JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AUDITING KRISHIMITR AND PILOTING CROP-SPECIFIC LEAF IMAGE CLASSIFIERS UNDER DATA AND COMPUTATIONAL CONSTRAINTS

Ayush Kumar Pandey, Ayush Mishra, Ayush Pandey,
Ashwani Singh, Prof- Aliza Raza Rizvi
1,2,3,4 Students, Department Of Computer Science And Engineering, Babu Banarasi Das Northern
India Institute Of Technology, Lucknow, India.
1Assistant Professor, Department Of Computer Science And Engineering, Babu Banarasi Das Northern
India Institute Of Technology, Lucknow, India.

ABSTRACT

KrishiMitr is a web application prototype that combines plant-leaf disease classification, crop recommendations, weather and rainfall guidance, and an agricultural conversational assistant. We audited the project artifacts and assessed whether available image data could support reproducible disease-classification experiments. RiceLeafDiseaseBD V3 was excluded because its provenance, labels, and unresolved duplicate relationships do not support a defensible modeling cohort. A pinned checkout of the PlantVillage tomato folders provided 4,500 readable 256 x 256 images across healthy, early-blight and late-blight source folders. Exact-duplicate handling and source leaf-group metadata produced a 2,918-image candidate cohort, split by source group into 2,042 training, 436 validation and 440 test images. A separately maintained wheat dataset supplied an existing 259-image healthy-versus-stripe-rust split. Because a 224-pixel model run exceeded practical CPU time, we trained one scratch-initialized compact CNN per task using 128-pixel inputs and a three-epoch cap. Test accuracy was 0.923 on wheat (n=39) and 0.911 on tomato (n=440). The tomato split has 12 unresolved same-class perceptual candidate pairs spanning partitions, so its score is a pilot result rather than a certified leakage-free estimate. Neither task has independent pathologist-confirmed labels or external field validation. Existing application models, synthetic crop/rainfall training paths and unverified checkpoints do not support predictive performance claims. The evidence supports a transparent prototype and limited source-label benchmark only, not a validated agricultural diagnostic service.

Index Terms : agricultural decision support; plant disease image classification; dataset audit; data leakage; reproducibility; controlled imagery

I. INTRODUCTION

Digital tools can combine several information workflows, but software integration alone does not demonstrate agronomic accuracy or usefulness. KrishiMitr is presented in the repository as a multi-module platform. Its code includes image-based plant disease handling, crop recommendation, weather and rainfall components, and an optional agricultural assistant. This study audits the available data and implementation evidence, then reports a narrow, reproducible pilot for two separate crop-specific image tasks.

Plant-leaf classifiers can perform well on curated, controlled image collections, but controlled capture and regularized backgrounds can create a gap between benchmark performance and field conditions. Mohanty et al. describe the PlantVillage image dataset as collected under controlled conditions and discuss concerns about background information and regularized collection [2, 4]. Consequently, performance on such imagery should not be treated as field validation.

The original KrishiMitr project outline proposed a combined seven-class disease task spanning wheat, rice and tomato. The available evidence did not justify that scope: rice cohort conflicts remain unresolved, tomato sources previously pursued were inaccessible, and the production API's output taxonomy did not match the proposed taxonomy. We therefore report two independent tasks and exclude rice. We do not pool “healthy” across plant species.

A. CONTRIBUTIONS

1. A project-wide evidence audit identifying which existing model and data artifacts can support research claims.
2. A read-only acquisition and audit of a pinned PlantVillage tomato source subset, including exact duplicate, perceptual candidate, and source-group checks.

3. A conservative separation of wheat and tomato into independent source-label tasks, with frozen split inputs and a documented CPU-limited pilot.
4. Explicit reporting of unresolved labels, licensing, source-group completeness, compute limitations, and potential residual leakage.

II. MATERIALS AND METHODS

A. PROJECT AUDIT

The repository audit inspected source code, model artifact references, dataset manifests, split files, reports, and configuration without reading secret environment values. The FastAPI backend exposes disease, weather, rainfall, crop recommendation, and assistant routes; the frontend is React/Vite. Static inspection found the disease trainer initializes and saves a new model rather than demonstrating a trained checkpoint. Existing `.pt` and `.h5` model files lack verified training provenance. Disease inference code combines CNN outputs with fixed color heuristics, uses four output labels that do not match the seven-class research proposal, and clamps reported confidence. Crop recommendation training uses rule-generated labels and random perturbations. Rainfall training can default to synthetic records or fall back to synthetic data. Thus, the application is treated as a prototype; no real-world predictive claims are inferred from its model artifacts or synthetic fit scores.

B. WHEAT IMAGE TASK

The source record advertises 208 stripe-rust, 102 healthy, and 97 Septoria images and describes collection at Holeta wheat farm, Ethiopia, using a Canon EOS 5D Mark III, with class sorting assisted by plant pathologists [1]. The 97 Septoria images were excluded because Septoria falls outside the selected binary classification task, which includes only Healthy and Yellow/Stripe Rust. This study uses only the existing audited Healthy and Yellow/Stripe Rust split. The split has 181 training, 39 validation, and 39 test records, after 50 pending expert-review exclusions and one exact duplicate-copy exclusion. The known two-image related rust pair is grouped within train. All 310 eligible source manifest rows were independently hash-checked in the split audit; no plant or capture-session identifiers exist. The split generator and its limitations are documented in `wheat_independent_split_audit.md`.

C. TOMATO IMAGE TASK

The official researcher-maintained PlantVillage repository [3] was checked out sparsely at commit `7f7ecc7e1eaca78107e3affe7cb5abd9427e139a`. The broader PlantVillage image collection is documented in an open-access repository paper [4]. Only raw/color Tomato Healthy, Early Blight, and Late Blight source folders were used. The local scan found 1,591, 1,000, and 1,909 JPEGs, respectively (4,500 total); all were decoded and measured at 256 x 256. Encoded-byte SHA-256 and decoded-RGB hashes identified 14 same-class exact-duplicate pairs (28 affected records, 14 duplicate-copy exclusions); no cross-class exact identity was found.

Perceptual screening used 64-bit pHash and dHash with a candidate threshold of Hamming distance ≤8 for either hash. It produced 313 candidate pairs. Contact-sheet screening covered cross-class candidates and the highest-priority within-class pairs. Seventeen of 18 cross-class pairs appeared visually distinct; one filename-linked pair and its complete connected component remain quarantined. Four filename-linked, non-exact within-class pairs were flagged for expert/source review. Other candidate pairs remain unreviewed. Similarity alone was not used to delete or relabel any image.

The source leaf map yielded 730 strict, class-consistent groups for 2,918 candidate records (999 Healthy, 999 Early Blight, 920 Late Blight), all with status `ELIGIBLE_PENDING_LABEL_VALIDATION`. A further 1,565 records lack a unique group map, and three unique records are quarantined in the unresolved cross-class component; 14 duplicate copies are excluded. The selected cohort split by whole source leaf groups with seed 42 into 2,042 train, 436 validation and 440 test images. Partition group IDs were disjoint. The split algorithm greedily assigned whole groups within class to minimize normalized deviation from 70/15/15 image-count targets; exact fractions differ due to group sizes.

The source-linked dataset card [3] declares CC BY-SA 3.0. The original paper's CC BY 4.0 license applies to the article and does not alone establish rights in each image. Underlying image-level rights were not independently verified; redistribution and commercial use require institutional rights review. See the source-specific provenance and licensing audit.

D. EXPERIMENTAL PROTOCOL

The initial intended comparison of a compact CNN, ResNet-18 and MobileNetV3-Small at 224×224 and up to 30 epochs proved impractical: after approximately 31 minutes, the first wheat model had not completed. It was stopped before test data were opened; its partial checkpoint is preserved but excluded. Before test access, the executable protocol was amended to one compact CNN per task, 128 x 128 inputs, maximum three epochs, patience two, and CPU execution. This amendment removes architecture-comparison claims.

Images were decoded as RGB, resized bilinearly to 128 x 128, converted to `[0,1]`, and normalized with ImageNet channel means and standard deviations. Training-only augmentation used horizontal flip and brightness/contrast jitter up to ±0.10, each with probability 0.5. The compact CNN has four convolutional blocks (32, 64, 128 and 192 channels), batch normalization, ReLU and max pooling, adaptive average pooling, dropout and a linear classifier. It was initialized from scratch. Optimization used AdamW (learning rate 0.001; weight decay 0.0001), cross-entropy, seed 42, batch sizes 32 (wheat) and 64 (tomato), and deterministic PyTorch algorithms on CPU.

The model with the best validation macro-F1 was selected; validation loss broke ties. After both task models had completed training and validation selection, the test partitions were scored. Metrics include accuracy, balanced accuracy, macro and per-class precision/recall/F1, confusion matrix, ROC-AUC when defined, 10,000-replicate class-stratified bootstrap intervals for accuracy and macro-F1, parameter count, checkpoint size and measured CPU latency. The completed run read test CSV metadata before training to build a sorted class-index map and compute split hashes; no test images or performance metrics affected gradients or validation choices, and every class was represented in training. A separate forward-only pass over up to 100 test images was then made for latency; those outputs were not used for selection or metric computation. These are procedural deviations from strict test isolation and single-pass access. The reusable training script has since been corrected to defer reading the test CSV until all training/validation runs finish; the existing results were not rerun. The bootstrap describes image-sampling variability only and does not represent label or provenance uncertainty.

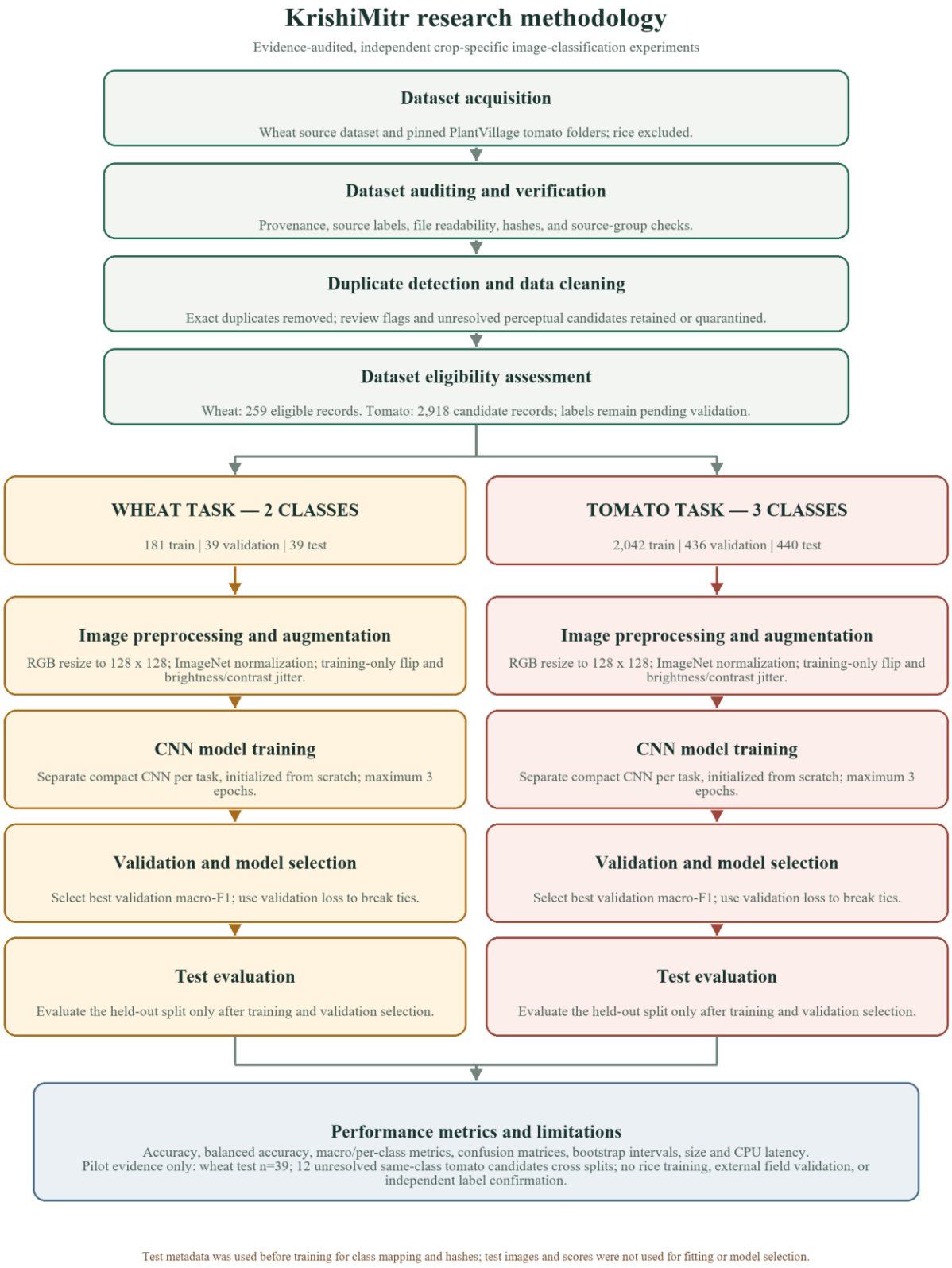

Figure 1. Evidence-audited methodology and independent wheat and tomato evaluation pathways. Rice was excluded from model training.

E. REPRODUCIBILITY

The project stores source manifests, split CSVs and their hashes, the pinned PlantVillage commit, per-file source SHA-256 values, scripts, audit reports, methodology, training histories, checkpoints, and evaluation JSON. The Python environment was Windows 11, Python 3.14.3, PyTorch 2.13.0+cpu, torchvision 0.28.0+cpu, NumPy 2.4.6, Pillow 12.3.0, SciPy 1.18.0 and scikit-learn 1.9.0; CUDA was unavailable. Run metadata records seed, split hashes, methodology hash, package versions, checkpoint hashes, and output hashes. Test predictions were not saved per image, limiting later error auditing.

III. RESULTS

A. DATASET AUDIT

All 4,500 PlantVillage tomato files were readable. All were 256 x 256 JPEGs; none contained a camera model in EXIF. The group map covered 999 of 1,591 Healthy, 999 of 1,000 Early Blight, and 920 of 1,909 Late Blight records. The total candidate cohort contained 730 mapped groups. The cross-class candidate screen contained 18 pairs in 17 connected components; 17 pairs were judged

visually distinct, and one filename-linked relationship remained unresolved. Among eligible records, 12 same-class perceptual candidate pairs have endpoints in different tomato partitions. They have different source group IDs but were not visually adjudicated; this leaves residual leakage risk.

B. PILOT CLASSIFICATION RESULTS

Both tasks selected epoch 2 by validation macro-F1. Results are from one seed, one small model, and a three-epoch compute-bounded run.

Table 1. Pilot test performance by crop-specific task.

Task	n	Accuracy	Balanced accuracy	Macro precision	Macro recall	Macro F1	95% bootstrap CI for macro F1
Wheat Healthy / Yellow Rust	39	0.923	0.940	0.912	0.940	0.920	0.819 to 1.000
Tomato Early Blight / Healthy / Late Blight	440	0.911	0.905	0.929	0.905	0.907	0.879 to 0.932

Wheat confusion matrix in `[Healthy, Yellow Rust]` order was `[[14, 0], [3, 22]]`. Tomato confusion matrix in `[Early Blight, Healthy, Late Blight]` order was `[[150, 1, 1], [0, 152, 0], [37, 0, 99]]`. Tomato Late Blight recall was 0.728; 37 of 136 source-labeled Late Blight test images were predicted as Early Blight. This cannot establish whether the model or source labels explain those errors.

The wheat test set has only 39 examples and produces broad uncertainty. Tomato's larger test set still has unresolved cross-partition same-class perceptual candidates; its score is a pilot, not a certified leakage-free estimate. Scores across the two crop tasks are not directly comparable.

IV. DISCUSSION

The repository contains a functioning application structure but not enough validated evidence to claim a complete agricultural decision-support system. The crop recommendation and rainfall training paths rely on synthetic or rule-generated targets, the disease API taxonomy differs from the research tasks, and existing disease model provenance is not established. These components require real outcome datasets, clearly defined target variables, and independent evaluation before any empirical efficacy claim.

The image pilot demonstrates that the stored splits and a compact scratch CNN can be executed reproducibly, but the results should not be overinterpreted. Wheat's small test set, lack of plant/session IDs, and unverified image-level labels limit inferential value. PlantVillage's leaf groups improve leakage control, but mapping is incomplete and the post-split perceptual audit identified 12 same-class candidate pairs straddling partitions. These pairs were not confirmed duplicates, but until adjudicated, the holdout cannot be called fully independent. The controlled imagery and source-folder labels also do not test field generalization or expert diagnostic performance.

The key contribution is therefore the evidence boundary and traceable data audit, not a novel model or state-of-the-art result. A confirmatory study should first resolve the tomato candidate pairs, obtain plant-pathologist review of a blinded sample or use authoritative annotations, verify image rights, then pre-register a fresh group-aware evaluation on new or untouched field data. Multiple random seeds and a broader model comparison should follow only with sufficient compute and a prespecified resource budget.

V. LIMITATIONS AND ETHICS

1. Labels are source-assigned and have not been independently verified by this project; visual duplicate review did not diagnose disease.
2. PlantVillage source groups are missing or ambiguous for 1,582 records; no camera EXIF was present in this tomato subset.
3. Twelve same-class tomato perceptual candidate pairs cross partition boundaries and remain unresolved, compromising a strong leakage-free test claim.
4. Wheat has only 39 test images and no plant/session identifiers.
5. Only one seed and one compact CNN per task were completed; the planned multi-architecture comparison was abandoned for compute reasons.
6. No per-image test prediction artifact was saved, limiting error analysis without another evaluation pass.
7. No external field dataset, user study, API runtime evaluation, calibration study, robustness evaluation, or clinical/agronomic utility study was performed.
8. PlantVillage's dataset card declares CC BY-SA 3.0, but image-level rights are not established; distribution rights need institutional review.
9. The rice dataset was excluded, and crop recommendation and rainfall claims remain unsupported by verified real-world training data.

The study used public image datasets and did not recruit participants. This work does not provide agronomic advice or disease diagnoses and should not be deployed for farmer decisions.

VI. CONCLUSION

KrishiMitr remains an application prototype whose system-level predictive performance has not been validated. A narrow wheat and PlantVillage tomato image-classification pilot was completed with source-specific cohort handling and reproducibility records. The resulting test scores are exploratory: wheat uncertainty is high, tomato has unresolved cross-partition candidate pairs, both datasets lack this project's independent label validation, and field generalization was not measured. A defensible next paper should

center on the audit and prototype status unless rights, label, leakage, external-data, and compute limitations are resolved before a confirmatory experiment.

GitHub repo :- https://github.com/ayushpandey3357/KRISHIMITR

VII. DATA AND CODE AVAILABILITY

Audit scripts and reports are contained in this project repository. PlantVillage files are a sparse checkout at the cited commit; users should obtain and attribute the dataset according to the source's current terms rather than redistribute the local subset. Wheat source data are governed by the source record's terms. The generated split and experiment manifests are included with the project. No test predictions by image are published in this run.

REFERENCES

[1]. Getachew, H. (2021). *Wheat Leaf Dataset*, Version 1. Mendeley Data. https://doi.org/10.17632/wgd66f8n6h.1.

[2]. 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.

[3]. Mohanty, S. P. (maintainer). *PlantVillage-Dataset*. GitHub repository, commit `7f7ecc7e1eaca78107e3affe7cb5abd9427e139a`. https://github.com/spMohanty/PlantVillage-Dataset/tree/7f7ecc7e1eaca78107e3affe7cb5abd9427e139a.

[4]. Hughes, D. P., & Salathé, M. (2015; version 2 revised 2016). An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv:1511.08060v2* [cs.CY]. https://doi.org/10.48550/arXiv.1511.08060.

[5]. Mohanty, S. P. (maintainer). PlantVillage Dataset Card (README_HF.md), repository commit 7f7ecc7e1eaca78107e3affe7cb5abd9427e139a. URL:https://github.com/spMohanty/PlantVillage-Dataset/blob/7f7ecc7e1eaca78107e3affe7cb5abd9427e139a/README_HF.md.