

An Integrated Computer Vision and Machine Learning Framework for Sorghum Disease Severity Assessment and Yield Prediction

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 a climate-resilient cereal of considerable importance in semi-arid agriculture, yet its productivity is constrained by foliar and stalk diseases, environmental stress, and the difficulty of obtaining timely field-level estimates of crop condition and expected yield. Conventional disease scoring and yield assessment depend heavily on visual observation, destructive sampling, and expert judgement. These practices remain valuable, but they are laborious and can vary among observers, locations, and crop stages. This paper presents an integrated, evidence-based framework that combines computer vision with machine-learning regression for two linked tasks: quantitative assessment of sorghum disease severity and prediction of grain yield. The framework is designed around field RGB imagery, optional unmanned aerial system imagery, agronomic observations, weather and soil variables, and measured yield. Disease severity is derived from segmentation of symptomatic tissue and expressed as the proportion of affected plant area, while yield prediction uses image-derived disease measures together with canopy, phenological, environmental, and management features. The paper explains data acquisition, annotation, preprocessing, segmentation, feature engineering, model training, validation, explainability, and statistical evaluation without reporting fabricated experimental results. Existing sorghum studies demonstrate that deep neural networks can quantify disease symptoms, detect panicles and grains, and extract phenotypic traits that are useful for yield modelling. The central argument is that disease assessment and yield prediction should not be treated as isolated computer-vision problems. Their integration provides a more agronomically meaningful representation of crop performance because disease burden is interpreted alongside crop structure, growth stage, and environmental conditions. The proposed framework therefore offers a reproducible methodological basis for field experimentation and decision-support research in precision sorghum production.*

Keywords: Sorghum; computer vision; disease severity; image segmentation; machine learning; yield prediction; precision agriculture; phenotyping.

I. INTRODUCTION

Sorghum is one of the major cereal crops cultivated in dry and semi-arid environments because of its comparatively strong adaptation to heat, intermittent water limitation, and diverse production systems. It is used for human food, livestock feed, fodder, brewing, and industrial purposes. Its apparent resilience, however, should not be confused with immunity to production constraints. Sorghum fields are exposed to fungal, bacterial, and viral diseases as well as drought, nutrient imbalance, insect damage, and lodging. Disease symptoms can reduce effective photosynthetic area, interfere with translocation, weaken stalks, and ultimately alter grain formation. For a farmer or breeder, the important question is therefore not merely whether a disease is present, but how much tissue is affected, how the disease progresses, and whether the observed crop condition is associated with a meaningful reduction in yield.

Field diagnosis traditionally relies on visual inspection and disease-rating scales. Expert scoring remains indispensable for confirming disease identity and creating reliable reference labels, but it has practical limitations when hundreds or thousands of plants must be assessed. Severity scores may vary among observers, particularly when symptoms are irregular, overlapping, or distributed across leaves at different canopy positions. A categorical score can also compress a continuous biological phenomenon into a small number of classes. Image analysis offers a complementary route

because a digital image can be revisited, segmented, measured, and linked to other observations. Recent reviews of plant disease severity assessment show that convolutional neural networks and segmentation methods have increasingly been used to move from simple disease recognition toward quantitative estimation of affected area (Shi et al., 2023; Macedo-Cruz et al., 2024).

The same shift is occurring in crop phenotyping and yield estimation. High-resolution RGB, multispectral, and unmanned aerial system imagery can provide repeatable measurements of canopy cover, vegetation indices, plant height, panicle number, panicle dimensions, and grain-related traits. In sorghum specifically, deep-learning pipelines have been used to detect and count panicles from aerial imagery (Lin & Guo, 2020), recognize panicles and estimate grain number from smartphone imagery (Santiago et al., 2024), and extract image-derived phenotypes for yield forecasting (Bari et al., 2026). Pugh et al. (2026) further showed that canopy-derived remote-sensing features can contribute substantially to plot-level sorghum grain-yield estimation. These studies establish that image-based phenotyping is no longer restricted to laboratory demonstrations; it can support field-scale measurement when experimental design and validation are carefully handled.

A methodological gap nevertheless remains between disease-severity research and yield-prediction research. Disease studies often end after classification or segmentation accuracy is reported, while yield studies frequently use canopy or panicle features without explicitly incorporating quantified disease burden. From an agronomic perspective, these are connected processes. Disease severity is one component of crop stress, and its effect on final yield depends on timing, susceptible tissue, genotype, weather, crop stage, and management. An integrated model can therefore be more informative than two independent systems. It can preserve the visual evidence of disease while testing whether image-derived severity adds predictive value beyond conventional crop and environmental variables.

This paper develops such an integrated framework. It is presented as a research methodology rather than as a report of an experiment that has not been conducted. Accordingly, no artificial accuracy values, invented sample sizes, or simulated findings are represented as empirical results. The framework is grounded in established computer-vision, machine-learning, plant-phenotyping, and sorghum research and is intended to guide a real field study in which predictions are evaluated against expert disease assessments and harvested grain yield. The emphasis is on scientific reproducibility, avoidance of data leakage, biologically meaningful validation, and interpretation of model outputs.

Integrated computer vision and machine-learning framework

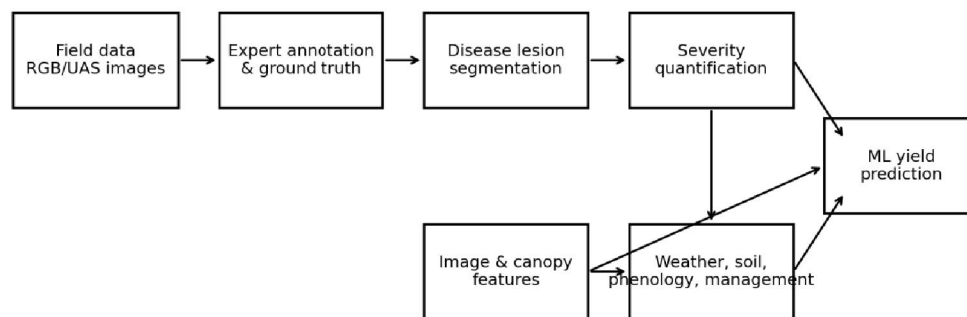


Figure 1. Integrated workflow linking image-based disease severity assessment with machine-learning yield prediction. Author-developed conceptual figure.

II. BACKGROUND AND RATIONALE

Plant disease analysis has evolved from handcrafted image processing toward data-driven representation learning. Earlier systems commonly separated the workflow into image enhancement, color-space transformation, lesion segmentation, extraction of texture or shape descriptors, and classification with algorithms such as support vector machines or random forests. These methods remain useful when datasets are modest and interpretability is important.

Deep learning altered the field by allowing convolutional networks to learn hierarchical visual features directly from images. Large-scale demonstrations such as Mohanty et al. (2016) showed the potential of deep convolutional models for plant disease recognition, while subsequent work emphasized the need to evaluate models under field conditions rather than only on clean, controlled backgrounds.

Severity assessment is more demanding than disease classification. A classifier can label an image as diseased even if only a small lesion is visible, but severity requires a quantitative estimate of how much relevant tissue is affected. This distinction makes segmentation particularly important. If L is the number of pixels assigned to symptomatic tissue and P is the number of pixels representing the plant organ under assessment, image-derived severity can be expressed as $\text{Severity (\%)} = (L/P) \times 100$. The formula is simple; obtaining trustworthy L and P is not. Shadows, senescent tissue, soil, overlapping leaves, insect injury, and illumination differences can all create pixels that resemble disease symptoms. For that reason, segmentation labels should be prepared or reviewed by trained annotators using an explicit disease-scoring protocol.

Sorghum provides a strong case for integrating severity estimation with yield modelling. Charcoal rot, anthracnose, grain mold, rust, leaf blight and other diseases can affect different organs and growth stages. Gonzalez et al. (2024), for example, used field RGB imagery and deep neural networks to classify and quantify charcoal-rot symptoms in sorghum, demonstrating that image-based symptom measurement is technically feasible under field-grown conditions. The broader implication is that disease pixels can be converted into measurable phenotypes rather than treated only as diagnostic labels.

Yield prediction is similarly a multivariable problem. Grain yield reflects genotype, plant population, crop duration, water availability, temperature, soil conditions, nutrient status, disease and pest pressure, and the formation and filling of panicles. Remote sensing compresses several of these effects into canopy and spectral features, but a model may still benefit from explicit disease information. Sorghum yield studies have reported useful associations between image-derived traits and harvested yield. In a tropical environment, multispectral vegetation indices and topographic information were successfully combined with artificial neural networks for sorghum grain-yield estimation (2024 study in Smart Agricultural Technology). More recent sorghum phenotyping work has integrated panicle detection, canopy features, and machine-learning regression, reinforcing the value of multimodal feature sets (Bari et al., 2026; Pugh et al., 2026).

The rationale for the present framework is therefore biological as well as computational. A model that predicts yield from disease severity alone would be incomplete because disease impact is conditional on crop stage and environment. Conversely, a model that ignores disease may attribute disease-related canopy deterioration to generic vegetation or spectral variables. The integrated design makes disease burden explicit, allowing its contribution to be tested while retaining other agronomic predictors.

III. OBJECTIVES AND RESEARCH QUESTIONS

The principal objective is to develop and evaluate an integrated computer-vision and machine-learning methodology for quantitative sorghum disease-severity assessment and grain-yield prediction under field conditions. The specific objectives are:

To acquire and curate field images representing healthy and diseased sorghum plants across relevant growth stages and disease-severity levels.

To segment symptomatic plant tissue and calculate continuous image-derived disease-severity measurements.

To extract complementary image-based phenotypic features, including color, texture, canopy, plant or panicle attributes where appropriate.

To integrate disease, image, agronomic, weather, soil, phenological, and management variables in machine-learning models for grain-yield prediction.

To compare classical machine-learning and deep-learning components using leakage-resistant validation and appropriate classification, segmentation, and regression metrics.

To interpret model predictions and determine whether disease-severity variables add measurable predictive information beyond non-disease features.

The study is guided by three research questions: (1) How accurately can symptomatic sorghum tissue be segmented from field RGB images when compared with expert annotations? (2) Which combination of disease, image-derived, environmental, and agronomic variables provides the most reliable prediction of harvested grain yield? (3) Does quantitative disease severity improve yield prediction after controlling for crop structure, phenology, and environmental conditions?

IV. MATERIALS AND METHODS

4.1 Research Design

The appropriate design is a quantitative, experimental and observational field-phenotyping study with repeated image acquisition and final harvest measurements. Experimental plots should be organized according to the agronomic trial being studied, such as a randomized complete block design when genotypes or treatments are compared. The computer-vision dataset and the yield-prediction dataset are related but not identical. The first contains images and pixel- or image-level disease labels; the second contains plot- or plant-level predictor variables linked to measured yield. A unique identifier should connect image, plant, plot, genotype, date, treatment, weather record and harvest observation. The sampling unit must be defined before data collection. If yield is measured per plot, model validation should occur at plot level rather than treating multiple images from the same plot as independent observations. Similarly, if repeated images are captured from the same plant, images from that plant should not be split across training and test sets. This grouped design prevents data leakage, a common source of overly optimistic performance in agricultural image studies.

4.2 Field Image Acquisition and Ground Truth

RGB images can be collected using a calibrated smartphone or digital camera at predetermined crop stages. The acquisition protocol should record date, time, camera or sensor, approximate distance, viewing angle, plant organ, growth stage, plot identifier, and disease status. Images should include natural field variation rather than only detached leaves on uniform backgrounds. A controlled subset may be collected for calibration, but external evaluation should include variable illumination, partial occlusion, overlapping leaves, and realistic backgrounds.

Where aerial phenotyping is available, UAS imagery can complement close-range images by capturing canopy-level traits and panicle distribution. Lin and Guo (2020) demonstrated the feasibility of sorghum panicle detection and counting from UAS images, and later work has combined aerial and laboratory imagery to derive yield-related features (Bari et al., 2026). Aerial data should be orthorectified where plot-level mosaics are used, and flight altitude, ground-sampling distance, overlap, sensor settings, and weather conditions should be documented.

Disease identity and severity ground truth should be established by plant pathologists or trained scorers. The annotation protocol should distinguish disease lesions from natural senescence, mechanical injury, nutrient symptoms, and insect damage. For segmentation, annotators delineate the relevant plant organ and symptomatic region. A subset should be independently annotated by a second expert to quantify inter-rater agreement. Where a standard field rating scale exists for the target disease, both categorical expert score and pixel-derived continuous severity should be retained because their agreement can be examined rather than assumed.

4.3 Image Annotation, Preprocessing and Augmentation

Annotations may be prepared as semantic masks, instance masks, bounding boxes, or image-level labels depending on the task. Disease-severity estimation benefits most directly from masks because the affected area can be calculated. If several leaves overlap, instance segmentation may be preferable when severity is required per leaf. Annotation files should be versioned, and any corrections made after expert review should be logged.

Preprocessing should be conservative. Images can be resized to the input dimensions required by the selected network and normalized using either dataset-specific statistics or the normalization scheme associated with a pretrained backbone. Excessive enhancement should be avoided because it may create color distributions that do not occur in field use. Training-time augmentation can include horizontal or vertical flipping when biologically acceptable, modest rotation, random cropping, brightness and contrast variation, and limited scale changes. Augmentation must be applied only to the training partition.

Color constancy is important because disease lesions are often distinguished partly by chromatic information. When acquisition spans different cameras or lighting conditions, a reference color target can be included periodically.

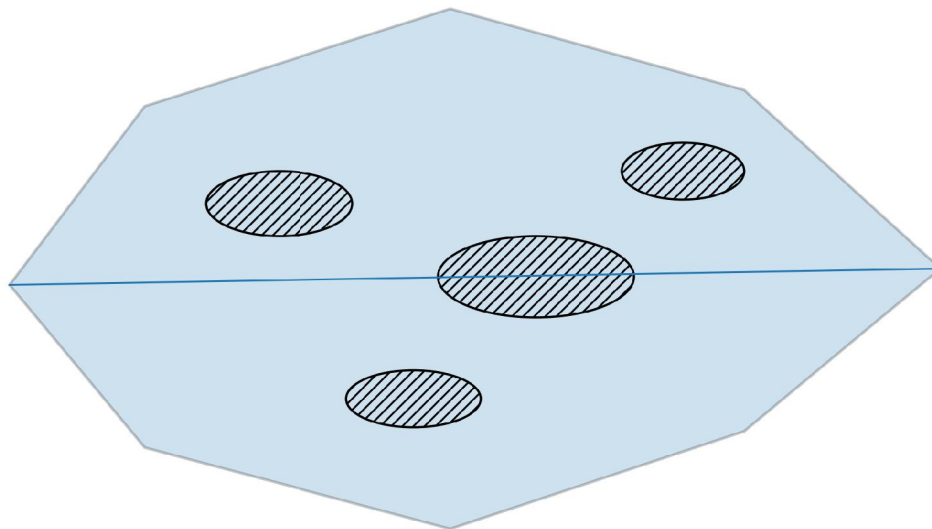
Alternatively, color-normalization procedures may be evaluated, but their effect should be reported because normalization can also suppress diagnostically useful color differences.

4.4 Computer-Vision Module for Disease Severity

The severity module consists of two linked segmentation problems: identifying the plant tissue of interest and identifying diseased tissue within it. A baseline can be constructed with conventional color and texture features followed by a classifier such as support vector machine or random forest. The deep-learning comparison can use an encoder-decoder segmentation architecture such as U-Net, DeepLabV3+, a fully convolutional network, or an instance-segmentation model where individual leaves are required. The purpose of including a classical baseline is methodological: a complex neural model should demonstrate an advantage over a simpler and more interpretable alternative rather than being assumed to be superior.

For each image i , severity is calculated as $S_i = 100 \times A_{d,i} / A_{p,i}$, where $A_{d,i}$ is the segmented diseased area and $A_{p,i}$ is the segmented assessable plant area. If severity is aggregated to plant or plot level, the aggregation rule must be specified. An area-weighted mean is generally preferable to an unweighted mean when leaves differ substantially in visible area. When disease severity is reported in ordered categories, continuous S_i can be mapped to predetermined epidemiological intervals, but the continuous value should be retained for statistical analysis.

Conceptual lesion-mask representation



$$\text{Disease severity (\%)} = \text{diseased lesion area} / \text{assessable plant area} \times 100$$

Figure 2. Conceptual representation of lesion segmentation and percentage disease-severity calculation. Author-developed schematic.

Segmentation performance should be assessed using intersection over union (IoU), Dice similarity coefficient, precision, recall, and pixel-level F1 score. Boundary-sensitive measures may be added when lesion margins are important. Severity error should be evaluated separately using mean absolute error (MAE), root mean squared error (RMSE), correlation with expert assessment, and Bland-Altman analysis or equivalent agreement analysis. High pixel accuracy alone is insufficient because a large healthy background can dominate the metric.

4.5 Feature Engineering for Yield Prediction

The yield model should use a structured feature table in which each row represents the independent prediction unit, preferably a plot or plant with a measured yield outcome. Candidate predictors are grouped below.

Feature group	Examples	Agronomic role
Disease	Mean severity, maximum severity, severity at key growth stages, lesion progression	Explicit disease burden
RGB image	Green fraction, hue statistics, texture, canopy cover, senescence fraction	Crop condition and structure
Panicle	Panicle count, visible panicle area, panicle dimensions, grain count where available	Yield-component information
Phenology	Days to flowering, growth stage, maturity class	Timing and developmental context
Weather	Rainfall, maximum/minimum temperature, humidity, heat units	Environmental stress and disease environment
Soil	Moisture, pH, organic carbon, available nutrients, elevation	Resource and site variability
Management	Sowing date, irrigation, fertilizer, treatment	Management effects
Genotype/block	Genotype or family, replication/block	Genetic and experimental structure

Handcrafted image features can include color moments in RGB, HSV or Lab spaces; vegetation or greenness indices derivable from RGB; gray-level co-occurrence matrix texture statistics; and geometric descriptors. Deep features may be extracted from a pretrained convolutional backbone, but dimensionality should be controlled through feature selection, principal components, regularization, or pooling. The use of thousands of latent features with a small number of plots creates a serious overfitting risk.

Panicle and grain traits deserve particular attention because they connect imagery to biological yield components. Santiago et al. (2024) demonstrated that smartphone imagery can support sorghum panicle recognition and grain-number estimation. Bari et al. (2026) used computer-vision models to derive panicle and seed-related traits before integrating those features into regression models. Pugh et al. (2026) found that canopy-derived feature sets produced more consistent yield estimates than panicle-only features in their breeding-program data, illustrating why the present framework treats yield as a multimodal rather than single-feature prediction problem.

4.6 Machine-Learning Models for Yield Prediction

At least one linear or regularized regression model should be included as a statistical baseline. Suitable candidates include ridge regression, LASSO and elastic net. Non-linear comparisons can include support vector regression, random forest regression, gradient-boosted trees such as XGBoost, and a multilayer perceptron when the dataset is sufficiently large. The response variable is measured grain yield expressed in a consistent unit such as kg plot⁻¹ or t ha⁻¹ after moisture correction.

Model selection should be based on cross-validated performance in the training data, not on repeated inspection of the final test set. Hyperparameters are tuned inside an inner validation loop, while the outer test fold estimates generalization. When observations come from several seasons, farms, or genotypes, the strongest validation is a grouped or external design in which an entire farm, season, or genotype group is withheld. Random image-level splitting is not adequate for claiming field generalization.

Performance should be reported with R^2 , RMSE, MAE and, where useful, normalized RMSE or mean absolute percentage error. Scatter plots of observed versus predicted yield and residual diagnostics should accompany summary metrics. Confidence intervals obtained through bootstrap resampling can show the uncertainty around model performance. A model should not be selected solely because it has the highest R^2 if it exhibits systematic bias at low or high yields.

4.7 Integration Strategy and Ablation Analysis

The central methodological test is whether quantified disease severity contributes useful information to yield prediction. This can be examined through staged feature sets: Model A uses agronomic and environmental variables; Model B adds general image and canopy features; Model C adds disease-severity variables; and Model D adds panicle

or grain features where available. Performance differences across the same validation folds provide a direct ablation test of the incremental value of each information source.

Interactions should be considered because the same severity percentage may have different consequences at different growth stages or under different water conditions. Tree-based models can learn some interactions automatically, while linear models require explicit terms such as severity $\times$ growth stage or severity $\times$ genotype. The interpretation should remain agronomically cautious: predictive association does not establish that a particular percentage of disease caused a specific yield loss unless the experimental design supports causal inference.

4.8 Explainability and Model Interpretation

Interpretability is necessary for both modules. For image models, Grad-CAM or related saliency methods can be used as a diagnostic aid to determine whether classification or feature extraction is responding to symptomatic tissue rather than irrelevant background. For segmentation, the predicted mask itself is inherently more interpretable because the spatial output can be compared with expert annotation. Explainability maps should not be treated as proof of causal reasoning; they are supporting evidence about model attention.

For yield models, permutation importance and SHAP values can identify variables associated with individual and global predictions. Lundberg and Lee (2017) provided the widely used SHAP framework for additive feature attribution. Interpretation should emphasize consistency across folds and biological plausibility. If disease severity appears highly influential only in one season or one genotype, that instability should be reported rather than concealed by an overall average.

4.9 Statistical Analysis and Reproducibility

Descriptive statistics should summarize disease severity, yield, weather, and key phenotypic variables by season, genotype and treatment. Relationships between expert severity and image-derived severity can be examined with Pearson or Spearman correlation according to distributional assumptions, but correlation should be complemented by agreement analysis. Mixed-effects models may be used to account for block, season, location, or genotype structure when the inferential objective extends beyond prediction.

Reproducibility requires a fixed data dictionary, documented preprocessing, version-controlled code, recorded random seeds, saved model configurations, and a clear description of hardware and software. Dataset partitions should be stored as identifiers so that every model is evaluated on the same units. Where ethical and ownership constraints permit, anonymized images, annotations, code, and metadata should be deposited in a research repository. Farm geolocation and personally identifying information should not be released without permission.

V. PROPOSED ANALYTICAL WORKFLOW

The complete workflow can be implemented in seven operational stages:

Field acquisition: collect close-range RGB images, optional UAS imagery, expert disease observations, phenology, weather, soil and management data.

Ground-truth construction: annotate plant tissue and disease lesions; verify labels; link every image to plant, plot, date and genotype identifiers.

Vision modelling: train and validate segmentation models; generate disease masks; calculate continuous severity and other image-derived phenotypes.

Feature integration: aggregate image features to the yield-prediction unit and merge them with agronomic and environmental records.

Predictive modelling: fit baseline and non-linear regression models using nested or grouped validation.

Ablation and interpretation: compare feature groups, inspect SHAP or permutation importance, and test the incremental value of disease severity.

External validation: evaluate the locked pipeline on a genuinely unseen season, farm, genotype group or independent dataset before making claims of field generalizability.

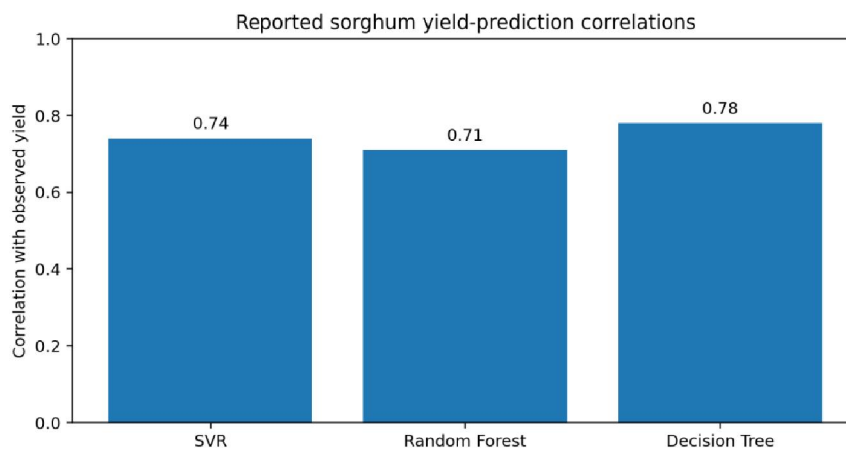
VI. EVALUATION CRITERIA

Task	Primary metrics	Reason
Disease segmentation	IoU, Dice, precision, recall, F1	Measures overlap and lesion-detection quality
Severity estimation	MAE, RMSE, correlation, agreement	Measures error in continuous severity
Disease class (if used)	Accuracy, macro-F1, per-class recall	Handles class-specific diagnostic performance
Yield prediction	R^2 , RMSE, MAE	Measures explained variance and prediction error
Generalization	External/grouped test performance	Tests transfer beyond familiar plots or seasons
Interpretability	Feature importance stability; visual mask review	Checks biological plausibility and model reliance

VII. DISCUSSION

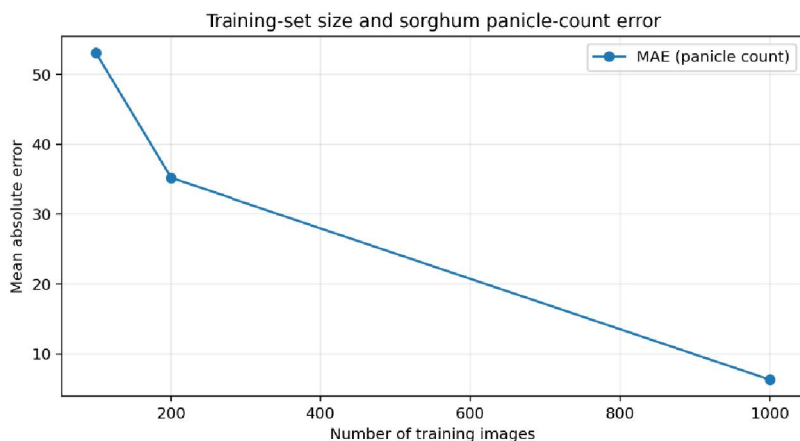
The integrated framework addresses a practical weakness in many agricultural artificial-intelligence studies: the separation of visually detectable stress from the final agronomic outcome. A disease classifier may be technically impressive but of limited management value if it cannot indicate severity. Likewise, a yield model may predict accurately within one dataset while offering little insight into the crop processes represented by its predictors. By converting disease symptoms into a quantitative phenotype and carrying that phenotype into the yield model, the framework creates a transparent link between crop health observation and productivity assessment.

Evidence from sorghum research supports each component of the pipeline. Lin and Guo (2020) showed that deep learning can support high-throughput panicle detection from UAS imagery. Gonzalez et al. (2024) demonstrated field-image classification and segmentation for sorghum charcoal-rot symptoms, with EfficientNet-B3 and a fully convolutional network among the strongest models in their comparison. Santiago et al. (2024) extended image-based phenotyping to smartphone recognition of sorghum panicles and grain counting. Bari et al. (2026) combined computer vision with regression for sorghum yield forecasting, while Pugh et al. (2026) integrated remote-sensing features and machine-learning models in a breeding program. Together, these studies indicate that the required technical pieces already exist; the research opportunity lies in integrating them under a coherent, leakage-resistant experimental design.



Source: Bari et al. (2026). Values reproduced from the article abstract.

Figure 4. Reported correlations between observed and predicted sorghum yield for three regression models in Bari et al. (2026).



Source: Lin & Guo (2020). Selected values reported in the published study.

Figure 3. Effect of training-set size on sorghum panicle-count MAE using selected published values from Lin and Guo (2020).

A major strength of segmentation-based severity is that it preserves spatial evidence. An expert can inspect whether the predicted lesion mask corresponds to actual symptoms. This is preferable to relying exclusively on a black-box severity class. Nevertheless, segmentation is not automatically objective. The model learns the annotation conventions used during training, and experts themselves may disagree at lesion boundaries. Inter-rater analysis and explicit annotation guidelines are therefore part of the measurement system, not optional administrative details.

The yield component also requires restraint in interpretation. Machine-learning models can exploit correlations that are specific to a season, camera, genotype set or field. For example, a model might associate a background color or planting pattern with yield if those visual cues are unevenly distributed across experimental treatments. Grouped validation, external testing and ablation studies reduce this risk. Pugh et al. (2026) reported that canopy-based features were more consistently useful than panicle-only features in their data, whereas Bari et al. (2026) showed that image-derived yield attributes could support predictive modelling. The difference is instructive: the best feature set is context dependent, so a research paper should report comparative evidence rather than assume one universal model.

Disease severity should be interpreted in time. A late-season lesion that develops after most grain filling may have a different yield relationship from a similar percentage observed before flowering. Repeated imaging therefore has greater scientific value than a single snapshot. Longitudinal severity can be summarized through maximum severity, rate of progression, area under a disease-progress curve, or severity at biologically important stages. These temporal features can then be tested against simpler single-date measures.

Another important issue is scale. Close-range leaf images are well suited to lesion segmentation, whereas plot-level yield is influenced by stand density, tillering, panicle number and spatial variability. Aggregating leaf-level severity to plot level must therefore be designed carefully. Sampling only the most visibly diseased leaves would bias severity upward; sampling only upper-canopy leaves could miss diseases concentrated elsewhere. A standardized plant and leaf sampling protocol is necessary so that image-derived severity represents the experimental unit whose yield is measured. The framework also provides a reasonable balance between classical machine learning and deep learning. Deep networks are particularly useful for segmentation and object detection because they learn complex visual representations. Yield tables, however, are often modest in sample size compared with image datasets. In such cases, regularized regression, support vector regression, random forest and gradient boosting may be more appropriate than a large end-to-end neural network. The methodological principle is to match model capacity to sample size and data structure rather than select algorithms because they are fashionable.

Human expertise remains central. Plant pathologists define valid symptom categories, agronomists determine meaningful sampling stages, breeders interpret genotype effects, and field technicians establish reliable harvest measurements. Computer vision does not replace these roles; it converts expert-defined visual phenomena into

repeatable digital measurements. This human-in-the-loop approach is especially important when symptoms overlap or when abiotic stress resembles disease.

From a practical perspective, the framework can be implemented with relatively accessible RGB cameras for the disease module and conventional agronomic records for the yield module. UAS and multispectral data can improve spatial coverage but are not mandatory. This makes the design adaptable to institutions with different levels of infrastructure. The essential requirements are reliable ground truth, consistent identifiers, representative sampling, and rigorous validation.

The paper deliberately does not present a fabricated results section. A methodological framework becomes scientifically useful only when it is tested on observed data and when performance metrics are derived from locked test observations. Reporting invented accuracy, R^2 or yield values would undermine the research norms that the framework is intended to support. Once field data are available, the analysis plan described here can be executed and the resulting tables, figures, confidence intervals, error analyses and model comparisons can be reported transparently.

VIII. LIMITATIONS

The framework has several limitations. First, RGB symptoms are not always pathogen-specific. Different diseases and abiotic stresses can produce similar discoloration or necrosis, and laboratory confirmation may be required for ambiguous cases. Second, visible severity does not capture latent infection or internal stalk damage unless the target disease produces externally observable features. Third, environmental variation in illumination and canopy structure can reduce segmentation accuracy. Fourth, yield is influenced by many factors that may not be measured, creating irreducible prediction error.

Generalization is also a central challenge. A model trained on one cultivar, camera, region or season may not transfer reliably to another. External validation is therefore more important than marginal improvements on an internal random split. Finally, interpretability methods such as SHAP and saliency maps describe model behavior but do not establish biological causation. Causal claims about disease-induced yield loss require experimental or epidemiological designs capable of separating disease effects from correlated environmental and management factors.

IX. CONCLUSION

An integrated computer-vision and machine-learning framework provides a scientifically coherent way to connect sorghum disease measurement with grain-yield prediction. The disease component transforms field images into quantitative estimates of affected tissue through segmentation, while the yield component combines those estimates with canopy, panicle, phenological, environmental, soil and management variables. The integration is valuable because it treats disease severity as an agronomic phenotype rather than an isolated image label.

Existing research demonstrates that sorghum panicles, grains, canopy traits and disease symptoms can be extracted from RGB or aerial imagery and incorporated into predictive analyses. The next methodological requirement is rigorous integration: expert ground truth, grouped data partitioning, transparent feature aggregation, comparison with simple baselines, ablation analysis, uncertainty reporting and external validation. When these principles are followed, the resulting system can support reproducible phenotyping research and provide a defensible basis for decision-support tools without overstating what artificial intelligence can infer from images alone.

REFERENCES

1. Bari, M. A. A., Bakshi, A., Choton, J. C., Pramanik, S., Witt, T. D., Caragea, D., Bean, S., Jagadish, S. V. K., & Felderhoff, T. (2026). Deep learning for sorghum yield forecasting using uncrewed aerial systems and lab-derived imagery. *Plant Phenomics*, 8(1), 100133. <https://doi.org/10.1016/j.plaphe.2025.100133>
2. Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145, 311–318. <https://doi.org/10.1016/j.compag.2018.01.009>
3. Gonzalez, J. B., et al. (2024). Quantifying leaf symptoms of sorghum charcoal rot in images of field-grown plants using deep neural networks. *The Plant Phenome Journal*, 7, e20110. <https://doi.org/10.1002/ppj2.20110>

4. 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. <https://doi.org/10.1109/CVPR.2016.90>
5. Lin, Z., & Guo, W. (2020). Sorghum panicle detection and counting using unmanned aerial system images and deep learning. *Frontiers in Plant Science*, 11, 534853. <https://doi.org/10.3389/fpls.2020.534853>
6. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
7. 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>
8. Pugh, N. A., Young, A., Nesbitt, M., & Hayes, C. (2026). Remote sensing data and machine learning models estimate sorghum grain yield in a plant breeding program. *Plant Phenomics*, 8(2), 100179. <https://doi.org/10.1016/j.plaphe.2026.100179>
9. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28
10. Sajitha, P., Andrushia, A. D., Anand, N., & Naser, M. Z. (2024). A review on machine learning and deep learning image-based plant disease classification for industrial farming systems. *Journal of Industrial Information Integration*, 38, 100572. <https://doi.org/10.1016/j.jii.2024.100572>
11. Santiago, G. N., et al. (2024). Deep learning methods using imagery from a smartphone for recognizing sorghum panicles and counting grains at a plant level. *Plant Phenomics*.
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. <https://doi.org/10.1109/ICCV.2017.74>
13. Shi, T., Liu, Y., Zheng, X., Hu, K., Huang, H., Liu, H., & Huang, H. (2023). Recent advances in plant disease severity assessment using convolutional neural networks. *Scientific Reports*, 13, 2336. <https://doi.org/10.1038/s41598-023-29230-7>
14. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6, 60. <https://doi.org/10.1186/s40537-019-0197-0>
15. 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. <https://doi.org/10.1016/j.compag.2018.03.032>
16. Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
17. Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
18. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297. <https://doi.org/10.1007/BF00994018>
19. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. <https://doi.org/10.1145/2939672.2939785>
20. Lei, L., Yang, Q., Yang, L., Shen, T., Wang, R., & Fu, C. (2024). Deep learning implementation of image segmentation in agricultural applications: A comprehensive review. *Artificial Intelligence Review*, 57, 149. <https://doi.org/10.1007/s10462-024-10775-6>
21. Faye, D., Diop, I., Mbaye, N., Dione, D., & Diedhiou, M. (2023). Plant disease severity assessment based on machine learning and deep learning: A survey. *Journal of Computer and Communications*, 11, 57–75. <https://doi.org/10.4236/jcc.2023.119004>
22. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations*.
23. Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

24. Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1251–1258. <https://doi.org/10.1109/CVPR.2017.195>
25. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning, 6105–6114.