Literature Review and Data Set Identification

By Samie Ahmad

Introduction

Advances in precision agriculture are transforming traditional farming into a highly automated and data-driven enterprise. As global demand for food increases and environmental pressures mount, researchers and practitioners alike are turning to artificial intelligence (AI), machine learning (ML), and big data analytics to optimize resource use, improve yields, and reduce environmental impact. The collection of works by Fountas et al., Kamilaris and Prenafeta Boldú, Wolfert et al., Viscarra Rossel et al., Folorunso et al., Ajina et al., and Jha et al. collectively underscores the potential of these emerging technologies. From leveraging digital infrastructure and big data to deploying deep learning for soil and crop analysis, these papers highlight strategies to guide the development of AI-driven solutions that offer personalized and site-specific recommendations to farmers.

Individual Paper Summaries and Relevance

Fountas, Spyros, et al. "The Future of Digital Agriculture: Technologies and Opportunities."

Fountas and colleagues explore how digital technologies—particularly those powered by AI and big data analytics—are reshaping agricultural practices. Their discussion emphasizes the shift from conventional methods to data-intensive farming, underscoring the potential benefits for productivity and sustainability. This paper is highly relevant as it provides a high-level overview of emerging precision agriculture tools and frames the context for how AI can be incorporated into decision-making, thus laying foundational insights for an AI-driven soil and crop recommendation platform.

Kamilaris, Andreas, and Francesc Prenafeta Boldú. "Deep Learning in Agriculture: A Survey."

Kamilaris and Prenafeta Boldú survey 40 research efforts employing deep learning in agricultural challenges, ranging from image-based disease detection to yield prediction. Their findings show how deep learning outperforms many traditional image processing and ML methods. For a personalized farming recommendation platform, their survey underscores the significance of sophisticated neural network architectures in solving complex tasks like disease diagnosis or crop health monitoring, thereby informing model selection and data requirements.

Wolfert, Sjaak, et al. "Big Data in Smart Farming – A Review."

Wolfert and colleagues delve into the role of big data in the evolution of smart farming, examining the societal and economic implications of data-driven decisions. They highlight how predictive insights can be derived from large volumes of diverse data and emphasize organizational and governance concerns. These points are directly pertinent to developing an AI-driven platform, as data governance and privacy must be considered when aggregating vast soil, weather, and farm management records to generate individualized recommendations.

Viscarra Rossel, Raphael, et al. "Visible, Near Infrared, Mid Infrared or Combined Diffuse Reflectance Spectroscopy for Simultaneous Assessment of Various Soil Properties."

Viscarra Rossel et al. investigate the efficacy of diffuse reflectance spectroscopy for rapid and cost-effective soil analysis. Their work compares different spectral ranges—visible, near infrared, and mid-infrared—and highlights how these methods can efficiently characterize multiple soil properties. This is highly relevant for an AI-driven platform that relies on timely, accurate soil data to personalize recommendations on fertilization or irrigation, as spectroscopy-based insights can feed directly into machine learning models.

Folorunso, Olusegun, et al. "Exploring Machine Learning Models for Soil Nutrient Properties Prediction: A Systematic Review."

Folorunso and co-authors present a systematic review of how machine learning techniques are applied for predicting soil nutrient properties. By detailing various ML models, data sources, and predictive accuracies, they illustrate how smarter soil mapping and nutrient assessments aid precision agriculture. These insights are essential for developing the soil-analysis component of a personalized AI system, informing model choice, data preprocessing, and potential areas of innovation.

Ajina, A., et al. "Advancements in Crop Yield Prediction Using Deep Learning Algorithms."

In this book chapter, Ajina and co-authors explore the use of deep learning models—such as convolutional neural networks (CNN), long short-term memory (LSTM), and deep neural networks (DNN)—for predicting crop yields under changing environmental conditions. As yield prediction is central to providing tailored recommendations (e.g., planting times, fertilizer use, and irrigation schedules), their discussion on model architectures and implementation strategies offers practical guidance for incorporating deep learning into a farming recommendation platform.

Jha, Kirtan, et al. "A Comprehensive Review on Automation in Agriculture Using Artificial Intelligence."

Jha and collaborators provide an overview of AI and automation efforts in agriculture, looking at core applications (e.g., disease detection, pesticide control, and irrigation management). They highlight internet-of-things (IoT) sensors, ML, and deep learning as key enablers. This review helps situate the role of AI-driven soil analysis within a broader ecosystem of automated solutions, underscoring the potential of integrated systems that combine sensor networks with AI models for actionable, farm-specific insights.

Dataset Identification

For the dataset, we can begin by gathering detailed soil attribute data—such as pH, texture, organic matter, and nutrient levels—from local or national soil surveys (for instance, the National Soil Survey of Pakistan or the FAO's Harmonized World Soil Database), ensuring all records are assigned uniform geographic coordinates. Then, we can source meteorological and vegetation indices from platforms like NASA Earth Observations, the Pakistan Meteorological Department, or SUPARCO, capturing rainfall, temperature, and NDVI/EVI values that align both spatially and temporally with the soil data. Afterward, incorporate yield figures from government or research bodies such as the Pakistan Bureau of Statistics or provincial agricultural departments, again mapping them to the same reference grid. Once these raw datasets are collected:

- Clean and Normalize: Remove or impute missing values, reconcile any discrepancies in units, and convert categorical information (like crop type) into appropriate numeric or encoded form.
- Feature Engineer: Derive new indicators relevant to local conditions, such as salinity risk, average growing degree days for specific crop cycles, or accumulated seasonal rainfall that directly influences yield.
- Merge and Index: Combine all datasets into one unified table or database keyed by location (longitude and latitude) and time (weekly, monthly, or growing-season intervals), preserving source metadata and data quality annotations to maintain clarity about how each attribute was obtained.

Currently, I have started checking where I can feasibly source this data, and will update once the entire dataset has been created.

References

Ajina, A., et al. “Advancements in Crop Yield Prediction Using Deep Learning Algorithms.” *Advancements in Crop Yield Prediction Using Deep Learning Algorithms*, 14 Feb. 2025, pp. 245–64, doi:10.4018/979-8-3693-7250-0.ch008.

Folorunso, Olusegun, et al. “Exploring Machine Learning Models for Soil Nutrient Properties Prediction: A Systematic Review.” *Big Data and Cognitive Computing*, vol. 7, 2023, p. 113, doi:10.3390/bdcc7020113.

Fountas, Spyros, et al. “The Future of Digital Agriculture: Technologies and Opportunities.” *IT Professional*, vol. 22, 2020, pp. 24–28, doi:10.1109/MITP.2019.2963412.

Jha, Kirtan, et al. “A Comprehensive Review on Automation in Agriculture Using Artificial Intelligence.” *Artificial Intelligence in Agriculture*, vol. 2, 2019, doi:10.1016/j.aiia.2019.05.004.

Kamilaris, Andreas, and Francesc Prenafeta Boldú. “Deep Learning in Agriculture: A Survey.” *Computers and Electronics in Agriculture*, vol. 147, 2018, doi:10.1016/j.compag.2018.02.016.

Viscarra Rossel, Raphael, et al. “Visible, Near Infrared, Mid Infrared or Combined Diffuse Reflectance Spectroscopy for Simultaneous Assessment of Various Soil Properties.” *Geoderma*, vol. 131, 2006, pp. 59–75, doi:10.1016/j.geoderma.2005.03.007.

Wolfert, Sjaak, et al. “Big Data in Smart Farming – A Review.” *Agricultural Systems*, vol. 153, 2017, pp. 69–80, doi:10.1016/j.agsy.2017.01.023.