Soil Analysis and Crop Recommendation using Machine Learning

Dr. Sunitha K, G. Tarun Kumar, Shaikh Saifulla, Sharath Kumar B S, Venugopal M Bhat

Department of Information Science and Engineering
RNS Institute of Technology
Bengaluru-98, India.

Abstract- India constitutes a part of the top three producers of several different crops worldwide and serves as a well-known agricultural center. While Indian farmers play a crucial role in agriculture industry, a large proportion of them are still at the lower end of the socioeconomic scale. Even with a few technological fixes, farmers still struggle to recognize the most lucrative and viable crops for their soil given the diversity of soil types in different parts of the world. This investigation introduces a crop recommendation system that forecasts the best crop derived from a thorough examination of several criteria, such as geography, soil type, yield, selling price, and more. It does this by using both a Convolutional Neural Network (CNN) architecture and a Random Forest Model. It is anticipated that the CNN design would provide an accuracy rate of 95.21%, while the Random Forest Algorithm may produce an accuracy rate of 75%.

Index Terms: Random Forest, Image classification, Deep learning, Convolutional Neural Network, MobileNet v2.

1. INTRODUCTION:
Two-thirds of India’s people directly rely on agriculture for their livelihoods, rendering it a traditional and ongoing pillar of the country's economy. Interestingly, it accounts for 20% of India's GDP (GDP). At the heart of the agriculture industry is the farmer, our country's Anna Datta (Food Provider), who faces a number of difficulties these days:
1) Since there are so many different kinds of soil within the nation, farmers often struggle to choose the best crop for their particular soil, climate, and area, which may result in large losses.
2) As weather patterns are unpredictable, farmers have significant difficulties in calculating profitability and forecasting yields for certain planting seasons.
3) The 'farm to market' method, which involves several middlemen who eat a sizeable percentage of income via the transportation and selling of products, is the cause of the meagre profits obtained by farmers for their goods.

Artificial intelligence and machine learning technologies have several uses in the contemporary agriculture sector. For example, crop recommender systems and precision agriculture can forecast yields, identify plant pests, improve overall quality of harvests, and solve nutrient deficits in farms. AI system integration has the capacity to revitalise the struggling agriculture industry.

There exist significant apprehensions regarding the future of the agriculture industry in India given the present state of affairs. The agriculture industry confronts difficulties even though it employs about two thirds of the workers and contributes 20% of the GDP. Because they lack collateral, almost all Indian farmers—about 85% of them—operate on fewer than five acres of land, assuming significant product and market risks every season. Due to poor literacy rates, smallholding farmers—who account for half of agricultural output and 46% of cultivated land—often find themselves shut out of contemporary market structures like contract farming and direct purchasing.

1.1 BACKGROUND:
The model's objective was to use the most widely available technology, such SMS and email, to suggest crops to even the tiniest farmer at the extent of his or her smallest agricultural plot.

1.2 OBJECTIVES:
This research project's primary objectives are to solve persistent issues in the agriculture industry, regardless of local or regional circumstances. The goal is to enhance the ordinary farmer's profitability by using machine learning. The main objectives include creating a cutting-edge crop recommendation system that uses cutting-edge technology to enable farmers to choose crops wisely. This study also describes the conception and execution of a website that serves as a
marketplace between agricultural purchasers and farmers, removing the need for middlemen.

1. Optimizing Crop Selection: Resolving the conundrum Indian farmers have in choosing the most lucrative and appropriate crops for their soil, especially considering the variety of prevalent soil types in different regions of the country.

2. Leveraging Advanced Technologies: Ensuring that farmers can profit from the newest advancements in agriculture by overcoming the current barriers to obtaining modern technical solutions.

3. Site-Specific Features: Addressing the requirement for site-specific data, in particular the chemical composition of the soil, but also noting the difficulties and financial ramifications that farmers may have in gathering this kind of information.

4. Improving Profitability: By giving farmers useful information via a strong crop recommendation system, it can assist them in becoming more profitable in their farming pursuits.

5. Cutting Out Middlemen: In a bid to simplify the agricultural supply chain and lessen reliance on middlemen, a digital platform that acts as a straightforward marketplace for farmers and prospective crop purchasers has been introduced.

2. RELATED WORK:
2.1 LITERATURE SURVEY:

The research on crop prediction and soil classification in agriculture is compiled and analyzed in this part.

Using data mining methods, Pudumalar et al. [1] created a recommender system for precision agriculture. Their strategy included the CHAID, Random Tree, Naive Bayes, and K-Nearest Neighbor models in an ensemble technique with majority voting. The advice received support from studies on crop yield, soil properties, and kinds.

An ensemble model that integrates Support Vector Machine and Artificial Neural Network, Random Tree, and Naive Bayes models with a majority voting method was proposed by Rajak et al. [2]. The dataset included several soil properties, such as pH and water density, gathered from university and soil testing facilities.

A two-step strategy for crop recommendation and soil categorization was presented by Reddy et al. [3]. During the subsequent phase, consideration was given to characteristics like soil series, temperature, humidity, rainfall, and pH, the initial phase used chemical properties of the soil to forecast the soil series or type. For crop suggestions, classification techniques including Support Vector Machine, K-Nearest Neighbors, and Bagging were proposed.

Venugopal et al. [4] used the Random Forest, Naive Bayes, and Logistic Regression algorithms to forecast yield values and suggest appropriate crops. To train their models, historical data on temperature, weather, and other variables were gathered; the Random Forest Algorithm produced the best accuracy.

A Crop Selection Method (CSM) was presented by Kumar et al. [5] aiming to increase crop net yield rates. The strategy suggested a sequence of crops across a season, considering various factors such as crop type, weather, soil type, and water density. The significance of crop selection was emphasized, along with its influencing variables, encompassing market price, government policy, and production rate.

Data mining methods were utilized by Ahamed et al. [6] to estimate the crop production of cereal crops in Bangladesh. Their technique used KNN, linear regression, and artificial neural networks in the quick miner tool for categorization and clustering to create district clusters.

The approach presented by Mahendra et al. [7] forecasts the ideal crop based on soil properties such as pH, moisture content, and rainfall. The rainfall was predicted using the Support Vector Machine (SVM) technique, while the prediction of crop yield was conducted using the Decision Tree technique.

3. METHODOLOGY:

The two-step crop suggestion model in the suggested technique is easy to utilise. On a specific webpage, users have the choice to submit soil photos or enter reports on certain soil features, such as moisture, phosphorus as well as nitrogen levels. Following the required inputs, the system analyses the information to provide a list of crop suggestions based on expected gross income, quality, and yield. This approach offers a smooth experience for both image-based and feature-based inputs, accommodating users spanning a broad spectrum of preferences.

3.1 SYSTEM OVERVIEW:

The technique can be segmented into two integral components that are meant to provide accurate crop recommendations
and efficient soil categorization.

Step 1: Classifying Soil using MobileNetV2 Initially, this system uses the MobileNetV2 Convolutional Neural Network (CNN) to interpret input from soil characteristics or pictures. This sophisticated neural network assesses the soil input based on many characteristics, such as moisture, phosphate, and nitrogen, and analyzes it to create multiple soil types. The next sections go into further depth about the MobileNetV2 implementation. After classifying the soil, the algorithm selects the crops with the highest likelihood of thriving in the designated soil type, ensuring that the crops are suggested match the properties of the soil for optimum results.

Step 2: Leveraging the Random Forest Algorithm to Predict Income and Recommend Crops: Predicting the revenue and production of the crops that made the short list is the next phase. Numerous factors are considered by this technique, including the anticipated kind of soil, the hectares of land under cultivation, geographical considerations like the state and district, the particular crop, and the cultivation season. The Random Forest Algorithm is employed by the system to predict the quintal yield of any cultivated crop nominated. Moreover, it retrieves local agricultural prices via an API created by the Indian government [5]. The system uses the unitary technique to assess the revenue earned by each specific crop based on crop prices and expected yield. Crop recommendation is grounded on greatest revenue, so quantity and profitability are balanced. By combining soil properties and economic viability, this methodical approach allows the model to provide complete and customized crop suggestions, facilitating an all-encompassing decision-making process in agriculture. In the next parts of this survey article, the Random Forest Algorithm and MobileNetV2 implementation details are explained.

![System Workflow](image)

3.2 SYSTEM ARCHITECTURE
1) CNN, or convolutional neural network:
   i) Collection of Datasets: Convolutional Neural Network (CNN) training dataset is carefully selected and includes photos of several soil types, including clay, alluvial, black, and red soil. The dataset includes 150–200 photos for each kind of soil in the training set and an extra 50 images in the test set to guarantee robust model training. The Soil Classification Image Dataset [6] on Kaggle and other relevant web sources were carefully gathered to create this dataset.

   ii) Pre-processing Data and Implementing Algorithms: Preparing the data is an important phase in the process because of the variations in sizes and dimensions of the gathered photographs. Pictures are downsized to a common 300x300x3 size, which takes into consideration the RGB format's three channels (Red, Green, and Blue). Afterwards, the machine learning model converts these pictures into numerical arrays. This numerical data is then sent into the algorithm, which is tailored to a particular use case, for further processing and training, making sure the CNN is properly trained to categorize different kinds of soil. The in-depth information on the preparation of data and the employing algorithms are explained in the following segments of this survey paper.
Mobilenetv2:
An architecture for deep learning called MobileNetV2 is intended for embedded and mobile devices. It uses inverted residual blocks and depthwise separable convolutions to perform efficient picture recognition. Below is an explanation of the algorithm and how it works:

Algorithm:
1. Input picture: An input image is used to start the process.
2. Fully convolutional layer: To extract the first features from the picture, it is processed through a fully convolutional layer.
3. Inverted residual blocks: After that, a number of inverted residual blocks are run through the features. After applying depth-wise filters and expanding the input channels, each block shrinks the channels to the intended output dimension. The network's effective information flow is guaranteed by the shortcut connection.
4. Bottleneck residual blocks: These blocks find application in downsampling. Before the 3x3 depthwise convolution, these blocks incorporate an extra 1x1 pointwise convolution to further minimize the number of channels.
5. Average pooling layer: An average pooling layer is used to minimize the dimensions of the final feature map.
6. Classification layer: To predict the category of the input picture, the pooled features are subsequently fed into a classification layer.

2) Random Forest:
The Random Forest method is a supervised learning technique that enhances overall accuracy and manages overfitting by aggregating predictions from numerous decision trees.

Algorithm:
- Data preparation:
The training set is split up into many replacement subsamples.
A random selection of characteristics is chosen for every subsample.
- Decision tree construction:
On each subsample, a decision tree is constructed using the chosen characteristics.
The tree continues to grow until a certain depth or other condition is satisfied.
The optimal split is selected at each node according to an impurity measure.
- Prediction:
For a new data point, each decision tree in the forest makes a prediction.
For classification, the majority vote of the trees determines the final prediction.
For regression, the average prediction of the trees is used.
4. ADVANTAGES OF CROP RECOMMENDATION SYSTEM USING MACHINE LEARNING:

1. **Accuracy:** The machine learning-based Crop Recommendation System's effectiveness depends on how well it forecasts appropriate crops in relation to actual crop yields in a given area. Using large-scale datasets and complex algorithms improves the system's crop suggestion accuracy. Moreover, the system's flexibility is further enhanced by ongoing updates that are founded on real-time agronomic data. This guarantees that farmers get the most precise and timely crop suggestions for optimal agricultural practices.

2. **Personalization:** The capacity of machine learning models to consider specific factors like soil composition, weather trends, past crop yields, and other relevant data points is what makes them so strong. With this feature, the system may provide tailored crop suggestions that satisfy particular needs and increase its effectiveness.

3. **Scalability:** Models based on machine learning are characterised by their scalability, which enables them to analyse large amounts of historical data with ease. They are particularly well-suited to managing large datasets. Their smooth incorporation of fresh data points also guarantees an ongoing improvement of suggestions over time, which adds to the system's scalability.

4. **Adaptability:** Machine learning models distinguish themselves with their capability to adjust suggestions in reaction to changing circumstances. With the system's ability to incorporate fresh data insights, elements like climate change or improved farming techniques may be accommodated, leading to more accurate and relevant crop suggestions. This flexibility maintains the system's effectiveness in changing agricultural environments and strengthens its capacity to adapt.

5. **CONCLUSION**

In conclusion, the study report emphasizes how cutting-edge technology have transformed India’s agricultural environment. Convolutional neural network (CNN) and random forest models are integrated into the suggested crop recommendation system, which is predicted to have accuracy rates of 95.21% and 75%, respectively. These models, considering a variety of variables including soil types and local circumstances, are crucial for enhancing crop choices. The study examines the Random Forest Algorithm's strong performance in crop result prediction as well as for effectiveness of MobileNetV2 in real-time applications.

In keeping with the goals, the study highlights how small farmers may easily access cutting-edge technology via user-friendly channels like email and SMS. By serving as a direct marketplace, the suggested online platform has the ability to cut out middlemen, simplify the supply chain, and give farmers more control. The benefits of utilizing machine learning in crop recommendation crop recommendation, including precision, customization, scalability, and flexibility, emphasize the potential of these technologies to completely change conventional agricultural methods.

Looking forward, the survey's conclusion emphasizes the need conducting additional research on creative fixes for India's agricultural ecosystem to make it more robust and sustainable. With its innovative technology and farmer-focused approach, the suggested crop selection system is a big step toward reviving the agriculture industry and providing hope for a better future.

REFERENCES:


