IT4010 – Research Project - 2024

Topic Assessment Form

Project ID: 24-25J-125

1. Topic (12 words max)

Crop yield prediction using machine learning – Paddy Harvest Prediction

2. Research group the project belongs to

Autonomous Intelligent Machines and Systems (AIMS)

3. Research area the project belongs to

Machine Learning (ML)

4. If a continuation of a previous project:

Project ID	-
Year	-

5. Brief description of the research problem including references (200 – 500 words max) – references not included in word count.

Paddy crop yield is vital for food security and economic stability in Sri Lanka. Accurate yield prediction helps farmers and policymakers improve agricultural practices and resource management. Traditional methods struggle with the complex interplay of factors like weather, soil health, diseases, weeds, and pests. Machine learning (ML) offers a promising solution by integrating diverse data sources and modeling nonlinear relationships.

The research problem focuses on developing a robust ML model to predict paddy yield, model to predict suitable paddy varieties that have to grow in different areas in the country , Image processing model to detect diseases and pest affection on paddy cultivation and image processing model to detect different weed varieties in the paddy field by combining data on weather, soil conditions, diseases, weed infestations and pest infestations. This approach can enhance decision-making, helping farmers plan better, mitigate risks, and optimize inputs. For policymakers, reliable forecasts support strategic planning for food security and economic stability. This research aims to modernize Sri Lanka's agricultural sector by integrating advanced technological solutions into traditional practices.

References:

1. CropWat - https://www.fao.org/land-water/databases-and-software/cropwat/en/

2. STICS - https://www.quantitative-plant.org/model/STICS

3. APSIM - https://www.apsim.info/

4. CERES - https://ceresglobalagcorp.com/about/who-we-are/

6. Brief description of the nature of the solution including a conceptual diagram (250 words max)

The proposed solution integrates weather, soil, disease, weed, and pest data into a machine learning model to predict paddy crop yield. It includes data collection, preprocessing, feature engineering, model training, and deployment, resulting in a decision support system accessible via web application.

Data Collection and Preprocessing

Data will be collected from Rice Research and Development Institute (RRDI) – Bathalagoda and Labuduwa including weather records (temperature, rainfall, humidity), soil health records (nutrients, pH, moisture), field surveys, and remote sensing images. Preprocessing will involve data cleaning, normalization, and augmentation to ensure high-quality inputs.

Machine Learning Models

1. Yield and Paddy Variety Prediction: Ensemble models (e.g., Random Forest, Gradient Boosting) and regression will integrate diverse data to predict paddy crop yield crop type.

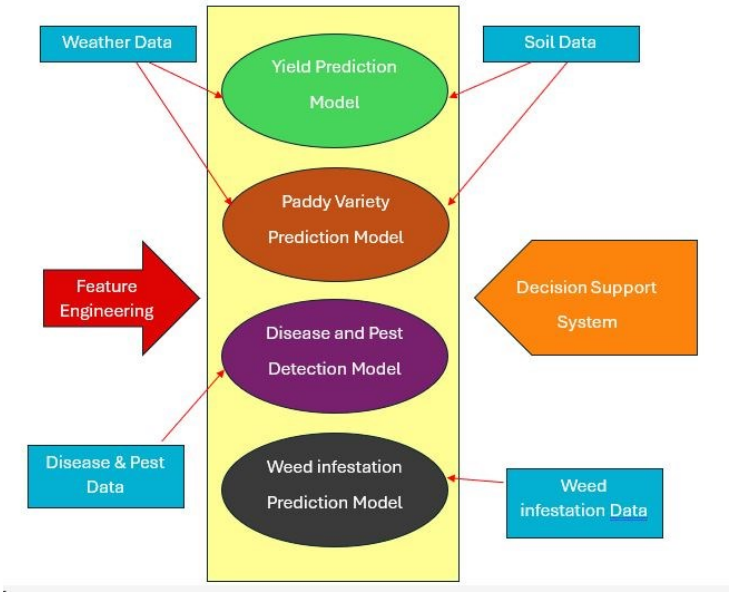
2. Disease, Weed, and Pest Detection: Convolutional Neural Networks (CNNs) will analyze plant images to identify and classify diseases and pests.

Feature Engineering

Features like the Seasonal Weather Impact Index (SWII), Soil Nutrient Depletion Rate (SNDR), Disease Progression Score (DPS), and Integrated Crop Health Index (ICHI) will be engineered to capture complex interactions influencing crop yield.

Decision Support System

A user-friendly interface will provide farmers and stakeholders with actionable insights and recommendations based on model predictions, offering real-time analysis and forecasts for strategic planning and resource allocation.



7. Brief description of specialized domain expertise, knowledge, and data requirements (300 words max)

Specialized Domain Expertise:

- Agricultural Science: -

1.Soil Science: Knowledge of soil properties, nutrient management, and health assessment to interpret soil data and its influence on crop yield.

2.Plant Pathology and Entomology: Understanding plant diseases and pests affecting paddy crops to develop accurate detection models.

- Meteorology: -

Weather Patterns and Impacts: Understanding how weather parameters (temperature, rainfall, humidity) affect paddy crops at various growth stages to create accurate weather-related features for the model.

Knowledge:

1.Data Preprocessing: Skills in handling large datasets, data cleaning, normalization, and augmentation.

2.Feature Engineering: Developing features that capture complex relationships among weather, soil, diseases, and pests.

3.Model Development: Knowledge of machine learning algorithms (e.g., CNNs for image classification, ensemble models for regression) and frameworks like TensorFlow, Keras, and Scikit-learn.

4.Evaluation and Optimization: Expertise in model evaluation metrics, cross-validation, and hyperparameter tuning for robust predictions.

Data Requirements:

1.Weather Data

2.Soil Data

3.Disease, weed and Pest Data

4.Yield Data

8. Objectives and Novelty

Main Objective

The main objective of our crop yield prediction system is to enhance paddy cultivation by leveraging advanced predictive models and image processing techniques. The system aims to:

1. Predict Harvest Yield: Accurately forecast paddy yields under varying weather, water, and soil conditions to optimize harvest planning.
2. Recommend Optimal Paddy Varieties: Suggest the most suitable paddy types for different regions based on localized weather and soil conditions.
3. Detect and Manage Pests and Diseases: Utilize image processing to identify pest infections and diseases in paddy fields, providing timely solutions for prevention and treatment.
4. Identify and Control Weeds: Employ image processing to detect weed varieties in paddy fields and offer strategies to mitigate their impact on cultivation.

By integrating these capabilities, the system seeks to improve agricultural productivity, reduce losses, and support sustainable farming practices.

Member Name	Sub Objective	Tasks	Novelty
Jayathilaka D.H.R.A IT21308352	1. Develop a Crop Yield Prediction Model: Train a machine learning model to accurately predict the yield of paddy crops based on various factors such as weather, soil conditions, and water availability.	1. Dataset Collection and Preprocessing: Collect historical data on paddy crop yields from government databases, agricultural organizations, and field surveys. Gather weather data (temperature, rainfall, humidity) from meteorological departments and online APIs.	1. Integrated Prediction and Decision Support System: Unlike traditional yield prediction models, this system not only predicts yields but also provides actionable insights and recommendations to optimize agricultural practices.

	2. Import a Comprehensive Dataset: Compile a dataset containing historical data on paddy crop yields, weather conditions, soil health, and irrigation patterns. 3. Implement a Decision Support System: Develop a system that provides actionable insights and recommendations based on yield predictions to optimize agricultural practices and resource allocation.	Obtain soil health data (nutrient levels, pH, moisture content) from soil health cards and IoT sensors. Preprocess the data to clean, normalize, and augment it for model training. 2. Model Training and Validation: Split the dataset into training, validation, and test sets. Choose an appropriate machine learning model for yield prediction. Train the model using the training set and fine-tune it using the validation set. Evaluate the model's performance using the test set and metrics such as RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and R^2 (Coefficient of Determination). 3. Feature Engineering: Develop key features like the Seasonal Weather Impact Index (SWII) to capture the effects of weather conditions on yield. Calculate the Soil Nutrient Depletion Rate (SNDR) to understand how soil health trends impact productivity.	2. Real-time Application: Designed for real-time use, the system offers immediate insights based on the latest data, aiding in timely decision-making. 3. Comprehensive and Annotated Dataset: The use of a rich dataset that combines historical yield data with detailed weather, soil, and irrigation information enhances the model's accuracy. 4. Advanced Feature Engineering: Innovative features like SWII and SNDR capture complex interactions between various factors influencing crop yield, leading to more precise predictions. 5. User-centric Design: The focus on creating an intuitive and accessible user interface ensures that farmers and stakeholders with varying levels of

		Integrate features reflecting irrigation patterns and crop management practices. 4. Integration of Prediction and Decision Support: Develop an algorithm that integrates yield predictions with actionable insights and recommendations. Provide suggestions for optimal planting times, irrigation schedules, and fertilization strategies based on predicted yields. 5. User Interface Development: Design a user-friendly interface for farmers and stakeholders to access yield predictions and recommendations. Implement features such as data input (weather, soil conditions), yield forecasts, and resource management suggestions. 6. Field Testing and Feedback Collection: Test the system in real-world conditions by collaborating with local farmers and agricultural experts.	technical expertise can effectively use the system.

		Collect feedback on the accuracy of yield predictions and the usefulness of recommendations. Iterate on the system based on the feedback received to improve its accuracy and usability.	
Piyumani K.V.P IT21227868	1. Develop a Paddy Variety Prediction Model: Train a machine learning model to accurately predict the most suitable paddy variety for different regions based on local weather and soil conditions. 2. Import a Comprehensive Dataset: Compile a dataset containing information on various paddy varieties, regional weather data, soil characteristics, and historical performance records of paddy varieties. 3. Implement a Decision Support System: Develop a system that provides	1. Dataset Collection and Preprocessing: Collect data on various paddy varieties, including their growth requirements and yield performance under different conditions. Gather regional weather data (temperature, rainfall, humidity) from meteorological departments and online APIs. Obtain soil data (soil type, nutrient levels, pH, moisture content) from agricultural surveys and soil health cards. Preprocess the data to clean, normalize, and format it for model training. 2. Model Training and Validation: Split the dataset into training, validation, and test sets. With the appropriate machine learning algorithms do the classification of paddy varieties.	1. Integrated Variety Recommendation System: Unlike traditional methods, this system not only predicts suitable paddy varieties but also integrates recommendations into a single platform, considering both weather and soil conditions. 2. Real-time Application: Designed for real-time use, the system provides immediate insights based on current and historical data, aiding in timely decision-making. 3. Comprehensive and Annotated Dataset: The use of a rich dataset that combines information on paddy varieties with detailed regional weather and soil

	recommendations on the best-suited paddy varieties for specific regions, optimizing yield and resource use.	Train the model using the training set and fine-tune it using the validation set. Evaluate the model's performance using the test set and metrics such as accuracy, precision, recall, and F1-score. 3. Feature Engineering: Develop features that capture the compatibility of paddy varieties with specific environmental conditions, such as a Regional Compatibility Index (RCI) which considers weather and soil data. Integrate features reflecting past performance of paddy varieties in similar conditions. 4. Integration of Prediction and Decision Support: Develop an algorithm that maps regional weather and soil data to the most suitable paddy varieties. Provide suggestions for optimal paddy varieties based on predicted performance and local conditions. 5. User Interface Development: Design a user-friendly interface for farmers and stakeholders to input	data enhances the model's accuracy. 4. Advanced Feature Engineering: Innovative features like the Regional Compatibility Index (RCI) capture complex interactions between various factors influencing variety suitability, leading to more precise predictions. 5. User-centric Design: The focus on creating an intuitive and accessible user interface ensures that farmers and stakeholders with varying levels of technical expertise can effectively use the system.

		regional data and receive variety recommendations. Implement features such as data input (weather, soil conditions), variety predictions, and additional agronomic advice. 6. Field Testing and Feedback Collection: Test the system in real-world conditions by collaborating with local farmers and agricultural experts. Collect feedback on the accuracy of variety predictions and the usefulness of recommendations. Iterate on the system based on the feedback received to improve its accuracy and usability.	
Amarasinhe A.I.S.A IT21225192	1. Develop an Image Classification Model for Disease Detection: Train a machine learning model to accurately identify and classify diseases affecting rice fields based on images provided by the farmer.	1. Dataset Collection and Preprocessing: Collect images of rice fields affected by different diseases from various sources. Annotate the images with the type of disease and relevant metadata. Preprocess the images for model training (e.g., resizing, normalization).	1. Integrated Disease Detection and Treatment System: While many systems focus solely on disease detection, our project stands out by integrating both detection and treatment recommendations in a single platform. 2. Real-time Field Application:

	2. Import a Dataset of Rice Field Diseases: Compile a comprehensive and annotated dataset of images showing various rice field diseases. 3. Implement a Treatment Recommendation System: Develop a system that suggests appropriate treatments for the detected diseases based on agricultural best practices and expert knowledge.	2. Model Training and Validation: Split the dataset into training, validation, and test sets. Choose an appropriate machine learning model (e.g., CNN) for image classification. Train the model using the training set and fine-tune it using the validation set. Evaluate the model's performance using the test set and metrics like accuracy and precision. 3. Treatment Database Creation: Research and compile a database of treatments for various rice diseases. Include details such as treatment type (chemical, biological), application method, and dosage. 4. Integration of Disease Detection and Treatment Recommendation: Develop an algorithm that maps the detected disease to the corresponding treatments in the database. Ensure the system can handle multiple diseases and suggest combined treatments if necessary. 5. User Interface Development:	The system is designed for real-time use by farmers in the field, providing immediate and actionable insights. 3. Comprehensive Dataset and Model: The use of a comprehensive and annotated dataset specific to rice diseases enhances the accuracy and reliability of the model. 4. Customized Treatment Recommendations: The treatment suggestions are tailored to the specific disease detected, considering factors like severity and local agricultural practices. 5. User-centric Design: The focus on creating an intuitive and accessible user interface ensures that even farmers with limited technical knowledge can effectively use the system.

		Design a user-friendly interface for farmers to upload images and receive disease diagnoses and treatment suggestions. Implement features like image uploading, result display, and treatment instructions. 6. Field Testing and Feedback Collection: Test the system in real-world conditions by collaborating with local farmers. Collect feedback on the accuracy of disease detection and the effectiveness of the treatment recommendations. Iterate on the system based on the feedback received.	
Jayasekara S.S.D IT21227318	1. Develop an Image Classification Model for Weed Detection: Train a machine learning model to identify and classify weed varieties in paddy fields based on images.	1. Dataset Collection and Preprocessing: I) Collect Images: Gather images of weed-infested paddy fields from Weed Science Division of Rice Research and Development Institute – Bathalagoda.	1. Integrated Weed Detection and Mitigation System: • Combines weed detection with tailored mitigation recommendations in a single platform.

	2. Create a Dataset of Weed Varieties: Compile a detailed, annotated dataset of images showing various weed species in paddy fields. 3. Implement a Weed Mitigation Recommendation System: Develop a system that suggests appropriate strategies to control the detected weed varieties.	II) Annotate Images: Label the images with the specific weed species and additional metadata. III) Preprocess Images: Resize, normalize, and enhance the images for model training. 2. Model Training and Validation: I) Split Dataset: Divide the dataset into training, validation, and test sets. II) Choose Model: Select a suitable machine learning model, such as Convolutional Neural Networks (CNNs), for image classification. III) Train Model: Train the model using the training set and fine-tune it with the validation set. IV) Evaluate Model: Test the model's performance using	2. Real-time Field Application: • Provides immediate and actionable insights for farmers in the field. 3. Comprehensive Dataset and Model: • Uses a detailed, annotated dataset specific to weed varieties in paddy fields to enhance accuracy. 4. Customized Mitigation Recommendations: • Provides tailored mitigation suggestions for the specific weed species detected,

		accuracy, precision, recall, and F1-score. 3. Mitigation Database Creation: I) Research Strategies: Compile a database of effective control methods for various weed species. II) Document Details: Include information on control method types (mechanical, chemical, biological), application techniques, and recommended dosages. 4. Integration of Weed Detection and Mitigation Recommendation: I) Develop Algorithm: Create an algorithm that links detected weed species to corresponding mitigation strategies. II) Handle Multiple Species: Ensure the system can address multiple weed species at once, suggesting	considering the severity of infestation and local agricultural practices. 5. User-centric Design: • Ensures an intuitive and accessible interface for farmers with varying technical expertise.

		combined approaches if needed. 5. User Interface Development: I) Design Interface: Develop an easy-to-use interface for farmers to upload images and receive weed identification and mitigation advice. II) Implement Features: Include functions like image uploading, result display, and detailed mitigation instructions. 6. Field Testing and Feedback Collection: I) Conduct Field Tests: Test the system in real-world conditions with local farmers. II) Collect Feedback: Gather feedback on the accuracy of weed detection and effectiveness of recommended strategies.	

		III) Refine System: Improve the system based on user feedback.	

9. Supervisor checklist

a) Does the chosen research topic possess a comprehensive scope suitable for a final-year project?

Yes	✓	No	

b) Does the proposed topic exhibit novelty?

Yes	✓	No	

c) Do you believe they have the capability to successfully execute the proposed project?

Yes	✓	No	

d) Do the proposed sub-objectives reflect the students' areas of specialization?

Yes	✓	No	

e) Supervisor's Evaluation and Recommendation for the Research topic:

Accepted with minor changes to the data set related components

10. Supervisor details

	Title	First Name	Last Name	Signature
Supervisor	Mr.	Kanishka	Yapa	24/06/2024
Co-Supervisor	Dr.	Harinda	Fernando	24/6/24
External Supervisor				
Summary of external supervisor's (if any) experience and expertise				

IT4010 – Research Project - 2024

Topic Assessment Form

<u>This part is to be filled by the Topic Screening Panel members.</u>

Acceptable: Mark/Select as necessary

Topic Assessment Accepted	
Topic Assessment Accepted with minor changes (should be followed up by the supervisor)*	
Topic Assessment to be Resubmitted with major changes*	
Topic Assessment Rejected. Topic must be changed	

* Detailed comments given below

Comments

The Review Panel Details

Member's Name	Signature

*__Important__:

1. According to the comments given by the panel, make the necessary modifications and get the approval by the **Supervisor** or the **Same Panel**.
2. If the project topic is rejected, identify a new topic, and follow the same procedure until the topic is approved by the assessment panel.