers, transformer-based systems, severity assessment, mobile deployment, thermal imaging and multimodal decision support. Because the source literature differs substantially in disease classes, dataset size, train-test design and evaluation metrics, reported accuracies are treated as study-specific results rather than as directly comparable rankings. Particular attention is given to whether a study addresses field realism, independent validation, model efficiency, interpretability and actionable use.

Conventional, hybrid and deep-learning approaches: The earliest automated pipelines followed a sequence of acquisition, preprocessing, segmentation, feature extraction and classification. Images were resized or enhanced, backgrounds or lesions were segmented, and descriptors were calculated from RGB, HSV or Lab colour spaces, texture matrices, local binary patterns, histogram of oriented gradients (HOG), Hu moments or morphological measurements. These systems were conceptually transparent and computationally manageable, but their accuracy depended on selecting features that remained stable across lighting, disease stage and cultivar. Giri *et al*. (2023), for example, combined mean-shift segmentation and Hu moment features with gradient boosting to separate powdery mildew and sooty mould. Fuzzy-neural reasoning has also been explored to accommodate ambiguity in symptom appearance (Bolano *et al*., 2024). Such approaches are useful when datasets or hardware are limited, although fixed descriptors are rarely sufficient for the full variation of orchard imagery.

Hybridisation became an intermediate step between handcrafted vision and end-to-end deep learning. CNN-SVM systems use the convolutional network as a feature extractor and the SVM as the final classifier. MangoSpot classified six leaf-spot severity categories and reported 95.68% accuracy, while a mango scab framework reported 92.86% (Baresary *et al*., 2023; Kaur *et al*., 2024). Choudhary *et al*. (2024) paired CNN features with random forest classification on 3,633 images and obtained 80.89% accuracy. The lower result is informative because it shows that hybridisation alone does not guarantee improvement; dataset composition, class overlap, feature quality and validation design remain decisive.

End-to-end CNNs are now the state-of-the-art in the field, as they reduce the need for hand-engineered features. Saravanan *et al*. (2023) used augmentation to increase the diversity available to a CNN. Thaseentaj and Ilango (2023) developed a customised deep CNN for South Indian mango diseases and reported 93.34% accuracy. This work was extended to transfer learning using pretrained networks such as ResNet, VGG, Inception, MobileNet and EfficientNet. ResNet uses residual connections to improve deep architectures , MobileNet uses depth-wise separable convolutions to reduce the computational cost and EfficientNet balances the depth, width and image resolution . Jain and Saini (2024) proposed a seven-class ResNet50 framework for early detection, while Mathur *et al*. (2024) compared multiple pretrained models and observed difficulty distinguishing morphologically similar classes despite high overall performance.

Transformer, multimodal and explainable approaches: The last stage is a combination of local feature learning and global attention, multimodal information and explainability. GLCM texture information for fungal disease detection . Patiño Ortiz *et al*. (2026) used vision-transformer representations and deep classification . In the ViX-MangoEFormer study, convolutional modules, transformer attention and statistical texture features were combined, reporting an F1-score of 99.78% and Matthews correlation coefficient of 99.34% on a merged dataset of 25,530 images (Zenodo, 2025). Srivastava *et al*. (2025) linked visual classification with language-oriented outputs in a framework for anthracnose and sooty mould. These systems illustrate the potential of richer representations, but their complexity also raises questions about reproducibility, computational demand and whether explanatory outputs are biologically valid rather than visually persuasive. The principal architectural trends and research evolution from 2024 to 2026 are summarised in Fig. 1.

Recent mango-leaf studies have increasingly moved from conventional CNNs and transfer-learning architectures toward lightweight models, vision transformers and explainable AI. Representative architectures and publications are summarised in Table 1.

Table 1. AI models and representative studies used for mango leaf disease classification

Combined AI Models	Model Category	Representative Publication	Authors and Year
Custom CNN, AlexNet, ResNet and EfficientNet	CNN and transfer-learning CNNs	Deep Learning-Based Disease Detection in Potato and Mango Leaves: A Comparative Study of CNN, AlexNet, ResNet, and EfficientNet	Mishra *et al.* (2026)
DenseNet121, EfficientNet, MobileNetV2, ShuffleNet, SqueezeNet and Data-Efficient Image Transformer (DeiT)	CNNs, lightweight CNNs and Transformer	Deep Learning for Mango Leaf Disease Identification: A Vision Transformer Perspective	Hossain *et al.* (2024)
MobileNetV3	Lightweight CNN	MobileNetV3 for Mango Leaf Disease Detection: An Efficient Deep Learning Approach for Precision Agriculture	Puranik *et al.* (2024)
VGG19, ResNet50, DenseNet121, EfficientNet-B0, InceptionV3, ConvNeXt, ViT-B/16, DeiT and Swin Transformer	CNNs and Vision Transformers	An Investigation on Advances in Transfer Learning and Explainable AI for Mango Leaf Disease Detection	Prabhu *et al.* (2025)
MangoLeafViT	Lightweight Vision Transformer	MangoLeafViT: Leveraging Lightweight Vision Transformer with Runtime Augmentation for Efficient Mango Leaf Disease Classification	Chowdhury and Ahmed (2025)
DenseNet201, InceptionV3, ResNet152V2, SE-ResNet152 and Xception	Fine-tuned transfer-learning CNNs	Fine-Tuned CNN-Based Approach for Multi-Class Mango Leaf Disease Detection	Ahmmed *et al.* (2025)
VGG16, ResNet50, ResNet101, Xception and Lightweight CNN (LCNN)	Standard and lightweight CNNs	From Code to Field: Evaluating the Robustness of Convolutional Neural Networks for Disease Diagnosis in Mango Leaves	de Andrade *et al.* (2025)
Custom CNN, DenseNet121, EfficientNet-B3, ResNet50 and Xception	CNN and transfer-learning CNNs	Utilizing Deep Learning Models for Early Detection and Classification of Fruit Diseases: Towards Sustainable Agriculture and Enhanced Food Quality	Alrashdi *et al.* (2026)
GourNet	Custom lightweight CNN	GourNet: A CNN-Based Model for Mango Leaf Disease Detection	Alam *et al.* (2026)

Models investigated in more than one publication are retained under each relevant author group to accurately represent the comparative scope of the individual studies.

Datasets, acquisition and validation: Dataset design is the strongest determinant of whether an AI model learns disease biology or merely recognises the circumstances under which images were collected. A useful dataset must represent healthy tissue, relevant diseases, early and advanced stages, cultivar diversity, seasonal conditions, different orchards, varied camera devices and realistic backgrounds. It should also contain confounders such as nutrient deficiency, insect injury, dust, shadows and senescent tissue. Many available datasets are much narrower: they contain detached leaves photographed against uniform surfaces, images captured during one season or data from a single geographical area. Such collections are valuable for method development, but they cannot by themselves establish field robustness.

Dataset quality, annotation and validation: MangoLeafBD is one of the most popular public resources (Ahmed *et al*., 2023). It consists of about 4000 images of high-resolution in various orchards in Bangladesh that have been captured using smart mobile cameras and are composed of healthy images, as well as images of several classes of diseases. However, Puranik *et al*. (2024) retrained MobileNetV3 and achieved nearly 98% accuracy in the context of an Android-based diagnostic system. Similarly, Swapno *et al*. (2024) utilized a dataset of approximately

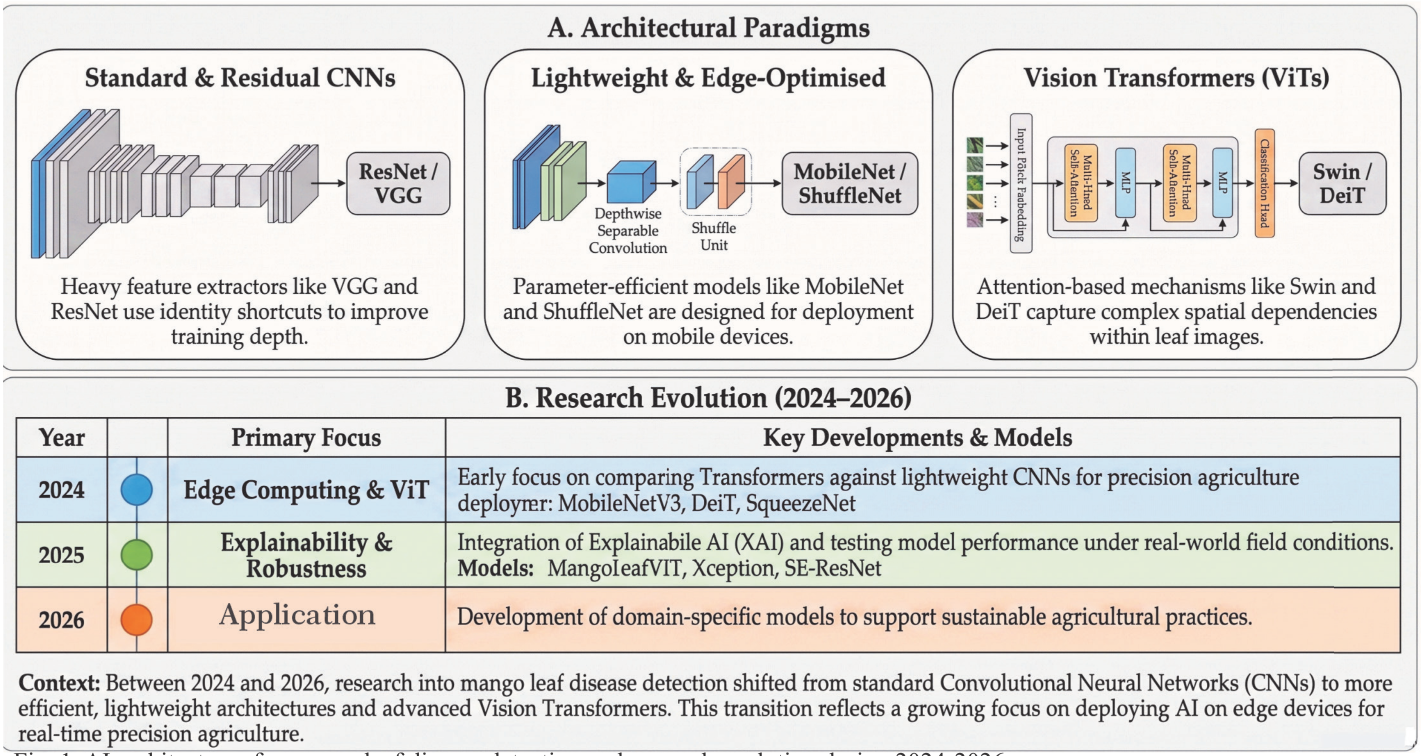

Fig. 1. AI architectures for mango leaf disease detection and research evolution during 2024-2026.

4,000 images categorized into eight classes, achieving training, validation and testing accuracy of 96.87%, 93.12% and 94.99%, respectively. Reproducibility and comparison are supported by making the data public, but using the same data repeatedly may sow the seeds of "benchmark overfitting". Models can become very specialised for a specific image and not show good transfer to other orchards and/or devices.

Multi-source collections are good for representation with larger collections, but come with its own risks. The previous ViX-MangoE model combined four public data sources to access 25,530 images and tested settings of both balanced and imbalanced data (Zenodo, 2025). Combining sources can expose a model to more symptom and imaging variation, yet it can also create source-specific shortcuts. If one disease class is dominated by images from a particular website, background or camera, the network may associate that source pattern with the label. Duplicate and near-duplicate images can leak between training and testing subsets, producing inflated estimates. Dataset audits should therefore include image hashing, similarity screening, source-level documentation and split rules that keep related images together.

Annotation quality is equally important. Mango diseases may share dark lesions, chlorosis, necrosis, curling or fungal coatings, and mixed infections are possible. Labels derived only from visual appearance may be uncertain, particularly at an early stage. Expert verification should record the basis of diagnosis and, where feasible, use laboratory confirmation for ambiguous cases. Severity labels require explicit definitions, such as percentage of affected area or an ordinal scale with reference images. Chhabra (2023), Baresary *et al*. (2023) and Kaur *et al*. (2024) demonstrate the value of severity-oriented datasets, but comparison across studies is difficult when seriousness levels are defined differently. Standardised annotation protocols would make severity models more diagnostically and agronomically meaningful.

Data augmentation is widely used to increase apparent diversity through rotation, flipping, cropping, scaling, brightness change or noise. It can reduce overfitting and teach invariance to harmless transformations, as illustrated by Saravanan *et al*. (2023). However, augmentation cannot replace genuine biological and geographical diversity. Excessive colour transformation may also distort disease cues, while arbitrary flipping or cropping can remove contextual information. Synthetic generation could help rare classes, but generated images should not be treated as independent evidence of field performance. The most useful role of synthetic data is to supplement, not substitute for, expert-labelled orchard observations.

Validation protocols should match the intended claim. Random image-level splitting is acceptable for preliminary model development but may place highly similar photographs from the same leaf, tree or acquisition session in both training and testing sets. Stronger evidence comes from leaf-level, tree-level, orchard-level, seasonal or geographical holdouts. External validation should use data collected independently from model development, ideally with different devices and operators. Reporting should include per-class precision, recall, F1-score, confusion matrices and uncertainty intervals rather than accuracy alone. For imbalanced datasets, balanced accuracy and Matthews correlation coefficient are useful. Calibration is also important because a model used for decisions should attach meaningful confidence to its predictions and recognise when an image is outside its training distribution.

Deep learning architectures and comparative performance: CNNs are well suited to leaf-image analysis because convolution preserves local spatial relationships and shares parameters across the image. Early layers respond to edges, colour transitions and simple textures; deeper layers combine these signals into disease-relevant structures. Pooling or strided convolution reduces spatial dimensions, and a classification head maps the learned representation to disease labels. In a well-designed pipeline, the network is trained with augmentation and regularisation, evaluated on unseen data and inspected to determine whether attention falls on symptomatic tissue. The principal limitation is that local pattern recognition can be vulnerable to background shortcuts and may not fully capture long-range relationships across a large leaf.

CNN, transfer learning and lightweight models: Custom CNNs allow architecture size and receptive field to be adapted to a specific dataset. Thaseentaj and Ilango (2023) reported 93.34% accuracy for anthracnose, powdery mildew and leaf blight using a customised network. Saravanan *et al*. (2023) similarly showed that a CNN combined with augmentation can support automated mango disease prediction. Custom models may be smaller and easier to deploy than large pretrained networks, but they require careful tuning and can overfit limited data. Their value is best assessed against strong transfer-learning baselines under the same split and preprocessing conditions.

Transfer learning is attractive because mango datasets are small relative to general computer-vision corpora. A pretrained model begins with filters that already encode edges, shapes and textures, and fine-tuning adapts higher-level features to disease symptoms. ResNet50 is frequently used because residual connections stabilise deep optimisation. Jain and Saini (2024) proposed ResNet50 for seven classes and targeted approximately 95% accuracy. Mathur *et al*. (2024) compared InceptionV3, MobileNetV3 variants and ResNet50, with InceptionV3 showing strong classification capability. Their observation that healthy leaves, gall midge and anthracnose remained difficult to separate is more important than a single average score because it identifies the classes most likely to cause practical errors.

MobileNet prioritises efficiency through depthwise separable convolution, making it suitable for smartphones and embedded devices. Puranik *et al*. (2024) reported approximately 98% accuracy with MobileNetV3 and integrated the model into an Android workflow. The result demonstrates feasibility, but deployment quality depends on more than the trained network. Camera guidance, image-quality checks, inference latency, offline capability, confidence thresholds and a safe response to uncertain images are essential. A mobile system should not present a low-confidence prediction as definitive; it should request a clearer photograph or recommend expert consultation.

EfficientNet uses compound scaling to balance network depth, width and input resolution. Mane *et al*. (2025) applied EfficientNet-B0 to mango fruit diseases and evaluated Float32, Float16 and INT8 quantisation. The INT8 version reduced model size by about 70% and inference time by about 37%, showing how post-training optimisation can support edge deployment. Although fruit and leaf diagnosis are not identical tasks, the deployment lesson is directly relevant: practical systems should evaluate the accuracy-size-latency trade-off on the target device rather than assuming that a high-performing server model will remain usable after compression.

Hybrid models, severity assessment and alternative sensing: Hybrid CNN-classifier systems attempt to combine representation learning with robust classical decision boundaries. MangoSpot used CNN-derived features and an SVM to classify six leaf-spot severity levels, reporting 95.68% accuracy (Baresary *et al*., 2023). Kaur *et al*. (2024) reported 92.86% for a CNN-SVM scab severity model. Choudhary *et al*. (2024) used CNN features with a random forest and obtained 80.89% overall accuracy. These studies show that SVM or random forest heads may be effective when learned embeddings are discriminative, but they add an additional tuning stage and do not automatically solve class overlap. Comparisons should include an end-to-end softmax head and should use identical splits to determine whether the hybrid component provides a genuine gain.

Severity assessment extends classification from naming a disease to estimating its extent. Chhabra (2023) trained ResNeXt50 on 25,000 images for sooty mould severity and reported 94.61% accuracy. Severity labels can better support decisions because treatment urgency depends on progression, yet discretising a continuous process into classes can conceal uncertainty near category boundaries. Segmentation or regression may be preferable where affected area can be estimated directly. A useful system could report both a disease probability and an estimated percentage of affected tissue, with confidence intervals and a visual overlay showing the measured region.

Alternative sensing can reveal information unavailable in RGB images. Nalawade *et al*. (2023) combined CNN analysis of visible images with ANN analysis of thermal data for anthracnose and powdery mildew. Thermal patterns may reflect physiological stress before extensive visible damage, but temperature is influenced by sunlight, wind, water status and acquisition geometry. Reliable use therefore requires calibration and contextual data. The broader implication is that early detection may depend on multimodal evidence rather than increasingly complex processing of a single RGB photograph.

Transformers model relationships between image patches through self-attention and can capture global context. Patiño Ortiz *et al*. (2026) combined vision-transformer features, GLCM texture descriptors and deep classification for fungal disease recognition. The ViX-MangoEFormer framework similarly integrated convolutional modules, attention and statistical texture features, with very high reported F1 and MCC values (Zenodo, 2025). Hybrid CNN-transformer designs are attractive because convolution provides local inductive bias while attention links distant symptom regions. Nevertheless, these architectures can be data-hungry and computationally expensive. Their advantage should be demonstrated through external validation and ablation studies showing which components contribute to performance.

Explainability is increasingly treated as part of model evaluation. Gradient-weighted class activation mapping can highlight the areas that most influenced a prediction. If the model consistently focuses on lesions, fungal growth or necrotic margins, the visualisation provides some evidence that classification is based on plausible cues. If it focuses on a background, label marker or image border, the explanation reveals shortcut learning. The ViX-MangoEFormer work incorporated such visual explanations, and Srivastava *et al*. (2025) moved further by coupling visual diagnosis with natural-language output. However, saliency maps are not proof of causality; they should be reviewed by plant pathologists and assessed for stability across small image changes.

Across the literature, reported accuracy ranges from about 80% to almost 100%. This spread does not establish a simple hierarchy of algorithms because the studies differ in disease number, image quality, dataset size, class balance and test design. A 95% result on a difficult external orchard set may be more valuable than 99% on a random split of curated images. Comparative research should therefore use common benchmarks, source-separated validation and transparent preprocessing. Model selection for field use should consider sensitivity to serious diseases, false-alarm cost, inference speed, calibration, memory use and interpretability in addition to average accuracy.

Applications in mango disease management: The most mature application is image based classification of healthy and diseased leaves. Multi-class models can reduce the time between observation and preliminary diagnosis, especially in cases where extension personnel cover large areas. VGG16, MobileNet, GoogLeNet, YOLOv8 and EfficientNet architectures were assessed for multiple mango leaf diseases by Pratap and Kumar (2024), whereas Swapno *et al*. (2024) demonstrated eight class classification on a public dataset. Classification is useful for screening, but deployment should clearly differentiate between preliminary AI advice and confirmed diagnosis. A model should also have an unknown/referral pathway for symptoms not in the training classes. Severity assessment may help in more proportionate management.

Field deployment and decision support Detecting binary disease does not tell you whether a few lesions should be watched, or whether a widespread infection calls for urgent action. Chhabra (2023), Baresary *et al*. (2023) and Kaur *et al*. (2024) severity models show movement toward management-relevant output. Future systems should provide linkages between severity estimates and crop stage, weather and locally approved control guidance. Integration may reduce blanket spraying of routine interventions, targeting intervention to disease type, affected area and risk of progression . Mobile and edge AI are especially relevant for smallholders and remote orchards. Puranik *et al*. (2024) demonstrated that a lightweight MobileNetV3 model can be integrated into a smartphone workflow, and Mane *et al*. (2025) showed the benefits of quantisation for resource-constrained devices.

Challenges and limitations: Local inference reduces reliance on continuous internet access and can improve privacy and response time. A robust application should guide image capture, detect blur and poor illumination, store anonymised feedback with consent, support local languages and provide clear uncertainty messages. Offline operation is valuable, but periodic model updating is still needed as new cultivars, symptoms and devices are encountered.

AI can also support monitoring at orchard scale. Repeated smartphone observations, fixed cameras or drone imagery could map disease distribution and identify emerging hotspots. Close-range leaf models cannot be transferred directly to aerial imagery because resolution, scale and background differ, but the same principles of representation learning and validation apply. Combining spatial disease maps with weather, irrigation and management records could support targeted scouting and treatment. The objective should be to direct human attention more efficiently, not to eliminate expert observation.

Decision support is the logical next step beyond classification. Srivastava *et al*. (2025) combined visual analysis with language-oriented responses, illustrating how a system might explain symptoms and management options. The value of such output depends on the quality and locality of the knowledge base. Recommendations must reflect the diagnosed disease, severity, crop stage, local regulations and resistance-management principles. An AI model that correctly labels a leaf but gives inappropriate treatment advice could cause more harm than a classifier that simply recommends expert review. Therefore, diagnostic and advisory components should be validated separately.

Challenges and limitations: Generalisation remains the most important unresolved problem. Disease appearance changes with cultivar, leaf age, pathogen strain, climate, camera, lighting and background. A model trained in one region may perform poorly elsewhere even when the same disease is present. Public datasets such as MangoLeafBD have accelerated research, but regional concentration limits the breadth of evidence (Puranik *et al*., 2024). Cross-region studies should train on multiple locations and reserve entire unseen locations for testing. Performance should then be reported separately for each orchard, cultivar and device to reveal where the model fails.

Generalisation, evaluation and practical limitations: Visual similarity and biological ambiguity create a second challenge. Anthracnose, leaf spots, scab and some insect or nutrient injuries can share dark or chlorotic patterns, especially early in development. Mathur *et al*. (2024) noted confusion among healthy, gall midge and anthracnose classes despite strong model performance. Mixed infection further complicates single-label classification. Multi-label models, lesion segmentation and hierarchical diagnosis may better reflect reality. At minimum, datasets should include common non-disease confounders and ambiguous samples rather than discarding them during curation.

Early detection is difficult because the visible signal may be weak. RGB models are limited when pathophysiological change precedes obvious discoloration or lesion development. Thermal, hyperspectral or multispectral sensing can add information, as suggested by Nalawade *et al*. (2023), but each modality introduces cost, calibration and environmental sensitivity. Research should identify whether the additional sensor materially improves lead time and decision quality, not merely classification accuracy under controlled conditions.

Evaluation practices frequently overstate readiness. Random splits can allow near-duplicate images or images from the same leaf to appear in both training and testing sets. Class imbalance may be hidden by overall accuracy, and hyperparameter selection may indirectly use the test set. Small studies can also report unstable results without confidence intervals. Strong reporting should document data sources, duplicate removal, split units, preprocessing, augmentation, class distribution and all metrics. Independent replication is especially important for results approaching 100%, because tiny leakage or source bias can have a large effect on headline performance.

Model complexity and deployment requirements create practical trade-offs. Transformers and ensembles may provide richer representations but require memory and processing capacity that are unavailable on low-cost devices. Lightweight networks and quantisation help, but compression can reduce sensitivity to subtle classes. Device-level tests should measure latency, battery consumption, memory use and accuracy after conversion. Cloud processing can support large models but depends on connectivity, incurs operating cost and raises data-governance questions. The appropriate architecture is therefore determined by the deployment setting, not by benchmark performance alone.

Explainability and trust require more than attractive heat maps. A saliency method may identify the general area of interest without explaining the specific pathological feature or differentiating two visually similar conditions. Explanations can also vary across methods. Plant pathologists should assess whether highlighted regions correspond to plausible symptoms, and user studies should test whether explanations improve decisions rather than merely confidence. Systems should communicate uncertainty and limitations plainly. A calibrated message such as ‘possible anthracnose; image quality moderate; expert confirmation recommended’ is safer than an unqualified label.

Actionability is another limitation. Many papers end with a disease name and do not connect the prediction to safe management. Farmers need information about urgency, monitoring, cultural controls and when professional or laboratory confirmation is required. Recommendations vary across countries and may change with pesticide registration or resistance guidance. Consequently, the advisory layer should be modular, locally governed and regularly updated. Image models should not embed static treatment advice that becomes obsolete.

Economic and social barriers can prevent technically successful systems from being adopted. Smartphone quality, connectivity, digital literacy, language and trust in institutions vary among farming communities. Applications designed without farmer participation may have complex interfaces, require unrealistic image capture or provide outputs that do not match decision needs. Participatory design should involve growers, extension workers and pathologists from the beginning. Performance should be evaluated not only by classification metrics but also by successful image capture, comprehension, time saved, appropriate referrals and changes in disease-management outcomes.

Finally, current evidence is fragmented. Studies use different disease names, severity scales, architectures and evaluation protocols, making synthesis difficult. Reference lists sometimes contain duplicate or incomplete records, while datasets may have unclear provenance. The field would benefit from shared reporting standards, dataset cards, model cards and preregistered evaluation plans. These measures would make claims easier to verify and would shift attention from isolated accuracy records to reproducible agricultural value. The main domain-shift, deployment and architecture trade-offs are summarised in Fig. 2.

Framework for field-ready evaluation: A field-ready evaluation should begin by defining the intended use before selecting an architecture. A system designed for preliminary farmer screening has different requirements from a tool used by extension officers or a model that triggers automated spraying. The target population, disease classes, image devices, operating environment and acceptable error rates should be specified in advance. For high-consequence diseases, sensitivity may be prioritised, whereas a pesticide recommendation system must also control false positives. The development dataset should then be sampled to reflect the intended use, including difficult images and realistic confounders rather than only clear textbook symptoms.

Technical, prospective and agronomic validation: Technical validation should proceed through increasingly demanding data splits. An initial internal split can support model selection, but related photographs must remain in the same partition. A second test should hold out complete trees, acquisition sessions

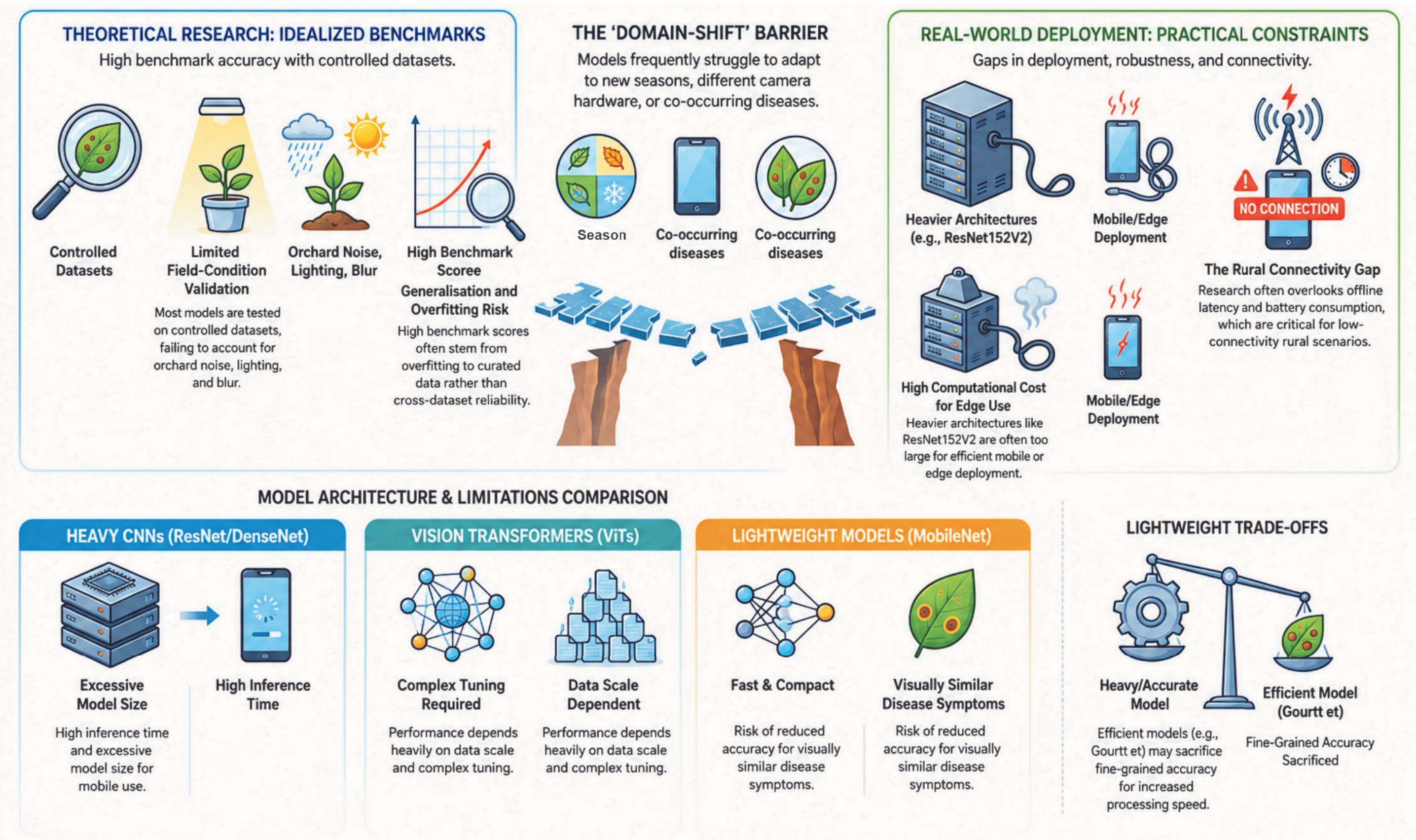

Fig. 2. Principal challenges in AI-driven plant disease detection, including domain shift, deployment constraints and model trade-offs.

or orchards. A third should use an external region, season or device that was not represented during training. Results should be reported by disease class and subgroup, with confidence intervals, calibration curves and an analysis of common errors. An abstention option should be evaluated so that the model can decline uncertain or out-of-distribution images instead of forcing a potentially misleading label.

Prospective orchard evaluation is the next requirement. Farmers or extension workers should capture images using ordinary devices and without laboratory-style supervision. The study should record failed image attempts, blur, poor lighting, internet limitations, inference time and user interpretation of the output. Predictions should be compared with an expert reference diagnosis, and uncertain cases should undergo additional confirmation where feasible. This design tests the complete workflow rather than only the classifier. A technically accurate model may still fail if users cannot obtain a suitable image or misunderstand confidence and referral messages.

Agronomic usefulness should be assessed separately from diagnostic accuracy. Relevant outcomes include time to detection, proportion of appropriate referrals, change in scouting effort, treatment timing, pesticide use, disease progression, yield, fruit quality and cost. A controlled field trial could compare standard practice with AI-assisted management while ensuring that both groups retain access to expert care. Such studies would show whether earlier or more consistent identification produces measurable benefits. They would also reveal unintended effects, such as over-treatment caused by false alarms or reduced scouting because users place excessive trust in the application.

Post-deployment monitoring is necessary because data and operating conditions change. New phone models, cultivars, pathogens, seasons and user behaviours can shift the distribution of incoming images. Applications should log performance indicators with appropriate consent, provide a mechanism for expert correction and maintain versioned model documentation. Updates should be tested against a fixed safety set and external datasets before release. Clear responsibility is needed for correcting harmful advice, withdrawing a poorly performing version and communicating changes to users. These governance measures turn a research prototype into an accountable agricultural service.

A minimum reporting checklist would make field studies easier to interpret and reproduce. Authors should identify the number of unique leaves, trees, orchards and acquisition sessions rather than reporting images alone; describe who assigned labels and how uncertain cases were handled; state whether duplicate and near-duplicate images were removed; publish the exact train, validation and test rules; and report class-specific results with confidence intervals. Model size, parameter count, inference time and target hardware should accompany accuracy metrics. For deployed systems, reports should describe image-quality controls, offline behaviour, data storage, language support, referral messages and update procedures. Publishing code, model weights and a representative external test set where ethical and legally permissible would enable independent verification. A consistent checklist would reduce accidental leakage, discourage selective reporting and allow future reviews to compare methods on evidence that reflects practical reliability rather than isolated headline scores.

Future perspectives and research priorities: Future progress should begin with coordinated data infrastructure. International datasets should include multiple mango-producing regions, cultivars, seasons, devices and disease stages. Images should be linked to expert diagnosis, severity, date, location at an appropriate privacy level, weather and management context. A common label ontology is needed so that disease names and severity categories are consistent. Benchmark partitions should include an internal test set and at least one external region or season. Open data should be encouraged, while governance must respect farmer consent, ownership and commercial sensitivity.

Data-intensive and adaptive learning: Multimodal and temporal learning can move diagnosis from visible symptom recognition toward risk prediction. Images can be combined with temperature, humidity, rainfall, leaf wetness, soil moisture and crop stage because disease development depends on both pathogen presence and environmental suitability. Thermal or spectral sensing may detect physiological stress earlier than RGB photography. Repeated observations of the same trees can reveal progression and help distinguish transient stress from expanding infection. Models should be evaluated on how early and accurately they predict an actionable event, not only on single-image class accuracy.

Self-supervised and semi-supervised learning are promising because large numbers of orchard images can be collected more easily than they can be expertly labelled. A model can first learn general mango-leaf representations from unlabelled data and then be fine-tuned on a smaller verified set. Active learning can prioritise uncertain or unusual images for expert review, making annotation more efficient. Synthetic data may help balance rare classes, but generated examples should be validated for biological plausibility and should never replace external testing on real field images.

Continual and federated learning could help models adapt after deployment. Continual learning allows a system to incorporate newly verified cases without forgetting earlier classes, while federated approaches can update a shared model using data stored locally at institutions or devices. These strategies may improve regional adaptation and privacy, but they require safeguards against label noise, device bias and malicious or erroneous updates. Version control and post-update validation are essential so that performance improvements in one region do not create regressions elsewhere.

Efficient, explainable and integrated systems: Efficient edge AI will remain central to accessibility. MobileNet-type architectures, pruning, knowledge distillation and quantisation should be compared on representative low-cost phones, not only on desktop hardware. The best model may be a cascade: a small on-device network performs image-quality assessment and preliminary screening, while uncertain cases are optionally sent to a more capable server or extension service. Such a design preserves offline usefulness while reserving expensive computation for difficult cases. Puranik *et al*. (2024) and Mane *et al*. (2025) provide practical foundations for this direction.

Explainable AI should evolve from generic saliency to evidence that is meaningful for plant pathology. Future systems could segment lesions, quantify affected area, describe observed features and show comparable verified examples. Explanations should be evaluated for localisation accuracy, stability and usefulness to farmers and experts. Multimodal assistants, like the direction explored by Srivastava *et al*. (2025), may translate predictions into local-language guidance, but language generation must be grounded in validated agricultural knowledge and constrained to avoid unsupported treatment claims.

Integration with precision agriculture can broaden impact. Disease predictions can be combined with IoT sensors, scouting records, drone or satellite observations and farm-management systems to create spatial risk maps and targeted alerts. The aim is not continuous technology use for its own sake; it is better timing and placement of intervention. Field trials should compare AI-assisted management with standard practice using outcomes such as diagnostic lead time, disease incidence, pesticide use, yield, cost and farmer workload. Environmental benefits should be measured rather than assumed.

A staged validation pathway would accelerate responsible translation. Stage one should establish technical validity on diverse, source-separated datasets. Stage two should test usability and robustness in prospective orchard studies with farmers and extension staff. Stage three should evaluate whether recommendations improve management and economic outcomes. Stage four should support post-deployment monitoring, model updates and incident reporting. At every stage, uncertainty, failure modes and referral criteria should be documented. This pathway would help distinguish promising prototypes from dependable agricultural tools.

Interdisciplinary collaboration is indispensable. Computer scientists contribute model design and evaluation; plant pathologists establish diagnostic validity; horticulturists and extension workers define actionable outputs; social scientists examine trust and adoption; and farmers identify practical constraints. Funders and journals can encourage this integration by requiring external validation, transparent data documentation and realistic deployment metrics. The future of mango disease AI will depend less on marginal increases in benchmark accuracy and more on whether systems are reproducible, interpretable, affordable and embedded in trustworthy support networks.

AI-based mango disease detection has progressed from segmented images and handcrafted descriptors to deep CNNs, transfer learning, hybrid classifiers, transformers, multimodal sensing and explainable decision support. The reviewed studies demonstrate that automated systems can classify several mango diseases, estimate severity and operate within mobile or edge-oriented workflows. CNN-SVM severity models, MobileNetV3 deployment, quantised EfficientNet and CNN-transformer frameworks illustrate the breadth of current innovation (Baresary *et al*., 2023; Puranik *et al*., 2024; Mane *et al*., 2025; Zenodo, 2025).

The evidence also shows why laboratory accuracy should not be equated with field readiness. Dataset duplication, restricted geography, image-level splitting, symptom similarity and uncontrolled environmental variation can all weaken generalisation. Practical systems must recognise uncertainty, withstand changes in orchard conditions and devices, and connect diagnosis to locally valid management without overclaiming. External validation, expert-reviewed explanations, model calibration and farmer-centred design are therefore essential.

Future research should build diverse and well-documented datasets, exploit multimodal and temporal information, develop efficient models for low-cost devices and establish continual updating under clear governance. Integration with weather, sensors, remote sensing and extension services could support earlier and more targeted intervention. When validated through prospective field trials, accurate and accessible AI tools have the potential to reduce diagnostic delay, support more precise chemical use and strengthen sustainable mango production. Their success, however, will depend on collaboration among AI researchers, plant pathologists, horticulturists, extension systems and farming communities.

References

Ahmed, S.I., M. Ibrahim, M. Nadim, M.M. Rahman, M.M. Shejunti, T. Jabid and M.S. Ali, 2023. MangoLeafBD: A comprehensive image dataset to classify diseased and healthy mango leaves. Data Brief, 47: 108941. doi:10.1016/j.dib.2023.108941.

Ahmmed, J., F. Ahmed, R.H. Shohan, M.M. Rana and M. Hasan, 2025. Fine-tuned CNN-based approach for multi-class mango leaf disease detection. arXiv preprint, arXiv:2510.05326. doi:10.48550/arXiv.2510.05326.

Alam, E., J. Sanyal, A.K. Das, A. Bhatyacharya and F. Sultana, 2026. GourNet: A CNN-based model for mango leaf disease detection. p. 269-280. In: Proceedings of the Second International Conference on Advanced Computing and Systems (AdComSys 2025), W. Ding, A. Chakrabarti, M. Chakraborty and S. Chakraborty (eds.). Lecture Notes in Networks and Systems, Vol. 1887. Springer, Cham. doi:10.1007/978-3-032-20253-6_21.

Alrashdi, I., M. Sharawi, A.M. Ali, K.M. Sallam, B. Arain and M. Abdel-Basset, 2026. Utilizing deep learning models for early detection and classification of fruit diseases: Towards sustainable agriculture and enhanced food quality. Sci. Rep., 16: 8167. doi:10.1038/s41598-026-38259-3.

Anand, A., 2025. Leaf disease detection system through advanced AI and machine learning integration. SSRN Electron. J. doi:10.2139/ssrn.5018301.

Baresary, D., A. Saini, R.K. Sharma, D. Bordoloi, R. Sharma and V.K. Kukreja, 2023. MangoSpot: A hybrid CNN-SVM model for multi-classification of mango leaf spot disease based on seriousness levels. In: Proceedings of the International Conference on Intelligent Technologies (CONIT). doi:10.1109/CONIT59222.2023.10205376.

Bolano, J.I.S., B.P.J.E. Pilares and A.N. Yumang, 2024. Mango leaf disease identification using a fuzzy neural approach. In: Proceedings of ELTICOM. doi:10.1109/ELTICOM64085.2024.10865311.

Chhabra, D., 2023. Mango leaf disease severity scale classification using ResNext50 model: A deep learning-based approach. In: Proceedings of ICTACS. doi:10.1109/ICTACS59847.2023.10389844.

Choudhary, S., M. Choudhary, S. Kaur and V.K. Kukreja, 2024. Integrating CNN and random forest for accurate classification of mango leaf diseases. In: Proceedings of AUTOCOM. doi:10.1109/AUTOCOM60220.2024.10486108.

Chowdhury, R.H. and S. Ahmed, 2025. MangoLeafViT: Leveraging lightweight vision transformer with runtime augmentation for efficient mango leaf disease classification. arXiv preprint, arXiv:2505.23961. doi:10.48550/arXiv.2505.23961.

de Andrade, G.V., S.R. dos Santos, I.P.C.A. da Silva, E.A.M. Pereira and E.A. Barboza, 2026. From code to field: Evaluating the robustness of convolutional neural networks for disease diagnosis in mango leaves. p. 291-305. In: Intelligent Systems: 35th Brazilian Conference, BRACIS 2025, Proceedings, Part IV, R. de Freitas and D. Furtado (eds.). Lecture Notes in Computer Science, Vol. 16182. Springer, Cham. doi:10.1007/978-3-032-15993-9_20.

Giri, G.A.V.M., I.A. Musdar, H. Angriani and M. Taruk, 2023. Enhancing disease management in mango cultivation: A machine learning approach to classifying leaf diseases. Indones. J. Data Sci., 4(3): 160–168. doi:10.56705/ijodas.v4i3.111.

Hossain, M.A., S. Sakib, H.M. Abdullah and S.E. Arman, 2024. Deep learning for mango leaf disease identification: A vision transformer perspective. Heliyon, 10(17): e36361. doi:10.1016/j.heliyon.2024.e36361.

Jain, E. and R. Saini, 2024. ResNet50-based deep learning model for early detection and classification of mango leaf diseases. In: Proceedings of ICCES. doi:10.1109/ICCES63552.2024.10859342.

Kaur, A., V.K. Kukreja, D.S. Rana, A. Garg and R. Sharma, 2024. AI-driven strategies for identifying and classifying mango scab disease. In: Proceedings of IITCEE. doi:10.1109/IITCEE59897.2024.10467992.

Mane, D., H. Nandane, A. More, R. More, T. Mali and M. Jadhao, 2025. Multistage deep learning-based mango fruit disease detection and deployment using quantized models. In: Proceedings of AIC. doi:10.1109/AIC66080.2025.11212077.

Mathur, P., F. Sheth, D. Goyal and A.K. Gupta, 2024. Deep insight: Mathematical modeling and statistical analysis for mango leaf disease classification using advanced deep learning models. J. Interdiscip. Math., 27(2): 317–342. doi:10.47974/jim-1830.

Mishra, U., A. Pandey, G. Logeswari and K. Tamilarasi, 2026. Deep learning-based disease detection in potato and mango leaves: A comparative study of CNN, AlexNet, ResNet, and EfficientNet. Sci. Rep., 16: 2788. doi:10.1038/s41598-025-32607-5.

Nalawade, R.R., S.D. Sawant, M.S. Joshi, P.M. Ingle, V.G. More and J.J. Kadam, 2023. Mango anthracnose and powdery mildew disease detection using convolutional neural network and artificial neural network. J. Plant Dis. Sci., 18(1): 11–19. doi:10.48165/jpds.2023.1801.03.

Patiño Ortiz, J., R. Carreño Aguilera, S. Juárez Vázquez, M. Patiño Ortiz and M.A. Chávez Benítez, 2026. Early detection of fungal diseases in mango leaves: Fusion of GLCM features and vision transformers with deep neural network classification. Fractals, 34(1): 2650004. doi:10.1142/S0218348X26500040.

Prabhu, O., M. T and S. Shetty, 2025. An investigation on advances in transfer learning and explainable AI for mango leaf disease detection. Smart Agric. Technol., 12: 101348. doi:10.1016/j.atech.2025.101348.

Pratap, V.K. and N.S. Kumar, 2024. Deep learning-based mango leaf disease detection for classifying and evaluating mango leaf diseases. Fusion: Pract. Appl., 15(2): 261–277. doi:10.54216/fpa.150222.

Puranik, S.S., S.R. Hanamakkanavar, A.P. Bidargaddi, V.V. Ballur, P.T. Joshi, S.M. Meena and U. Kulkarni, 2024. MobileNetV3 for mango leaf disease detection: An efficient deep learning approach for precision agriculture. In: 2024 5th International Conference for Emerging Technology (INCET), IEEE, Belgaum, India, p. 1-7. doi:10.1109/INCET61516.2024.10593318.

Rajan, S., 2021. Mango: The King of Fruits. p. 1-11. In: The Mango Genome, C. Kole (ed.). Compendium of Plant Genomes. Springer, Cham. doi:10.1007/978-3-030-47829-2_1.

Ramkumar, G., 2025. Early identification of crop leaf disease detection system through artificial intelligence (AI)-assisted deep learning methodology. In: Proceedings of ICESC. doi:10.1109/ICESC65114.2025.11212217.

Saravanan, T.M., M. Jagadeesan, P.A. Selvaraj, M. Aravind, G.D. Raj and P. Lokesh, 2023. Prediction of mango leaf diseases using convolutional neural network. p. 1-4. In: Proceedings of the International Conference on Computer Communication and Informatics (ICCCI). doi:10.1109/ICCCI56745.2023.10128578.

Srivastava, V., M. Srivastava, S. Rajan, P. Sagar, P.K. Mishra and V.K. Singh, 2024. Leveraging web-based tool for phenological data management in climate-sensitive fruit crops like mango. J. Appl. Hortic., 26(1): 112-116. doi:10.37855/jah.2024.v26i01.21.

Srivastava, V., M. Srivastava and S. Rajan, 2025. Deep learning and GCC-MA-based visual system for mango leaf disease analysis with anthracnose and sooty mold as case study. J. Appl. Hortic., 27(3): 553-558. doi:10.37855/jah.2025.v27i03.100.

Swapno, S.M.M.R., S.M.N. Nobel, M.B. Islam, R. Haque, V.P. Meena and F. Benedetto, 2024. A novel machine learning approach for fast and efficient detection of mango leaf diseases. In: Proceedings of ICMI. doi:10.1109/ICMI60790.2024.10585939.

Thaseentaj, S. and S. Ilango, 2023. Deep convolutional neural networks for South Indian mango leaf disease detection and classification. Comput. Mater. Contin. doi:10.32604/cmc.2023.042496.

Zenodo, 2025. Interpretable mango leaf disease detection using a hybrid CNN-transformer model with GLCM features. Zenodo Repository. doi:10.5281/zenodo.17173342.