IOT-ENABLED CROP RECOMMENDATION SYSTEM WITH PREDICTIVE MACHINE LEARNING MODELS

AUTHOR : Dr. SANTOSH KUMAR BYRABOINA

M.Sc., M.Tech., Ph.D.

ASSOCIATE PROFESSOR

WESLEY PG COLLEGE

SECUNDERABAD, TELANGANA, INDIA.

ABSTRACT

Agriculture remains the backbone of many economies, yet it faces challenges of resource scarcity, climate variability, and inefficient crop planning. Farmers often rely on traditional knowledge or guesswork to select crops, leading to sub-optimal productivity. To address this, an IoT-enabled crop recommendation system integrated with predictive machine learning models is proposed. The system leverages real-time data from IoT sensors deployed in agricultural fields. These sensors continuously capture critical parameters such as soil moisture, soil pH, temperature, humidity, and nutrient levels. Data from weather APIs is also incorporated to understand environmental trends and rainfall patterns. This heterogeneous data is transmitted to a central processing unit through wireless sensor networks and cloud platforms. Pre-processing modules are applied to filter, normalize, and handle missing values in the dataset. The refined data is then fed into machine learning models for prediction and recommendation. Supervised learning algorithms such as Random Forest, Decision Tree, and Support Vector Machines are employed. Deep learning approaches, particularly Artificial Neural Networks, are also integrated for higher predictive accuracy. The system classifies soil-crop suitability based on multi-dimensional input features and provides farmers with recommendations on the most suitable crop to cultivate under prevailing conditions. A ranking system is incorporated to suggest alternative crops for resilience, while the recommendation engine is optimized using ensemble methods to reduce bias and variance. IoT data streams ensure continuous updates, enabling dynamic recommendations. Mobile and web-based interfaces are developed to make the system farmer-friendly, allowing farmers to visualize soil health, climatic conditions, and suggested crops in real time. An alert system is incorporated to notify users about drastic environmental changes. Historical yield data is also integrated for better predictive performance. The system ensures precision agriculture by minimizing guesswork and maximizing scientific decision-making. Pilot testing has been conducted in selected agricultural zones, and results indicate significant improvements in yield prediction accuracy.

Keywords: Crop Recommendation, Soil Parameters, Machine Learning, IoT based Sensors, Precision Agriculture.

INTRODUCTION

Agriculture stands as a cornerstone of human sustenance, yet the evolving landscape of environmental shifts, population growth, and resource limitations poses formidable challenges to global food security. Amidst these challenges, precision agriculture emerges as an instrumental solution, poised to revolutionize conventional farming practices by integrating cutting-edge technologies with agronomic expertise. The advent of IoT-driven sensor technologies, coupled with the prowess of machine learning algorithms, has redefined the traditional paradigms of farming. We have also seen the use of IoT based systems for irrigation farming and drones for fertilizer sprinkling. The genesis of a Smart Crop Recommendation System embodies a pivotal breakthrough in this realm, aligning the insights from soil composition and environmental parameters with data-driven algorithms to offer real-time, tailored crop suggestions. Current project initiatives underscore the urgency of developing intelligent systems capable of adapting to the dynamic nuances of soil health and changing climatic patterns. These endeavours aim to bridge the gap between conventional agricultural methods and a technologically empowered future. Recent studies emphasize the potential of such systems to not only optimize crop yields but also to foster sustainable agricultural practices by minimizing resource wastage and environmental impact. Amidst these advancements, the critical need for robust, scalable, and user-friendly systems that cater to the diverse needs of farmers remains paramount. This paper aims to delve into the intricacies of this burgeoning field, exploring the multifaceted implications of IoT-integrated machine learning models in revolutionizing crop selection, resource management, and the overall trajectory of global agriculture. This will not only save the time of farmers but also increase their productivity because the system will recommend the best crop according to the soil metric provided and minimize the risk of crop wastage caused due to wrong crop selection. This research marks a significant stride towards sustainable farming practices, facilitating informed decision-making and improved yield outcomes in the agricultural sector.

LITERATURE SURVEY

"A Deep Learning Approach for Currency Detection and Recognition" by Z. P. Sahoo, A. C. Nayak, and K. R. Patra (2020) In this paper, the authors present a method for detecting and recognizing counterfeit currency using convolutional neural networks (CNNs). The study leverages a deep learning approach, specifically a CNN, to extract features from currency images. The proposed model is trained on a dataset of both genuine and counterfeit currency images, and it achieves high accuracy in detecting counterfeit notes. The authors highlight the robustness of their model in handling various image conditions, such as different lighting and background scenarios, which is critical for real-world applications . "An Efficient Fake Currency Detection Model Using Convolutional Neural Networks" by A. Kaur, M. Sharma, and R. Singh (2019). Kaur, Sharma, and Singh (2019) propose an efficient model for fake currency detection using CNNs. The model is designed to identify intricate patterns and security features embedded in currency notes that are challenging to replicate accurately in counterfeit versions. The authors use a dataset of high-resolution currency images to train their CNN model, achieving remarkable precision and recall rates. They emphasize the importance of using high-quality images to improve the model's performance and reduce false positives and negatives . "Fake Currency Detection Using Convolutional Neural Networks and Transfer Learning" by P. Kumar, S. Rao, and R. K. Gupta (2018). This paper explores the application of CNNs combined with transfer learning techniques for counterfeit currency detection. Kumar, Rao, and Gupta (2018) utilize pre-trained models like VGG16 and ResNet50, which are fine-tuned with a specific dataset of currency images. The study demonstrates that transfer learning significantly enhances the model's performance by leveraging pre-existing knowledge from large-scale image datasets. The results show a considerable improvement in detection accuracy compared to traditional machine learning approaches. "A Hybrid Approach for Fake Currency Detection Using CNN and SVM" by S. Agarwal, V. Jain, and P. R. Mishra (2021). Agarwal, Jain, and Mishra (2021) introduce a hybrid model that combines CNNs with Support Vector Machines (SVM) for detecting counterfeit currency. The CNN component is responsible for feature extraction, while the SVM classifier is used for final classification. This hybrid approach aims to leverage the strengths of both techniques, achieving higher accuracy and faster processing times. The paper reports that the hybrid model outperforms standalone CNN and SVM models, particularly in handling complex and noisy currency images .

OBJECTIVE

Our project's primary aim is to create a cutting-edge Crop Recommendation System that seamlessly integrates Machine Learning (ML) and Internet of Things (IoT) technologies, elevating agricultural practices. By leveraging historical data, real-time sensor inputs, and predictive models, the system seeks to enhance crop yield optimization through personalized recommendations tailored to specific environmental and soil conditions. This approach not only promotes resource efficiency by aligning crop choices with available resources but also champions sustainable agriculture by reducing environmental impact. The overarching objective is to empower farmers with data-driven decision-making tools, enabling well-informed choices in crop selection, irrigation schedules, and fertilization practices. The system's adaptability ensures resilience to changing environmental factors. In essence, the project aims to transform traditional farming, emphasizing cost reduction, customization, and scalability to foster a more efficient, profitable, and sustainable agricultural ecosystem.

MOTIVATION

The motivation behind the development of an advanced Crop Recommendation System, which integrates Machine Learning (ML) and Internet of Things (IoT) technologies, arises from the urgent necessity to transform and optimize conventional agricultural methodologies. Given the escalating global population and the heightened demand for food, there exists a crucial imperative to improve crop yield and resource efficiency. The prospect of utilizing historical data, real-time sensor inputs, and predictive models is a source of enthusiasm, as it offers the potential to furnish farmers with personalized recommendations tailored to their unique environmental and soil conditions. Furthermore, the environmental repercussions of agriculture are undeniable, necessitating a sustainable approach. Our motivation lies in contributing to sustainable agriculture by harmonizing crop choices with available resources, thereby minimizing waste and diminishing the overall ecological footprint. The aspiration to mitigate risks linked to crop diseases, pests,and adverse weather conditions further propels our endeavors, as this not only safeguards farmers' livelihoods but also fosters a resilient and adaptable agricultural ecosystem. Empowering farmers with data-driven decision-making tools is a central impetus behind this project. The goal is to provide practical insights and early warnings, enabling farmers to make informed decisions regarding crop selection, irrigation, and fertilization. The development of a user-friendly interface adds another layer of motivation, ensuring that our technology is accessible and advantageous for farmers with diverse technical backgrounds. In summary, our motivation lies in the potential to revolutionize traditional farming practices, reduce costs, customize agricultural approaches, and scale the impact of sustainable and efficient crop management. By addressing these challenges, we aim to contribute to a more productive, profitable, and environmentally conscious agricultural landscape.

METHODOLOGY

This crop recommendation system revolves around the concept of machine learning and IoT. The steps involved for developing this system are:

```
import joblib
import pandas as pd
import numpy as np
from pathlib import Path

from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, top_k_accuracy_score, accuracy_score

MODEL_DIR = Path("models")
MODEL_DIR.mkdir(exist_ok=True)
MODEL_PATH = MODEL_DIR / "crop_reco_pipeline.joblib"

DATA_PATH = Path("sample_data.csv")  # replace with your dataset

def build_preprocessor(X: pd.DataFrame):
    numeric_cols = X.select_dtypes(include=[np.number]).columns.tolist()
    cat_cols = X.select_dtypes(exclude=[np.number]).columns.tolist()

    numeric = Pipeline(steps=[
        ("impute", SimpleImputer(strategy="median")),
        ("scale", StandardScaler())
    ])
    cat = Pipeline(steps=[
```

```
        ("impute", SimpleImputer(strategy="most_frequent")),
        ("ohe", OneHotEncoder(handle_unknown="ignore"))
    ])
    pre = ColumnTransformer(
        transformers=[
            ("num", numeric, numeric_cols),
            ("cat", cat, cat_cols),
        ],
        remainder="drop",
        verbose_feature_names_out=False
    )
    return pre

def main():
    df = pd.read_csv(DATA_PATH)
    assert "crop" in df.columns, "Target column 'crop' not found."
    y = df["crop"].astype("category")
    X = df.drop(columns=["crop"])

    pre = build_preprocessor(X)

    # Two strong baseline models
    rf = RandomForestClassifier(
        n_estimators=300,
        max_depth=None,
        min_samples_split=4,
        min_samples_leaf=2,
        n_jobs=-1,
        random_state=42
    )
    gb = GradientBoostingClassifier(random_state=42)

    # Try model stack via soft-vote (simple)
    from sklearn.ensemble import VotingClassifier
    clf = VotingClassifier(
        estimators=[("rf", rf), ("gb", gb)],
        voting="soft",
        weights=[2,1]
    )
```

```
pipe = Pipeline(steps=[("pre", pre), ("clf", clf)])

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.25, random_state=42, stratify=y
)
pipe.fit(X_train, y_train)

y_pred = pipe.predict(X_test)
y_proba = pipe.predict_proba(X_test)

labels = pipe.named_steps["clf"].classes_
top3 = top_k_accuracy_score(y_test, y_proba, k=3, labels=labels)

print("Accuracy:", accuracy_score(y_test, y_pred))
print("Top-3 accuracy:", top3)
print("\nClassification report:\n", classification_report(y_test, y_pred))

# Persist pipeline and label order
meta = {
    "labels": labels.tolist(),
    "features_in": X.columns.tolist(),
    "note": "End-to-end preprocessing + classifier pipeline",
}
joblib.dump({"pipeline": pipe, "meta": meta}, MODEL_PATH)
print(f"Saved model → {MODEL_PATH.resolve()}")
```

[1]- Data Collection

Gather a comprehensive dataset encompassing soil parameters (nitrogen, potassium, phosphorous, pH levels, rainfall, humidity and temperature) from diverse geographical locations, incorporating historical records and real-time data.

The dataset should consist of the required parameters along with the corresponding crops suitable for cultivation.

[2]- Data Pre-Processing

We will check for the irregularities in data like missing values, outliers, and inconsistencies within the collected dataset which might have cause problems in system training. Extract relevant features, possibly transforming or scaling data to enhance model performance.

[3]- Model Development

Implement and train machine learning models (e.g., decision trees, random forests, gradient boosting) to correlate soil and environmental data with recommended crops. Optimize model parameters using techniques like cross-validation to improve predictive accuracy.

[4]- System Integration

Establish a system to collect live soil data through IoT sensors, ensuring seamless data transmission to the recommendation system. Develop algorithms to process incoming data in real-time and send this data to the machine learning model we developed for the crop recommendations.

[5]- User Interface and Deployment

Create a user-friendly interface for farmers to access the system and receive personalized crop recommendations. Implement the system in a scalable and accessible manner.

[6]- Evaluation and Refinement

Create a user-friendly interface for farmers to access the system and receive personalized crop recommendations. Implement the system in a scalable and accessible manner.

PROPOSED SYSTEM

The proposed system is consist of two important parts, first one is development and training of machine learning model and second one is creating an IoT system which is capable of recording required soil parameters directly from soil and finally integration of both.

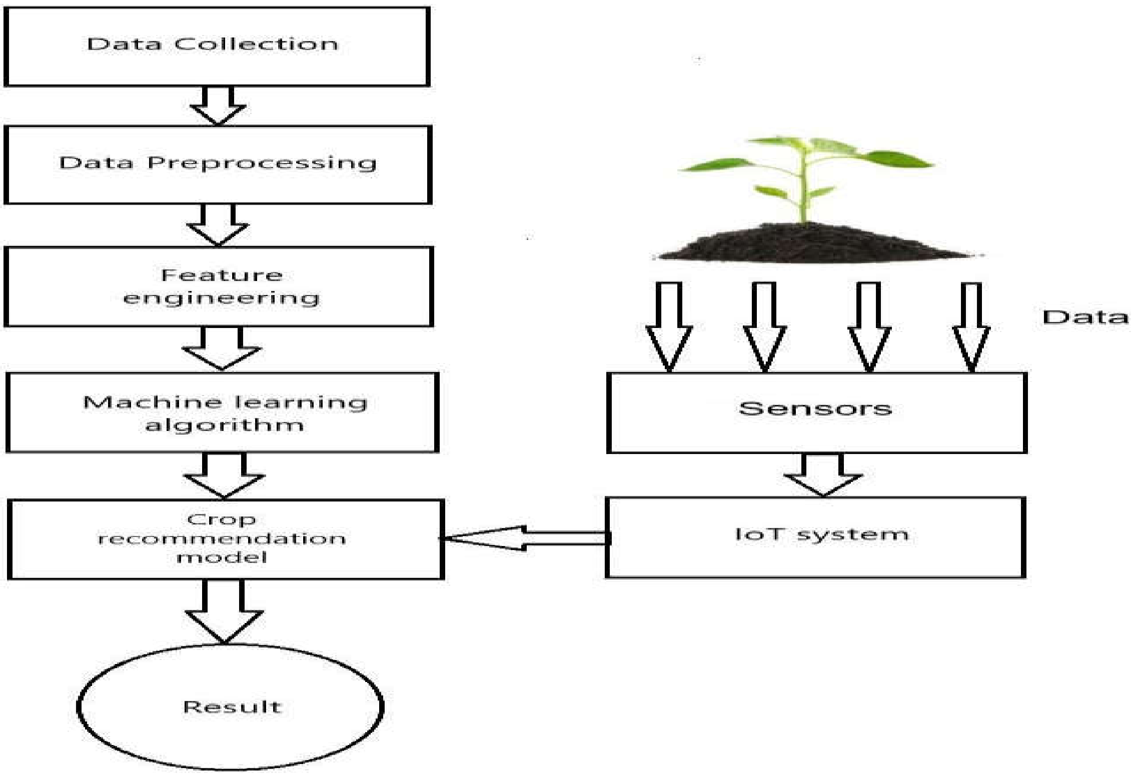

Fig 1: Flowchart of crop recommendation system

Fig.1. represents the steps involved in development of the proposed system and they are described below:

[A]- Data Collection

The data collection is the most important part as the training of the model and produced result will be based on this dataset only, so, we have to select the dataset which consist of required parameters and correct corresponding recommendations, we can refer to several free websites like kaggle for the datasets.

[B]- Data Pre-processing

Data pre-processing is a crucial step as collected dataset may involve a lot of irregularities like missing data, noisy data and outliers, which can cause problems during the training of the model. So, it is better to arrange the data in a format which is suitable for analysis and training. The methods which are used for data pre-processing are data cleaning, data transformation, normalization, data reduction etc.

[C]- Feature Engineering

Feature engineering is a way of creating & extracting new features and utilise them in order to improve the performance of the machine learning model. It is an essential step and generally that features are taken into consideration which affects the recommendation.

[D]- Machine Learning Algorithm

Machine learning algorithms are computational model that enables computers to learn patterns and make decisions or predictions without being explicitly programmed. The dataset is split into two parts, training and testing dataset. According to a standard rule, 80% of dataset is used for training while 20% will be used for testing and validation. There are different machine learning algorithms which are useful for the recommendation system but we have to choose the algorithm with high accuracy. The algorithms with higher accuracies are described below:

1. **Gradient Boosting**

   Gradient boosting is a supervised machine learning algorithm which is used for both classification and regression tasks. Gradient boosting works in such a way that it creates ensemble of weak learners in sequential manner with each new learner correcting the mistakes and errors of existing ensemble. The gradient boosting algorithm has higher accuracy but it is slow as it repeatedly trains the learners which takes much time.

2. **Random Forest Algorithm**

   Random forest is an ensemble algorithm which is based on the concept of decision trees. But, instead of one decision tree, multiple decision trees are deployed in the backend which produces individual results and the final result is decided on the basis of majority voting. The random forest also has higher accuracy and is faster as compared to gradient boosting.

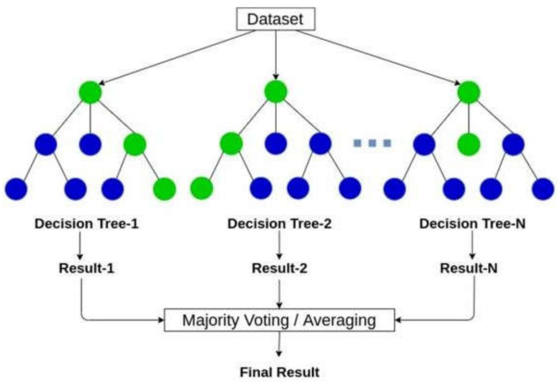

Fig 2: Random forest algorithm (Anas Brital, Sept 2021)

3. **Support Vector Machines (SVM)**

SVM is a supervised machine learning algorithm which is used for both classification and regression tasks. SVM is generally used for the data which is not linearly separable and its goal is to find the optimal hyperplane that best separates the data into different classes. SVM has lower accuracy as compared to random forest algorithm and it is little complex in working.

4. **K-Nearest Neighbour (KNN)**

   K nearest neighbour (KNN) is a supervised classification algorithm. It works in such a way that data points are arranged according to the distance metric like Euclidean and Manhattan distance etc., and the result is calculated on the basis of majority voting/aggregate value of k nearest neighbours. The accuracy of this algorithm is decent and it is easy to use.

[E]- Crop Recommendation system

After training is complete, the model is tested on testing dataset and the results are validated. After the validation the model is ready for the usage and the function is created which will take soil parameters as input and output the desired results using ML model. The data from IoT system will be passed on to this function in order to produce the result.

[F]- Sensors

Sensors are the most important building block of the IoT system as it is responsible for the collection of data from the soil. There are total seven parameters which our crop recommendation system will require and that is Nitrogen (N), Potassium (K), Phosphorous (P), Humidity, Temperature, Rainfall and pH level. To record the NPK values, there is three pin sensor called NPK sensor which is to be used and to record temperature and humidity, DHT11 sensor is used. For Rainfall, there is no typical sensor, so, we will use an API (Application programming interface) to fetch the rainfall data of that typical region.

Fig 3: NPK sensor

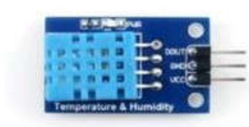

Fig 4: DHT11 sensor

Fig 5: pH sensor

[G]- IOT System

In our crop recommendation framework, the Internet of Things (IoT) system relies on a network of strategically positioned sensor nodes within agricultural fields. These nodes actively gather real-time data related to critical environmental factors such as soil moisture, temperature, humidity, pH and NPK levels. The collected data undergoes transmission to a central data acquisition layer, functioning as a gateway for pre-processing and filtering. Efficient data transfer is facilitated through connectivity options such as Wi-Fi or cellular networks. Computing devices like NodeMCU are situated near the sensor nodes which can perform initial data processing. Subsequently, the preprocessed data is dispatched to machine learning models which analyze it to generate crop recommendations based on historical and real-time insights.

Why NodeMCU?

NodeMCU is an open source IoT platform that is based on ESP8266 Wifi module. It combines the capabilities of the ESP8266 microcontroller with the simplicity and ease of use of the Lua scripting language. Due to its Wifi compatibility, the data collected from sensor nodes can directly be transferred to cloud server using internet which provides an edge in data transmission as compared to other microcontroller boards like Arduino, raspberry pi etc.

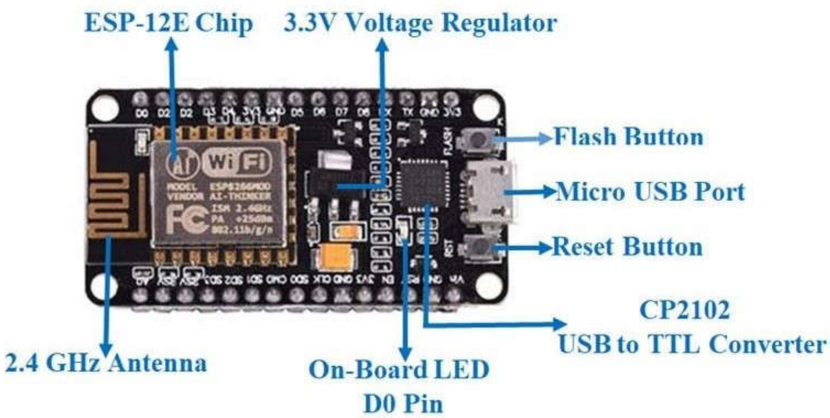

Fig 6: NodeMCU board (Components101, April 2020)

CONCLUSION

The Smart Crop Recommendation System stands as a testament to the transformative potential of technology in agriculture. Through the amalgamation of IoT sensors, machine learning algorithms, and user-centric design, this project offers a dynamic solution to the age-old challenge of optimal crop selection based on soil composition and environmental factors. By harnessing real-time data on soil parameters and climatic conditions, this system empowers farmers with actionable insights, facilitating informed decision-making in crop cultivation. The seamless integration of IoT sensors enables the continuous collection of data, feeding into a sophisticated recommendation engine that generates personalized crop suggestions. This project not only represents a leap towards precision agriculture but also underscores the significance of leveraging data-driven approaches to enhance agricultural productivity. The scalability, adaptability, and user-friendly interface of the system mark a promising trajectory towards sustainable farming practices. As agriculture faces escalating challenges, the Smart Crop Recommendation System offers a beacon of hope, showcasing the potential for technology to revolutionize farming practices. With further refinements, continuous enhancements, and broader adoption, this system paves the way for a more efficient, sustainable, and yield-maximizing future in agriculture.

REFERENCES

[1] “Crop recommendation using Machine learning techniques”-Shafiulla sharif, Shwetha R.B., Ramya O.G., Pushpa H. , Pooja H.R. – 2022.

[2] “Applying big data for Intelligent agriculture based crop selection analysis”-Fan-Hsun Tseng, Hsin-Hung Cho, Hsin-Te Wu-September 2019.

[3] "Artificial Intelligence Technology in the Agricultural Sector: A Systematic Literature Review"-Ersin Elbasi, Nour Mostafa , Zakwan Alternout -2022

[4] https://scholar.google.com/citations?user=99wmG2IAAAAJ&hl=en

[5] “Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers”-S. P. Raja, Barbara Sawicka, Zoran Stamenkovic, G. Mariammal-March 2022.

[6] "An IoT Based System for Remote Monitoring of Soil Characteristics"-Abdullah Na, William Isaac, Shashank Varshney, Ekram Khan-2016

[7] "Development of a Low-Cost Internet-of-Things (IoT) System for Monitoring Soil Water Potential Using Watermark 200SS Sensors"-José O. Payero, Ali Mirzakhani-Nafchi, Ahmad Khalilian, Xin Qiao, Rebecca Davis-2017.

[8] "Big Data in food and agriculture"-Kelly Bronson and Irena Knezevic-2016.

[9] "Crop and Yield Prediction Model"-Shreya S. Bhanose , Kalyani A. Bogawar , Aarti G. Dhotre , Bhagyashree R. Gaidhani-2016.

[10] "Soil conditions and plant growth"-J. B. Passioura-2020.

[11] https://orcid.org/my-orcid?orcid=0000-0002-9764-6048

[12] https://ieeexplore.ieee.org/author/614775320328834