**Agri-GNN: Advanced Graph Neural Network Framework for Optimized Yield Prediction**

Team Members:  
1. Grandhi Shriya (23P81A0519)  
2. Vemula Abhigna (23P81A0562)  
3. Abbani Manasa (23P81A6901)  
4. Jakinalapally Babitha (23P81A0523)  
5. Karedla Jahnavi (23P81A6927)  
  
Mentor: Dr. A. Shilpa Gupta  
College: Keshav Memorial College of Engineering

## Abstract:

Agriculture plays a crucial role in sustaining the global economy and ensuring food security. However, the increasing effects of climate change, soil degradation, and unpredictable weather patterns make accurate yield prediction a challenging task. This project introduces an advanced Graph Neural Network (GNN)-based model designed to optimize crop yield prediction by effectively capturing relationships between various agricultural parameters such as soil type, climate, crop variety, and regional conditions.

## Problem Statement:

In an era characterized by escalating climate change, which is resulting in unpredictable weather patterns and increasing environmental stresses, the agricultural sector faces significant challenges. Unforeseen climatic events such as droughts, floods, and extreme temperatures are impacting crop yields, highlighting the imperative for advanced, precise, and resilient crop yield prediction models. Amidst this backdrop of climatic uncertainties, the necessity for accurate and comprehensive crop yield predictions is more critical than ever.  
  
The agricultural ecosystem is inherently complex and interconnected, with numerous factors playing a pivotal role in determining crop yields. Traditional machine learning models, while effective in handling structured tabular data, often treat each data sample independently and fail to capture interrelations between agricultural features such as soil type, rainfall, and crop characteristics. Therefore, a more advanced approach that can model these relationships holistically is needed to improve yield prediction accuracy.

## Proposed System :

## The proposed system employs a Graph Neural Network (GNN) architecture to represent agricultural datasets as graphs, where nodes represent farms or regions, and edges denote similarity based on soil, rainfall, and crop characteristics. The model processes these relationships to predict yield values with improved accuracy. The data preprocessing involves encoding categorical features, normalization, and graph construction. The GNN is then trained using node embeddings and aggregation functions to learn complex interactions among agricultural features.

## Traditional Machine Learning vs Graph Neural Network:

Traditional machine learning models treat each data record as an independent data point without considering relationships among them. In agriculture, this means that each farm or region is analyzed separately, ignoring dependencies such as soil or climate similarities between nearby regions. This independence assumption limits model accuracy and generalization.  
  
In contrast, Graph Neural Networks (GNNs) can capture the relational nature of agricultural data. Each farm is represented as a node, and edges represent connections based on soil, rainfall, or crop similarities. The model learns from both the farm's own data and its neighboring nodes to make context-aware yield predictions. This relational learning improves robustness and prediction accuracy, especially for large and complex datasets.

## GraphSAGE and Its Implementation:

GraphSAGE (Graph Sample and Aggregate) is a powerful variant of GNN that enables inductive learning on large graphs. Instead of using the entire graph during training, GraphSAGE samples a fixed number of neighboring nodes and aggregates their information. This makes it efficient and scalable for real-world agricultural datasets.  
  
In our project, GraphSAGE was used to aggregate neighboring node features like soil type, rainfall, and crop data to generate meaningful embeddings. These embeddings were then passed through multiple layers to predict crop yield outcomes efficiently, enabling the model to generalize to new data points that were not present during training.

## Technologies Used:

-Programming Language: Python

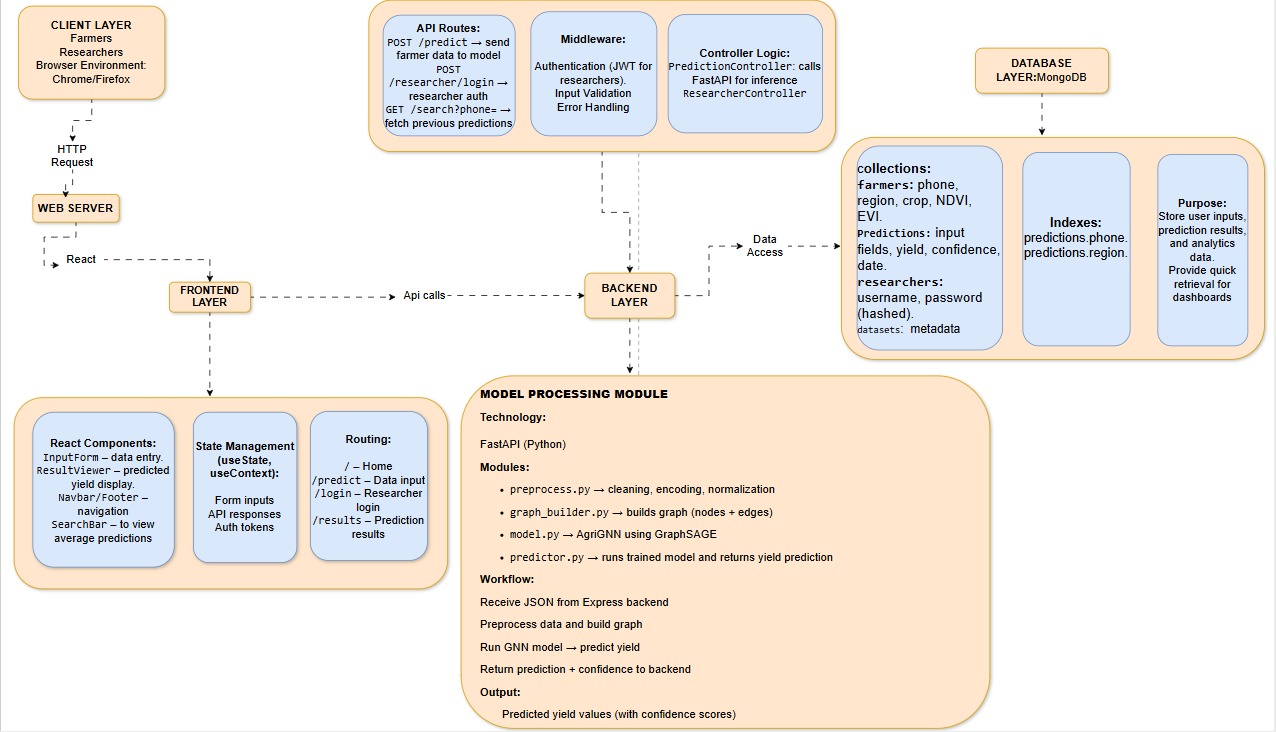
- Frameworks/Libraries: PyTorch, NumPy, Pandas, Scikit-learn

- Data Handling: CSV datasets with preprocessing and normalization

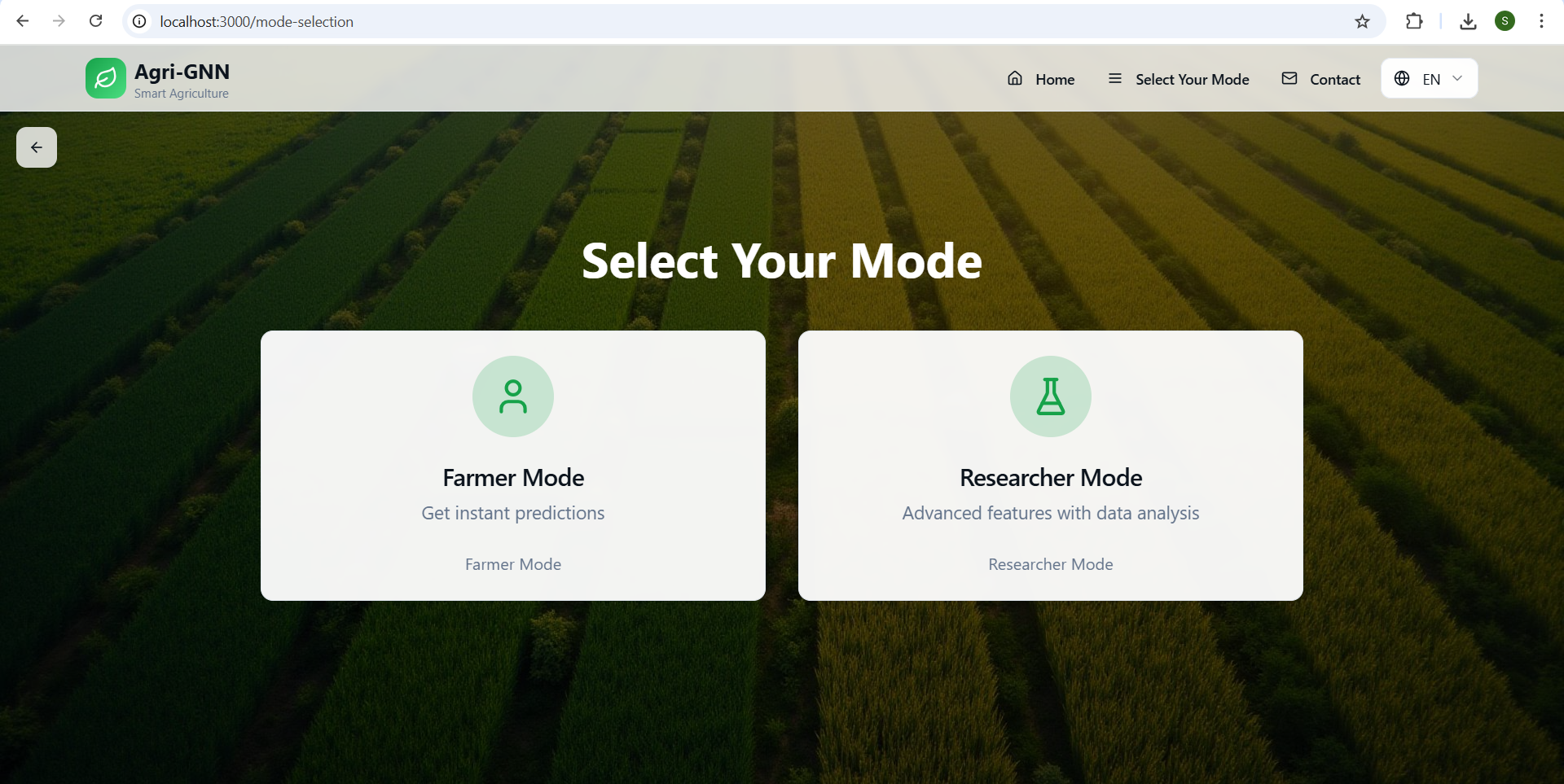
- Visualization: Matplotlib

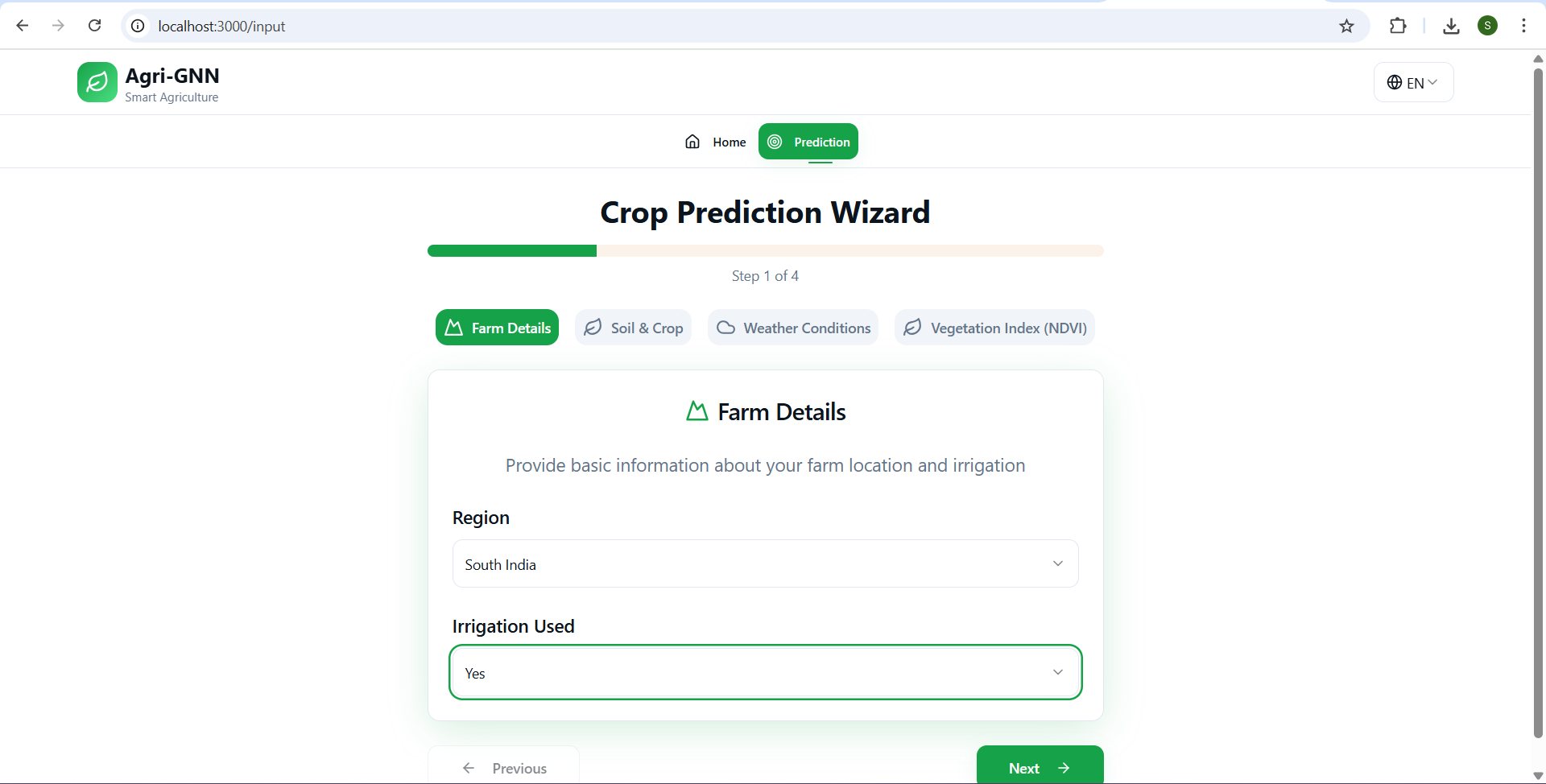
- Development Environment: Google Colab

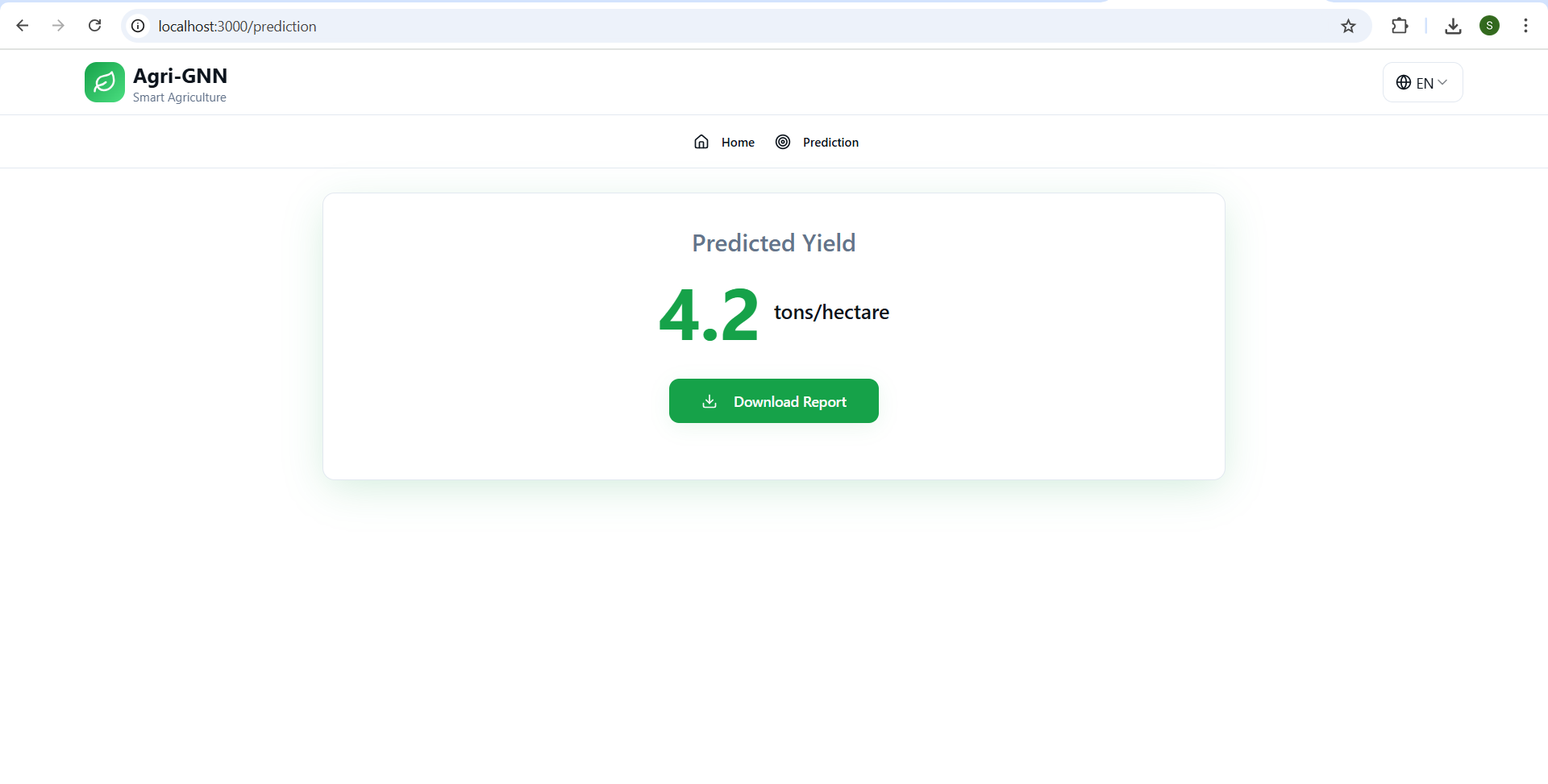
## Architecture Diagram:

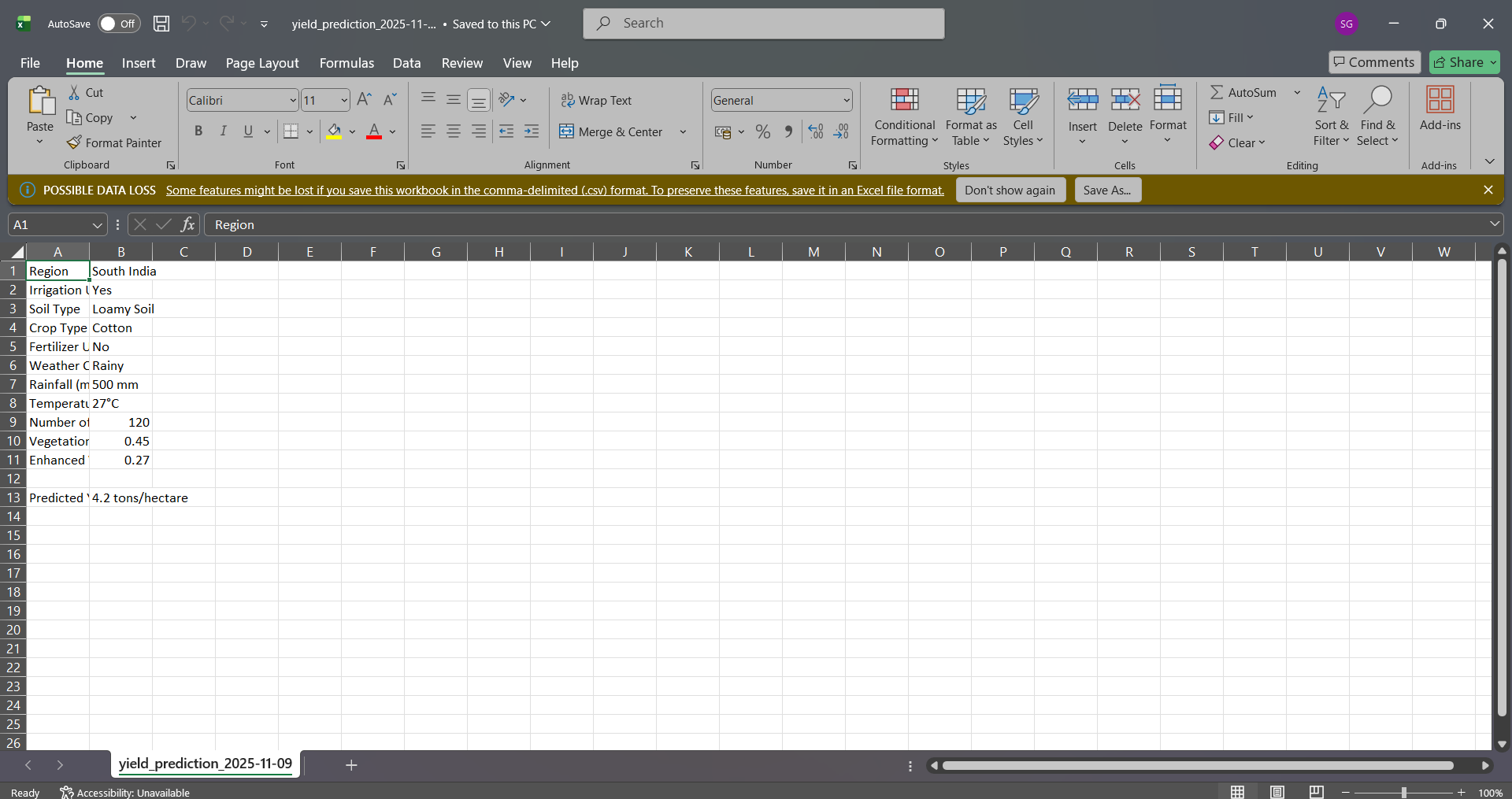


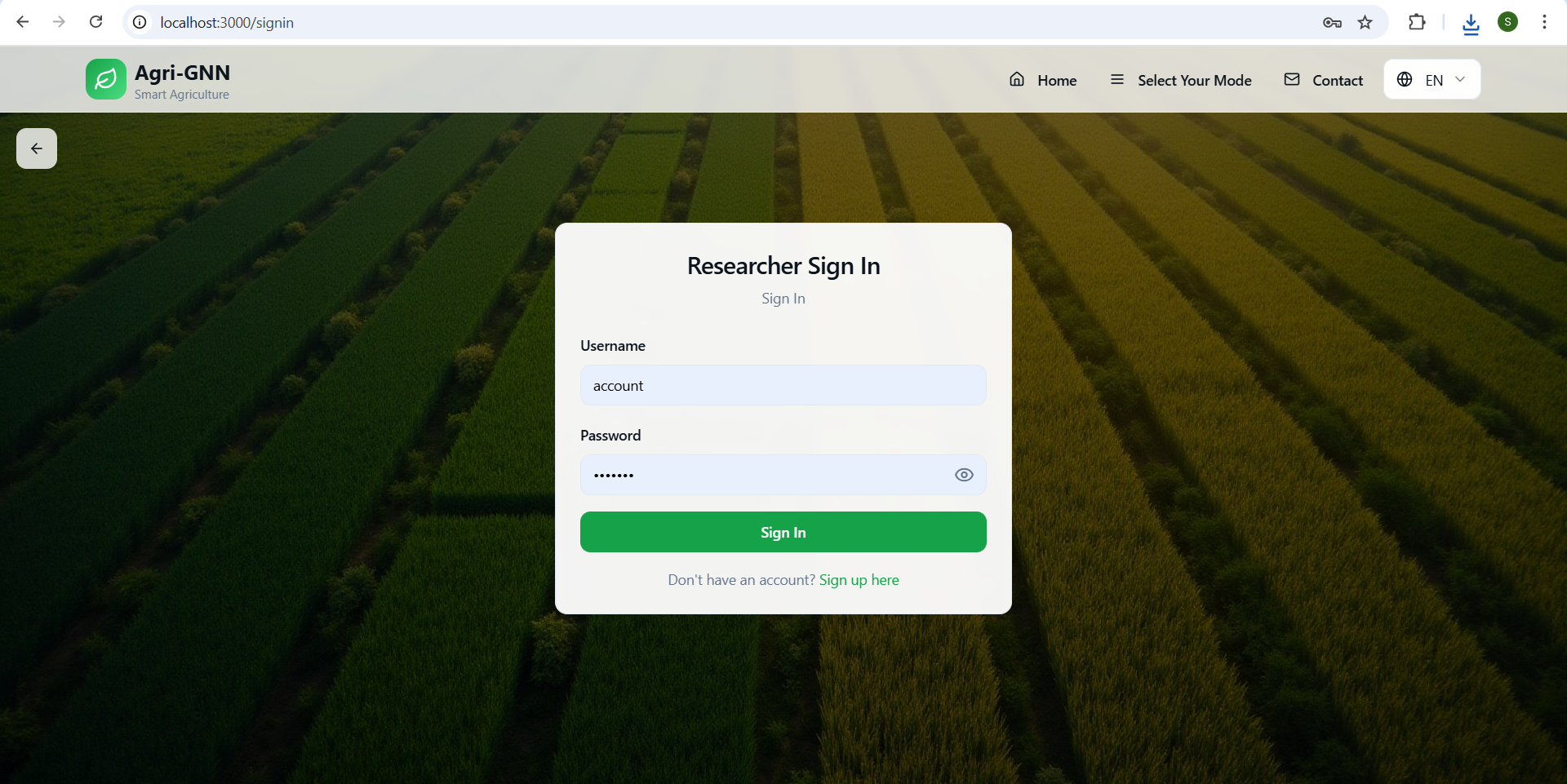
## 













## Mathematical Foundation:

Our model uses Graph Neural Networks (GNN) with the GraphSAGE algorithm.  
  
Each node represents a farm, and edges link farms with similar soil or weather conditions.  
  
Main Formula:  
h\_v^k = sigma(W^k \* AGGREGATE({h\_v^(k-1), h\_u^(k-1), u in N(v)}))  
  
This combines information from nearby farms to predict yield more accurately.  
  
Activation Function: ReLU -> f(x) = max(0, x)  
It helps the model learn faster and ignore negative values.  
  
Loss Function:  
MSE = (1/n) \* SUM(y\_i - y\_pred\_i)^2  
Smaller values indicate better accuracy.  
  
Optimizer: Adam -> adjusts learning rate automatically for efficient training.  
  
Result:  
The model learns relationships between climate, soil, and crops, and provides accurate yield predictions even when some data is missing.

## Improvements Over the Paper:

In our project, we made several improvements over the base paper:  
  
- We added NDVI and EVI features to make the model understand vegetation better.  
- The model was first trained on 10,000 records, and we plan to expand it to 10 lakh rows for better learning.  
- We developed a complete web application with two logins: farmers (no authentication) and researchers (secure login).  
- A clear dashboard was created to show predicted yield and confidence scores.  
- We used MongoDB for data storage and Docker for smooth deployment.  
- All modules were connected using FastAPI and Express.js.  
  
These improvements made our system more practical, accurate, and easy to use for both farmers and researchers.

## Results:

The Agri-GNN model successfully predicts yield values with higher accuracy than conventional methods. The inclusion of NDVI and EVI enhanced vegetation understanding, and GraphSAGE improved learning efficiency. The model achieved low Mean Squared Error (MSE) and high R^2 scores, confirming its reliability for yield estimation. Additionally, the integrated web application allows farmers and researchers to interact with predictions easily and view confidence scores for informed decision-making.