

Data-Driven Decisions: A Review of Machine Learning Models, Soil Sensors and Crop Recommendation Systems

Gujjeti Nagaraju¹, Nayakoti Rishika², Paruchuri Pranav³, Pothala Mukesh⁴ and Talluri Anthyush⁵

¹⁻⁵Department of Computer Science & Engineering (AIML& IoT), Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering & Technology, Hyderabad, Telanagana - 500090, India.

Email: nagaraju_g@vnrvjiet.in, rishika.nayakoti3781@gmail.com, pranav.paruchuri6@gmail.com, mukeshpothala189@gmail.com, tallurianthyush02@gmail.com

Abstract— The integration of advanced technologies in revolutionary agriculture transformed decision-making processes through the implementation of data-driven techniques. This study presents a thorough examination of the connection of machine learning models, soil sensors, and crop recommendation systems, with the purpose of demonstrating their combined impact on agricultural practices. The research delves into the crucial role of machine learning algorithms in agricultural data collecting, highlighting their ability to analyze complex datasets and derive helpful insights for informed decision-making. Furthermore, this paper explores the ever-evolving environment of soil-sensing technologies, emphasizing their significance to precision agriculture by providing real-time, high-resolution data on soil conditions. In addition, crop recommendation integration into systems is investigated, with a focus on its effectiveness in improving crop choices based on data-driven insights acquired from historical and real-time data. The combination of these technical aspects has been identified as a turning point in the development of sustainable agriculture. Embracing these advances in technology not only makes agriculture more efficient, but also accounts for a more resilient and environmentally conscientious future, demonstrating the transformative potential of data-driven techniques in influencing the trajectory of global food supply.

Index Terms— Soil Monitoring, Mobile application, Sensors, Crop Recommendation, Image processing, Machine learning.

I. INTRODUCTION

The escalating global population presents an imminent challenge to existing agricultural practices, necessitating a profound transformation to ensure sustained food security. Traditional farming methods, proven to be inherently unsustainable, demand a paradigm shift towards a data-driven agricultural revolution. This review endeavors to explore the pivotal role of cutting-edge technologies, specifically soil sensors, image processing, and mobile applications, in fostering informed decision-making and optimizing resource utilization for the advancement of agriculture.

One of the cornerstones of this agricultural revolution is the deployment of soil sensors. These devices exhibit a

capability to monitor critical parameters such as NPK (Nitrogen, Phosphorus, and Potassium) levels, organic content, and moisture in real-time. The acquisition of this real-time data empowers farmers with the precision needed for tailored crop management and optimized fertilizer application. In this review, we delve into the extensive research assessing soil sensor accuracy and investigate the synergistic potential of combining soil sensor data with advanced image processing techniques.

Complementary to soil sensors, image processing emerges as a powerful tool to glean insights into soil characteristics. Through digital analysis, image processing unlocks information on pH levels, nitrogen content, and organic matter composition. This review critically evaluates specific studies to discern the strengths and weaknesses of image processing in achieving high data accuracy. Furthermore, we explore the burgeoning field of machine learning, which leverages image-derived soil data to predict crucial factors such as nutrient levels, crop health, and potential yields. This integration underscores the transformative impact of data-driven decision-making in modern agriculture. In parallel, the ubiquity of mobile technology has birthed a suite of applications designed to democratize data dissemination within the agricultural community. These mobile applications serve as conduits for real-time information, equipping farmers with actionable insights for optimizing resource utilization and maximizing yield. This review meticulously examines the multifaceted functionalities of these applications, illustrating how they contribute to a more connected and informed farming community.

As we conclude this exploration, attention is drawn to emerging trends poised to further propel the precision agriculture frontier. Weather API integration and the incorporation of drone technology for aerial data collection represent promising avenues for advancing our understanding of agricultural ecosystems. By embracing these innovations, the agricultural landscape stands to gain unprecedented levels of insight, paving the way for a sustainable and resilient future.

II. LITERATURE REVIEW

In recent years, the agriculture sector has endured a transformative shift towards data-driven decision-making, fuelled by breakthroughs in technology. This study focuses on the usage of soil sensors, machine learning models and the development of crop recommendation systems specifically.

Zhang et al. [1] used hyperspectral image analysis to estimate soil pH with an excellent accuracy of 93.7% in their 2021 study. This approach catches a wide variety of wavelengths, allowing for thorough spectral data collecting. The high accuracy shows that the spectral data acquired efficiently corresponded with fluctuations in soil pH, demonstrating the promise of hyperspectral imaging for precise and non-invasive soil pH assessment. This technology has the potential to improve agricultural soil management methods by delivering quick and precise assessments of soil conditions, resulting in better informed decision-making.

Emrullah et al. [2] examines the application of machine learning approaches to the soil surface humidity (SSH) determination over moderately vegetated landscapes. The work extracts polarimetric SAR features from Radarsat-2 data, which gives high-resolution views of the Earth's surface. To predict SSH, these features are fed into an ELM-based regression model. The dataset used in the study is made up of 156 SSH measurements obtained from two different Dicle University terrains. The experimental results show that the suggested ELM-based model can predict SSH accurately with a RMSE of 2.19%.

Khanal et al. [3] examines how remote sensing data and machine learning techniques can be utilised to estimate soil properties and maize production. The authors collected experimental data on five soil parameters, including soil organic matter (SOM), cation exchange capacity (CEC), Mg, K, and pH, as well as yield monitor-based maize yield data, from seven Ohio fields. The fields also produced multispectral aerial photos and digital elevation model (DEM) data. To forecast soil parameters and maize yield, the authors employed machine learning techniques such as random forest, SVM with radial and linear kernels, cubist, and others. They discovered that machine learning methods outperformed linear regression algorithms in most circumstances, with greater R² and lower RMSE. Random Forest regularly beat other models, providing higher accuracy for maize yield where R² is 0.53 and RMSE is 0.97. Topographic parameters were less important than soil and vegetation indices based on soil imaging in depicting on-field variability in maize yields and soil attributes.

Trontelj et al. [4] uses spectroscopic approaches to forecast soil nutrients using machine learning (ML) algorithms. The performance of common ML approaches like Random Forest (RF), Decision Tree (DT), Artificial Neural Network (ANN), etc. were compared. Their research revealed that, surprisingly, the most frequently used and intricate prediction strategy wasn't necessarily the most accurate. The impact of the chosen category on nutritional characterisation was explored, with a multi-component technique indicating superior prediction. Conversely, the accuracy of predicting soil parameters using a single component was lower. Furthermore, the

impact of category levels was not as great as anticipated when using a 3 different levels classification for each nutrient. This can be applied to develop a nutritional characterisation technique that is more accurate. The soil from the adjacent farm produced a better forecast than soils obtained from different parts of Slovenia when compared to soil from a nearby farm with a similar texture. Ultimately, 5, 10, 20, and 50 first principal components were used to assess the effect of principle component analysis; the machine learning performance was shown to be improved with the use of the 50 first principal components.

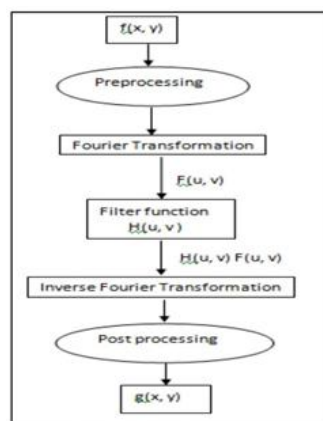


Figure 1. Image Enhancement [5]

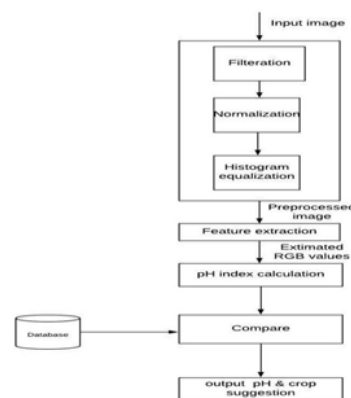


Figure 2. General Framework [8]

In their paper [6], Bulan et al. investigate the application of machine learning algorithms and Vis-NIR spectroscopy to soil nutrient forecasting in Indonesia's Aceh Province. After air drying and powdering the soil samples (0–60 cm) from ten different Aceh Province locations, the scientists examined their Vis-NIR spectra (400–2150 nm). The Kjeldahl method was used to determine the soil's nitrogen and phosphorus content, and the Bray method was used to evaluate the soil's total potassium content. Using three machine learning algorithms—SVM regression (SVR), partial least squares (PLS-ANN), and gradient-boosted tree regression (GBRT)—their Vis-NIR spectra were utilised to create calibration models that could predict TN, TP, and TK content. The results showed that all machine learning algorithms that were evaluated in a chemometric method generated good models for quantitative evaluations of TN, TP, and TK content. When considering the complete spectrum, the accuracy of the SVR method was similar to that of PLS-ANN models in general. The ANN algorithm (PLS-ANN) beat the SVR algorithm, albeit, depending on the processing technique. The most accurate simulations were produced by the GBRT approach, with RPDs for TN, TP, and TK content of 2.64, 3.93, and 2.38, respectively. The findings demonstrate that a dependable technique for determining the concentrations of TN, TP, and TK in soil is to use Vis-NIR in combination with a machine learning system.

Puno et al. [7] present a unique method for assessing soil pH and nutrient levels that makes use of image processing and an artificial neural network (ANN). Under controlled lighting, this system took photos of soil samples and retrieved attributes such as HSV and Lab values. The trained ANN correctly identified samples based on nutrient and pH levels. While promising, the system has some drawbacks, including a small dataset (670 photos), reliance on controlled lighting, manual color interpretation, concentration on six specific nutrients, and testing on a small sample size. Extending the dataset, integrating environmental variables, and investigating alternate picture attributes could improve the system's generalizability and resilience.

Sharma et al. [5] presents an innovative system for soil categorization and characterisation based on easily available smartphone or tablet technologies. The system supports farmers in finding viable crops for their land by employing image processing techniques such as HSV colour space analysis, enhancement algorithms, and support vector machine (SVM) classifiers. This real-time soil study equipment offers huge promise for rural areas due to its portability and accessibility. However, there are limitations due to the restricted training data concentrated in Udaipur, Rajasthan, which may hamper accuracy for diverse soil types or lighting circumstances. More research into expanding the dataset and integrating environmental variables would improve the system's generalizability and durability. In spite of these constraints, the possibility for smartphone-based soil analysis for customized crop recommendations is an encouraging move towards empowering farmers with easily available crop management tools.

Vinay Kumar et al. [8] describes a method for estimating soil pH that makes use of digital image processing techniques. The researchers collected 50 soil samples from the Nathnagar block of India's Bhagalpur district and used a pH metre to evaluate their pH. After photographing the soil samples digitally, the red, green, and blue (RGB) values of each pixel in the images were extracted using a colour segmentation method. The authors identified a relationship between soil sample RGB values and pH levels. They developed a model to determine soil pH based on the RGB values of the soil samples. The model predicted soil pH with 92% accuracy. The researchers also discovered that the colour of the soil sample accurately predicted its pH. Soil samples with a dark brown tint had a higher pH than soil samples with a pale yellowish or greenish colour.

M Arun et al. [9] deals with the incorporation of image processing to determine the physical and chemical qualities of soil. The authors propose a method for analysing soil samples that uses fractal dimensions and RGB model analysis. The box counting method is used for calculating the fractal dimension, and RGB model analysis involves extracting the pixel values of each plane (R, G, and B). The soil pH index is calculated as $\text{Soil pH index} = \text{Red}/\text{Green}/\text{Blue}$. The authors put their approach to the test on a soil sample from Watrap in the Virudhunagar district and discover that it can reliably identify the soil's moisture content, plastic limit, shrinkage limit, coefficient of curvature, specific gravity, coefficient of uniformity, and field density. The authors further note that their method accurately determines soil's pH.

III. COMPARATIVE ANALYSIS

A proposed system [10] measures soil moisture and NPK levels with soil moisture sensors and colour sensors, respectively. Data from the sensors is transmitted to a cloud database and shown via a mobile application. This technique can help farmers conserve resources while increasing cultivation by producing a large quantity of high-quality crops. Data acquired from sensors can also help farmers make better decisions by pinpointing the best times for planting and harvesting. Furthermore, accurate tracking of production rates allows estimate future yield and run the system more efficiently.

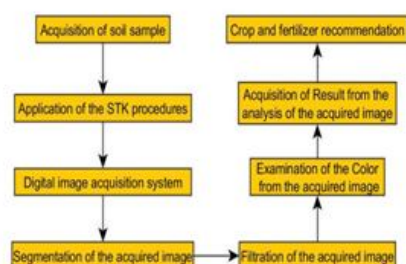
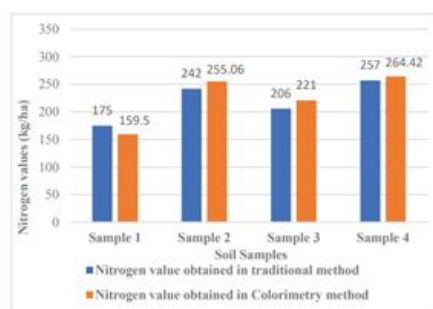


Figure 3. Comparison of Nitrogen values for Traditional and Colorimetry Method [10] Figure 4. System's Model Flowchart [12]

Using multi-class labels, a different proposed system [11] describes an intelligent way of evaluating soil for crop cultivation. Four processes make up the technique, which helps farmers boost crop production: (i) collecting soil data; (ii) preprocessing; (iii) extracting soil features; and (iv) implementing classifiers using various machine learning methods. Naive Bayes, KNN, random forest, decision tree, SVM, and artificial neural network are some of the machine learning classifiers that are used to classify soil features. We later construct an ensemble of the earlier classifiers as well. The ensemble approach method (decision tree + random forest + multi-layer perceptron) yields the best accuracy of 98.03% over the experiments, while the decision tree yields the second-highest accuracy of 87.09%.

A study [12] developed a mobile application that assesses the pH and macronutrient (potassium, phosphate, and nitrogen) contents of soil using image processing. The application uses the OpenCV library for image processing, while Eclipse is used as the Android programming tool. The outcomes showed that the smartphone app can precisely monitor the pH and NPK levels of the soil and can also recommend crops and fertilisers. The t-test for pH level and the Wilcoxon signed ranks test for NPK levels were used to confirm this. Farmers can make time and financial savings by using the smartphone app. In the paper [13] authors have utilized Convolution Neural Networks as the preferred method for image analysis, specifically in face recognition. They achieved superior accuracy as compared to the other techniques.

IV. PROPOSED SYSTEM

Imagine empowering farmers with the ability to unlock the contents, nutrients and materials hidden within their soil, simply by snapping a picture with their phone. This transformative vision becomes reality with the proposed system, a sophisticated blend of mobile technology and cutting-edge machine learning.

At the heart of the system lies a user-friendly mobile application, designed to be accessible and intuitive for farmers of all technical backgrounds. With a few simple clicks, users can capture high-resolution images of their soil, transforming their smartphones into powerful analysis tools. But the most important and viable thing truly happens behind the scenes, within the robust backend ensemble of machine learning models. Think of this ensemble as a team of machine learning models and deep learning models based on the selection, each specializing in different aspects of soil assessment. This diverse team might include Convolutional Neural Networks, adept at deciphering intricate patterns within the images, or Support Vector Machines, skilled at identifying subtle variations in nutrient levels. By combining the strengths of these algorithms, the system achieves a high level of accuracy and comprehensiveness.

The size or in other words the hyperparameter of this ensemble isn't fixed – it's carefully optimized through experimentation. Just like selecting the perfect team size for a project, the system meticulously evaluates different configurations (ranging from 2 to 10 models) to find the perfect point or hyperparameter that balances accuracy with efficiency. The process continues with each individual model within the ensemble undergoes rigorous training, fine-tuned through a process called hyperparameter tuning. Imagine this like equipping each team member with specialized tools and knowledge; it empowers them to excel in their specific roles. Finally, the system selects the goal is to select a model with the highest "efficiency value," a metric that considers both accuracy and speed. This ensures that farmers receive not only highly reliable information but also get it in a timely manner, crucial for making informed decisions during critical growth stages.

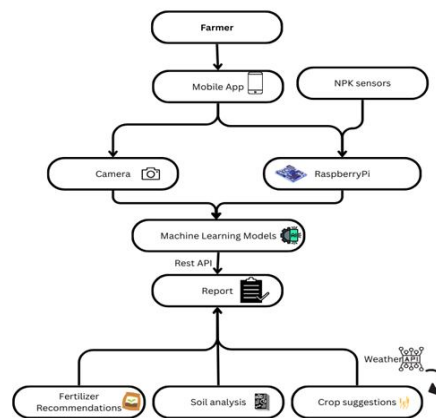


Figure 5. Proposed System

In essence, this system transcends the limitations of previous approaches. It offers not just a glimpse into the soil's health, but a comprehensive understanding, empowering farmers to become true masters of their land. From optimizing irrigation practices to tailoring fertilizer application, the information gleaned from this system has the potential to revolutionize agricultural practices, leading to increased yields, reduced costs, and a more sustainable future for all.

V. FUTURE SCOPE AND DISCUSSIONS

A combination of machine learning algorithms, soil sensors, and crop recommendation systems has the potential to change agriculture. These advancements, by providing farmers with more accurate and up-to-date information, can help to increase crop yields, reduce the cost of inputs, and make agriculture more sustainable. This study examined the present state of the art in this field and suggested many promising areas for future research.

A. Developing more accurate and efficient machine learning models to predict soil conditions and agricultural yields. This could be accomplished by the use of more complicated models, such as deep learning models, or through the training of models on larger and more diverse datasets.

B. The combination of soil sensors. The combination of soil sensors with additional sources of data, such as meteorological data and satellite photography. This might provide farmers with a more complete picture of their farms, allowing them to make more informed crop management decisions.

C. Designing and developing the personalised crop recommendation systems that take into consideration the unique requirements of farms. Machine learning could be used to find the optimum crops for each farm depending on soil type, climate, and other characteristics.

D. The utilisation of mobile applications to provide farmers with real-time agricultural advice. This could help farmers make better crop management decisions and increase yields.

VI. CONCLUSION

Ultimately, the incorporation of machine learning models, soil sensors, and crop recommendation systems has altered agricultural practices by enabling data-driven decision-making for better resource utilisation, optimum crop yields, and sustainable farming methods. Soil sensors provide real-time information on soil conditions to farmers, allowing them to make informed decisions about irrigation, fertilisation, and insect management. Image processing techniques extract useful information from soil samples, enabling for the estimation of nutrient levels, pH, and other soil properties. Massive amounts of agricultural data are analysed by machine learning algorithms in order to forecast crop yields, predict probable pest infestations, and recommend the best crop choices. Precision agriculture, a paradigm shift in crop management that prioritises data-driven precision, has been paved by the convergence of these technologies. By placing actionable insights at farmers' fingertips, mobile applications have further democratised access to agricultural data. The precision agriculture revolution is poised to reach new heights as emerging trends such as weather API integration and drone-based aerial data collection gain acceptance, ensuring the agricultural industry's long-term viability and productivity.

REFERENCES

- [1] X. Meng, Y. Bao, Q. Ye, H. Liu, X. Zhang, H. Tang, X. Zhang, Soil organic matter prediction model with satellite hyperspectral image based on optimized denoising method, *Remote Sensing* 13 (12) (2021) 2273
- [2] E. ACAR, M. S. OZERDEM, B. B. USTUNDAG, Machine learning based regression model for prediction of soil surface humidity over moderately vegetated fields, in: 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), 2019, pp. 1–4. doi:10.1109/Agro-Geoinformatics.2019.8820461.
- [3] S. Khanal, J. Fulton, A. Klopfenstein, N. Douridas, S. Shearer, Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield, *Computers and electronics in agriculture* 153 (2018) 213–225.
- [4] J. Trontelj ml, O. Chambers, Machine learning strategy for soil nutrients prediction using spectroscopic method, *Sensors* 21 (12) (2021) 4208.
- [5] H. K. Sharma, S. Kumar, Soil classification & characterization using image processing, in: 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), IEEE, 2018, pp. 885–890.
- [6] R. Bulan, A. Sitorus, et al., Vis-nir spectra combined with machine learning for predicting soil nutrients in cropland from aceh province, indonesia, *Case Studies in Chemical and Environmental Engineering* 6 (2022) 100268.
- [7] J. C. Puno, E. Sybingco, E. Dadios, I. Valenzuela, J. Cuello, Determination of soil nutrients and ph level using image processing and artificial neural network, in: 2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), IEEE, 2017, pp. 1–6.
- [8] V. Kumar, B. K. Vimal, R. Kumar, M. Kumar, Determination of soil ph by using digital image processing technique, *Journal of Applied and Natural Science* 6 (1) (2014) 14–18.
- [9] C. ManikandaBabu, M. A. Pandian, Determination of physical and chemical characteristics of soil using digital image processing, *DIGITAL IMAGE PROCESSING* 1 (2016) 2.
- [10] R. Madhumathi, T. Arumuganathan, R. Shruthi, R. S. Iyer, Soil nutrient analysis using the colorimetry method, in: 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), IEEE, 2020, pp. 252–256.
- [11] M. S. J. Jamil, A. Anika, M. S. H. Mukta, M. N. Huda, Ensemble of classifiers-based soil evaluation for crop cultivation, in: 2022 IEEE Region 10 Symposium (TENSYP), IEEE, 2022, pp. 1–5.
- [12] J. J. F. Montañez, Level identification of soil ph and macronutrients using mobile application through image processing, in: 2021 2nd International Conference for Emerging Technology (INCET), IEEE, 2021, pp.
- [13] Kalyan, L.P., Nagaraju, G., Reddy, R.K., Mulkalapalli, S. (2022). Identification of Face Mask Detection Using Convolutional Neural Networks. In: Sanyal, G., Travieso-González, C.M., Awasthi, S., Pinto, C.M., Purushothama, B.R. (eds) *International Conference on Artificial Intelligence and Sustainable Engineering. Lecture Notes in Electrical Engineering*, vol 837. Springer, Singapore. https://doi.org/10.1007/978-981-16-8546-0_25