Literature Review on Agricultural Yield Prediction and Soil Data Analysis Using Machine Learning

Mrs. Ashvini Ganapati Patil[1], Mrs. Sampada S. Gulwani[2]
[1]Research Student, Bharati Vidyapeeth (Deemed to be University), Institute of Management, Kolhapur, India
ashvini.patil18@gmail.com
[2]Bharati Vidyapeeth (Deemed to be University), Institute of Management, Kolhapur, India
sampada.gulwani@bharatividyapeeth.edu

Abstract -

This paper presents a comprehensive review of current research on agricultural yield prediction and soil data analysis, emphasizing the application of machine learning techniques. It discusses how soil characteristics, weather variables, and fertilizer management influence crop productivity. The review explores various machine learning models, including decision trees, random forests, support vector machines, and deep learning architectures, highlighting their effectiveness in improving yield forecasts and optimizing resource use. Additionally, the integration of Internet of Things (IoT) technologies with ML models for real-time monitoring and precision agriculture is examined. Key research gaps identified include the need for integrated fertilizer recommendation systems, real-time IoT data incorporation, regional specificity in models, and accessible farmer-centric decision support tools. Addressing these gaps offers significant potential to enhance agricultural sustainability, productivity, and economic viability.

Keywords: Soil Fertility, Machine Learning, Classification, Random Forest, SVM, Agriculture Data Mining.

I. INTRODUCTION

India, being primarily agrarian, relies heavily on agriculture, constituting 15.96% of its GDP. Numerous factors, including soil composition, weather patterns, precipitation, and the application of fertilizers and pesticides, significantly influence crop yield. Soil, comprised of finely divided rocks, minerals, and organic matter, stands as a pivotal element affecting agricultural productivity, with each soil type possessing distinct characteristics. The soil's nutrient composition directly impacts productivity, where deficiencies can markedly reduce output, necessitating comprehensive soil analysis.

Water, sourced from bore wells, open wells, irrigation systems, rivers, and canals, serves as a critical resource in agriculture, with varying pH levels based on the source and seasonal fluctuations, thereby requiring water analysis. Fertilizers play a pivotal role in augmenting natural soil fertility and fostering plant growth and development. Addressing nutrient deficiencies in the soil, fertilizers—categorized as organic, inorganic, bio, and soluble—come from living or artificially synthesized materials. Organic fertilizers enhance soil texture, moisture retention, and erosion resistance, providing plants with usable nitrogen that supports growth. Conversely, inorganic fertilizers can increase soil acidity, degrade physical properties, necessitating an integrated fertilizer approach.

ISSN PRINT 2319 1775 Online 2320 7876
 Journal UGC CARE Listed (Group-I) Volume 11, Issue , 10 2022

The lack of awareness among farmers about soil and water chemical properties often results in the indiscriminate and manual application of fertilizers, sometimes based on anecdotal advice. Such practices can adversely affect plant life, diminish yield, and compromise soil health due to the excessive or insufficient use of synthetic fertilizers. Therefore, a scientific approach is imperative in recommending appropriate fertilizers, incorporating soil, water, and leaf analyses. Safeguarding soil health and optimizing yield requires a balanced and controlled application of fertilizers, as excessive synthetic fertilizer use contributes to soil erosion.

In conclusion, a strategic and scientific fertilizer recommendation process, encompassing thorough soil, water, and leaf analyses, is vital for bolstering agricultural yield and nurturing soil vitality. Responsible and limited fertilizer application is crucial in mitigating soil erosion and fostering sustainable agricultural practices. Agricultural productivity stands as a cornerstone of global food security, economic stability, and sustainable development.

II. AGRICULTURAL YIELD AND SOIL DATA

Agricultural yield refers to the measurement of the output of agricultural crops per unit area, which is a critical factor in food security and economic stability. Various elements influence agricultural yield, with soil data playing a crucial role. Soil data encompasses essential characteristics such as nutrient content, pH levels, moisture levels, organic matter, and texture, all of which determine the suitability of soil for specific crops. The composition of soil affects root growth, water retention, and microbial activity, making soil analysis indispensable for effective farming.

Understanding soil properties through detailed chemical and physical assessments enables farmers to optimize fertilizer application, irrigation practices, and crop selection. Traditional agricultural practices often rely on generalized knowledge, leading to inefficient resource utilization and suboptimal yields. By leveraging soil data, precision agriculture can enhance productivity, ensuring that farming practices are tailored to the specific needs of different soil types.

Additionally, soil health monitoring helps mitigate issues like nutrient depletion, erosion, and contamination from excessive fertilizer or pesticide use. Sustainable agricultural methods depend on accurate soil data, which, when integrated with advanced analytical techniques, can lead to improved crop yield predictions and better resource management.

III. LITERATURE REVIEW

[1] P.D. Malwe, B. Gawali, M. Deshpande (September 2022) conducted an in-depth energy analysis of grape cultivation in Maharashtra, India, particularly in the Sangli district, a key global hub for grape and raisin production. The study quantifies energy consumption in essential farming activities, drawing data from prominent farms in the region. Findings indicate that diesel (29%), electricity (22%), and human labor (15%) constitute the primary energy sources. The research reveals an energy output nearly six times the input, with an energy ratio of 5.92 and a productivity ratio of 0.56 kg/MJ, demonstrating the efficiency of local farmers in energy utilization.

ISSN PRINT 2319 1775 Online 2320 7876
 Journal UGC CARE Listed (Group-I) Volume 11, Issue , 10 2022

[2] H.A. Burhan (July 2022) explores the effects of global events, especially the early COVID-19 pandemic, on agriculture, causing fluctuations in demand and pricing. Despite these vulnerabilities, Turkey is working toward enhancing data-driven agricultural strategies as part of its 11th development plan, with an emphasis on accurate crop yield forecasting. The study utilizes meteorological data, pesticide usage, and crop statistics from 1990 to 2019, implementing machine learning techniques. Decision Tree Regression (DTR) and Random Forest Regression (RFR) show promise in predicting wheat, barley, and maize yields, while Support Vector Regression (SVR) produces mixed results. The research recommends expanding datasets and exploring alternative machine learning models such as ANN, DNN, and kNN for better predictions.

[3] A. Mhetre, R. Chavan, and S. Chaudhari (2020) conducted an economic analysis of grape production in Sangli, Maharashtra. Using a multistage sampling method, the study focused on Tasgaon and Miraj, two tehsils with significant grape cultivation. Data was collected from 96 grape farmers through structured interviews during the 2019-20 period. Cost analysis revealed an establishment cost of Rs. 568,303.2 per hectare, while cultivation costs totaled Rs. 1,124,862, largely attributed to land rental, plant protection, and labor. The study found an average yield of 355.98 quintals per hectare, with gross returns of Rs. 2,036,205.6 and net profits of Rs. 911,343.6 per hectare. The output-input ratio of 1.81 highlights the financial viability of grape farming in the region.

[4] Chen and Wang (2020) discuss the role of machine learning (ML) in agriculture, examining various algorithms and their practical applications. Decision trees (DT) are highlighted for their simplicity and ease of interpretation, though they tend to overfit small datasets. To counter this, ensemble techniques such as random forests and gradient boosting machines are utilized to improve predictive accuracy. Support vector machines (SVM) perform well in classification and regression tasks but require careful hyperparameter tuning and are less efficient for very large datasets. Neural networks, especially deep learning (DL) models, excel in detecting complex patterns but present challenges in interpretability and computational resource demands. The study underscores ML's transformative potential in agriculture, particularly in yield forecasting, disease identification, pest control, and resource management, while also acknowledging obstacles related to data quality and the need for model customization based on agricultural contexts.

[5] Ersin Elbasi et al. (May 2023) investigate the integration of machine learning with IoT sensors to enhance agricultural efficiency. The research aims to optimize crop production and minimize waste through data-driven decision-making in planting, irrigation, and harvesting. The study explores the challenges and opportunities associated with machine learning and IoT applications in agriculture, presenting empirical evidence on the impact of data labeling on algorithm accuracy. The findings demonstrate the potential of real-time farm data in improving agricultural decision-making. Experimental results show high classification accuracy, with Bayes Net achieving 99.59% accuracy, Naïve Bayes Classifier 99.46%, and Hoeffding Tree algorithms. These results highlight the effectiveness of machine learning in enhancing crop productivity and reducing losses.

[6] Brown and Wilson (2019) analyze how soil properties influence crop performance, focusing on pH levels, nutrient composition, texture, and organic matter content. They

highlight the importance of soil health for sustainable agricultural productivity. Soil pH determines nutrient availability, with optimal ranges differing across crops. Essential macronutrients—nitrogen, phosphorus, and potassium—are critical for growth and yield. Soil texture impacts water retention, drainage, and root penetration, while organic matter improves soil structure and nutrient supply. The study discusses soil management techniques, including conservation practices, organic amendments, and precision agriculture strategies, to enhance soil fertility and productivity. The authors stress the necessity of continuous monitoring and adaptive strategies to maintain soil quality amid environmental changes. Lee and Park (2018) further explore how machine learning techniques can be integrated into soil analysis for precision agriculture, reviewing predictive modeling, classification, and clustering methods to improve decision-making in soil management.

IV. PREDICTION AND STRATEGY FORECAST OUTCOMES

Machine learning models in agriculture not only predict crop yields but also aid in formulating strategic decisions to enhance productivity. Prediction models analyze factors such as soil nutrients, water availability, temperature variations, and past yield data to forecast agricultural output with high accuracy.

The outcomes of predictive models include:

- Yield Prediction: Estimating future crop production based on soil conditions, weather forecasts, and historical data.
- Fertilizer Optimization: Providing precise recommendations for fertilizer types and quantities to prevent overuse and soil degradation.
- Water Management: Suggesting irrigation schedules to maximize water efficiency and minimize wastage.
- Crop Disease and Pest Detection: Using image recognition and data analysis to identify disease symptoms and recommend treatments.
- Seasonal Strategy Formulation: Advising farmers on the best planting and harvesting times based on climatic patterns and soil data.

By integrating these predictive insights, agricultural strategies can shift from reactive to proactive, leading to improved yield consistency and reduced losses. Furthermore, such approaches contribute to sustainable farming practices, ensuring long-term soil health and resource conservation.

V. RESEARCH GAP

Despite extensive research in agricultural yield prediction and soil data analysis, several gaps persist in the existing body of knowledge. These gaps highlight areas where further investigation is required to enhance prediction accuracy and optimize agricultural strategies using machine learning models. One major gap is the lack of integrated approaches in fertilizer recommendation. Most studies either focus on soil analysis or crop yield prediction separately, without considering a comprehensive integration of soil composition, water quality, weather conditions, and crop-specific requirements to recommend the most effective fertilizer combinations.

ISSN PRINT 2319 1775 Online 2320 7876
Research Paper © 2012 IJFANS. All Rights Reserved, Journal UGC CARE Listed (Group-I) Volume 11, Issue , 10 2022

Furthermore, there is inadequate consideration of real-time data and IoT integration. Most existing studies focus on analyzing historical data without incorporating real-time sensor-based soil and weather data. The integration of IoT-based soil monitoring with machine learning can significantly enhance real-time prediction accuracy and support better decision-making in agricultural management. Additionally, research tends to emphasize broad agricultural trends rather than region-specific crop yield predictions. Since soil composition, climate, and farming practices vary across different geographical regions, models must be tailored to accommodate regional variations for more accurate and practical results.

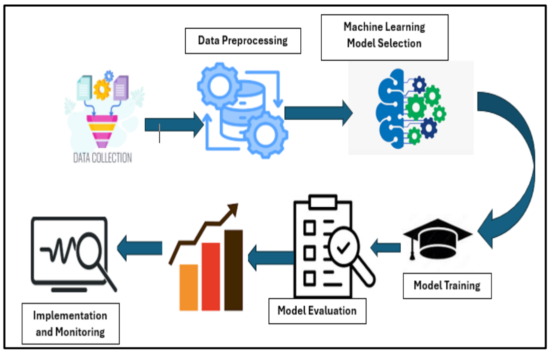
A crucial gap exists in the lack of accessible and farmer-friendly decision support systems. While machine learning models can predict yield and suggest strategies, there is a need to translate these predictions into user-friendly applications for farmers.

Another challenge in the current research landscape is data collection and quality issues. Many existing models suffer from poor data quality, missing values, and inconsistent records, which affect the reliability of predictions. There is a strong need for a standardized data collection framework that ensures high-quality and reliable datasets.

Finally, there are limited studies on the economic impact of AI-driven agriculture. While several studies highlight the technological feasibility of machine learning models in agriculture, few research efforts quantify their economic benefits, such as cost savings, yield improvements, and reductions in fertilizer misuse. Conducting an in-depth economic analysis would help justify the adoption of AI-driven techniques in agriculture, making it more appealing for policymakers and farmers alike.

VI. PROPOSED METHODOLOGY

Fig. 1 presents a conceptual framework that leverages machine learning techniques in conjunction with soil analysis to optimize agricultural yield prediction and strategic planning. The model is designed to promote sustainable and resource-efficient farming methods. Each phase plays a pivotal role in translating raw agricultural data into meaningful insights for informed decision-making.



VII. CONCLUSION

Machine learning has emerged as a powerful tool to enhance agricultural yield prediction by effectively analyzing diverse and complex datasets related to soil composition, weather patterns, and crop conditions. Models such as Random Forest, Support Vector Machines, and deep learning have demonstrated superior accuracy compared to traditional methods, enabling better forecasting and resource optimization. The integration of IoT technologies for real-time data acquisition further advances precision agriculture by allowing dynamic and timely decision-making.

However, several challenges remain. Most existing approaches separately address soil analysis, yield prediction, and fertilizer recommendation without comprehensive integration, limiting their practical utility. The scarcity of real-time IoT-based monitoring integration hinders adaptive management in rapidly changing field conditions. Additionally, regional and crop-specific tailor-made models are needed to account for geographic variability. Another critical limitation is the lack of accessible, user-friendly decision support tools that can translate complex model outputs into actionable guidance for farmers, particularly smallholders.

Data quality issues and inconsistent datasets pose additional barriers, highlighting the necessity for standardized, high-quality data collection frameworks. Furthermore, economic impact analyses assessing the cost-effectiveness and sustainability benefits of AI-driven agriculture remain limited. Addressing these gaps through interdisciplinary collaboration and technological innovation will solidify machine learning's role in fostering sustainable, efficient, and profitable agriculture, ultimately supporting global food security and environmental stewardship.

References

1. P.D. Malwe, B. Gawali, M. Deshpande et al., Energy nexus for grapes production: A case study of Sangli region in India, Energy Nexus 8 (2022) 100145, https://doi.org/10.1016/j.nexus.2022.100145, 19 September 2022, 2772-4271/©2022, Elsevier Ltd.
2. A.V. Mhetre, R. V. Chavan, S. A. Chaudhari, Economic Analysis of Grape Production in Sangli District of Maharashtra, International Journal of Current Microbiology and Applied Sciences, ISSN: 2319-7706 Special Issue-11 pp. 1439-1444
3. Menaga, S. Vasantha, Smart Sustainable Agriculture Using Machine Learning and AI: A Review, Springer Nature Singapore Pte Ltd. 2022, Y.-C. Hu et al. (eds.), Ambient Communications and Computer Systems, Lecture Notes in Networks and Systems 356, https://doi.org/10.1007/978-981-16-7952-0_42
4. Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Montesinos-López, J. C., Mota-Sanchez, D., Estrada-González, F., et al. (2018). Prediction of multiple trait and multiple environment genomic data using recommender systems. G3 (Bethesda) 8, 131–147. doi: 10.1534/g3.117.300309
5. Patil, B.V., Gala, D.M., Pawar, B., Chakraborty, A., Enhancing Agricultural Resilience in Sangli District: Leveraging ML for Soil-based Yield Forecasting and Strategy Development (2024) Proceedings of the 18th INDIAcom; 2024 11th International Conference on Computing for Sustainable Global Development, INDIACom 2024, pp. 427-433
6. Rumpf, T., Mahlein, A. K., Steiner, U., Oerke, E. C., Dehne, H. W., & Plümer, L. (2010), Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Computers and Electronics in Agriculture, 74(1), 91-99. https://doi.org/10.1016/j.compag.2010.06.009
7. Zhang, X., Zhao, L., Sun, Y., Song, A., & Shen, Y. (2018). Application of ML techniques for predicting the nutrient status of soil in China. Agriculture, 8(7), 96. https://doi.org/10.3390/agriculture8070096
8. S. Ajitha, T. V. Suresh Kumar, K. Rajanikanth, Framework for Multi-Agent Systems Performance Prediction Process Model: MASP3, BIJIT - BVICAM's International Journal

ISSN PRINT 2319 1775 Online 2320 7876
 Journal UGC CARE Listed (Group-I) Volume 11, Issue , 10 2022

of Information Technology, BIJIT – 2014; July - December 2014; Vol. 6 No. 2; ISSN 0973 – 5658, PP No -774- 778

9. Smith, J. R., & Johnson, A. B. (2019), A review of statistical models for crop yield prediction. Journal of Agricultural Science, 7(3), 112-126.
10. Chen, L., & Wang, Y. (2020), ML applications in agriculture: A review. Computers and Electronics in Agriculture, 176, 105627.