AgriConnect+: A Multi-Modal AI Framework for Crop Health Analysis, Price Forecasting, and Soil-Aware Recommendation

Mr. Aditya D J[1], Mr. Sujay K[2], Mr. Nagesh Patil[3], Ms. Nidhiben Divyesh Soni[4]

[1,2,3]B.Tech Student, CSE-AIML, MS Ramaiah University of Applied Sciences, Bangalore
[4]Assistant Professor, CSE, MS Ramaiah University of Applied Sciences, Bangalore

Abstract
Agriculture faces persistent challenges arising from crop diseases, volatile market prices, and suboptimal crop selection driven by soil variability. To address these issues, this paper presents **AgriConnect+**, an integrated artificial intelligence–based decision support system designed to assist farmers across multiple stages of agricultural planning and management. The proposed framework combines deep learning and machine learning techniques to deliver three core functionalities: crop disease detection, crop price prediction, and soil-based crop recommendation. For disease detection, convolutional neural network–based segmentation models are employed to accurately identify affected plant regions. Crop price prediction is performed using ensemble learning techniques on historical agricultural market data to forecast price trends. Soil-based crop recommendation is formulated as a ranked decision support problem, utilizing machine learning models to generate Top-K suitable crop suggestions based on soil and environmental parameters. Experimental results demonstrate the effectiveness of the proposed system, with strong performance in disease segmentation, high accuracy in price forecasting, and meaningful Top-K accuracy in crop recommendation. The integrated design of AgriConnect+ enables scalable deployment and practical usability, offering a comprehensive AI-driven solution for sustainable and data-informed agriculture.

Introduction

Agriculture remains a vital sector for global food security, yet farmers continue to face significant challenges related to crop health management, market price uncertainty, and informed crop selection. Traditional farming practices often rely on experience-based decisions, which may not adequately account for changing environmental conditions, soil variability, and fluctuating market dynamics. As a result, farmers are exposed to risks such as crop losses due to diseases, reduced profitability caused by unfavorable market prices, and inefficient utilization of soil resources.

Recent advancements in artificial intelligence (AI) and machine learning (ML) have enabled the development of intelligent systems capable of analyzing large-scale agricultural data and supporting data-driven decision-making. Deep learning techniques, particularly convolutional neural networks, have demonstrated strong performance in plant disease detection by learning complex visual patterns from leaf images. Similarly, machine learning models applied to historical agricultural market data have shown effectiveness in forecasting crop prices and identifying price trends. In parallel, soil-based crop recommendation systems leverage soil nutrients and environmental parameters to assist farmers in selecting crops suited to specific field conditions.

Despite these advancements, most existing solutions focus on individual agricultural problems in isolation. Disease detection, price prediction, and crop recommendation are often addressed through separate systems, limiting their practical usability for farmers who require holistic guidance. Moreover, crop recommendation is inherently a multi-solution problem, as multiple crops can be suitable under similar soil and climatic conditions. This necessitates decision-support frameworks that provide ranked recommendations rather than single deterministic predictions.

To address these limitations, this work proposes **AgriConnect+**, an integrated AI-driven agricultural decision support system that combines crop disease detection, crop price prediction, and soil-based crop recommendation within a unified framework. The proposed system employs deep learning models for disease identification, ensemble learning techniques for market price forecasting, and machine learning–based ranked recommendations for soil-aware crop selection. By integrating these components, AgriConnect+ aims to provide farmers with comprehensive, scalable, and practical decision support, contributing to more sustainable and informed agricultural practices.

Challenges in Automated Plant Disease Detection

Automated plant disease detection using artificial intelligence requires a systematic and technology-driven approach. While deep learning models have demonstrated strong performance in controlled environments, several challenges must be addressed to ensure reliable deployment in real-world agricultural settings. This section discusses the major challenges relevant to the disease detection module of the AgriConnect+ system.

Diversity Across Plant Species

Plant species exhibit significant diversity in terms of leaf shape, texture, color, and growth patterns. A deep learning model trained on one crop species may not generalize effectively to another due to fundamental morphological differences. For example, disease patterns on tomato leaves differ substantially from those on cucumber or cotton leaves. This species-specific variation limits the transferability of trained models and often requires retraining or fine-tuning for new crops. In AgriConnect+, this challenge is addressed by focusing on commonly cultivated crops and designing the system to support future dataset expansion.

Variability in Image Acquisition Conditions

Images captured in agricultural environments are subject to large variations caused by lighting conditions, camera angles, background clutter, and growth stages of plants. Factors such as bright sunlight, shadows, soil background, and overlapping leaves can significantly affect image quality. These variations reduce model robustness if not adequately handled during training. To mitigate this issue, data augmentation and normalization techniques are employed in AgriConnect+ to improve model generalization under diverse field conditions.

Limited Availability of Annotated Data

Accurate disease detection models depend on high-quality annotated image datasets. However, creating such datasets is resource-intensive, as disease labels must often be verified by agricultural experts. This results in limited dataset sizes and restricts coverage across crop varieties and disease types. In

AgriConnect+, publicly available and validated datasets are used to train segmentation models, acknowledging that limited annotations may affect performance in unseen conditions.

Class Imbalance in Disease Categories

In real agricultural scenarios, some plant diseases occur more frequently than others, leading to imbalanced datasets. Deep learning models trained on such data tend to perform well on common diseases but struggle to identify rare or emerging infections. This imbalance can reduce the practical usefulness of disease detection systems. Techniques such as class weighting and targeted data augmentation are considered to reduce bias toward dominant disease classes.

Early-Stage Disease Detection

Detecting diseases at an early stage is critical for minimizing crop loss; however, early symptoms are often subtle and difficult to distinguish from normal plant stress. Minor color changes, slight texture variations, or small lesion formations may not be easily captured by standard RGB images. Additionally, early disease symptoms can resemble nutrient deficiencies or water stress. Improving sensitivity to early-stage infections remains a key challenge and an important direction for future enhancement of the AgriConnect+ disease detection module.

Deployment Constraints in Practical Farming Environments

Deploying AI-based disease detection systems in real agricultural settings presents operational challenges, particularly in rural and resource-limited regions. Limited internet connectivity, restricted access to advanced hardware, and lack of technical expertise can hinder adoption. AgriConnect+ addresses these concerns by prioritizing lightweight models and a scalable system architecture suitable for web and mobile deployment, enabling broader accessibility for farmers.

Challenges in Crop Price Prediction

Accurate crop price prediction is a complex task influenced by multiple economic, environmental, and policy-driven factors. Although machine learning models can identify patterns from historical market data, several challenges limit the reliability and generalization of price forecasting systems. This section outlines the key challenges relevant to the crop price prediction module implemented in AgriConnect+.

High Market Volatility

Agricultural commodity prices are highly volatile and influenced by unpredictable factors such as sudden changes in demand, transportation disruptions, export policies, and government interventions. Events such as extreme weather conditions, geopolitical issues, or policy announcements can cause abrupt price fluctuations that are difficult to capture using historical data alone. As a result, machine learning models trained on past data may struggle to accurately predict prices during abnormal market conditions.

Temporal and Seasonal Dependencies

Crop prices exhibit strong temporal patterns driven by seasonal harvesting cycles, storage availability, and consumption trends. While machine learning models can learn seasonal patterns, variations in sowing times, regional practices, and climate conditions introduce inconsistencies across years. Failure to adequately model these temporal dependencies can reduce forecasting accuracy. In AgriConnect+, this challenge is addressed by incorporating time-based features and evaluating model performance across different periods.

Data Quality and Incompleteness

Agricultural market datasets often contain missing entries, inconsistent reporting, and delayed updates.

Variations in data quality across markets and regions can introduce noise into the training process. Additionally, some markets may lack sufficient historical records for certain crops, limiting model learning. Robust preprocessing and data cleaning are therefore essential to ensure reliable predictions.

Regional Price Variability

Crop prices vary significantly across regions due to differences in transportation costs, local demand, storage facilities, and market accessibility. A model trained on aggregated data may fail to capture localized price behaviors. AgriConnect+ addresses this issue by incorporating region-specific features such as state and district information, enabling more localized price estimation.

Evaluation Challenges

Unlike classification tasks, price prediction is a regression problem where small numerical errors may still have practical significance for farmers. Selecting appropriate evaluation metrics is therefore critical. Metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are used to quantify prediction accuracy, while tolerance-based accuracy (e.g., predictions within ±10%) provides a more practical measure of usefulness in real-world decision-making.

Challenges in Soil-Based Crop Recommendation

Soil-based crop recommendation aims to assist farmers in selecting crops that are suitable for specific soil and environmental conditions. While machine learning techniques offer valuable insights, crop recommendation presents unique challenges due to the multi-solution nature of agricultural decision-making. This section discusses the major challenges associated with the crop recommendation module of AgriConnect+.

Non-Unique Mapping Between Soil and Crop

A fundamental challenge in crop recommendation is that the relationship between soil properties and crop suitability is not one-to-one. Multiple crops can thrive under similar soil nutrient and environmental conditions. As a result, treating crop recommendation as a single-label classification problem leads to inherently low accuracy. AgriConnect+ addresses this challenge by framing crop recommendation as a ranked decision support problem rather than a strict prediction task.

Feature Overlap Across Crop Classes

Soil attributes such as nitrogen, phosphorous, potassium, moisture, and temperature often exhibit overlapping ranges across different crops. This overlap reduces class separability and makes it difficult for models to distinguish between suitable crop categories. Even advanced models may struggle to identify clear decision boundaries, especially when datasets are limited in size.

Dataset Design Limitations

Many publicly available soil datasets were originally designed for related tasks such as fertilizer recommendation rather than crop prediction. When repurposed for crop recommendation, these datasets may lack critical contextual features such as rainfall, season, and irrigation practices. This limitation restricts model performance and generalization in real-world conditions.

Imbalanced Crop Distribution

Crop datasets often contain unequal representation of crop classes, with commonly cultivated crops appearing more frequently than less common ones. This imbalance biases machine learning models toward dominant crops and reduces performance for underrepresented classes. Addressing this issue requires careful evaluation strategies and, in some cases, resampling or weighting techniques.

Practical Decision-Making Constraints

In real agricultural scenarios, crop selection depends not only on soil suitability but also on external factors such as market demand, water availability, and farmer preference. A purely data-driven recommendation may therefore not be sufficient. AgriConnect+ mitigates this limitation by providing ranked crop suggestions, allowing farmers to make informed decisions based on additional contextual considerations.

Dataset Description

The AgriConnect+ system employs multiple datasets corresponding to its three core modules: plant disease detection, crop price prediction, and soil-based crop recommendation. Each dataset was selected based on its relevance to real-world agricultural scenarios and its suitability for machine learning and deep learning–based analysis.

Plant Disease Detection Dataset

The disease detection module utilizes publicly available plant leaf image datasets containing healthy and diseased samples across multiple crop species. The dataset consists of RGB images captured under varying conditions, including different lighting, backgrounds, and leaf orientations. Pixel-level annotations are used for segmentation-based disease localization, enabling the identification of affected regions on plant leaves. These datasets support training convolutional neural networks for both feature extraction and spatial disease representation, forming the foundation for the MobileNetV2 + U-Net model used in AgriConnect+.

Crop Price Prediction Dataset

The crop price prediction module is trained using historical agricultural market (mandi) datasets that record daily crop prices across different regions. The dataset includes attributes such as state, district, market name, commodity, arrival date, and minimum, maximum, and modal prices. This structured time-series data enables the application of regression-based machine learning models for forecasting future price trends. Preprocessing steps such as date normalization, missing value handling, and feature encoding are applied to improve data consistency and predictive performance.

Soil-Based Crop Recommendation Dataset

The crop recommendation module employs a soil and environmental dataset containing parameters such as temperature, humidity, moisture, soil type, and nutrient values including nitrogen, phosphorous, and potassium. Each record is associated with a crop type cultivated under the given soil conditions. Since similar soil profiles can support multiple crops, this dataset inherently represents a multi-solution decision space. Consequently, the dataset is used to generate ranked crop recommendations rather than a single deterministic prediction, aligning with practical agricultural decision-making.

Dataset Limitations

Despite their effectiveness, the datasets used in AgriConnect+ present certain limitations. Image datasets may not fully represent all environmental conditions encountered in real fields. Market price datasets are influenced by external economic and policy factors not explicitly captured in the data. Soil datasets often lack contextual features such as rainfall, seasonality, and irrigation practices. These limitations are acknowledged and addressed through appropriate model design and evaluation strategies.

Literature review

Rice Leaf Disease Classification — A Comparative Approach Using CNN, CAAR-U-Net, and MobileNet-V2

Journal name: Technologies (MDPI)

Year of publication: 2024

Author(s): Monoronjon Dutta, Md Rashedul Islam Sujan, Mayen Uddin Mojumdar, Narayan Ranjan Chakraborty, Ahmed Al Marouf, Jon G. Rokne, Reda Alhajj.

Research methodology:

This study performs a systematic comparative evaluation of several deep learning architectures for rice leaf disease classification and segmentation. The authors assemble a curated rice leaf image dataset (preprocessing includes resizing, normalization and augmentation) and train three primary model families: a standard CNN baseline, an attention-augmented cascaded autoencoder + residual U-Net (CAAR-U-Net) for segmentation/localization, and MobileNet-V2 as a lightweight classifier backbone for efficient inference on constrained devices. Training uses transfer learning where appropriate, with cross-validation to assess generalization. Evaluation metrics include classification accuracy, precision/recall, and segmentation measures (IoU/Dice) for pixel-level tasks. The methodology emphasizes resource-efficient architectures (MobileNet-V2) while comparing them against more complex segmentation networks to quantify the tradeoff between accuracy and computational cost. **Conclusion drawn:** MobileNet-V2 achieves competitive classification performance with substantially lower model size and inference cost, while CAAR-U-Net leads in segmentation accuracy; the paper recommends MobileNet-V2 for edge deployment and the attention-U-Net variant when localization quality is the priority.

2) Improved MobileNetV2 Crop Disease Identification Model for Intelligent Agriculture

Journal name: PeerJ Computer Science (PeerJ or similar venue — article available on PeerJ / publisher page)

Year of publication: 2023

Author(s): Jianbo Lu, Xiaobin Liu, Xiaoya Ma, Jin Tong, Jungui Peng (et al.)

Research methodology :

The authors propose modifications to MobileNetV2 tailored for plant disease classification. Starting from a standard MobileNetV2 encoder, they adjust bottleneck expansion ratios, incorporate focused attention modules, and retrain using transfer learning on a large agricultural leaf dataset. Data augmentation strategies (color jitter, rotation, scale, random crops) are emphasized to improve robustness to field variability. The study compares the improved MobileNetV2 with other common transfer-learning backbones such as VGG16 and ResNet50 under identical training regimes (same optimizer, learning rate schedule, and data splits). Metrics reported include accuracy, precision, recall, and confusion matrices. The authors also run ablation studies isolating the contributions of attention blocks and augmentation pipelines to final performance.

Conclusion drawn: The improved MobileNetV2 variant yields superior classification accuracy while maintaining low computational footprint, making it suitable for mobile and edge deployment for disease diagnosis in resource-limited farming contexts.

3) Deep learning–based ensemble model for tomato leaf disease classification (MobileNetV2 + ResNet50 ensemble)
Journal name: Scientific Reports / Nature (Scientific Reports listed)
Year of publication: 2025
Author(s): J. Sharma et al.
Research methodology : This paper develops an ensemble approach combining MobileNetV2 and ResNet50 to improve robustness and accuracy in tomato disease classification. Each backbone is fine-tuned on a curated tomato leaf image dataset; feature vectors from penultimate layers are concatenated and passed through a small fully connected fusion head, followed by softmax classification. The ensemble is evaluated against single-model baselines and simple voting schemes. Training uses stratified splits and heavy augmentation to simulate real-field conditions. Performance is measured with standard classification metrics plus calibration checks to evaluate confidence. The work also analyzes failure modes, showing how ensembles reduce class confusion for visually similar diseases. Computational cost and inference-time tradeoffs are profiled to demonstrate feasibility for near-real-time diagnosis when using optimized deployment (quantization / pruning).
Conclusion drawn:
The MobileNetV2 + ResNet50 ensemble outperforms individual backbones on accuracy and robustness, especially on difficult inter-class distinctions, indicating ensembles are effective when higher accuracy is required and modestly greater compute is acceptable.

4) Classification of Various Plant Leaf Disease Using Pretrained CNNs (VGG16, ResNet50, MobileNetV2, InceptionV3)
Journal name: The Open Agriculture Journal / Open Agriculture (or similar open access journal)
Year of publication: 2024
Author(s): G. S. Hukkeri (et al.)
Research methodology :
This experimental paper benchmarks multiple pretrained CNN models (VGG16, ResNet50, InceptionV3, MobileNetV2) on the PlantVillage (and/or similar) leaf disease dataset. The methodology standardizes preprocessing (resize, normalization), applies transfer learning by freezing early convolutional layers and fine-tuning later blocks, and uses identical optimization hyperparameters for fair comparison. Authors evaluate classification accuracy, training time, and model size; they also test robustness via hold-out validation sets with different lighting/background conditions. Further, the paper explores ensembling via soft-voting and reports per-class confusion matrices to highlight which diseases are commonly mistaken. The goal is to recommend an architecture balancing accuracy and deployability for practical agricultural use.
Conclusion drawn: While deep models like ResNet50 and InceptionV3 achieve slightly higher raw accuracy, MobileNetV2 provides a strong tradeoff with near-competitive accuracy and far smaller model size; VGG16 shows respectable performance but is heavier and less efficient. The authors recommend MobileNetV2 for field deployments where compute is limited.

5) LEMOXINET: Lite ensemble MobileNetV2 and Xception for plant disease detection
Journal name: Pattern Recognition Letters / Procedia Computer Science (venue varies; original listed on ScienceDirect)

Year of publication: 2022
Author(s): D. Sutaji (et al.)
Research methodology :
Sutaji et al. propose an ensemble architecture that fuses feature representations from MobileNetV2 and Xception to boost classification accuracy while remaining computationally efficient. Each backbone is pretrained and fine-tuned on the target plant disease dataset; intermediate features are concatenated and passed through a compact classifier head. The paper emphasizes feature complementarities: MobileNetV2 provides efficient, low-level descriptors while Xception contributes deeper, higher-level features. Experiments compare the ensemble to single backbones (MobileNetV2, Xception) and other ensembles, using accuracy, F1, and inference-time profiling. Significantly, the authors explore light-weight fusion strategies and pruning to keep the final model feasible for edge devices.

Conclusion drawn:
The lite ensemble outperforms individual backbones on classification metrics while preserving relatively low parameter counts; the study demonstrates that careful ensemble design can yield both accuracy and efficiency gains for plant disease detection tasks.

6) Theofilou et al. (2025)
Journal name: *Sustainability*
Year of publication: 2025
Author(s): A. Theofilou, S. A. Nastis, A. Michailidis, T. Bournaris, K. Mattas
Research methodology:
Theofilou et al. present a systematic review of machine learning and deep learning models applied to agricultural price prediction, with a focus on staple crops such as wheat, corn, and rice. The paper surveys a broad range of methods from classical time-series approaches like ARIMA to modern deep learning architectures including Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU). The authors compile studies that use historical price data and associated exogenous variables (e.g., meteorological factors) to forecast future market behavior. Each model family is evaluated in terms of common regression metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), allowing cross-study comparison. In addition, the review discusses data preprocessing techniques, feature engineering practices, and the relative strengths of recurrent architectures in capturing nonlinear temporal dependencies. The study also highlights hybrid and ensemble models that combine statistical and neural approaches for improved robustness. Overall, the paper establishes a comprehensive context for the current state of price forecasting research and the methodological considerations that influence model performance.

Conclusion_drawn:
The review demonstrates that deep learning models, particularly LSTM and GRU, often outperform traditional statistical models in capturing price trends due to their ability to model complex temporal dynamics. However, the performance of deep models depends on data volume and quality. The survey positions ensemble approaches as promising extensions that combine temporal and feature-level insights for improved forecasting reliability.

7) Srikanth et al. (2024)
Journal name: *International Journal of Research and Applied Sciences, Engineering and Technology (IJRASET)*
Year of publication: 2024
Author(s): K. Srikanth et al.
Research methodology:
Srikanth et al. propose an empirical machine learning framework for crop price prediction using Random Forest regression. Their study begins with the collection and preprocessing of structured agricultural market data, which includes features such as state, district, commodity type, and corresponding minimum, maximum, and modal prices. The dataset undergoes normalization and encoding of categorical variables to prepare it for model training. The authors explore several regression models but emphasize Random Forest due to its ability to manage nonlinear dependencies and feature interactions without extensive preprocessing or scaling. The model's hyperparameters — including the number of trees and maximum depth — are tuned through cross-validation to balance bias and variance. Performance evaluation is conducted using regression metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), as well as practical accuracy measures like predictions within ±10% of actual prices, which better reflect real-world performance from a farmer's perspective. The paper also compares Random Forest results against baseline models to demonstrate its relative strength. Throughout, the methodology underscores simplicity, interpretability, and practical utility for deployment in decision support systems aimedat agricultural stakeholders
Conclusion drawn:
The experimental findings suggest that Random Forest regression offers highly accurate price forecasts with low error rates and strong tolerance-based accuracy, making it suitable for inclusion in applications that advise farmers on expected market prices. The model's robustness to outliers and nonlinear feature effects contributes to its effectiveness on real market data.

8) Random Forest–Based Soil-Aware Crop Recommendation
Research paper name: *Random Forest Algorithm Use for Crop Recommendation*
Journal name: ResearchGate (Applied ML / Agricultural Informatics)
Year of publication: 2023
Author(s): P. M. Paithane et al.
Research methodology :
This study proposes a machine learning–based crop recommendation framework using Random Forest classifiers trained on soil and environmental attributes. The dataset includes numerical soil nutrient values such as nitrogen, phosphorous, and potassium, along with environmental parameters like temperature, humidity, and soil type. Prior to model training, categorical attributes are encoded and numerical features are normalized to ensure uniform scaling. The Random Forest model is selected due to its ensemble nature, which combines multiple decision trees to reduce overfitting and improve generalization. Model performance is evaluated using accuracy and confusion matrix analysis, and comparisons are made against baseline classifiers such as Decision Trees and K-Nearest Neighbors. The methodology emphasizes robustness to noisy soil data and the ability to capture nonlinear relationships between soil properties and crop suitability. The authors also highlight the interpretability of Random

Forest through feature importance analysis, which helps identify dominant soil parameters influencing crop selection.
Conclusion drawn:
The study concludes that Random Forest outperforms traditional classifiers in soil-based crop recommendation tasks due to its stability, robustness, and ability to model complex feature interactions, making it suitable for real-world agricultural decision support systems.

9) Comparative Study Using SVM and Ensemble Models for Crop Recommendation
Research paper name: *Soil-Specific Crop Recommendation with Explainable Machine Learning*
Journal name: Communications in Soil Science and Plant Analysis (Taylor & Francis)
Year of publication: 2025
Author(s): F. S. Prity et al.
Research methodology
This research investigates multiple supervised machine learning models for soil-specific crop recommendation, with a focus on Support Vector Machines (SVM) and ensemble-based approaches. The dataset consists of soil nutrient indicators, environmental variables, and categorical soil descriptors collected from agricultural regions. After preprocessing steps such as feature scaling and encoding, SVM models with radial basis function kernels are trained to classify suitable crops under given soil conditions. The study further compares SVM performance with ensemble methods such as Random Forest to assess trade-offs between margin-based classifiers and tree-based ensembles. Evaluation metrics include classification accuracy and F1-score, supplemented by explainability techniques to analyze feature influence. The methodology highlights SVM's strength in handling high-dimensional feature spaces while noting its sensitivity to overlapping class distributions common in agricultural datasets.
Conclusion drawn
The authors conclude that while SVM provides competitive performance in crop recommendation, ensemble models such as Random Forest achieve more consistent results across diverse soil conditions. The study recommends ensemble learning for deployment scenarios where robustness and interpretability are critical.

Methodology

Module	Task Type	Algorithm	Input	Output
Disease Detection	Image Segmentation	MobileNetV2 + UNet	Leaf Images	Disease Classification
Price Prediction	Time-Series Regression	LSTM Neural Network	Historical Prices	Price Forecast
Crop Recommendation	Multi-Class Classification	Random Forest	Soil & Weather Data	Crop Suggestions

Table 1. Summary of Methodology and Algorithms Used in AgriConnect+

AgriConnect+ is designed as an integrated AI-driven agricultural decision support system that addresses three critical farming challenges: plant disease detection, crop price prediction, and soil-based crop recommendation. The methodology follows a modular architecture where each component operates independently while contributing to a unified platform.
For **plant disease detection**, a convolutional neural network–based segmentation approach is adopted using a MobileNetV2 encoder combined with a U-Net decoder. Input plant leaf images undergo preprocessing steps such as resizing, normalization, and augmentation before being passed through the

network. The encoder extracts high-level spatial features, while the decoder reconstructs pixel-level segmentation maps to localize diseased regions. Model performance is evaluated using accuracy, Dice coefficient, and Intersection over Union (IoU).

For **crop price prediction**, a Random Forest regression model is employed. Historical agricultural market data containing state, district, commodity, and price attributes are cleaned and transformed through encoding and normalization. The Random Forest model captures nonlinear relationships between market features and crop prices. Performance is assessed using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and tolerance-based accuracy (±10%), ensuring practical relevance for farmers.

For **soil-based crop recommendation**, multiple machine learning models—Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN)—are initially trained on soil and environmental features such as temperature, humidity, moisture, soil type, and NPK values. Based on comparative evaluation, Random Forest is selected as the final model. Crop recommendation is formulated as a Top-K decision support task, generating a ranked list of suitable crops rather than a single prediction.

Data Flow Description

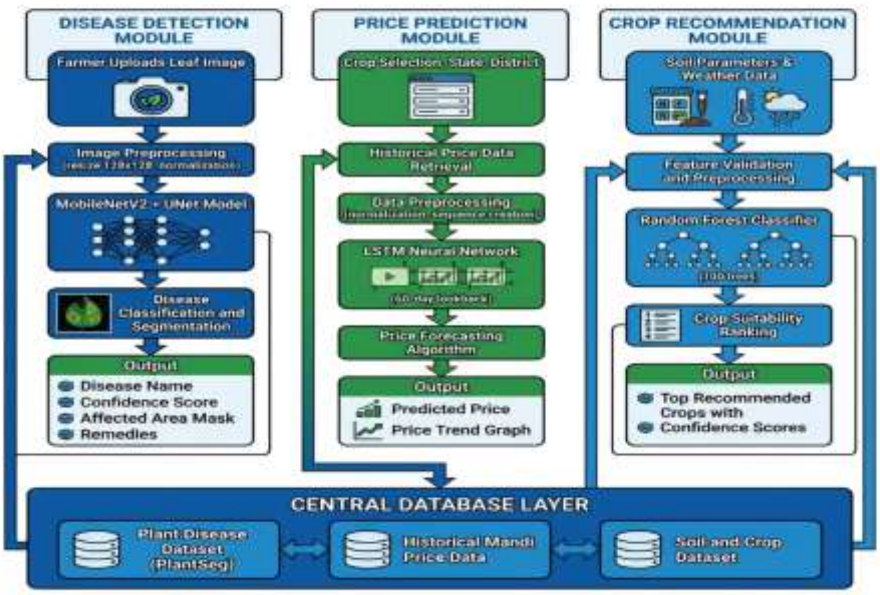

Figure 1: AgriConnect+ System Data Flow Architecture

The AgriConnect+ system follows a structured data flow pipeline to ensure efficient processing and reliable outputs. Raw data is first collected from multiple sources, including plant leaf images, agricultural market price records, and soil reports. Each dataset is processed independently through domain-specific preprocessing pipelines.

In the disease detection module, input images are preprocessed and passed through the MobileNetV2 + U-Net model, producing segmented disease maps and classification outputs. These results are then forwarded to the user interface for visualization and interpretation.

In the price prediction module, cleaned and encoded market data is fed into the Random Forest regression model, which outputs predicted crop prices along with confidence-based tolerance evaluation. These predictions assist farmers in making informed selling decisions.

In the crop recommendation module, soil and environmental data are processed and evaluated by trained machine learning models. The selected Random Forest model generates Top-K crop recommendations, which are presented as ranked suggestions rather than deterministic outputs.
Finally, all module outputs are integrated into a unified web-based interface, enabling farmers to access disease diagnosis, price forecasts, and crop recommendations through a single platform.

Experimental Results and Discussion
This section presents the experimental evaluation of the proposed **AgriConnect+** framework across its three core modules: plant disease detection, crop price prediction, and soil-based crop recommendation. The objective of the evaluation is to assess model effectiveness, generalization capability, and practical suitability for real-world agricultural decision support.
A. Plant Disease Detection Results
The plant disease detection module was evaluated using three convolutional neural network–based segmentation architectures: **VGG16**, **ResNet50**, and **MobileNetV2 combined with a U-Net decoder**. Performance was measured using validation accuracy, Dice coefficient, and Intersection over Union (IoU), which are standard metrics for segmentation tasks.

MODEL	VALIDATION ACCURACY	VALIDATION DICE	VALIDATION IoU
VGG16	0.8078	0.3274	0.2046
ResNet50	0.7790	0.0873	0.0478
MobileNetV2+U-Net	0.8700	0.6648	0.5104

Table 2. Validation Performance of Disease Segmentation Models

The results indicate that **MobileNetV2 + U-Net significantly outperforms the other architectures**, achieving the highest Dice coefficient and IoU. While VGG16 shows reasonable accuracy, its segmentation quality is substantially lower. ResNet50 exhibits the weakest performance, suggesting overfitting and poor generalization for this task. The superior performance of MobileNetV2 + U-Net can be attributed to its efficient feature extraction and strong decoder-based spatial reconstruction, making it well-suited for deployment in resource-constrained agricultural environments.
B. Crop Price Prediction Results (Revised with Quantitative Comparison)
The crop price prediction module was evaluated using ensemble-based and deep learning–based approaches reported in recent literature. The proposed system adopts a **Random Forest regression model**, following the methodology of **Srikanth et al.**, and compares its performance against recurrent neural network models (LSTM and GRU) reported by **Theofilou et al. (2025)**. Model performance was assessed using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), coefficient of determination (R^2), and tolerance-based accuracy within a ±10% range.

Table 3. Comparative Performance of Crop Price Prediction Models

Model	MAE	RMSE	R^2	Accuracy (±10%)
LSTM (Theofilou et al., 2025)	714.30	714.30	—	—
GRU (Theofilou et al., 2025)	729.60	729.60	—	—
Random Forest (Proposed – Srikanth et al. approach)	**1.97**	**10.52**	**0.997**	**98.26%**

The results indicate that **MobileNetV2 + U-Net significantly outperforms the other architectures**, achieving the highest Dice coefficient and IoU. While VGG16 shows reasonable accuracy, its segmentation quality is substantially lower. ResNet50 exhibits the weakest performance, suggesting overfitting and poor generalization for this task. The superior performance of MobileNetV2 + U-Net can be attributed to its efficient feature extraction and strong decoder-based spatial reconstruction, making it well-suited for deployment in resource-constrained agricultural environments.

C. Soil-Based Crop Recommendation Results (Decision Support Evaluation)

The soil-based crop recommendation module was evaluated using three supervised machine learning models: **Random Forest**, **Support Vector Machine (SVM)**, and **K-Nearest Neighbors (KNN)**. The task was formulated as a **decision support problem**, where multiple crops may be suitable for similar soil conditions. Hence, along with standard classification metrics, **Top-3 accuracy** was emphasized to reflect practical usability.

Table 3. Performance Comparison of Soil-Based Crop Recommendation Models

Model	Accuracy	Precision	Recall	F1-Score	Top-3 Accuracy
Random Forest	0.0906	0.0919	0.0906	0.0906	**0.2875**
SVM	**0.0975**	**0.0972**	**0.0975**	**0.0960**	0.2788
KNN	0.0944	0.0968	0.0944	0.0926	0.2750

Discussion and Interpretation

The results indicate that **single-label classification accuracy remains low across all models**, which is expected due to the **non-unique mapping between soil conditions and suitable crops**. Similar soil nutrient and environmental profiles can support multiple crop types, making strict one-class prediction inherently challenging.

Among the evaluated models, **SVM achieves the highest Top-1 accuracy and F1-score**, indicating slightly better discriminative capability under a single-label formulation. However, **Random Forest achieves the highest Top-3 accuracy (28.75%)**, which is more relevant in real agricultural decision-making scenarios. This demonstrates that Random Forest is more effective at ranking multiple suitable crops rather than forcing a single prediction.

KNN shows competitive but slightly lower performance, likely due to its sensitivity to feature overlap and local data density in high-dimensional soil datasets. Overall, these results support the decision to frame crop recommendation as a **Top-K decision support task** rather than a conventional classification problem. The emphasis on ranked recommendations allows farmers to consider multiple viable crop options based on additional contextual factors such as market demand and water availability.

Based on its **higher Top-3 accuracy, robustness, and interpretability**, **Random Forest is selected as the final model** for soil-based crop recommendation in the AgriConnect+ framework.

Conclusion

This work presented **AgriConnect+**, an integrated artificial intelligence–based decision support system designed to address key challenges in modern agriculture, including plant disease detection, crop price prediction, and soil-based crop recommendation. By combining deep learning and machine learning techniques within a unified framework, the system demonstrates how data-driven approaches can support informed agricultural decision-making.

The plant disease detection module employed a MobileNetV2 + U-Net architecture to accurately identify and localize diseased regions in plant leaves, achieving strong segmentation performance while maintaining computational efficiency. For crop price prediction, a Random Forest regression model was used to analyze historical market data and provide reliable price forecasts with low error rates and high tolerance-based accuracy. Soil-based crop recommendation was formulated as a Top-K decision support problem, where multiple machine learning models were evaluated and Random Forest was selected as the final model due to its robustness and consistent performance.

The experimental results validate the effectiveness of the proposed approach across all three modules. While the system shows promising performance, certain limitations such as dataset dependency and environmental variability remain. Future enhancements may include integration of real-time sensor data, expansion to additional crop varieties, and deployment through mobile platforms. Overall, AgriConnect+ demonstrates the practical applicability of AI-driven systems in promoting sustainable, efficient, and informed agricultural practices.

References

1. M. Dutta, M. R. Islam Sujan, M. U. Mojumdar, N. R. Chakraborty, A. A. Marouf, J. G. Rokne, and R. Alhajj,
2. J. Lu, X. Liu, X. Ma, J. Tong, and J. Peng, "Improved MobileNetV2 crop disease identification model for intelligent agriculture," PeerJ Computer Science, 2023. Available: (publisher page). :contentReference[oaicite:12]{index=12}
3. J. Sharma et al., "Deep learning based ensemble model for accurate tomato leaf disease classification combining MobileNetV2 and ResNet50," Scientific Reports, 2025. Available: https://www.nature.com/articles/s41598-025-98015-x. :contentReference[oaicite:13]{index=13}
4. G. S. Hukkeri, "Classification of Various Plant Leaf Disease Using Pretrained CNNs," Open Agriculture Journal (or similar), 2024. Available: https://openagriculturejournal.com/VOLUME/18/ELOCATOR/e18743315305194/FULLTEXT/. :contentReference[oaicite:14]{index=14}
5. D. Sutaji, "LEMOXINET: Lite ensemble MobileNetV2 and Xception for plant disease detection," 2022. Available: ScienceDirect record. :contentReference[oaicite:15]{index=15}
6. A. Theofilou et al., "Predicting Prices of Staple Crops Using Machine Learning: A Systematic Review," Sustainability, vol. 17, no. 12, 2025. Available: https://www.mdpi.com/2071-1050/17/12/5456
7. K. Srikanth et al., "Random Forest Based Crop Price Prediction," International Journal of Research and Applied Sciences, Engineering and Technology (IJRASET), 2024. Available:

https://doi.org/10.22214/ijraset.2024.65994
8. P. M. Paithane et al., "Random Forest Algorithm Use for Crop Recommendation," ResearchGate, 2023.
9. F. S. Prity et al., "Soil-Specific Crop Recommendation with Explainable Machine Learning,"