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

Nepal-Localized Climate-Smart Crop Advisory from Crop-Recommendation, Cereal-Yield, Climate, Boundary, and Soil-Parent Data: A Small-Data Evaluation

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.2023.30705

Keywords

Climate-smart agriculture; Nepal; crop recommendation; cereal yield prediction; low-yield risk; soil parent material; NASA POWER; ridge regression; naive Bayes; small-data evaluation.

Abstract

This paper evaluates a Nepal-localized climate-smart crop advisory workflow using small public datasets that can be inspected and rerun locally. The workflow separates three tasks: crop-suitability classification from a compact soil-weather crop-recommendation table; district-crop cereal yield regression from Nepal cereal statistics, a fiscal-year NASA POWER climate index, district boundaries, and soil-parent features; and low-yield risk screening. The crop-recommendation dataset contained 2,200 records, seven numeric soil-weather attributes, and 22 crop labels. The Nepal panel contained 11,760 clean district-crop-year yield records for 75 districts and five cereals. A 12,053-row daily POWER point series was converted into 33 fiscal-year climate-index records, and a 75-district boundary layer was overlaid with 942 soil-parent polygons to add static terrain and parent-material descriptors. Using seed 42, Gaussian naive Bayes achieved 0.993 test accuracy and 0.993 macro-F1 on the crop-suitability task. The best temporal yield model, a ridge model with climate, district, crop, geo-soil, and crop-climate interaction features, achieved RMSE 727.90 kg ha-1, MAE 506.70 kg ha-1, and R2 0.486 on fiscal years 2010/11 to 2013/14. The low-yield risk task remained difficult, but the ridge risk classifier improved the screening result to F1 0.330, recall 0.592, precision 0.229, and AUC 0.664. The results support a conservative advisory design: strong crop-suitability screening from the clean benchmark dataset, Nepal-wide district-crop yield screening for cereals, and low-yield risk flags that require field review before intervention.

Introduction

Climate-smart agriculture (CSA) links productivity, adaptation, and mitigation in agricultural decision making [1]. It is particularly relevant for smallholder systems in mountain, hill, and Terai districts, where rainfall variability, heat exposure, remoteness, and limited advisory access can interact with food-security risk [2]-[6]. For Nepal, previous work has also emphasized farmer adaptation practices, mountain livelihood constraints, and the role of environmental and socio-economic conditions in food security [8]-[11]. A useful CSA advisory workflow for such a setting must therefore be local enough to reflect district crop history and environmental context, while remaining simple enough for institutional review.

The study focuses on tabular advisory evidence rather than plant-disease image recognition. Disease diagnosis would require a clean image collection whose acquisition, location, disease labels, and crop varieties match the intended advisory area. The available evidence in this experiment is stronger for soil-weather crop suitability and district-crop cereal yield screening, so the empirical claims are limited to those tasks. This narrower scope makes the results easier to interpret and prevents a strong result in one data modality from being used to support a different claim.

The central question is how well transparent small-data models can support a Nepal-localized CSA advisory workflow when the inputs are a compact crop-recommendation table, Nepal cereal-yield statistics, a daily climate index, administrative boundaries, and soil-parent descriptors. The question is intentionally practical. The paper does not claim to replace agronomists, causal crop models, or official extension systems. It tests whether a low-cost pipeline can classify suitable crops from soil-weather attributes, predict district-crop cereal yield in later fiscal years, and flag low-yield risk with enough transparency to guide further local review.

The revised data design strengthens the Nepal part of the analysis. The crop-recommendation dataset is retained as a generic crop-suitability benchmark, while the cereal-yield and low-yield risk tasks use all 75 districts that can be matched to the Nepal district boundary layer. The Nepal panel also includes static geo-soil descriptors from the district boundary and soil-parent overlay. This design preserves the original small-data character of the work but avoids limiting the Nepal yield experiment to a small subset of joined districts.

The study makes three contributions. First, it constructs a Nepal-wide district-crop cereal panel by joining cereal statistics, fiscal-year climate indices, district boundary coordinates, and soil-parent features. Second, it evaluates transparent classifiers and ridge models with fixed train, validation, and temporal test splits. Third, it converts measured test predictions into advisory-style review records while keeping the risk result clearly separate from automatic agronomic prescription. The evaluation questions are therefore separated by data source and purpose: crop-suitability classification, temporal cereal-yield regression, and low-yield risk screening.

Method

Table 1 summarizes the empirical data used in the experiments. The crop-recommendation file provides the soil-weather benchmark for 22-class crop-suitability classification. The Nepal-localized cereal workflow uses district-level cereal statistics, a NASA POWER daily point file converted to fiscal-year climate summaries, the Nepal ADM2 boundary layer, and a soil-parent polygon layer. Figure 1 shows how these components enter the advisory workflow, and Figure 2 shows the data construction and split logic used for evaluation.

Table 1. Dataset inventory used in the experiments.

Dataset	Rows used	Variables used	Experimental role
Crop recommendation	2,200 records	N, P, K, temperature, humidity, pH, rainfall, crop label	Crop-suitability classification
NASA POWER daily point series	12,053 daily rows; 33 fiscal-year records	Precipitation, temperature, humidity, pressure, wind, wet-bulb temperature	National fiscal-year climate index
Nepal ADM2 boundary layer	75 district polygons	District names, centroid latitude/longitude, area, boundary identifiers	District matching and geo features

Dataset	Rows used	Variables used	Experimental role
Soil-parent polygon layer	942 polygons; 75 district summaries	Parent material, dominant soil class, landform, elevation, slope, relief	Static geo-soil descriptors
Nepal cereal yield panel	11,760 clean district-crop-year rows	District, crop, fiscal year, area, production, yield	Yield regression and low-yield risk

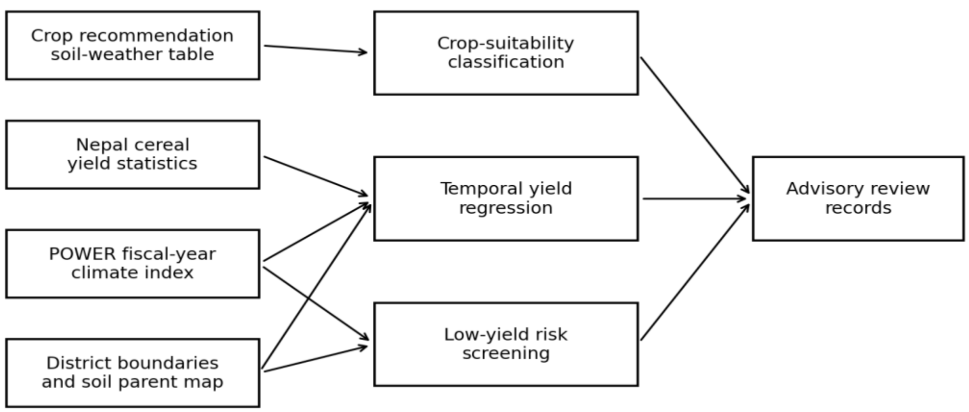

Figure 1. System architecture for the Nepal-localized climate-smart advisory workflow.

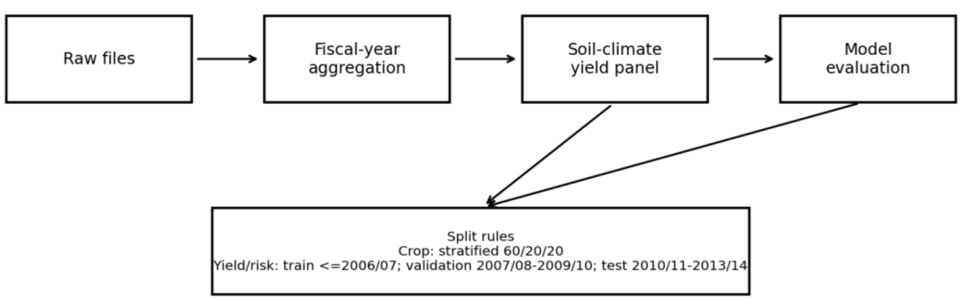

Figure 2. Data construction and temporal evaluation pipeline.

The daily climate records were assigned to Nepal fiscal years using the rule July 16 to July 15. For each fiscal year, precipitation was summarized by sum and mean; temperature, humidity, pressure, and wind were summarized by means or extremes; and dry, wet, hot, cold, and humid-hot day counts were computed from fixed thresholds. Because the available POWER file is a single point series, the climate variables are interpreted as a national fiscal-year climate index rather than as district-specific weather. District variation is instead represented by crop identity, district identity, centroid coordinates, boundary-derived area, and soil-parent summaries.

The cereal table was reshaped from wide crop-year fields to district-crop-year rows. Paddy, maize, millet, wheat, and barley were retained because they have long-history coverage in the source table. Buckwheat was excluded from the predictive models because the available records did not provide the same long-history structure. Yield records with non-positive yield, area, or production, and yield records above 10,000 kg ha-1, were removed before the temporal split. Area and production were retained for audit but not used as yield predictors because they are directly tied to the target yield.

The boundary and soil-parent layers were used to construct district-level static features. Each district boundary was matched to the cereal district names, centroid latitude and longitude were computed, and soil-parent polygons were overlaid with district boundaries to identify the dominant parent material, dominant soil group, landform, elevation range, median slope, and median relief. The resulting 75 district rows were joined to every district-crop-year record in the cereal panel.

Table 2. Feature construction and data-quality rules.

Component	Rule used	Reason
Fiscal year	Dates on or after July 16 are assigned to the current fiscal start year; earlier dates are assigned to the previous fiscal start year.	Aligns daily climate records with Nepal agricultural fiscal-year yield records.
Climate summaries	Means and extremes for temperature, humidity, pressure, wind, and wet-bulb temperature; sums and means for precipitation; threshold counts for dry, wet, hot, cold, and humid-hot days.	Produces interpretable seasonal resource and stress indicators.
Cereal records	Paddy, maize, millet, wheat, and barley retained; buckwheat excluded from model evaluation.	Maintains comparable long-history coverage for training, validation, and test years.
Geo-soil features	District centroids and boundary area joined with dominant soil-parent and terrain descriptors from polygon overlay.	Adds Nepal-specific static environmental context without synthetic observations.
Yield filter	Yield <=0, area <=0, production <=0, and yield >10,000 kg ha-1 removed before splitting.	Removes impossible and extreme data-entry values before evaluation.

Component	Rule used	Reason
Leakage control	Standardization, historical district-crop means, and risk thresholds learned from training rows only.	Keeps later fiscal years out of model fitting and threshold construction.

Modeling and evaluation

The crop-recommendation task was treated as 22-class classification. For each crop label, 60 records were assigned to training, 20 to validation, and 20 to testing, producing 1,320 training, 440 validation, and 440 test records. Numeric features were standardized from the training set. Five transparent classifiers were evaluated: nearest centroid, Gaussian naive Bayes, k-nearest neighbours, one-vs-rest ridge classification, and multinomial softmax regression. The validation set selected k for k-nearest neighbours and the regularization level for ridge and softmax models.

The Nepal yield task was evaluated as temporal regression. Training used fiscal years through 2006/07, validation used 2007/08 to 2009/10, and testing used 2010/11 to 2013/14. The dependent variable was cereal yield in kg ha-1. Baselines included a global mean, crop mean, and district-crop mean computed from the training period. Ridge models then added climate-index features, district and crop identity, geo-soil descriptors, and crop-specific interactions with fiscal rainfall and mean temperature. The temporal split is stricter than a random split because it asks the models to predict later fiscal years from earlier fiscal years.

The low-yield risk task converted each district-crop training series into a local threshold: the first quartile of training yield. Validation and test observations were labelled as low-yield risk when observed yield was at or below that district-crop threshold. Four risk methods were compared: a majority baseline, a threshold on the yield-regression prediction, a Gaussian climate/soil-risk naive Bayes classifier, and a ridge risk classifier. Classification used accuracy and macro-F1 for the crop task, regression used RMSE, MAE, bias, and R2, and risk screening used accuracy, precision, recall, F1, and AUC.

Table 3. Experimental splits and sample counts.

Task	Training	Validation	Test	Split logic
Crop recommendation	1,320	440	440	Stratified 60/20/20 by crop label
Cereal yield regression	9,219	1,095	1,446	Temporal: <=2006/07, 2007/08-2009/10, 2010/11-2013/14
Low-yield risk	9,219	1,095	1,446	Same temporal split with train-quartile risk threshold

Results and Discussion

Crop-suitability classification

The crop-suitability task produced the strongest empirical result. Table 4 reports the held-out test comparison, and Figure 3 visualizes the corresponding top-1 accuracies. Gaussian naive Bayes achieved 0.993 test accuracy and 0.993 macro-F1. The test set contains 440 records, so this result corresponds to 437 correct predictions and three errors. k-nearest neighbours was also strong, while the linear one-vs-rest ridge model was substantially weaker, indicating that several crop classes occupy class-conditional regions that are not well represented by one linear boundary.

Table 4. Crop recommendation model comparison on the held-out test set.

Model	Selected parameter	Validation accuracy	Test accuracy	Macro-F1
Gaussian naive Bayes	var=1e-09	0.998	0.993	0.993
Multinomial softmax	C=100	0.984	0.980	0.979
k-nearest neighbours	k=1	0.982	0.970	0.970
Nearest centroid	none	0.868	0.873	0.863
One-vs-rest ridge	alpha=10	0.718	0.732	0.674

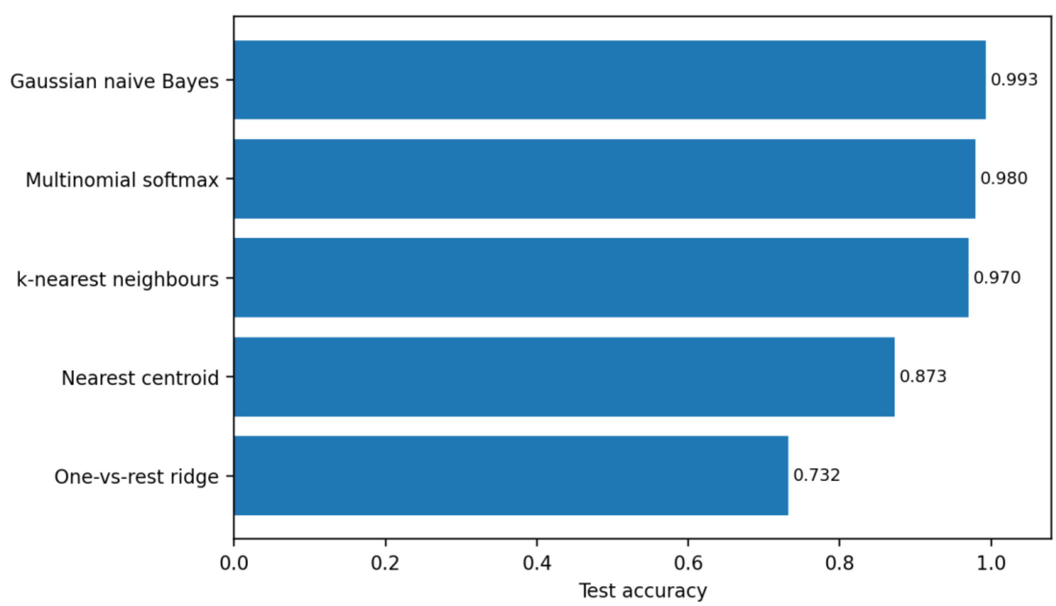

Figure 3. Measured top-1 crop recommendation accuracy for five classifiers.

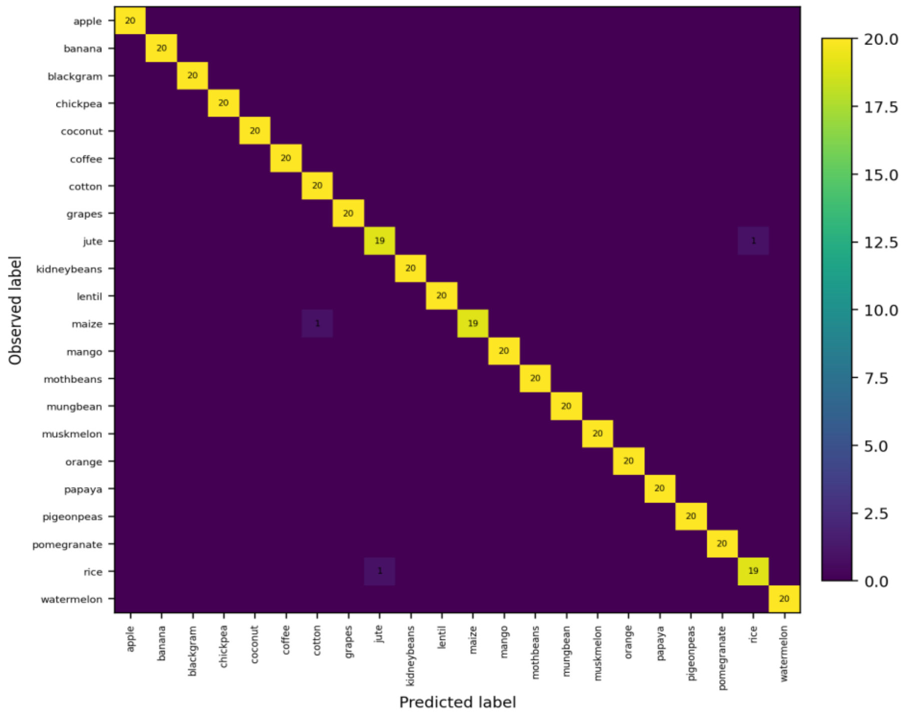

Figure 4. Confusion matrix of the best crop classifier.

Figure 4 shows that the best classifier made only three off-diagonal predictions. The high score is therefore interpreted as evidence that this cleaned benchmark is well separated for a simple probabilistic classifier, not as proof that every Nepal farm can be classified from seven variables alone. In a deployed advisory workflow, the crop-suitability score should be combined with local crop calendars, soil tests, irrigation access, elevation, market objectives, and agronomist review.

Nepal cereal-yield regression

The Nepal cereal-yield task was harder because the test period is later in time and because district-crop yield depends on management, irrigation, seed, pests, and measurement factors that are not available in the panel. Table 5 reports the temporal test results, Figure 5 compares RMSE values, and Figure 6 plots observed versus predicted yield for the best temporal ridge model. The climate + identity + geo-soil + crop-interaction ridge model achieved RMSE 727.90 kg ha-1, MAE 506.70 kg ha-1, and R2 0.486 on 2010/11 to 2013/14. Relative to the global mean baseline, RMSE fell by 34.9%; relative to the district-crop mean baseline, RMSE fell by 11.2%.

Table 5. Nepal cereal yield regression results on temporal test years 2010/11-2013/14.

Model	Selected parameter	Validation RMSE	Test RMSE	MAE	R2
Climate + identity + geo-soil + crop interactions	alpha=0.01	464.6	727.9	506.7	0.486

Model	Selected parameter	Validation RMSE	Test RMSE	MAE	R2
Climate + identity + geo-soil ridge	alpha=0.01	479.3	734.3	519.8	0.477
Climate + identity ridge	alpha=0.01	479.3	734.3	519.8	0.477
Identity ridge	alpha=0.01	478.2	739.3	518.8	0.470
District-crop mean	none	419.5	819.7	576.2	0.349
Crop mean	none	557.7	915.6	607.9	0.187
Climate + geo-soil ridge	alpha=0.01	750.4	977.0	766.6	0.075
Climate ridge	alpha=10000	780.9	1063.5	818.1	-0.096
Global mean	none		1117.6	844.9	-0.211

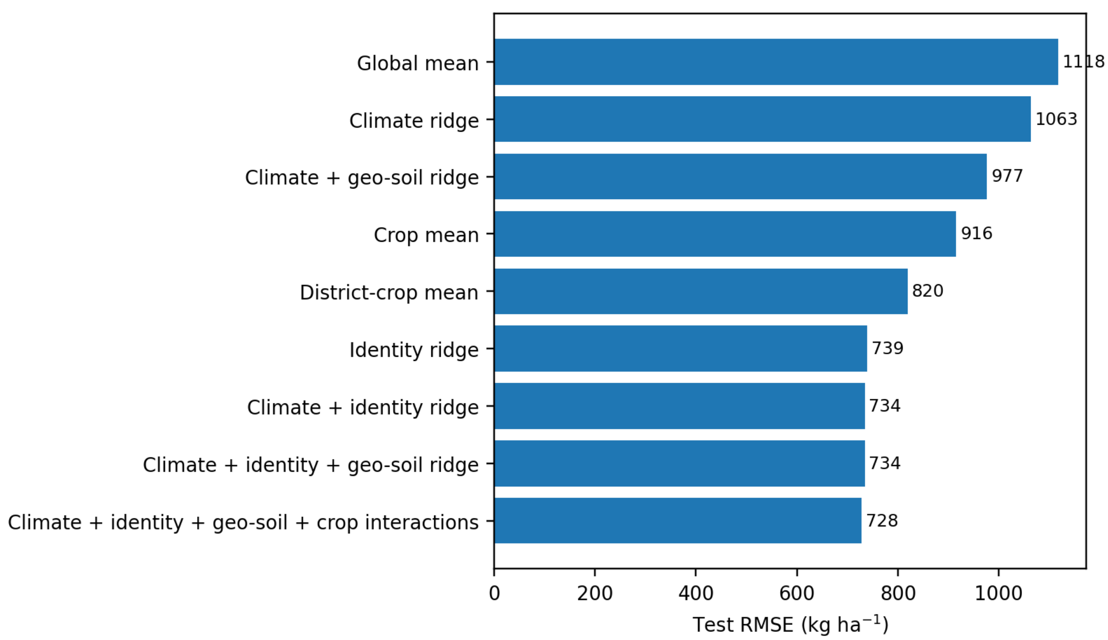

Figure 5. Yield regression RMSE comparison on temporal test years.

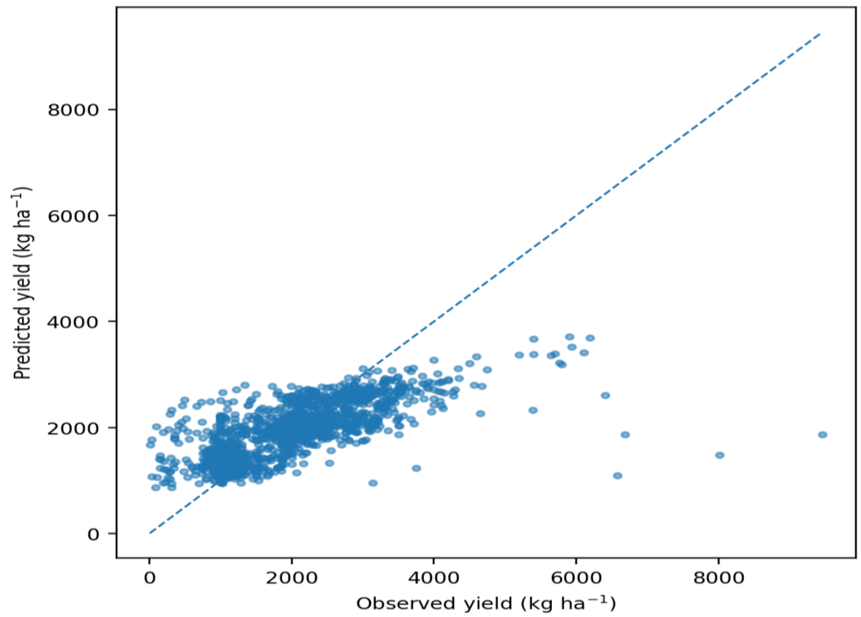

Figure 6. Observed versus predicted cereal yield for the best yield model.

Table 6 isolates the effect of the feature sets used in the ridge models. District and crop identity explained a large share of the predictable variation. Adding the national climate index and geo-soil descriptors produced a small additional improvement, while the crop-climate interaction terms gave the best test RMSE. This pattern is consistent with a conservative screening interpretation: persistent district-crop history is the main anchor, and climate terms modestly adjust the expectation in later fiscal years.

Table 6. Ridge feature-set ablation for yield prediction.

Feature set	District identity	Geo-soil features	Crop-climate interactions	Penalty	RMSE	R2
Climate ridge	no	no	no	alpha=10000	1063.5	-0.096
Climate + geo-soil ridge	no	yes	no	alpha=0.01	977.0	0.075
Identity ridge	yes	no	no	alpha=0.01	739.3	0.470
Climate + identity ridge	yes	no	no	alpha=0.01	734.3	0.477
Climate + identity + geo-soil ridge	yes	yes	no	alpha=0.01	734.3	0.477
Climate + identity + geo-soil +	yes	yes	yes	alpha=0.01	727.9	0.486

Feature set	District identity	Geo-soil features	Crop-climate interactions	Penalty	RMSE	R2
crop interactions						

Crop-level results in Table 7 show where the yield model was most useful and where it was weaker. Barley had the lowest absolute RMSE but negative R2 because its test-year variation was small and several low-yield cases were over-predicted. Paddy, maize, and wheat had positive R2 but still showed large absolute errors. Millet was unstable, reflecting the wider range and noisier observations in the test years. These crop-level differences reinforce that the model is suitable for screening and ranking, not for exact production accounting.

Table 7. Crop-level yield performance of the best temporal model.

Crop	Test rows	RMSE	MAE	Bias	R2
Barley	284	483.0	395.4	256.3	-1.178
Wheat	300	614.7	473.7	-249.6	0.105
Millet	270	768.8	499.7	277.8	-0.305
Paddy	292	824.1	617.0	-235.9	0.236
Maize	300	873.1	544.1	-187.9	0.106

The feature correlations in Table 8 and Figure 7 are descriptive, not causal. Fiscal year had the largest association with district-crop yield anomaly, reflecting long-run productivity change. Wet-bulb temperature, hot-day counts, vapour quantity, minimum temperature, and maximum temperature also appeared among the largest climate associations. None of these variables should be interpreted as a single-feature decision rule because management and measurement variables remain outside the panel.

Table 8. Top climate and temporal feature correlations with district-crop yield anomaly in training data.

Feature	Correlation	Absolute correlation
fy_start	0.422	0.422
T2MWET_mean	0.208	0.208
hot_days	-0.203	0.203
QV2M_mean	0.196	0.196

Feature	Correlation	Absolute correlation
T2M_MIN_mean	0.183	0.183
T2M_MAX_max	-0.182	0.182
PS_mean	0.152	0.152
WS10M_mean	-0.145	0.145
RH2M_mean	0.128	0.128
dry_days	-0.124	0.124

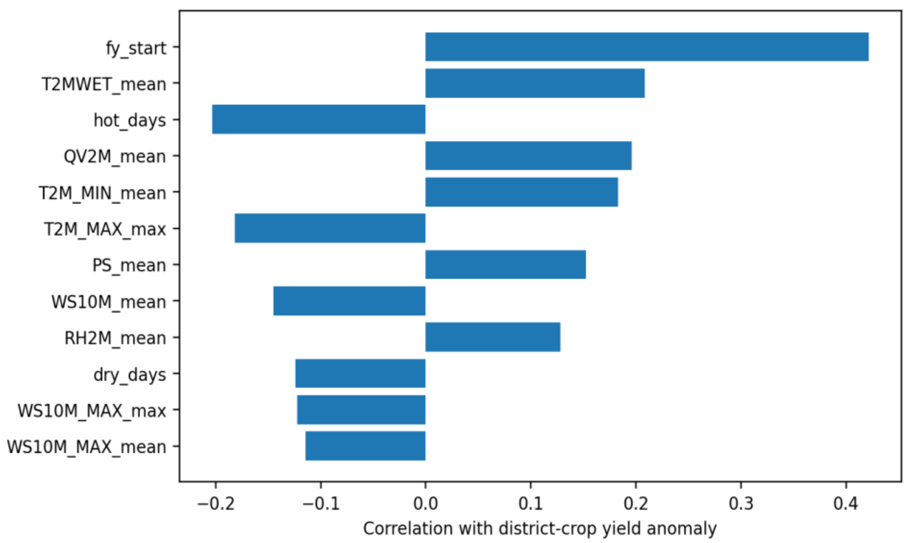

Figure 7. Feature correlations used for interpretation of the yield model.

Low-yield risk screening and advisory review records

The low-yield risk task was the most imbalanced part of the workflow. The temporal test set contained 179 positives among 1,446 rows under the district-crop training-quartile definition. Table 9 shows that the ridge risk classifier achieved the best measured F1, 0.330, with recall 0.592, precision 0.229, and AUC 0.664. It retrieved 106 of the 179 low-yield cases but also produced 357 false positives. Figure 8 visualizes the F1 comparison. These values are appropriate for prioritizing review, not for automatic input, credit, or insurance decisions.

Table 9. Low-yield risk classification on temporal test years.

Model	Selected parameter	Accuracy	Precision	Recall	F1	AUC
Ridge risk classifier	alpha=10; threshold=-0.1228	0.703	0.229	0.592	0.330	0.664
Gaussian climate/soil-risk NB	var=1e-09; threshold=0.0000	0.557	0.164	0.631	0.261	0.609
Yield-regression threshold	predicted_yield <= train Q1	0.868	0.000	0.000	0.000	0.677
Majority baseline	train majority	0.876	0.000	0.000	0.000	0.500

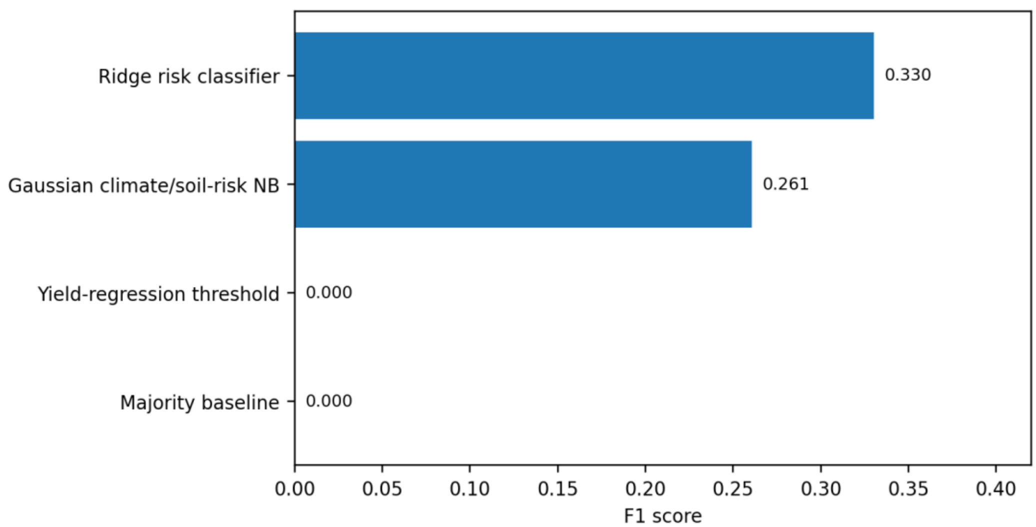

Figure 8. F1 comparison for the low-yield risk classifiers.

Table 10 converts measured temporal-test rows into advisory review records. Each row reports the fiscal year, district, crop, observed yield, predicted yield, risk flag, and measured trigger. Some rows were true low-yield cases under the training-quartile threshold, while others were risk-score flags that would need field confirmation. The table is written as review evidence rather than as a prescriptive advisory, because the panel does not contain fertilizer, seed, pest, irrigation, or household-level constraints.

Table 10. Advisory review records selected from temporal-test predictions.

FY	District	Crop	Observed yield	Predicted yield	Risk flag	Measured trigger
2010/11	Saptari	Millet	1000.0	1293.9	1	Observed yield below train Q1
2010/11	Rolpa	Millet	534.0	1068.0	1	Observed yield below train Q1
2010/11	Bardiya	Millet	1000.0	1357.4	1	Observed yield below train Q1
2010/11	Nawalparasi	Millet	1000.0	1376.3	1	Observed yield below train Q1
2010/11	Jhapa	Millet	1200.0	1388.9	1	Risk score above validation threshold; review needed
2010/11	Parsa	Millet	1025.0	1630.3	1	Risk score above validation threshold; review needed
2010/11	Okhaldhunga	Millet	1351.0	1114.4	1	Risk score above validation threshold; review needed
2010/11	Bhaktapur	Millet	1400.0	2220.7	1	Risk score above validation threshold; review needed

The data, result, and figure files are small enough for local inspection. Table 11 lists the approximate footprint of the files used by the experiment. The largest processed file is the Nepal climate-soil-yield panel, because it contains all 75 district-crop fiscal-year records after joins and filters.

Table 11. Data, result, and figure footprint.

Package component	Size KB
raw/Crop_recommendation.csv	146.5
raw/NepalAgriStats_Cereal.csv	192.7
raw/POWER_Point_Daily_19810716_20140715_027d72N_085d32E_LST.csv	820.3
raw/geoBoundaries-NPL-ADM2-all.zip	258.3
raw/parentsoil.zip	2658.2
processed/crop_recommendation_splits.csv	158.9
processed/nepal_fiscal_climate_index.csv	8.4
processed/nepal_district_geo_soil_features.csv	11.0
processed/nepal_cereal_yield_long.csv	543.0
processed/nepal_climate_soil_yield_panel.csv	4876.1
results/*.csv and experiment_summary.json	230.0
figures/*.png	605.4

Limitations

The experiment has five concrete limitations. First, the POWER file used here is a single daily point record near Kathmandu, so the climate variables are a national fiscal-year climate index rather than district-specific weather. This choice preserves the available data exactly but cannot represent local rainfall gradients across the mountains, hills, and Terai. Second, the crop-recommendation dataset is a generic soil-weather benchmark rather than a Nepal farm survey. It supports a transparent classifier, but local deployment would require Nepal-specific crop calendars, local soil tests, and agronomist review. Third, the soil-parent layer provides useful static environmental context, but it is not a substitute for field-level soil chemistry, fertilizer use, irrigation access, seed variety, pest incidence, or management variables. Fourth, the low-yield risk classifier still has low precision and should be used only to prioritize follow-up. Fifth, plant-disease image recognition is outside the scope of the empirical claims because no clean Nepal-matched disease-image dataset was used.

These limitations also identify the clearest path for improvement. District-specific daily climate extraction would strengthen the temporal yield panel. Field soil chemistry and management variables would reduce the unexplained yield variation. More positive low-yield cases would support calibrated risk thresholds. A Nepal-native crop-suitability survey would allow the crop classifier to move from benchmark evidence to local advisory evidence.

Conclusion

This study evaluates a Nepal-localized climate-smart crop advisory workflow using small tabular datasets and transparent models. Gaussian naive Bayes achieved 0.993 top-1 accuracy and 0.993 macro-F1 on the crop-recommendation benchmark. A climate + identity + geo-soil + crop-interaction ridge model achieved RMSE 727.90 kg ha-1 and R2 0.486 on later Nepal fiscal years for 75 districts and five cereals. A ridge risk classifier improved low-yield screening to F1 0.330 and recall 0.592, but its precision of 0.229 means that risk flags require field review. The coherent interpretation is that crop-suitability classification is strong on the clean benchmark dataset, while Nepal cereal yield and risk outputs are best used as conservative decision support for extension review.

References

[1] Food and Agriculture Organization of the United Nations, Climate-Smart Agriculture Sourcebook. Rome, Italy: FAO, 2013.

[2] L. Lipper et al., "Climate-smart agriculture for food security," Nature Climate Change, vol. 4, no. 12, pp. 1068-1072, 2014.

[3] Intergovernmental Panel on Climate Change, Climate Change 2014: Impacts, Adaptation, and Vulnerability. Cambridge, U.K.: Cambridge University Press, 2014.

[4] D. B. Lobell, W. Schlenker, and J. Costa-Roberts, "Climate trends and global crop production since 1980," Science, vol. 333, no. 6042, pp. 616-620, 2011.

[5] A. J. Challinor et al., "A meta-analysis of crop yield under climate change and adaptation," Nature Climate Change, vol. 4, no. 4, pp. 287-291, 2014.

[6] C. Rosenzweig et al., "Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison," Proceedings of the National Academy of Sciences, vol. 111, no. 9, pp. 3268-3273, 2014.

[7] D. B. Lobell and M. B. Burke, "On the use of statistical models to predict crop yield responses to climate change," Agricultural and Forest Meteorology, vol. 150, no. 11, pp. 1443-1452, 2010.

[8] S. Manandhar, D. S. Vogt, S. R. Perret, and F. Kazama, "Adapting cropping systems to climate change in Nepal," Regional Environmental Change, vol. 11, no. 2, pp. 335-348, 2011.

[9] P. Gentle and T. N. Maraseni, "Climate change, poverty and livelihoods: Adaptation practices by rural mountain communities in Nepal," Environmental Science & Policy, vol. 21, pp. 24-34, 2012.

[10] P. C. Tiwari and B. Joshi, "Natural and socio-economic factors affecting food security in the Himalayas," Food Security, vol. 4, no. 2, pp. 195-207, 2012.

[11] A. Maharjan, S. de Campos, A. Singh, M. Dasgupta, and S. Srinivasan, "Understanding rural outmigration and agricultural land use change in the Gandaki Basin, Nepal," Applied Geography, vol. 124, p. 102278, 2020.

[12] B. Ghimire, District Level Cereal Crop Production in Nepal. Kathmandu, Nepal: Ministry of Agricultural Development, 2014.

[13] Jiaying Jin and Tina Huang, "Market Microstructure Risk Forecasting from Limit Order Books: Multi-Horizon Price-Move Classification and Volatility Estimation with DeepLOB-Style CNN–LSTM and Temporal Transformers", JACS, vol. 3, no. 12, pp. 34–44, Dec. 2023, doi: 10.69987/JACS.2023.31205.

[14] P. W. Stackhouse, D. Westberg, W. Chandler, T. Zhang, and C. Hoell, Prediction of Worldwide Energy Resource (POWER) Release 8 Methodology. Hampton, VA, USA: NASA Langley Research Center, 2018.

[15] D. Runfola et al., "geoBoundaries: A global database of political administrative boundaries," PLoS ONE, vol. 15, no. 4, e0231866, 2020.

[16] Soil parent material map of Nepal, SoilParentFINAL/Simplified Polygon layer metadata, 2019.

[17] A. Ingle, "Crop Recommendation Dataset," Kaggle, 2020.

[18] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, vol. 12, no. 1, pp. 55-67, 1970.

[19] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," IEEE Transactions on

Information Theory, vol. 13, no. 1, pp. 21-27, 1967.
[20] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," Machine Learning, vol. 29, no. 2-3, pp. 103-130, 1997.
[21] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, 2nd ed. New York, NY, USA: Wiley, 2001.
[22] 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.
[23] K. P. Murphy, Machine Learning: A Probabilistic Perspective. Cambridge, MA, USA: MIT Press, 2012.
[24] 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.