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

COMPARATIVE EVALUATION OF MACHINE LEARNING ALGORITHMS FOR MULTICLASS CROP RECOMMENDATION USING SOIL AND CLIMATIC PARAMETERS, WITH A SEPARATE CNN BASELINE FOR SOIL-IMAGE CLASSIFICATION

MAMATHA C
Sweetu.888@gmail.com
Assistant Professor
Bachelor of Computer science
Govt Science college, Hassan (Autonomous)
Orcid id:0009000535864014

ABSTRACT
Crop recommendation based on soil and climatic measurements is a multiclass classification task. Because performance on public benchmark datasets can approach 100%, careful validation and transparent reporting are essential. This study presents two independent experiments. In the first, Random Forest, Extra Trees, Decision Tree and a support vector machine with a radial basis function kernel (SVM-RBF) were evaluated for 22-class crop recommendation using seven features: nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH and rainfall. The dataset contained 2,200 records and was evaluated using an 80:20 stratified train–test split, with five-fold stratified cross-validation on the training set. Random Forest achieved a mean cross-validation accuracy of 99.32% (SD 0.53), while Extra Trees achieved 99.26% (SD 0.39). On the 440-record test set, both models achieved 99.55% accuracy and a 99.55% macro F1-score (95% Wilson confidence interval: 98.4–99.9%). The small difference between their cross-validation scores and their identical test results do not support a claim that either model is superior. Decision Tree and SVM-RBF achieved test accuracies of 97.96% and 98.41%, respectively. In the second experiment, a convolutional neural network (CNN) classified 552 soil images into five categories and achieved 91.89% accuracy and a 91.40% macro F1-score on 111 test images (95% Wilson confidence interval: 85.3–95.7%). Eight of the nine errors were predictions of type_5. The datasets are independent, and no model combining numerical and image data was developed. Crop labels were not treated as measures of soil health, and independent field validation was not conducted.
Keywords: crop recommendation; precision agriculture; multiclass classification; Random Forest; Extra Trees; support vector machine; convolutional neural network; soil-image classification.

I. INTRODUCTION

Agriculture plays an important role in food security, economic development and environmental sustainability. Selecting a suitable crop depends on several interacting factors, including soil nutrient levels, pH, temperature, humidity and rainfall. An unsuitable choice may reduce productivity and increase input costs. For this reason, data-driven decision-support methods are receiving growing attention in precision agriculture. Machine learning (ML) can identify patterns in historical agricultural data and use them to recommend crops for particular soil and climatic conditions.

Numerical crop-recommendation datasets describe soil nutrients and climatic conditions as structured measurements. Soil images provide a different type of information, such as visible colour, texture and particle appearance, which can be analysed using convolutional neural networks (CNNs). These approaches address different prediction tasks. A crop recommendation based on numerical measurements does not establish soil health, while assigning a category to a soil image does not determine which crop will perform best. The reliability of both approaches depends on data quality, label provenance and the evaluation procedure, especially when reported performance is close to 100%.

This study therefore examines two separate tasks. The first compares four standard classifiers for recommending crops across 22 classes using seven numerical features. The second evaluates a CNN for classifying soil images into five categories using a separate dataset. The datasets are not paired, and no model combining the numerical and image inputs was trained or evaluated.

1.1 Research objectives

1. To compare Random Forest, Extra Trees, Decision Tree and SVM-RBF for multiclass crop recommendation from numerical soil and climatic features under one stratified split and five-fold cross-validation protocol.
2. To establish a CNN baseline for five-class soil-image classification, evaluated on a held-out test set.
3. To report uncertainty (standard deviations, 95% intervals) and class-level errors for both experiments rather than accuracy alone.
4. To state explicitly what the results do and do not support, including the absence of fusion, soil-health labels and field validation.

1.2 Contributions

This study makes three contributions: (i) a comparison of four classifiers for numerical crop recommendation using cross-validation, held-out testing, confidence intervals and error analysis; (ii) a CNN baseline for soil-image classification, including class-level metrics derived from its confusion matrix; and (iii) a clear account of the scope and limitations of both experiments. The study does not claim a novel algorithm or a validated multimodal framework. Its contribution is a documented baseline for two distinct classification tasks.

II. LITERATURE SURVEY

2.1 Numerical crop recommendation

Dey et al. [1] investigated machine-learning methods for recommending agricultural and horticultural crops in India using NPK values, soil pH and three climatic variables. Fattepur et al. [2] examined precision crop recommendation using IoT-derived data alongside machine-learning and deep-learning methods. These studies are relevant to the numerical experiment in this paper, although their datasets, class definitions, models and validation procedures may differ. Their reported accuracies should therefore be compared only when the evaluation settings are sufficiently similar. In addition, recommendations learned from a dataset should not be treated as field-validated agronomic advice, because crop performance also depends on factors such as location, irrigation, soil texture and farm management.

2.2 Soil-image classification

Gyasi and Purushotham [3] reviewed deep-learning approaches to soil classification, providing background for CNN-based image analysis. Sreelatha et al. [4] presented a hybrid deep-learning approach for automated soil classification and crop recommendation. Image-based classification can identify visual soil categories, but appearance alone does not establish nutrient concentrations or crop suitability.

2.3 Sensor-based soil analysis

Senapaty et al. [5] presented an IoT-enabled approach to soil nutrient analysis and crop recommendation, while Choudhari et al. [6] investigated soil-parameter monitoring and soil-health prediction using machine learning and data fusion. The former is directly related to crop recommendation; the latter addresses a related but different prediction target and is therefore not treated as a direct performance benchmark.

2.4 Multimodal agricultural learning

Youssef et al. [7] describe a crop-recommendation framework that combines soil-image analysis with geospatial, climatic and tabular information. Pandey et al. [8] present an arXiv preprint that combines soil-image analysis with nutrient profiling for a similar purpose; it has not undergone peer review as a journal article. Since earlier work has already explored combining images with numerical data, this study does not claim that such integration is novel. Other related studies examine crop-health monitoring and yield prediction using deep learning and spectral analysis [9], soil IoT and UAV multispectral data [10], multispectral–IoT fusion for pest and soil-health prediction [11], and image processing with IoT sensor data for crop-health assessment [12]. These studies address different outputs and are not direct benchmarks for crop classification.

2.5 Comparison of related studies

Table 1. Related studies and their relevance to the present work. Only information supported by the cited titles and the discussion above is shown.

Study	Objective	Input data	Comparability to this study
Dey et al. 2024 [1]	ML-based crop recommendation, India	NPK, soil pH, climatic variables	Direct (numerical task)
Fattepur et al. 2024 [2]	Crop recommendation with IoT, ML and DL	IoT-derived data	Related (data source differs)
Gyasi & Purushotham 2023 [3]	Review of DL soil classification	Soil images	Background for image task
Sreelatha et al. 2026 [4]	Automated soil classification and crop recommendation	Not stated here	Related (combined task)
Senapaty et al. 2023 [5]	IoT soil nutrient analysis and crop recommendation	Sensor data	Related (sensor-based)
Choudhari et al. 2024 [6]	Soil monitoring and health prediction	Sensor readings	Related (different target)
Youssef et al. 2026 [7]	Crop recommendation framework	Soil images, geospatial, climatic, tabular	Overlaps multimodal direction

Study	Objective	Input data	Comparability to this study
Pandey et al. 2025 [8] (preprint)	Crop recommendation via soil images and nutrient profiling	Soil images, nutrient profile	Overlaps multimodal direction
Jauhari et al. 2026 [9]	Crop-health monitoring and yield prediction	Images and spectral data	Adjacent (different target)
Kadam et al. 2025 [10]	Predictive crop-health analytics	Soil IoT, UAV multispectral	Adjacent (different target)
Girisha et al. 2026 [11]	Early pest and soil-health prediction	Multispectral and IoT	Adjacent (different target)
Mishra et al. 2026 [12]	Crop-health assessment	IoT sensor data and images	Adjacent (different target)

A numerical ranking of results across studies is not presented because the reviewed papers do not establish that they use the same dataset version, class definitions, data split and evaluation protocol. Table 1 instead identifies the studies most relevant to each task. Cross-study performance comparisons are meaningful only when the necessary results and evaluation details can be verified in the original publications; values have therefore not been inferred from abstracts or secondary sources.

2.6 Research gap and aim

The reviewed studies use different inputs and address different prediction tasks. Their reported scores cannot be compared directly unless the dataset composition, class definitions, train–test separation and evaluation metrics are sufficiently aligned. High performance on a public benchmark also does not, by itself, demonstrate generalisation to independent field conditions.

The study focuses on two questions: how well do four standard classifiers distinguish among 22 crop classes when cross-validation, uncertainty and class-level errors are reported, and what baseline performance can a CNN achieve on a separate five-class soil-image dataset? Combining numerical and image data, assessing soil health and validating predictions in the field are outside the scope of this work.

III. RESEARCH METHODOLOGY

Two datasets were used in two independent experiments (Table 2). Records in the numerical dataset and images in the image dataset are not paired.

Table 2. Datasets and data partitions.

Dataset	Samples	Input	Classes	Partition	Task
Numerical crop dataset	2,200	Seven soil and climatic features	22	1,760 training / 440 test (80:20 stratified); 5-fold CV on training	Crop recommendation
Soil-image dataset	552	Soil images	5	352 training / 89 validation / 111 test	Soil-image classification

3.1 Numerical dataset

The numerical dataset contains 2,200 records and seven input features: nitrogen (N), phosphorus (P), potassium (K), temperature, relative humidity, soil pH and rainfall. The target variable is the recommended crop, represented by 22 classes. The stratified test set contains 440 records, with 20 records per class (Fig. 3), consistent with the source dataset's 100 records per class.

The file used for this experiment, Crop_recommendation.csv, contains 2,200 rows and eight columns. Its schema and class distribution match the Crop Recommendation Dataset attributed to Atharva Ingle on Kaggle [14]. The dataset contains 22 crop labels, with 100 records per class. The input features are N, P, K, temperature, humidity, pH and rainfall. The dataset description reports N, P and K as nutrient-content values or ratios, temperature in degrees Celsius, humidity as a percentage, pH as dimensionless and rainfall in millimetres [14]. Inspection of the supplied CSV found no missing values or exact duplicate rows. The observed ranges were N: 0–140, P: 5–145, K: 5–205, temperature: 8.83–43.68 °C, humidity: 14.26–99.98%, pH: 3.50–9.94 and rainfall: 20.21–298.56 mm. A formal outlier analysis was not recorded. The crop labels were supplied with the dataset; the available materials do not establish that they were independently validated through field trials or generated independently of the input-feature construction. The results should therefore be interpreted as benchmark classification performance rather than evidence of field-level validity.

3.2 Soil-image dataset

The image dataset contains 552 images in five classes. The confusion matrix labels them type_1 to type_5 (Fig. 4); the test partition holds 22, 19, 28, 21 and 21 images respectively. Example images are shown in Fig. 1.



Fig. 1. Example images from the soil-image dataset (six samples; class membership is not indicated).

The image dataset appears to correspond to *SoilClass: A Novel Image Dataset for Soil Classification*, released as Mendeley Data Version 1 (DOI: 10.17632/6mx34nshm4.1) [13]. The source describes 552 images across five categories under a CC BY 4.0 licence. It reports that images were captured using mobile cameras under different angles and lighting conditions and that the five soil categories were tested by a soil-testing laboratory in West Midnapore, India [13]. In this study, the categories are retained as Type 1–Type 5 (type_1–type_5 in the model output); they are not assigned soil-taxonomy names because the available documentation does not establish such mappings. The source reports class counts of 109, 96, 137, 107 and 103 for Types 1–5, respectively [13]. The notebook resizes images to 160 × 160 pixels, decodes them as three-channel images and scales pixel values to [0, 1]. The test partition contains 22, 19, 28, 21 and 21 images for Types 1–5. The available files do not record the camera models, camera-to-sample distances or full class counts for the training and validation partitions. In addition, a duplicate or near-duplicate image audit has not been completed, so possible sample-level leakage must be considered when interpreting the test results.

3.3 Preprocessing and data partitioning

For the numerical data, the seven features formed the input and the crop category the target. The data were divided 80:20 using stratified sampling, giving 1,760 training and 440 test records. Five-fold stratified cross-validation was run on the training partition only, and the test partition was used once for final evaluation of each model.

The supplied implementation uses a scikit-learn Pipeline with a ColumnTransformer. Numerical features are processed using median imputation followed by StandardScaler; categorical features, if present, are processed using most-frequent imputation and one-hot encoding. All seven features in the supplied CSV are numerical, and no missing values were found. Because preprocessing is included in the pipeline passed to cross-validation, its steps are fitted separately within each training fold. The hold-out split uses test_size=0.20, stratify=y and random_state=42. Five-fold StratifiedKFold uses shuffle=True and random_state=42. Random Forest, Extra Trees and Decision Tree use random_state=42. SVM-RBF uses kernel='rbf' and class_weight='balanced'; because C and gamma were not explicitly set, the scikit-learn defaults apply (C=1.0 and gamma='scale').

For the image data, 552 images were divided into 352 training, 89 validation and 111 test images. The validation set was used for training decisions and the test set for final evaluation.

The image pipeline obtains class labels from the parent-folder names and sorts them to assign integer labels. Images are decoded as three-channel images, resized to 160 × 160 pixels and divided by 255 to scale pixel values to [0, 1]. The batch size is 32. The training dataset is shuffled using seed 42, while the validation and test datasets are not shuffled. The split is stratified: 20% of the images are reserved for testing, and 20% of the remaining images are used for validation. This produces 352 training images, 89 validation images and 111 test images. The supplied notebook does not apply data augmentation and does not implement sample-level grouping or a duplicate/near-duplicate check. The split uses random seed 42.

3.4 Numerical models

Random Forest combines multiple decision trees trained on bootstrap samples, with feature subsampling used when selecting splits. Extra Trees introduces additional randomness into split-threshold selection. A single Decision Tree provides a simpler baseline, while SVM-RBF models nonlinear decision boundaries using a radial basis function kernel and is sensitive to feature scaling. The configurations used in this study are summarised in Table 3.

Table 3. Hyperparameters and software for the numerical models.

Model	Hyperparameters	Tuning procedure
Random Forest	n_estimators=300; class_weight="balanced"; random_state=42	Fixed configuration
Extra Trees	n_estimators=300; class_weight="balanced"; random_state=42	Fixed configuration
Decision Tree	class_weight="balanced"; random_state=42	Fixed configuration
SVM-RBF	kernel="rbf"; C=1.0; gamma="scale"; class_weight="balanced"	Fixed/default configuration

All four numerical models were evaluated using fixed hyperparameter settings; no systematic hyperparameter search was conducted. The reported results therefore describe the selected configurations, not the best performance that might be obtained through tuning. This distinction should be considered when interpreting the comparison.

The implementation uses Python and the pandas, NumPy, Matplotlib, scikit-learn, Pillow and TensorFlow/Keras libraries. The supplied notebook does not record package versions, hardware specifications or an environment manifest. These details should be reported from the actual execution environment where available; they cannot be reliably inferred from the exported result files alone.

3.5 CNN architecture and training

The CNN was trained to classify soil images into five categories. Its architecture and training settings are summarised in Table 4. The training curve shown in Fig. 5 covers 24 epochs; this epoch count should be checked against the original run log before submission.

Table 4. CNN architecture and training settings.

Item	Setting
Convolutional layers, pooling, activations	Conv2D(32, 3×3, ReLU) → MaxPooling2D → Conv2D(64, 3×3, ReLU) → MaxPooling2D → Conv2D(128, 3×3, ReLU) → GlobalAveragePooling2D → Dense(64, ReLU) → Dense(5, softmax).
Regularisation (dropout, weight decay)	Dropout(0.30); no weight decay specified.
Input size	160 × 160 × 3; images resized and pixel values scaled to [0, 1].
Optimiser and learning rate	Adam; learning rate not explicitly set in code (Keras default).
Batch size	32.
Loss function	Sparse categorical cross-entropy.
Epochs and stopping rule	Maximum 25 epochs; EarlyStopping monitors val_loss, patience=5, restore_best_weights=True. Manuscript training curve shows 24 epochs; verify against the original run log.
Checkpoint used for test evaluation	EarlyStopping restores best validation-loss weights; no separate ModelCheckpoint file is configured.
Pretrained weights or training from scratch	Training from scratch; no pretrained weights are specified in the supplied notebook.

3.6 Evaluation metrics

For class c, let TP_c, FP_c and FN_c be its true positives, false positives and false negatives, N the number of evaluated samples and C the number of classes.

$$Accuracy = (Number\ of\ correct\ predictions) / N$$
$$Precision_c = TP_c / (TP_c + FP_c)$$
$$Recall_c = TP_c / (TP_c + FN_c)$$
$$F1_c = 2 \times Precision_c \times Recall_c / (Precision_c + Recall_c)$$
$$Macro\text{-}metric = (1 / C) \times \Sigma_{c=1..C}\ metric_c$$

Macro-averaged precision, recall and F1-score assign equal weight to each class. Because the numerical test set contains the same number of records for each class (20 per class), macro recall equals accuracy. Confusion matrices were used to examine class-level errors. The 95% confidence intervals for held-out accuracy were calculated using Wilson score intervals based on the number of correct predictions and the total number of test samples.

3.7 Scope of the experiments

The numerical and image experiments are independent. Neither dataset provides laboratory-measured soil-health indicators, and no model combining the two was built. Results from one experiment must not be read as evidence about the other.

IV. RESULTS

4.1 Cross-validation on the numerical data

Table 5 summarises five-fold stratified cross-validation on the 1,760-record training partition. Random Forest achieved the highest mean accuracy (99.318%) and macro F1-score (99.315%), followed closely by Extra Trees, with a mean accuracy of 99.261% and macro F1-score of 99.258%. Decision Tree and SVM-RBF achieved mean accuracies of 98.523% and 97.898%, respectively. The standard deviations reflect variation across the five folds. Although Random Forest recorded slightly higher mean scores than Extra Trees, the difference is small and does not establish statistically significant superiority.

Table 5. Five-fold stratified cross-validation results on the training partition (mean ± SD across folds, %).

Model	Accuracy	Macro precision*	Macro recall*	Macro F1
Random Forest	99.32 ± 0.53	99.39 ± 0.45	99.32 ± 0.53	99.31 ± 0.53
Extra Trees	99.26 ± 0.39	99.31 ± 0.35	99.26 ± 0.39	99.26 ± 0.39
Decision Tree	98.52 ± 0.68	98.63 ± 0.63	98.52 ± 0.68	98.52 ± 0.68
SVM-RBF	97.90 ± 1.03	98.22 ± 0.80	97.90 ± 1.03	97.88 ± 1.06

Values are reported as the mean ± population standard deviation (ddof = 0) across the five fold-level scores. The metrics were recalculated using the supplied numerical-model workflow, Crop_recommendation.csv and random_state=42. Precision and recall standard deviations are rounded to two decimal places.

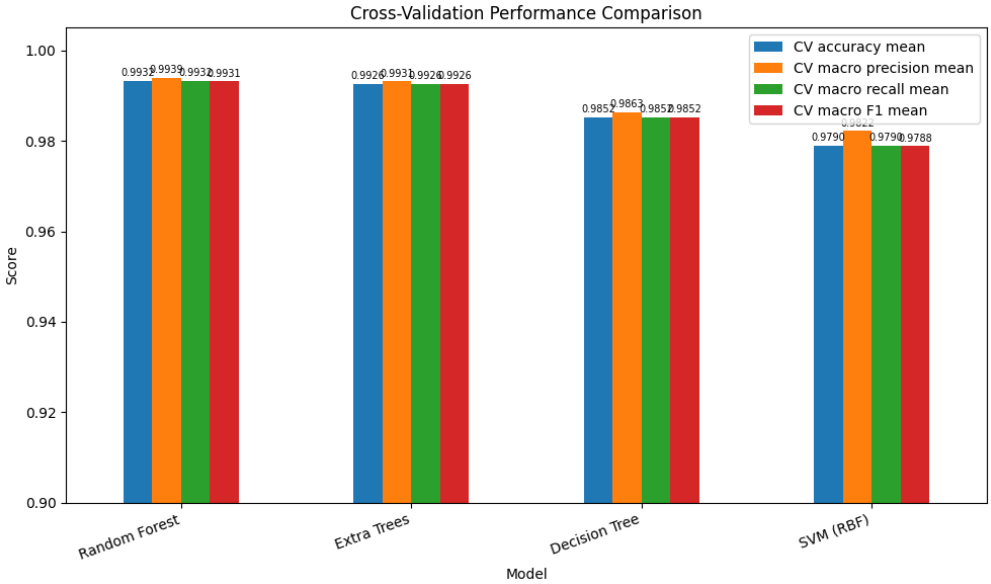

Fig. 2. Mean five-fold cross-validation accuracy, macro precision, macro recall and macro F1-score for the four numerical models. The vertical axis starts at 0.90, which visually enlarges small differences.

4.2 Held-out test results on the numerical data

Table 6 summarises performance on the 440-record held-out test set. Random Forest and Extra Trees achieved identical results, with 99.55% accuracy and a 99.55% macro F1-score. Decision Tree achieved 97.96% accuracy, while SVM-RBF achieved 98.41%. The Random Forest confusion matrix shows two misclassifications: one blackgram record was predicted as maize and one rice record was predicted as jute (Fig. 3).

Table 6. Held-out test performance of the numerical models (%). Intervals are 95% Wilson intervals for accuracy computed from the number of correct predictions out of 440.

Model	Accuracy	Macro precision	Macro recall	Macro F1	Accuracy 95% interval
Random Forest	99.55	99.57	99.55	99.55	98.4 – 99.9
Extra Trees	99.55	99.57	99.55	99.55	98.4 – 99.9
Decision Tree	97.96	98.06	97.96	97.94	96.2 – 98.9
SVM-RBF	98.41	98.56	98.41	98.40	96.8 – 99.2

The confidence intervals overlap, and no paired significance test was performed. The observed ranking should therefore be interpreted cautiously; the results do not establish statistically significant differences among the models. Random Forest and Extra Trees also cannot be distinguished by their aggregate test scores. Decision Tree performed better than SVM-RBF in cross-validation (98.52% versus 97.90%), whereas SVM-RBF performed better on the test set (98.41% versus 97.96%). Differences of this size may reflect sampling variation.

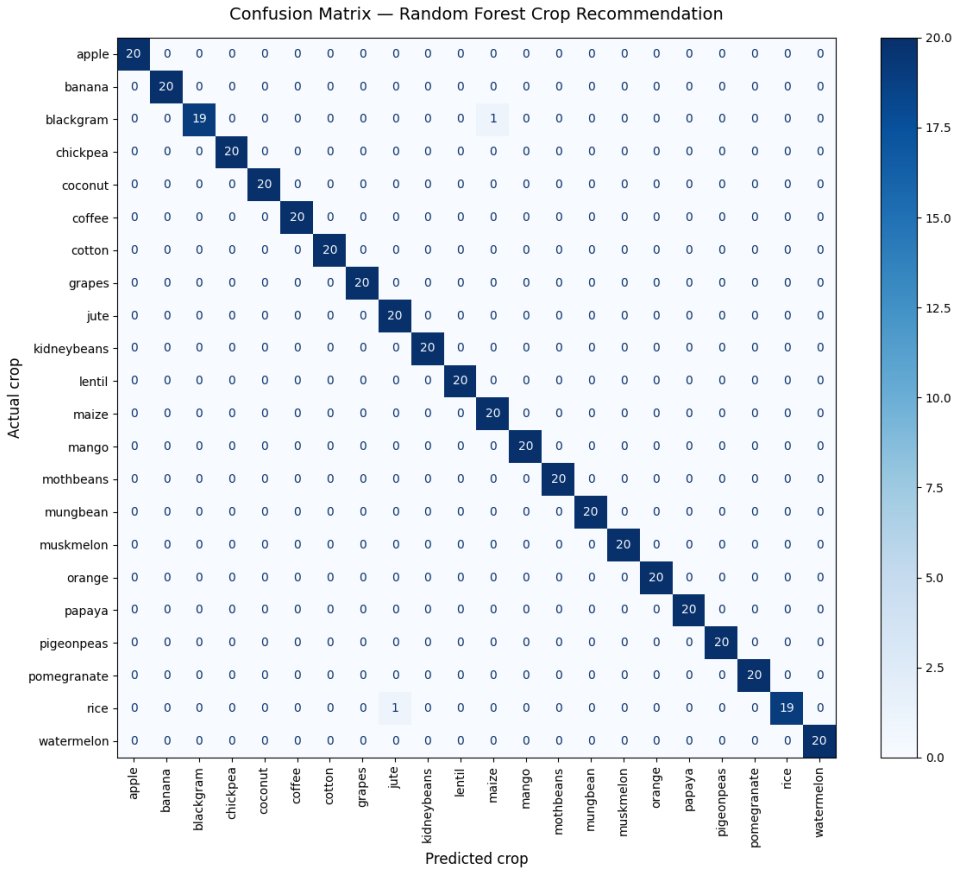

Fig. 3. Confusion matrix of Random Forest on the 440-record held-out test set (22 crop classes, 20 records per class).

The exported results report identical aggregate test metrics for Random Forest and Extra Trees. However, an Extra Trees confusion matrix was not available among the exported files, so it is not included here. The matching aggregate scores should be verified against the original Extra Trees predictions, and its confusion matrix should be retained with the study records.

4.3 CNN soil-image classification

On the 111-image test set, the CNN achieved 91.89% accuracy (102 correct predictions; 95% Wilson confidence interval: 85.3–95.7%), 93.11% macro precision, 91.13% macro recall and 91.40% macro F1-score (Table 7). These values agree with the confusion matrix shown in Fig. 4.

Table 7. CNN performance on the held-out soil-image test set (n = 111).

Metric	Result (%)
Accuracy	91.89
Macro precision	93.11
Macro recall	91.13
Macro F1-score	91.40

The class-level results in Table 8 were calculated from the confusion matrix. The model classified type_1 and type_3 without error. Eight of the nine errors were predictions of type_5: three type_2 images and five type_4 images were assigned to type_5. This pattern reduced type_5 precision to 71.43% and type_4 recall to 76.19%. One type_5 image was incorrectly classified as type_2.

Table 8. Per-class CNN test metrics (%) computed from the confusion matrix in Fig. 4.

Class	Precision	Recall	F1-score	Test images
type_1	100.00	100.00	100.00	22
type_2	94.12	84.21	88.89	19
type_3	100.00	100.00	100.00	28
type_4	100.00	76.19	86.49	21
type_5	71.43	95.24	81.63	21
Macro average	93.11	91.13	91.40	111

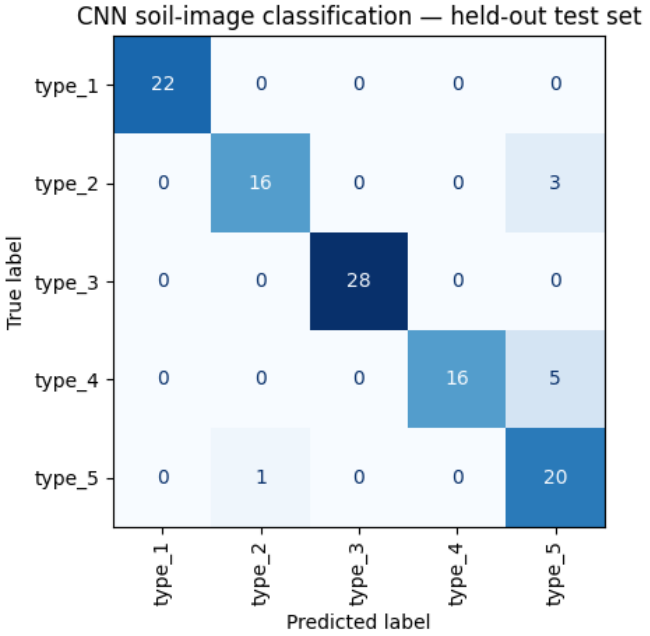

Fig. 4. Confusion matrix of the CNN on the 111-image held-out test set. Classes are labelled type_1 to type_5 as in the experiment output.

Figure 5 shows training and validation accuracy over 24 epochs. Training accuracy rose to approximately 0.89–0.91. After the early epochs, validation accuracy fluctuated between about 0.76 and 0.89 and ended near 0.82. This variation is unsurprising given the small validation set of 89 images. The final test accuracy (91.89%) was higher than the final validation accuracy; this difference should be interpreted cautiously because both estimates are based on limited samples.

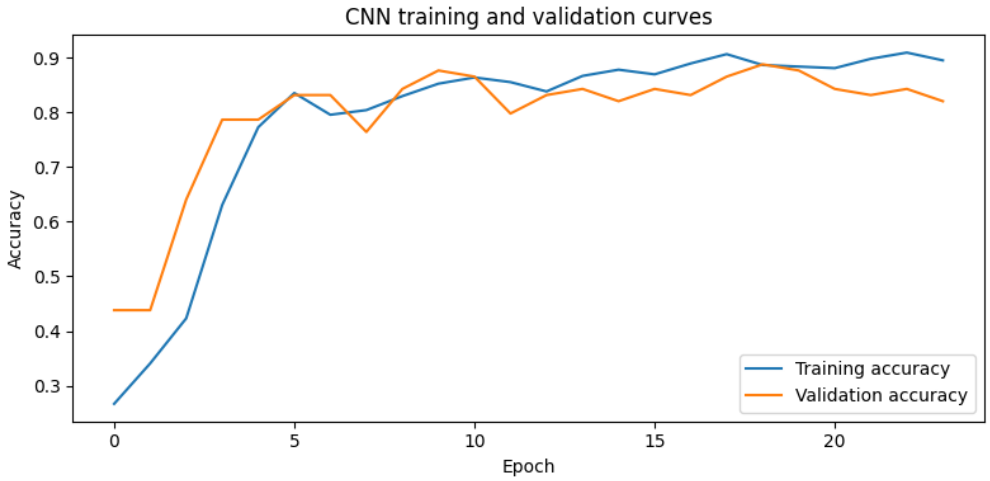

Fig. 5. CNN training and validation accuracy over 24 epochs.

V. DISCUSSION

5.1 Numerical crop recommendation

All four classifiers achieved at least 97.9% accuracy on the 22-class numerical task, with Random Forest and Extra Trees producing the highest scores. However, the differences were small. The two ensemble models differed by only a small amount in cross-validation and achieved identical aggregate results on the test set. The overlapping confidence intervals and absence of a paired significance test do not support declaring one model definitively superior. The results indicate that tree ensembles are strong candidates for this particular dataset.

Performance close to 100% may indicate that the crop classes are highly separable within this benchmark, rather than demonstrating performance under challenging real-world conditions. The two Random Forest errors involved blackgram being classified as maize and rice being classified as jute. These errors may reflect overlap between the feature patterns of particular crop classes. However, because the dataset's label-generation process is not fully documented and duplicate or near-duplicate records have not been ruled out, these results should not be interpreted as evidence of generalisation to independent farms, seasons or measurement devices.

5.2 Soil-image classification

The CNN achieved 91.89% accuracy on the five-class task, but the test set contains only 111 images, so the confidence interval remains relatively wide. Errors were concentrated in type_5: several type_2 and type_4 images were predicted as type_5. Visual similarity, differences between partitions or class composition may contribute to this pattern, but the current results do not establish the cause. The validation accuracy of 0.438 in the first two epochs corresponds to 39 correct predictions out of 89 and may reflect a temporary tendency to predict one class. Reporting class counts for each partition would help clarify this behaviour. Until duplicate and near-duplicate images have been checked across partitions, the test result may overestimate performance on genuinely unseen samples.

The image result cannot be compared with the numerical results, because the datasets, labels and tasks differ.

5.3 Relation to previous work

Studies [1] and [2] address numerical crop recommendation, while studies [7] and [8] explore combining soil images with numerical information. The dataset used in the present numerical experiment is the widely distributed 2,200-record, 22-class crop-recommendation benchmark [14]. A numerical comparison across studies is not provided because the available sources do not establish that the experiments use matching dataset versions, class mappings, splits and evaluation protocols. Cross-study comparisons should be added only when the relevant metrics and methods can be verified directly from the original publications.

5.4 What the results do not show

These experiments do not demonstrate that combining numerical and image data improves crop recommendation, because no combined model was developed. They also do not establish soil-health status: the crop labels are not laboratory-validated soil-health measurements. Finally, they do not demonstrate performance on real farms, as no independent field dataset was used.

VI. CONCLUSION

Random Forest and Extra Trees each achieved 99.55% test accuracy on the numerical dataset covering 22 crop categories. In the separate image experiment, the CNN achieved 91.89% test accuracy for five-class soil-image classification. These results provide baseline performance on the datasets used, but high scores on public benchmarks should be interpreted cautiously. Crop recommendation based on benchmark features cannot replace direct soil testing, and results from two independent datasets do not validate a multimodal fusion framework. Future work should evaluate the models on independent, multi-season field data, document dataset provenance, assess sensor calibration and investigate numerical–image fusion using paired samples and validation by agronomic experts.

REFERENCES

[1] B. Dey, J. Ferdous, and R. Ahmed, “Machine learning based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables,” Heliyon, vol. 10, no. 3, Art. no. e25112, 2024, doi: 10.1016/j.heliyon.2024.e25112.

[2] B. Fattepur et al., “Cultivating prosperity: A fusion of IoT data with machine learning and deep learning for precision crop recommendations,” in 2024 5th International Conference for Emerging Technology (INCET), 2024, doi: 10.1109/INCET61516.2024.10593556.

[3] E. K. Gyasi and S. Purushotham, “Advancements in soil classification: An in-depth analysis of current deep learning techniques and emerging trends,” Air, Soil and Water Research, vol. 16, 2023, doi: 10.1177/11786221231214069.

[4] A. Sreelatha, P. Praveen, R. K. Gatla et al., “Hybrid deep learning model for automated soil classification and crop recommendation to enhance agricultural productivity,” Scientific Reports, vol. 16, 2026, doi: 10.1038/s41598-026-66201-0.

[5] M. K. Senapaty, A. Ray, and N. Padhy, “IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture,” Computers, vol. 12, no. 3, p. 61, 2023, doi: 10.3390/computers12030061.

[6] A. Choudhari, D. Bhoyar, and W. P. Badole, “Design of an improved model for soil parameter monitoring and health prediction using machine learning and data fusion techniques,” in 2024 2nd International Conference on Emerging Trends in Engineering and Medical Sciences (ICETEMS), 2024, doi: 10.1109/ICETEMS64039.2024.10965059.

[7] S. Youssef, K. Gamage, and F. Zablith, “An integrated AI framework for crop recommendation,” Horticulturae, vol. 12, no. 4, p. 416, 2026, doi: 10.3390/horticulturae12040416.

[8] V. Pandey, R. Das, and D. Biswas, “AgroSense: An integrated deep learning system for crop recommendation via soil image analysis and nutrient profiling,” arXiv preprint arXiv:2509.01344, 2025, doi: 10.48550/arXiv.2509.01344.

[9] S. Jauhari, K. K. Agrawal, S. P. Yadav, and A. Quezada, “Integrating deep learning and spectral analysis for multi-modal data fusion in precision agriculture for enhancing crop health monitoring and yield prediction,” FMDB Transactions on Sustainable Computing Systems, vol. 4, no. 1, pp. 63–77, 2026, doi: 10.69888/FTSCS.2026.000610.

[10] A. Kadam, N. Bansod, S. Bakal, S. Ghodke, F. Shaikh, and S. Shelke, “Predictive crop health analytics via sensor fusion: Integrating soil IoT, UAV multispectral data, and machine learning,” in 2025 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS), 2025, doi: 10.1109/ICBDS67396.2025.11377223.

[11] G. S. Girisha, B. A. Mala, K. Maruthi, K. Murugesh, M. R. Nadaf, and A. Mohan Kumar, “CropGuard AI: Multispectral-IoT fusion framework for early pest and soil health prediction,” in Proc. 9th International Conference on Trends in Electronics and Informatics (ICOEI), 2026, pp. 1407–1414, doi: 10.1109/ICOEI68323.2026.11542058.

[12] S. Mishra et al., “Fusion of IoT sensor data and image processing for comprehensive crop health assessment,” International Journal of Drug Delivery Technology, vol. 16, no. 40s, 2026, doi: 10.25258/ijddt.16.40s.93.

[13] K. G. Dhal, A. Das, B. Chakraborty, and S. Pan, “SoilClass: A Novel Image Dataset for Soil Classification,” Mendeley Data, Version 1, June 11, 2026, doi: 10.17632/6mx34nshm4.1. Licence: CC BY 4.0.

[14] A. Ingle, “Crop Recommendation Dataset,” Kaggle. Available: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset. Accessed October 29, 2026.