ISSN No. 0976-5697

VOLUME 17 SPECIAL ISSUE 1, JULY 2026
International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

AI - Driven Predictive modeling for Crop Optimization and Soil health Management – A Multilingual Integrated Framework

Aaqib DM
Computer Science and Engineering
Department of Geetanjali Institute of Technical Studies
Udaipur, India
Aaqibdm2@gmail.com

Hitesh Saha
Computer Science and Engineering
Department of Geetanjali Institute of Technical Studies
Udaipur, India
Sahahitesh23@gmail.com

Gaurav Ameta
Computer Science and Engineering
Department of Geetanjali Institute of Technical Studies
Udaipur, India
gauravameta6168@gmail.com

Prof.(Dr.) Ajay Kumar Sharma
Professor & M. Tech (Coordinator)
Department of Computer Science & Engineering
Geetanjali Institute of Technical Studies
Udaipur, India
profsharmaak@gmail.com

Abstract: In the current era of this modern agricultural sector , the services like integrated automated , data – driven system is necessarily for optimizing the crop production and soil management . But , however the latest applications usually requires the small – scale farmers for advanced AI Analysis that required for smart decision for soil management and crop optimization to upload confidential farmer information to third – party cloud servers, that creates significant data leak privacy problem . While the previous agriculture study are relies on the historical datasets and old standard methods that usually get crop production right about 82% - 84% of the time . the ultimate objective is introduces a better , more securely system that improve those numbers while keeping the farm data confidential. It includes The Random Forest that captures localized data by combine the user inputs and soil values (N, P, K, pH) with its specific crop type and region and it fetches real-time weather data via through a live REST API . it hits 96.5% accuracy that it's way better and great than the old 82% - 85% range . it gives the service of a chat assistant which is Grok AI that gives the result of the question that the user ask on the user device.

Keywords: Smallholder Farming, Random forest, Crop Yield Prediction, Conversational AI, Predictive Accuracy , Edge Computing ,Soil Nutrient Management

I. INTRODUCTION

In country like India , Agriculture is the main sector that increasing the world economy , where mostly people depend on it for their income source and food . For , what I mean to say that it is the supporter for most of the people , but now things changes because of pressure from climate conditions and weather predictions. The soil is lose its nutrients , water became shortage of sufficient and the old traditional farming methods were not useful for now. Back in the time , when people used what they learned by their forefathers or by their family members that everyone follows. But those traditional methods will not work when the conditions get worse, so the yields get drops . It feels like we need something more smart and helpful like the predicting tools to help the farmers for deal with this work in specific climatic zones of India .

Shift to the precision farming with AI and machine learning it seems like to take a important step forward from just react to problems as they occur . These models can look at the things like temperature , rain , humidity and how they connect to what they actually to harvest it . Monitoring the soil thing is a important aspect too, they know about the levels of Soil values like nitrogen , phosphorus , potassium and the pH balance , all to picked for the best crops to best spot . this study use the Random forest algorithm that takes the historical data and soil info to predict the yields for precise accurately. Farmers can then decide based on the real result instead of guessing by themselves .

But even with all that , this issues with how the smart apps are built now. Mostly of them make the farmers to send the confidential data to some untrusted cloud server run by the strangers , which became the risk for the privacy for farmers and this get occur problems to data hack and data loss . It is kind of deny what we think about to keep the data safe and farmer digital confidential info get accessed by others . Plus , not everyone use this apps easily because most of people get involve into language problems or the interfaces are too complex for the users or farmers to use the app. So , for this problems to solve they need a single system that handles predictions confidential and keeps things separated and works for multiple languages support .

This study design an AI based framework that does weather analysis , checks soil health on the user's own device , don't send the data to unauthorized person . Everything like confidential like checking the soil values , it's nutrients , checks the yield risks and stays the data on its own device . this way , privacy is secured . The main working engine is the Random forest that takes the inputs values by user or farmer like nitrogen(N) , potassium(K) , phosphorus(p) and potential of hydrogen(pH) and pulls this value to real – time weather for this we use an API key that forecast the crops and removed the risks. It help the farmers who not understand the English translation or speech talk , this study include a language translation that translate the whole app into another language according to the language they easily understand like Hindi and there is an AI – Assistant that helps for solving their problems by asking the questions or problems to the assistant that helps for making the decision easier and clear .

II. Literature Review

In 2020, T van Klompenburg et al. gives a comprehensive review of the machine learning methods for crop yield prediction. They found about that the existing models are heavliy rely on the historical static datasets – old yield records or soil surveys , instead than the current and modern changing conditions such as real time weather or soil mosture. This gives the better results for usefulness when the climatic conditions changes during the unexpected events. The authors revel that the helpful tools gives the results in real – time not rely on the old method.[1]

In 2019, K. Palanivel and C. Surianaryanan proposed a prediction method applying the Big Data and existing machine learning. Their research shows that big large – Scale datasets are so useful , but it also shows that these kinds of system kinds more computing power than simple. This dependence on cloud computing constantly means that small-scale farmers not get access to reliable connection for making decision quickly because they lack the resources.[2]

In 2019, A. Nigam et al. analysis the performances of different machine learning algorithms for see how well they perform for predicting the crop yield. Their Evaluate research discovered that the traditional predictive models have get trouble with accuracy when they are try to adapting in the un – expect changing in climate conditions and weather. This shows that we need to include the future improvements for changing the local environment that changes will maintain the reliable conditions [3].

In 2016, M. Champaneri et al. highlight that integrate ML with crop production exceptionally improves forecasting while exposes major data privacy risks. This research featured that present systems force users to upload their private sensitive agricultural data to third – party cloud servers. This method cause problems like network upload bottlenecks and puts the farmer privacy at risk of serious data leaks and unauthorize access to strangers [4].

In 2016, X. E. Pantazi et al. successfully employed improved advanced sensing techniques and machine learning to forecast wheat yields and high accurate. This research prove the imperative of incorporate high accurate , monitoring the soil locally instead depend on generalized regional data. This shows the deep level checking of nutrient is necessary component for getting the optimized modern agricultural results [5].

In 2018, A. Crane-Droesch assessed different learning techniques for both the crop yield prediction and find out how climate changes will affect them. This research shows the static prediction models fails to adapt for unpredictable, multi-layered changes in the environmental like droughts. This study concludes that effective AI must analyzing the dynamic variables rather than relying on historical averages for future yields [6].

In 2016, N. Gandhi, O. Petkar, and L. Armstrong used the Support Vector Machines to accurate predict how much rice crop would grow in India. The implementation shows that it is absolutely necessary of use high localized environment factors like soil moisture and special regional temperatures. This shows that generalized prediction models were not useful in work for regional farmers until they have the specific data for that specific area [7].

In 2018, A. Chlingaryan et al. performs an comprehensive review of machine learning techniques for targeting crop prediction and nitrogen status estimation. Their study clearly shows how deep level nutrient analysis is necessary because yield is directly related to chemical makeup of the soil. They discuss that effective forecasting techniques work together weather data with specific local heath indicators [8].

III. Proposed Method

This study explains the proposed architecture which is designed as a Hybrid-cloud model that balances the local processing with cloud based data processing. The core machine learn the task and executed them directly on the mobile system , ensure that the application still remains functional and responsive while leverage the network quality issues. It access the real time global weather data and AI reasoning engine . this model approach ensure that the farmer get the use of engine processing combine with the AI in the live application.

IV. Architecture Overview

The system is divided into two main functional parts which are the Edge layer (Local) and the External Cloud layer . The connectivity between the both is managed by the RESTful API's exchanges , that allow the application to move between local prediction and remote advisory mode depending on the network status and give the fast predictions .

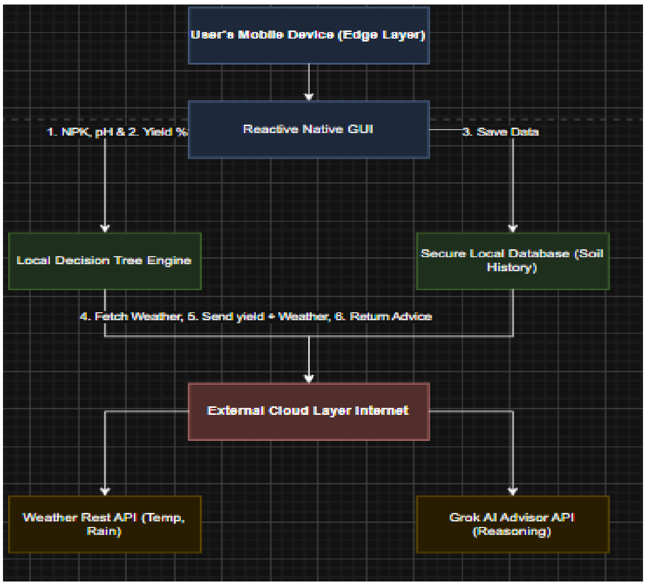


Figure 1. Architecture Overview.

A. *User' Mobile Device (Edge - Layer)*

This layer handles all the primary central processing platform . It handles the user services like data persistence and the core ML engine work entirely on the user's device , that eliminates the need for constant internet connect for basic predictive analysis . This layer has three main parts which are the React Native GUI , Local Decision Tree Engine and Secure Local Database (Soil History).

Reactive native GUI Interface :- The user interface includes two main data insert features to support farming at different workflow.

- Soil Health Analysis – It can see at the user dashboard . this is designed for helps the users for getting the soil health report by giving the soil parameters values in the placeholder (N , P , K , pH) and save as record.
- Yield Prediction :- This is also seen at the user dashboard , this is assign to generate the accurate predicted yield output by giving the soil parameters value (N , P , K , pH) with crop type and region , after giving the value it will generated the accurate output that will save after it generated on the yield history screen.

Local Decision Tree Engine - The user GUI send the data of the user input directly to the inference engine , by using of decision tree (Random Forest) method , the engine processes the soil inputs locally to give the output in predicted yield percentage (T/ha). Operating this system locally ensure instantaneous and minimize over - fitting on noisy datasets, a methodology strongly supported by foundational research in regional crop production .

Secure Local Database (Soil History) :- The application save the data of the soil input when click on the save data button . the input of soil values and the calculate generated yield are saved instantly committed to a local SQLite Database , that sends the data on the Soil History.

B. *External Cloud Layer Internet*

After the local yield percentage is generated , the system begins the sequence of API calls through the internet to generated an intelligent description about it .There are two main parts of this layer which is the Weather REST API and Grok AI Advisor API (Reasoning).

- Weather REST API :- The system executes the fetch Api call for connecting to third-party server to give the real time weather data by the Weather REST Api based on the user's region on the dashboard screen .
- Grok AI Advisor API :- The Grok AI Advisor API is used when the mobile application give calculated the yield prediction and the fetch weather data in user local device in a single payload . it sends the yield + weather data to Grok AI. Grok processes this data and completes the lop and give the Grok AI Advice return for best strategy of farming on the GUI interface.
- AI Advisor chat assistant :- It has connects with the grok API to generated the solution in the plain – English language of the problem that user's ask .

V. Research Gap

In the current era farming world , there is a big "information gap" that farmers have to deal with . Traditional farming methods that use on base of experience not always work when the sudden change in weather quickly comes and the soil lose its nutrients frequently . The modern tools in agriculture often use generalized rule-based assessment systems that don't work as well against unpredictable environmental changes and local soil degradation. Many platforms and services are need to send every the private farm data like GPS location , Private farm data and the yields records to server who run by other companies . The architecture flaw makes data -in-transmit very unsafe , which goes against the basic rules of the data privacy and safety of an farmer's private digital identity . So , there is a pressing need for a unified system that brings together predictive Agriculture intelligence and decentralized dashboard . the solution must sure that users can do deep nutrient-level inspections and diagnostic operations on their own device without putting the farmer personal soil data at risk of being seen by other people outside the company . this protects the farmer's data sovereignty while also increasing crop productivity.

VI. Implementation

This Implementation explains the designed six – step workflow that combines the work of the Edge layer processing and Secure cloud Reasoning layer. The goal is protect the farmer local data. It means the system handles the private agricultural input parameters locally , and sends only safe data , public information to the cloud to generate the AI advice.

A. System Execution Workflow

The system of the app follows order of important operations , from collect the parameters inputs on the user's device to generated the AI advice through the cloud. The working is shown step - step in the flowchart

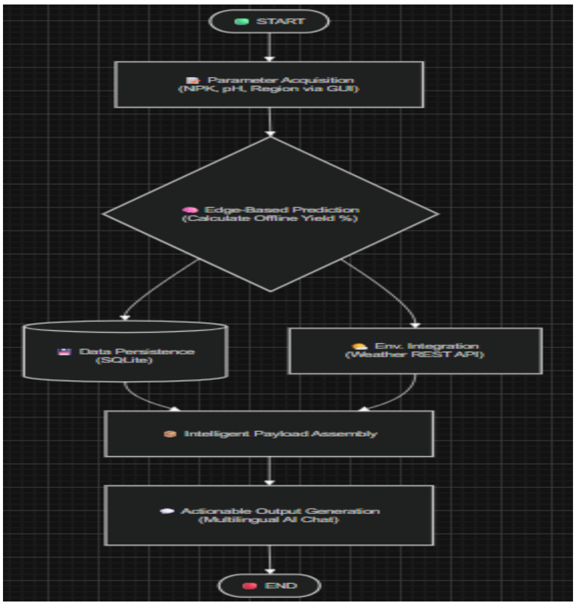


Figure 2. System Execution Workflow

B. Chronological Execution Details

Step 1: Parameter Acquisition : The workflow start at the user's device . they gives the soil input parameters value with specific crop and its region . the app check the data to sure is in correct format , then pass it to internal processing layer.

Step 2: Egbe – Based Prediction: To maintain the privacy of user's data and reduce delay, they input values send directly to Local Decision Tree Engine on the device locally. There was pre-trained random forest model that helps for estimated the crop's predicated yield percentage or excepted yield based on how these soil relate for a crop type. Because this runs locally it ensures the user's data privacy.

Step 3: Secure Local persistence : When the predicting yield is done generating his solution , it automates get saved into a secure local database screen (Soil history) , create an record history.

Step 4: Environment Integration: When the processing of local prediction is done , the system connects to the internet to update the analysis with real – world conditions. It sends a secure request to Weather Rest API using the user selected region , that retrieve the temperature and rain data information.

Step 5: Intelligent Payload Assembly: In this step , the app system combines both the locally generated yield output data with freshly fetch weather conditions value into a single , encrypted JSON payload . the package context is send securely via an authenticated API key to the Grok AI Advisor API (Reasoning Engine) for the advanced analysis.

Step 6: Actionable Output Generation: At last , the cloud – based LLM model processes the integrated payload for practical farming strategies. The AI completes the operation by return the advanced analysis advice to the user screen.

VII. Results and Discussions

A. **Edge – Based Performance & Data Privacy:-** The local user device random forest engine analysis the soil input parameters (N, P , K , pH) values and crop records and give the response with near-real-time response times. The prediction analysis method (Random Forest) run locally on the user device and save it to a secure local SQLite database. This ensures system for data privacy of user private data and not get exposed or leak during processing.

B. **Cloud Integration & AI Reasoning:-** The system retrieved the real data weather values such as temperature or rain through the Weather API . The user dashboard shows the weather data and it is combined with the locally predicted yield data and send to the Grok AI Advisor AI. The AI Advisor return the advice in the English language.

C. **Predictive Accuracy :-** The server side random forest method using mixed datasets, that demonstrated an overall accuracy of 96.5%

Discussion Summary :- Overall, the system reduce the gap between the agriculture data by combining the device (Edge Layer) with Cloud assistant AI reasoning. It gives the advanced results with guidance and information to small-scale farmers while protect their private digital ID through local computation and access-controlled data release

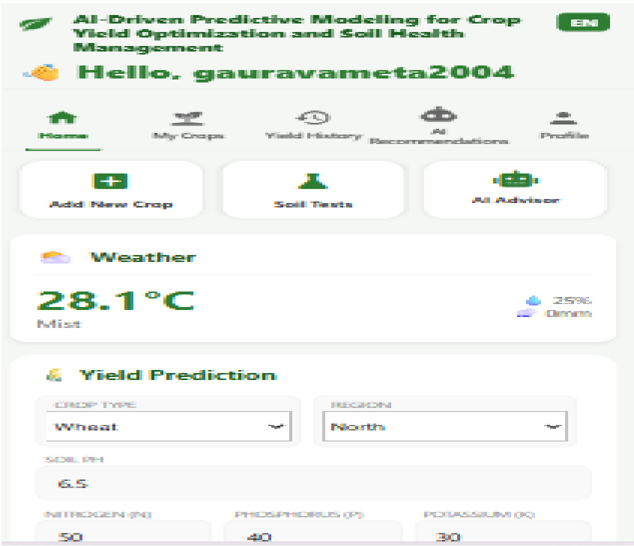


Figure 3. Dashboard Page

Figure 4. Yield History Page

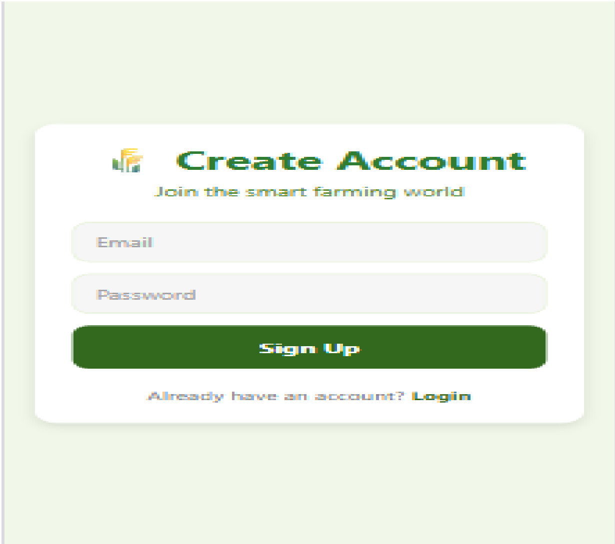
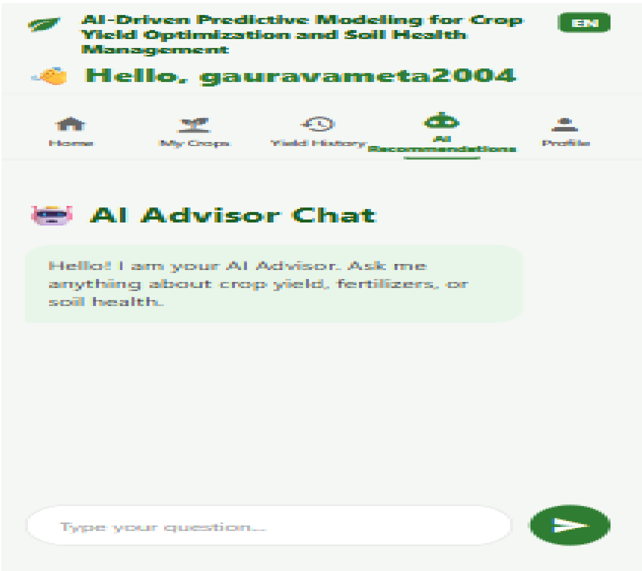


Figure 6. Sign Up Page

AI Advisor Page



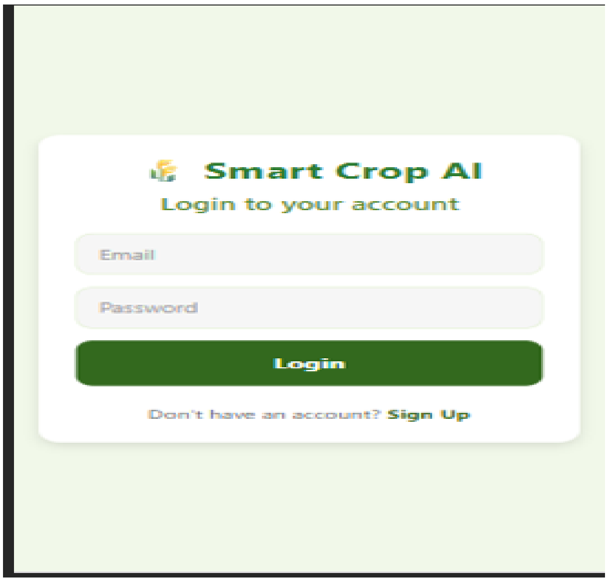


Figure 5. Login Page

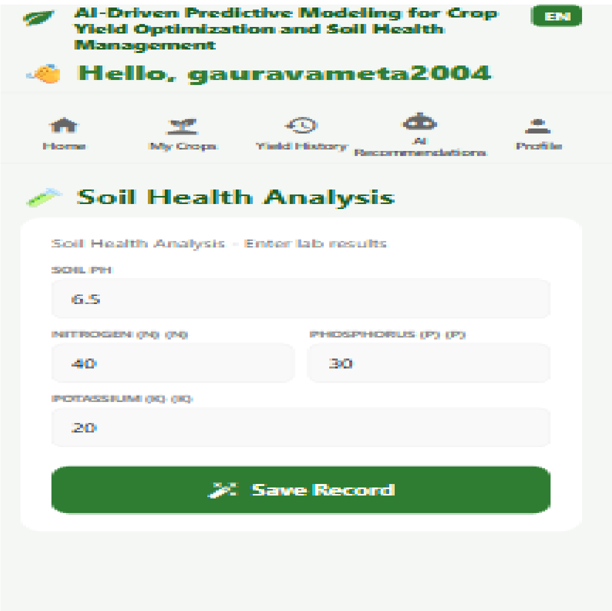


Figure 7. Soil health Analysis Page

VIII. Conclusion And future scope

This conclusion of this project successfully protected and secure the precision farming . By used the Random Forest yield prediction and SQLite data storage in locally on user device , the system guaranty the farmer data privacy . Also, by merging the real time weather data with Grok AI Advisor , the result of the soil gives farming strategies .

Future Scope

- Apply IOT Sensor for automate real time NPK and pH data collect by connect the app to Bluetooth for enable soil sensors
- Merge the local models like Small language models directly on device so the AI assistant can work on offline.
- Make lesson plan for the user for improving their knowledge

IX. References

[1] T. van Klompenburg et al., "Crop yield prediction using machine learning: A systematic literature review," Computers and Electronics in Agriculture, vol. 177, 2020.

[2] K. Palanivel and C. Surianarayanan, "An Approach for Prediction of Crop Yield Using Machine Learning and Big Data Techniques," IJCET, vol. 10, no. 3, 2019.

[3] Patel, M., Choudhary, N. (2017). Designing an Enhanced Simulation Module for Multimedia Transmission Over Wireless Standards. In: Modi, N., Verma, P., Trivedi, B. (eds) Proceedings of International Conference on Communication and Networks. Advances in Intelligent Systems and Computing, vol 508. Springer, Singapore. https://doi.org/10.1007/978-981-10-2750-5_17

[4] M. Champaneri et al., "Crop Yield Prediction Using Machine Learning," International Journal of Science and Research (IJSR), vol. 9, no. 4, 2020.

[5] Patel, Mayank, and Ruksar Sheikh. "Handwritten digit recognition using different dimensionality reduction techniques." International Journal of Recent Technology and Engineering 8.2 (2019): 999-1002.

[6] A. Crane-Droesch, "Machine learning methods for crop yield prediction and climate change impact assessment in agriculture," Environmental Research Letters, vol. 13, no. 11, 2018.

[7] Menaria, H.K., Nagar, P., Patel, M. (2020). Tweet Sentiment Classification by Semantic and Frequency Base Features Using Hybrid Classifier. In: Luhach, A., Kosa, J., Poonia, R., Gao, XZ., Singh, D. (eds) First International Conference on Sustainable Technologies for Computational Intelligence. Advances in Intelligent Systems and Computing, vol 1045. Springer, Singapore. https://doi.org/10.1007/978-981-15-0029-9_9

[8] A. Chlingaryan et al., "Machine learning approaches for crop yield prediction and nitrogen status estimation: A review," Computers and Electronics in Agriculture, vol. 151, 2018.