

A Comparative Study of Deep Learning Architectures for Plant Leaf Disease Detection: A Systematic Review

¹Atul Raj Dhairya, ²Prof. Nagendra Patel

¹M.Tech. Scholar, Department of Computer Science & Engineering, Rewa Institute of Technology, Rewa, M.P., India
atulrajdhairya@gmail.com

²Assistant Professor (Guide), Department of Computer Science & Engineering, Rewa Institute of Technology, Rewa, M.P., India
nagendra.nriist@gmail.com

Abstract—

Agriculture remains a primary source of livelihood for a large share of the Indian population and is essential to food security and national economic development. Variations in climate, soil condition and local environment increase the risk of crop disease, and because the earliest symptoms of most infections appear on the leaf, an infection that is not identified in time spreads through the plant and across the field. Manual inspection by trained experts is slow, costly, inconsistent and impractical over large cultivation areas, so automated detection has become a necessity rather than a convenience. This paper reviews the development of automated plant leaf disease detection from classical image processing and machine learning through deep convolutional architectures to the vision transformer, hybrid and deployment-oriented models reported up to 2026. Thirty-six representative studies are organised into a four-family taxonomy and compared in terms of technique, dataset, crop coverage and reported accuracy. A generic five-stage detection pipeline — image acquisition, pre-processing and augmentation, segmentation, feature extraction and selection, and classification — is presented together with a model architecture in which convolutional, transfer-learning and transformer backbones are interchangeable. The review shows that reported accuracy on the standard benchmark has risen from the high eighties for early convolutional and image-processing systems to above ninety-nine per cent for recent transformer and lightweight convolutional designs, but that this progress is measured almost entirely under laboratory conditions. Persistent gaps are identified in field-condition robustness, dataset diversity, class imbalance, early-stage and multi-disease detection, computational cost at the edge, and interpretability, and these gaps are consolidated into a set of directions for future work.

Keywords—deep learning, plant disease identification, plant health assessment, convolutional neural network, vision transformer, precision agriculture, image classification, transfer learning.

I. INTRODUCTION

Agriculture is essential not only to human survival but also to national growth and economic stability. In India it employs a very large share of the working population and contributes a substantial proportion of gross domestic product. As the global population continues to rise while the number of active farmers gradually declines, maintaining and improving agricultural productivity has become central to food security and to sustained economic development [1], [2].

Crops are frequently affected by disease arising from variation in weather, soil condition and local environmental factors. Such diseases retard plant growth, cause excessive leaf drop, lower product quality and ultimately reduce yield, and infections spread rapidly through seeds, insects and microbial agents. Fungi, bacteria, viruses and pests are the common causes. Timely and accurate identification is therefore critical to protecting both the quantity and the quality of produce; where disease is not detected and managed early, it can damage entire cultivation regions and contribute to food shortage and price inflation [3].

Before the widespread adoption of information technology, disease detection relied on manual inspection with the naked eye. That approach requires experienced plant pathologists, repeated field visits and continuous monitoring, which makes it costly and inefficient at scale [4]. Many cultivators lack the expertise to distinguish one pathogen from another, and inconsistent diagnosis leads to incorrect pesticide application, which reduces productivity and raises cost without controlling the infection. Manual monitoring is also labour-intensive and simply impractical across large agricultural regions.

Automated detection addresses these limitations directly. Most automated systems operate on visible leaf symptoms — single or multi-coloured spots, abnormal colour change, deformation of leaf shape, early or late blight patterns, and brown or yellow lesions — because the leaf carries the earliest and most legible evidence of infection. Over the last decade the underlying technique has moved through three clear generations: hand-engineered colour and texture descriptors fed to a classical classifier; deep convolutional networks that learn the descriptors themselves; and, most recently, transformer and hybrid architectures that model global context across the whole image [5], [6].

This paper contributes the following. First, it consolidates the literature into a four-family taxonomy that separates classical image processing, deep convolutional architectures, transformer and hybrid models, and deployment-oriented models, and situates each surveyed work within it. Second, it presents a generic five-stage detection pipeline and a corresponding model architecture in which the classification backbone is interchangeable, so that the contribution of each stage can be reasoned about independently. Third, it compares reported accuracy across the surveyed studies and shows how the benchmark figure has evolved. Fourth, it identifies the gaps that remain open once laboratory accuracy has saturated, and states the research directions that follow from them.

The remainder of the paper is organised as follows. Section II presents the taxonomy. Section III reviews the literature. Section IV describes the detection methodology. Section V presents the model architecture and the backbone families. Section VI compares reported

performance. Section VII states the research gaps and expected outcomes, and Section VIII concludes.

II. TAXONOMY OF DETECTION APPROACHES

Fig. 1 organises the field into four families. The classical family extracts hand-engineered colour, texture and shape descriptors and passes them to a support vector machine, k-nearest neighbour, artificial neural network or random forest, usually after K-means or Otsu segmentation. The deep convolutional family replaces the descriptor stage with learned filters, and in practice relies heavily on transfer learning from AlexNet, VGG, GoogLeNet, ResNet, DenseNet or EfficientNet backbones pre-trained on large natural-image corpora. The transformer and hybrid family applies self-attention to image patches, either alone as a vision transformer or in combination with a convolutional front end. The deployment-oriented family is concerned less with headline accuracy than with the constraints of practical use: lightweight models that run on a handset or edge device, explainable models whose decisions can be justified to an agronomist, and few-shot models that function where labelled data are scarce.

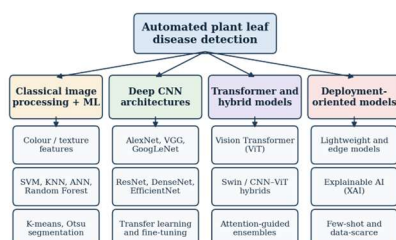


Fig. 1. Four-family taxonomy of automated plant leaf disease detection.

The families are cumulative rather than mutually exclusive. Segmentation techniques from the classical family remain in use as a pre-processing stage for deep models, and the deployment family draws its backbones from the two families that precede it. The practical consequence is that a contemporary system is best described by the combination of stages it selects, which is the view adopted in Sections IV and V.

III. LITERATURE SURVEY

Researchers have approached plant disease detection through artificial intelligence, image pre-processing, machine learning, deep learning, transfer learning, genetic algorithms, big data and the Internet of Things. The models proposed carry recognisable limitations: pre-processing is difficult to standardise, feature selection is model-dependent, and diagnostic accuracy is not uniformly high once the range of pests and diseases widens.

A. Classical Image Processing and Machine Learning

Ghaiwat et al. compared artificial neural networks, support vector machines, self-organising maps and fuzzy logic for leaf disease classification, and concluded that the k-nearest-neighbour rule was the simplest and most generally applicable of the alternatives; they also noted that where the training data are not separable, selecting the optimal parameter of a support vector machine becomes difficult [7]. Singh and Misra combined image segmentation with soft computing and established a pipeline — segment,

describe, classify — that most later classical work reproduced [8]. Adhikari et al. applied image processing to a corpus of 520 tomato leaf images covering bacterial canker, grey spot and late blight together with healthy samples, and reported 89% accuracy; the method is restricted to three diseases and one healthy class, and extending it to further pathogens requires retraining on new data [9]. Saradhambal et al. used Otsu classification with K-means clustering to isolate the affected region of the leaf, and added an integrated voice navigation system to guide the user, but reported neither the accuracy achieved nor the size of the dataset used, which makes the result difficult to place [10].

B. Deep Convolutional Architectures

Mohanty et al. established the reference result for the deep convolutional family by training on the standard benchmark corpus and demonstrating that a transfer-learned network could separate a large number of crop-disease pairs, while also observing that performance fell sharply on images captured under conditions different from the training set [1]. Sardogan et al. applied a convolutional network whose output was passed to a learning vector quantisation stage, using a dataset of 500 images split 80/20 across five classes — one healthy and four infected — at a resolution of 512×512 with the RGB channel matrix supplied as input, and reported approximately 88% accuracy on tomato-specific disorders [11]. Park et al. used two convolutional layers followed by three fully connected layers to identify powdery mildew on strawberry leaves, reporting 89.7% accuracy on a CPU with images resized to 227×227 , although the categorisation was limited to healthy and powdery-mildew classes over only 500 images [12]. Wu et al. combined AlexNet and GoogLeNet within a transfer-learning framework across 38 classes of normal and diseased leaf samples, improving identification accuracy at the cost of a complex model that converges slowly [13]. Jiang et al. addressed the detection rather than the classification problem, proposing an improved convolutional detector for apple leaf disease in real time [14], and Zhang et al. used multi-feature fusion with Faster R-CNN on synthetic soybean leaf images to compensate for limited real data [15].

C. Transformer and Hybrid Models

Borhani et al. introduced a vision transformer for automated plant disease classification and also evaluated conventional convolutional models and CNN-ViT combinations, coupling convolutional blocks with attention blocks to accelerate prediction; the highest convergence was obtained on wheat rust, rice leaf and the standard benchmark, although the study reported convergence rather than a clean accuracy figure, which is not an ideal primary metric [16]. Later work extended the idea in several directions: FormerLeaf adapted an efficient transformer to cassava leaf disease [17]; Devi et al. applied EfficientNetV2 to joint disease classification and pest recognition [18]; Ali et al. evaluated vision transformers directly against convolutional baselines [19]; and Jia et al. proposed ConvTransNet-S, a convolutional-transformer hybrid designed specifically for complex field environments rather than laboratory imagery [20]. A multi-kernel inception-enhanced vision transformer reported

99.41% on the standard benchmark together with strong results on apple, rice, ibean and cassava corpora, and notably a much lower 76.51% on cassava, which is the clearest published evidence that benchmark accuracy does not transfer to harder field-collected data [21].

D. Deployment-Oriented and Review Studies

The most recent literature has shifted from accuracy to deployability. Guo and Guerra-Manzanares proposed an explainable, lightweight, attention-guided ensemble aimed at data-scarce conditions [22], and Kafi et al. presented an interpretable detection system targeted explicitly at edge computing [23]. Recent comparative studies have begun to benchmark convolutional, residual and transformer models under identical pre-processing, augmentation and optimisation conditions, which is the control that most earlier comparisons omitted [24]. At the review level, Das et al. conducted a systematic analysis of deep learning for plant disease detection [25]; Rathore and Bhati reviewed the field with a specific focus on transformers and explainable artificial intelligence [26]; and Ghimire et al. examined forty-one peer-reviewed studies published between 2008 and 2025 under PRISMA guidelines, summarising convolutional networks, vision transformers, transfer and few-shot learning and multimodal sensing, and concluding that reported performance is consistently obtained under conditions more favourable than those of real deployment [27].

E. Related Deep Learning Work by the Authors' Group

The image-classification methodology reviewed here has been applied by the present group across several adjacent domains. Transfer-learning-based object detection using convolutional networks [28] and convolutional glaucoma detection from retinal images [29] use the same backbone-plus-fine-tuning strategy that dominates Section III-B. Disease screening from images has been addressed for monkeypox using GoogLeNet [30] and for cancer prediction using random forest combined with deep learning [31]. Detection and recognition pipelines have been built for face-mask identification with OpenCV [32], multi-stage convolutional face-mask detection [33], activity detection and people counting with Mask-RCNN and bidirectional ConvLSTM [34], and facial expression recognition using both eigenface [35] and neural network [36] formulations. Kernel support vector machines have been applied to brain image classification [37], machine learning to rare disease identification from electronic health records [38], and an AIoT device to real-time object recognition with voice conversion [39]. Sequence models have been used for network traffic prediction with long short-term memory [40], and principal component analysis with random forest for intrusion detection [41]. The methodological continuity across these studies — careful pre-processing, explicit feature construction and comparative evaluation of several backbones — is the perspective from which the present review is written.

TABLE I. Summary of Representative Reviewed Studies

Ref.	Technique	Crop / dataset	Outcome and limitation
[7]	ANN, SVM, SOM, fuzzy, KNN	Generic leaf	KNN simplest; SVM parameter hard to tune
[8]	Segmentation + soft computing	Multiple	Established segment-describe-classify pipeline

Ref.	Technique	Crop / dataset	Outcome and limitation
[9]	Image processing	Tomato, 520 images	89%; only three diseases covered
[10]	Otsu + K-means, voice guide	Generic leaf	Accuracy and dataset size not reported
[1]	Transfer-learned CNN	Benchmark corpus	Reference deep result; fails on field images
[11]	CNN + LVQ	Tomato, 500 images	≈88%; small dataset, five classes
[12]	CNN (2 conv + 3 FC)	Strawberry, 500 images	89.7%; binary task only
[13]	AlexNet + GoogLeNet transfer	38 classes	Higher accuracy; complex, slow convergence
[14]	Improved CNN detector	Apple leaf	Real-time detection rather than classification
[15]	Multi-feature Faster R-CNN	Soybean (synthetic)	Compensates for limited real data
[16]	ViT and CNN-ViT	Wheat, rice, benchmark	Reports convergence, not clean accuracy
[18]	EfficientNetV2	Disease + pest	Joint disease and pest recognition
[20]	ConvTransNet-S hybrid	Field imagery	Designed for complex field conditions
[21]	Multi-kernel inception ViT	Five corpora	99.41% benchmark vs 76.51% cassava
[22]	Explainable lightweight ensemble	Data-scarce	Targets scarcity and interpretability
[23]	Interpretable edge model	Field deployment	Optimised for edge computing
[27]	PRISMA review, 41 studies	2008–2025	Performance obtained under favourable conditions

IV. DETECTION METHODOLOGY

Across the families of Section II, the reviewed systems share a common skeleton. Fig. 2 presents that skeleton as a five-stage pipeline, extended with the augmentation and feature-selection stages that recent work has shown to be decisive.

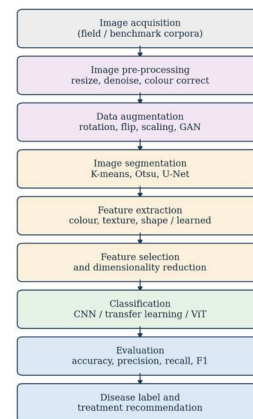


Fig. 2. Generic pipeline for automated leaf disease detection.

A. Image Acquisition

The first stage obtains leaf images, either from a publicly available repository or by direct field capture. Repository corpora contain healthy leaves together with samples damaged by disease or insect attack, usually photographed against a uniform background under controlled illumination. Field capture introduces variable lighting, occlusion, overlapping leaves and cluttered backgrounds, and the divergence between the two acquisition regimes is

the single largest source of the generalisation gap discussed in Section VI.

B. Pre-processing and Augmentation

Noise is removed using filtering: water spots, spores and dust on the leaf surface alter pixel values and would otherwise be learned as spurious evidence. Images are resized to a fixed input resolution to reduce computation, geometric deformity is corrected, and a smoothing filter limits distortion. Because most public corpora are small relative to the capacity of a deep network, augmentation — rotation, horizontal and vertical flipping, scaling, brightness and contrast jitter, and in some studies generative synthesis — is now a standard part of this stage rather than an optional addition.

C. Image Segmentation

Segmentation separates the leaf from its background and partitions it into regions. K-means clustering and the Otsu threshold are the classical instruments; the RGB image is commonly converted to the LAB colour model first, because the separation of luminance from chrominance makes the clustering of lesion regions substantially more stable. Segmentation yields the lesion boundary, the affected area and the colour statistics of the region, all of which feed the descriptor stage [8], [10]. In deep pipelines this stage is increasingly performed by a learned segmentation network rather than by thresholding.

D. Feature Extraction and Selection

In the classical family, the colour co-occurrence method is the dominant descriptor: the RGB image is transformed into hue-saturation-intensity space and a co-occurrence matrix is constructed from the pixel map, from which homogeneity, contrast, energy, correlation, variance, standard deviation and mean are derived, alongside shape and surface descriptors. In the deep family this stage is absorbed into the network, and the convolutional filters learn the descriptors directly from data. Feature selection then reduces dimensionality before classification, which matters most where the training corpus is small relative to the descriptor count.

E. Detection and Classification

The final stage assigns the leaf to a healthy or diseased class and, where the model is multi-class, to a specific pathogen. Classical systems apply a minimum-distance criterion or a conventional classifier to the selected descriptors; deep systems pass the learned feature map to fully connected layers with a softmax output. The class vector supplied as the target during training determines the granularity of the diagnosis, and it is at this point that most reviewed systems make their principal simplification — restricting the label set to a small number of pathogens on a single crop.

V. MODEL ARCHITECTURE

Fig. 3 shows the corresponding architecture, in which the classification backbone is deliberately left interchangeable so that convolutional, transfer-learned and transformer models can be substituted without altering the surrounding stages.

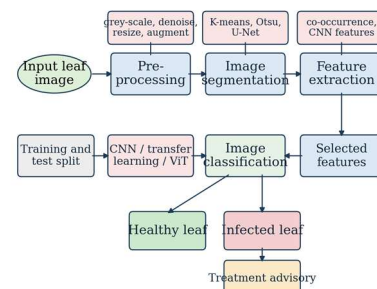


Fig. 3. Reference architecture with an interchangeable classification backbone.

A convolutional network comprises an input layer, an output layer and a set of hidden layers; the hidden portion contains convolution layers, pooling layers, activation functions — most commonly rectified linear units — and fully connected layers [11]. The number of layers, their arrangement and the additional processing units introduced between them are what distinguish one architecture from another.

A. AlexNet

AlexNet comprises five convolutional layers, three max-pooling layers, two normalisation layers, two fully connected layers and a softmax output. Each convolutional layer contains convolutional filters followed by a rectified linear activation, and the pooling layers perform maximum pooling. Because the fully connected layers fix the input dimensionality, the input size is constant. The architecture carries approximately sixty million parameters, which makes it heavy relative to its accuracy by contemporary standards.

B. GoogLeNet

GoogLeNet stacks nine inception modules across twenty-two deep layers with pooling interleaved, and replaces the large fully connected block with a global average pooling layer after the inception stack. The inception module applies several filter sizes in parallel and concatenates the result, which captures features at multiple scales within one layer. The architecture carries approximately four million parameters, an order of magnitude fewer than AlexNet, and it is this efficiency that made it the preferred transfer-learning backbone in the studies of Section III-B [13].

C. Residual, Dense and Efficient Networks

ResNet introduced identity shortcut connections that allow gradients to bypass layers, which made networks of a hundred or more layers trainable. DenseNet extended the idea by connecting every layer to every subsequent layer within a block, improving parameter efficiency and feature reuse. EfficientNet and EfficientNetV2 apply compound scaling of depth, width and resolution and currently give the best accuracy-per-parameter among convolutional backbones, which is why they dominate recent work on disease classification and pest recognition [18].

D. Vision Transformer and Hybrid Backbones

The vision transformer divides the image into fixed-size patches, embeds each patch linearly and processes the sequence with multi-head self-attention, so that every patch can attend to every other patch from the first layer. This global receptive field is the main advantage over

convolution, whose receptive field grows only with depth, and it is particularly relevant where symptoms are distributed across the leaf rather than localised. The cost is data hunger: transformers require substantially more training data or stronger augmentation than convolutional networks of comparable capacity. Hybrid designs address this by placing a convolutional front end before the attention stack, retaining the inductive bias of convolution while adding global context [16], [20].

TABLE II. Comparison of Classification Backbones

Backbone	Depth	Parameters	Principal characteristic
AlexNet	8 layers	≈60 million	Early deep CNN; heavy FC block
GoogLeNet	22 layers	≈4 million	Inception modules; multi-scale
ResNet	50–152	≈25–60 million	Residual shortcuts; very deep
DenseNet	121–201	≈8–20 million	Dense connectivity; feature reuse
EfficientNetV2	scaled	≈21–120 million	Compound scaling; best CNN efficiency
Vision Transformer	12–24 blocks	≈86 million+	Global self-attention; data hungry
CNN–ViT hybrid	variable	variable	Convolutional bias plus global context

VI. COMPARATIVE PERFORMANCE AND DISCUSSION

Fig. 4 plots the accuracy reported by representative studies across the period surveyed. The trend is unambiguous.

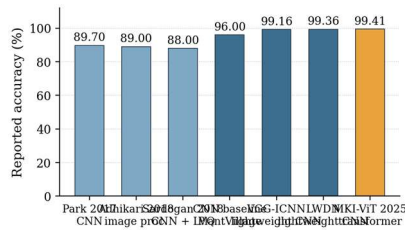


Fig. 4. Accuracy reported by representative studies. Datasets differ across studies, so the comparison is indicative rather than strict.

Early convolutional and image-processing systems cluster in the high eighties: 88.0% for the convolutional network with learning vector quantisation on tomato [11], 89.0% for image processing on a three-disease tomato corpus [9], and 89.7% for the strawberry powdery mildew classifier [12]. Convolutional baselines on the standard benchmark now reach approximately 96%, while lightweight convolutional designs and the multi-kernel inception transformer reach 99.16%, 99.36% and 99.41% respectively [21]. Measured on the benchmark, therefore, the problem appears close to solved.

That conclusion does not survive contact with field data. The same transformer that reports 99.41% on the benchmark reports 76.51% on the cassava corpus, which is collected under field rather than laboratory conditions [21]. Mohanty et al. observed the same effect a decade earlier, when accuracy fell sharply on images captured outside the training distribution [1]. The PRISMA review of Ghimire et al. reaches the same position from a systematic base [27]. The gap of more than twenty percentage points between benchmark and field performance, rather than the last fraction of a per cent on the benchmark, is the quantity that current research should be reducing.

Three further observations follow from Table I. First, dataset size in the early literature is very small — 500 to 520 images in several influential studies [9], [11], [12] — so the reported figures carry wide confidence intervals that are almost never stated. Second, class coverage is narrow: binary or five-class problems on a single crop are common, whereas an operational system must discriminate among many pathogens across many crops and must also reject inputs it has not seen. Third, evaluation protocols differ in split ratio, augmentation and pre-processing, so cross-study comparison of the kind plotted in Fig. 4 is indicative only; the controlled benchmarking now beginning to appear [24] is the correct remedy.

VII. RESEARCH GAPS AND EXPECTED OUTCOME

The following gaps remain open.

- Field-condition robustness. Accuracy obtained on uniformly lit, single-leaf laboratory images does not transfer to cluttered field imagery with variable illumination and occlusion.
- Dataset diversity and size. Public corpora are dominated by a small number of crops and are geographically narrow, and several influential results rest on a few hundred images.
- Class imbalance and rare pathogens. Diseases of low prevalence are under-represented, so accuracy is a misleading summary and the models that matter most operationally are the least well trained.
- Early-stage detection. Most systems recognise established symptoms; detecting infection before visible lesions develop, which is where the agronomic value lies, remains largely unaddressed.
- Multi-disease and co-infection. Real plants carry more than one condition at a time, while most reviewed systems assume a single label per image.
- Computational cost. Transformer and deep convolutional backbones exceed the budget of the handset or edge device on which a cultivator would actually run them, which motivates the lightweight and distilled designs of Section III-D [22], [23].
- Interpretability. An agronomist acting on a diagnosis needs to know which region of the leaf drove it; explainable formulations are only now being integrated systematically [26].

The expected outcome of addressing these gaps is a detection system that analyses and diagnoses plant disease accurately enough to support healthy crop growth under real cultivation conditions rather than under laboratory conditions. Such a system would overcome the limitations of manual identification — slow, expensive, and inconsistent where expertise is limited — and would deliver early-stage diagnosis together with an actionable treatment recommendation, improving both crop productivity and farm profitability.

VIII. CONCLUSION

This paper reviewed the development of automated plant leaf disease detection from classical image processing and machine learning, through deep convolutional architectures and transfer learning, to vision transformers,

hybrid models and the deployment-oriented designs reported up to 2026. Thirty-six studies were organised into a four-family taxonomy, a generic five-stage detection pipeline and a reference architecture with an interchangeable classification backbone were presented, and the principal backbones were compared by depth, parameter count and characteristic behaviour. The comparative analysis showed that accuracy reported on the standard benchmark has risen from the high eighties to above ninety-nine per cent, but that the same models lose more than twenty percentage points on field-collected corpora, which establishes generalisation rather than benchmark accuracy as the binding constraint on the field.

Future work should therefore concentrate on field-condition robustness through domain adaptation and stronger augmentation; on larger, more diverse and better balanced corpora covering more crops and geographies; on early-stage and multi-disease detection; on lightweight and distilled architectures that meet an edge inference budget; and on explainable formulations that justify each diagnosis to the agronomist who must act on it. Controlled benchmarking under identical pre-processing and evaluation protocols should become the default reporting standard, so that the contribution of each architectural choice can be measured rather than asserted.

REFERENCES

- [1] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, Art. no. 1419, 2016, doi: 10.3389/fpls.2016.01419.
- [2] V. Singh, N. Sharma, and S. Singh, "A review of imaging techniques for plant disease detection," *Artificial Intelligence in Agriculture*, vol. 4, pp. 229–242, 2020, doi: 10.1016/j.aiaa.2020.10.002.
- [3] S. B. Dhaygude and N. P. Kumbhar, "Agricultural plant leaf disease detection using image processing," *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.*, vol. 2, no. 1, pp. 599–602, 2013.
- [4] S. R. Maniyath et al., "Plant disease detection using machine learning," in *Proc. Int. Conf. Design Innovations for 3Cs Compute Communicate Control (ICDI3C)*, 2018, pp. 41–45, doi: 10.1109/ICDI3C.2018.00017.
- [5] P. Sharma, P. Hans, and S. C. Gupta, "Classification of plant leaf diseases using machine learning and image preprocessing techniques," in *Proc. 10th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, 2020, pp. 480–484, doi: 10.1109/Confluence47617.2020.9057889.
- [6] R. Gandhi, S. Nimbalkar, N. Yelamanchili, and S. Ponkshe, "Plant disease detection using CNNs and GANs as an augmentative approach," in *Proc. IEEE Int. Conf. Innovative Res. Develop. (ICIRD)*, 2018, pp. 1–5, doi: 10.1109/ICIRD.2018.8376321.
- [7] S. N. Ghaiwat and P. Arora, "Detection and classification of plant leaf diseases using image processing techniques: A review," *Int. J. Recent Adv. Eng. Technol.*, vol. 2, no. 3, pp. 1–7, 2014.
- [8] V. Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," *Information Processing in Agriculture*, vol. 4, no. 1, pp. 41–49, 2017, doi: 10.1016/j.inpa.2016.10.005.
- [9] S. Adhikari, B. Shrestha, B. Baiju, and S. K.C., "Tomato diseases detection system using image processing," in *Proc. 1st KEC Conf. Eng. Technol.*, 2018, pp. 81–85.
- [10] G. Saradhambal, R. Dhivya, S. Latha, and R. Rajesh, "Plant disease detection and its solution using image classification," *Int. J. Pure Appl. Math.*, vol. 119, no. 14, pp. 879–884, 2018.
- [11] M. Sardogan, A. Tuncer, and Y. Ozen, "Plant leaf disease detection and classification based on CNN with LVQ algorithm," in *Proc. 3rd Int. Conf. Comput. Sci. Eng. (UBMK)*, 2018, pp. 382–385, doi: 10.1109/UBMK.2018.8566635.
- [12] H. Park, J.-S. Eun, and S.-H. Kim, "Image-based disease diagnosing and predicting of the crops through the deep learning mechanism," in *Proc. Int. Conf. Inf. Commun. Technol. Convergence (ICTC)*, 2017, pp. 129–131, doi: 10.1109/ICTC.2017.8190957.
- [13] J. Wu, B. Li, and Z. Wu, "Detection of crop pests and diseases based on deep convolutional neural network and improved algorithm," in *Proc. 4th Int. Conf. Mach. Learn. Technol. (ICMLT)*, 2019, pp. 20–27, doi: 10.1145/3340997.3341010.
- [14] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks," *IEEE Access*, vol. 7, pp. 59069–59080, 2019, doi: 10.1109/ACCESS.2019.2914929.
- [15] K. Zhang, Q. Wu, and Y. Chen, "Detecting soybean leaf disease from synthetic image using multi-feature fusion Faster R-CNN," *Computers and Electronics in Agriculture*, vol. 183, Art. no. 106064, 2021, doi: 10.1016/j.compag.2021.106064.
- [16] Y. Borhani, J. Khoramdel, and E. Najafi, "A deep learning-based approach for automated plant disease classification using vision transformer," *Scientific Reports*, vol. 12, Art. no. 11554, 2022, doi: 10.1038/s41598-022-15163-0.
- [17] T. Thai, K. Nguyen, and L. Tran, "FormerLeaf: An efficient vision transformer for cassava leaf disease detection," *Computers and Electronics in Agriculture*, vol. 204, Art. no. 107518, 2023, doi: 10.1016/j.compag.2022.107518.
- [18] R. S. Devi, V. R. Kumar, and P. Sivakumar, "EfficientNetV2 model for plant disease classification and pest recognition," *Computer Systems Science and Engineering*, vol. 45, no. 2, pp. 2249–2263, 2023, doi: 10.32604/csse.2023.032231.
- [19] M. Ali, M. Salma, M. Haji, and B. Jamal, "Plant disease detection using vision transformers," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 15, no. 2, pp. 2334–2344, 2025, doi: 10.11591/ijece.v15i2.pp2334-2344.
- [20] S. Jia, G. Wang, H. Li, Y. Liu, L. Shi, and S. Yang, "ConvTransNet-S: A CNN-transformer hybrid disease recognition model for complex field environments," *Plants*, vol. 14, no. 15, Art. no. 2252, 2025, doi: 10.3390/plants14152252.
- [21] "Multi-kernel inception-enhanced vision transformer for plant leaf disease recognition," *Scientific Reports*, 2025, doi: 10.1038/s41598-025-16142-x.
- [22] M. Guo and A. Guerra-Manzanarez, "CROPPER: An explainable lightweight attention-guided ensemble for robust plant disease diagnosis in data-scarce conditions," 2026.
- [23] M. A. A. Kafi, S. K. Banshal, R. Moni, A. L. Aziz, M. Aljuaid, and R. Bansal, "LeafAI: Interpretable plant disease detection for edge computing," *PLOS ONE*, 2026.
- [24] Anirudh, Srujan, Sai, and G. Swathika O. V., "Comparative deep learning approaches for bean leaf disease recognition," *Frontiers in Plant Science*, vol. 17, Art. no. 1703449, 2026, doi: 10.3389/fpls.2026.1703449.
- [25] P. K. Das, S. S. Rupa, S. Purnin, U. C. Das, and M. K. Hossen, "Deep learning for plant disease detection and classification: A systematic analysis and review," *Current Applied Science and Technology*, vol. 24, no. 4, Art. no. e0259016, 2024, doi: 10.55003/cast.2024.259016.
- [26] M. Rathore and H. Bhati, "Plant leaf disease detection with deep learning techniques: A review with a focus on transformers and explainable AI," in *Expert Applications and Security (NICE-TEAS 2025)*, *Lecture Notes in Networks and Systems*, vol. 1854, Cham, Switzerland: Springer, 2026, doi: 10.1007/978-3-032-18846-5_10.
- [27] S. Ghimire, R. Lamsal, A. Sankuratri, P. Lakkarsu, and S. Vutla, "Artificial intelligence-driven plant disease detection and diagnosis: A comprehensive review of deep learning approaches, multimodal sensing technologies, and future perspectives in precision agriculture," *The Plant Pathology Journal*, vol. 42, no. 2, pp. 121–137, 2026, doi: 10.5423/PPJ.RW.01.2026.0004.
- [28] B. Bamne, N. Shrivastava, L. Parashar, and U. Singh, "Transfer learning-based object detection by using convolutional neural networks," in *Proc. Int. Conf. Electron. Sustain. Commun. Syst. (ICESC)*, Coimbatore, India, 2020, pp. 328–332, doi: 10.1109/ICESC48915.2020.9156060.
- [29] A. Saxena, A. Vyas, L. Parashar, and U. Singh, "A glaucoma detection using convolutional neural network," in *Proc. Int. Conf. Electron. Sustain. Commun. Syst. (ICESC)*, Coimbatore, India, 2020, pp. 815–820, doi: 10.1109/ICESC48915.2020.9155930.
- [30] U. Singh and L. S. Songare, "Analysis and detection of monkeypox using the GoogLeNet model," in *Proc. Int. Conf. Autom. Comput.*

- Renew. Syst. (ICACRS), Pudukkottai, India, 2022, pp. 1000–1008, doi: 10.1109/ICACRS55517.2022.10029125.
- [31] M. Ranjan, A. Shukla, K. Soni, S. Varma, M. Kuliha, and U. Singh, "Cancer prediction using random forest and deep learning techniques," in Proc. IEEE 11th Int. Conf. Commun. Syst. Netw. Technol. (CSNT), Indore, India, 2022, pp. 227–231, doi: 10.1109/CSNT54456.2022.9787608.
- [32] R. Baghel, P. Pahadiya, and U. Singh, "Human face mask identification using deep learning with OpenCV techniques," in Proc. 7th Int. Conf. Commun. Electron. Syst. (ICES), Coimbatore, India, 2022, pp. 1051–1057, doi: 10.1109/ICES54183.2022.9835884.
- [33] S. P. Singh, J. Shukla, S. Sharma, K. Daga, B. S. Bhalavi, and U. Singh, "Face mask detection using multi-stage CNN architecture," in Proc. Int. Conf. IoT Based Control Netw. Intell. Syst. (ICINIS), 2021, doi: 10.2139/ssrn.3884022.
- [34] U. Singh, P. Gupta, and M. Shukla, "Activity detection and counting people using Mask-RCNN with bidirectional ConvLSTM," J. Intell. Fuzzy Syst., 2022, pp. 6505–6520.
- [35] P. Shakyawar, P. Choure, and U. Singh, "Eigenface method through facial expression recognition," in Proc. Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA), 2017, doi: 10.1109/ICECA.2017.8212714.
- [36] K. Kushwah, V. Sharma, and U. Singh, "Neural network method through facial expression recognition," in Proc. Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA), 2017, doi: 10.1109/ICECA.2017.8212721.
- [37] Y. Mathur, P. Jain, and U. Singh, "Foremost section study and kernel support vector machine through brain images classifier," in Proc. Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA), 2017, doi: 10.1109/ICECA.2017.8212726.
- [38] H. Soni, A. Vyas, and U. Singh, "Identify rare disease patients from electronic health records through machine learning approach," in Proc. Int. Conf. Inventive Res. Comput. Appl. (ICIRCA), 2018, doi: 10.1109/ICIRCA.2018.8597203.
- [39] P. Gupta, M. Shukla, N. Arya, U. Singh, and K. Mishra, "Let the blind see: An AlloT-based device for real-time object recognition with voice conversion," in Machine Learning for Critical Internet of Medical Things, Cham, Switzerland: Springer, 2022, doi: 10.1007/978-3-030-80928-7_8.
- [40] S. Nihale, S. Sharma, L. Parashar, and U. Singh, "Network traffic prediction using long short-term memory," in Proc. Int. Conf. Electron. Sustain. Commun. Syst. (ICESC), Coimbatore, India, 2020, pp. 338–343, doi: 10.1109/ICESC48915.2020.9156045.
- [41] S. Waskle, L. Parashar, and U. Singh, "Intrusion detection system using PCA with random forest approach," in Proc. Int. Conf. Electron. Sustain. Commun. Syst. (ICESC), Coimbatore, India, 2020, pp. 803–808, doi: 10.1109/ICESC48915.2020.9155656.