Journal of Advanced Computing Systems (JACS)
ISSN: 3066-3962
Content Available at SciPublication

Lightweight Multimodal Rice Leaf Disease Screening and Nepal-Localized Climate-Smart Advisory Support: Evaluations on Real Rice Images, Nepal SOTER Soil Profiles, and WorldClim Climate Data

Lumeng Han[1], Derek Zhou[2]
[1] *International Agricultural Development, University of California, Davis, CA, USA*
[2] *Computer Science, University of California, Berkeley, Berkeley, CA, USA*
lmhan@ucdavis.edu

DOI: 10.69987/JACS.2023.30405

Keywords

rice leaf disease; Nepal agriculture; climate-smart advisory; image classification; retrieval-augmented generation; soil-climate risk; SOTER; WorldClim; TF-IDF.

Abstract

This study evaluates a lightweight multimodal pipeline for rice leaf disease screening and Nepal-localized climate-smart advisory support. The revised evaluation uses directly downloadable public datasets rather than procedurally generated image or soil-weather records. The vision branch uses the UCI Rice Leaf Diseases dataset, which contains 120 real JPG images in three balanced classes: bacterial leaf blight, brown spot and leaf smut. The environmental branch combines Nepal SOTER soil profile data with WorldClim 2.1 monthly climate rasters for precipitation, minimum temperature, maximum temperature, water vapor pressure and elevation, producing 645 profile-season records from 129 georeferenced Nepal soil profiles. The advisory branch uses 24 curated English-Nepali rice management cards and 96 farmer-style retrieval queries grounded in Nepal rice disease, soil fertility and crop-management references. Five image classifiers were compared using 136 engineered color, lesion and edge features. Logistic regression achieved the strongest clean hold-out performance, with 0.917 accuracy and 0.918 macro-F1. Stress tests showed that the model is reasonably stable to small rotations and moderate blur, but accuracy declines under strong sensor noise and large brightness shifts. In the advisory retrieval experiment, character TF-IDF performed best with 0.875 top-1 accuracy, 0.969 top-3 accuracy and 0.922 mean reciprocal rank. For the soil-climate risk task, random forest achieved the best hold-out macro-F1 of 0.869 and cross-validated macro-F1 of 0.851. The results support the paper's central claim: a small, inspectable system can connect leaf-image evidence, Nepal soil-climate context and grounded farmer advice, while remaining explicit about the limits of non-field diagnostic evaluation.

1. Introduction

Rice disease screening is a useful applied machine-learning problem because leaf diseases affect farmer income, food security and the efficiency of fertilizer and pesticide use. Many image-based plant disease studies evaluate larger deep-learning models, but smallholder advisory tools also need transparent behavior, local agronomic context and clear limits on the advice that follows a visual prediction. Nepal is a relevant setting for this kind of evaluation because rice is central to production, food security and rural livelihoods, and Nepal rice pathology references repeatedly discuss blast, bacterial blight, brown spot, seed health, nutrient balance and field sanitation as practical management issues.

The revised study therefore keeps the architecture lightweight but strengthens the evidence base. The image experiment now uses real rice leaf images from the UCI Rice Leaf Diseases dataset. The environmental experiment replaces a constructed table with a derived soil-climate table built from Nepal SOTER soil profiles and WorldClim climate rasters. The advisory layer remains retrieval-based so that a farmer query returns a traceable card rather than an unconstrained generated answer. Figure 1 summarizes this pipeline, and Table 1 lists the datasets used in the evaluation.

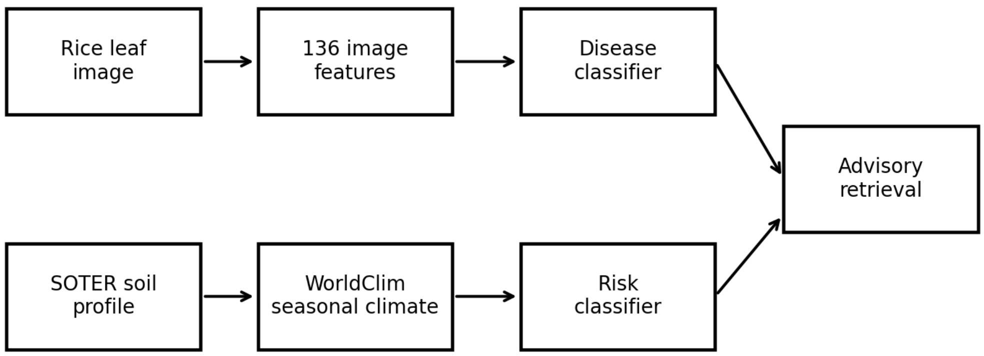

Figure 1. Multimodal rice disease and Nepal-localized advisory pipeline.

Table 1. Datasets used for empirical evaluation.

Dataset	Rows or images	Variables	Task	Source file
UCI Rice Leaf Diseases	120 images	JPG leaf images; 3 disease classes; 40 images per class	3-class image classification	rice+leaf+diseases.zip
Nepal SOTER + WorldClim soil-climate table	645 records	SOTER topsoil properties + monthly precipitation, temperature, vapor pressure and elevation	Soil-climate risk classification	NP-SOTER.zip; wc2.1_10m_*.zip
Advisory knowledge cards	24 cards	English-Nepali rice disease, soil, weather and IPM advice	Retrieval grounding	Curated from Nepal rice references

Advisory evaluation queries	96 queries	Four farmer-style queries per advisory card	Top-k retrieval evaluation	Curated query set

2. Materials and Methods

2.1 Data sources and preprocessing

The image branch uses the UCI Rice Leaf Diseases dataset. As shown in Table 2 and Figure 2, the dataset is balanced across bacterial leaf blight, brown spot and leaf smut. Images were resized to 224 x 224 pixels and converted to RGB before feature extraction. The class balance is intentionally preserved in the stratified split so that macro-F1 can be interpreted directly across the three disease labels.

Table 2. UCI Rice Leaf Diseases class distribution.

label	class_name	images
bacterial_leaf_blight	Bacterial leaf blight	40
brown_spot	Brown spot	40
leaf_smut	Leaf smut	40

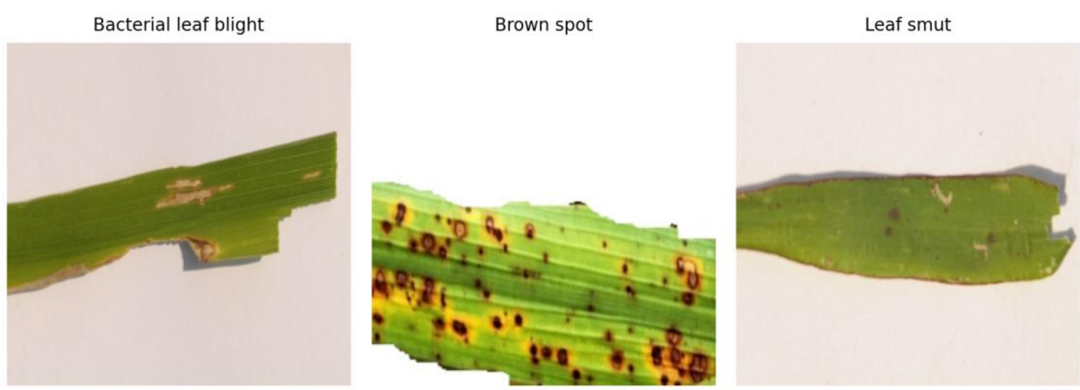

Figure 2. Sample images from the UCI Rice Leaf Diseases dataset.

The soil-climate branch uses Nepal SOTER profile data and WorldClim 2.1 climate rasters. Topsoil values were taken from the first available horizon of georeferenced profiles. Monthly WorldClim rasters were sampled at each profile coordinate and summarized into five rice-relevant seasonal windows: pre-monsoon, early monsoon, peak monsoon, late monsoon and post-monsoon. Precipitation was summed within each window, while temperature and vapor pressure were averaged. A relative-humidity proxy and dew-risk-hours proxy were derived from vapor pressure and temperature to represent disease-conducive moisture conditions.

Table 3. Public-source alignment and experimental use.

Component	Public source	Experimental use

Image disease classes	UCI Rice Leaf Diseases: 120 JPG images; bacterial leaf blight, brown spot and leaf smut; 40 per class	Used directly for all image training, testing and robustness measurements.
Soil properties	ISRIC Nepal SOTER: landform and soil-property database with representative Nepal profiles	Used to obtain georeferenced topsoil pH, organic carbon, total nitrogen, phosphorus, exchangeable potassium, texture and water-holding capacity.
Climate variables	WorldClim 2.1 monthly climate rasters for 1970-2000, including precipitation, temperature, vapor pressure and elevation	Sampled at SOTER profile coordinates and summarized by rice-season windows.
Advisory grounding	Rice Science and Technology in Nepal and Nepal rice disease literature	Used to curate concise advisory cards for disease, nutrient, weather and IPM retrieval.

2.2 Image feature extraction and classifiers

Each image was represented by 136 engineered features. The extractor segmented plant tissue against the mostly light background, then computed RGB and HSV histograms, color-channel summary statistics, lesion-color ratios, connected-component summaries for dark, brown and yellow lesion masks, edge density, plant-area ratio and a green-excess statistic. These features are intentionally simple and inspectable; they test whether a lightweight non-deep model can capture separable visual evidence in a small real-image benchmark.

Five classifiers were compared: logistic regression, linear support vector machine, random forest, gradient boosting and k-nearest neighbor. A 70/30 stratified hold-out split was used for the confusion matrix and per-class metrics. Five-fold stratified cross-validation was also run on the full image dataset. The best clean hold-out model was stress-tested on the same hold-out images using Gaussian noise, brightness shifts, blur and rotation. Stress transformations were not used for training.

2.3 Advisory retrieval

The advisory task evaluates retrieval rather than open-ended text generation. The knowledge base contains 24 concise bilingual cards covering symptoms, disease-conducive weather, nitrogen management, soil pH, organic matter, phosphorus and potassium balance, seed health, drainage, sanitation, dry-spell and flood response, scouting, chemical caution and expert referral. Four query variants were written for each card, giving 96 farmer-style test queries. Word TF-IDF, character TF-IDF, word TF-IDF with latent semantic indexing and a word-character hybrid retriever were evaluated using top-1 accuracy, top-3 accuracy, top-5 accuracy, mean reciprocal rank and median rank.

2.4 Soil-climate risk classification

The soil-climate table contains SOTER topsoil measurements and WorldClim seasonal summaries. Risk classes are derived from agronomic risk rules that emphasize high rainfall, high relative-humidity proxy, long dew-risk-hours proxy, disease-conducive temperature ranges, higher topsoil nitrogen, soil reaction outside a moderate range, water-holding capacity and elevation-season interactions. The target is therefore an interpretable soil-climate risk index, not an observed field disease incidence label. This distinction is important: the model tests the pipeline's ability to learn and reproduce a Nepal soil-climate risk index from real geospatial datasets, while field forecasting would require observed disease incidence data.

Table 4. Model inputs and fixed hyperparameters.

Module	Model	Features	Hyperparameters
Image classifier	Logistic regression	136 image features	StandardScaler; C=2.0; max_iter=1000; random_state=17
Image classifier	Linear SVM	same	StandardScaler; C=1.0; class_weight=balanced; max_iter=3000
Image classifier	Random forest	same	80 trees; min_samples_leaf=2; random_state=17
Image classifier	Gradient boosting	same	80 estimators; learning_rate=0.06; max_depth=3
Image classifier	k-NN	same	StandardScaler; k=5; distance weighting
Advisory retrieval	Character TF-IDF	card topic + bilingual card text	character 3-5 grams
Advisory retrieval	Hybrid TF-IDF	word + character cosine	0.45 word cosine + 0.55 character cosine
Risk classifier	Random forest	15 soil-climate features	100 trees; min_samples_leaf=3; random_state=17

3. Results and Discussion

3.1 Image classification on real rice leaf images

Table 5 reports the image classifier comparison and Figure 3 visualizes the five-fold macro-F1 scores. Logistic regression had the best clean hold-out performance with 0.917 accuracy and 0.918 macro-F1. Its cross-validated macro-F1 was 0.889. The three best image models were close on hold-out performance, but logistic regression provided the strongest combination of accuracy, macro-F1 and training speed in this feature setting.

Table 5. Image model comparison on UCI Rice Leaf Diseases.

model	test_accuracy	test_macro_f1	cv_accuracy_mean	cv_accuracy_std	cv_macro_f1_mean	cv_macro_f1_std	fit_time_s
Logistic regression	0.917	0.918	0.892	0.057	0.889	0.058	0.012
Linear SVM	0.889	0.892	0.867	0.049	0.864	0.051	0.013

k-NN	0.889	0.892	0.85	0.077	0.846	0.08	0.002
Gradient boosting	0.861	0.864	0.867	0.055	0.864	0.059	0.606
Random forest	0.861	0.863	0.842	0.067	0.841	0.066	0.148

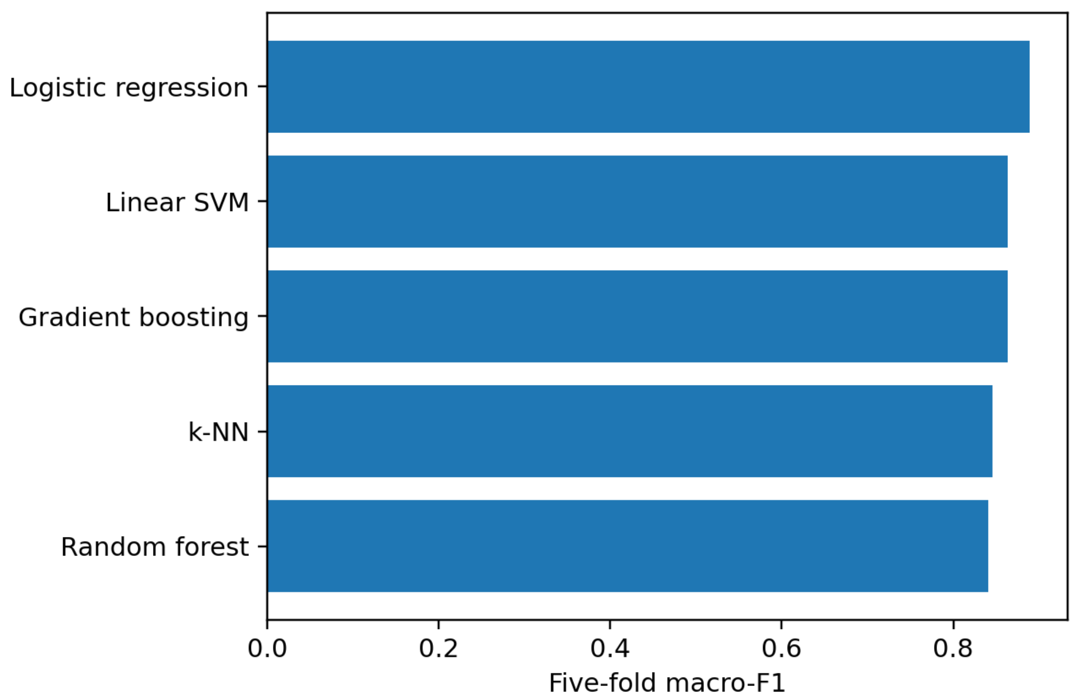

Figure 3. Image classification model comparison by five-fold macro-F1.

The hold-out confusion matrix in Table 6 and Figure 4 shows that all bacterial leaf blight test images were correctly recognized, while two brown spot images and one leaf smut image were classified as bacterial leaf blight. The per-class metrics in Table 7 reflect this pattern: bacterial leaf blight had perfect recall but lower precision because it absorbed the three errors; leaf smut had the highest F1 among the three classes.

Table 6. Hold-out confusion matrix for the best image classifier.

True class	Bacterial leaf blight	Brown spot	Leaf smut
Bacterial leaf blight	12	0	0
Brown spot	2	10	0
Leaf smut	1	0	11

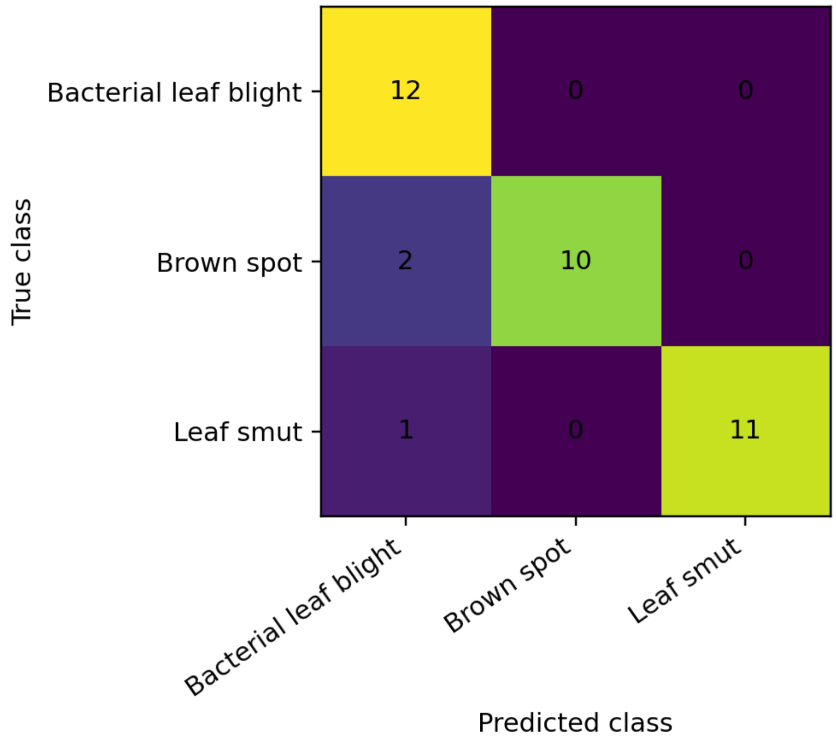

Figure 4. Hold-out confusion matrix for the best image classifier.

Table 7. Per-class clean hold-out performance.

class	precision	recall	f1	support
Bacterial leaf blight	0.8	1	0.889	12
Brown spot	1	0.833	0.909	12
Leaf smut	1	0.917	0.957	12

The stress tests in Table 8 and Figure 5 provide a more cautious interpretation than the clean hold-out score alone. Moderate blur and rotations up to 15 degrees preserved most performance, while severe noise and large brightness shifts were harmful. Accuracy fell from 0.917 on unmodified hold-out images to 0.444 at noise sigma 16 and to 0.333 at noise sigma 32. Strong darkening also reduced accuracy to 0.583, and strong brightening reduced accuracy to 0.444. These results indicate that practical capture guidance, exposure checks and image-quality rejection would be necessary before field deployment.

Table 8. Stress-test results for the best image classifier.

stress_test	level	accuracy	macro_f1

noise_sigma	0	0.917	0.918
noise_sigma	8	0.861	0.86
noise_sigma	16	0.444	0.348
noise_sigma	24	0.417	0.311
noise_sigma	32	0.333	0.235
brightness_factor	0.7	0.583	0.464
brightness_factor	0.85	0.667	0.643
brightness_factor	1	0.917	0.918
brightness_factor	1.15	0.75	0.745
brightness_factor	1.3	0.444	0.37
blur_radius	0	0.917	0.918
blur_radius	0.75	0.889	0.892
blur_radius	1.5	0.833	0.83
blur_radius	2.25	0.833	0.83
blur_radius	3	0.778	0.771
rotation_degrees	0	0.917	0.918
rotation_degrees	5	0.917	0.918
rotation_degrees	10	0.917	0.918
rotation_degrees	15	0.917	0.918
rotation_degrees	20	0.889	0.89

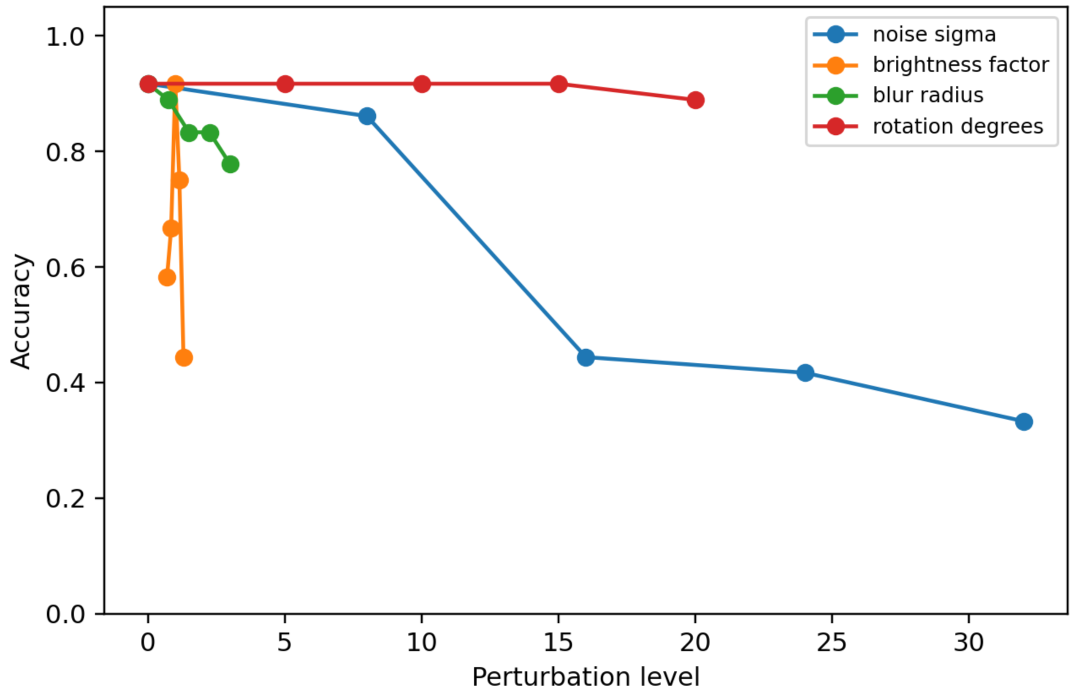

Figure 5. Image classifier robustness on transformed hold-out images.

3.2 Advisory retrieval performance

Table 9 and Figure 6 summarize advisory retrieval performance. Character TF-IDF was the best retriever, with 0.875 top-1 accuracy, 0.969 top-3 accuracy and 0.922 mean reciprocal rank. The hybrid word-character retriever was close, with 0.854 top-1 accuracy and the same 0.969 top-3 accuracy. Word-only TF-IDF remained useful but was weaker than character-level matching, which is expected when queries include short phrases, spelling variation and Nepali strings.

Table 9. Advisory retrieval method comparison.

method	top1_accuracy	top3_accuracy	top5_accuracy	mrr	median_rank
Char TF-IDF	0.875	0.969	0.979	0.922	1
Hybrid word+char TF-IDF	0.854	0.969	0.979	0.908	1
Word TF-IDF	0.75	0.875	0.938	0.824	1
Word TF-IDF + LSA	0.635	0.865	0.906	0.746	1

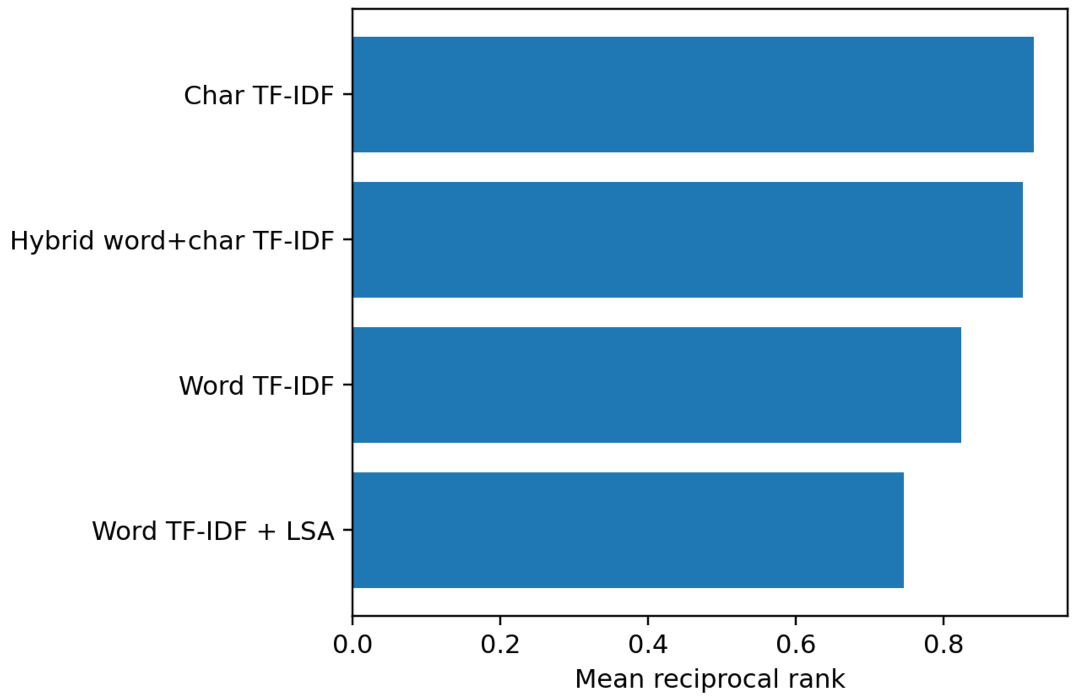

Figure 6. Advisory retrieval comparison by mean reciprocal rank.

The main retrieval errors occurred where broad management concepts naturally overlapped. For example, queries involving seed health, nursery management, nitrogen reduction or drainage can match more than one reasonable card. In a deployed retrieval-augmented advisory workflow, this behavior is manageable because the correct card appeared in the top three for 96.9% of queries. Figure 7 shows the advisory logic used to keep disease prediction, environmental context and retrieved agronomy guidance separate before advice is returned.

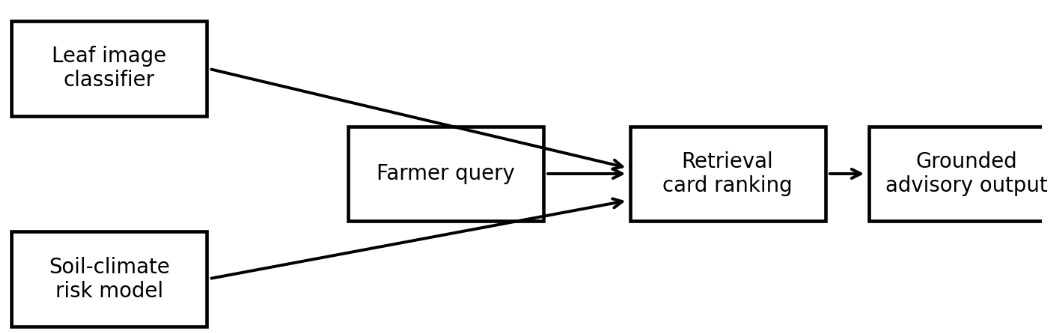

Figure 7. Grounded advisory generation logic.

3.3 Soil-climate risk classification

The soil-climate branch produced 645 profile-season records from 129 georeferenced Nepal SOTER profiles. Table 10 summarizes the resulting dataset. Medium-risk records were the largest class, followed by high and low risk. Figure 8 shows how risk counts vary by season: early, peak and late monsoon periods contain most of the high-risk records, while pre-monsoon and post-monsoon records contain more low-risk cases. Figure 9 shows the expected moisture gradient, with higher-risk records concentrated at higher seasonal rainfall and humidity-proxy values.

Table 10. Soil-climate risk dataset summary.

Item	Value
SOTER profiles used	129
Profile-season records	645
Low-risk records	152
Medium-risk records	310
High-risk records	183
Season windows	pre-monsoon, early monsoon, peak monsoon, late monsoon, post-monsoon
Model features	15 soil, climate and derived moisture variables

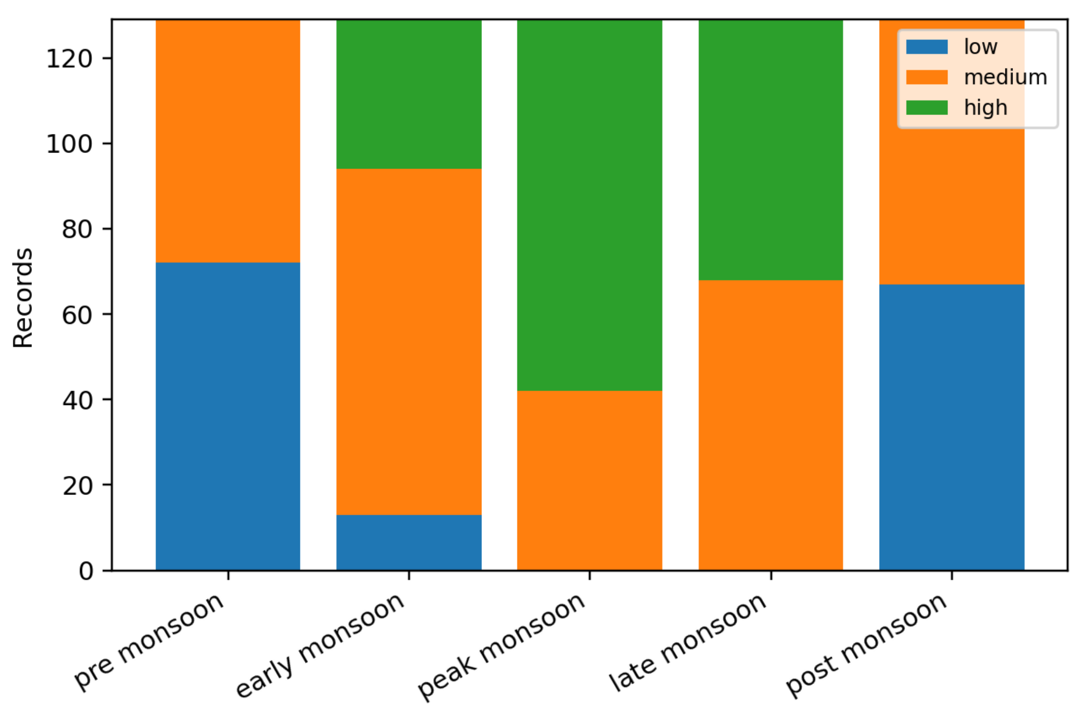

Figure 8. Soil-climate risk class counts by season.

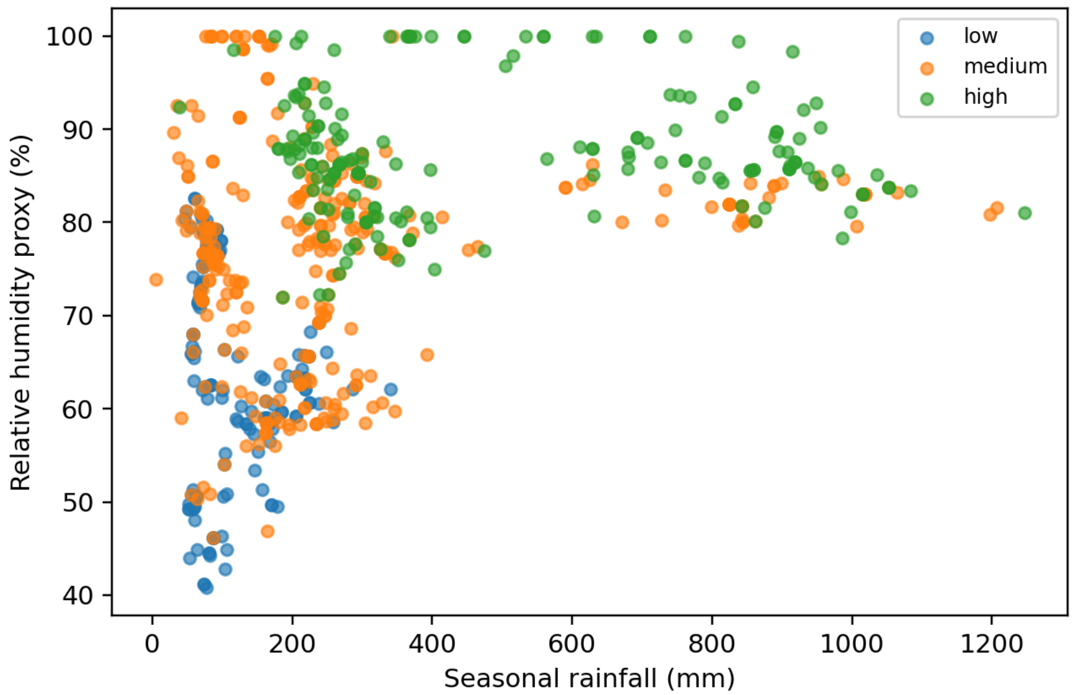

Figure 9. Rainfall and relative-humidity proxy by risk class.

Table 11 and Figure 10 compare the risk classifiers. Random forest achieved the strongest hold-out macro-F1 at 0.869 and a cross-validated macro-F1 of 0.851. Gradient boosting had the highest mean cross-validated macro-F1 at 0.860 but a slightly lower hold-out macro-F1 of 0.856. The tree-based models outperformed multinomial logistic regression and Gaussian naive Bayes, which suggests that interactions among rainfall, humidity proxy, temperature, elevation and soil properties are useful for this risk-index task.

Table 11. Soil-climate risk model comparison.

model	test_accuracy	test_macro_f1	cv_accuracy_mean	cv_accuracy_std	cv_macro_f1_mean	cv_macro_f1_std
Random forest	0.866	0.869	0.85	0.022	0.851	0.019
Gradient boosting	0.856	0.856	0.86	0.022	0.86	0.024
Decision tree	0.814	0.809	0.808	0.031	0.807	0.033
Multinomial logistic	0.747	0.751	0.73	0.048	0.734	0.048
Gaussian NB	0.711	0.719	0.664	0.029	0.672	0.029

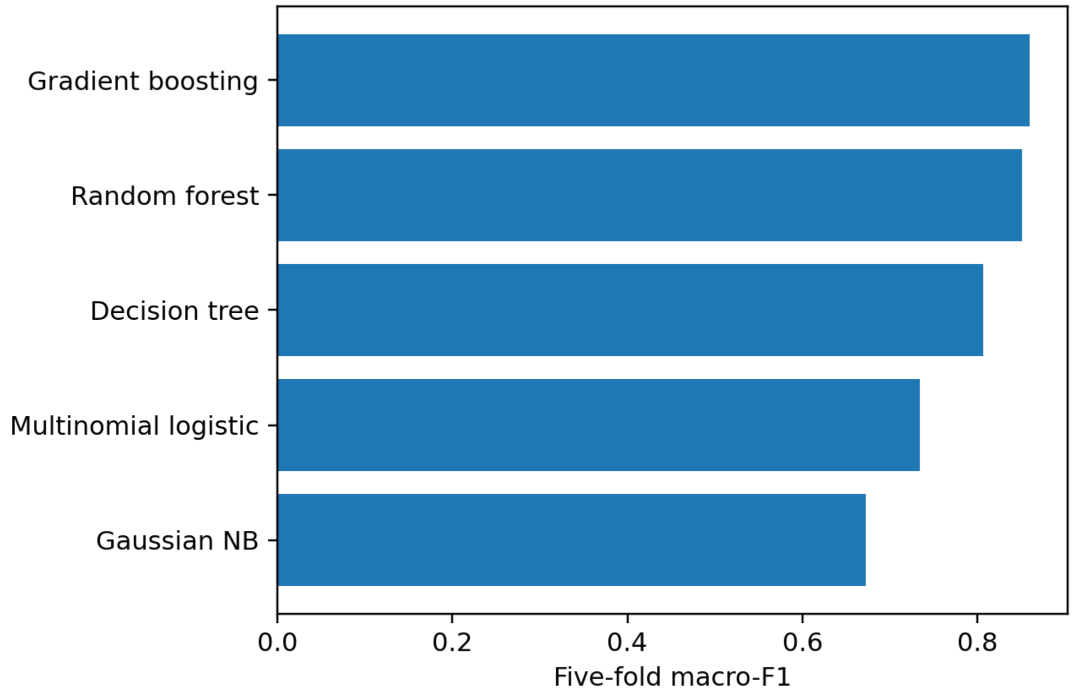

Figure 10. Soil-climate risk model comparison.

The risk confusion matrix in Table 12 and Figure 11 shows that the random forest separated the three risk classes reasonably well. Low-risk recall was 0.913, high-risk recall was 0.891 and medium-risk recall was 0.828. Table 13 gives the per-class precision, recall and F1 scores. Unlike the previous rule-constructed small table, this experiment derives the feature matrix from Nepal SOTER and WorldClim records; however, the risk class remains an index label rather than an observed disease-incidence measurement.

Table 12. Hold-out confusion matrix for the best soil-climate risk classifier.

True risk	high	low	medium
high	49	0	6
low	0	42	4
medium	8	8	77

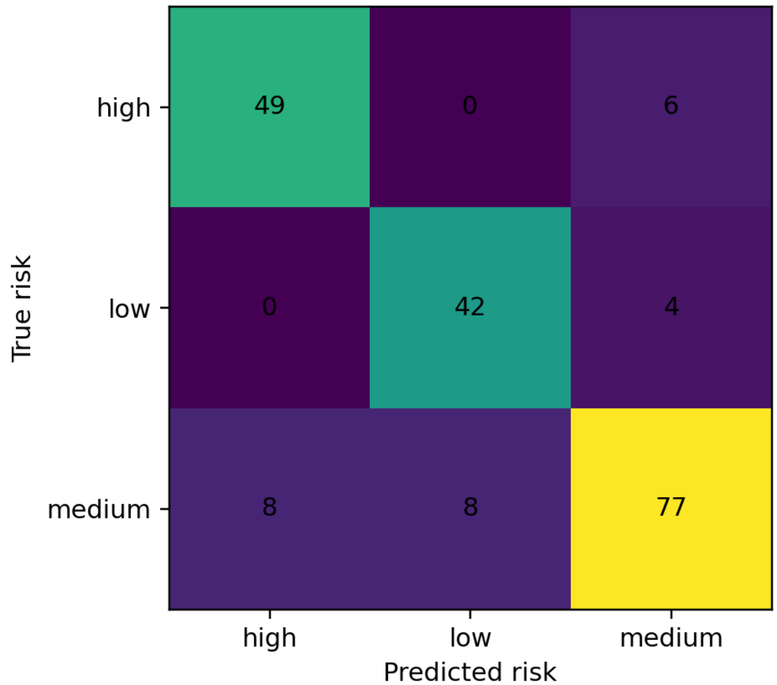

Figure 11. Hold-out confusion matrix for the best soil-climate risk classifier.

Table 13. Per-class hold-out performance for the best soil-climate risk classifier.

risk_class	precision	recall	f1	support
high	0.86	0.891	0.875	55
low	0.84	0.913	0.875	46
medium	0.885	0.828	0.856	93

3.4 Cross-module interpretation

The three modules give complementary evidence rather than a single opaque answer. The image classifier supplies a probable disease label from leaf appearance, the soil-climate module supplies a seasonal environmental risk index, and the advisory retriever supplies management text grounded in curated rice agronomy cards. The measured error patterns also suggest practical safeguards. Severe image noise and exposure shifts should trigger recapture or low-confidence handling. Broad farmer queries should return more than one card when the top-ranked card is uncertain. Medium risk should trigger monitoring and scouting language rather than definitive treatment advice.

4. Limitations

The image dataset is real but small. It contains three classes and 120 images, so the results should be interpreted as a transparent benchmark rather than evidence of national field-readiness. Field photographs from Nepal would include complex backgrounds, mixed symptoms, shadows, water droplets, partial occlusions, different phone cameras and expert-label disagreement. A field system should be retrained and validated on locally collected photographs before diagnostic use.

The advisory evaluation measures retrieval rank, not final language quality. A generative Nepali-language layer would still need hallucination checks, translation evaluation, farmer comprehension testing and safety constraints around chemical recommendations. The retrieval results show that the correct advice card is usually near the top of the ranked list; they do not prove that a generated response would always be safe or understandable.

The soil-climate experiment uses real SOTER and WorldClim covariates, but its target is a risk-index class derived from agronomic rules. It is therefore suitable for evaluating a soil-climate modeling workflow and for explaining seasonal risk patterns, but it is not an official disease forecast. Future work should connect the same covariates to observed disease incidence, cultivar susceptibility, crop calendars and extension-confirmed field reports.

5. Conclusion

This revised study evaluates a lightweight rice disease and advisory pipeline on real directly downloadable datasets. The image experiment uses UCI rice leaf images and reaches 0.917 hold-out accuracy and 0.918 macro-F1 with logistic regression on 136 engineered features. Robustness testing exposes a clear weakness under severe noise and strong brightness shifts. The advisory retriever performs strongly, with character TF-IDF achieving 0.875 top-1 accuracy and 0.969 top-3 accuracy on bilingual farmer-style queries. The soil-climate module uses Nepal SOTER and WorldClim data to produce 645 profile-season records and a learned risk-index classifier; random forest achieves 0.869 hold-out macro-F1.

The main contribution is not a field-ready diagnostic product. It is a compact, inspectable evaluation showing how leaf-image evidence, Nepal soil-climate context and grounded advisory retrieval can be integrated without relying on artificial image corpora or constructed soil-weather tables. The next step is to add Nepal field images, observed disease incidence records and farmer-facing safety evaluation.

References

[1] J. Shah, H. Prajapati, and V. Dabhi, "Rice Leaf Diseases," UCI Machine Learning Repository, 2019, doi: 10.24432/C5R013.
[2] H. B. Prajapati, J. P. Shah, and V. K. Dabhi, "Detection and classification of rice plant diseases," Intelligent Decision Technologies, vol. 11, no. 3, pp. 357-373, 2017.
[3] ISRIC - World Soil Information, "Soil and Terrain Database for Nepal (SOTER_Nepal)," version 1.0.
[4] S. E. Fick and R. J. Hijmans, "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas," International Journal of Climatology, vol. 37, pp. 4302-4315, 2017.
[5] CDD and ASoN, Rice Science and Technology in Nepal, Crop Development Directorate and Agronomy Society of Nepal, Lalitpur, Nepal, 2017.
[6] H. K. Manandhar, "Rice diseases and management in Nepal," in Rice Science and Technology in Nepal, CDD and ASoN, 2017, pp. 454-474.
[7] N. Neupane and K. Bhusal, "A review of blast disease of rice in Nepal," Journal of Plant Pathology & Microbiology, vol. 11, no. 528, 2020.
[8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE CVPR, 2005, pp. 886-893.
[9] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE TPAMI, vol. 24, no. 7, pp. 971-987, 2002.
[10] Eric Wang and Heyu Wang, “QoE-Driven Reinforcement Learning for Joint Bitrate, Rebuffering, and TTFF Optimization in HLS/DASH”, JACS, vol. 3, no. 2, pp. 50–59, Feb. 2023, doi: 10.69987/JACS.2023.30204.
[11] L. Breiman, "Random forests," Machine Learning, vol. 45, pp. 5-32, 2001.
[12] Shenghan Lu and David Zhou, “LLM-Augmented Customer Representation Learning for Next-Purchase Prediction in Online Retail”, JACS, vol. 3, no. 3, pp. 50–64, Mar. 2023, doi: 10.69987/JACS.2023.30305.
[13] Kai Zhang, Siquan Meng, and Eric Zhou, “Evidence-Grounded Trading Desk Risk Memos over SEC Filings: Retrieval-Augmented Generation with XBRL Numeric Verification”,

JACS, vol. 3, no. 2, pp. 60–76, Feb. 2023, doi: 10.69987/JACS.2023.30205.
[14] K. Sparck Jones, "A statistical interpretation of term specificity and its application in retrieval," Journal of Documentation, vol. 28, no. 1, pp. 11-21, 1972.
[15] FAO, Climate-Smart Agriculture Sourcebook, Food and Agriculture Organization of the United Nations, Rome, 2013.