

Deep Learning-Based Detection and Classification of Sorghum Leaf Diseases Using Computer Vision and Transfer Learning

Umesh K. Joshi and Dr. Sushma Agarwal

Research Scholar, Jayoti Vidyapeeth Women's University, Jaipur

Professor, Jayoti Vidyapeeth Women's University, Jaipur

Abstract: *Sorghum (Sorghum bicolor (L.) Moench) is an important cereal crop in semi-arid and drought-prone regions because of its adaptability, nutritional value, and capacity to maintain production under comparatively harsh environments. Its productivity is nevertheless constrained by foliar diseases, including anthracnose, leaf blight, rust, and other leaf-spotting disorders. Conventional diagnosis depends heavily on visual inspection by farmers, extension workers, and plant pathologists. Expert examination remains essential for definitive diagnosis, but manual scouting is time-consuming, subjective, and difficult to scale across large production areas. This research paper presents a rigorous framework for deep learning-based detection and classification of sorghum leaf diseases using computer vision and transfer learning. The framework integrates field-oriented image acquisition, expert-supported labelling, image quality control, preprocessing, augmentation, convolutional neural networks, and transfer-learning architectures including ResNet50, MobileNetV2, EfficientNetB0, and DenseNet121. It emphasizes independent dataset partitioning, class imbalance management, confusion-matrix analysis, interpretability, and external validation rather than relying on overall accuracy alone. Published research demonstrates that convolutional neural networks can learn disease-related colour, texture, lesion, and spatial patterns directly from images, while transfer learning reduces dependence on very large crop-specific datasets. The paper also distinguishes classification under controlled image conditions from dependable diagnosis in field environments, where illumination, occlusion, cultivar differences, symptom severity, background clutter, and mixed stresses affect performance. It concludes that deep learning can provide a valuable decision-support mechanism for sorghum disease screening when dataset design, diagnostic labels, validation, and reporting are scientifically controlled. Artificial intelligence should complement rather than replace plant-pathological expertise.*

Keywords: Sorghum; Sorghum bicolor; leaf disease; deep learning; convolutional neural network; computer vision; transfer learning; ResNet50; MobileNetV2; EfficientNetB0; precision agriculture.

I. INTRODUCTION

Sorghum is one of the major cereal crops cultivated in dry and semi-arid regions of the world. It is valued for grain, fodder, feed, brewing, and industrial uses and is particularly important where rainfall is uncertain and high temperatures restrict the productivity of less resilient cereals. The crop's physiological tolerance to drought, however, does not protect it from biological stresses. Diseases can reduce photosynthetic area, weaken plants, affect grain filling, and diminish both yield and quality. Foliar diseases are especially important from a computer-vision perspective because infection frequently produces visible changes in colour, texture, lesion shape, lesion distribution, and tissue integrity that can be recorded in digital images.

Disease identification in conventional crop management normally begins with visual scouting. Farmers and field workers observe leaves and compare symptoms with their practical experience, while plant pathologists may use lesion morphology, host information, environmental history, microscopy, culturing, serological methods, or molecular tests when confirmation is necessary. This diagnostic process is scientifically valuable, but expert support is not uniformly available at every farm. Large cultivation areas also make repeated plant-by-plant inspection expensive and slow. When symptoms are subtle or several disorders appear similar, visual diagnosis by non-specialists can be inconsistent.

Computer vision provides a complementary approach by converting digital images into measurable information. Earlier image-processing systems usually depended on manually designed features such as colour histograms, texture descriptors, edge measures, and shape statistics. These features were then supplied to conventional classifiers. Deep learning altered this workflow by allowing models to learn useful visual representations directly from labelled images. Convolutional neural networks (CNNs), in particular, became influential because their filters can capture local spatial patterns and combine them hierarchically into increasingly complex representations. The large-scale success of CNNs in general image recognition created a methodological foundation for agricultural image analysis (Krizhevsky et al., 2012; LeCun et al., 2015).

Plant disease recognition became an important application of this approach. Mohanty, Hughes, and Salathé (2016) demonstrated that deep neural networks could classify a large collection of healthy and diseased crop-leaf images with very high performance under the conditions of the PlantVillage dataset. Their study was important not simply because of the reported accuracy, but because it also exposed the difference between controlled image classification and performance on images captured under different conditions. This distinction is central to sorghum research. A model trained on isolated leaves photographed against uniform backgrounds may not behave similarly when it encounters overlapping leaves, soil, shadows, variable sunlight, insects, senescence, motion blur, or mixed symptoms in an actual field.

Transfer learning is particularly relevant because sorghum-specific datasets are normally much smaller than general image collections. Instead of training a deep network from random initialization, transfer learning begins with a network previously trained on a large source dataset such as ImageNet. The pretrained layers already contain filters that respond to edges, textures, shapes, and other reusable image structures. These representations can then be adapted to the target agricultural task. The present paper therefore develops a research-oriented framework for sorghum leaf disease detection and classification based on transfer learning, while giving equal attention to dataset quality, biological validity, model evaluation, and field generalization.

II. BACKGROUND AND RATIONALE

Plant disease recognition lies at the intersection of plant pathology, computer vision, and statistical learning. A disease is a biological condition, whereas a photograph records only its visible expression at a particular time and under particular imaging conditions. A computer-vision model does not directly observe a pathogen. It learns statistical associations between image patterns and the labels supplied during training. This distinction prevents an important conceptual error: a high classification score should not automatically be interpreted as proof that the causal organism has been identified.

Sorghum foliar symptoms can be visually distinctive, but they can also overlap. Anthracnose may produce characteristic necrotic lesions; rust is associated with pustular symptoms; and leaf blight may form elongated or irregular necrotic areas. Yet symptom expression changes with cultivar, pathogen strain, disease stage, leaf age, temperature, humidity, crop nutrition, and simultaneous stresses. A leaf damaged by drought, insects, nutrient deficiency, or physical injury may contain colour and texture patterns that partially resemble disease. A scientifically credible image dataset therefore requires expert-supported labels and clear inclusion criteria.

Illustrative sorghum leaf symptom patterns (schematic, not diagnostic photographs)

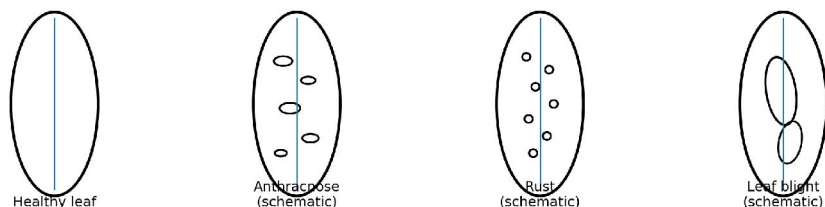


Figure 1. Schematic visual patterns associated with major sorghum leaf conditions. These are illustrative drawings, not diagnostic photographs.

The rationale for deep learning is that disease symptoms are spatially structured. Lesions are not merely collections of individual pixel values; they have boundaries, internal textures, colour transitions, shapes, orientations, and distributions across leaf tissue. CNNs are designed to learn such local spatial structures. Early network layers often respond to edges and simple colour transitions, while deeper layers combine them into higher-level patterns. This makes CNNs appropriate for differentiating healthy tissue from disease-associated visual structures without requiring the researcher to define every discriminative feature manually.

A second rationale is operational scalability. Once a model has been trained and validated, image inference can be performed repeatedly and consistently. This can support screening of large image collections, prioritization of suspicious samples, and structured field monitoring. The value of automation is therefore not that it removes the need for plant pathology, but that it can reduce the burden of repetitive visual screening and make expert attention more targeted.

III. RESEARCH PROBLEM

The principal research problem is the reliable classification of visually expressed sorghum leaf diseases from digital images under conditions that reflect genuine agricultural variability. The problem contains several connected challenges. First, disease labels must be sufficiently reliable for supervised learning. Second, the image collection must contain enough variation to prevent the network from learning accidental correlations. Third, training and test samples must be independent. Fourth, the selected architecture must achieve useful predictive performance without excessive computational cost. Finally, the model must be evaluated in a way that reveals disease-specific strengths and weaknesses rather than hiding them behind a single accuracy figure.

Data leakage is a particularly serious concern. If multiple photographs of the same leaf or nearly identical frames from one photographic sequence are divided between training and testing subsets, the test set is no longer genuinely independent. The model may appear to generalize when it is partly recognizing details already represented during training. Similar problems arise when disease classes are photographed systematically against different backgrounds. A network can exploit background information rather than pathological features. Robust experimental design must therefore control the unit of partitioning and, where possible, separate data by plant, farm, season, or location.

Avoiding data leakage through grouped dataset partitioning

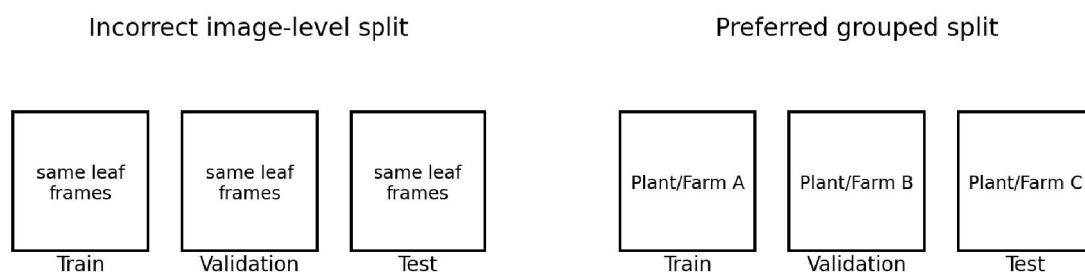


Figure 2. Illustration of data leakage risk and grouped train-validation-test separation.

Another problem is class imbalance. Some diseases may be common and easy to photograph, while others may be rare or seasonal. A dataset dominated by healthy leaves or one major disease can produce a high overall accuracy even when the model performs poorly on minority classes. Because missing a disease may have practical consequences, class-specific recall, precision, and F1-score are required alongside overall accuracy.

IV. OBJECTIVES AND RESEARCH QUESTIONS

The objectives of the study are to develop a systematic computer-vision workflow for sorghum leaf disease recognition; examine preprocessing and augmentation procedures appropriate for leaf imagery; evaluate transfer-learning architectures represented by ResNet50, MobileNetV2, EfficientNetB0, and DenseNet121; assess performance through



accuracy, precision, recall, F1-score, confusion matrices, and computational measures; examine the influence of class imbalance and dataset design; and identify requirements for reliable field-oriented validation.

The study addresses four main questions: How effectively can deep CNNs discriminate visually expressed sorghum leaf diseases? What advantages does transfer learning offer when crop-specific labelled data are limited? How should the selected architectures be compared when both predictive performance and computational efficiency matter? Which methodological factors most strongly determine whether a model generalizes beyond the images on which it was developed?

V. REVIEW OF LITERATURE

The development of deep learning for visual recognition was accelerated by large-scale image-classification research. Krizhevsky, Sutskever, and Hinton (2012) demonstrated that a deep convolutional architecture could achieve major improvements on ImageNet. Subsequent research established deeper and more efficient architectures. ResNet introduced residual connections that made very deep networks easier to optimize (He et al., 2016). DenseNet promoted extensive feature reuse through dense connectivity (Huang et al., 2017). MobileNet families focused on computational efficiency for constrained devices (Howard et al., 2017; Sandler et al., 2018), while EfficientNet introduced coordinated scaling of depth, width, and image resolution (Tan & Le, 2019).

Agricultural researchers rapidly adapted these methods. Mohanty et al. (2016) trained deep models on 54,306 PlantVillage images representing healthy and diseased leaves. Their best controlled-dataset result demonstrated the technical potential of CNN-based disease recognition, but the reduced performance under different image conditions also became an important warning about domain shift. Ferentinos (2018) evaluated deep learning models across a large collection of plant disease images and showed that CNNs could distinguish numerous crop-disease combinations. Too et al. (2019) compared fine-tuned deep learning architectures and demonstrated the importance of architecture selection and transfer learning for plant disease identification.

Kamilaris and Prenafeta-Boldú (2018) reviewed deep learning applications in agriculture and documented the expanding role of neural networks across classification, detection, and related tasks. Their review highlighted the capacity of deep models to learn features automatically while also noting the importance of sufficient training data. Shorten and Khoshgoftaar (2019) examined image data augmentation more broadly and explained why controlled transformations can improve generalization when labelled data are limited.

The literature also moved beyond classification accuracy toward understanding model behaviour. Toda and Okura (2019) investigated how CNNs diagnose plant disease and showed the importance of visualizing the regions and characteristics used by a network. Grad-CAM, introduced by Selvaraju et al. (2017), became widely used for this purpose because it produces class-discriminative localization maps from convolutional networks. Such methods are valuable in plant pathology because a model that focuses on lesions is more credible than one whose decision is dominated by irrelevant background objects.

Disease severity represents another related research direction. Wang, Sun, and Wang (2017) explored image-based severity estimation using deep learning, demonstrating that computer vision can move beyond categorical recognition toward quantitative assessment. Severity estimation is distinct from disease classification, however. A model may identify a disease correctly without accurately measuring the proportion of tissue affected. For sorghum research, these tasks should be evaluated separately unless the experimental design explicitly integrates them.

Transfer learning has become especially important in crop disease research because many agricultural datasets are relatively small. Features learned from large image datasets can provide a useful starting point even though the source images are not agricultural. Fine-tuning then adapts deeper representations to the target disease classes. The success of transfer learning does not imply that ImageNet contains plant-pathological knowledge; rather, it demonstrates that visual primitives such as edges, contours, textures, and object parts can be reused across domains.

The existing literature nevertheless reveals a recurring limitation: many studies use datasets captured under controlled or semi-controlled conditions. High internal test performance may therefore coexist with poor external generalization. Agricultural images vary across cameras, farms, seasons, cultivars, disease stages, and weather conditions. Reliable

sorghum research requires deliberate attention to these sources of domain shift rather than treating them as noise to be removed from the dataset.

VI. MATERIALS AND METHODS

The study adopts a quantitative experimental design in which digital sorghum leaf images constitute the principal observational units. The methodological workflow consists of image acquisition, expert-supported disease labelling, quality assessment, preprocessing, data partitioning, augmentation, transfer-learning model development, validation, independent testing, and error analysis. The design treats classification as the assignment of an image to a predefined disease category and detection as localization of relevant symptomatic objects or regions. These terms should not be used interchangeably when reporting results.

Research workflow for sorghum leaf disease classification

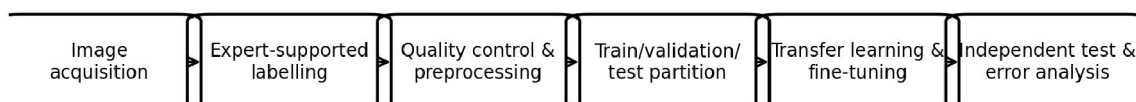


Figure 3. Proposed research workflow for deep learning-based sorghum leaf disease classification.

Images should be acquired from healthy plants and plants displaying major target diseases, with metadata recorded wherever feasible. Useful metadata include farm or plot identifier, cultivar, date, growth stage, disease label, diagnostic confidence, camera type, and environmental conditions. Such metadata make it possible to evaluate whether performance changes across locations, seasons, or devices. Images with corrupted files or extreme blur should be excluded according to predetermined quality rules, but realistic field complexity should not be systematically removed merely because it makes classification harder.

Reference labels should be assigned or reviewed by personnel with plant-pathological expertise. When visual symptoms are ambiguous, the sample should be marked uncertain, excluded from strict supervised training, or confirmed by an appropriate diagnostic method. This approach recognizes that label noise can become a hidden source of model error. A CNN trained on incorrect labels may learn consistently from them and still produce apparently stable validation curves.

Preprocessing begins with image decoding, resizing to the input dimensions required by each architecture, and normalization according to the pretrained model's convention. Resizing should preserve enough lesion detail for classification. Excessive downsampling can remove small pustules or early lesions, while unnecessarily large images increase computation. The chosen resolution should therefore be documented and applied consistently.

Data augmentation is applied only to the training subset. Appropriate transformations can include moderate rotation, translation, zoom, reflection when biologically reasonable, small changes in brightness or contrast, and limited cropping. Augmentation is intended to reduce sensitivity to photographic variation rather than create new biological observations. Transformations that distort lesion morphology or produce implausible leaf colours should be avoided because they may teach the network patterns that would not occur naturally.

The dataset is partitioned into training, validation, and independent test sets. A representative division is 70 percent, 15 percent, and 15 percent, although independence is more important than the exact ratio. Images originating from the same leaf, plant, or photographic burst should remain in a single subset. Where sufficient data exist, farm-level or season-level separation provides a stronger evaluation of generalization. The test set should remain untouched during architecture selection and hyperparameter tuning.

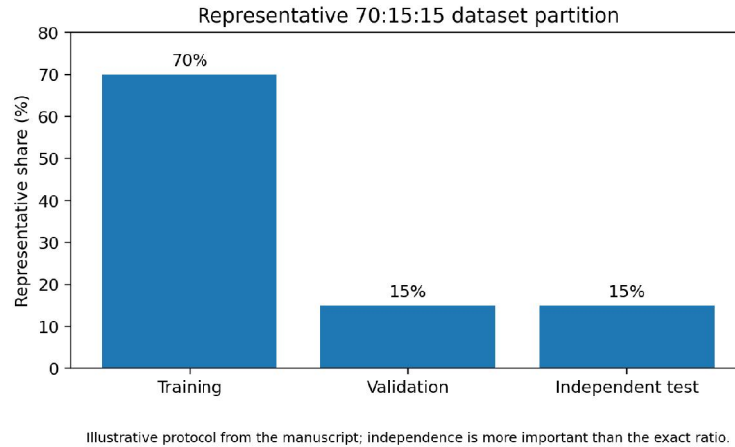


Figure 4. Representative dataset partition described in the manuscript (70% training, 15% validation, 15% independent test).

Four pretrained CNN architectures are considered. ResNet50 represents a deep residual architecture in which shortcut connections help information and gradients propagate through the network. DenseNet121 connects layers densely and encourages reuse of features learned at different depths. MobileNetV2 uses depthwise separable convolution and inverted residual structures to reduce computational demand. EfficientNetB0 uses a carefully balanced architecture developed through compound scaling principles. Together these models allow comparison between representational capacity and deployment efficiency.

For each architecture, the original ImageNet classification head is removed and replaced by a sorghum-specific head. A common configuration uses global average pooling, optional dropout for regularization, a dense layer where required, and a final softmax layer containing one output for each disease class. Initial training can freeze the convolutional base while the new head learns the target labels. Fine-tuning then unfreezes selected deeper layers and continues optimization using a smaller learning rate.

Transfer-learning strategy used in the proposed framework

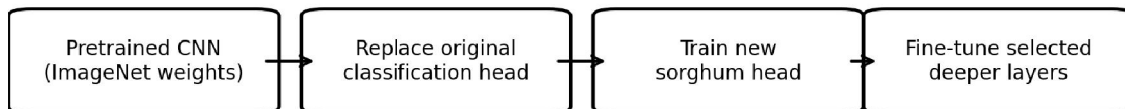


Figure 5. Transfer-learning and fine-tuning sequence for the sorghum-specific classifier.

For mutually exclusive classes, categorical cross-entropy is an appropriate loss function. Optimization may use Adam or another documented optimizer. Early stopping based on validation loss can reduce unnecessary overtraining, while model checkpoints preserve the best validation state. Learning-rate reduction may be used when validation performance plateaus. All such decisions should be recorded because reproducibility requires more than naming the architecture.

Class imbalance can be addressed through class weighting, balanced sampling, or targeted augmentation. If N represents the number of training images, C the number of classes, and n_i the number of samples in class i , a simple inverse-frequency weight can be expressed as $w_i = N/(C n_i)$. Weighting should be applied only after examining whether imbalance reflects genuine prevalence or a sampling artifact.

Performance evaluation includes overall accuracy together with class-specific and macro-averaged precision, recall, and F1-score. Accuracy is defined as the proportion of correctly classified observations. Precision measures the proportion of predicted positives that are correct, recall measures the proportion of actual positives recovered by the

model, and F1-score represents their harmonic mean. A confusion matrix is essential because it shows which disease pairs are repeatedly confused. For deployment-oriented comparison, model size, parameter count, and inference time should also be measured under the same hardware conditions.

VII. MODEL EVALUATION AND RESULT REPORTING

A research paper should report only values produced by the actual experiment. Because no raw sorghum dataset, training logs, or independent test predictions were supplied for the present manuscript, numerical results for ResNet50, MobileNetV2, EfficientNetB0, and DenseNet121 are not fabricated here. This is an important research-norm requirement. Plausible-looking accuracy values cannot substitute for measured results and would make the paper scientifically unreliable.

When the experiment is conducted, the results section should first describe the dataset after quality control, including the number of images per class and the number of independent plants or sampling units. Training and validation curves should then be presented to show convergence and possible overfitting. Final model comparison should use the untouched test set. Confidence intervals or repeated runs with different random seeds can strengthen the comparison by showing whether small differences between architectures are stable.

The confusion matrix should be interpreted at the disease level. If anthracnose is repeatedly confused with leaf blight, the researcher should inspect the corresponding images to determine whether the errors occur at early disease stages, under poor illumination, or in samples with overlapping symptoms. A model that performs well on healthy leaves but poorly on a rare disease should not be described simply as highly accurate. Macro-averaged F1-score is useful in this context because it gives equal importance to each class.

Confusion-matrix reporting template

Actual class	Healthy	—	—	—	—
	Anthrachnose	—	—	—	—
	Rust	—	—	—	—
	Leaf blight	—	—	—	—
		Predicted class			
		Healthy	Anthrachnose	Rust	Leaf blight

Cells intentionally left blank until independent test predictions are available.

Figure 6. Confusion-matrix template for reporting class-level errors; numerical cells remain blank because no test predictions were supplied.



Published values may be used only as comparative evidence. Mohanty et al. (2016), for example, reported 99.35 percent accuracy in their best PlantVillage experiment, but this value belongs to their dataset and experimental protocol. It is not a result of the present sorghum study. The same rule applies to every value reported in previous sorghum or plant-disease research. Maintaining this separation between literature evidence and original findings is essential to research integrity.

VIII. DISCUSSION

The literature and methodological analysis indicate that transfer learning is a strong candidate for sorghum leaf disease recognition, but architecture alone does not determine scientific quality. A well-designed experiment using a moderately sized network and independent field data can be more informative than a larger model evaluated on a leaky or overly controlled test set. Dataset design should therefore be treated as part of the model rather than as a preliminary clerical step.

ResNet50 offers substantial representational capacity and has a mature residual architecture. DenseNet121 provides efficient feature reuse and may preserve useful low-level disease characteristics across deeper layers. EfficientNetB0 provides a balanced option when predictive quality and computational cost are both important. MobileNetV2 is particularly relevant where inference must occur on limited hardware. The preferred model should consequently be selected through a multi-criteria comparison rather than by accuracy alone.

Field generalization is likely to remain the principal challenge. A leaf photographed on a laboratory background presents a relatively clean visual object. In the field, the network encounters overlapping leaves, variable sun angle, shadows, soil, stems, weeds, water droplets, insects, senescent tissue, and motion blur. Disease severity also changes continuously. A model may learn strong representations of advanced lesions while missing early symptoms. External testing across farms, seasons, cultivars, and devices is therefore a stronger indicator of practical value than random image-level testing alone.

The consequences of different errors also matter. In screening, a false negative may allow a diseased sample to be overlooked, while a false positive may lead to unnecessary inspection. The preferred operating point depends on the intended use. If the model is designed to prioritize leaves for expert review, higher recall may be more important than maximizing precision. If it is used to produce epidemiological counts, the balance may be different. Evaluation should therefore be linked to the decision that follows the prediction.

Model confidence deserves similar attention. A softmax classifier always produces a highest-probability class even when the image is unlike anything in the training data. A responsible system should therefore distinguish high-confidence predictions from ambiguous cases and provide a route for expert review. Calibration analysis can determine whether predicted probabilities correspond reasonably to observed correctness. This is particularly important in agricultural environments where out-of-distribution images are common.

IX. EXPLAINABILITY AND ERROR ANALYSIS

Interpretability techniques can help determine whether a model relies on biologically meaningful regions. Grad-CAM uses gradients flowing into a convolutional layer to create a class-discriminative localization map (Selvaraju et al., 2017). In a sorghum classifier, a heat map concentrated on characteristic lesions provides evidence that the prediction is associated with symptomatic tissue. A heat map concentrated on background objects, labels, or image borders indicates a possible shortcut learned from the dataset.

Explainability should not be overstated. A saliency map does not prove that a neural network understands disease mechanisms, nor does it identify a pathogen independently of the training labels. Its strongest role is model auditing. Researchers can use it to detect spurious associations, compare correctly and incorrectly classified samples, and determine whether the network's attention changes with disease severity.

Error analysis should categorize mistakes rather than merely count them. Useful categories include visually similar symptoms, early infection, severe tissue destruction, multiple symptoms on one leaf, poor illumination, blur, occlusion, senescence, insect injury, uncertain reference labels, and domain shift. Persistent errors in a particular category can

guide improvement of the dataset. For example, repeated failure on early disease suggests the need for more early-stage examples rather than simply a deeper network.

Misclassification patterns can also reveal limitations of RGB imaging. Some diseases may be difficult to separate visually before characteristic symptoms develop. In such cases, additional information such as temporal observations, environmental variables, multispectral data, or laboratory confirmation may be more valuable than further optimization of the same RGB classifier. Recognizing such limits is part of rigorous model development.

X. PRACTICAL SIGNIFICANCE

A validated sorghum disease classifier can support several agricultural workflows. Researchers can use automated screening to organize large image collections. Extension personnel can use predictions to prioritize plants requiring closer inspection. Farmers may obtain preliminary screening support when expert assistance is not immediately available. In each case, the model is most defensible as a decision-support tool rather than a replacement for diagnostic expertise.

Lightweight architectures are important for practical implementation because connectivity and computing resources may be limited in agricultural settings. MobileNetV2 and similarly efficient models can reduce memory and inference requirements. If a lightweight model performs comparably to a heavier network, its lower computational burden may make it more useful in practice. Deployment decisions should therefore consider latency, storage, energy use, and hardware compatibility together with predictive metrics.

The same image-analysis pipeline can also support longitudinal crop monitoring. Repeated observations can provide structured information about when symptoms appear and how they spread. Such data may later be combined with weather, soil, cultivar, and management information, but the disease label should remain traceable to its diagnostic basis. Integrating additional data should strengthen interpretation rather than conceal uncertainty in the original image prediction.

XI. LIMITATIONS

The framework has several limitations. Deep learning depends on the quantity, diversity, and correctness of labelled data. Rare diseases may remain underrepresented, and augmentation cannot replace genuinely independent observations. Visible symptoms may be insufficient to identify a causal pathogen in every case. Cultivar differences, disease stage, environmental stress, and mixed infection can alter symptom appearance. Images from a limited geographical region may not represent conditions elsewhere.

Transfer learning also has limits. Features learned from ImageNet provide useful visual initialization, but the source domain differs substantially from crop pathology. Fine-tuning can adapt the representation, yet the resulting model remains dependent on the target dataset. A high internal test score therefore cannot be assumed to represent performance on unseen farms or seasons.

Another limitation is that classification does not automatically provide lesion localization or disease severity. If the practical objective requires measuring affected area, a segmentation or severity-estimation component should be developed and validated separately. Similarly, a model that identifies a diseased leaf within a cropped image is not equivalent to a detector that can find multiple symptomatic leaves in a complex field scene.

XII. RECOMMENDATIONS

Future empirical work should prioritize dataset diversity and independent validation. Images should be collected across multiple farms, cultivars, growth stages, seasons, cameras, and illumination conditions. Plant- or farm-level identifiers should be retained so that data leakage can be prevented. Disease labels should be expert reviewed, and ambiguous cases should be documented rather than forced into a class.

Model comparison should include repeated training runs, macro and class-specific metrics, confusion matrices, model size, and inference time. Explainability maps should be inspected systematically, not only for a few visually attractive examples. External test sets should be preserved until the final model is selected. Where possible, the dataset partitioning protocol and code should be documented sufficiently for replication.

Finally, the intended decision context should guide threshold selection and reporting. A screening system may reasonably favor sensitivity, while a system intended to trigger costly interventions may require greater precision and expert confirmation. Human oversight should remain part of the workflow, especially for uncertain predictions and symptoms that are visually atypical.

XIII. CONCLUSION

Deep learning provides a technically credible approach to the image-based recognition of sorghum leaf diseases when it is embedded within a sound research design. CNNs can learn colour, texture, lesion, and spatial characteristics directly from images, while transfer learning allows these capabilities to be adapted when sorghum-specific datasets are limited. ResNet50, DenseNet121, EfficientNetB0, and MobileNetV2 provide useful alternative architectures representing different balances of depth, feature reuse, efficiency, and deployment cost.

The central conclusion is that reliable performance depends as much on dataset construction and validation as on network architecture. Expert-supported labels, independent train-test separation, class-aware evaluation, realistic augmentation, confusion-matrix analysis, explainability, and external field testing are essential. Without these controls, very high accuracy can reflect the structure of the dataset rather than dependable recognition of disease.

Computer vision should therefore be viewed as an extension of agricultural observation rather than a substitute for plant pathology. A carefully validated model can accelerate screening, improve consistency, and help organize disease information, while uncertain and biologically complex cases remain appropriate for expert examination. The strongest research contribution is not a single headline accuracy value, but a reproducible system whose strengths, errors, and domain of validity are clearly documented.

REFERENCES

1. Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145, 311–318.
2. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.
3. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv:1704.04861*.
4. Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708.
5. Hughes, D. P., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv:1511.08060*.
6. Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90.
7. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
8. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.
9. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7, 1419. <https://doi.org/10.3389/fpls.2016.01419>
10. Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., & Hughes, D. P. (2017). Deep learning for image-based cassava disease detection. *Frontiers in Plant Science*, 8, 1852.
11. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4510–4520.
12. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision*, 618–626.

13. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6, 60.
14. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*.
15. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning*, 97, 6105–6114.
16. Toda, Y., & Okura, F. (2019). How convolutional neural networks diagnose plant disease. *Plant Phenomics*, 2019, 9237136.
17. Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture*, 161, 272–279.
18. Wang, G., Sun, Y., & Wang, J. (2017). Automatic image-based plant disease severity estimation using deep learning. *Computational Intelligence and Neuroscience*, 2017, 2917536.