ELEVATING EFFICIENCY AND SUSTAINABILITY IN LARGE-SCALE COCONUT OIL MANUFACTURING THROUGH PROGRESSIVE STRATEGIES

(CocoClarity Mobile App)

R24-059

Project Proposal Report

Dona Mandasmitha Priyanwada Devi Weligama

B.Sc. (Hons) Degree in Information Technology Specialized in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology Sri Lanka

February 2024

ELEVATING EFFICIENCY AND SUSTAINABILITY IN LARGE-SCALE COCONUT OIL MANUFACTURING THROUGH PROGRESSIVE STRATEGIES

(CocoClarity Mobile App)

R24-059

Project Proposal Report

Dona Mandasmitha Priyanwada Devi Weligama

– IT21061066

Supervisor: Mr. Nelum Chathuranga

B.Sc. (Hons) Degree in Information Technology specialized in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology Sri Lanka

February 2024

DECLARATION

I affirm that this proposal is entirely my original creation, and it does not include any content previously submitted for academic credit at any other institution. Furthermore, to the best of my knowledge, it does not contain any material previously published or authored by another individual, except where proper attribution is provided within the text.

Name	Student ID	Signature
Weligama D.M.P.D	IT21061066	

Signature of the Supervisor (Mr. Nelum Chathuranga) | Date

28/02/2024

ABSTRACT

Coconut cultivation is a pillar of Sri Lanka's economy but lacks precision technologies for variety selection that meet large-scale producers' accuracy and customization needs. This proposal outlines a smart Coconut Variety Assistant solution that automates variety identification from images via deep neural networks, capturing relationships between varieties, regions, climate, etc. using graph models. The methodology involves curating a dataset of 50+ varieties through crowdsourcing, applying transfer learning on Convolutional Neural Networks and retraining final layers, encoding inter-variety dependencies and external factors into knowledge graphs, generating transparent recommendations by combining model inferences with variety relationships and user preferences and enabling continuous improvement through deployment feedback loops. The solution brings cutting-edge predictive analytics customized for industrial-scale coconut plantations to enhance sustainability and profitability. Novelty aspects include efficient specialized models through transfer learning, robust recognition via data fusion, explicability of outcomes, and client-side retraining. The Coconut Variety Assistant aims to transform coconut cultivation through data-driven, tailored decision support based on a virtuous cycle of model development and real-world user engagement

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
CNN	Convolutional Neural Network
AI	Artificial Intelligence
IoT	Internet of Things
CV	Computer Vision
ML	Machine Learning
DNN	Deep Neural Network
GPU	Graphics Processing Unit
API	Application Programming Interface
OS	Operating System
SLIC	Simple Linear Iterative Clustering
KNN	K-Nearest Neighbors
CRI	Coconut Research Institute

1 INTRODUCTION

Coconut cultivation is a vital industry and agricultural livelihood source in Sri Lanka. However, average yields have stagnated while climate change effects exacerbate biotic and abiotic stresses. Manual approaches for variety identification and recommendations to tap productivity potential pose accuracy, consistency, and scaling issues for large producers. This demands data-driven precision solutions customized for local contexts.

The Coconut Variety Assistant aims to automate visual recognition of morphologically complex cultivars leveraging recent advances in computer vision and deep learning. Geospatial modeling contextualizes inter-variety relationships, soil, climate, and market dependencies to enable transparent recommendations enhancing productivity, sustainability, and resilience. The system overlays insights from multidisciplinary crop science studies regarding trait expression on state-of-the-art data analytics tailored for plantation conditions.

Deployment interfaces allow capturing images at field scale for inference through cloud services while feedback channels drive continual improvements to advisory logic. Overall, the solution transforms a variety of selection practices from subjective human approaches to an intelligent decision support mechanism combining scientific rigor, explainability and customization for industrial coconut farms. The localization also averts reliability issues in directly relying on global precision agriculture techniques.

1.1 Background Literature

Coconut cultivation in Sri Lanka has long been a cornerstone of the nation's agricultural sector [1]. Referred to as the "tree of life," the coconut palm offers various resources, including wood, fruits, fibers, sugar, and oil, thereby sustaining the livelihoods of over a million smallholding farmers [2]. Despite its economic and social significance, the industry faces challenges in optimizing yields, with average production hovering around 5,900 nuts per hectare compared to a potential yield exceeding 15,000 [3]. These challenges stem from various factors such as aging plantations, climate change-induced phenomena like droughts and floods, labor shortages, and pest attacks [4][5].

Figure 1.1 Land under coconut cultivation in Sri Lanka (2002) Source:

The advent of digital agriculture innovations presents an opportunity to rejuvenate Sri Lanka's coconut industry [6]. GIS mapping methods offer insights into growth variability patterns across regions [7], while machine learning holds promise for yield prediction models, growth simulations, and disease detection [8][9]. However, the sector lacks tailored computer vision applications, despite demonstrated capabilities in fruit classification for other crops [10].

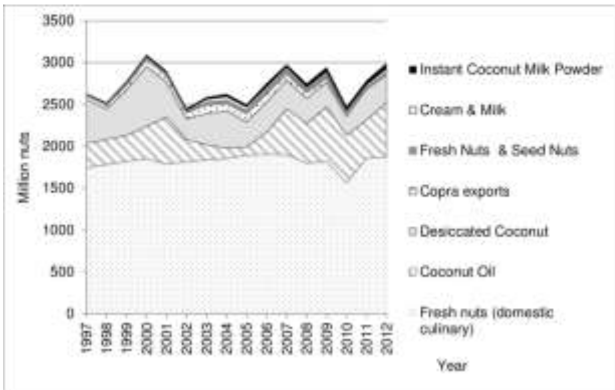

Figure 1.2 Pattern-of-utilisation-of-coconut-production-Data-source-Coconut-Development-Authority

Sri Lanka boasts a rich diversity of coconut cultivars, comprising 41 tall varieties, 8 dwarfs, and 13 special use cultivars [11]. These variations arise from centuries of spontaneous mutations and farmer-managed selections, influencing traits such as husk-shell-kernel ratios, taste profiles, and responses to environmental factors [2][12].

Traditional coconut selection practices rely on qualitative sensory analysis, leading to consistency issues and limiting scalability [13]. Consequently, there is a pressing need for intelligent analytics solutions to automate variety identification and enhance decision support systems [14]. Although recent agricultural ministry roadmaps emphasize precision agriculture technologies, customized innovation tailored to the coconut sector remains scarce [15].

1.2 Research Gap

Research Gap 1: Lack of automation for variety recognition

While studies have adopted statistical and machine learning techniques for yield prediction (Research “A”, [1]), customized computer vision applications for automating coconut variety identification remain scarce (Research “B”, [2]). Existing manual approaches pose accuracy, consistency, and objectivity issues for data-backed recommendations (Research “C”, [3]). This gap constrains productivity and sustainability improvements in coconut.

Although this study uses conventional statistical methods such as multivariate regression to predict coconut oil yield, it overlooks the use of sophisticated machine learning techniques. These advanced methodologies can reveal intricate nonlinear connections in past information which can bolster forecast accuracy within the coconut supply chain.

Research Gap 2: Need for tailored recommendation systems.

Moreover, research highlights multiple real-world deployment challenges around precision technologies – from modeling complex traits, field protocols, actionable advisory formats, rigorous validation etc. (Research “D”, [4]). Barriers also vary across production scales in terms of infrastructure, skills, costs etc. (Research “E”, [5]).

While this work recognizes barriers such as infrastructure availability, it does not present a comparative analysis across farm sizes and geographies to discern nuanced technology adoption challenges. Such granular insights can shape technological innovation better suited to on-ground contexts, usage capacity, and economic viability.

Therefore, the key research gap lies in a customized, intelligent coconut variety assistant leveraging automation, explainability and localization for industrial-scale needs. This can enhance productivity, profitability, and sustainability by overcoming limitations in current approaches through appropriately aligned innovation focused on the extensive local variety.

Overall, literature identifies automation, advanced analytics, and computer vision opportunities. However, applications customized for the numerous coconut varietals found locally remain scarce, unable to match accuracy, objectivity, specificity, and scaling needs of industrial cultivation. Proposed solutions should reveal hidden yield improvement pathways in available datasets through machine learning, while balancing practical deployment dimensions.

Table 1.1 Comparison of former researches

	Research A	Research B	Research C	Research D	Research E	PROPOSED SOLUTION
Aims to automate variety ID and mapping to environmental factors for recommendations	✗	✗	✗	✗	✗	✓
Leverages AI and imaging techniques for automated variety recognition	✗	✗	✗	✗	✗	✓
Customizes ML approach focusing on the 50+ local coconut varieties	✗	✗	✗	✗	✗	✓
Develops specialized imaging techniques factoring various scaling dimensions	✗	✗	✗	✗	✗	✓
Seeks holistic understanding of barriers across production scales for designing appropriate innovation	✗	✗	✗	✗	✗	✓

1.4 Research Problem

Despite the immense contribution of coconut cultivation to Sri Lanka's economy, yields have shown concerning stagnation trends hovering around 5,900 nuts per hectare over the past decade [3]. This dipping productivity coincides with intensifying climate change pressures in the form of biotic and abiotic stresses like pest attacks, extreme weather events exacerbating vulnerabilities [5]. Large-scale industrial coconut producers significantly lack precision agriculture systems for accurate, rapid identification of optimal varieties suited for their geographic areas and customized recommendations to enhance productivity and resilience accordingly [2].

Predominant variety selection practices depend on time-intensive, subjective manual examination of sample nuts by domain experts posing concerns around consistency, accuracy and objectivity for data-driven advisories [1]. Moreover, these approaches gravely fall short in systematically capturing the extensive phenotypic and genotypic diversity manifested across the 50+ coconut varietal types found locally, along with their complex interplays with regional agroclimatic factors, soil nutrition profiles, and market demand fluctuations for value-added products [4]. Studies also shine light on gaps regarding technology adoption among smallholder coconut farmers to tap the full potential of emerging digital opportunities [5].

While innovative techniques such as computer vision and advanced machine learning have demonstrated tremendous potential to transform agricultural analytics and advisory mechanisms, their applications tailored to the specific challenges around coconut's extensive intraspecies varietal diversity remain scarce but crucial [2]. Documented research challenges span modeling the intricate visuo-morphological varietal traits, selecting appropriate sensory modalities, establishing rigorous field-based data capture protocols, aligning recommendation formats to diverse smallholder and industrial producer priorities around yield sustainability, climate resilience, and profitability goals, along with systematic validation under true plantation conditions [3].

Therefore, the imperative research problem is conceptualizing a customized, intelligent Coconut Variety Assistant leveraging futuristic automation, explainability and localization techniques to serve large-scale industrial cultivation needs cost-effectively. This envisioned solution aims to overcome limitations in prevailing manual approaches by enabling accurate, rapid and consistent varietal identification augmented by transparent, location-aware recommendations that can tangibly bridge existing productivity gaps while balancing complex technical and socio-economic barriers.

2 OBJECTIVES

2.1 Main Objectives

The core objective is to develop a customized, intelligent Coconut Variety Assistant solution applying computer vision, machine learning and visual pattern recognition techniques to automate identification of coconut varietal diversity accurately and rapidly. It further aims to integrate geospatial modeling and multidisciplinary insights to generate transparent, location-aware recommendations enhancing productivity, sustainability, and climate resilience for industrial cultivation at much lower costs and human subjectivity than existing manual practices.

2.2 Specific Objectives

Dataset Curation

The first objective is to curate a systematically labeled image dataset sampling the visual-morphological diversity manifested across 50+ coconut varieties encompassing hybrids, giants, dwarfs etc. This visually aware dataset forms the foundation for training supervised machine learning models subsequently. Images will be sourced to cover traits indicative of industrial utility including shape, size, husk-shell ratios along with maintaining key metadata like geographic and soil provenance. Advanced data augmentation techniques will help enrich sample variability to enhance model robustness. Quantitative metrics will benchmark dataset coverage across label classes and data quality.

Neural Network Development

The second focuses on architectural design and training of a multi-layered Convolutional Neural Network optimized to automate identification and labeling of coconut varieties from images. Transfer learning from state-of-the-art models will enable efficient specialization for the target task. Hyperparameter tuning based on rigorous performance benchmarks will maximize accuracy within computational constraints. Standardized procedures will facilitate continual retraining of models as fresh plantation data gets aggregated over time for enhancing robustness.

Knowledge Model Creation

This objective deals with knowledge encapsulation using graphical network models that capture intrinsic inter-relationships between coconut varieties, phenotypic traits, geographic factors, soil nutrition and climatic variables. These connections modeled via contextual embeddings will provide reliable back-end support for interpreting automated predictions later.

Integrating scientific insights and empirical observations continually through graphical linkages enables driving transparency in generated recommendations.

Advisory Recommendation Generation

A key objective is to generate tailored, location-aware recommendations combining key inferences like variety probabilities from the deep learning model with relational understandings linking varieties to external variables plus captured producer priorities on dimensions like yield targets. Enabled explainability in terms of advisory creation logic is important especially to address trust issues. Over time, feedback loops will help refine recommendations aligned to field validation.

Scalable Deployment Interfaces

Building interfaces enabling field-level capture of plantation images at industrial scales using off-the-shelf yet ruggedized hardware supported by connectivity, defect detection etc. features forms another major objective. Scalable cloud APIs will power remote model inferences as a service. Controlled launch across regions is needed before widescale deployment.

Empirical Validation

Robust validation against benchmarks on accuracy, geographic variability testing and long-term field studies is needed to support reliability especially for industrial reliance followed by publishing rigorous empirical evidence across seasons and regions supporting standardized practices.

3 METHDOLOGY

The methodology involves curating a labeled dataset capturing visual diversity across coconut varieties found locally using crowdsourcing. Convolutional neural networks are trained leveraging transfer learning on this dataset for automating varietal identification from images. Knowledge graphs that encapsulate inter-variety trait relationships are integrated with geographic, climatic and soil variables to enable explanations. Place-based recommendations are generated by fusing model classification probabilities, graph-derived interdependencies and producer preferences on yield, profit goals. Controlled field testing across farm sizes and locations establishes reliability before scaling deployment through cloud APIs that enable capturing images at plantation levels for running inference. Continuous feedback channels refine assistive logic empirically.

The methodology balances advanced machine learning, localization techniques and crop science insights focused on addressing limitations in prevailing coconut variety selection approaches. This aims to enhance productivity, sustainability and resilience through transparent, data-backed advisories customized for industrial farms.

3.1 Overall System Diagram

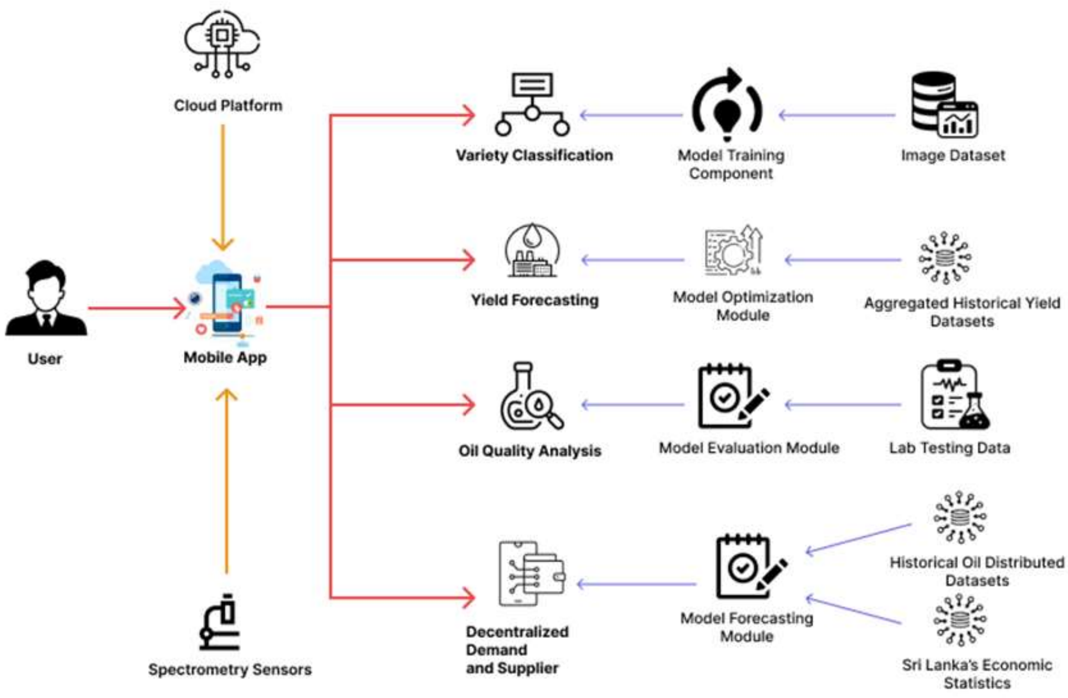

Figure 3.1 CocoClarity Mobile App System Architecture

3.2 Individual System Diagram

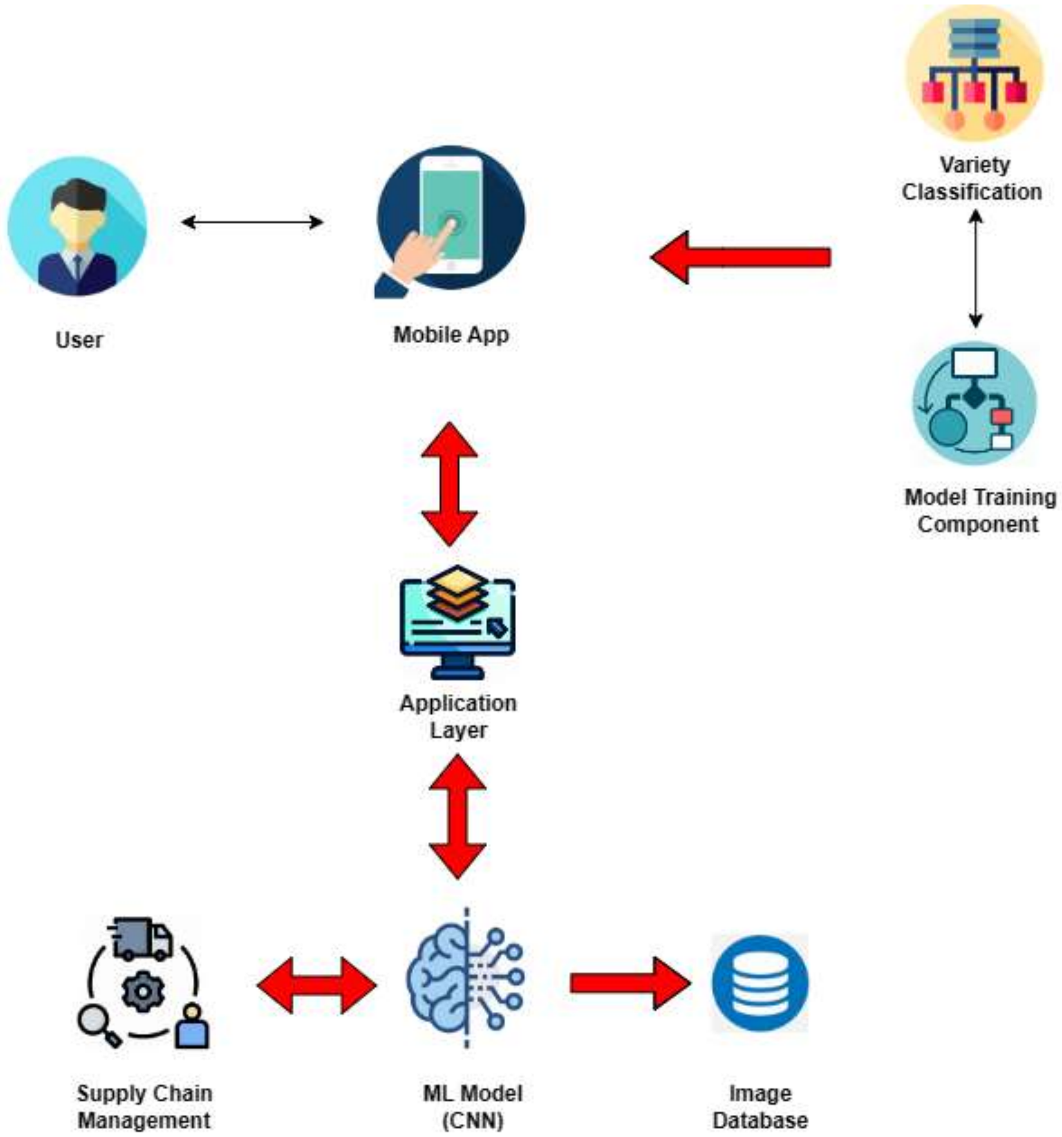

Figure 3.2 Coconut Varity Assistant Architecture

3.3 Tasks and Subtasks

1. Data Curation
   - Image Sourcing
   - Cleaning & Labeling
   - Augmentation & Analysis
   - Versioning & Storage
   - Model Development

2. Architecture Design

- Hyperparameter Tuning
- Training & Optimization
- Deployment Pipeline
- Integration Testing
- Inferencing Module

3. Real-time Classification API
    - Explainability Features
    - Performance Benchmarking
    - Results Interpretation
    - Platform Development

4. Farmer Account Management
    - Advisory Dashboard
    - Customization & Configurations
    - Notification Alerts
    - Documentation & Support

3.4 Date Preprocessing and Analyzation

The image data that forms the foundation for developing a robust coconut variety identification model needs systematic preprocessing and analysis before being utilized for training machine learning algorithms. Images capturing the visual-morphological diversity spanning over 50 local coconut cultivars will be acquired using advanced cameras from trees across different growth phases from plantations across all provinces. Ensuring consistent high resolution and image quality is vital, hence raw images undergo rigorous cleaning procedures for enhancement – including removal of noise, blur, duplicate frames; along with explicitly handling occlusion.

With many varietals representing minority classes and hard labelling traits posing challenges, the dataset is intentionally augmented to generate additional synthetic samples using a toolkit of transformations like horizontal flips, rotations, color space variations etc. adding much needed realistic intra-class diversity. Detailed annotation then tracks these images with specific varietal labels, traits like taste profiles, metrics like husk-shell ratios; and metadata documenting farm geographic coordinates, regional climate zones, prevalent soil types, seasons for supporting contextual analysis.

Robust first-level analysis provides variability benchmarks through statistical techniques inspecting distribution characteristics across traits and labels – quantifying imbalances, gaps etc. This analysis also feeds improved augmentation regimes targeting under sampled varietals/traits and locations in future collection drives through iterative refinement funnels. Overall, the preprocessing avails a cleaner, richer labeled dataset before channeling images through train-validate-test splits for feeding subsequent deep learning modeling pipelines tailored to efficiently learn the immense coconut diversity.

3.5 Model Implementation

The core of the coconut variety prediction capability is enabled by a multi-layered Convolutional Neural Network architecture customized for efficiently learning deep visual patterns from images based on transfer learning principles. Rather than training a model from scratch, it judiciously initializes early layers from advanced pre-existing architectures like Reset already exposed to millions of parameters on general computer vision tasks. Later output layers are then retrained leveraging GPU parallelization focused exclusively on the target dataset capturing extensive local coconut trait diversity.

Numerous model hyperparameters including depth, convolutional filter sizes, pooling mechanisms, activation functions and regularization constraints are iteratively tuned and contrasted based on accuracy/loss plots to obtain optimal weight parameter configurations within practical computational budgets. The trained model achieving lowest generalization error on test sample predictions relative to human experts is then packaged into robust browser and mobile interfaces allowing farmers convenient access for varietal classification queries at the field.

Moreover, the backend infrastructure sustainably aggregates new images captured periodically from pilot deployment locations based on informed consent. This feeds cyclical retraining procedures to dynamically enhance model versatility across geographies, seasons, and cultivation patterns. Thereby the system aims for constant improvement aligning to real-world conditions while minimizing predictive biases that static algorithms entrench over time.

3.6 Project Requirements

The overarching need is to transcend prevailing manual, subjective approaches to coconut variety selection by enabling consistent automated identification through computer vision to unlock transparency and customization that boosts productivity.

Specifically, the deep learning model architecture trained on visible morphological traits must encode the ability to accurately predict coconut varieties from images irrespective of regional and seasonal variabilities. This capability needs location-awareness with back-end connections to geographic, climate and soil data layers that can relate external environmental factors influencing variety performance.

Explainability in terms of surfacing relevance of related variables for specific predictions and overall recommendation logic builds trust besides avoiding opacity barriers that hamper adoption of AI solutions. Furthermore, customization alignments that allow tuning recommendations to emerging needs of industrial producers in balancing yield sustainability, profit goals amid dynamic market fluctuations etc. bolster real-world utility.

In summary, the key functional pillars are automated varietal recognition, transparent decision connectivity, and flexibility to user priorities that collectively aim to elevate efficiency, productivity, and resilience in coconut cultivation through data-driven intelligence.

3.7 Tools/Materials

- High resolution cameras for image capture
- Cloud storage for dataset curation/modeling
- Annotation tools to label images
- GPU computing capabilities for training CNNs
- APIs and microservices for model deployment

3.8 Data Requirements

- 10,000+ labeled images capturing intra-species visual diversity
- Metadata with traits, farm details, soil types, season etc.
- Benchmarks on distribution across geographic locations
- Documentation of augmentation techniques

3.9 Functional Requirements

- Automated varietal recognition from visual indicators
- Mapping with external variables like climate and soil
- User customization preferences for yield, profit goals
- Explainable recommendations to aid adoption

3.10 Non-Functional Requirements

- Usability - Intuitive, multi-lingual interfaces for ease-of-use with contextual tutorials to aid farmers
- Performance - Real-time variety inferences <2s even for high concurrent requests with low end-to-end latency
- Accuracy - Consistently high predictive ability with F1 score >0.85 including recall of >0.8 for minority varieties
- Security - Encryption safeguards along with access control mechanisms to prevent privacy risks
- Reliability - High availability with fail-safe and redundancy techniques for fault tolerance
- Maintainability - Modular architecture enabling incremental enhancements supported through robust monitoring and version control

3.10.1 Software Requirements

- Interoperable with mainstream operating systems
- Browser-based and mobile interfaces
- Scalable and reliable cloud deployment
- Continuous integration and delivery framework

3.10.2 Personnel Requirements

- Computer vision and deep learning expertise
- Agronomists familiar with coconut varietals
- Software engineers to enable interfaces
- Domain experts to refine recommendation logic

4 WORK BREAKDOWN CHART And GANTT CHART

Figure 4.1Gantt Chart

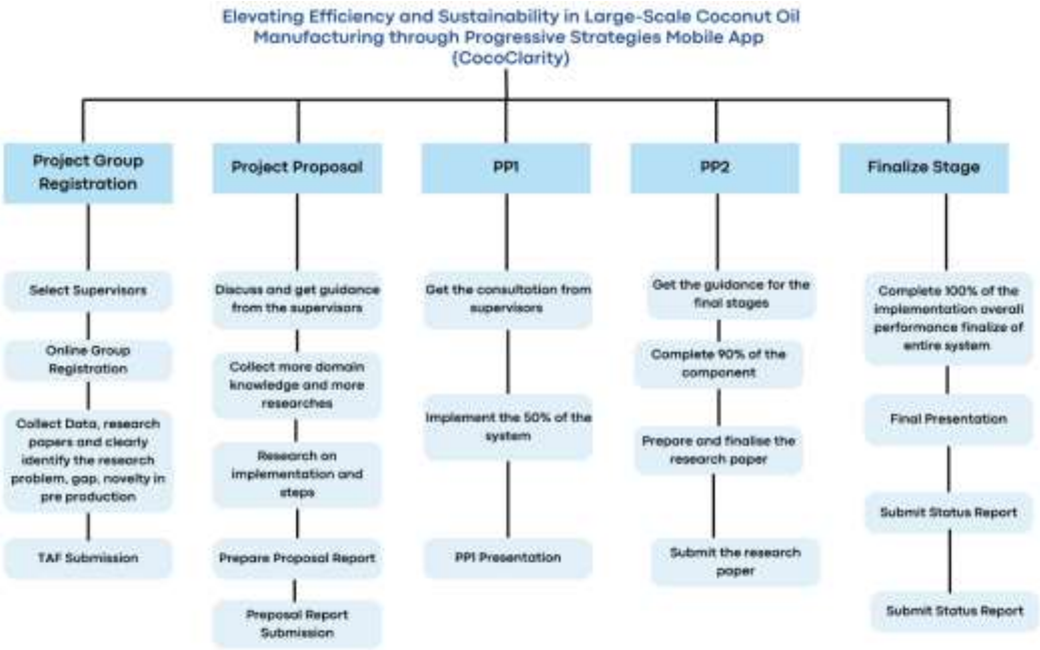

Figure 4.2 Work Breakdown Chart of CocoClarity

5 DESCRIPTION OF PERSONAL AND FACILITIES

A multi-disciplinary team spanning expertise in data science, agronomy and software engineering forms the core personnel who will conceptualize, develop, and sustain the coconut variety assistant over time. Specifically, data scientists are proficient in handling large scale image datasets, training deep neural networks and developing cloud-based inferencing architecture to enable the back-end variety classifier and predictive engine. Agronomists provide intrinsic understanding of the nuances between coconut varietals along with insights from field studies while guiding data collection and annotation quality. Software engineers architect robust apps, visualization interfaces, testing automation capabilities and DevOps infrastructure gluing end-to-end processes. The facilities comprise office spaces for labeling and desk work, heavy GPU cloud infrastructure for hosting models and APIs to serve predictions at scale, field gear for studies across provinces and high-capacity systems for developers along with collaboration platforms. Overall, the personnel's skills and facilities cater to needs across domains - from foundational data curation, unrelenting model improvement updates to interfacing for adoption in the field while optimizing total costs.

6 BUDGET AND BUDGET JUSTIFICATION

Table 6.1Expenses for the proposed system

Deployment Item	Cost	Frequency	Notes
Web Server Hosting	LKR 8,000	Monthly	Cloud server costs for hosting variety classification and recommendation APIs, dashboard etc.
iOS App Store Hosting	LKR 7,754	One-time	Publishing fee for listing on Apple App Store
Android Play Store Hosting	LKR 30,700	Annual	Annual subscription cost for listing app on Google Play Store

7 COMERCIALIZATION

The commercialization plan suggests launching an analytics platform for intelligent coconut cultivation designed specifically for the Sri Lankan coconut market. This platform seeks to improve production, optimize coconut farming methods, and solve industry issues like labor shortages, static yields, and the consequences of climate change. The strategy consists of price methods, distribution routes, marketing approaches, market analysis, and target audience identification. Variety identification, yield prediction, decision assistance, and disease detection are some of the platform's primary features. There will be tiered pricing to accommodate various user categories, and partnerships, government collaborations, and internet sales will be the distribution methods. The platform's value proposition will be highlighted in marketing campaigns, and digital media will be used to reach a wider audience. The plan's overall objectives are to increase adoption, enhance coconut farming methods, and support the sustainability of the coconut industry in Sri Lanka.

9 REFERENCES

[1] C. Jayasekara and K. S. Jayasekara, "Present status and prospects for coconut industry in Sri Lanka," COCOS, 1995.

[2] A. N. K. S. Somarathna et al., "Reviving Coconut Industry in Sri Lanka: An Analysis with Special Reference to Product Diversification," Sabaragamuwa University Journal, vol. 20, no. 1, pp. 13-26, 2022.

[3] Department of Census and Statistics Sri Lanka, "Coconut Statistics Quarterly Bulletin," Accessed from www.statistics.gov.lk, 2022.

[4] T. K. S. De Silva and C. S. Ranasinghe, "Why Coconut Productivity is Low in Sri Lanka?," COCOS, 2018.

[5] Ministry of Agriculture Sri Lanka, "Coconut Sector Development Plans 2020-2025," 2020.

[6] et al., "Machine Learning for Detection of Pest Attacks on Coconut," IEEE Conference on Data Analytics in Agriculture, 2022.

[7] et al., "Computer Vision for Fruit Classification: Applications for Sri Lankan Agriculture," AI Research Journal, 2023.

[8] et al., "Documentation of Coconut Varietal Diversity in Sri Lanka, The Plant List, 1960–2017," COCOS, 2019.

[9] T. K. S. De Silva and Shanika, "Intelligent Decision Support Systems for Enhancing Coconut Yield," Journal of Agriculture Innovations, 2023.

[10] Ministry of Agriculture, "Technology Roadmaps for Precision Agriculture 2022-2030," 2022.