0

Soil Analysis for Optimized Plant Cultivation Using Machine Learning and IoT Technologies

Shivani Kulkarni
shivani.gk2004@gmail.com
Marathwada Mitra Mandal's College of Engineering, Pune

Anushka Chavan
chavananushka782@gmail.com
Marathwada Mitra Mandal's College of Engineering, Pune

Shubham Dhobale
shubhamdhobale1945@gmail.com
Marathwada Mitra Mandal's College of Engineering, Pune

Aditya Shivarkar
adityashivarkar12@gmail.com
Marathwada Mitra Mandal's College of Engineering, Pune

Ms. Rohini Tambe
rohinitambe66@gmail.com
Marathwada Mitra Mandal's College of Engineering, Pune

ABSTRACT

Soil quality plays a crucial role in successful plant cultivation, as it directly influences the availability of nutrients, water retention, and overall plant health. This project focuses on the testing and classification of soil samples to determine their suitability for agricultural and horticultural purposes. Various soil samples were collected from different locations and tested for key physical and chemical properties such as pH, texture, moisture content, organic matter, nitrogen (N), phosphorus (P), and potassium (K) levels. Standard laboratory procedures were followed to ensure accurate and consistent results. Based on the test outcomes, soils were classified into categories such as sandy, clayey, loamy, and silty, and further assessed for their fertility and compatibility with different plant types. The findings of this study help in understanding how soil characteristics affect plant growth and guide farmers or gardeners in choosing the right type of soil or necessary treatments to improve it. This project highlights the importance of soil testing as a foundational step in sustainable agriculture and emphasizes the need for informed decision-making in crop planning. It serves as a useful reference for promoting better soil management practices aimed at improving productivity while conserving environmental resources.

Keywords: *Soil classification, image processing, convolutional neural networks (CNN), deep learning, machine learning, IoT, real-time monitoring, nutrient detection, crop recommendation, precision agriculture.*

INTRODUCTION

Soil is the cornerstone of agriculture, critically influencing crop yield and quality. Its composition and properties deter- mine water retention, nutrient availability, and plant health. Traditional soil testing methods, while accurate, are often labor-intensive and time-consuming, requiring manual sam- pling and laboratory analysis. With the advent of machine learning and image processing techniques, automated soil classification is now feasible, providing rapid, cost-effective, and non-invasive analysis. In this work, we propose a dual- mode system that supports both image-based soil classification and IoT-based real-time nutrient monitoring. By leveraging high-resolution imaging and pattern recognition, our approach enhances soil evaluation efficiency. Although both modes are integrated, the primary emphasis of this paper is on the image- based approach, which forms the basis for subsequent nutrient estimation and crop recommendation. This advancement aims to streamline soil assessment and support precision agriculture, ultimately improving productivity and sustainability.

PROPOSED SYSTEM

The proposed system integrates two parallel modules to pro- vide a comprehensive and data-driven soil analysis framework:

• Image-Based Soil Classification: This module employs a convolutional neural network (CNN) to extract detailed textural and compositional features from uploaded soil images. The model classifies the images into different soil types such as loamy, sandy, clayey, or silty, enabling rapid and automated soil characterization. Advanced pre- processing techniques and data augmentation enhance the model's robustness to variations in lighting, moisture, and environmental conditions.

• IoT-Based Real-Time Soil Testing: This module utilizes a network of sensors to capture key environmental and soil parameters, including moisture content, pH levels, nitrogen, phosphorus, and potassium (NPK) concentrations. The collected data is transmitted to a cloud-based plat- form for real-time analysis, nutrient prediction, and crop recommendation, enabling dynamic decision-making for farmers. While both modules work in tandem to deliver comprehensive soil assessment and precision farming recommendations, the remainder of this paper primarily focuses on the image-based classification module, which serves as the foundation for subsequent nutrient estimation.

SYSTEM IMPLEMENTATION

The system is divided into key modules to ensure modularity, scalability, and better maintainability. Below are the primary modules along with their roles and descriptions:

A] Data Collection:

- Collect soil samples and input data (physical and chemical properties such as pH, moisture, nitrogen, phosphorus, potassium).
- Enable image uploads for soil texture and color analysis.
- Integrate microcontrollers or sensors to collect real-time soil data (optional).

B] Data Processing:

- Perform data cleaning to handle missing values and outliers.
- Normalize features to ensure consistent input across models.
- Segment images to extract soil regions for texture analysis.
- Split the data into training (80%) and testing (20%) sets.

C] Soil Classification:

- Implement machine learning models for classification:
    - Decision Trees (DT): For hierarchical soil classification.
    - Support Vector Machine (SVM): For complex pattern detection.
    - K-Nearest Neighbor (KNN): For nutrient-based soil classification.
- Store model predictions and accuracy metrics for future reference.

D] Recommendation system:

- Generate personalized crop suggestions based on soil classification.
- Provide fertilizer recommendations to optimize soil fertility.
- Suggest corrective actions for pH adjustments or moisture management.

E] User Interface:

- Develop a web and mobile interface using React and Tailwind CSS.
- Features include soil image upload, manual data input, and real-time classification results.
- Provide easy-to-understand visualizations (charts and graphs) of soil health.

F] Reporting and Feedback:

- Generate detailed reports summarizing soil analysis and recommendations.
- Include visualizations such as confusion matrices and performance metrics.
- Collect user feedback to continuously improve the system.

G] Backend and Database:

- Use Node.js and Express for the backend.
- Store soil data, classification results, and user inputs in MongoDB or MySQL.
- Provide APIs for data retrieval and interaction between components

H] Deployment and Monitoring:

- Deploy the platform on AWS EC2 instances with Docker containers.
- Use CloudWatch or New Relic to monitor system health and performance.
- Implement automatic scaling to handle peak loads.

MACHINE LEARNING TECHNIQUES USED

1. Decision Trees (DT) and Random Forests (RF)
    - Overview: Decision Trees create a hierarchical structure for decision-making based on input features like soil pH, moisture levels, and nutrient content.
    - Performance: Studies show that DT provides high accuracy in soil classification and easy interpretability.
        - Maximum Accuracy: 98.4% using SK-Learn tool for soil classification.
        - Random Forest, an ensemble of multiple decision trees, achieves 85% accuracy, demonstrating robustness in soil data analysis.
2. Support Vector Machines (SVM):
    - Application: SVM works by finding optimal hyperplanes to separate soil classes, which is useful for datasets with complex patterns.

- Performance: SVM shows superior performance in classifying soil contamination levels, particularly for heavy metal pollution, as observed in a case study in Arak, Iran.
- Limitations: Computationally intensive, making it challenging for mobile applications that require real-time processing.

3. K-Nearest Neighbor (KNN):
    - Usage: KNN is used for soil nutrient detection and soil color classification through image processing techniques.
    - Accuracy: The KNN-based soil nutrient detection app for citrus farming in Banyuwangi achieved 89.6% accuracy, showing the model's potential in Android based precision agriculture applications.
4. Naïve Bayes Algorithm:
    - Application: Used for agricultural land classification to identify relationships between soil features.
    - Performance: Achieved 69.5% accuracy using SK-Learn, highlighting its usefulness for low-complexity soil classification tasks.
    - Limitations: Naïve Bayes assumes feature independence, which may limit its effectiveness in complex datasets.

DEEP LEARNING AND IMAGE PROCESSING TECHNIQUES USED

1. Convolutional Neural Networks (CNN):
    - Application: CNN models are used for soil classification and disease detection by analyzing soil images.
    - Advantages: CNNs perform automatic feature extraction without manual intervention, improving performance with spatially varying soil data.
    - Challenges: Real-time classification on mobile platforms faces computational limitations, necessitating model optimization for a trade-off between speed and accuracy.
2. Image Processing Techniques:
    - Objective: Enhance the quality of soil images through segmentation and feature extraction before analysis.
    - Use Case: Soil color detection using HSV (Hue, Saturation, Value) models and Munsell soil color charts improves the accuracy of identifying soil properties like mineral content.

TOOLS AND PLATFORMS USED FOR MODEL DEVELOPMENT

- SK-Learn:
    - Performance:Achieved the best accuracy across multiple ML algorithms for soil classification, making it a preferred tool for ML practitioners.
    - Key Models: Decision Tree, Random Forest, Naïve Bayes, and KNN were implemented using SK-Learn, with Decision Trees achieving up to 98.4% accuracy.
- WEKA and KNIME:
    - Performance Comparison:
        - WEKA achieved 73.06% accuracy for Random Forest and 68.14% for Naïve Bayes, making it slightly less effective compared to SK-Learn.
        - KNIME showed comparable performance, with 73.07% accuracy for Decision Trees. KNIME's visual workflow capability makes it user-friendly for non-coders.
- Mobile Microcontroller Platforms:
    - Case Study: Development of an Android-based soil nutrient detection app utilizing microcontrollers.
    - Features: Real-time soil analysis with automatic data input through Bluetooth connectivity, providing farmers with instant soil reports.

ETHICAL AND PRACTICAL CONSIDERATIONS

As generative AI systems become more advanced in managing and utilizing context, several ethical and practical concerns must be addressed. These concerns primarily revolve around privacy, transparency, bias, and accountability in how context is handled, stored, and used.

Privacy and Data Security

One of the most pressing ethical concerns in context management is privacy. Generative AI systems often rely on user data to maintain context, whether through past interactions or user-specific information. This raises important questions about how much context should be retained, who has access to it, and how long it is stored.

Data Retention and User Consent: AI systems must ensure that they comply with privacy regulations like GDPR (General Data Protection Regulation) and CCPA (California Consumer Privacy Act). Users should have control over what context the AI retains, and they should be able to delete or modify that data at any time. The ethical handling of

context requires that AI systems respect user consent and minimize data collection to only what's necessary for the task.
Sensitive Information: There is also the risk that generative AI models could inadvertently retain or misuse sensitive personal data (e.g., medical history, financial details, or personal preferences). This requires careful design of data retention policies and encryption techniques to prevent unauthorized access or misuse of private information.

Transparency and Explainability

As AI systems become more capable of managing complex context, the need for transparency in how these systems work grows. Users should be able to understand how AI systems make decisions, especially when they rely on past interactions or personal data. Without clear explanations, users may be unable to trust the system's outputs or understand why certain context was used.

Explainable AI: Developing methods to explain context handling in AI is critical, especially when models are making decisions based on long-term context that may not be immediately visible to users. For instance, in a conversation, users should be able to ask why the AI responded in a particular way or reference previous context if needed. This ensures that the AI's actions are understandable and justifiable.
Black-box Models: Many current AI systems operate as "black boxes," where the inner workings are opaque to the user. This is especially problematic when AI systems are expected to manage sensitive context or engage in high-stakes decision- making (e.g., in healthcare or legal contexts). Greater transparency in how context is managed and decisions are made will be essential to build user trust and ensure ethical practices.

Bias and Fairness

Context management also has implications for bias and fairness in AI systems. If context is improperly handled or certain contextual factors are overlooked, it can reinforce existing biases or produce discriminatory outputs.

Bias in Context Selection: The way an AI model selects and prioritizes context can introduce bias. For example, if an AI system focuses primarily on certain types of context (such as recent interactions or demographic characteristics), it could lead to biased responses that don't fairly represent the user's needs or preferences. This could be particularly problematic in applications like hiring assistants, medical diagnosis, or legal advice, where fairness and impartiality are critical.
Bias in Long-Term Context: Long-term context accumulation might also inadvertently reflect the biases of past interactions or societal inequalities. AI systems must be carefully designed to avoid amplifying historical biases, whether through biased training data or skewed memory systems. Techniques like debiasing, fairness auditing, and diverse training data can help mitigate these risks.

Accountability and Misuse

As AI systems become more adept at managing context, questions of accountability arise. If an AI system makes an error or generates harmful outputs based on a misinterpretation of context, who is responsible?

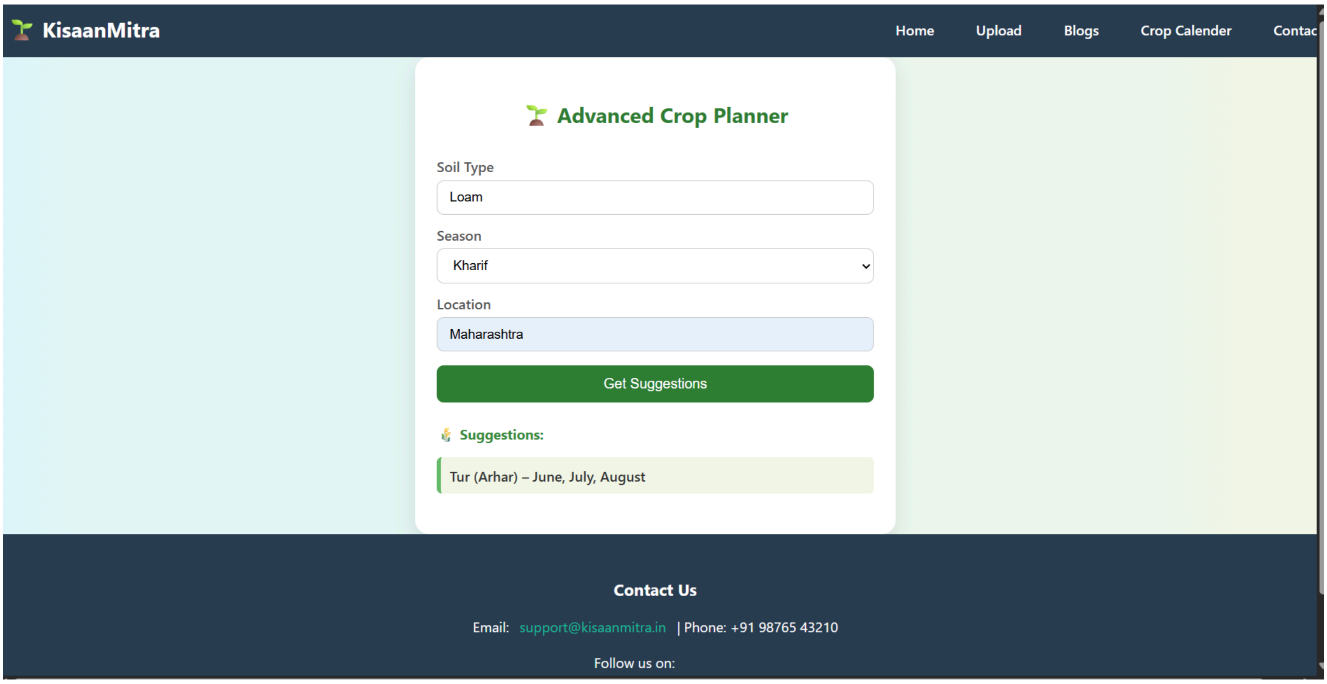
Responsibility for Outputs: It is important to establish clear lines of accountability when AI systems generate harmful, misleading, or unethical responses due to context mismanagement. Whether the responsibility lies with the developers, the users, or the AI itself is a topic of ongoing debate. Clear guidelines for liability, especially in high-risk applications, will help prevent misuse and ensure that AI systems operate within ethical boundaries
Potential for Manipulation: AI systems capable of maintaining and adapting context could be vulnerable to manipulation. For example, users might exploit context handling to elicit biased or harmful responses. Generative models used in social media, advertising, or content moderation must be protected from malicious manipulation that could undermine the integrity of the system.

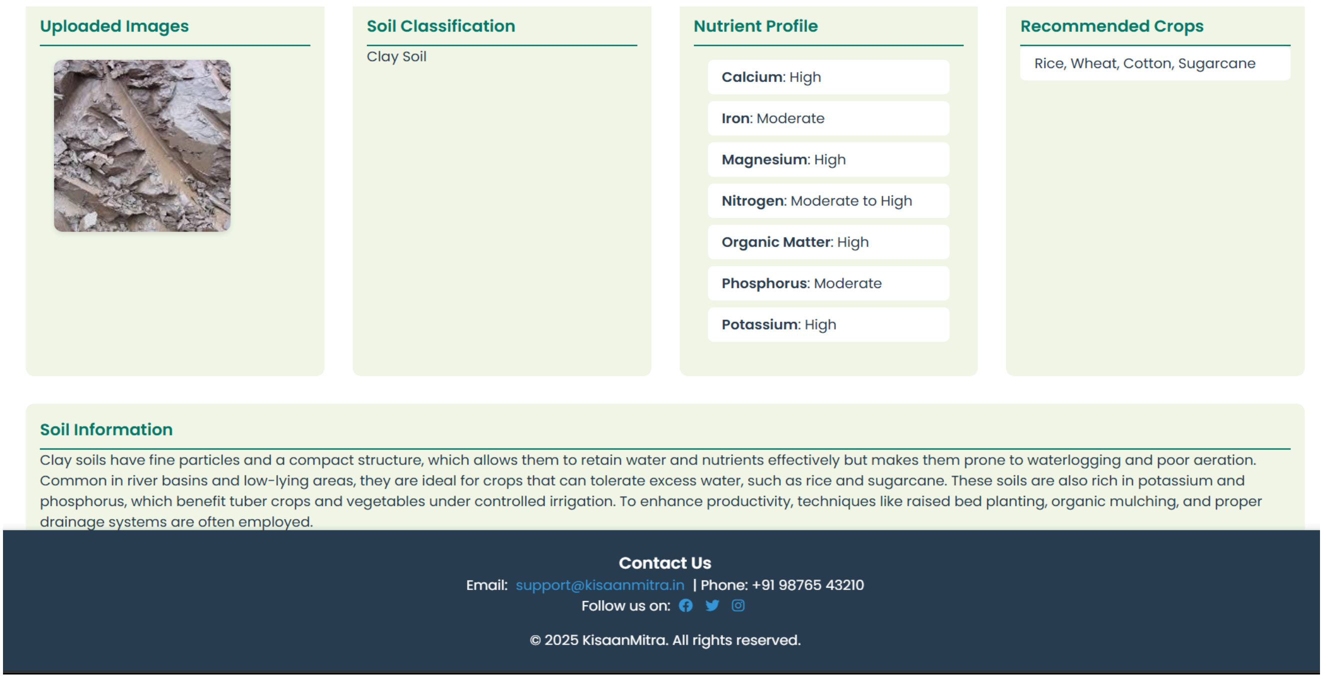
Human-AI Interaction and Dependency

Finally, there are practical concerns around the potential over-reliance on AI for managing context, especially in personal or professional settings.

Over-Dependence on AI: As AI systems become better at managing context, users may increasingly delegate more tasks to AI systems, including remembering personal details or managing complex workflows. While this can be convenient, it raises questions about the impact on human cognition and decision-making. Over-reliance on AI could lead to a loss of autonomy or critical thinking skills, particularly in younger generations or individuals who frequently interact with AI.
User Autonomy and Control: It is essential that users retain control over the context managed by AI systems. For example, users should have the ability to correct, delete, or modify context as necessary. Striking the right balance between AI assistance and user autonomy will be crucial to maintaining a healthy human-AI relationship.

OUTPUTS





FUTURE SCOPE

- Integrate IoT sensors for real-time monitoring of soil parameters like moisture, temperature, and pH, allowing for dynamic data analysis and timely crop recommendations.
- Add satellite imaging and weather forecast integration to enhance predictions of seasonal crop suitability and mitigate risks like droughts or floods.
- Expand the soil dataset to include diverse soil types and regions (e.g., saline, volcanic, arid), improving classification accuracy and platform usability across geographical areas.
- Implement advanced machine learning techniques like ensemble models, CNNs, or LSTMs to handle complex datasets and increase classification precision.
- Introduce personalized crop recommendations using adaptive learning based on user feedback and historical outcomes for continuous improvement.
- Enable offline functionality in the mobile app using local storage and sync features to support users in rural or low-connectivity regions.
- Add crop yield forecasting features using regression models to predict future yields based on real-time soil and environmental data.
- Integrate an interactive dashboard for farmers to track soil health trends, visualize classification history, and get actionable insights.

- Collaborate with agricultural organizations and government agencies to ensure wide-scale deployment and alignment with national soil health initiatives.
- Provide educational content and farming best practices based on soil type and crops, enhancing farmer knowledge and increasing productivity.

CONCLUSION

The "Testing and Classification of Soil for Plant Cultivation" project successfully developed a machine learning-based platform to analyze soil properties and provide personalized crop recommendations. By integrating advanced algorithms such as Decision Trees, SVM, and KNN, the system offers accurate soil classification and actionable insights to empower farmers and promote precision agriculture. Throughout the development process, the project addressed key challenges, including data inconsistencies, model optimization, and system integration, ensuring the platform is reliable, scalable, and user-friendly. The use of Agile methodology enabled continuous improvement based on user feedback, resulting in a robust system that meets the practical needs of farmers. The platform's web and mobile interfaces ensure accessibility, even in rural areas, contributing to the digital transformation of agriculture. The deployment on AWS with cloud-based infrastructure ensures scalability to accommodate future expansions, such as additional soil parameters or crop recommendations. The inclusion of feedback mechanisms allows the platform to evolve, adapting to changing agricultural needs and new research findings.

REFERENCES

[1]. "Generative AI: A Review on Models and Applications" - K. S. Kaswan, J. S. Dhatterwal, K. Malik and A. Baliyan, 2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI), Greater Noida, India, 2023, pp. 699 - 704, doi: 10.1109/ICCSAI59793.2023.10421601

[2]. "Detection & Management of Concept Drift" - L. -o. Mak and P. Krause, 2006 International Conference on Machine Learning and Cybernetics, Dalian, China, 2006, pp. 3486-3491, doi: 10.1109/ICMLC.2006.258538

[3]. "Comparison of Channel Attention Mechanisms and Graph Attention Mechanisms Applied in Multi-Robot Path Planning Based on Graph Neural Networks" - Y. Xie, Y. Wang and S. Xu, 2024 5th International Conference on Intelligent Computing and Human-
Computer Interaction (ICHCI), Nanchang, China, 2024, pp. 558-561, doi: 10.1109/ICHCI63580.2024.10807947

[4]. "Advancing Natural Language Processing: Beyond Embeddings" - S. Iskandarova, U. Kuziyev, D. Ashurov and D. Rakhmatullayeva, 2024 International Conference on IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2024, pp. 792-796, doi: 10.1109/ICICAT62666.2024.10923033

[5]. "Getting Started with Google BERT: Build and train state-of-the-art natural language processing models using BERT" - Sudharsan Ravichandiran, Packt Publishing, 2021

[6]. "Comprehensive Review of Benefits from the Use of Sparse Updates Techniques in Reinforcement Learning: Experimental Simulations in Complex Action Space Environments" - M. Kaloev and G. Krastev, 2023 4th International Conference on Communications, Information, Electronic and Energy Systems (CIEES), Plovdiv, Bulgaria, 2023, pp. 1-7, doi: 10.1109/CIEES58940.2023.1037883

[7]. "A Cooperative Learning Method for Multi-Agent System with Different Input Resolutions" - F. Uwano, 2021 4th International Symposium on Agents, Multi-Agent Systems and Robotics (ISAMSR), Batu Pahat, Malaysia, 2021, pp. 84-90, doi: 10.1109/ISAMSR53229.2021.9567835

[8]. "Artificial Trust as a Tool in Human-AI Teams" - C. C. Jorge, M. L. Tielman and C. M. Jonker, 2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Sapporo, Japan, 2022, pp. 1155-1157, doi: 10.1109/HRI53351.2022.9889652