Capstone Project Concept Note and Implementation Plan

Project Title: AI-Powered Crop Disease Detection & Advisory System For Smallholder Farmers In Liberia

Team Members

1. Lydia L Labala
2. Michael Senkao

Concept Note

1. Project Overview

Agriculture remains the primary source of livelihood for the majority of Liberians, with cassava, rice, and pepper serving as key staple crops. However, productivity in this sector continues to be undermined by the prevalence of crop diseases such as cassava mosaic disease, rice blast, and fungal infections in pepper. These diseases significantly affect food security, household income, and the national economy.

Conventional disease detection methods, such as manual scouting and extension services, are often inaccessible to farmers in remote rural communities. As a result, disease outbreaks are frequently detected too late, leading to severe yield losses and increased vulnerability among smallholder farmers.

This project proposes **AgriGuard: An AI-Powered Crop Disease Detection and Advisory System for Smallholder Farmers in Liberia**, aimed at addressing the challenge of late and inaccurate detection of crop diseases such as cassava mosaic disease, rice blast, and fungal infections in pepper. By leveraging artificial intelligence and computer vision, the system will enable farmers to identify diseases from simple leaf images using smartphones, even in offline conditions, and provide multilingual, practical advisory support.

Alignment with the Sustainable Development Goals (SDGs):

- **SDG 2 (Zero Hunger):** Enhance food security through improved crop health and productivity.
- **SDG 1 (No Poverty):** Reduce farmer losses and increase household income.
- **SDG 13 (Climate Action):** Promote sustainable agricultural practices by minimizing excessive pesticide use through targeted interventions.
- **SDG 9 (Industry, Innovation, and Infrastructure):** Strengthen innovation and digital transformation in agriculture through the application of AI and mobile systems.

2. Objectives

The overarching aim of this project is to improve food security and resilience among smallholder farmers in Liberia by designing and implementing an accessible AI-powered crop disease detection and advisory system. The specific objectives are:

1. To develop and train lightweight machine learning models capable of accurately detecting diseases in cassava, rice, and pepper under diverse real-world field conditions.
2. To deploy the trained models on mobile devices using TensorFlow Lite in order to ensure accessibility in offline and low-connectivity environments.
3. To construct a localized dataset of cassava, rice, and pepper diseases by collecting and annotating images from Liberian farms.
4. To integrate multilingual and user-friendly advisory support to ensure inclusivity for farmers with varying levels of literacy.
5. To conduct comprehensive field testing and evaluation of the system to assess its accuracy, usability, and potential for large-scale adoption.

3. Background

Agriculture is the backbone of Liberia's economy, employing nearly 70% of the population and contributing significantly to household livelihoods and national food supply. However, the sector faces persistent challenges, particularly from crop diseases that threaten yields of staple crops such as cassava, rice, and pepper. Outbreaks of cassava mosaic disease, rice blast, and fungal infections continue to reduce productivity and undermine food security, leaving smallholder farmers vulnerable to income losses and market instability.

Traditional methods of disease detection, which rely heavily on visual inspection by farmers or the limited agricultural extension services available, are often inadequate due to delays, high costs, and the inaccessibility of expert support in rural areas. As a result, many farmers either misdiagnose crop health issues or identify them too late for effective intervention.

Existing Approaches:

- The **PlantVillage application** has demonstrated the feasibility of AI-based crop disease detection via mobile devices. However, it relies on curated datasets that do not reflect the real-world conditions of Liberian farms.
- **Convolutional Neural Network (CNN) models** such as ResNet, MobileNet, and EfficientNet have achieved accuracies above 90% in controlled environments but require adaptation to local conditions.
- **IoT-integrated agricultural systems** have achieved remarkable success in large-scale contexts but remain unaffordable and inaccessible to smallholder farmers in Liberia.

Identified Gaps:

- Over-reliance on global datasets such as PlantVillage and Kaggle, which fail to capture local environmental variability, lighting, and mixed infections.
- Limited evidence of successful field validation and adoption of AI-powered agricultural tools in Sub-Saharan Africa.
- Insufficient consideration of multilingual support and offline deployment to address inclusivity in rural contexts.

Rationale for Machine Learning:
Machine learning, particularly deep learning models such as CNNs, provides a scalable, cost-effective, and real-time solution to crop disease detection. Transfer learning techniques further enable the adaptation of pre-trained models to localized datasets, ensuring accuracy and applicability in Liberia. This approach offers significant advantages over traditional methods, including speed, scalability, and accessibility.

4. Methodology

The project will be implemented through a structured methodological framework consisting of the following phases:

1. Data Collection and Preprocessing

- **Sources:** PlantVillage dataset, cassava-specific datasets from Kaggle, and newly collected images from Liberian farms.
- **Dataset Size:** Approximately 500–1000 annotated images across cassava, rice, and pepper.
- **Preprocessing:** Image resizing, normalization, and augmentation (rotation, flipping, brightness/contrast adjustments). Data will be partitioned into training, validation, and test sets.

2. Model Development

- **Algorithms and Architectures:**
    - Convolutional Neural Networks (CNNs) and lightweight object detection models such as YOLOv5 and YOLOv8 will be employed to detect and classify crop diseases.
    - Transfer learning techniques will be applied to leverage pretrained models, reducing training time and improving accuracy on relatively small datasets.
    - Hybrid CNN-transformer architectures to ensure robustness in variable field conditions.

- **Techniques:**
    - Transfer learning techniques will be applied to leverage pretrained models, reducing training time and improving accuracy on relatively small datasets.
    - Confidence thresholding with an “unsure” class to mitigate misclassification risks.

3. Advisory Module

- Model predictions will be mapped to actionable recommendations, including cultural practices, treatment methods, and preventive measures.
- Confidence-based outputs will be incorporated; low-certainty cases will prompt referrals to agricultural extension services.

4. System Deployment

- Models will be deployed on smartphones using TensorFlow Lite for offline inference.
- A cross-platform mobile application will be developed using React Native.

5. Pilot Testing and Evaluation

- Pilot studies will be conducted with smallholder farmers in selected counties.
- System evaluation will be based on accuracy, precision/recall, inference time, usability, and adoption rate.
- Feedback from farmers and agricultural stakeholders will be used to iteratively refine the system, ensuring it is practical, inclusive, and user-centered.

5. Architecture Design Diagram

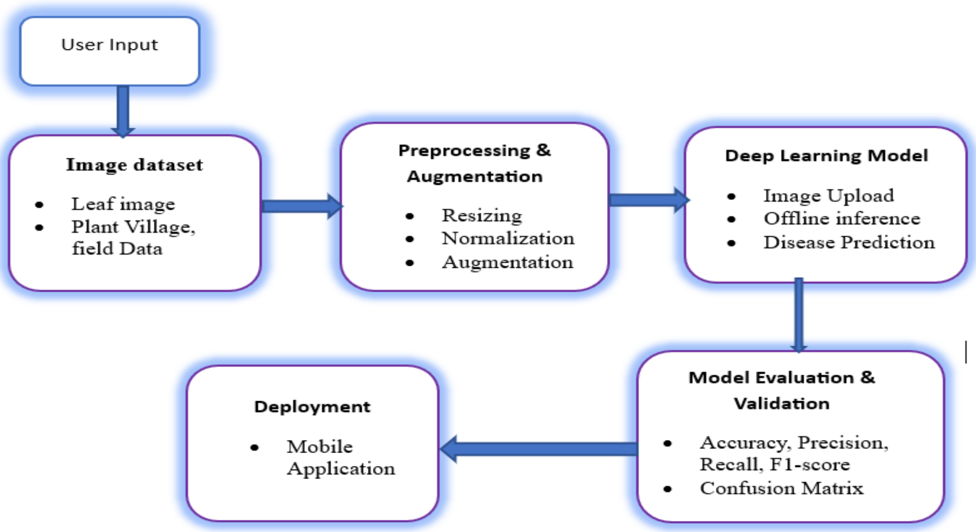

Figure 1: Architecture Design Diagram

6. Data Sources

The project will utilize labeled crop-image datasets from publicly available repositories such as PlantVillage (via Kaggle) and agricultural research archives, supplemented with locally gathered field photos from smallholder farms in Liberia to capture real-world variation in lighting, occlusion, and mixed disease symptoms. These datasets are directly relevant to the task of automated plant disease detection and advisory systems, as they include both healthy and diseased plant images across multiple crop types. Preprocessing will involve de-duplication, class balancing, format standardization, resizing, and augmentation techniques such as flips, rotations, and color adjustments to improve model robustness and generalization. In addition, metadata such as crop type, location, and growth stage will be preserved where possible to support deeper error analysis and facilitate future domain adaptation. This rigorous preparation ensures that the YOLO-based deep learning model is trained on diverse and representative data, enabling accurate detection and classification of crop diseases in real-world farming conditions.

7. Literature Review

Recent advances in artificial intelligence (AI) and computer vision have significantly enhanced crop disease detection, with a growing body of literature demonstrating both technical feasibility and practical applicability in agricultural contexts. Lightweight convolutional neural network (CNN) models optimized for mobile deployment have shown strong accuracy in real-time plant disease detection, validating the viability of smartphone-based diagnostic tools for smallholder farmers. This demonstrates that artificial intelligence (AI), particularly deep learning and computer vision, has matured as a powerful tool for crop disease detection, with convolutional neural networks (CNNs), YOLO-based detectors, and hybrid transformer architectures achieving accuracies exceeding 90% across diverse datasets (Khan et al., 2024; Lee & Zhang, 2025; Wang & Li, 2025).

However, much of this progress is based on curated datasets such as PlantVillage, which fail to capture the variability of real-world farming conditions where poor lighting, overlapping leaves, and mixed infections are common (Patel et al., 2024). To address these limitations, recent studies have focused on lightweight, mobile-friendly models optimized for offline use, enabling real-time disease detection via smartphones in low-connectivity contexts (Rahman & Chen, 2024). Others have demonstrated the integration of AI with IoT-based soil and climate sensors, providing a pathway toward combined detection and advisory platforms that extend beyond classification to actionable guidance (Nguyen et al., 2025; Osei et al., 2025). Importantly, emerging systems emphasize embedding treatment recommendations, multilingual support, and user-friendly dashboards, recognizing that diagnosis without advisory features limits utility for smallholder farmers (Doe & Kumar, 2024; Khan et al., 2024).

Despite these advances, systematic reviews highlight that most solutions remain prototypes, with limited validation in real-world field conditions or adoption at scale (Patel et al., 2024). For Liberia, where cassava mosaic disease, rice blast, and fungal infections in pepper significantly threaten productivity, these insights underscore the need for solutions that combine robust YOLO-based real-time detection with offline deployment and localized advisory features. The proposed project directly addresses these gaps by building localized datasets, optimizing for smartphone compatibility, and integrating decision-support tools tailored to the needs of Liberian smallholder farmers.

Implementation Plan

1. Technology Stack

The implementation of the AI-powered crop disease detection and advisory system will rely on a carefully selected set of technologies that ensure accuracy, efficiency, and accessibility in low-resource environments. The technology stack is organized into four main layers:

I. **Data Management and Preprocessing**
- **Python (NumPy, Pandas, OpenCV):** For data cleaning, preprocessing, and augmentation of crop leaf images.
- **Roboflow / LabelImg:** For image annotation and dataset management.
- **Google Colab / Jupyter Notebooks:** For collaborative development and cloud-based training.
- **Justification:** These tools provide open-source, cost-effective solutions suitable for academic and low-budget contexts.

II. **Model Development and Training**
- **Deep Learning Frameworks:** TensorFlow, PyTorch: For building and training CNN and YOLO-based models.
- **Keras:** For high-level experimentation with CNN architectures.
- **Transfer Learning (ResNet, DenseNet, MobileNet):** To improve accuracy while reducing training time.
- **Justification:** TensorFlow and PyTorch are widely used, well-documented, and support both research and production workflows.

III. **Deployment and Application Development**
- **TensorFlow Lite / ONNX:** For optimizing AI models to run efficiently on mobile devices.
- **Android Studio (Java/Kotlin):** For developing the mobile application.
- **Flutter (Dart):** As an alternative for cross-platform (Android/iOS) support.
- **SQLite / Firebase (optional):** For storing user interaction data and updating advisory content.
- **Justification:** TensorFlow Lite ensures offline inference, while Flutter allows broader accessibility. Offline-first design is crucial for Liberia’s rural areas.

IV. **Advisory and User Interaction**
- **Rule-Based Expert System (Python backend):** For providing disease management recommendations.
- **Justification:** Ensures that farmers not only receive diagnoses but also practical, easy-to-understand advice.

2. Timeline

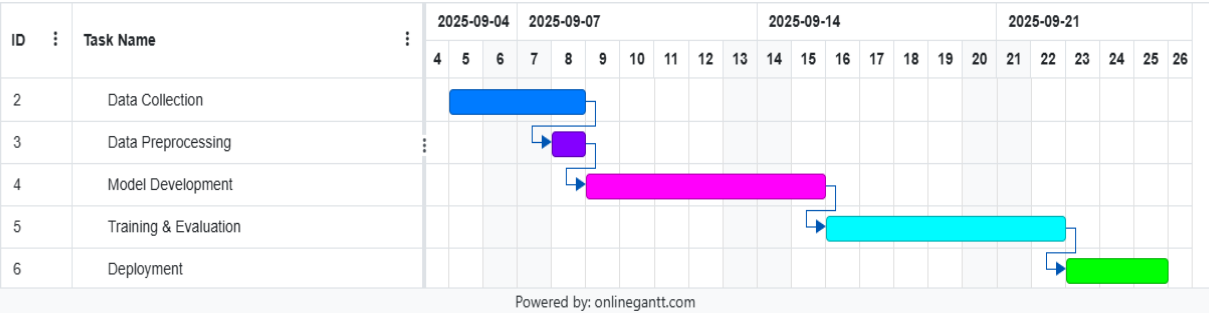

Figure 2: Project Timeline

Task Distribution Matrix:

Task	Lydia	Michael	McCess	Cyrus	Melvin	Nelson
Data Collection & Cleaning	✓	✓	✗	✗	✗	✗
Data Preprocessing & Augmentation	✓	✓	✗	✗	✗	✗
Model Development	✓	✓	✗	✗	✗	✗
Model Training & Evaluation	✓	✓	✗	✗	✗	✗
Deployment	✓	✓	✗	✗	✗	✗

Phase 1: Data Collection and Preprocessing (Day 1 – Day 4)

- Collect datasets (PlantVillage, cassava, pepper, rice datasets).
- Annotate and preprocess images (resizing, cleaning, augmentation).
- Organize datasets for training and testing.

Phase 2: Model Development and Training (Day 5 – Day 15)

- Implement baseline CNN and YOLO architectures.
- Apply transfer learning using pretrained models.
- Train models and evaluate using accuracy, precision, recall, and F1-score.
- Select the best-performing model for deployment.

Phase 3: Evaluation and Deployment (Day 15 – Day 25)

- Validate system accuracy on benchmark datasets.
- Conduct small-scale field testing with Liberian farmers (pilot trial).
- Collect feedback for improvements.

3. Milestones

I. **Dataset Acquisition and Preparation (End of Day 4)**
- Completion of data collection from PlantVillage and other online sources.
- Dataset cleaned, annotated, and augmented for training.

II. **Baseline Model Development (End of Day 9)**
- Initial CNN and YOLO models implemented and tested on prepared dataset.
- Performance metrics (accuracy, precision, recall, F1-score) established.

III. **Optimized Model Selection (End of Day 14)**
- Best-performing model identified and optimized with TensorFlow Lite for mobile deployment.
- Model ready for integration into mobile app.

IV. **Advisory Module Integration (End of Day 15)**
- Rule-based multilingual advisory module added to the mobile application.
- Advisory content validated against agricultural extension resources.

V. **Pilot Testing and Evaluation (End of Day 25)**
- Field testing conducted with smallholder farmers in Liberia.
- User feedback and performance results documented.

VI. **Future Expansion Roadmap (Day 25 onwards)**
- Plans outlined for incorporating locally collected datasets.
- Roadmap developed for scaling the system to additional crops and regions.

4. Challenges and Mitigations

Challenge	Description	Mitigation Strategy
Limited Local Datasets	The current system relies heavily on global datasets (PlantVillage, online sources), which may not fully capture the diversity of local crop diseases in Liberia.	Begin with transfer learning on global datasets, while planning for future fine-tuning with locally collected data during pilot phases. Collaborate with agricultural institutions for data collection.
Low Internet Connectivity	Many rural farmers lack stable internet access, which can hinder cloud-based services.	Deploy models on mobile devices using TensorFlow Lite to ensure offline functionality. Design the app to sync updates when connectivity is available.
Farmer Digital Literacy	Limited exposure to smartphones and digital tools among rural farmers may reduce adoption.	Develop a simple, intuitive mobile interface with icons and audio support in local languages. Provide training sessions through extension officers and community groups.
Model Generalization Issues	Models trained on curated datasets may underperform in real-world farm conditions due to noise, lighting, or background clutter.	Use data augmentation to simulate real-world conditions. Continuously retrain with field-collected images from pilot studies.

Challenge	Description	Mitigation Strategy
Advisory Content Relevance	Rule-based recommendations may not reflect local farming practices or input availability.	Collaborate with local agricultural extension services to adapt advisory content. Update recommendations based on farmer feedback.
Sustainability of the System	Long-term sustainability may be challenged without local ownership and funding.	Partner with universities, NGOs, and government agencies to integrate the system into existing agricultural programs. Explore public–private partnerships for scaling.

5. Ethical Considerations

The implementation of an AI-powered crop disease detection and advisory system raises several ethical considerations that must be addressed to ensure fairness, transparency, and sustainability.

1. **Data Privacy and Security:**
   Although the initial datasets will be sourced from public repositories, future phases will involve collecting images directly from farmers. Ensuring informed consent, secure data storage, and anonymization of sensitive information will be prioritized to protect farmer identities and data rights.
2. **Equity and Accessibility:**
   Many smallholder farmers in Liberia have limited access to smartphones and digital tools. To avoid deepening digital divides, the system will be designed for low-cost devices, with offline functionality and multilingual support to ensure inclusivity.
3. **Bias and Fairness:**
   Models trained primarily on global datasets may inadvertently introduce bias, reducing performance on local crops and conditions. This limitation will be mitigated through fine-tuning with locally collected data and continuous evaluation in real-world contexts to ensure fairness across different farmer groups.
4. **Transparency and Explainability:**
   AI models can often function as “black boxes,” making it difficult for users to trust their outputs. To address this, the mobile app will display not only the disease diagnosis but also confidence levels and simple visual cues, enabling farmers to understand and verify results.
5. **Responsible Advisory Content:**
   Providing accurate and context-appropriate advice is critical. Recommendations will be validated through collaboration with agricultural extension officers and relevant institutions to avoid misleading farmers or suggesting interventions that are not feasible locally.
6. **Sustainability and Ownership:**
   Ethical responsibility also extends to sustainability. To ensure long-term impact, the system will be developed in collaboration with local stakeholders, including

universities, NGOs, and government agencies, fostering local ownership and avoiding overdependence on external actors.

6. References

Doe, J., & Kumar, R. (2024). AI-powered crop care: Transforming farming with disease detection and sustainable practices. *Journal of Agricultural AI, 12*(3), 45–58.

Rahman, A., & Chen, L. (2024). Crop disease detection using lightweight deep learning models for smartphones. *Computers and Electronics in Agriculture, 210,* 107–115.

Osei, K., Mensah, F., & Park, J. (2025). AI-driven crop disease prediction and management system integrating IoT data. *Smart Agriculture, 8*(1), 22–39.

Nguyen, H., Patel, S., & Yadav, R. (2025). AI-powered crop health monitoring using IoT and machine learning. *Sustainable Computing, 14*(2), 89–103.

Lee, M., & Zhang, Y. (2025). Real-time tomato leaf disease detection using YOLO models. *Computers in Agriculture, 34*(4), 56–70.

Khan, S., Ali, M., & Roberts, P. (2024). A deep learning-based model for plant disease detection. *Applied AI in Agriculture, 6*(1), 112–125.

Wang, T., & Li, X. (2025). Hybrid transformer-CNN models for robust plant disease detection. *Nature Scientific Reports, 15*(3), 2234–2249.

Patel, R., Singh, A., & Brown, C. (2024). Deep learning and computer vision in plant disease detection: A comprehensive review. *Artificial Intelligence Review, 47*(2), 301–325.

PlantVillage. (n.d.). *Open access plant disease dataset.* Retrieved from https://plantvillage.psu.edu

TensorFlow. (n.d.). *TensorFlow Lite.* Retrieved from https://www.tensorflow.org/lite

PyTorch. (n.d.). *PyTorch open-source deep learning framework.* Retrieved from https://pytorch.org

Roboflow. (n.d.). *Roboflow: Annotate, preprocess, and manage datasets.* Retrieved from https://roboflow.com