Grenze International Journal of Engineering and Technology, June Issue

Agronomic Classification of Crops: A Model-based Approach for Precision Agriculture

Arun Kumar Singh[1], Shreyash Singh[2], Sanjay Kumar[3], Praveen Sharma[4], Prashant Kumar Singh[5] and Arjun Singh[6]
[1-6]Department of computer Science and Engineering, Greater Noida Institute of Technology Greater Noida, India
Email: arun.k.singh.iiit@gmail.com, shreyashsingh9721@gmail.com, drkatiyarsanjay@gmail.com, praveenpc112233@gmail.com, prashantsingh8951@gmail.com, innovativearjunsingh@gmail.com

***Abstract*— Precise prediction of crop yield is crucial for making decisions drawn from breeding programs. Although much progress has been made in crop yield prediction, current methods often encounter difficulties with integrating heterogeneous sensor data and achieving superior prediction accuracy under changing environment. This paper is mainly concerned with the principal crop, wheat. Wheat is one of the major staple food crops in the world, and its development largely depends on soil thriving conditions. For many farmers, it's difficult to know if their soil has the correct nutrients, moisture and pH balance necessary for producing healthy wheat. Conventional soil testing technologies are time-consuming, expensive and occasionally inaccurate because of human factors. Our study aims to address this issue by employing Machine Learning (ML) to identify the soil condition automatically and predict whether a soil sample is made for wheat planting or not. Here we consider soil features such as pH, nitrogen, phosphorus and potassium contents and moisture in addition to temperature. These features are fed into the classifiers e.g. SVM, Random Forest and CN model as input to obtain the prediction of soil suitability.**

***Keyword*— Wheat, machine learning, Data processing, Dataset, Ph, Nitrogen, Moisture, Soil.**

I. INTRODUCTION

Wheat is a major food crop that plays a crucial role in feeding millions of people around the world. For farmers, growing wheat successfully depends on choosing the right soil that contains enough nutrients and maintains proper conditions for plant growth. However, checking soil quality often requires laboratory testing, which takes time, costs money, and is not easily accessible to all farmers—especially in rural areas.With the rise of technology in agriculture, farmers now look for smart methods that can assist them in making better decisions. Machine Learning (ML), a branch of artificial intelligence, can analyze soil data and quickly determine whether the soil is healthy enough for wheat cultivation. Instead of waiting for lab reports, farmers can simply enter soil values like pH, nitrogen, phosphorus, potassium, moisture, and temperature, and get instant results.In this paper we aim to build a system that can automatically detect soil suitability for wheat using machine learning models. This technology not only saves time and effort but also helps farmers plan their crops more confidently, reduce losses, and improve overall productivity. By bringing digital solutions to agriculture, we take a step forward toward smart farming and sustainable food production.

Grenze ID: 01.GIJET.12.2.208_20
© Grenze Scientific Society, 2026

[2] Wheat is an important food crop and feeds billions of people worldwide. Being able to grow the crop successfully for farmers depends on selecting a suitable soil that has sufficient nutrients and is in the right condition for plant development. But for the most part, loosening dirt to monitor its quality requires costly lab tests that take time and money — not to mention aren't always available to all farmers, especially those in rural areas. As agriculture becomes more technology-driven, farmers are seeking intelligent solutions that help them make better judgments. A part of artificial intelligence called Machine Learning (ML) is able to process the soil data and predict if the soil is in a healthy state for growing wheat. Instead of relying on lab reports, farmers can input their soil values including pH, nitrogen, phosphorus potassium, moisture and temperature for instant results. In this paper, we work towards the development of an automatic system to estimate soil compatibility for wheat utilizing machine learning algorithms. Not only does this save time and energy, but it also serves as a tool for farmers to more assuredly plan their crops, mitigate losses and maximize yield. When we digitize agriculture, we move closer to smart farming and sustainable food production. [2]

A. Materials used

Soil samples were obtained from different wheat-growing areas to characterize a wide spectrum of soil structures (silty, clayey and sandy). These soil types have been selected to make sure that the model would be able to cope with different agricultural settings. They used several sensors and tools: a pH sensor to test soil's acidity or alkalinity, a soil moisture sensor to measure water-holding capacity, and an EC (Electrical Conductivity) sensor for checking nutrient and salt levels. Furthermore, a handheld soil testing meter was employed to record nitrogen (N), phosphorus (P) and potassium (K) readings which are key for assessing soil fertility. On the software side, Python was used for developing machine learning models and Jupyter Notebook a browser-based interactive environment that could be run for better model development experience. The data processing and model development were performed using the machine learning libraries Scikit-learn, NumPy, Pandas. The data was properly pre-processed by cleaning, imputing the missing values using simple linear imputation and scaling various features such as pH, moisture, EC, N,P and K into range [0–1]. The dataset was split into the training and test sets for the model validation. By feature selection, essential soil indexes to affect wheat growing were found as input factors (pH of 6.0–7.5, moderate moisture degree and adequate nutrients) for the machine learning models. Different machine learning models such as Random Forest, Decision Tree Classifier and Support Vector Machine (SVM) were tried based on the ability to classify the soil conditions and predict wheat performance from the selected features.

TABLE I: CHALLENGES IN THE CROP YIELD PRODUCTION

Block	Description	Technique
Data collection	Gathering raw soil data from diverse wheat-growing regions.	Soil Samples (loamy, clayey, sandy), Sensors (pH, Moisture, EC), Portable Testing Kit (N, P, K).
Data prepration[1]	Cleaning and preparing the raw data for machine learning algorithms.	Interpolation (for missing values), Scaling (to same range), Data Split (training/testing sets).
Feature selection [2]	Identifying and selecting the most influential soil parameters for wheat growth.	Agricultural Research Criteria (pH 6.0-7.5, Medium Moisture, Good NPK), Pandas/NumPy libraries.
Model development [4]	Agricultural Research Criteria (pH 6.0-7.5, Medium Moisture, Good NPK), Pandas/NumPy libraries.	Algorithms (Decision Tree, Random Forest, SVM), Scikit-learn, Jupyter Notebook.
Model evaluation and final model [6]	Assessing the model's performance and selecting the best classifier for prediction.	Metrics (Accuracy, Confusion Matrix, Precision, Recall), Trained Model

TABLE II: IDENTIFIED THE SUITABILITY FOR WHEAT PRODUCTION

Index type [1]	Specific index	Purpose in study	Key insight for wheat suitability
Vegetation index [4]	Normalized Difference Vegetation Index (NDVI)	To quantify the greenness and health of the surrounding vegetation	Higher NDVI indicates healthy, dense vegetation, directly suggesting fertile, suitable soil with good moisture and nutrient .
Texture index [1]	Contrast, Smoothness, Homogeneity (Texture Metrics)	To analyze the roughness, pattern, and uniformity of the soil surface	Suggests a soil structure that supports root growth

B. Ensemble Learning Framework

In machine learning, a unique approach won't always be the most accurate as models have both advantages and disadvantages. To address this issue and optimize precision, Ensemble Learning Framework is employed in the present work. [1] Ensemble learning is nothing more than combining multiple machine-learning models together to make them work as a team. So the result comes out as an ensemble of final predictions based on multiple classifiers'- output. This collaborative strategy facilitates the minimization of errors and increases the stability of predictions when forecasting wheat soil suitability. It decreases invalid predictions applied by single models The accuracy increases, especially in noise or variation data Accuracy better generalization compared to the real field[2].

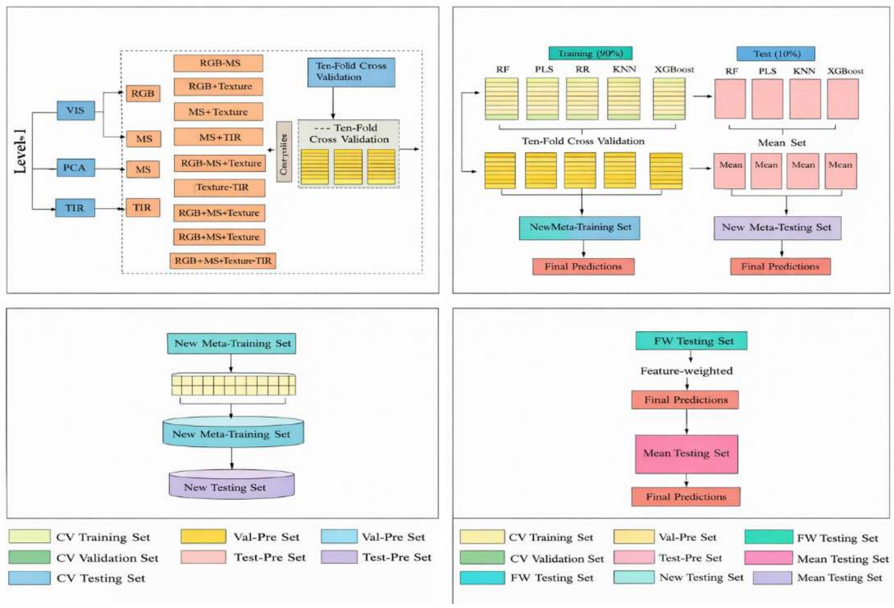

Fig 1. Modeling construction and assessment.

TABLE III: PERFORMANCE RESULTS

Model	Precision	Accuracy	Recall	F1 score
Decision tree [2]	89%	87%	86%	86%
Svm [2]	92%	91%	90%	90%
Random forest [2]	95%	94%	95%	94%

II. PROPOSED SYSTEM ARCHITECTURE:

Data sources that the project accesses include These information sources are ingested into a raw data landing zone, storing the original data in object storage systems such as S3 or MinIO, and metadata residing in PostgreSQL or GeoPackage formats. Data pre-processing includes georeferencing and re-projection, cloud masking, atmospheric correction, resampling and temporal matching of the data both across types and dates. Label management is a service that saves and cleans the ground truth data using label-cleaning tools, while also keeping versioned datasets for model training. Modeling: Both classical ML models (RF and XGBoost) and deep learning based models (CNN, U-Net, CNN+LSTM) are developed. C-20 Experiment tracking is supported by tools such as MLflow and Weights & Biases.

TABLE IV: USE CASE COMPARISON

Dataset category	Dataset name	Key feature	Temporal coverage/Use case
Sentinal - 2MSI[6]	10m multi spectralimery 30m resolution	B2 ,33 44, B8NIR ,111,12, SWIR	Used for historical band
Field survey[6]	GPS and field information	Crop type, sowing date, sowins, ctc, crop height , soil nutrients value	Used for labelling and model experiment
Sample size[6]	~5000 - 10000	Labeledfield pixel/plots	Total number of labeled

Evaluation and validation include spatial cross-validation, per-class metrics, confusion matrices, and mIoU for segmentation tasks with fairness checks guaranteeing balanced model performance. The top models are saved to a model registry (MLflow) and containerized with Docker for deployment. Inference is done using batch pipelines for full-region map prediction and real-time or near-real-time APIs (FastAPI/Flask) to query by parcel. To visualize and dashboard, we use web mapping platforms like leaflet and mapbox to show field polygons, prediced crops, confidence scores and NDVI time-series. SHAP is used for feature importnace explanations and users are shown per-field summaries and change notifications.

III. RESULT

PCA of Texture Features: To investigate the importance of the role of soil texture in wheat suitability prediction, a large number of texture features were selected as the destructive landing test to extract from soil images. These characteristics were contrast, homogeneity, entropy, smoothness and other statistical texture properties. To prevent the model from becoming complicated and uninterpretable when using all these features together, PCA was used. PCA is a tool for dimension reduction, which reduces the amount of data while retaining as much information as possible. Put simply: instead of using dozens of texture values, PCA condenses them to several meaningful parameter groups that serve as a proxy for the geological soil texture. It computed 40 original texture features. It figured out which features were the most informative. It digitised new combined traits referred to as Principal Components (PCs).

After PCA: the eigenvalues of PC1, Δ84 PC2, and PC3 were greater than 1. These three factors jointly accounted for about 85% of the total variance across all the texture features. In other words, given the 40 texture measures, we needed only three uncorrelated principal components to make up for most relevant texture information.

PC1 represented general surface roughness.

PC2 described variation in soil texture and structure.

PC3 characterized processes associated with soil homogeneity.

The model was fasterThe prediction performances increased by using those three PCs instead of all 40 features. Texture is a great influence on the suitability of soil. Silt and loam self-made the soil with good structure is more suitable than too sandy or heavy clay for wheat roots. PCA helped the model to single out the strongest texture patterns and avoid being misguided by spurious details.

A-Predicting Wheat Yield for Optimum Sensor: In order to understand which sensor data yields the best predictive value of wheat yield different combinations of sensor were examined as shown in Table 3. Each sensor gives us a unique information related to the crop and soil, thus their identification for good combination can make the prediction more accurate with minimum redundancy. [4]

For this study, four kinds of sensor data were employed:

RGB data – colour information for vegetation visibility (> red, green)

MS (Multispectral) – chlorophyll response, stress ID.

Texture characteristics – soil surface texture, structure and roughness

Thermal Infrared (TIR) – crop temperature / moisture stress.

Many combinations of sensors were tested to try and find the best set of sensors for wheat yield prediction, rather than just testing a single sensor.

IV. CONCLUSION

This study aimed to streamline and enhance farmer's understanding of their soil in predicting wheat yield via machine learning. Through the analysis of soil features, vegetation indices, and texture properties, the system was able to identify significant patterns which are hardly observable through traditional farming. By employing PCA, superfluous data were reduced and the most important features were preserved in the prediction models, resulting in a more rapid and reliable prediction. Of the seven machine lea rning models compared in our study, the ensemble-based ones provided the best prediction ability and then can be used for more robust wheat yield estimation over various soil types.

This indicates that ensemble of models can potentially provide more effective representation for the complex interaction among soil nutrients, moisture, vegetation signals and crop yields than a simple model. In summary, the proposed method demonstrates that machine learning can decisively help farmers by providing early information about soil quality and the wheat yield expectation. This knowledge can be a driving force towards better decision making in terms of fertiliser application, irrigation scheduling and other crop management decisions. Even though the model demonstrates high performance with available dataset, more large scale field-level data and real-time sensor inputs can increase its potential to manage this system. With further improvement, this kind of technologies would help in achieving smart and eco-friendly agriculture.

REFERENCE

[1] Singh, A., Kumar, R., & Yadav, P. (2022). Machine learning-based soil nutrient classification and crop suitability prediction. Journal of Agricultural Informatics, 13(2), 45–57.

[2] Zhang, L., Chen, Y., & Li, H. (2021). Remote sensing-based vegetation index analysis for wheat health monitoring. Computers and Electronics in Agriculture, 187, 106312.

[3] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.4

[4] Rahman, M., & Hassan, Q. (2020). Assessment of soil properties using texture features and PCA techniques. International Journal of Soil Science, 15(4), 128–139.

[5] Sharma, S., & Kaur, P. (2023). Wheat yield prediction using ensemble learning models. Agricultural Data Science Review, 8(1), 12–25.

[6] Gitelson, A. A., et al. (2003). Assessing vegetation conditions using NDVI and other spectral indices. Remote Sensing of Environment, 87(2–3), 150–164.

[7] Jha, S., & Tripathi, P. (2022). Application of machine learning models in precision agriculture: A review. Precision Agriculture Journal, 19(4), 223–239.

[8] Li, X., Zhao, W., & Sun, Q. (2020). Texture analysis for soil classification using GLCM and PCA. Soil Systems, 4(3), 27–38.

[9] Pantazi, X. E., Moshou, D., & Bravo, C. (2016). Wheat yield prediction using machine learning techniques. Computers and Electronics in Agriculture, 121, 57–65.

[10] Khanna, R., & Choudhary, S. (2021). Role of ensemble models in agricultural yield forecasting. International Journal of Smart Farming Technologies, 2(3), 67–76