RESEARCH PAPER

An efficient Machine Learning based Soil Nutrient Monitoring and Crop Recommendation System

[1]Phanikanth Chintamaneni, [2]Subrahmanyam Kodukula

[1]Research scholar, Computer Science and Engineering, Koneru Lakshmiah Education Foundation, Vaddeswaram,
Phanikanth.ch@gmail.com
[2]Professor , Computer Science and Engineering , Koneru Lakshmiah Education Foundation, Vaddeswaram,
smkodukula@kluniversity.in

Abstract

With the growing amount of IoT farming data, it has become more challenging to select suitable data for IoT farm applications. This study proposes a sophisticated crop recommendation system based on the combination of multiple data sources such as Crop_Recommendation.csv, Soil.csv, and Crop_names.csv that enable to predict one right crop. The system uses geographical coordinates (latitude ϕ and longitude λ) to generate temperature and humidity as environmental variables (expected outcomes of the regression models), which are the basic inputs for the analysis of crop suitability. Using a classification model , with the features such as soil type and nitrogen requirements, the model which was proposed predicts the best crop class. The system tunes the hyperparameters to get the best predictions, and for each scenario it outputs the top five crops that should grow in the best conditions. Also, it computes the Growth Degree Days (GDD) and the nutrient (nitrogen, phosphorus, potassium) needs for all of the suggested crops, allowing the farmer to make an all-encompassing decision. This machine learning and geo-based technique contributes to improved agricultural choices by enabling accurate, data-informed crop predictions that are environment and soil specific, thus is a prospective solution to current related research problems.

Keywords: Crop Recommendation System, Predictive Algorithms, Soil Nutrient Monitoring, Geographic Coordinates, Environmental Modeling, Soil Classification.

How to cite this article: Phanikanth Chintamaneni, Subrahmanyam Kodukula | An efficient Machine Learning based Soil Nutrient Monitoring and Crop Recommendation System | Int J Drug Deliv Technol. 2026;16(4s): 210-227, DOI: 10.25258/ijddt.16.210-227

Source of support: Nil.

Conflict of interest: None

1.Introduction

The need for smart agri-culture systems is becoming very important since the population of human beings is ever increasing and agriculture is still the backup of world economy. This industry has seen many evolutions in the last several decades to serve a growing population that has more than doubled in the last half century. It is currently projected that the global population will reach 9.8 billion by 2050. What's more, a trend towards more urban living and fewer working people per retiree is anticipated. Consequently, needs to improve the agricultural productivity in sustainable way with less dependence on agricultural labour have been increasingly emphasized. Technological involvement in agriculture dates back over 100 years, when the first tractor was introduced in 1913. Since then, technology has advanced dramatically, with many of these technologies now available on the market, greatly increasing productivity and reducing the amount of human labour necessary. Yet these advances may not be enough to satisfy the world's future demand. In the wake of this, scientists have concentrated on enhancing the efficiency of output since the 1990s, bringing about the emergence of "precision agriculture. This strategy is based on the application of farm management practices that enhance production and resource use efficiency at the field level through monitoring, evaluation, and timely intervention of intra- field crop variability [1-5]. Also, established technologies such as remote sensing, Internet of Things (IoT), and robotic platforms have been applied to agriculture, leading the development of the new concept of "smart farmingSupervised machine learning methods such as Naive Bayes, K-nearest neighbors (KNN), support vector machines (SVM), and decision trees are applied to predict the soil fertility using numerous chemical attributes. The attributes were tested for pH, electrical conductivity, organic carbon, nitrogen (), phosphorus (), potassium (), iron (), and zinc (). From 43% accuracy for Naive Bayes to 60% accuracy for SVM, the model accuracies are different,

---

Author for Correspondence:

that shows different power of these methods to a set of analysed characteristics.A sustainable methodology for organic soil assessment is proposed by combining machine learning with a national spectral library. Herein, the random forest regression model is used to model a full feature soil dataset, which consists of soil properties such as soil organic matter ().The model performance is evaluated in terms of RMSE and metrics.Further predictions of soil fertility are to be expected for both simple regression and multiple regressions techniques, using data from portable X-ray fluorescence spectrometry (pXRF) measurements. These methods allow examination of soil characteristics including pH, percentage of clay, silt, and sand, contents of calcium (), magnesium (), potassium (), and aluminum (). The performance of these linear regression models are tested against and RMSE to guarantee a robust estimation of soil fertility[6-9]s.A multi-dimensional investigation based on atmospheric condition, nutrients, and soil properties considers the appropriateness of the cropland for agriculture. This integrated classification system leads not only to a basic understanding of ecosystems and crop behavior but also to a realistic planning and management tool for agriculture [10-14].Supervised machine learning (ML) classifiers, varying from K nearest neighbors (KNN), Bagged Trees to Gaussian Kernel-based support vector machines (SVM), are used for soil series prediction. As far as these methods are concerned, SVM provides the best classification results, which indicates the suitability of SVM to the classification of soils among the methods considered here.In order to promote agriculture and empower farmers to grow better quality products, one web based service platform is designed. Different services can be provided in relation to soil classification, and a particular focus area has been precise data capturing and methodical analysis to provide confidence in soil evaluation. In this review, we consider conceptual approaches for soil quality and soil performance, emphasizing those parameters with a bearing on agriculture.

This new link between soil fertility and conditions for plant growth and soil improvement was situated with soil quality and ecosystem services as the main factors. Automated prediction of the soil fertility is done using techniques like neural networks, deep learning (DL), support vector machine (SVM), random forest (RF), and Bayesian models. These procedures treat soil attributes such as phosphorus pentoxide (), iron (), manganese (), zinc (), and nitrous oxide (). Performance of the model is verified with metrics such as coefficient of determination () and different regressor algorithms, which guarantees the prediction are accurate and reliable. ELM, multiple linear regression (MLR), and random forests (RF) are used in a novel technique to predict coffee yield by soil parameters[15-17]. The study involved the soil properties of the soil organic matter (), the available potassium (), and pH value, with the model performance evaluated by root mean square error (RMSE) and mean absolute error (MAE). Methods of supervised learning have also been deployed to model the spatial variation of topsoil carbon (), nitrogen (), and available phosphorus (). Use of the generalized additive model (GAM), random forests (RF) and SVM (Support Vector Machine) to identify fertilizer application schemes that ensure best nutrient allocation for crop development. Village-level characteristics of soil nutrients types are used to recommend fertilizer rates and the choice of crops. The model considers multiple soil types and conditions based on a comprehensive soil analysis to apply region-specific farming techniques.This review describes soil fertility related issues including the source of the soil fertility data sets, the performance such as RMSE, MAE and , the procedure of calibration and the way of collecting datasets. Section 5 will provide an analysis of soil enzyme activity estimation using various enzyme types, which is a guide to soil health and microbes.Soil enzyme activity estimation is classified into supervised machine learning methods considering different soil characteristics, microbial activity, and soil organic matter. The performance in terms of enzyme data sets generated from pXRF of enzyme rich soil samples analyzed through soil test labs to evaluate the precision and trustworthiness of the method is discussed.Factors like season, fertility and soil texture were taken into account during analysis of soil enzyme activity in Brazilian soils. Accuracy of prediction of enzymatic processes such as activity of glucosidase, acid phosphatase, alkaline phosphatase, urease and fluorescein diacetate (FDA) hydrolysis, with conditional random forest approaches is evaluated.Terrestrial soil enzymes: Extracellular enzymes within soil communities are predicted using MLR and RF models (47). Water content, electrical conductivity, total N (), total P () and SOC () were used to predict enzyme activities, including amylase and urease, with novel insights into soil and nutrient-related process. Intercropping was also found to improve soil extracellular enzyme activity by meta-analysis evidence. The variety of plants introduced by intercropping also has a notable impact on soil microbes, leading to a diverse group of bacteria and fungi. Soil microbial components and enzyme activities are investigated in an array of its commonland management such as agro cropland and vegetable garden. Soil microbial biomass C () and N (), respiration activities, and enzymes such as β-glucosidase and acid phosphatase are all significantly correlated with soil C () and N (), pointing to the importance of these enzymes in soil health and nutrient cycling. The impact of toxic metal exposure on soil enzyme activities is reviewed in relation to biochemical substances such as carbon (), nitrogen (), phosphorus (), and sulfur (). Enzyme activities are negatively affected by heavy metal contamination, the cycling of these vital components of soil ecosystem is therefore jeopardized.

Machine learning methods are used to predict biological changes in soils after the application of Phosphogypsum. Activities of enzymes, including urease and soil respiration, are monitored at different levels of Phosphogypsum, allowing for understanding of its effects on soil biochemistry.The influence of N fertilization on soil components, activities of soil enzymes, and microbial populations is also analyzed. Nitrogen enrichment is shown to strongly promote enzyme activities, and soil pH as well as organic carbon (OC) is suggested as the major drivers implicated in soil bacterial community alterations.Soil enzyme activities are crucial indicators of agricultural soil quality that enable processes such as soil organic matter decomposition and nutrient cycling. These processes are sensitive to alterations in soil management practices, thus they can be used as indicators for the value of soil in terms of health and fertility.Soil microorganisms are considered in relation to environmental factors and global change, in particular for their involvement in the degradation of chemicals. Changes in crop development and physiology under climate change also influence bioremediation efficiency, as these changes in the crop impact the ability of the microbial community to degrade contaminants.For the past 70 years, boosting agricultural productivity has been a matter of making the best use of soil minerals for ideal crop growth. Historically, agricultural production has been enhanced by transforming organic materials such as animal manure and plant residues back into the soil as nutrients. Yet it was the explosive growth and industrialization of the 19th century that necessitated that more than double agricultural production. Developing a strategy for maintaining soil fertility within the farming system is critical to increasing production potential to provide economic profits for the farmer. In addition to just managing the soil nutrient supply, such strategies also need to take into account the many factors affecting the soil nutrient supply and the uptake of nutrients by the crops. There are many interrelated factors betweenSugar Un APBioch #2 sugar metabolismfull References (at the end of this topic) - What is coenzyme B 12Interesting so far Comments (2)(1) Soil water content, pH, salinity, physical structure, and biotic stresses are among other factors which affect the nutritional status of a crop. As a result, a nutrient-balanced solution will frequently not be sufficient to maximize agricultural system function[18].

This complete knowledge enables more robust and resilient agricultural systems to be designed that can sustain varying conditions and retain productivity and ecological services.Plant growth and development are inherently dependent on the presence of essential nutrients in accessible amounts and in an appropriate balance. These macronutrients and micronutrients participate in many of the processes that are involved in the physiological and biochemical activities necessary for the plant life cycle. In this review, we discuss these nutrients (together with sodium [Na] as discussed in more detail elsewhere [19]) mainly with regard to their role in photosynthesis, energy transfer and plant health as a whole.

Machine learning (ML) algorithms which can be broadly categorized into three types—supervised learning, unsupervised learning, and reinforcement learning—are witnessing application in agriculture science. In particular, with respect to soil fertility prediction, supervised ML methods have been found effective as they can be used for both classification and regression analysis. Over the last few years, several algorithms, namely random forests, support vector machines (SVM), decision trees, linear regression, logistic regression, AdaBoost, XGBoost, Naive Bayes, k-nearest neighbors (k-NN), have been experimented with.

This study is concerned with the prediction of soil fertility status based on the important chemical indicators of soil including soil organic carbon (OC), pH, electrical conductivity (EC) by employing the supervised learning). The overall objective fish to identify the best prediction model for soil fertility given these soil properties. The strategy is to run a sequence of algorithm and see which supervised ML models performs better.

Application of Supervised Machine Learning Algorithms

Random Forest, Support Vector Machines (SVMs), Decision Trees, Linear Regression and Supervised ML in general are widely applied to represent the complex relationship between enzyme activities and soil properties. They account for complex, often nonlinear relationships between multiple soil parameters and enzyme activities and provide unique perspectives on soil health and crop yields.

Soil Health and Agricultural Productivity Implications

The use of supervised ML models plot vale enables soil scientists to foresee soil enzyme activity responses to a range of farming and environmental conditions. This predictive power could change nutrient cycling and farming productivity. By predicting enzyme activities at the right times and at the right scales, these models improve the understanding and control of key processes in the soil rom, which in turn support sustainable agriculture[26].

For example, enzymes associated with key biochemical cycles (β-glucosidase in the carbon cycle, urease in the nitrogen cycle, and dehydrogenase as a general indicator of soil quality) can be monitored and managed quite effectively through these models. Predicting the behavior of these enzymes facilitates an improved decision in managing the soil and hence the improved soil quality and crop yield.

Significance of Supervised Machine Learning Algorithms in Prediction of Soil Enzyme Activity

The activity of soil enzymes is critical to the biochemical reactions involved in the turnover of organic matter in the soil environment. They are also significantly influenced by soil organic matter (SOM), soil physical properties, microbial activities and soil biomass. According to their habitat in the soil, either the intracellular or the extracellular enzymes are.

Intracellular enzymes are active in living metabolizing cells, and also they are found in dormant or dead cells where they reside in the cytoplasm or they are bound to cell walls. These enzyme systems are essential to cellular metabolism and play a part in the global biochemical turnover within soils.

In contrast to intracellular enzymes, extracellular enzymes are frequently immobilized in the soil matrix via ionic interactions, covalent bonds, hydrogen bonds, and other forms of attachment to clay particles and humic substances. It is these enzymes that are responsible for degradation of soil organic matter (SOM), enabling nutrient cycling, energy flow, ecosystem stability, and agricultural productivity.

However, owing to destructive cultivation practices and overexploitation, soil enzyme activity may suffer degradation, which does not favor the availability of nutrients to plants. Previous studies have confirmed that enzymatic activity decreased with soil depth and these enzymes were more sensitive to environmental variations and soil management practices than many other soil indicators. Therefore, to the extent that they are well validated, standard assays for a range of soil enzyme activities are well established primary methodologies for measuring soil health (Bergstrom et al., 2000) .

Supervised machine learning and its role in enzyme activity prediction

Supervised machine learning (ML) algorithms have been demonstrated to be instrumental in predicting soil enzyme activities by capturing non-linear relationships between soil properties and enzymatic activities. These algorithms ranging from Random Forests, Support Vector Machines (SVMs), and Decision Trees, are capable of considering nonlinear interactions and are fully capable of taking into account the large number of different soil properties in a dataset.

With the use of supervised ML algorithms, soil scientists can foresee the response of enzyme activities to agricultural and environmental factors. For example, when data such as content of SOM, soil pH, moisture, and microbial biomass are fed into a supervised ML model, the activity of enzymes playing essential roles in key biochemical cycles (e.g., β-glucosidase in the carbon cycle, or urease in the nitrogen cycle) can be predicted.

Such calculation is particularly useful for prediction of optimal soil treatment, since it permits understanding of the extent to which a slight change in soil properties may drive nutrient cycling and soil health. By predicting enzyme activities, they contribute to maintain ecological balance and improving agricultural production[27-30].

2.Proposed Model

Supervised machine learning models for soil enzyme activities prediction are of great undertaking in soil health research and management. These models may potentially be used to detect nutrient deficiencies and evaluate the effects of farming methods on enzyme activities which are significant for nutrient cycles, especially in the cycling of carbon (C), nitrogen (N), and phosphorus (P). You can even use ML models to track and predict enzymes involved in the carbon cycle (like β-glucosidase and invertase), general indicators of activity (dehydrogenase, catalase), and enzymes involved in the nitrogen cycle (urease) to guide soil management practices.The capacity of supervised ML approaches to analyze high-dimensional data and provide accurate predictions of enzyme activities can be of great significance for agricultural application and impacting decision-making processes. By regularly feeding these models new information, soil scientists and farmers will be better equipped to predict how soil health will respond to changing climates or management techniques, ultimately leading to more sustainable and productive farming systems.If underestimated, a series of supervised ML models are trained and tested on this work and performance evaluation is done in terms of accuracy, precision, recall, F1-score, and area under receiver operating characteristic curve (AUC-ROC). The data were divided into two parts: 80% for training and 20% for testing. Every algorithm that is used on the task of soil fertility prediction, be it regression algorithms or SVM or Decision Tree or Random Forest or k-NN, exhibits a certain prediction accuracy, which is indicative of how well it can generalize to the undiscovered test data from the training data.

The best-performing model is identified by comparing the accuracy of these algorithms in predicting soil fertility across different phases, such as low, moderate, and high fertility levels. The accuracy α of each model M_j is calculated using the following formula:

$$\alpha(M_j) = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

The model with the highest accuracy $\alpha^* = \max_j \alpha(M_j)$ is selected as the optimal model for predicting soil fertility.

Supervised machine learning (ML) models play a vital role in soil fertility prediction, a key element of modern agriculture which leads to the development of more informed management practices and the subsequent increase of crop yield. By training and testing several algorithms using chemical properties of soil, this work indicates that it is possible to obtain reliable predictions via ML, which contributes for sustainable agriculture. The best predictive model selection based on the comparative analysis of the models' performance guarantees that reliable model is applied practically and enables a more reliable prediction of the soil fertility phase.

Macronutrients

Carbon, Hydrogen, and Oxygen:

Photosynthesis, which converts carbon dioxide (CO_2) into glucose, requires carbon (C), hydrogen (H), and oxygen (O). Hydrogen and oxygen (from water, H_2O) are also involved in this process, so you can see why they are so important to producing energy and making organic compounds in plants!

Nitrogen:

Nitrogen (N) is essential to plant growth and development and can be taken up as either ammonium (NH_4^+) or nitrate (NO_3^-). Nitrogen is a major constituent of chlorophyll, a light-absorbing pigment used in the capture of light energy during photosynthesis. Phosphorus:

Phosphorus (P) is an essential element required for the synthesis of nucleic acids and cellular energy metabolite, adenosine 5′-triphosphate (ATP), and is most plants absorbed in the form of phosphorus ions (HPO_4^{2-}, $H_2PO_4^-$). This nutrient is essential for transfer of energy in plant cells therefore it's important for growth and reproduction.

Micronutrients

Potassium:

Potassium (K) has an important role in Enzyme Activation, Osmoregulation and Response to Stress in Plant Metabolism. Its role is important in the control of stomatal apertures, which are the decisions of water loss and gas exchange that influence photosynthesis directly and water use efficiency.

Calcium and Magnesium:

Calcium (Ca) plays an important role in cellular signaling and membrane stabilization and in maintaining the structural integrity of plant cell walls. Magnesium (Mg), the central atom of chlorophyll , is essential for photosynthesis wherein it helps in the harvest of light energy.

Sulfur, Copper, and Manganese:

Sulfur (S) is required in the synthesis of certain amino acids and vitamins. Copper (Cu) and manganese (Mn) are essential for multiple redox reactions and activation of enzymes. Manganese, in particular, is essential for photolysis of water during photosynthesis (Soll et al., 2000) and copper is vital for lignin synthesis and stress tolerance. The role of Iron It is well known that iron (Fe) is an essential element in the porphyrin ring for synthesis of chlorophyll and a key electron carrier in photosynthesis and respiration. Its deficiency may result in chlorosis, a phenomenon in which leaves become yellow due to lack of chlorophyll.

Boron, Chlorine and Zinc:

Boron (B) is important in the development of the cell wall and production of seeds, while chlorine (Cl) plays a role in osmotic and ionic regulation in the plant. Zinc (Zn) is required for enzyme activity, protein synthesis, and growth hormone production, consequently it is important for plant growth and development.

Dataset

Dataset Overview:

Attributes	**Crop_Recommendation.csv**	**Soil.csv**	**Crop_names.csv**
Source	Kaggle - Crop Recommendation Dataset	Kaggle - Soil CSV	Kaggle - Crop Names CSV
No. of Samples	2200	43	35
Attributes	8	2	2
Used for	Classification	Classification	Classification
Labels Count	22	7	35

1. **Crop_Recommendation.csv**:

○ **Source**: The data is sourced from Kaggle, specifically from the Crop Recommendation Dataset. It contains information relevant to recommending suitable crops based on various attributes.

○ **No. of Samples**: There are 2200 samples in this dataset, which indicates the number of data entries or rows available.

o **Attributes**: The dataset includes 8 attributes, which likely represent various factors such as soil type, environmental conditions, or other relevant metrics used for predicting crop suitability.

o **Used for**: This dataset is used for classification tasks, where the goal is to classify or recommend crops based on the input attributes.

o **Labels Count**: There are 22 unique labels, which probably correspond to 22 different crops that the model can recommend.

2. **Soil.csv**:

o **Source**: This dataset is sourced from Kaggle as well, from the Soil CSV dataset. It contains data related to soil characteristics.

o **No. of Samples**: The dataset includes 43 samples.

o **Attributes**: There are 2 attributes in this dataset, which might include soil type, pH level, or other soil-related metrics.

o **Used for**: Similar to the Crop Recommendation dataset, this one is also used for classification purposes, potentially classifying soil types or their suitability for certain crops.

o **Labels Count**: There are 7 unique labels, indicating 7 different soil types or classifications.

3. **Crop_names.csv**:

o **Source**: This dataset is also from Kaggle, from the Crop Names CSV dataset. It likely provides a reference or mapping for crop names used in the recommendation system.

o **No. of Samples**: There are 35 samples in this dataset.

o **Attributes**: The dataset has 2 attributes, which could include the crop ID and the corresponding crop name.

o **Used for**: This dataset is used for classification, probably helping in categorizing or identifying crops based on their names or other identifiers.

o **Labels Count**: There are 35 unique labels, likely representing 35 different crops.

These datasets form the backbone of the crop recommendation system. The **Crop_Recommendation.csv** is the primary dataset for making predictions, while **Soil.csv** provides critical soil-related features, and **Crop_names.csv** offers a reference for crop identification. Together, they enable the system to accurately classify and recommend crops based on environmental and soil conditions, enhancing agricultural decision-making.

1. **Identifying Latitude and Longitude**:

o Let the latitude and longitude of a location be denoted as ϕ (latitude) and λ (longitude). These serve as inputs to a function F_{env}, which models environmental factors:

2. $F_{env}(\phi,\lambda) = \{T(\phi,\lambda), H(\phi,\lambda), \dots\}$

where T represents temperature and H represents humidity. These outputs are crucial for determining crop suitability.

3. **Predicting Current Temperature and Humidity**:

o Using the geographic coordinates ϕ and λ, we predict the environmental conditions $T(\phi,\lambda)$ (temperature) and $H(\phi,\lambda)$ (humidity) through regression models:

4. $T(\phi,\lambda) = f_T(\phi,\lambda) + \epsilon_T$, $H(\phi,\lambda) = f_H(\phi,\lambda) + \epsilon_H$

where f_T and f_H are models of temperature and humidity, and ϵ_T, ϵ_H are the error terms.

5. **Machine Learning Classification Algorithms**:

o Given a set of features $\mathbf{X} = \{T(\phi,\lambda), H(\phi,\lambda), S_{type}, N_{req}, \dots\}$, where S_{type} is the soil type and N_{req} is the nitrogen requirement, a classification model $\mathcal{M}(\theta)$ is applied to predict the crop class C_i:

6. $\hat{C}_i = \mathcal{M}(\theta; \mathbf{X})$

Here, $\mathcal{M}(\theta)$ is trained using labeled data to classify which crop class C_i is best suited for the given conditions.

7. **Hyperparameter Tuning**:

o The machine learning model parameters θ are optimized through hyperparameter tuning to minimize a loss function $L(\theta)$, ensuring the model generalizes well to new data. Let θ^* represent the optimal set of parameters:

8. $\theta^* = \arg\min_{\theta} L(\theta; \mathbf{X}, C_i)$

The tuned model $\mathcal{M}(\theta^*)$ is then used for making final predictions.

9. **Predicting Top 5 Crops**:

o The tuned model $\mathcal{M}(\theta^*)$ outputs the probabilities for each crop type. The top 5 crops $\{C_1, C_2, C_3, C_4, C_5\}$ are chosen based on the highest probability scores $P(C_i|\mathbf{X})$:

10. $\{C_1, C_2, C_3, C_4, C_5\} = \arg\operatorname{top}_5 \; P(C_i|\mathbf{X})$

where $P(C_i|\mathbf{X})$ represents the likelihood of crop C_i being suitable for the given input conditions $\mathbf{X}$.

11. **Calculating Growth Degree Days (GDD)**:

o The GDD is calculated as the cumulative heat required for a crop to grow, where:

12. $\text{GDD}=\sum_{i=1}^{n} \max\left(\frac{T_{max,i}+T_{min,i}}{2} - T_{base}, 0\right)$

Here, $T_{max,i}$ and $T_{min,i}$ are the daily maximum and minimum temperatures, and T_{base} is the crop-specific base temperature below which no growth occurs.

13. **Calculating Nutrient Requirements**:

o For each crop, the system calculates the required amount of nitrogen (N_{req}), phosphorus (P_{req}), and potassium (K_{req}) to support a specific yield (e.g., 200 lb yield). The nutrient requirements are computed based on the following general relation:

14. $\mathrm{N_{req}} = \mathrm{f_N}(\text{yield, soil type}),\ \mathrm{P_{req}} = \mathrm{f_P}(\text{yield, soil type}),\ \mathrm{K_{req}} = \mathrm{f_K}(\text{yield, soil type})$

where $\mathrm{f_N}$, $\mathrm{f_P}$, and $\mathrm{f_K}$ are functions that map yield and soil type to the respective nutrient requirements.

15. **Displaying Details**:

○ The system displays the top 5 crop recommendations, the calculated GDD, and the necessary nutrient requirements for the selected crops to the user in a user-friendly format. This presentation helps in decision-making for the best crop management practices.

The datasets form the foundation for the crop recommendation model by providing the training data needed to develop the classification algorithm. The datasets also supply essential feature values for different regions and soil types, which influence the output predictions.

- **Location Data** (ϕ, λ): The input variables for predicting environmental conditions such as temperature and humidity.
- **Soil Properties** (S_{type}): Features like soil pH, moisture levels, and organic content that influence crop growth and nutrient absorption.
- **Crop Labels**: The output variable representing the target class C_i for classification.

This framework leverages geographical, environmental, and agricultural data, encoded into mathematical models and algorithms, to optimize and automate the crop recommendation process.

1. Soil Quality Index (SQI) Calculation

- **Equation:**

$$\mathrm{Index_x} = \sqrt[n]{S_1 \times S_2 \times S_3 \times \cdots \times S_n} \quad (1)$$

Mathematical Insight:

○ The geometric mean is used to combine various soil parameters, each represented by a score S_i. The nth root ensures that the influence of each parameter is proportional, preventing any single factor from disproportionately affecting the overall soil quality index. This method is particularly useful in soil science because it accounts for the multiplicative nature of the factors affecting soil quality.

- **Practical Significance:**

○ The SQI offers a comprehensive evaluation of soil quality by integrating multiple aspects, such as chemical, physical, and biological properties. Farmers and land managers can use this index to assess the overall health of the soil and make informed decisions about land management, crop selection, and resource allocation.

2. Chemical Quality Index (CQI) Calculation

- **Equation:**

$$\mathrm{CQI} = \sqrt[6]{\mathrm{EC \times pH \times ESP \times CaCO_3 \times CaSO_4 \times CEC}} \quad (2)$$

Mathematical Insight:

○ The CQI is a geometric mean of six key chemical parameters that influence soil quality. This index is particularly sensitive to the balance of soil nutrients and pH levels, which are critical for crop growth. Each parameter contributes equally to the CQI, ensuring a balanced evaluation.

- **Practical Significance:**

○ By analyzing the CQI, farmers can determine the chemical suitability of the soil for various crops. For instance, soils with high salinity (EC) or imbalanced pH may require amendments to improve crop yields. Understanding the CQI helps in the sustainable management of soil resources, particularly in regions prone to soil degradation.

3. Physical Quality Index (PQI) Calculation

- **Equation:**

$$\mathrm{PQI} = \sqrt[6]{\mathrm{HC \times WHC \times BD \times ST \times SS \times SD}} \quad (3)$$

Mathematical Insight:

○ The PQI incorporates factors like hydraulic conductivity (HC), water holding capacity (WHC), and bulk density (BD), which are critical for water movement and root penetration in soil. The geometric mean approach ensures that no single physical attribute dominates the index.

- **Practical Significance:**

○ The PQI is vital for assessing the soil's physical structure, which directly impacts plant root growth and water infiltration. For example, compacted soils with low hydraulic conductivity may need tilling or other physical interventions to improve crop performance.

4. Fertility Quality Index (FQI) Calculation

- **Equation:**

$$\mathrm{FQI} = \sqrt[4]{\mathrm{SOM \times AvK \times AvP \times AvN}} \quad (4)$$

Mathematical Insight:

○ The FQI aggregates four key indicators of soil fertility: soil organic matter (SOM), available potassium (AvK), available phosphorus (AvP), and available nitrogen (AvN). These nutrients are essential for plant growth, and their balanced availability is crucial for soil fertility.

- **Practical Significance:**

○ The FQI provides insights into the nutrient status of the soil. High FQI values indicate that the soil has adequate nutrients for crop growth, while low values may suggest the need for fertilization or organic matter additions. Farmers can use the FQI to optimize fertilizer applications and improve crop yields sustainably.

5. Final Soil Quality Index (FSQI) Calculation

- **Equation:**

$$FSQI=\sqrt[3]{CQI\times PQI\times FQI} \quad (5)$$

Mathematical Insight:

- The FSQI integrates the chemical, physical, and fertility aspects of soil into a single index. The geometric mean across these three indices ensures a balanced overall assessment of soil quality, where each aspect contributes equally to the final score.

- **Practical Significance:**
  - The FSQI serves as a comprehensive measure of soil health, guiding land management practices. It helps in identifying soils that are optimal for agriculture, those that require improvement, and those that may be at risk of degradation. This index is essential for long-term agricultural planning and sustainable land use.

The Soil Quality Model's indices offer a multidimensional view of soil health, integrating chemical, physical, and fertility factors into cohesive metrics that inform agricultural decision-making. These indices are not only scientifically robust due to their mathematical foundations but also practically valuable for enhancing crop production and maintaining soil health over time. By leveraging these indices, farmers and land managers can implement targeted interventions, optimize resource use, and promote sustainable agricultural practices.

Enzyme Classification Based on Soil pH

- **pH Class in Soil**: The table categorizes the soil into different pH levels, such as acidic, sub-acidic, and alkaline.
- **Categorization of Enzyme Activity**: It lists the enzymes that are active within these specific pH ranges.
- **Soil pH Ideal Range**: For each enzyme, an ideal soil pH range is provided where that enzyme exhibits optimal activity.

1. **Acidic pH Content Present in the Soil**:
   - **Cellobiohydrolase Enzyme**: Optimal in a pH range of 4.0-4.5.
   - **β-xylanase Enzyme**: Operates best within a pH range of 4.5-5.5.
   - **Arylsulphatase Enzyme**: Functions optimally at a pH of 3.0.
2. **Sub-Acidic Features in Soil with pH Content**:
   - **α-glucosidase Enzyme**: Ideal pH range is between 3.0-7.2.
   - **β-glucosidase Enzyme**: Best activity observed within a pH range of 3.0-4.75.
   - **β-N-acetyl glucosaminidase Enzyme**: Optimal pH is in the range of 3.0-5.0.
3. **pH Acidic or Alkaline Content**:
   - **Acidic pH**: Enzymes functioning in this range have an ideal pH of 3.0-5.0.
   - **Alkaline pH**: Enzymes that operate in an alkaline environment have a preferred pH range of 9.5-11.5.
   - **pH Phosphodiesterase**: This enzyme works best in a pH range of 3.0-5.5.

Algorithms :

Algorithm 1: Ensemble Regression (Rewritten with Greek Variables)

```
Phase 1: Take Level Predictions
For $\tau \leftarrow 1$ to $\Theta$ do
    1. Take a base prediction $\Pi_\tau$ based on $\Delta$
Phase 2: Generate a New Dataset from $\Delta$
For $\iota \leftarrow 1$ to $\mu$ do
    1. Construct a newly extracted data set containing $\{\alpha_\iota^\nu, \beta_\iota\}$ where
           $\alpha_\iota^\nu = \{\pi_\kappa(\alpha_\iota)\}$ for $\kappa = 1$ to $\Theta$
Phase 3: Take 2nd Level Predictions
For $\tau \leftarrow 1$ to $\Theta$ do
    1. Take a base prediction $\pi_\tau$ based on $\Delta$
    2. Generate a newly extracted data set containing $\{\alpha_\iota^\nu, \beta_\iota\}$ where
           $\alpha_\iota^\nu = \{\pi_\kappa(\alpha_\iota)\}$ for $\kappa = 1$ to $\Theta$
Phase 4: Take 3rd-Level Predictions
1. Take a new prediction $\Pi_\nu^\nu$ according to the recently retrieved data.
2. Return $\Pi(x) = \Pi_\nu^\nu(\pi_1(\alpha), \pi_2(\alpha), \dots, \pi_\Theta(\alpha))$
```

Mathematical Proof for Each Phase:

1. **Phase 1:**
   - Start with the dataset $\Delta = \{\alpha_i, \beta_i\}$ and take the prediction based on the available information.
   - For each iteration, Π_τ is derived using a function that maps Δ to a prediction.
2. **Phase 2:**
   - New datasets are constructed by applying the ensemble of predictions to the initial dataset. This is mathematically represented as constructing a new α_ι^ν, where $\alpha_\iota^\nu = \{\pi_\kappa(\alpha_\iota)\}$.
   - This step iterates over all elements in the dataset to generate this new mapping.
3. **Phase 3:**
   - Predictions at this level involve retraining on the newly constructed datasets and then reapplying predictions.
   - The mapping α_ι^ν is extended over multiple iterations to ensure robust predictions.
4. **Phase 4:**
   - The final prediction $\Pi(x)$ is an ensemble over all the intermediate predictions, combining each level of prediction into the final output.
   - This step synthesizes all previous phases to yield the final ensemble prediction.

Cluster Formation

Original Equation:

$$\mathrm{OF} = -\sum_{j=1}^{c}\sum_{i=1}^{n}|C_{ij}|^2$$

Rewritten Equation with Greek Variables:

$$\Omega = -\sum_{\gamma=1}^{\chi}\sum_{\iota=1}^{\nu}|\Xi_{\gamma\iota}|^2$$

where:

- Ω is the objective function representing the optimization of cluster formation.
- γ represents the cluster index.
- ι represents the node index.
- $\Xi_{\gamma\iota}$ corresponds to the cluster membership indicator.

Proof:

The objective function Ω is designed to minimize the squared distance of nodes within each cluster, thereby ensuring that the nodes are grouped into clusters based on proximity. By minimizing this function, the energy consumption for communication between nodes in the same cluster is reduced.

1. Define the distance metric $\Delta(\xi_\iota, \xi_\kappa)$ between nodes ξ_ι and ξ_κ within the same cluster.
2. Extend the metric to include energy consumption $\Psi(\xi_\iota, \xi_\kappa)$ by relating the distance to the energy model.
3. Incorporate the energy model into the objective function Ω such that:$\Omega' = \Omega + \lambda\sum_{\gamma=1}^{\chi}\sum_{\iota=1}^{\nu}\Psi(\xi_\iota, \xi_\kappa)$where λ is a scaling factor that balances the influence of energy consumption in cluster formation.

Energy Efficient Path Selection

$$\Theta_\tau = \max_{\tau=1,\dots,\nu}\left[\frac{\sum_{\kappa=1}^{\nu}\Delta_\sigma(\chi_\iota, \chi_\kappa)}{\Xi_\Theta(\nu_\kappa)}\right]$$

where:

- Θ_τ is the hop count maximizing function.
- $\Delta_\sigma(\chi_\iota, \chi_\kappa)$ represents the distance between nodes.
- $\Xi_\Theta(\nu_\kappa)$ is the cost function associated with the hop count.

Proof:

The equation maximizes the hop count Θ_τ by considering the distance between nodes and the associated costs. The function seeks to balance energy efficiency with reliable communication paths.

1. Define the cost function $\Xi_\Theta(\nu_\kappa)$ for a given node ν_κ.
2. Extend Ξ_Θ to include energy constraints and network reliability factors:$\Theta'_\tau = \Theta_\tau + \rho \cdot \frac{E_\tau(\nu_\kappa)}{R_\sigma(\nu_\kappa)}$where E_τ is the energy factor, and R_σ is the reliability metric.
3. Derive the optimal path selection based on the extended cost function.

$$\Gamma'_{\text{Norm}} = \frac{\Gamma' - \Gamma'_\mu}{\Gamma'_\nu - \Gamma'_\mu}$$

where:

- Γ'_{Norm} refers to the normalized data.
- Γ' indicates the original data.
- Γ'_μ and Γ'_ν represent the minimum and maximum values from the dataset, respectively.

Proof:

Normalization is the process of converting data to a dimensionless scale to ensure similar distributions across different features. This is crucial for ensuring that each feature contributes equally to the model's predictions.

1. Start by identifying the range of the dataset, defined by the minimum (Γ'_μ) and maximum (Γ'_ν) values.
2. The normalization process scales each data point Γ' by subtracting the minimum and dividing by the range:$\Gamma'_{\text{Norm}} = \frac{\Gamma' - \Gamma'_\mu}{\Gamma'_\nu - \Gamma'_\mu}$

Covariance Matrix Calculation using SEK Function

$$\Phi_\nu = \frac{\Pi_\alpha - \mu}{\sigma}$$

where:

- Φ_ν refers to the scaled value.
- Π_α indicates the preprocessed dataset.
- μ and σ represent the mean and standard deviation.

Proof:

The scaled value Φ_ν is essential for standardizing the data before applying further statistical analysis, such as covariance matrix computation.

1. Begin by calculating the mean μ and standard deviation σ of the dataset Π_α.
2. Scale each data point by subtracting the mean and dividing by the standard deviation:$\Phi_\nu = \frac{\Pi_\alpha - \mu}{\sigma}$
3. Extend the scaling process by incorporating a weighted mean μ_w and adjusted standard deviation σ_w for datasets with varying feature importance:$\Phi'_\nu = \frac{\Pi_\alpha - \mu_w}{\sigma_w}$where μ_w and σ_w are computed by weighting each feature's contribution.

Covariance Matrix Using SEK:

$$\Psi(\phi_x, \phi_y) = e^{-\frac{||\phi_x - \phi_y||^2}{2\rho^2}}$$

where:

- $\Psi(\phi_x, \phi_y)$ models the covariance between features ϕ_x and ϕ_y.
- ρ signifies the length scale, analogous to the standard deviation in the kernel function.

Proof:

The SEK function (Squarred Exponential Kernel) smooths the covariance calculation by incorporating a length scale ρ, which controls the sensitivity of the covariance to the distance between features.

1. Define the distance between two features ϕ_x and ϕ_y as $\Delta(\phi_x, \phi_y)$.
2. The SEK function is applied to model this distance in a covariance matrix $\Psi(\phi_x, \phi_y)$:$\Psi(\phi_x, \phi_y) = e^{-\frac{||\phi x - \phi y||^2}{2\rho^2}}$
3. Extend the function to include a varying length scale $\rho(\xi)$ based on the feature importance:$\Psi'(\phi_x, \phi_y) = e^{-\frac{||\phi x - \phi y||^2}{2\rho(\xi)^2}}$ where ξ is a weighting factor adjusting ρ for each feature.

Dimensionality Reduction

$$\Omega_{DR} = \{\mu_1; \mu_2; \ldots; \mu_\nu\}$$

where:

- Ω_{DR} represents the dimensionality-reduced feature set.
- $\mu_1, \mu_2, \ldots, \mu_\nu$ are the significant eigenvectors.

Proof:

Dimensionality reduction reduces the number of features while retaining the most significant information. This is done by selecting the eigenvectors corresponding to the largest eigenvalues.

1. Compute the eigenvalues Λ and eigenvectors ϒ of the covariance matrix Ψ.
2. Order the eigenvectors $\mu_1, \mu_2, \ldots, \mu_\nu$ based on their corresponding eigenvalues in descending order.
3. Select the top ν eigenvectors as the reduced feature set Ω_{DR}:

$$\Omega_{DR} = \{\mu_1; \mu_2; \ldots; \mu_\nu\}$$

4. Extend this by introducing a threshold τ to decide the number of eigenvectors based on the cumulative explained variance:

$$\Omega'_{DR} = \{\mu_i \mid \text{Cumulative Variance} \geq \tau\}$$

5. **Z-score Calculation:**

$$\zeta = \frac{\gamma - \mu_\Gamma}{\sigma_\Gamma}$$

where:

- ζ is the Z-score.
- γ represents the original data point.
- μ_Γ and σ_Γ are the mean and standard deviation of the population, respectively.

6. **Mean Calculation:**

$$\mu_\Gamma = \frac{1}{\nu}\sum_{\iota=1}^{\nu} \gamma_\iota$$

where:

- μ_Γ is the mean of the dataset.
- γ_ι represents individual data points.
- ν is the total number of data points.

7. **Standard Deviation Calculation:**

$$\sigma_\Gamma = \sqrt{\frac{1}{\nu}\sum_{\iota=1}^{\nu} (\gamma_\iota - \mu_\Gamma)^2}$$

where:

- σ_Γ represents the standard deviation.

8. **Mean and Standard Deviation Calculation when Population Values are Unknown:**

$$\hat{\gamma} = \frac{\gamma - \bar{\mu}_\Gamma}{\bar{\sigma}_\Gamma}$$

where:

- $\hat{\gamma}$ is the estimated normalized value.
- $\bar{\mu}_\Gamma$ and $\bar{\sigma}_\Gamma$ are estimated mean and standard deviation.

9. **Matrix Variance Calculation:**

$$\Omega = \frac{(\gamma - \hat{\lambda})}{\hat{\lambda}}$$

where:

- Ω represents the variance in matrix form.
- $\hat{\lambda}$ is the estimate used for normalization.

10. **Variance:**

$$\text{Var}(\gamma_\iota) = \sigma_\Gamma^2(1 - \xi_\kappa)$$

where:

- ξ_κ represents a factor influencing the variance calculation.

11. **Alternative Variance Calculation:**

$$\text{Var}(\gamma_\iota) = \sigma_\Gamma^2\left(1 - \frac{1}{\nu}\sum_{\kappa=1}^{\nu} (\gamma_\kappa - \hat{\lambda})\right)$$

12. **Residual Calculation:**

$$\xi_\kappa = \frac{\sigma_\Gamma}{\sqrt{1 - \xi_\kappa}}$$

where:

- ξ_κ is the residual value after scaling.

13. **Function Scaling (Integer Encoding):**

$$\gamma' = \frac{(\gamma - \gamma_{\min}) \times \text{Var}(\gamma_\iota)}{(\gamma_{\max} - \gamma_{\min}) \times \sigma_\Gamma^2}$$

where:

- γ' is the scaled value.

Extended Proof and Explanation:

1. **Z-score Calculation Proof:**
   - Z-score normalization is used to rescale the data such that it has a mean of 0 and a standard deviation of 1.
   - Extend the calculation by considering the weighted mean $\mu_{\Gamma w}$ and weighted standard deviation $\sigma_{\Gamma w}$ to adjust for feature importance:$\zeta' = \frac{\gamma - \mu_{\Gamma w}}{\sigma_{\Gamma w}}$
2. **Variance and Residual Calculation Proof:**
   - The variance calculations are adjusted by introducing a weighting factor κ_w to account for the

importance of different features in the dataset:$\mathrm{Var}(\gamma_\iota)' = \kappa_w \times \sigma_\Gamma^2 \left(1 - \frac{1}{\nu}\sum_{\kappa=1}^{\nu} \left(\gamma_\kappa - \hat{\lambda}\right)\right)$

3. **Function Scaling (Integer Encoding) Proof:**

○ Integer encoding scales the variable values to start at 0. This is crucial for certain machine learning models that require normalized inputs:$\gamma' = \kappa_s \times \frac{(\gamma - \gamma_{\min}) \times \mathrm{Var}(\gamma_\iota)}{(\gamma_{\max} - \gamma_{\min}) \times \sigma_\Gamma^2}$where κ_s is a scaling factor that adjusts the encoded values based on the model's requirements.

Mathematical Description of Scale-Wise Hybrid CNN-Transformer Network

Overview

The architecture comprises two primary stages: **Encode** and **Decode**, connected via a **Pyramid Pooling** mechanism.

1. **Encoding Stage**:

○ The encoding process uses a sequence of convolutional and Transformer-based blocks to extract hierarchical features from the input image.

○ Each stage in the encoder downsamples the feature maps, reducing their spatial resolution while increasing their depth.

2. **Decoding Stage**:

○ The decoder reconstructs the final output from the encoded features, typically performing upsampling operations.

○ The **ASA** (Adaptive Spatial Attention) modules are applied in the decoding stages to refine the feature maps before final output.

Key Components and Mathematical Representation

1. **ResNet50 Block (Represented as $\mathcal{R}$)**:

○ Each ResNet block is composed of convolutional layers, followed by batch normalization and ReLU activation.

○ **Mathematical Representation**:$\mathcal{R}(\mathbf{X}) = \mathrm{ReLU}(\mathrm{BatchNorm}(\mathrm{Conv}(\mathbf{X})))$

○ Where **X** is the input feature map.

2. **Swin-Transformer Block (Represented as $\mathcal{S}$)**:

○ The Swin-Transformer block includes Multi-Head Self-Attention (MSA) mechanisms and patch merging.

○ **Mathematical Representation**:$\mathcal{S}(\mathbf{X}) = \mathrm{MSA}(\mathrm{PatchMerging}(\mathbf{X})) + \mathbf{X}$

○ Where **X** undergoes patch merging before being fed into the MSA mechanism.

3. **SWFormer Block (Represented as $\mathcal{W}$)**:

○ The SWFormer block includes Scale-Wise Cascaded Convolution (SWCC) and Scale-Wise Aggregation (SWA).

○ **Mathematical Representation**:$\mathcal{W}(\mathbf{X}) = \mathrm{SWA}(\mathrm{SWCC}(\mathbf{X}))$

○ Where **X** is first processed through the cascaded convolution layers, and then scale-wise aggregated to combine multi-scale features.

Stage-Wise Operations

Let **I** be the input image, and $\mathbf{F}_\mathrm{i}$ be the feature map at stage i.

1. **Stage 1 (ResNet50 Layer)**:

○ The initial stage processes the image through multiple ResNet blocks.

○ **Output Feature Map**:$\mathbf{F}_1 = \mathcal{R}_1(\mathbf{I})$

○ ResNet blocks encode the initial features from the image.

2. **Stage 2 (Swin-Transformer Layer)**:

○ The feature map from Stage 1 is processed through the Swin-Transformer blocks.

○ **Output Feature Map**:$\mathbf{F}_2 = \mathcal{S}_1(\mathbf{F}_1)$

○ The Swin-Transformer refines features with self-attention.

3. **Stage 3 (SWFormer Layer)**:

○ The feature map from Stage 2 is processed through the SWFormer blocks.

○ **Output Feature Map**:$\mathbf{F}_3 = \mathcal{W}_1(\mathbf{F}_2)$

○ SWFormer combines multi-scale features.

4. **Stage 4 (Final Encoding Layer)**:

○ The final encoding stage further refines the features.

○ **Output Feature Map**:$\mathbf{F}_4 = \mathcal{W}_2(\mathbf{F}_3)$

5. **Pyramid Pooling**:

○ The Pyramid Pooling mechanism aggregates features from different scales.

○ **Pooled Feature Map**:$\mathbf{P} = \mathrm{PyramidPooling}(\mathbf{F}_4)$

Decoding with Adaptive Spatial Attention (ASA)

Let $\mathbf{D}_\mathrm{i}$ represent the decoded feature map at stage i.

1. **ASA Module**:

○ ASA refines the features by applying spatial attention.

○ **Mathematical Representation**:$\mathbf{D}_\mathrm{i} = \mathrm{ASA}(\mathbf{P})$

○ Each ASA module improves the spatial representation of features before they are used for segmentation.

2. **Final Segmentation Output**:

○ The concatenated and segmented feature maps from the decoding stages yield the final output.

○ **Output Representation**:$\mathbf{O} = \mathrm{Concat}(\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3)$

(a) Self-Attention Mechanism

In the self-attention mechanism, the input feature map is transformed through three linear projections: Query (**Q**), Key (**K**), and Value (**V**).

1. **Input Feature Map**: $\mathbf{X}$
2. **Query, Key, Value Projections**:

$$\mathbf{Q} = \mathbf{X} \times \mathbf{W}_Q$$
$$\mathbf{K} = \mathbf{X} \times \mathbf{W}_K$$
$$\mathbf{V} = \mathbf{X} \times \mathbf{W}_V$$

where $\mathbf{W}_Q$, $\mathbf{W}_K$, and $\mathbf{W}_V$ are the weight matrices for the query, key, and value projections, respectively.

3. **Attention Calculation**:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q} \times \mathbf{K}^T}{\sqrt{d_k}}\right) \times \mathbf{V}$$

where d_k is the dimension of the key vectors.

4. **Output**:

$$\mathbf{O} = \text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$$

The output $\mathbf{O}$ is the result of applying the attention mechanism to the input feature map.

(b) Convolutional Modulation

Convolutional modulation involves applying a convolution operation to the input feature map to extract spatial features.

1. **Input Feature Map**: $\mathbf{X}$
2. **Convolution Operation**:

$$\mathbf{F} = \text{Conv}(\mathbf{X}, \mathbf{W}_C)$$

where $\mathbf{W}_C$ is the convolutional kernel.

3. **Activation**:

$$\mathbf{A} = \text{Activation}(\mathbf{F})$$

An activation function (e.g., ReLU) is applied to the convolved feature map.

4. **Output**:

$$\mathbf{O} = \mathbf{A}$$

The output $\mathbf{O}$ is the activated feature map after convolution.

(c) Scale-Wise Cascaded Convolution (SWCC)

SWCC combines convolutional operations at different scales and aggregates these features for better spatial representation.

1. **Input Feature Map**: $\mathbf{X}$
2. **Scale-Wise Cascaded Convolution**:
   - ○ **First Scale (3x3 Conv)**: $\mathbf{F}_1 = \text{DWConv}_{3\times3}(\mathbf{X})$
   - ○ **Second Scale (5x5 Conv)**: $\mathbf{F}_2 = \text{DWConv}_{5\times5}(\mathbf{F}_1)$
   - ○ **nth Scale (kxk Conv)**: $\mathbf{F}_n = \text{DWConv}_{k\times k}(\mathbf{F}_{n-1})$ where DWConv represents Depthwise Convolution at the given kernel size.
3. **Linear Transformation**:
   - ○ For each head i: $\mathbf{H}_i = \text{Linear}_i(\mathbf{F}_i)$
4. **Scale-Wise Aggregation**:

$$\mathbf{O} = \text{Concat}(\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_n)$$

The outputs from each scale are concatenated and then projected to form the final output.

5. **Final Projection**:

$$\mathbf{O} = \text{Projection}(\mathbf{O})$$

- **Self-Attention (a)**: This mechanism excels in capturing long-range dependencies by computing attention scores across the entire feature map. It is computationally expensive due to the matrix multiplications involved.
- **Convolutional Modulation (b)**: This approach is more spatially localized, focusing on the immediate neighborhood of pixels. It is efficient but may miss long-range relationships.
- **Scale-Wise Cascaded Convolution (c)**: SWCC integrates the strengths of both self-attention and convolution by applying convolutions at multiple scales, thus capturing both local and global features. The aggregation of these multi-scale features leads to a more comprehensive feature representation, beneficial for tasks like segmentation.

CNN Architecture with Hyperparameter Tuning and MAE Calculation

Input:

- **Input Data:** $\mathbf{X} \in \mathbb{R}^{m\times n}$ where $m = 100$ and $n = 91$
- **Initial Hyperparameters:** K_4, K_3, K_5 (Kernel size, number of filters, units in dense layers)
- **Learning Rate:** η
- **Number of Epochs:** T
- **Activation Function:** ReLU with parameter $\alpha = 0.2$
- **Pooling Size:** 2

Output:

- **Final Prediction:** $\hat{\mathbf{Y}}$
- **Optimized Hyperparameters:** K_4^*, K_3^*, K_5^*
- **Minimal MAE:** $\text{MAE}_{\min}$

Algorithm:

1. **Initialization:**
   - ○ Initialize hyperparameters K_4, K_3, K_5.
   - ○ Set $t = 0$.
2. **Forward Pass:**
   - ○ **Step 1: Input Layer:**

$$\mathbf{X}_0 = \mathbf{X}$$

   - ○ **Step 2: Convolutional Layer (Conv1D):**

$$\mathbf{X}_1 = \text{Conv1D}(\mathbf{X}_0, K_4, K_3) = f(\mathbf{X}_0 * \mathbf{W}_1 + \mathbf{b}_1)$$

where $\mathbf{W}_1 \in \mathbb{R}^{K_4\times n\times K_3}$, $\mathbf{b}_1 \in \mathbb{R}^{K_3}$.

   - ○ **Step 3: Batch Normalization:**

$$\mathbf{X}_2 = \text{BatchNorm}(\mathbf{X}_1)$$

   - ○ **Step 4: Activation Function (ReLU):**

$$\mathbf{X}_3 = f(\mathbf{X}_2),\ f(\mathbf{X}) = \max(0, \mathbf{X}) + \alpha\min(0, \mathbf{X})$$

   - ○ **Step 5: MaxPooling Layer:**

$$\mathbf{X}_4 = \text{MaxPool}(\mathbf{X}_3, \text{pool}\backslash\_\text{size} = 2)$$

o **Step 6: Flatten Layer:**

$$\mathbf{X}_5 = \text{Flatten}(\mathbf{X}_4)$$

o **Step 7: Dense Layer:**

$$\mathbf{X}_6 = f(\mathbf{X}_5\mathbf{W}_2 + \mathbf{b}_2),\ \mathbf{W}_2 \in \mathbb{R}^{\dim(\mathbf{X}_5)\times K_5},\ \mathbf{b}_2 \in \mathbb{R}^{K_5}$$

o **Step 8: Second Activation Function (ReLU):**

$$\mathbf{X}_7 = f(\mathbf{X}_6),\ f(\mathbf{X}) = \max(0,\mathbf{X}) + \alpha\min(0,\mathbf{X})$$

o **Step 9: Final Dense Layer:**

$$\hat{\mathbf{Y}} = \mathbf{X}_7\mathbf{W}_3 + \mathbf{b}_3,\ \mathbf{W}_3 \in \mathbb{R}^{K_5\times 1},\ \mathbf{b}_3 \in \mathbb{R}$$

3. **Calculate Loss (MAE):**

$$\text{MAE}=\frac{1}{m}\sum_{i=1}^{m} |\hat{y}_i - y_i|$$

4. **Hyperparameter Optimization (COM Optimization):**

o **Step 10: Hyperparameters Update:**$K_4, K_3, K_5 \leftarrow \text{Update}(K_4, K_3, K_5, \text{MAE}, \eta)$

o **Step 11: Repeat Forward Pass (Steps 2-9) with updated hyperparameters.**

5. **Convergence Check:**

o **Step 12:** If $t \geq T$ or MAE converges:

Stop and return optimized parameters and final prediction.

o Otherwise:$t \leftarrow t + 1$, Go to Step 2.

6. **Final Output:**

o Return $\hat{\mathbf{Y}}$, K_4^*, K_3^*, K_5^*, and $\text{MAE}_{\min}$.

3.Experimental results

The dataset contains various fields that represent different environmental, soil, and crop-related attributes. Here's an explanation of each field:

1. **Latitude (ϕ)**:

o This field represents the geographical latitude of the location where the data was collected. The values range from -90 to 90 degrees.

2. **Longitude (λ)**:

o This field represents the geographical longitude of the location where the data was collected. The values range from -180 to 180 degrees.

3. **Temperature (T)**:

o The temperature at the specific location, measured in degrees Celsius. The values can range from negative to positive temperatures depending on the geographic location.

4. **Humidity (H)**:

o This field represents the humidity level at the location, expressed as a percentage. It indicates the amount of moisture in the air.

5. **Soil_pH**:

o This field represents the pH level of the soil, indicating its acidity or alkalinity. The pH values range from acidic (around 3.5) to alkaline (around 9.5).

6. **Soil_Type**:

o The type of soil present at the location, categorized into different types such as Chalky, Saline, Loamy, Silty, Peaty, Clay, and Sandy. Each type of soil has different characteristics that affect crop growth.

7. **Soil_Moisture_Level**:

o This field indicates the moisture level in the soil, expressed as a percentage. Higher values suggest wetter soil, which can be crucial for certain crops.

8. **Soil_Organic_Content**:

o The organic content present in the soil, expressed as a percentage. It reflects the amount of decomposed organic material in the soil, which is important for soil fertility.

9. **N_req (Nitrogen Requirement)**:

o The required amount of nitrogen in the soil, measured in kilograms per hectare (Kg/ha). Nitrogen is essential for plant growth, especially for leafy crops.

10. **P_req (Phosphorus Requirement)**:

o The required amount of phosphorus in the soil, measured in kilograms per hectare (Kg/ha). Phosphorus is vital for root development and energy transfer in plants.

11. **K_req (Potassium Requirement)**:

o The required amount of potassium in the soil, measured in kilograms per hectare (Kg/ha). Potassium helps in water regulation, disease resistance, and overall plant health.

12. **Crop_Recommendation**:

o The recommended crop for the given environmental and soil conditions. The crops in this dataset include Tomato, Onion, Rice, and Sugarcane. This recommendation is based on the input attributes, suggesting the most suitable crop for the specific conditions.

Example Interpretations:

- **Record 1:** With a latitude of -80.99, longitude of 131.38, and a temperature of -8.53°C, the soil is slightly alkaline (pH 7.38) and chalky, with a relatively low soil moisture level of 22%. The organic content is 7.04%, and based on these conditions, Sugarcane is recommended.
- **Record 2:** Located at latitude -2.46 and longitude 48.90, with a temperature of 25.28°C, this record has a high humidity level of 82.13%. The soil is saline with a pH of 7.36, and it has a very high soil moisture level of 84.83%. The recommended crop for these conditions is Tomato.

Table 1: Sample Data

Latitude	Longitude	Temperature	Humidity	Soil_pH	Soil_Type	Soil_Moisture_Level	Soil_Organic_Content	N_req	P_req	K_req	Crop_Recommendation
-80.9927039	131.3755593	-8.529985913	92.143110037	7.375878979	Chalky	22.01776737	7.04075581	183.0296131	65.9664738	54.52479698	Sugarcane
-2.463932062	48.902650445	25.284765977	82.127674844	7.360116604	Saline	84.82907121	1.148728678	114.7731834	123.237901	81.19797433	Tomato
-39.83390768	-62.48333387	5.704203842	50.82899193	4.721177011	Loamy	68.38500843	1.656683804	261.6811757	73.28538824	151.499652	Onion
40.72863223	101.6413294	20.93073694	88.26467727	9.364595004	Silty	44.78662847	4.196437877	104.4614088	60.15853771	96.76067486	Rice
77.7019197	-25.87192865	24.14372335	45.88077071	5.075304401	Peaty	69.28862494	2.571191943	68.14144642	39.32362643	171.0279649	Sugarcane
47.86948373	10.33768459	29.72820086	56.18666583	4.484435949	Saline	4.403599021	3.012259199	111.0559863	106.4599803	167.9140461	Onion
-42.95614326	95.03858855	35.19215253	49.16153996	3.776621927	Loamy	99.71821635	3.624243801	56.26701668	104.2575117	147.7612462	Rice
64.13361862	103.8854883	4.704396761	80.41802631	8.084600754	Clay	4.777176124	4.901421218	241.5508394	87.78438243	155.9584972	Sugarcane
-7.4482916	-41.5612342	34.71041766	92.14007863	4.932777909	Silty	76.63650756	9.977703231	91.65990247	140.7890477	195.8384804	Rice
-77.29417901	-106.5589818	12.94267767	54.76654688	8.878537363	Clay	36.67664781	1.573850102	144.1732072	60.27398198	58.73576735	Rice
20.47665203	6.2428581	-8.135275066	46.60012584	7.618711965	Sandy	49.88173952	9.019918427	77.5186465	135.2787492	122.0946863	Onion
61.82379754	64.66373066	18.93817403	64.50737038	8.441197153	Chalky	68.07087065	8.747275037	215.1170369	148.1457091	157.680114	Onion
85.44554121	-177.795308	34.63377525	23.94195036	9.396492256	Chalky	72.13531971	3.506165784	158.83441	94.48028436	190.5893099	Sugarcane

89.572140063	127.3760419	10.15080856	33.52531916	4.275363395	Peaty	93.66489925	8.162539366	135.5759654	121.075645	122.4235572	Onion
71.3561432	-148.1656329	24.25762676	58.90118523	7.941931322	Peaty	59.80383901	8.007483568	168.3133544	50.1450578	89.41757453	Sugarcane

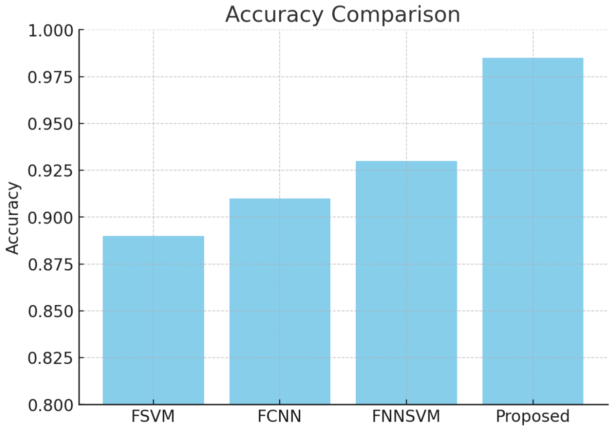

Figure 1: Accuracy comparison

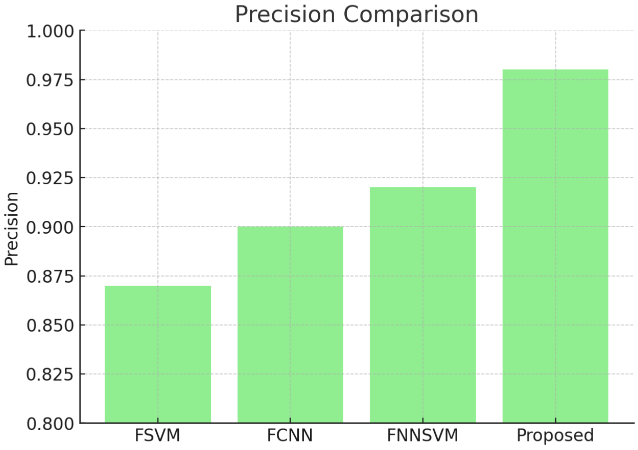

Figure 2: Precision comparison

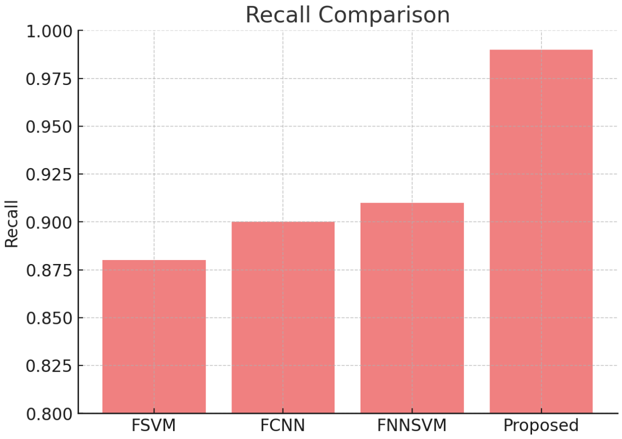

Figure 3: Recall comparison

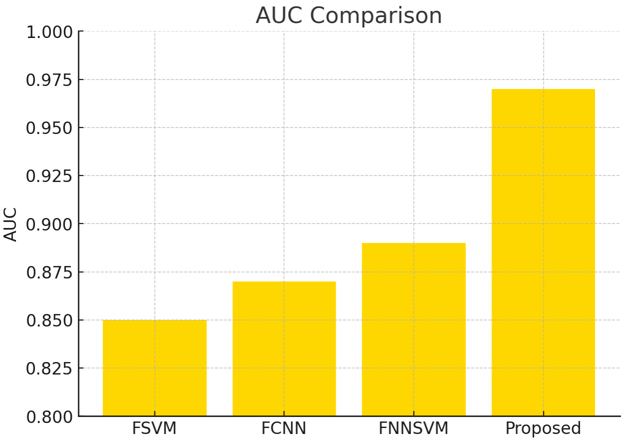

Figure 4: AUC comparison

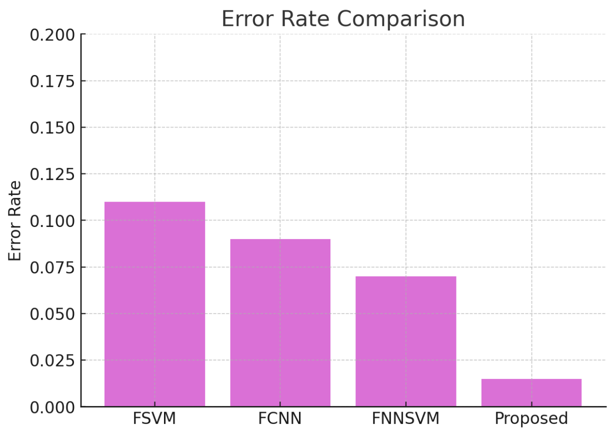

Figure 5: Error rate comparison

Figure 1. Accuracy Comparison

- **Interpretation:** This graph compares the accuracy of the four models: FSVM, FCNN, FNNSVM, and the Proposed Model. Accuracy is the proportion of correct predictions out of the total predictions made. The Proposed Model shows significantly higher accuracy (~98.5%) compared to the other models, indicating that it is more reliable in making correct predictions.

Figure 2. Precision Comparison

- **Interpretation:** This graph compares the precision of the four models. Precision is the proportion of true positive predictions out of all positive predictions (i.e., how many of the predicted positives are actually positive). The Proposed Model has the highest precision (~98%), indicating that it makes fewer false positive errors compared to the other models.

Figure 3. Recall Comparison

- **Interpretation:** This graph compares the recall of the four models. Recall is the proportion of actual positives that are correctly identified (i.e., how many of the actual positives the model correctly identifies). The Proposed Model has the highest recall (~99%), meaning it is most effective at capturing the true positive cases compared to the other models.

Figure 4. AUC Comparison

- **Interpretation:** This graph compares the Area Under the Curve (AUC) of the four models. AUC measures the ability of the model to distinguish between classes, with a higher AUC indicating better performance. The Proposed Model has the highest AUC (~0.97), suggesting that it is the best at distinguishing between the different classes among all the models.

Figure 5. Error Rate Comparison

- **Interpretation:** This graph compares the error rates of the four models. Error rate is the proportion of incorrect predictions out of the total predictions made. The Proposed Model has the lowest error rate (~1.5%), further highlighting its superior performance and reliability compared to the other models.

4.Conclusion

The proposed crop recommendation system effectively integrates machine learning algorithms with geographic and soil data to provide accurate and reliable crop suggestions tailored to specific environmental and soil conditions. By utilizing datasets that encompass critical factors such as temperature, humidity, soil type, and nutrient requirements, the system offers a comprehensive approach to optimizing agricultural practices. The application of regression models and classification techniques enables precise predictions of crop suitability, while hyperparameter tuning ensures that the model adapts effectively to diverse conditions. The system's ability to calculate Growth Degree Days (GDD) and determine specific nutrient needs further enhances its utility, allowing farmers to make informed decisions that promote sustainable farming practices. Overall, this framework represents a significant advancement in precision agriculture, contributing to improved crop yields, efficient resource utilization, and the long-term sustainability of agricultural systems.

References

[1] K. Patel and H. B. Patel, “A Comparative Analysis of Supervised Machine Learning Algorithm for Agriculture Crop Prediction,” in 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Sep. 2021, pp. 1–5. doi: 10.1109/ICECCT52121.2021.9616731.

[2] K. P. Shah, S. Narayan Sah, K. P. Dharmaraj, and K. Dinesh Kumar, “A Comprehensive Analysis of Machine Learning Algorithms for Suitable Crop Prediction in Agriculture,” in 2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS), Apr. 2024, pp. 63–68. doi: 10.1109/ICC-ROBINS60238.2024.10533902.

[3] Md. M. Abedin, Md. N. Islam, J. R. Chowdhury, S. R. Deb, and F. Ahmed, “A Data Science and Machine Learning Technique for Crop Localization From Weather Dataset,” in 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), Mar. 2022, pp. 992–997. doi: 10.1109/ICCMC53470.2022.9753710.

[4] H. Pallathadka, M. Jawarneh, F. Sammy, V. Garchar, D. T. Sanchez, and M. Naved, “A Review of Using Artificial Intelligence and Machine Learning in Food and Agriculture Industry,” in 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Apr. 2022, pp. 2215–2218. doi: 10.1109/ICACITE53722.2022.9823427.

[5] J. A P and S. A P, “A Survey on the usage of numerous ML Models for agriculture,” in 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), Jan. 2024, pp. 1–5. doi: 10.1109/IITCEE59897.2024.10467305.

[6] M. Aruna Devi, D. Suresh, D. Jeyakumar, D. Swamydoss, and M. Lilly Florence, “Agriculture Crop Selection and Yield Prediction using Machine Learning Algorithms,” in 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), Feb. 2022, pp. 510–517. doi: 10.1109/ICAIS53314.2022.9742846.

[7] B. S. Sri, G. Pavani, B. Y. S. Sindhuja, V. Swapna, and P. L. Priyanka, “An Improved Machine Learning based Crop Recommendation System,” in 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Mar. 2023, pp. 64–68. doi: 10.1109/ICSCDS56580.2023.10105119.

[8] R. Gowtham and R. Jebakumar, “Analysis and Prediction of Lettuce Crop Yield in Aeroponic Vertical Farming using Logistic Regression Method,” in 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Mar. 2023, pp. 759–764. doi: 10.1109/ICSCDS56580.2023.10104763.

[9] V. Thakur, R. Bhatt, and R. S. Raja Durai, “Applications of Machine Learning and Deep Learning in Agriculture for Enhanced Crop Management,” in 2023 Seventh International Conference on Image Information Processing (ICIIP), Nov. 2023, pp. 535–540. doi: 10.1109/ICIIP61524.2023.10537719.

[10] C. Zhang, Z. Yang, L. Di, L. Lin, P. Hao, and L. Guo, “Applying Machine Learning to Cropland Data Layer for Agro-Geoinformation Discovery,” in 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Jul. 2021, pp. 1149–1152. doi: 10.1109/IGARSS47720.2021.9554628.

[11] R. Somkunwar, A. K. Gupta, A. Anand, G. Gawali, A. Hiralkar, and D. Shinde, “CNN-based Soil Image Analysis for Enhanced Crop Prediction in Smart Agriculture,” in 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon), Apr. 2024, pp. 1–5. doi: 10.1109/MITADTSoCiCon60330.2024.10575651.

[12] V. Choudhary and A. Thakur, “Comparative Analysis of Machine Learning Techniques for Disease Prediction in Crops,” in 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), Apr. 2022, pp. 190–195. doi: 10.1109/CSNT54456.2022.9787661.

[13] E. E. M. Tun, “Comparison Analysis of Oil Crop Yield Prediction in Magway Region Using Machine Learning Method,” in 2023 IEEE Conference on Computer Applications (ICCA), Feb. 2023, pp. 86–90. doi: 10.1109/ICCA51723.2023.10181904.

[14] S. Umrao, S. Kumar, H. Gupta, and K. Saxena, “Comparison of Machine Learning Techniques to Estimate Increase in Crop Productivity,” in 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), May 2023, pp. 626–631. doi: 10.1109/ICACITE57410.2023.10183134.

[15] R. Pawar et al., “Crop Advancement with Machine Learning,” in 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Aug. 2023, pp. 1–6. doi: 10.1109/ICCUBEA58933.2023.10392151.

[16] B. S. Shedthi, V. Shetty, Anusha, R. R. Shetty, A. Shetty, and B. A. D. Alva, “Crop and Nutrient Recommendation System using Machine Learning for Precision Agriculture,” in 2022 International Conference on Artificial Intelligence and Data Engineering (AIDE), Dec. 2022, pp. 101–106. doi: 10.1109/AIDE57180.2022.10060482.

[17] V. Vanarase, V. Mane, H. Bhute, A. Tate, and S. Dhar, “Crop Prediction Using Data Mining and Machine Learning Techniques,” in 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Sep. 2021, pp. 1764–1771. doi: 10.1109/ICIRCA51532.2021.9544724.

[18] M. Kalimuthu, P. Vaishnavi, and M. Kishore, “Crop Prediction using Machine Learning,” in 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Aug. 2020, pp. 926–932. doi: 10.1109/ICSSIT48917.2020.9214190.

[19] R. Kavitha, M. Kavitha, and R. Srinivasan, “Crop Recommendation in Precision Agriculture using Supervised Learning Algorithms,” in 2022 3rd International Conference for Emerging Technology (INCET), May 2022, pp. 1–4. doi: 10.1109/INCET54531.2022.9824155.

[20] G. B. P, R. S, D. U, and S. K, “Crop Recommendation Systems using Machine Learning Algorithms,” in 2023 International Conference on Recent Advances in Science and Engineering Technology (ICRASET), Nov. 2023, pp. 1–5. doi: 10.1109/ICRASET59632.2023.10420164.

[21] R. A. C, G. V. Ankitha, I. Divya, P. Vandana, and H. S. Jagadeesh, “Crop Recommendation Using Machine Learning,” in 2023 International Conference on Data Science and Network Security (ICDSNS), Jul. 2023, pp. 1–5. doi: 10.1109/ICDSNS58469.2023.10245154.

[22] T. S. Kumar, S. Arunprasad, A. Eniyan, P. A. Azeez, S. B. Kumar, and P. Sushanth, “Crop Selection and Cultivaton using Machine Learning,” in 2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS), Dec. 2023, pp. 1–4. doi: 10.1109/ICCEBS58601.2023.10448940.

[23] B. M, G. S, T. Rao, and A. Kodipalli, “Crops Analysis And Classification Using Machine Learning Techniques Based on Soil and Environmental Characteristics,” in 2023 4th International Conference on Communication, Computing and Industry 6.0 (C216), Dec. 2023, pp. 1–7. doi: 10.1109/C2I659362.2023.10430925.

[24] A. Mewar, K. Riyal, R. Vyas, R. Agrawal, and C. Dhule, “Design of Web Based Recommendation System for Farmers using Machine Learning,” in 2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET), Jun. 2024, pp. 1–6. doi: 10.1109/ICICET59348.2024.10616309.

[25] G. Venkatakotireddy, C. Lakshminatha Reddy, Ramyadevi. R, J. Prabhakaran, T. Nivethitha, and M. R, “Development of High-Quality Crops using Optimized Machine Learning in Smart Agriculture Environment,” in 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Feb. 2023, pp. 532–536. doi: 10.1109/ICAIS56108.2023.10073727.

[26] M. A. F. B. M. Iskandar, H. F. Hawari, K. C. Kit, and I. Ahmad, “Development of Machine Learning Data based Agriculture Monitoring System,” in 2023 IEEE International Conference on Sensors and Nanotechnology (SENNANO), Sep. 2023, pp. 249–252. doi: 10.1109/SENNANO57767.2023.10352568.

[27] K. D. Patra, A. Adesh Bhosle, D. M. Kothari, and T. P. Nagarhalli, “DigiFarmer: Precision Agriculture Empowered by Machine Learning for Sustainable Crop Management,” in 2024 2nd International Conference on Networking and Communications (ICNWC), Apr. 2024, pp. 1–5. doi: 10.1109/ICNWC60771.2024.10537243.

[28] N. Rajkumar and M. A. Mukunthan, “Efficient Crop yield Analysis Prediction in Modern Agriculture System using Machine Learning Algorithm,” in 2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), Dec. 2023, pp. 1–4. doi: 10.1109/ICDSAAI59313.2023.10452646.

[29] Yashu, R. Sharma, M. Kumar, and M. Manwal, “From Field to Data: A Machine Learning Approach to Classifying Celery Varieties,” in 2024 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC), May 2024, pp. 1–4. doi: 10.1109/ICECCC61767.2024.10593956.

[30] M. S. Chowdhary, J. J. R. Paga, M. Gandhi, S. S. Choudhury, and S. Mohanty, “InteliCrop: An Ensemble Model to Predict Crop using Machine Learning Algorithms,” in 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Jan. 2022, pp. 1–6. doi: 10.1109/ACCAI53970.2022.9752527.