Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Integrating Machine Learning and IoT Sensors for Enhanced Soil Nutrient Monitoring and Crop Recommendation Systems

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

ARTICLE INFO

Received: 26 Dec 2024

Revised: 14 Feb 2025

Accepted: 22 Feb 2025

ABSTRACT

With the increase in the number of IoT farming datasets, identifying the appropriate data for IoT agriculture applications has become increasingly challenging. This research presents an advanced crop recommendation system developed by integrating various datasets, including Crop_Recommendation.csv, Soil.csv, and Crop_names.csv, which provide the foundation for accurate crop predictions. The system leverages geographic coordinates (latitude ϕ and longitude λ) to model environmental factors like temperature and humidity using regression models, forming essential inputs for crop suitability analysis. By applying a classification model , trained on features such as soil type and nitrogen requirements, the system predicts the most suitable crop class . Hyperparameter tuning optimizes the model to ensure robust predictions, and the system ranks the top five crops based on their likelihood of thriving under given conditions. Additionally, the system calculates Growth Degree Days (GDD) and nutrient requirements (nitrogen, phosphorus, potassium) for each recommended crop, offering a comprehensive decision-making tool for farmers. This framework, grounded in machine learning and geographical data, enhances agricultural decision-making by providing precise, data-driven crop recommendations tailored to specific environmental and soil conditions.

Keywords: Crop recommendation system, Machine learning algorithms, Soil nutrient monitoring, Geographic coordinates, Environmental modeling, Soil type classification, Nitrogen requirements, Hyperparameter tuning, Growth Degree Days (GDD), Nutrient requirements, Precision agriculture.

1.INTRODUCTION

Agriculture remains a cornerstone of the global economy, with the need for intelligent agricultural systems becoming increasingly crucial as the human population continues to grow. Over the past several decades, this sector has undergone numerous transformations to meet the needs of an expanding population, which has more than doubled in the last 50 years. Current projections suggest that by 2050, the global population will reach 9.8 billion. Additionally, a shift toward more urban living and a decrease in the ratio of working individuals to retirees are expected. As a result, agricultural productivity will need to increase in a sustainable manner while relying less on manual labor.The integration of technology into agriculture began over a century ago, with the introduction of the first tractor in 1913. Since then, mechanical advancements have surged, bringing numerous technologies to the market that have significantly boosted productivity while reducing the need for human labor. However, these advancements may not be sufficient to meet future global demand. In response, researchers have focused on improving production efficiency since the 1990s, leading to the development of "precision agriculture."[1-5] This approach involves farm management practices that optimize yields and resource use by observing, measuring, and responding to crop variability. More recently, existing technologies like remote sensing, the Internet of Things (IoT), and robotic platforms have been applied to agriculture, ushering in the era of "smart farmingSupervised machine learning algorithms, including Naive Bayes, K-nearest neighbors (KNN), support vector machines (SVM), and decision trees, are utilized to forecast soil fertility by analyzing a variety of chemical factors. The properties evaluated include pH content, electrical conductivity, organic carbon, nitrogen (N), phosphorus (P), potassium (K), iron (Fe),

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

and zinc (Zn). The model accuracies vary, ranging from 43% for Naive Bayes to 60% for SVM, indicating the varying effectiveness of these techniques based on the specific properties analyzed.A sustainable approach to organic soil analysis is developed by integrating machine learning techniques with a national spectral library. In this context, the random forest regression model is applied to a comprehensive dataset, analyzing soil features like soil organic matter (SOM). The performance of this model is measured using RMSE and R^2 metrics.Further predictions of soil fertility are anticipated using both simple and multiple regression techniques, based on data obtained from portable X-ray fluorescence spectrometry (pXRF) measurements. These techniques examine soil properties such as pH, clay content, silt, sand, calcium (Ca), magnesium (Mg), potassium (K), and aluminum (Al). The efficacy of these linear regression models is evaluated using R^2 and RMSE to ensure the robustness of soil fertility assessment[6-9]s.A comprehensive study evaluates the suitability of agricultural farmland by considering atmospheric conditions, nutrient levels, and soil properties. This classification system provides a holistic understanding of ecosystems and crop behavior, enabling better agricultural planning and management[10-14].Supervised machine learning algorithms, including K-nearest neighbors (KNN), Bagged Trees, and Gaussian Kernel-based support vector machines (SVM), are employed to predict soil series. Among these methods, SVM demonstrates superior performance, highlighting its effectiveness in soil classification tasks.To support agricultural development and enhance the quality of farm-grown products, a web-based service platform is developed. This platform offers various services related to soil classification, with a strong emphasis on precise data collection and methodological analysis, ensuring reliable and accurate soil assessments.The conceptual frameworks for soil quality and performance are reviewed, focusing on parameters that are crucial to farming. The proposed concept establishes a connection between soil fertility, nutrient availability, growing conditions, and soil improvement, linking these factors to broader soil quality and ecosystem services.Automated soil fertility prediction is carried out using advanced techniques such as neural networks, deep learning (DL), support vector machines (SVM), random forests (RF), and Bayesian models. These methods analyze soil properties, including phosphorus pentoxide (P_2O_5), iron (Fe), manganese (Mn), zinc (Zn), and nitrous oxide (N_2O). Model performance is evaluated using metrics like the coefficient of determination (R^2) and various regressor methods, ensuring the accuracy and reliability of the predictions.An innovative approach is employed to estimate coffee production by analyzing soil factors through extreme learning machines (ELM), multiple linear regression (MLR), and random forests (RF)[15-17]. The analysis includes soil features such as organic matter (OM), available potassium (K), and pH value, with model performance assessed using root mean square error (RMSE) and mean absolute error (MAE).Supervised learning methods are also applied to determine the spatial distribution of topsoil carbon (C), nitrogen (N), and available phosphorus (P). Techniques like the generalized additive model (GAM), random forests (RF), and SVM are used to define fertilizer application strategies, optimizing nutrient distribution for crop growth.Soil nutrient estimations based on village-specific characteristics are used to propose appropriate fertilizer levels and crop preferences. The model integrates various soil types and conditions according to detailed soil assessments, ensuring tailored agricultural practices for different regions.This review covers multiple aspects of soil fertility, including the sources of soil fertility datasets, performance metrics such as RMSE, MAE, and R^2, calibration methods, and dataset collection techniques. The subsequent section will explore soil enzyme activity estimation through different enzyme classifications, providing insights into soil health and microbial activity.Soil enzyme activity estimation is categorized using supervised machine learning algorithms, focusing on various soil properties, microbial activity, and soil organic matter. Performance metrics for enzyme datasets, collected via portable X-ray fluorescence (pXRF) and soil testing laboratories, are analyzed to assess the accuracy and reliability of these methods.In Brazilian regions, soil enzyme activity is analyzed considering factors such as season, fertility, and soil texture. Enzyme activities such as glucosidase, acid phosphatase, alkaline phosphatase, urease, and fluorescein diacetate (FDA) hydrolysis are investigated, with accuracy evaluated using conditional random forest methods.Extracellular enzymes within soil ecosystems are predicted using multiple linear regression (MLR) and random forest (RF) models. Soil characteristics like water content, electrical conductivity, total nitrogen (TN), total phosphorus (TP), and soil organic carbon (SOC) are utilized to estimate enzyme activities, including amylase and urease, providing valuable insights into soil health and nutrient cycling processes.Intercropping has been demonstrated to enhance soil extracellular enzyme activity, as supported by meta-analysis research. The diversity of plants introduced through intercropping significantly influences microbial communities within the soil, promoting a richer and more diverse ecosystem.Soil microbial elements and enzyme

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

activities are studied across various land types, including agricultural and vegetable fields. Metrics such as soil microbial biomass carbon (C_{MB}), respiration activities, and enzymes like β-glucosidase and acid phosphatase exhibit strong correlations with soil carbon (C) and nitrogen (N), indicating the pivotal role of these enzymes in soil health and nutrient cycling.The responses of soil enzyme activities to toxic metal exposure are analyzed, focusing on biochemical elements such as carbon (C), nitrogen (N), phosphorus (P), and sulfur (S). Heavy metal contamination adversely affects enzyme activities, disrupting the cycling of these essential elements within the soil ecosystem.

Machine learning techniques are employed to estimate biological modifications in soils following the addition of Phosphogypsum. Enzyme activities, such as urease and soil respiration, are assessed under varying concentrations of Phosphogypsum, providing insights into its impact on soil biochemistry.The effects of nitrogen fertilization on soil constituents, enzyme activities, and microbial communities are explored in detail. Nitrogen enrichment is shown to significantly enhance enzyme activities, with soil pH and organic carbon (OC) identified as key factors influencing changes in soil bacterial communities.Soil enzyme activities are highlighted as critical indicators of agricultural soil quality, facilitating processes such as decomposition and nutrient cycling. These activities respond dynamically to changes in soil management practices, making them practical indicators for assessing soil health and fertility.The impact of environmental factors and climate change on soil microorganisms is examined, particularly regarding the decomposition of chemicals by soil microorganisms. Variations in crop growth and physiological structure due to climate change also affect the efficacy of bioremediation efforts, as these changes influence the microbial community's ability to degrade pollutants.For decades, enhancing agricultural productivity has relied on the strategic utilization of soil minerals to ensure optimal crop growth. Historically, agricultural yield has been improved by converting organic materials, such as manure and crop residues, into essential nutrients. However, the 19th century's rapid industrialization and population expansion required a significant increase in agricultural output. This demand led to the development and widespread use of industrially produced phosphorus (P) fertilizers. Following the success of phosphorus fertilization, nitrogen (N) fertilization became an essential step, leading to the adoption of inorganic nitrogen (N), phosphorus (P), and potassium (K) fertilizers as foundational elements in crop development strategies. Given the complexity of soil fertility, we can model the interaction of these factors mathematically. Let $F_s(t)$ represent the soil fertility function over time t, where:

$$F_s(t) = f(N(t), P(t), K(t), h(C_{MB}(t), pH(t), OC(t), S(t)))$$

Here, $N(t)$, $P(t)$, and $K(t)$ represent the time-dependent concentrations of nitrogen, phosphorus, and potassium, respectively. The function $h(C_{MB}(t), pH(t), OC(t), S(t))$ captures the influence of soil microbial biomass carbon ($C_{MB}(t)$), pH, organic carbon (OC), and sulfur (S) on soil fertility.This equation underscores how these soil properties interact dynamically to influence soil fertility $F_s(t)$, ultimately impacting agricultural productivity.Effective soil management requires a thorough understanding of the complex interactions between these variables. By integrating multiple scientific disciplines and applying mathematical models to nutrient dynamics, we can develop optimized strategies that enhance crop yields while ensuring environmental sustainability.

The formulation of a soil fertility strategy within agricultural systems is essential for augmenting production capacities to sustain and improve economic returns. Beyond the simple management of soil nutrients, such strategies must also consider the myriad factors that influence the soil's capacity to supply nutrients and the efficiency with which crops assimilate these nutrients. Several additional elements, including soil moisture content, pH levels, salinity, physical structure, and biotic stresses, interact in complex ways to affect a crop's nutritional status. Consequently, a well-balanced nutrient profile alone is often insufficient to achieve the optimal functionality of an agricultural system[18].

Given the critical role of soil nutrients in crop productivity, the interaction of these variables can be modeled using mathematical equations. Let $F(t)$ represent the fertility function, which depends on various soil properties $S_i(t)$, where:

$$F(t) = f(N(t), P(t), K(t), g(M(t), pH(t), S(t), B(t)))$$

Here:

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

- $N(t)$, $P(t)$, and $K(t)$ represent the time-dependent concentrations of nitrogen, phosphorus, and potassium, respectively.
- $g(M(t), pH(t), S(t), B(t))$ is a function representing the interaction of soil moisture $M(t)$, pH levels $pH(t)$, salinity $S(t)$, and biotic stress factors $B(t)$ over time.

This equation illustrates how these soil properties dynamically interact to influence the fertility function $F(t)$, and by extension, agricultural productivity.

Effective soil fertility management necessitates a deep understanding of the intricate interactions between soil nutrients and other environmental factors. The integration of multiple scientific disciplines, combined with a mathematical approach to modeling nutrient dynamics, can lead to optimized strategies that not only enhance crop yields but also ensure economic and environmental sustainability.

This comprehensive understanding allows for the development of more robust and resilient agricultural systems, capable of adapting to varying conditions while maintaining productivity and ecological balance.Plant growth and development are fundamentally dependent on the availability and proper balance of essential nutrients. These nutrients, categorized into macronutrients and micronutrients, are integral to a range of physiological and biochemical processes that are vital to the plant life cycle. This paper delves into the roles of these nutrients, emphasizing their importance in photosynthesis, energy transfer, and overall plant health[19].

Machine learning algorithms, which are broadly classified into three categories—supervised learning, unsupervised learning, and reinforcement learning—are increasingly applied in the field of agricultural science. Specifically, in the context of soil fertility prediction, supervised machine learning techniques have proven to be highly effective due to their capability in handling both classification and regression tasks. These techniques encompass a variety of algorithms, including random forests, support vector machines (SVM), decision trees, linear regression, logistic regression, AdaBoost, XGBoost, Naive Bayes, and k-nearest neighbors (k-NN).

This study is centered on the application of supervised machine learning techniques for predicting soil fertility by leveraging key soil chemical properties such as soil organic carbon (OC), pH, and electrical conductivity (EC). The primary objective is to identify the most effective model that can accurately predict soil fertility based on these soil characteristics. The methodology involves evaluating multiple algorithms and conducting a comparative analysis to determine the most suitable supervised machine learning models.

To mathematically model soil fertility prediction, we define a supervised machine learning model M_θ as a function that maps input features X (soil characteristics such as OC, pH, EC) to an output prediction $\hat{y}$ (soil fertility level). This can be represented as:

$$\hat{y} = M_\theta(X)$$

where X represents the vector of input features, and θ denotes the model parameters that are optimized during the training process.

Given a dataset $\{(X_i, y_i)\}_{i=1}^{n}$, where X_i is the feature vector for the i-th observation and y_i is the corresponding soil fertility level, the goal is to minimize the loss function $L(y_i, \hat{y}_i)$ over the training set to find the optimal model parameters θ^*:

$$\theta^* = \arg\min_\theta \sum_{i=1}^{n} L(y_i, M_\theta(X_i))$$

Soil enzyme activity is a critical component of the biochemical cycles within soil ecosystems, significantly influencing the decomposition of organic matter, nutrient cycling, and overall soil health. These activities are closely associated with various soil properties, including soil organic matter (SOM), physical characteristics, microbial activity, and soil biomass. Accurately understanding and predicting soil enzyme activity is essential for optimizing agricultural practices and maintaining soil health. Supervised machine learning algorithms have emerged as powerful tools for

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

predicting soil enzyme activities based on these properties, offering a data-driven approach to enhancing agricultural productivity and sustainability[20-25].

The prediction of soil enzyme activity $\hat{y}$ can be mathematically formulated using a supervised learning model M_θ, which maps input features X (e.g., soil organic matter (SOM), soil pH, moisture content, microbial biomass) to an output prediction $\hat{y}$:

$$\hat{y} = M_\theta(X)$$

Here, X represents the vector of input features, and θ denotes the parameters of the machine learning model that are learned during the training process.

The objective is to find the optimal parameters θ^* that minimize the prediction error, which is typically quantified by a loss function $L(\hat{y}, y)$. This optimization problem can be expressed as:

$$\theta^* = \arg\min_\theta \frac{1}{n}\sum_{i=1}^{n} L(M_\theta(X_i), y_i)$$

where:

- n is the number of training examples,
- X_i is the feature vector for the i-th observation,
- y_i is the observed soil enzyme activity for the i-th observation.

A common loss function used in regression tasks is the mean squared error (MSE), defined as:

$$L(\hat{y}, y) = \frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

The goal is to adjust the model parameters θ such that the MSE is minimized, resulting in a model that accurately predicts soil enzyme activity based on the input features. This approach allows for the systematic prediction of soil enzyme activity, facilitating the optimization of agricultural practices and contributing to the overall health of soil ecosystems.

Application of Supervised Machine Learning Algorithms

Supervised machine learning (ML) algorithms, such as Random Forests, Support Vector Machines (SVMs), Decision Trees, and Linear Regression, are extensively utilized to model the intricate relationship between soil properties and enzyme activities. These models are designed to capture the complex, often nonlinear interactions between various soil factors and enzyme activities, offering valuable insights into soil health and agricultural productivity.

Random Forests

A Random Forest model M_{RF} is an ensemble learning method composed of multiple decision trees. Each tree T_j in the forest is trained on a random subset of the data, and the final prediction $\hat{y}$ is obtained by averaging the predictions from all trees:

$$\hat{y} = \frac{1}{m}\sum_{j=1}^{m} T_j(X; \theta_j)$$

where:

- m is the number of trees in the forest,
- T_j is the j-th tree in the ensemble,
- θ_j represents the parameters of the j-th tree,

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

- X is the vector of input features.

The Random Forest algorithm effectively reduces overfitting and improves predictive accuracy by aggregating the outputs of multiple decision trees.

Support Vector Machines (SVMs)

An SVM model M_{SVM} aims to find the optimal hyperplane that maximizes the margin between different classes (in classification tasks) or regression boundaries (in regression tasks). The prediction $\hat{y}$ for a given input X is given by:

$$\hat{y} = w \cdot X + b$$

where:

- w is the weight vector,
- b is the bias term.

The SVM model is particularly effective for high-dimensional spaces and is capable of handling both linear and nonlinear relationships by employing different kernel functions.

Impact on Soil Health and Agricultural Productivity

By applying these supervised ML models, soil scientists can predict how soil enzyme activities will respond to different agricultural practices and environmental conditions. This predictive capability has profound implications for nutrient cycling and agricultural productivity. For instance, enzymes involved in the carbon cycle (e.g., β-glucosidase), nitrogen cycle (e.g., urease), and overall soil health indicators (e.g., dehydrogenase) can be monitored and managed using these models.

Carbon Cycle Enzyme Prediction

Consider the prediction of β-glucosidase activity, denoted by $\hat{y}_{\beta\text{-gluc}}$. The model might use input features such as soil organic carbon (SOC), soil pH (pH), and soil moisture (M):

$$\hat{y}_{\beta\text{-gluc}} = M_{\theta}(\mathrm{SOC}, pH, M)$$

This prediction helps in understanding how changes in these soil properties affect the carbon cycling process. By analyzing the predicted enzyme activities, agricultural practices can be optimized to enhance soil health, ensuring sustainable productivity and ecological balance.

These models not only provide insights into the current state of soil health but also offer predictive capabilities that can guide future soil management practices, contributing to more efficient and sustainable agricultural systems.

Impact on Soil Health and Agricultural Productivity

The application of supervised machine learning (ML) models allows soil scientists to predict how soil enzyme activities will respond to various agricultural practices and environmental conditions. This predictive capability has profound implications for nutrient cycling and agricultural productivity. By accurately forecasting enzyme activities, these models help in understanding and managing essential processes within the soil ecosystem, thereby contributing to sustainable agricultural practices[26].

For example, enzymes involved in critical biochemical cycles—such as β-glucosidase in the carbon cycle, urease in the nitrogen cycle, and dehydrogenase as an overall indicator of soil health—can be effectively monitored and managed using these models. The ability to predict the behavior of these enzymes enables better decision-making in soil management, leading to enhanced crop yields and improved soil quality.

Carbon Cycle Enzyme Prediction

Consider the prediction of β-glucosidase activity, denoted by $\hat{y}_{\beta\text{-gluc}}$. This enzyme plays a crucial role in the carbon cycle by breaking down complex carbohydrates into simpler forms that plants can absorb and utilize. To predict its activity, the model might use input features such as soil organic carbon (SOC), soil pH (pH), and soil moisture (M):

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

$$\hat{y}_{\beta\text{-gluc}} = M_\theta(\mathrm{SOC}, pH, M)$$

where:

- $\hat{y}_{\beta\text{-gluc}}$ is the predicted activity of β-glucosidase,
- SOC represents the soil organic carbon content,
- pH is the soil pH level,
- M denotes soil moisture,
- M_θ is the supervised ML model with parameters θ.

The output of this model provides insights into how changes in soil organic carbon, pH, and moisture levels influence the carbon cycling process through the activity of β-glucosidase. By understanding these interactions, agricultural practices can be adjusted to optimize carbon cycling, thereby improving soil fertility and supporting sustainable crop production.

Role of Supervised Machine Learning Algorithms in Predicting Soil Enzyme Activity

Soil enzymes are integral to the biochemical processes that drive the recycling of organic materials within the soil ecosystem. Their activities are closely tied to various soil properties, including soil organic matter (SOM), physical characteristics, microbial activity, and soil biomass. Depending on their location within the soil, enzymes can be classified as either intracellular or extracellular.

Intracellular enzymes operate within living, metabolically active cells, as well as in dormant or dead cells where enzymes are located in the cytoplasm or attached to cell walls. These enzymes are crucial for maintaining cellular metabolism and contribute to the overall biochemical activity within the soil.

Extracellular enzymes, in contrast, are often immobilized within the soil matrix through mechanisms such as ionic interactions, covalent bonding, hydrogen bonding, and other forms of binding to clay particles and humic substances. These enzymes are pivotal in breaking down soil organic matter (SOM), facilitating nutrient cycling, energy transformation, ecological stability, and ultimately, agricultural productivity.

Despite their importance, agricultural practices such as mechanical tillage and excessive harvesting can negatively impact soil enzyme activity, leading to reduced nutrient availability for plants. It is well-documented that enzymatic activity tends to decrease with increasing soil depth and that these enzymes respond more rapidly to environmental changes and soil management practices than many other soil quality indicators. Consequently, well-established assays for a wide range of soil enzyme activities are commonly used as primary methods for assessing soil health (Bergstrom et al., 2000).

The Role of Supervised Machine Learning in Predicting Enzyme Activity

Supervised machine learning (ML) algorithms play a critical role in predicting soil enzyme activities by modeling the complex relationships between soil properties and enzymatic functions. These algorithms, such as Random Forests, Support Vector Machines (SVMs), and Decision Trees, are adept at handling nonlinear interactions and can effectively analyze large datasets containing diverse soil properties.

The application of supervised ML algorithms enables soil scientists to predict how enzyme activities will respond to various agricultural practices and environmental conditions. For instance, by inputting variables such as SOM content, soil pH, moisture levels, and microbial biomass into a supervised ML model, it is possible to predict the activity levels of enzymes involved in critical biochemical cycles, such as β-glucosidase in the carbon cycle or urease in the nitrogen cycle.

Such predictions are invaluable for optimizing soil management practices, as they provide insights into how changes in soil properties can affect nutrient cycling and overall soil health. By accurately predicting enzyme activities, these models help in maintaining ecological stability and enhancing agricultural productivity[27-30].

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supervised Machine Learning in Predicting Soil Enzyme Activity

Supervised machine learning (ML) algorithms have emerged as powerful tools for predicting soil enzyme activities, which are influenced by both biotic and abiotic factors. These algorithms leverage soil properties and environmental parameters as input features to predict enzymatic activity levels, which are critical for understanding soil health and nutrient cycling.

Mathematically, the relationship between soil enzyme activity $\hat{y}$ and its predictors can be modeled using a supervised learning algorithm M_θ, which maps a set of input features X (e.g., soil organic matter (SOM), soil pH, moisture content, microbial biomass) to an output prediction $\hat{y}$:

$$\hat{y} = M_\theta(X)$$

where:

- X is the feature vector comprising the relevant soil and environmental characteristics,
- θ represents the model parameters that are learned during the training process.

Algorithmic Approach to Soil Enzyme Activity Prediction

To predict soil enzyme activities, multiple supervised ML algorithms—such as random forests, support vector machines (SVMs), decision trees, and linear regression—can be applied. The primary objective is to develop a model that minimizes the prediction error by accurately capturing the complex interactions between soil properties and enzyme activities.

Given a training dataset $\{(X_i, y_i)\}_{i=1}^{n}$, where X_i represents the feature vector for the i-th observation and y_i is the observed enzyme activity, the supervised ML model is trained by optimizing the following objective function:

$$\theta^* = \arg\min_\theta \frac{1}{n}\sum_{i=1}^{n} L(M_\theta(X_i), y_i)$$

Here:

- $L(\hat{y}_i, y_i)$ is a loss function that quantifies the discrepancy between the predicted enzyme activity $\hat{y}_i = M_\theta(X_i)$ and the actual enzyme activity y_i,
- n is the number of observations in the training dataset.

Commonly used loss functions in regression tasks include:

- **Mean Squared Error (MSE):**

$$L(\hat{y}_i, y_i) = \frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

- **Mean Absolute Error (MAE):**

$$L(\hat{y}_i, y_i) = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i|$$

The choice of the loss function depends on the specific application and the nature of the data. MSE is sensitive to outliers, making it suitable when larger errors need to be penalized more heavily. MAE, on the other hand, provides a more robust measure of prediction accuracy when the impact of outliers needs to be minimized.

By optimizing the loss function, the supervised ML model learns to accurately predict soil enzyme activities based on the given input features. This predictive capability is vital for enhancing soil management practices, improving nutrient cycling, and ultimately contributing to the sustainability of agricultural ecosystems.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2.PROPOSED MODEL

Supervised machine learning models for predicting soil enzyme activities have significant implications for understanding and managing soil health. These models can help identify nutrient imbalances and assess the impact of agricultural practices on enzyme activities, which are crucial for nutrient cycling, particularly in the carbon (C), nitrogen (N), and phosphorus (P) cycles. For instance, carbon cycle-related enzymes such as β-glucosidase and invertase, general activity indicators like dehydrogenase and catalase, and nitrogen cycle enzymes such as urease, can all be monitored and predicted using ML models to inform better soil management practices.The ability of supervised ML models to integrate complex datasets and predict enzyme activities with high accuracy makes them invaluable for agricultural planning and decision-making. By continuously updating these models with new data, farmers and soil scientists can anticipate changes in soil health due to environmental shifts or management practices, thus enhancing the sustainability and productivity of agricultural systems.In this study, various supervised machine learning models were trained and evaluated using performance metrics such as accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). The dataset was split into two subsets: 80% for training and 20% for testing. Each algorithm applied to the soil fertility prediction task—whether regression models, SVMs, decision trees, random forests, or k-NN—yields different levels of prediction accuracy, reflecting its ability to generalize from the training data to unseen test 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 models are instrumental in predicting soil fertility, a crucial task in modern agriculture that informs better management practices and enhances crop yield. By training and evaluating multiple algorithms on soil chemical properties, this study demonstrates the effectiveness of machine learning in providing accurate predictions, thereby supporting sustainable agricultural practices. The comparative analysis of model performances ensures that the most accurate model is chosen for practical application, leading to more reliable predictions of soil fertility phases.

Macronutrients

Carbon, Hydrogen, and Oxygen:

Carbon (C), hydrogen (H), and oxygen (O) are integral to the process of photosynthesis, wherein carbon dioxide (CO_2) is converted into glucose. Hydrogen, derived from water (H_2O), and oxygen are also directly involved in this process, highlighting their crucial roles in energy production and organic molecule synthesis within plants.

Nitrogen:

Nitrogen (N), absorbed as ammonium (NH_4^+) or nitrate (NO_3^-), is a vital component of amino acids, the building blocks of proteins, and nucleic acids, which are essential for plant growth and development. Nitrogen is also a key element in chlorophyll, the pigment responsible for capturing light energy during photosynthesis.

Phosphorus:

Phosphorus (P), mainly taken up as phosphate ions (HPO_4^{2-}, $H_2PO_4^-$), is crucial for the formation of nucleic acids and the energy currency of cells, adenosine triphosphate (ATP). This nutrient is indispensable for energy transfer processes within plant cells, making it fundamental for growth and reproduction.

Micronutrients

Potassium:

Potassium (K) plays a multifaceted role in plant metabolism, including enzyme activation, osmoregulation, and stress

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

response. Its presence is critical for the regulation of stomatal openings, which control water loss and gas exchange, thus directly affecting photosynthesis and water use efficiency.

Calcium and Magnesium:

Calcium (Ca) is essential for the structural integrity of plant cell walls and membrane stabilization, playing a key role in cellular signaling. Magnesium (Mg), as the central atom in the chlorophyll molecule, is indispensable for photosynthesis, facilitating the capture of light energy.

Sulfur, Copper, and Manganese:

Sulfur (S) is important for synthesizing certain amino acids and vitamins. Copper (Cu) and manganese (Mn) are required for various redox reactions and enzyme activation. Manganese, in particular, is critical for the photolysis of water during photosynthesis, while copper is essential for lignin synthesis and stress resistance.

Iron:

Iron (Fe) is pivotal in the synthesis of chlorophyll and functions as an electron carrier in both photosynthesis and respiration. Its deficiency can lead to chlorosis, a condition characterized by yellowing leaves due to inadequate chlorophyll production.

Boron, Chlorine, and Zinc:

Boron (B) is crucial for cell wall formation and reproductive growth, while chlorine (Cl) is involved in osmotic regulation and ionic balance within the plant. Zinc (Zn) is necessary for enzyme function, protein synthesis, and the production of growth hormones, making it vital for overall plant health 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.
   - **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.
   - **Used for**: This dataset is used for classification tasks, where the goal is to classify or recommend crops based on the input attributes.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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

2. **Soil.csv**:
   - **Source**: This dataset is sourced from Kaggle as well, from the Soil CSV dataset. It contains data related to soil characteristics.
   - **No. of Samples**: The dataset includes 43 samples.
   - **Attributes**: There are 2 attributes in this dataset, which might include soil type, pH level, or other soil-related metrics.
   - **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.
   - **Labels Count**: There are 7 unique labels, indicating 7 different soil types or classifications.
3. **Crop_names.csv**:
   - **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.
   - **No. of Samples**: There are 35 samples in this dataset.
   - **Attributes**: The dataset has 2 attributes, which could include the crop ID and the corresponding crop name.
   - **Used for**: This dataset is used for classification, probably helping in categorizing or identifying crops based on their names or other identifiers.
   - **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**:
   - 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. $F_{\text{env}}(\phi,\lambda) = \{T(\phi,\lambda), H(\phi,\lambda), \dots\}$

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

3. **Predicting Current Temperature and Humidity**:
   - 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**:
   - Given a set of features $\mathbf{X} = \{T(\phi,\lambda), H(\phi,\lambda), S_{\text{type}}, N_{\text{req}}, \dots\}$, 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. $\hat{C}_i = \mathcal{M}(\theta; \mathbf{X})$

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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**:
    - 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**:
    - 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. $\{C_1, C_2, C_3, C_4, C_5\} = \arg\text{top}_5\ P(C_i|\mathbf{X})$

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

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

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**:
    - 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. $\mathrm{N}_{\text{req}} = f_N(\text{yield, soil type}),\ \mathrm{P}_{\text{req}} = f_P(\text{yield, soil type}),\ \mathrm{K}_{\text{req}} = f_K(\text{yield, soil type})$

where f_N, f_P, and 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

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

- **Equation:**

$$\text{Index}_x = \sqrt[n]{S_1 \times S_2 \times S_3 \times \cdots \times S_n} \ (1)$$

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:**

$$\text{CQI}=\sqrt[6]{\text{EC}\times\text{pH}\times\text{ESP}\times\text{CaCO}_3 \times \text{CaSO}_4 \times \text{CEC}} \ (2)$$

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:**

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

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:**

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

Mathematical Insight:

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

  - 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:**

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

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**:

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

- **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 Δ

```
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.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

- This step synthesizes all previous phases to yield the final ensemble prediction.

Cluster Formation

Original Equation:

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

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

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

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.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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

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 $\Delta(\phi_x, \phi_y)$.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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

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

where:

- Ω_{DR} represents the dimensionality-reduced feature set.
- $\mu_1, \mu_2, \dots, \mu_\nu$ 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 $\mu_1, \mu_2, \dots, \mu_\nu$ based on their corresponding eigenvalues in descending order.
3. Select the top ν eigenvectors as the reduced feature set Ω_{DR}:

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

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

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

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:**

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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

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):**

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

where:

- γ' is the scaled value.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

   - Where $\mathbf{X}$ 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 $\mathbf{X}$ is first processed through the cascaded convolution layers, and then scale-wise aggregated to combine multi-scale features.

Stage-Wise Operations

Let $\mathbf{I}$ be the input image, and $\mathbf{F}_i$ 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 $\mathbf{D}_i$ represent the decoded feature map at stage i.

1. **ASA Module**:
   - ASA refines the features by applying spatial attention.
   - **Mathematical Representation**:$\mathbf{D}_i = \mathrm{ASA}(\mathbf{P})$
   - 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.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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

(a) Self-Attention Mechanism

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

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**:

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

- **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:** MAE_{min}

Algorithm:

1. **Initialization:**
   - Initialize hyperparameters K_4, K_3, K_5.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

- Set $t = 0$.

2. **Forward Pass:**
   - **Step 1: Input Layer:**

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

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

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

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:**

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

   - **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:**

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

   - **Step 6: Flatten Layer:**

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

   - **Step 7: Dense Layer:**

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

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

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

   - **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):**

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

4. **Hyperparameter Optimization (COM Optimization):**
   - **Step 10: Hyperparameters Update:**$K_4, K_3, K_5 \leftarrow \mathrm{Update}(K_4, K_3, K_5, \mathrm{MAE}, \eta)$
   - **Step 11: Repeat Forward Pass (Steps 2-9) with updated hyperparameters.**

5. **Convergence Check:**
   - **Step 12:** If $t \geq T$ or MAE converges:Stop and return optimized parameters and final prediction.
   - Otherwise:$t \leftarrow t + 1$, Go to Step 2.

6. **Final Output:**
   - Return $\hat{\mathbf{Y}}$, K_4^*, K_3^*, K_5^*, and $\mathrm{MAE}_{\min}$.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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)**:
   - 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)**:
   - 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)**:
   - 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)**:
   - 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**:
   - 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**:
   - 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**:
   - 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**:
   - 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)**:
   - 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)**:
    - 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)**:
    - 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**:

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

- 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	92.143111037	7.375878979	Chalky	22.01776737	7.04075581	183.0296131	65.9664738	54.52479698	Sugarcane
-2.463932062	48.90265045	25.28476597	82.12767484	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	103.854883	4.704396761	80.41802631	8.084600754	Clay	4.777176124	4.901421218	241.5508394	87.78438243	155.9584972	Sugarcane

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

-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
89.57214063	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

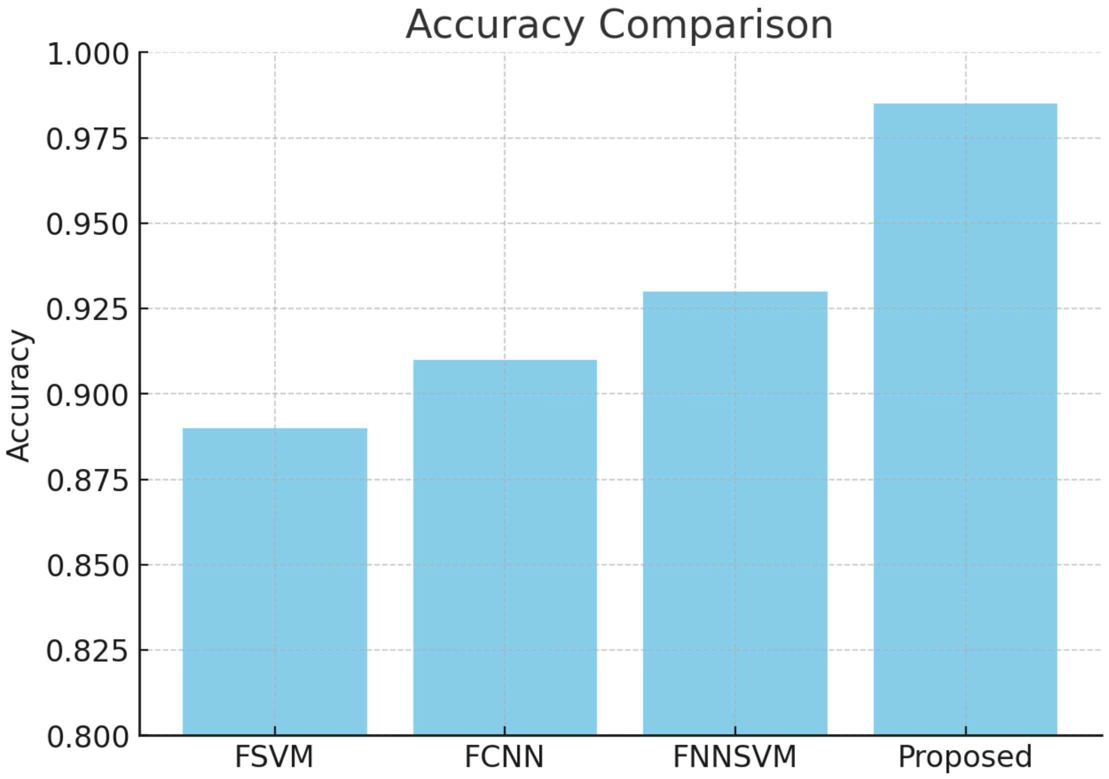

Figure 1: Accuracy comparison

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

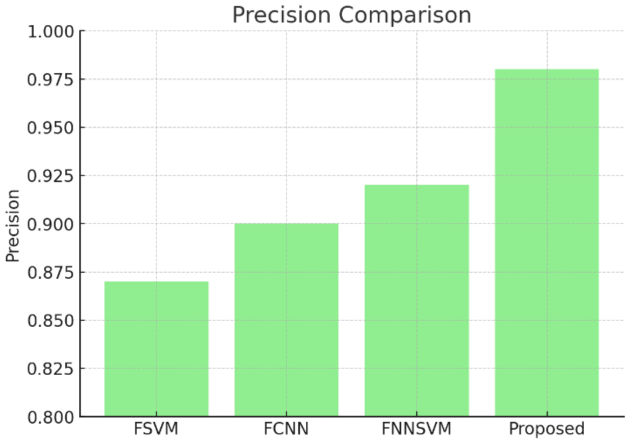

Figure 2: Precision comparison

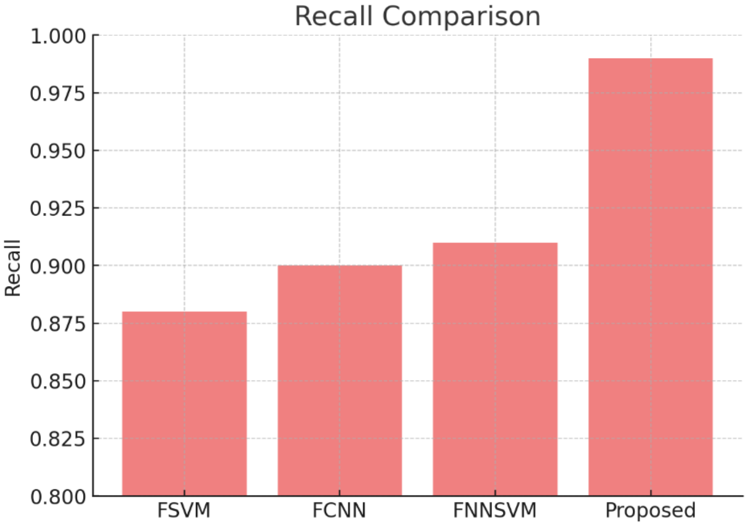

Figure 3: Recall comparison

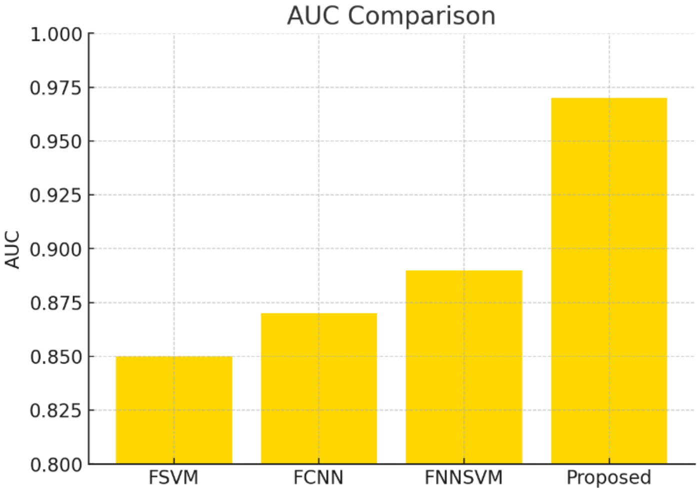

Figure 4: AUC comparison

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

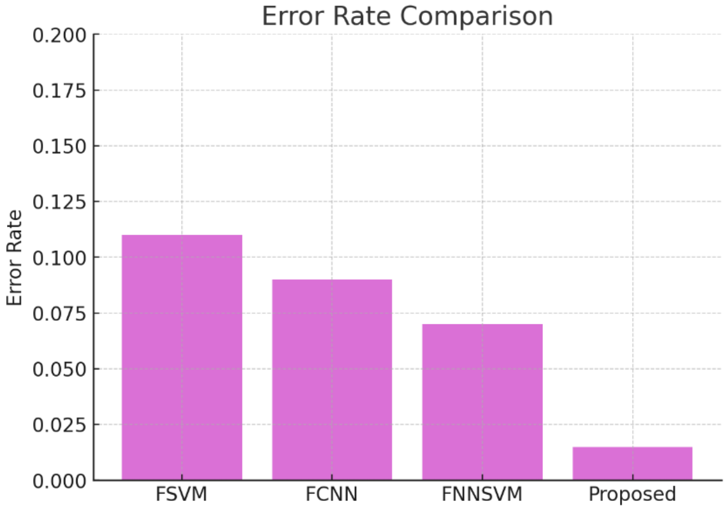

Figure 5: Error rate comparison

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

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[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.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[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.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.