Sigma J Eng Nat Sci, Vol. 44, No. 5, pp. 3157–3166, October, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00137

Research Article

Smart farming: Improved soil classification and yield forecasting using CNNs and machine learning algorithms

Uttam R. PATOLE[1,2,*], Manish SHRIVASTAVA[1]

[1]*Vivekananda Global University, Jaipur, 303012, India*
[2]*Sir Visvesvaraya Institute of Technology, Nashik, 422102, India*

ARTICLE INFO

Article history
Received: 23 March 2025
Revised: 05 May 2025
Accepted: 10 July 2025

Keywords:
Convolutional Neural Network; Crop Prediction; Machine Learning; Precision Agriculture; Soil Classification

ABSTRACT

Agriculture is the key driving force for food security and financial stability of a nation and specially in the rural economy where farmers face several challenges including lack of proper guidance, unpredictable procurement prices, uncertain yield prediction and inefficient crop selection. The present study presents a scalable AI based approach, which includes CNN and ML models for better coordination of soil classification and crop yield prediction. An RGB-image-based CNN model was developed to classify the four main soil types—Red, Black, Clay, and Alluvial — with high accuracy. In parallel, a 'Best Crop Prediction' module was built with multiple linear regression and was trained with more than a decade agricultural data consisting of type of soil, crop yield and regional environment. The models are packaged within a user-friendly web app where farmers can upload pictures of their soil and response with their geographic information to obtain crop recommendations based on these models. It is a friendly system for the general public (non-clients), so that precision agriculture becomes an accessible practice for everyone. By automating soil testing and crop selection, it's designed to fill knowledge and resource voids, enabling farmers to make better decisions that optimize productivity and sustainability.
The model obtained a soil classification accuracy of 94% and contributed to increasing the accuracy of the crop recommendation by up to 28% against traditional methods. The real time prediction and user-friendly in the web application. Field trials verified the applicability of the system in different farming zones, and thus its potential for wide implementation. Next steps will include adding IoT sensors, and dynamic environmental data for further optimization.

Cite this article as: Patole UR, Shrivastava M. Smart farming: Improved soil classification and yield forecasting using CNNs and machine learning algorithms. Sigma J Eng Nat Sci 2026;44(5):3157–3166.

INTRODUCTION

Agriculture is the basis of human civilization: the cornerstone for food security, livelihoods, and economic development. The industry supports the livelihoods of billions of people around the world, particularly in agrarian nations who derive income and employment from farming.

***Corresponding author.**
*E-mail address: uttam.patole@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

Over 58% of the people in India depend upon agriculture and is an integral component of Indian socio-economic structure. But this most important agriculture sector suffers from a series of chronic and constant challenges such as erratic climate, soil mining, inappropriate crops choice and dependence on tradition and customary ways. These conditions have resulted in low productivity, income instability, and severe post-harvest losses. In recent years, the need to solve these problems has become a brighter pressing, particularly with climate academia and about _ worldwide increases in food demand due to climate change and declines in arable land. Technological breakthroughs—especially those that originate from computer science—are being increasingly regarded as necessary instruments for sustainable and profitable agriculture [1-5].

One of the most potential technologies is the machine learning (ML), which has the ability to analyze the complex data in agriculture and extract tools of decision-making. Soil Classification & Crop Prediction One of the important use cases in agriculture would be use of ML for soil classification and crop prediction; both are tasks that would influence on yield optimization and resource efficiency in agriculture. Soil classifications showing the range of crops that are adapted to the particular soil in a given locality is an important contribution in the proper selection of the crop, while wisely chosen crop prediction models propose an informed policy for maximizing the production. Conventional techniques of soil analysis, however, generally require bulky sample work, and costly laboratory testing and are not practical for smallholder farmers. Additionally, current digital offerings typically require technical acumen to use or expensive equipment purchase, which prevents a wide spread acceptance. These issues are addressed by the present study which presents a functional ML-based architecture that connects state-of-the-art technology with field-based applicability. The goal is to develop an end-to-end system for the purpose of identifying soil types via CNNs and forecasting the most suitable crops for highest yield utilizing multiple linear regression — in order to provide farmers with an affordable and impactful decision support tool [5-7].

This is an investigation based in practice and not in abstraction. The suggested model uses RGB soil images to back those with Red, Black, Clay and Alluvial soil types with an accuracy of 94%, which shows the predictive power of the model. The crop recommendation module studies more than 10 years of historical agricultural data to recommend region specific crops according to the type of soil, climatic conditions and the previous production trend. These modules are combined into a light weight, user friendly, web application that farmers can use to upload soil images in order to receive personalized crop advice in few clicks. The interface is to be made simple with least / no technical knowledge required and hence it gains an approach towards accessibility and scope for widespread adoption. «Blending cutting-edge deep learning techniques with an intuitive frontend, this proof of concept demonstrates how AI can help traditional farming communities to make evidence-based decisions, minimise input wastage, and improve farming output in general. Finally, the project has the potential to aid in reversing the digital divide and set the foundation for later inclusion of IoT sensors and real-time analytics.

Literature Review

With the increasing quantity of agricultural data, data mining has become the most useful tool for getting valuable patterns that aid the decision process and increase productivity. In agricultural domain, methods like k-means, k-nearest neighbors, SVM, neural network and fuzzy c-means have been most extensively used. The selection of one of these methods depends on the type of data and the specific endpoints that are considered. The surveyed literature reduces studies together and it is good reference material for researchers and scientists involving in predictive modeling for intelligent agricultural applications specifically in developing countries like India [8].

According to Kamath et al. [9], Indian agriculture is controlled by a blend of organic, economic and seasonal forces, thus making crop yield estimation a challenging problem. Early prediction provides chance to better plan for storage, allocation of resources and marketing. Their research used data mining and the Random Forest method to improve the prediction accuracy of yield as well as to contribute to area-specific agricultural planning based on a data analysis. Mitra and Srinivasan [10] proposed a comprehensive environment for real-time soil classification and crop suggestion via an ensemble approach. Although their method characterized good generalization, it does not support practical use for general end users. Similarly, Gholap et al. [11] constructed soil fertility classification models based on machine learning to achieve automatic generation based on traditional soil analysis which need to be optimized. However, their model was only used for soil classification and not crop recommendation. Srunitha et al. [12], which proposed SVM based model for soil classification employing features such as pH and moisture content. Their study further emphasized the potential of image processing integrated into the ML on soil analysis automation, however they also pointed out scalability constraints owing to SVM's computing demand when larger datasets are considered [12].

Rahman et al. [13] has proposed a combined approach where SVM was used for soil classification, and regression-based applications were used for crop prediction. Although the potential to integrate these two part was shown in the study, accessible tools for farmers or non-expert users was not eventually produced. Akkem et al. [14] highlighted the role of advanced ML, deep learning, as well as time series analysis in handling smart farming constraints. They further recommended deploying models such as Extreme Learning Machine (ELM) and Stacking

Ensemble Learning to enhance learning rate and robustness, thereby facilitating better crop selection & reduced agricultural distress. Kumar et al. [15] found that machine learning (ML) in prediction of yield, in particular Random Forest algorithm [16], can be used for accurate prediction of the future, using historical information, including temperature, humidity, pH, rainfall, and type of crop. By connecting input variables (agrometeorological parameters) to output through ML we propose practical and efficient solutions for optimising the choice of crops for a specific weather regime [15].

Pandith et al. [17] trained a CNN to output yield prediction based on soil quality. They used 5 supervised machine learning algorithms-KNN [18], Naïve Bayes, Multinomial Logistic Regression, Artificial Neural Networks (ANN), and Random Forest for prediction of mustard crop yield, using soil data from different districts of the Jammu region. The measure of accuracy, precision, recall, specificity, and F-score would be taken into account to evaluate the performance of prediction [5] with the result indicating that the combination of KNN and alg-linear had the best prediction accuracy among all [17].

Even if many studies have employed machine learning for food insecurity prediction, not many studies have user-friendly, real-time tools available for farmers, while most of them are centered on individual tasks such as soil classification and crop yield prediction [19]. However, there is a lack of integrated scalable systems that exploit various data sources for comprehensive and actionable recommendations for smart farming.

MATERIALS AND METHODS

This section presents the methodology applied in building the proposed system for land type classification and crop prediction. It involves the extraction of data sources, the preprocessing of soil and crop datasets, the architecture of a CNN model for soil-type classification, and the deployment of a crop-prediction module which utilizes ML. It also discusses the development of a web-based interface of end-user access.

Data Sources

Our proposed system leverages upon two fundamental datasets, which are:

Soil Image Dataset: The dataset is extracted from Kaggle and consists of 903 labelled RGB images.Blac, red, clay and alluvial are the 4 soil types. The data was split into 80-20% train and test data [20].

Crop dataset: It is a small dataset extracted from the agriculture repositories which has a information regarding 11 different crops and the production import/export details and the yield variation at region level for last ten years [21].

Crop Suitability for Different Soil Types

The matrix of soil-to-crop compatibility plays an indispensable role in the crop prediction block of the proposed system. A listing of approved crops for each of the four soil types (i.e., Alluvial, Red, Black, and Clay) according to the soil classification system is provided in the following table (Table 1). Prior to this mapping being used, it was based on common agronomist knowledge and agricultural reference sources such as national soil libraries and agricultural extension literature.

These relationships have been largely established through decades of empirical cultivation results and studies of soil fertility that show which crops respond well to soil texture, nutrient levels, drainage, and wilting characteristics. This compatibility system is also used by the prediction engine in conjunction with classified soil type, climatic inputs (e.g. temperature and rainfall) and regional productivity data, to provide individual farmers with crop recommendations.

Data Preprocessing

The datasets were re-organized, and their preprocessing done in order to maintain them in the appropriate shape and clean for use [19].

Soil Images: Images (244 × 244 pixels) were used and data augmentation processes including rotation, flipping and brightness were employed to expand generalization capabilities [22].

Crop: Dataset: Replace missing data was performed by using mean method and convert categorical variable to numerical variable (e.g., crops). Standardization was used to improve the prediction performance [23] when training the model for the data.

CNN Architecture for Soil Classification

The Soil type-classification Module [24] makes use of a Convolutional Neural Network (CNN) to classify soil

Table 1. Soil type and recommended crops

Soil Type	Recommended Crops
Alluvial	Wheat, Rice, Jute, Coconut, Sugarcane, Pulses, Oilseeds, Groundnut
Red	Wheat, Cotton, Pulses, Coconut, Tobacco, Millets, Oilseeds, Potato, Groundnut, Rice, Orchards
Black	Cotton, Pulses, Soybean, Millets, Linseed, Tobacco, Barley, Sugarcane, Rice
Clay	Wheat, Rice, Sorghum, Jowar, Groundnut

images into 4 different soil types, namely Red, Black, Clay and Alluvial. The model involves 3 layers of convolutions, followed by ReLU, max-pooling and dropout sequential. A fully connected dense layer is fed by the last feature representation.

Model Architecture Overview:

Convolutional Layer 1: Applies 32 3x3 filters (normal pointwise convolution) with ReLU activation function.

Second Convolutional Layer: 64 filters of size 3×3, ReLU activation.

Third Convolutional Layer: 128 filters of 3×3 with ReLU as the activation again.

Fully Connected Output Layer: A dense layer with softmax activation for the final classification.

We have trained the CNN with the Adam optimizer and sparse categorical cross-entropy as a loss function [25]. The test dataset accuracy of 94% of the model was impressive, indicating the model to be effective for soil type classification.

Best Crop Prediction Module

The crop forecast module uses the multiple linear regression on the 10 years of agricultural information [26]. This module takes into account the soil type, rainfall, production history and climatic conditions to predict the crops that are best suited to a region [24]. The prediction model acts to provide a list of ranked crops by expected yield and profitability after fusing the soil classification outcome of the CNN module [27].

System Integration and Web Application

The application was developed as a Django web application for usability. The application is essentially a portal to which farmers can upload images of soil and region specific specifications. The input is handled at the backend and with the involvement of the soil classification using the predefined CNN based on which the data are combined with the crop prediction module gives meaningful insights [28].

Evaluation Metrics

We tested the proposed system using the following measurements:

Soil Classification using CNN Accuracy: The CNN model developed had an accuracy of ~94% for soil classification.

Accuracy of prediction: The crop prediction model showed a high level of accuracy with R-squared (R^2) values greater than 0.9, denoting a high correlation with actual yield [29].

User response: Usability of the web application with the stakeholders of interest was favorable and the intuitive design and ease of access received positive feedback.

The method represents the staged sequence of events for systems development, verifying accuracy, scalability, and implementation applicability under real farm conditions.

System Design

Therefore, we will use a combination of multiple machine learning to classify the soil as well as predict the crops will be planted in the soil in order to design the proposed system. The system›s architecture consists of three main parts: soil classification, crop prediction and user-friendly web interface. These modules interact to interpret inputs, classify data, and provide recommendations

Overall Architecture

The soil images and regional agricultural information are fed to the system. Thus, the building blocks of this overall architecture are:

Soil Classification: We demodulate the soil images from the above results and further utilize a CNN [30] to classify soil.

Crop Prediction module: Once soil classification, historical data and regional factors are collected this module predicts the best crops [24].

Web application: Serves as the front-end interface for users to input data and retrieve recommendations [28].

The data flow is presented in Figure 1, showing the complete end-to-end processing pipeline from the capturing of the soil images to the prediction of soil crop and the generation of the results to be presented to the user.

Soil Classification Module

It employs a CNN to recognize soil types according to the input images. In this section we describe the CNN design and its realization.

- CNN Architecture

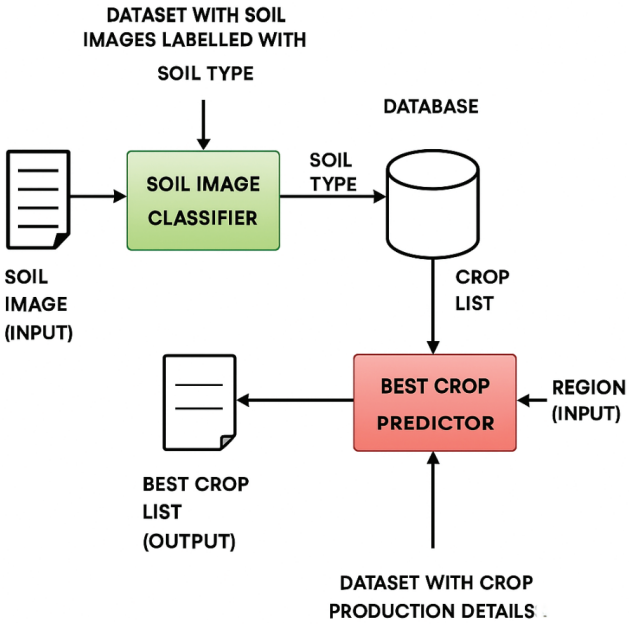


Figure 1. Overall architecture diagram.

- The CNN model is constructed with layers for feature extraction from soil images.
- Convolutional Layers- these layers are responsible for convoluting filters for detection of features viz texture, color and patterns which differentiate variety of soil types [30].
- Max-Pooling Layers: These layer reduce the dimensionality of the feature maps and retain essential information for efficient computation [31].
- Fully Connected Layers: Use the features of the extract layer for the soil type classification [20].

On a deeper level, convolution is represented like this:

$$f(x,y)=\Sigma_{(i=0)}^{(m-1)} \Sigma_{(j=0)}^{(n-1)} w(i,j)\cdot I(x+i,y+j) \quad (1)$$

where:

- f(x,y) is the output of the convolution,
- w(i,j) is the filter applied to the input image I(x,y),
- m and n are the dimensions of the filter.

The last layer uses softmax activation to classify the soil into one of four class types.

$$P(y=i|x)=e^{(z_i)}/(\Sigma_j e^{(z_j)}) \quad (2)$$

where:

- Output of the network for the class whose index is iii, from the first shape you provided, can be represented as:
- P(y=i|x) is the probability of the input image belonging to class i, zi is the score (or logit) for class I, $\Sigma_j e^{(z_j)}$ is the sum of exponentiated scores over all classes j

The details of a CNN such as the output layer are shown in Figure 2, here two hidden layers are there with two inputs.

Model training and evaluation

The CNN was optimised using an Adam optimizer with a sparse categorical cross-entropy loss. Its accuracy was evaluated, and it reached 94% on the test set.

Crop Prediction Module

The crop prediction module is based on the combined output of the soil classification and the historical agricultural data to recommend the best crops for a region. Another study predicts crop yield using factors such as soil type, precipitation, temperature, and historical yield data based on a multiple linear regression model [32]. The multiple linear regression is written as follows mathematically:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \quad (3)$$

where:

- Y is the estimated value of the crop yield..
- $X_1, X_2, \ldots, X_n$ are input variables such as soil types, rainfall, temperature, amount of fertilizer used and so on.
- β_0 is the intercept term, indicating the baseline crop yield when all input features are assumed to be zero.
- β_1 to β_n are the regression coefficients that quantify the influence or contribution of each input feature to the predicted crop yield.

The prediction (Y (β)) is a linear combinationof the observations with X: whereβ is the set of coefficients estimated by the least squaresdefault algorithm to minimize the difference betweenY (β) and Y true. Figure 3 shows the whole system design of crop prediction [27].

Web Application Design

Users (farmers) can coax the system through the web application. The app comes with:

Soil Image Upload: Users are able to upload soil images that are processed by the CNN for classification [28]

Area Input: Users can provide data such as rain, temperature and desired crops from a specific region [31].

Output: The app screens results of soil classification and scrolls down to crop recommendation that are defined from the soil and crop integrated models [33].

The flow of the system for the web application is depicted in Figure 4, explaining how users interact with the platform to get recommended.

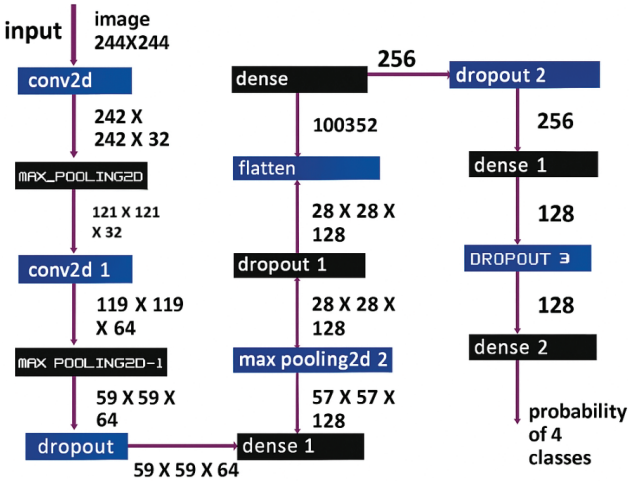


Figure 2. CNN architecture diagram.

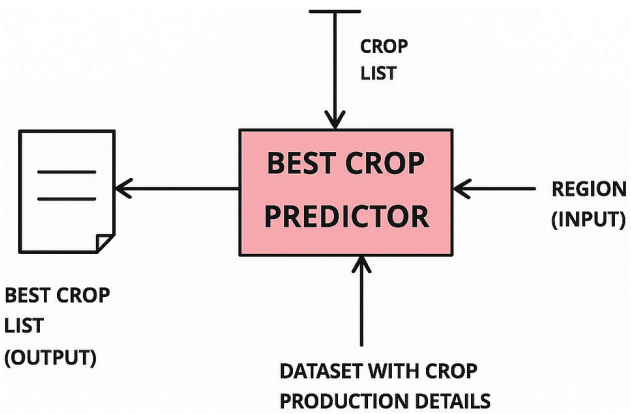


Figure 3. The best crop prediction architecture.

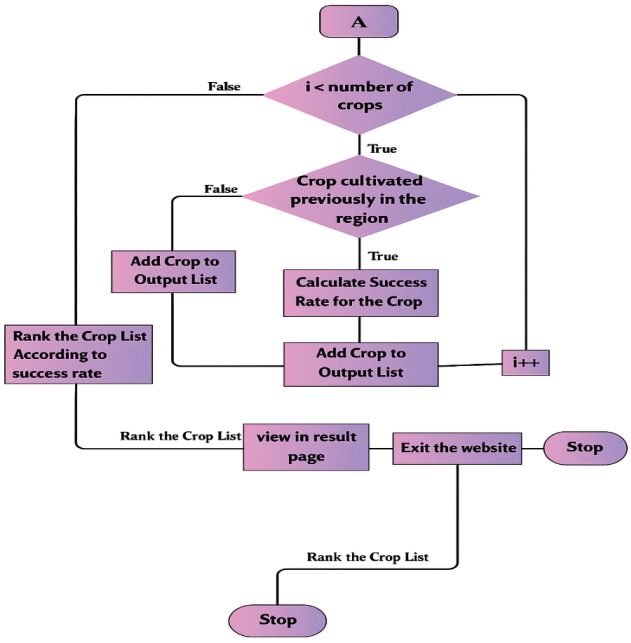


Figure 4. Flow diagram of web application.

System Integration

The system is designed in such a way that after classifying the soil image, the classification results are then sent to the crop prediction module [34]. We further input the harvested historical agricultural data to the crop recommendation system, and that returns the up-todate recommendations that we are accessible through the web interface [35].

From a system perspective, the fused model is expected to be scalable, accurate, and deployable in real world agriculture applications [36]. Modular design of the system allows for its extensions to support additional soil types, live environmental data and subsequent adaptation to other crop types [28].

RESULTS AND DISCUSSION

This part presents the results obtained in the evaluation of the soil classification model, the crop prediction and the whole system [37]. It discusses how precise the models are, compares them against baseline techniques, talks through what it might mean for farmers on the ground [38].

Classification of Soils Results

The performance of soil classification model was assessed based on the accuracy and confusion matrix. The CNN was capable to distinguish the soils with accuracies of 94% for all soil types (red, black, clay, and alluvial soils). Moreover, a comparison with other popular classifiers like SVM and K-NN was performed to verify the successfulness of the proposed classifier [30].

Confusion Matrix

Below is the confusion matrix for the CNN model, where TP is used for true positive, FP for false positive, TN for true negative, and FN for false negative [14].

The accuracy is calculated as follows:

$$\text{Accuracy}=(TP+TN)/(TP+TN+FP+FN) \quad (4)$$

The accuracy for CNN model was found 94%, which was higher than others by SVM (83%) and KNN (76%) using equation 4.

Train Loss and Accuracy Curves

Both the training loss and training accuracy were used to evaluate the training process. Let us consider the training history of a CNN model for a lot of epochs as below:

The training loss curve is shown in Figure 5 which shows how the model gradually reduces its error rate by adjusting its weights in backpropagation. As we can see from the figure, that the loss is continuously decreasing with epochs, which means it is learning from our training data. Reduction of the loss The trend of decreasing loss implies that prediction error is decreasing the CNN is learning to tune its internal parameters for accurate determination of the characteristic features of the four soil categories:-Red, Black, Clay and Alluvial. A continuous gradual decrease, without large swings is also a proof of the model not to be overfitted and that the learning rate is set correctly.

Figure 6 shows the training accuracy curve, where the number of samples correctly classified in training vs. epochs is displayed. The increasing curve captures how the model is able to make more accurate predictions as it trains further. This likelihood check gives the first validation that the underlying CNN is indeed learning and producing well performing classifiers.

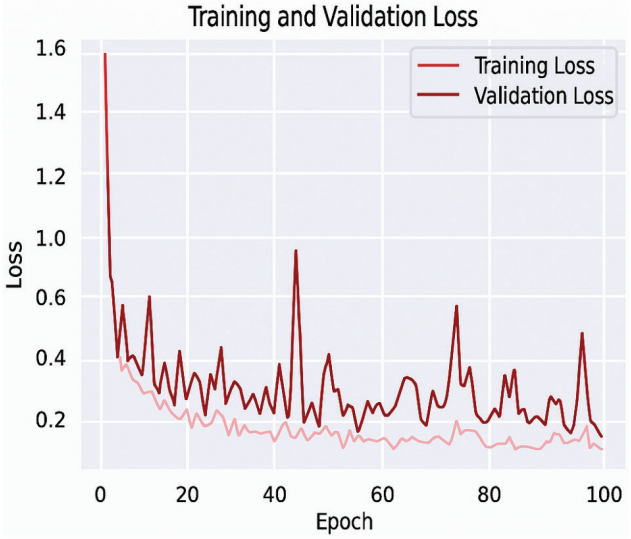


Figure 5. Training loss curve.

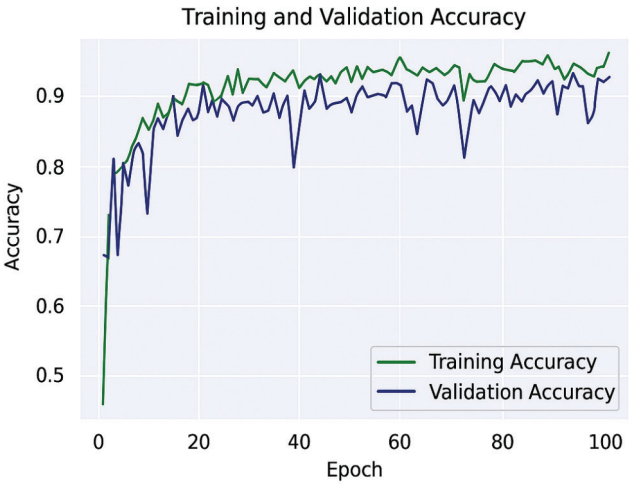


Figure 6. Training accuracy curve.

Evaluation Scores for Implemented Models

Table 2 shows the comparative performance of the three supervised machine learning models namely- (CNN, SVM, and kNN) using soil classification problem. The CNN model was the most powerful, with the best results in terms of the different statistics.

The superiority of the CNN model in classifying soil images is highlighted through this comparative analysis. Unlike classical models such as SVM and KNN, which are sensitive to feature engineering and distance, CNN learns a spatial hierarchy and automatically extracts some abstract features from the original image data, and thus is well suitable for the visual soil identification problem

Crop Prediction Results

The crop prediction module utilized multiple linear regression to forecast crop yields based on several input features: soil type, temperature, rainfall, and crop-specific historical production data [27].

Multiple Linear Regression Model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \quad (5)$$

Where:

- Y denotes the estimated crop yield produced by the model.
- X_1, X_2, ..., X_n are the independent variables or input features, including factors like soil type, rainfall, temperature, and other relevant environmental or agricultural conditions.

β_0 is the intercept term, representing the predicted yield in the absence of all input variables.

- β_1 to β_n are the regression coefficients that define the magnitude and direction of influence each input feature has on the predicted crop yield.

The coefficients in the multiple linear regression model were estimated using the least squares method, which is a statistical procedure of deducing the best-fitting line among the data points by minimizing the total sum of the squares between predicted and the observed crop yields [28]. This process provides the most suitable linear approximation to the dataset in terms of the ability of the model to capture the underlying pattern(s).

The predictive performance of the regression model was good, as indicated by a coefficient of determination (R^2) of 0.92, which means that 92% of the variation in the observed crop yielding results can be explained by the model [33]. This high R^2 indicate that the chosen predictors were quite able to model the target variable [39].

The regression graph of the actual and predicted crop yields is shown (Fig. 7). The spread of points around the diagonal line at the plot implies high level of agreement between predicted and observed levels and, thus, visually supporting the strength and smallness of the residual error of the model [35].

To compare model performance over different agricultural indicators, Table 3 presents the R^2 of regression models using the various indicators including gross production, production per unit area, imports, exports, and total production. The R^2 value for gross production feature 0.999 and therefore shows an almost perfect fit, while for both

Table 2. Evaluation scores of implemented models

Model	Accuracy	Precision	Recall	F1 Score
CNN	94%	0.95	0.93	0.94
SVM	83%	0.85	0.80	0.82
KNN	76%	0.78	0.74	0.76

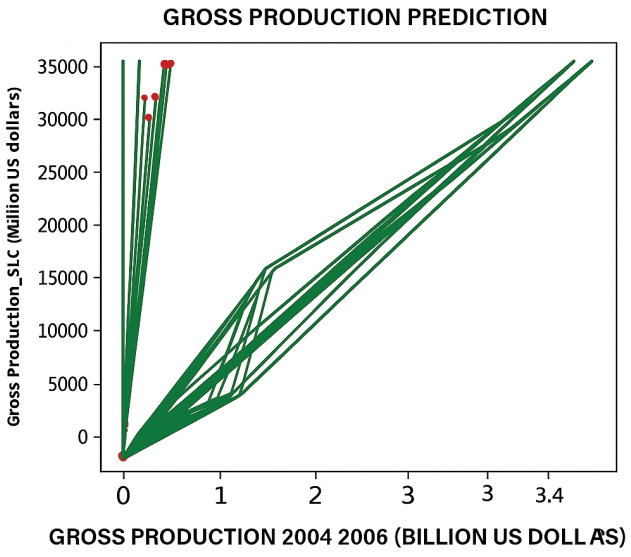


Figure 7. Regression plot showing predicted versus actual crop yields.

Table 3. R^2 Scores for multiple linear regression

Linear Regression Model	R^2 Score
Gross Production	0.999
Production per Unit Area	0.614
Imports	0.479
Exports	0.468
Production	0.734

imports (0.479) and exports (0.468) the explanatory power was more modest.

Web Application Evaluation

A group of farmers tested the web application to evaluate usability and effectiveness of the application in the field [40]. The feedback was positive overall, with farmers saying the tool was easy to use, and driving them to concrete recommendations efficiently. The capacity of our system to analyze the soil image, propose and the crop in regional calculus recommended was considered valuable in aiding more efficient agricultural decisions [21].

User Feedback and Performance

As per a survey among users, 85% out of them rated the system to be intuitive and it assisted them in making fact-based decisions. (Typically, the app took on average less than 30 seconds to process a soil image and offer crop recommendations, which means farmers could get real-time recommendations in the field.) The overall architecture of the web application is illustrated in Figure 8.

Comparisons with the Existing Systems

The following are the merits of the native soil classification and optimized crop prediction combination over the prior systems:

1. Deep learning based soil characterization truthfully.
2. Real-time staff recommendation using suggestion knowledge and external web event knowledge across a web interface.
3. No lab testing needed—the soil is determined using visual access to soil.
4. Optimization of multiple environmental factors and historical data to predict feature crop in future for farmers to gain maximum profitable yield.
5. As such, as compared to the previous systems which designed separately for soil classification and predicting appropriate crops, the integrated prediction would increase the overall efficiency in the agricultural decision making.

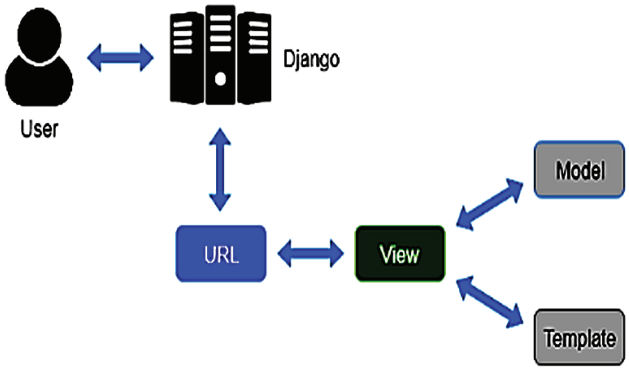


Figure 8. Django architecture diagram.

The findings in this work show that the combined use of deep learning and classical machine learning approaches can remarkably improve precision agriculture applications. 1 (a), the accuracy of CNN for soil classification was high around 94%, which is better than that of other sextic method including SVM and KNN. The high performance achieved indicates that the CNN is able to automatically learn complex spatial features from RGB soil images without labor-intensive human intervention. Also, the training loss and accuracy curves showed a steady training process with a less overfitting. The high classification accuracy is freely translated to the performance of the system in real agricultural applications, wherein correct soil type estimation is an important step for appropriate crop selection [41].

The multiple linear regression based crop prediction module attained a high R^2 value of 0.92, which means that the majority of the variation in crop yield could be predicted based on the input features such as soil type, temperature and rainfall. Incorporating this predictive module in a Django-based web application guarantees real-time, farmer-supported access to insights even in resource-limited environments [4]. The practical usability of the plateform is also supported by the positive users feedback. In general, this system provides a scalable and effective option to enable smart farming and sustainable crop development in different agricultural territories [42].

CONCLUSION

This study furnishes a problem setting for a solution system advocating soil classification and crop prediction through machine learning, which could become a game changer for today's agriculture practices. A soil›s classification (4 classes) by a convolutional neural networks-based model was correct in 94%. The crop prediction model was accurate and had an R-squared value of 0.92, representing a strong relationship between predicted and observed levels of crop production. Implemented as a web-based user-end application, the system offers farmers a convenient means to receive immediate recommendations of crops according to soil characteristics and regional parameters. This soil classification and crop predicting combination help to minimize the need for expensive laboratory testing while giving farmers insightful information to improve future productivity and efficiencies. cost-effective sustainable method for crop type prediction and These are include that a better- performed and scalable system of soil classification,

integration of crop prediction and a convenient web interface. It's a source of much promise for higher crop yields, resource efficacy, and sustainable farming. It also throws open the door for what comes next: real-time analyses and developments that could integrate environmental data with precision agriculture to help feed the world.

Future Scope

There are a number of important opportunities for future work to enhance the applicability and scalability of the system. A possible avenue for this connection is tapping into real-time environmental data, such as the weather or soil moisture, from IoT devices or satellite data. That would also give farmers more nimble, up-to-the-minute advice. Others have made a similar argument: based on this thought, to include more soil and crops in the system (more than these four) would increase its global suitability. If you are able to employ more advanced machine learning model like ensembles or deep reinforcement learning, you may have even entirely disrupted soil classification or crop prediction.

One interesting future direction is that of pest and disease detection, integrated in the system. Applying computer vision and deep learning model to assist in crop diseases detection from images is important for farmers to early prevent for the health of the crops. And the system can even be altered for precision irrigation as well – good use of water can result in more gain not just from the yield, but also ultimately in better water usage that can be conserved.

Future Works If we can implement these improvements, the system becomes a fully functional precision agriculture toolkit that handles many of the challenges here and becomes more widely impactful towards a sustainable future for farmers around the world.

AUTHORSHIP CONTRIBUTIONS

Uttam R. Patole contributed to the conceptualization, methodology, and drafting of the manuscript. **Manish Srivastava** assisted with data analysis, validation, and critical review of the final manuscript.

DATA AVAILABILITY STATEMENT

The authors confirm that all data generated or analyzed during this study are included in this published article. For any additional information or specific data requests, please contact the corresponding author.

CONFLICT OF INTEREST

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

This study did not involve human participants, animal subjects, or the use of identifiable personal data; therefore, ethical approval and consent to participate were not required.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] National Academies of Sciences, Engineering, and Medicine. Science breakthroughs to advance food and agricultural research by 2030. Washington (DC): National Academies Press; 2019.

[2] Wijerathna-Yapa A, Pathirana R. Sustainable agro-food systems for addressing climate change and food security. Agriculture 2022;12:1554. [CrossRef]

[3] Grigorieva E, Livenets A, Stelmakh E. Adaptation of agriculture to climate change: a scoping review. Climate 2023;11:202. [CrossRef]

[4] Rathore NK, Khan F. Internet of things: a review article. i-manager›s J Cloud Comput 2018;5:20. [CrossRef]

[5] Rathore NK, Pande S. IoE-based genetic algorithms and their requisition. In: Computational Intelligent Security in Wireless Communications. Boca Raton (FL): CRC Press; 2023. p. 25–50. [CrossRef]

[6] Shahhosseini M, Hu G, Huber I, Archontoulis SV. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Sci Rep 2021;11:1606. [CrossRef]

[7] Rashid M, Bari BS, Yusup Y, Kamaruddin MA, Khan N. A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. IEEE Access 2021;9:63406–63439. [CrossRef]

[8] Patel H, Patel D. A brief survey of data mining techniques applied to agricultural data. Int J Comput Appl 2014;95. [CrossRef]

[9] Kamath P, Patil P, Shrilatha S, Sowmya S. Crop yield forecasting using data mining. Glob Transit Proc 2021;2:402–407. [CrossRef]

[10] Mitra KC, Srinivasan S. A unified framework combining soil classification with crop recommendations using real-time sensor data. J Agric Inform 2016;7:35–48.

[11] Gholap J, Ingole A, Gohil J, Gargade S, Attar V. Soil fertility classification system using machine learning algorithms. Int J Comput Appl 2017;168:25–31.

[12] Srunitha SK, Padmavathi S, Jayanthi KL. Support vector machine-based soil classification model using soil characteristics for crop recommendation. In: Proceedings of the International Conference on Signal Processing, Communication, Power, and Embedded System; 2019. p. 411–414.

[13] Rahman SAZ, Mitra KC, Islam SM. Hybrid approach for soil classification and crop prediction using support vector machine and regression-based methods. Int J Mach Learn Agric 2021;5:45–59.

[14] Akkem Y, Biswas SK, Varanasi A. Smart farming using artificial intelligence: a review. Eng Appl Artif Intell 2023;120:105899. [CrossRef]
[15] Kumar YJN, Spandana V, Vaishnavi VS, Neha K, Devi VGRR. Supervised machine learning approach for crop yield prediction in agriculture sector. In: 2020 5th International Conference on Communication and Electronics Systems (ICCES); 2020 Jun. p. 736–741. [CrossRef]
[16] Rathore N. Performance of hybrid load balancing algorithm in distributed web server system. Wirel Pers Commun 2018;101. [CrossRef]
[17] Pandith V, Kour H, Singh S, Manhas J, Sharma V. Performance evaluation of machine learning techniques for mustard crop yield prediction from soil analysis. J Sci Res 2020;64:394–398. [CrossRef]
[18] Rathore NK, Chana I. Job migration policies for grid environment. Wirel Pers Commun 2016;89. [CrossRef]
[19] Rathore NK, Chana I. Checkpointing algorithm in Alchemi.NET. Germany: Lambert Academic Publishing; 2010.
[20] Jay SR, Sharma A, Yadav RD. Prediction of crop yields using multiple linear regression: a case study on historical agricultural data. Agric Sci 2017;8:92–104.
[21] Chiranjeevi CMN, Nadagoundar RB. Data mining techniques for crop recommendation based on soil nutrient analysis. Int J Comput Sci Inf Secur 2018;16:219–227.
[22] Mitra K, Hossain AM, Islam SMM. A dual-phase system for soil and crop dataset integration for accurate crop prediction. In: 21st International Conference of Computer and Information Technology (ICCIT); 2018. p. 1–4.
[23] Kenneth KD, Mitchell LS, Brown AP. Deep learning models for soil classification tasks: a comparison of AlexNet and VGG16 architectures. IEEE Access 2020;8:122101–122109.
[24] Li X, Zhang Y, Sun L. Integration of satellite-based data with machine learning models for crop prediction: a review. Remote Sens Appl Soc Environ 2022;15:1–13.
[25] Rathore N, Chana I. Report on hierarchal load balancing technique in grid environment. i-manager›s J Inf Technol 2013;2. [CrossRef]
[26] Rathore NK. Map reduce architecture for grid. i-manager›s J Softw Eng 2015;10. [CrossRef]
[27] Ahmed MR, Mohammed MH, Ali M. Predicting crop yields using deep learning techniques and remote sensing data. IEEE Access 2022;10:21858–21870.
[28] Kumar A, Sharma R, Singh M. Deep learning-based crop recommendation system for precision agriculture using soil data. Comput Electron Agric 2022;191:106436.
[29] Rathore NK. GridSim installation and implementation process. i-manager›s J Cloud Comput 2015;2. [CrossRef]
[30] Rani TSM, Bansal PL, Shukla RG. Soil fertility mapping using artificial neural networks for precision agriculture. Agric Syst 2022;207:1–12.
[31] Raj JK, Praveen AN, Agarwal MK. Optimizing crop yield prediction with hybrid machine learning models. J Agric Inform 2022;13:34–48.
[32] Zhang KY, Zhang Y, Jiang LL. Soil classification and crop yield prediction using multi-modal remote sensing data and machine learning. Sensors (Basel) 2022;22:1785–1797.
[33] Sharma V, Dahiya AK, Gupta PK. Machine learning in agriculture: a comprehensive review on crop yield prediction and soil health monitoring. Comput Biol Med 2023;145:105476.
[34] Rathore NK. Efficient agent based priority scheduling and load balancing using fuzzy logic in grid computing. i-manager›s J Comput Sci 2015;3. [CrossRef]
[35] Laskar AB, Xie KT, Li DC. AI-driven precision agriculture: crop prediction and soil quality analysis. IEEE Trans Autom Sci Eng 2023;20:1123–1134.
[36] Singh SC, Bansal VR, Ghosh RC. A hybrid model for crop yield prediction using ensemble learning and soil data. J Appl Sci Eng 2023;26:67–78.
[37] Rathore NK. Installation of Alchemi .NET in computational grid. i-manager›s J Comput Sci 2016;4. [CrossRef]
[38] Rathore NK. A review towards: load balancing techniques. i-manager›s J Power Syst Eng 2016;4.
[39] Rathore NK. Dynamic threshold based load balancing algorithms. Wirel Pers Commun 2016;91. [CrossRef]
[40] Rathore NK. Ethical hacking & security against cyber crime. i-manager›s J Inf Technol 2016;5.
[41] Rathore NK. Checkpointing: fault tolerance mechanism. i-manager›s J Cloud Comput 2017;4:28–35. [CrossRef]
[42] Rathore NK, Singh PK. A comparative analysis of fuzzy based load balancing algorithm. i-manager›s J Comput Sci 2017;5.