

**MAKERERE**



**UNIVERSITY**

**COLLEGE OF COMPUTING AND INFORMATION  
SCIENCES (CoCIS).**

**DEPARTMENT OF COMPUTER SCIENCE**

**Course Unit: MCS 7103 – Machine Learning.**

## **CONCEPT PAPER**

**Project Title: Coffee Disease Detection using image  
classification.**

**Presented by:**

STUDENT NAME	STUDENT NUMBER	REGISTRATION NUMBER
GRACE. RONALD. SEMWEZI	2500726370	2025/HD05/26370U

## 1. Introduction

Uganda sits at the heart of the global coffee landscape, ranking as Africa's second-largest producer and one of the world's top exporters of coffee. As of 2025, Uganda's coffee industry continues to flourish, recording an annual export volume exceeding 7.9 million bags and generating about US\$2.2 billion in revenue, a testament to its exponential growth and centrality to the country's socio-economic health. However, this success is persistently threatened by plant diseases, with pathogens causing substantial yield losses and compromising quality, market access, and farmer livelihoods. Early detection and precise management of coffee diseases have thus become crucial imperatives for ensuring industry sustainability and supporting millions of rural Ugandans.

Amidst these challenges, Digital Transformation through Machine Learning (ML) is rapidly advancing agriculture worldwide. The deployment of ML—particularly image-based classification powered by deep Convolutional Neural Networks (CNNs)—offers fast, cost-effective, and scalable solutions for the early detection and diagnosis of plant diseases. In Uganda, such innovations present a timely, transformative opportunity: supporting smallholder farmers with affordable disease surveillance tools, minimizing crop losses, optimizing input use, and empowering extension services. This concept paper proposes a research project to Design, Implement, and Evaluate a coffee disease detection system using state-of-the-art image classification methods.

The following sections expound on the background of Uganda's coffee industry, detail disease threats and their impacts, explore the role of ML in agriculture, describe the methodology for developing a coffee disease image classifier, outline expected outcomes, and consider the anticipated challenges, all supported by current academic and industry references.

## 2. Background

### 2.1. Overview of Uganda's Coffee Industry

Coffee is not merely a crop in Uganda—it is a foundation of the national economy and the backbone of rural livelihood. As of the 2024/2025 season, industry data shows that Uganda achieved its second-highest export level in history, reaching 6.3–7.9 million 60-kg bags and generating over USD \$2 billion in revenue. The sector comprises over 1.7 million smallholder farmers, with coffee-related activities providing income for more than 9 million Ugandan households.

There are two dominant coffee species in Uganda:

- **Robusta coffee (*Coffea canephora* Pierre ex A.)** accounts for about 80% of the nation's production, thriving in lower-altitude zones such as Central, Western, and Eastern Uganda.
- **Arabica coffee (*Coffea arabica*)** takes the remaining 20%, predominantly grown in cooler, high-altitude areas like Mt. Elgon and Rwenzori regions.

The export markets for Ugandan coffee are broad and robust. Europe claims the largest share (over 60%), followed by Africa, Asia, and North America. Top destinations include Italy, Germany, Sudan, and India,

with leading exporters such as Olam Uganda Ltd, IBERO Uganda Ltd and UGACOF Ltd ensuring global reach.

Despite industry growth and modernization efforts—like farmer registration and digital traceability to meet strict EU deforestation regulations—disease outbreaks and climate variability remain formidable obstacles. Persistent yield gaps, which are far below genetic potential, are attributed largely to pest and disease attacks; Sub-optimal agronomic practices, and climate-induced stress.

2.2. Key Coffee Varieties in Uganda

Ugandan coffee is prized for its diversity, with multiple grades and preparation methods:

Robusta Coffee Varieties

Robusta dominates the Central, Western, and Eastern regions, grown mostly on small mixed-crop farms. The crop is sold in several grades:

- Washed Robusta
- Screen grades (18, 17, 15, 12, etc.)
- BHP 1199 (Bulk Hand-Picked)

Arabica Coffee Varieties

Arabica thrives in high-altitude, cooler areas, mainly Mt. Elgon and Rwenzori. Key types include:

- Bugisu (A+, AA, A, AB, CPB)
- Mt. Elgon
- DRUGAR (dried Ugandan Arabica)
- WUGAR (washed Ugandan Arabica)
- Organic and Sustainable grades

Below summarized table of exports reflects this diversity for the month of August 2025 alone:

Coffee Type	Quantity (60-Kg Bags	%-age Quantity	Value in US\$	Unit Price US\$/Kg
Robusta	797,363	93.2	180,790,699	3.78
Arabica	58,078	6.8	21,957,846	6.30
Total All	855,441	100	202,748,545	3.95

Arabica generally fetches a higher price per kg, reinforcing the need for both quality and disease resilience in the existing production systems.

2.3. Major Diseases Affecting Coffee Plants in Uganda

Coffee plant diseases pose some of the gravest threats to production and export revenue in Uganda. The most significant diseases and pests include:

- **Coffee Wilt Disease (CWD):** A fungal disease with the potential for catastrophic losses—historically, CWD destroyed up to 45% of Robusta trees and afflicted 90% of farms in some regions.
- **Coffee Leaf Rust (CLR):** Affects both Arabica and Robusta, causing premature leaf fall, reduced photosynthesis, and substantial yield reduction.
- **Red Blister Disease:** Known as “berry blotch” or 'ekikaata', now recognized as the most prevalent disease in some regions, causing lesions, berry drop, and pronounced yield loss.
- **Black Coffee Twig Borer:** The most damaging pest, with incidences of 15–25%, causing dieback and yield loss.
- **Leaf miners, skeletonisers, tailed caterpillars, scales, berry moths, root mealybugs, stem borers, and other beetles:** These cause varying degrees of defoliation, berry damage, or tree death.

Findings from recent field surveys show considerable variation in disease incidence and severity across districts:

Disease	Regional Mean Incidence (%)	Severity Score (1–5)
Red Blister	51.6	1.92
Brown Eye Spot	47	1.5
Leaf Rust	32	1.5
CWD	2.5	–

Certain districts like Mpigi and Kayunga face especially high burdens, amplified by the presence of old, unpruned trees and poor agronomic practices.

## 2.4. Economic Impact of Coffee Diseases

The direct and indirect impacts of coffee diseases on Uganda are extensive:

- **Yield Losses:** Persistent pest and disease pressure leads to average yields significantly below the potential. Actual clean Robusta yields average 0.6 t/ha, far below the attainable 2.2 t/ha or the 4.8 t/ha of CWD-resistant varieties.
- **Export Declines:** Historical CWD outbreaks caused an estimated 7.4% per annum loss in yield, with total income, production, and land area under coffee all declining dramatically in high-disease zones.
- **Market Disruptions:** Disease-induced harvest shortfalls result in fluctuating farm-gate prices, export inconsistencies, and market instability, especially for high-value Arabica producers.
- **Livelihood Threats:** For Uganda’s 8–9 million coffee-dependent people, disease outbreaks equate to reduced income security, higher poverty rates, and even secondary effects—such as reduced school attendance also seen elsewhere in Africa during coffee sector declines.

The UCDA and Ministry of Agriculture estimate annual losses due to coffee pests and diseases in excess of US\$8 million, underscoring the immense national stakes involved.

## 2.5. Significance of Early Disease Detection

Timely diagnosis and mapping of diseases are critical for several reasons:

- i. **Prevention of Epidemic Spread:** Early intervention enables localized response—pruning, spot chemical treatment, or rouging—minimizing area-wide loss and input waste.
- ii. **Resource Optimization:** Accurate, early detection allows for targeted treatments rather than blanket pesticide/fungicide application, reducing costs and environmental harm.
- iii. **Sustainable Management:** Early alerts facilitate integrated pest management strategies and reduce the overuse of chemicals, supporting sustainable intensification and environmental conservation.
- iv. **Empowering Farmers:** Digital tools facilitate real-time monitoring, allowing even smallholders in remote areas rapid access to actionable information.

In practice, national programs now deploy farmer trainings and participatory surveillance; over 4,000 Ugandan farmers across multiple regions participated in pest and disease management sessions in 2025 alone. Yet, with the manual detection method remains labor-intensive and subjective; Scalable and automated solutions such as image-based classification using Machine Learning are urgently needed to bridge the detection gap.

## 3. Problem Statement

Despite Uganda’s robust coffee research tradition, significant gaps persist in universal, timely, and scalable disease surveillance and diagnosis. Existing approaches to plant disease detection are hampered by a number of persistent problems that include but not limited to:

1. **Manual Inspection Limitations:** Reliance on the naked-eye identification of symptoms by experts or trained farmers is slow, prone to human error, and not feasible for monitoring large plantations or providing real-time alerts.
2. **Diagnosis Complexity:** Many coffee diseases (such as Red Blister and Brown Eye Spot) present overlapping or ambiguous symptoms, especially in early stages, increasing misdiagnosis risk.
3. **Scarcity of Affordable Diagnostic Tools:** While laboratory tests are definitive, they are inaccessible and unaffordable for most smallholders.
4. **Data Gaps:** There is an absence of exhaustive, up-to-date disease maps and datasets for Uganda’s coffee regions, limiting extension service response and the systematic roll-out of management interventions.
5. **Lack of Scalable, Real-time Digital Solutions:** The current technological penetration is insufficient to provide real-time, robust disease identification at farm level.

Consequently, there is an urgent need for **technology-driven, scalable, affordable, and accurate disease detection platforms** that can empower smallholder farmers, extension agents, and stakeholders across the coffee value chain in Uganda.

## 4. Objectives

The overarching goal of this project is to harness Machine Learning—specifically, state-of-the-art image classification models—for the early, accurate, and scalable detection of coffee diseases in Uganda. The specific research objectives are:

1. To review and document the principal diseases and pests affecting coffee (Arabica and Robusta) in Uganda, their geographic incidence, and their economic impact.
2. To develop a robust, annotated image dataset of Ugandan coffee leaves with representative samples for all major disease classes, using field data and available public datasets.
3. To design, train, and validate an image classification model using Deep Convolutional Neural Network architectures capable of identifying the major coffee diseases with high accuracy.
4. To assess, compare, and optimize various image preprocessing, data augmentation, and annotation techniques for effective model training and generalization.
5. To evaluate model performance using relevant validation metrics (accuracy, precision, recall, F1-score, confusion matrix, etc.) and benchmark against recent literature.
6. To investigate the potential and limitations of deploying the trained model on mobile and edge devices, enabling practical, on-farm disease diagnostics.
7. To analyze the core implementation and deployment challenges, addressing technical, operational, and socio-economic barriers in real-world Ugandan contexts.
8. To demonstrate, through pilot testing, the effectiveness and readiness of the developed system for wider roll-out and integration with national extension services.

## 5. Methodology

### 5.1 Data Acquisition: *Datasets and Annotation*

#### Field Data Collection

The foundational requirement is a high-quality, representative, and well-annotated image dataset:

- **Field Data:** Digital images of coffee leaves will be collected from representative farms across Uganda’s main coffee-growing regions (Central, Western, Elgon, Rwenzori, etc.), ensuring:
  - All major disease phenotypes (CLR, CWD, Red Blister, etc.)
  - Both healthy and diseased samples
  - Natural lighting and background variability

- **Supplementary Data:** In a bid to enhance volume and diversity, publicly available datasets will be used where appropriate. Noteworthy datasets include:
  - **PlantVillage:** Over 54,000 images for 38 classes of crops/diseases, including some coffee leaf images.
  - **RoCoLe - Robusta Coffee Leaf Dataset:** Sourced from GTS AI portal (<https://gts.ai/dataset-download/rocole-a-robusta-coffee-leaf-images-dataset/>), this dataset contains 1,560 high-resolution images, categorized into healthy leaves, Coffee Leaf Rust (CLR) disease, Spider Mite infestation. These are captured under real-world Ugandan/Brazilian conditions with natural lighting and varied backgrounds. Ideal for training deep learning models.
  - **Kaggle Coffee Datasets:** Datasets with coffee leaf images classified for Coffee Leaf Rust (CLR), Leaf Miner, and other common pathologies.
  - **Uganda Coffee Leaf Disease Dataset:** Sourced from Mendeley Data portal (<https://data.mendeley.com/datasets/k36wnd6knb/1>); containing about 3,312 labelled images collected from farms in Uganda; categorized into Health, Coffee Leaf Rust (CLR) and Phoma disease. This dataset also includes images (rotation, flipping, brightness adjustment) that are excellent for localized research and model training.
  - **Coffee Leaf Disease Classification GitHub Repository:** Sourced from GitHub.com (<https://github.com/lenaAlenazi/Coffee-Leaf-Diseases-Classification-using-Deep-Learning-Models>), this dataset includes multiple disease types like Coffee Leaf Rust, Ceercospora Leaf Spot, Phoma Leaf Spot and Coffee Leaf Miner.

## 5.2 Data Labeling and Quality Control

- **Annotation Tools:** Software like Labelling, VGG Image Annotator (VIA), or custom annotation tools will be used to precisely label each image with disease class, severity, and region of interest (ROI) segmentation if required. Weakly supervised and semi-supervised annotation strategies may also be explored to reduce manual burdens.
- **Expert Curation:** Agronomists and plant pathologists will validate annotated samples, minimizing mislabeling and class crossover.
- **Data Augmentation:** To counter class imbalance and augment scarce disease samples, techniques such as rotation, flipping, color jittering, and the use of GANs/diffusion models for synthetic augmentation may be applied.

## 5.3 Image Preprocessing and Augmentation

Quality data preprocessing is pivotal for model performance:

- **Standardization:** All images will be resized to a uniform resolution (e.g., 224×224 pixels) and normalized.

- **Noise Reduction:** Techniques like Gaussian, median, or bilateral filtering to remove high-frequency noise and enhance leaf features.
- **Color Space Transformation:** Optionally convert RGB to HSV or grayscale for feature enhancement, particularly for texture or lesion identification.
- **Segmentation:** Foreground extraction (i.e., leaf segmentation from background) using U-Net or u^2 Net or conventional thresholding may improve focus on disease features and minimize distractors.

Augmentation methods—both conventional (geometric transforms, color jitter) and advanced (GAN or diffusion-generated synthetic images)—will be crucial to:

- Balance the class distribution and
- Enhance the robustness of the classifier to real-world visual variability and rare early-stage symptoms.

## 5.4 Model Architecture: Image Classification Methods.

### Model Selection

Convolutional Neural Networks (CNNs) have become the dominant architecture for plant disease classification tasks, offering automated feature extraction and high classification accuracy. The project will investigate and benchmark several architectures:

- **Baseline Architectures:** ResNet, VGG, InceptionV3, DenseNet, and MobileNet families—all proven on PlantVillage and in related studies for both accuracy and mobile efficiency.
- **Latest Models:** EfficientNet (balances accuracy and speed), Vision Transformers, Swin Transformers, and attention-enhanced CNNs for complex or subtle multi-class problems.

Transfer learning—fine-tuning pre-trained models on ImageNet or larger plant disease datasets—will increase model efficiency and performance, especially when field data is limited.

## 5.5 Training, Validation, and Hyperparameter Tuning

- **Dataset Splitting:** Data will be split into training (70%), validation (15%), and hold-out testing (15%) partitions balanced by class for unbiased performance estimates.
- **Training:** Optimization will use Adam, SGD, or RMSProp, with early stopping and learning rate schedulers to avoid overfitting.
- **Cross-validation:** K-fold cross-validation may be applied where sample size allows.
- **Hyperparameter Search:** Grid or random search for optimal batch size, learning rate, dropout, augmentation parameters, and model depth.
- **Loss Functions:** For multi-class classification, categorical cross-entropy is standard; class weighting can be implemented to address class imbalance.



## 5.6 Model Evaluation Metrics

Key evaluation criteria will include:

- **Accuracy:** Overall correct predictions out of total samples.
- **Precision, Recall, F1-score:** Per-class evaluation for sensitivity and specificity.
- **Confusion Matrix:** For insight on common misclassifications and confusion between similar diseases.
- **ROC-AUC and PR curves:** Particularly for binary and multi-class problems
- **Robustness:** Performance across lighting, noise, and disease severity levels

The project will compare results with recent benchmarks and document overfitting risks (e.g., models performing well on controlled datasets but dropping significantly in real-world field images).

## 5.7 Model Interpretability and Visualization

To increase trust and transparency:

- **Attention Maps:** Grad-CAM, Score-CAM, and related approaches will be used to visualize which image regions most influenced the model's decision, crucial for scientific and extension acceptance.
- **Ablation Studies:** Evaluating the impact of different pre-processing, augmentation, or model architecture components.

## 5.8 Deployment and Edge Implementation

For practical field use, model deployment strategies will emphasize accessibility and efficiency:

- **Mobile and Edge Deployment:** Models will be quantized and converted (TensorFlow Lite, ONNX, etc.) for efficient on-device inference in Android smartphones or affordable edge devices (e.g., Raspberry Pi).
- **Offline Capability:** Essential for rural Ugandan farmers with inconsistent internet access.
- **App Interface:** Simple user interface for image capture, disease prediction, and feedback. If feasible, the interface will support anonymous data upload to update the central database and enable continuous improvement (federated learning).
- **Feedback and Extension Integration:** Collaborate with the National Coffee Research Institute (NaCORI) and UCDA for extension worker piloting, farmer field school demonstrations, and integration with national traceability platforms.

## 5.9 Expected Outcomes:

1. **Comprehensive, Validated Dataset:** A robust, annotated image database of Ugandan coffee leaves covering most major diseases, severity levels, and region-specific variation, contributing to both academic and practical knowledge bases.
2. **Trained, Validated Disease Classifier:** A deep learning model (e.g., EfficientNet, DenseNet, or hybrid attention-based CNN) with high classification accuracy, capable of distinguishing between healthy leaves and the principal coffee diseases listed.
3. **Operational Deployment Prototype:** A cost-effective, user-friendly smartphone or edge device App that provides instant, on-farm disease identification—bridging the diagnostic gap for smallholders and extensionists alike.
4. **Enhanced Early Warning and Decision Support:** By enabling timely alerts and rapid diagnosis, yield losses from outbreaks can be significantly mitigated, reducing the need for blanket pesticide application and supporting environmental sustainability.
5. **Empowerment of Farmers and Extension Systems:** Democratized access to expert-level diagnostics, especially in resource-constrained, hard-to-reach, or previously underserved regions.
6. **Research and Policy Contribution:** Recommendations for national scale-up, integration with traceability requirements (e.g., EU Deforestation Regulation), and incorporation into Uganda’s Digital Agriculture strategy.
7. **Academic Output:** Peer-reviewed publications, open-source code contributions, datasets for future research, and capacity-building of Makerere University and national research partners.

## 5.10 Challenges.

The real-world implementation of image-based coffee disease detection in Uganda will need to confront several significant challenges:

### Dataset Diversity and Annotation

- **Data Scarcity in Rare Classes:** Some diseases are infrequent, leading to imbalanced datasets. Generating enough labeled early-stage or rare disease images is a persistent hurdle.
- **Annotation Quality:** Consistency of expert annotation is vital, yet inter-annotator variability and time-consuming manual curation may undermine model learning if not properly managed.
- **Environmental and Regional Variability:** Lighting, background, variety, and severity gradient changes across regions can confound image classification unless explicitly modeled during training.

### Model Generalization and Overfitting

- **Controlled vs. Field Images:** Models trained on neat, lab-based images often overfit and fail to generalize to noisy, field-collected data—where non-uniform backgrounds and subtle early-stage lesions are typical.

- **Robustness to Noise:** Variation in mobile phone cameras, image quality, or environmental artifacts (e.g., dust, shading) can reduce prediction accuracy.

#### Computational and Operational Constraints

- **Edge Deployment Trade-offs:** Achieving both speed and accuracy on low-cost devices requires careful model optimization and quantization, and sometimes, compression at the expense of performance.
- **Connectivity:** Many rural farmers have limited or no internet; offline functionality is mandatory.
- **User Training:** Deployment success depends on intuitive interfaces, multi-language support, and community sensitization.

#### Socio-economic and Institutional Barriers

- **Farmer Trust and Adoption:** Tech skepticism, digital literacy gaps, and concerns over privacy/data sharing require participatory design, capacity building, and transparent communication.
- **Integration with Policy/Tracing:** Sustainable impact hinges on integration with government extension systems, certification, and traceability frameworks.

#### Data Privacy and Ethics

- **Farmer Data Ownership:** Clear, transparent policies regarding data collection, usage, and protection are essential to retain user confidence and comply with ethical norms.

#### Future-proofing and Sustainability

- **Evolution of Pests and Diseases:** Ongoing climate change and mutation of pathogens may introduce novel disease presentations, necessitating an updateable and expandable framework.
- **Continuous Model Learning:** Systems must allow ongoing re-training as new images and disease variations emerge, ideally via federated learning or semi-supervised methods.

## 6. Literature Review

### In-text References

All references have been embedded contextually, spanning recent research articles, Ugandan official Coffee Authority reports, international NGO and Government studies, peer-reviewed works, and domain-specific datasets, including but not limited to:

- Uganda Coffee Development Authority (UCDA) reports (2023–2025).
- United States Department of Agriculture, Coffee: World Markets and Trade (2024–2025).
- Recent peer-reviewed articles on ML/DL in plant disease identification.
- Open datasets for plant disease and coffee leaf images.
- Reports on edge deployment, transfer learning, and the challenges specific to African smallholder systems.

- Signo et al. (2024) reviewed various models including SVM, VGG-19, and DenseNet for coffee disease detection, achieving accuracies up to 99.57%.
- Essoh et al. (2022) used texture descriptors and multi-class SVM to classify Arabica coffee leaf diseases, reaching 96.75% *F1-score*.
- Yebasse et al. (2021) emphasized the importance of visualization techniques like Grad-CAM to interpret model decisions and improve trust in AI systems.

## 7. Conclusion

The advent of accessible, robust, image-based machine learning tools is poised to revolutionize the fight against endemic coffee diseases in Uganda. Leveraging locally relevant data, expert knowledge, and state-of-the-art computational models, this project will advance both the science and practice of precision agriculture in East Africa. It will bolster national resilience to pest and disease outbreaks, empower millions of smallholder farmers, and contribute to the global drive for a sustainable, productive, and equitable coffee industry.

The key to meaningful, sustainable impact will lie not only in technological prowess but also in a participatory approach—continuous engagement with farmers, researchers, extensionists, and policy-makers, ensuring that every image, every prediction, and every decision is rooted in Ugandan realities and tailored to local needs.