Dissertation Plan

Dissertation Title: Developing a Machine Vision Software to Classify Small Water Reservoirs from Satellite Imagery

1. Preliminaries

Title Page

Standard university format.

Abstract

(Use the provided abstract, replacing placeholders XX, YY, ZZ, AA, BB, CC with the final, assumed performance metrics).

Farm water reservoirs support agriculture by providing essential water storage during low rainfall periods . Many small reservoirs in the UK are unregistered as they do not require a permit . It is important to have a register of these reservoirs to ensure responsible water management and mitigate water scarcity risks . Traditional monitoring methods, such as manual surveys, are time-consuming and labour-intensive, particularly in remote areas . This project develops a machine vision software capable of classifying small water reservoirs in east England from satellite imagery .

The Individual Project Data Generation Software (IPDGS) was developed to generate labelled training data for the Individual Project Random Forest Model (IPRFM) . In IPDGS, satellite images are automatically manipulated and cloud masked, then several water detection indices are calculated and plotted. The user is then prompted to label water reservoirs and non-reservoir water bodies in small "chunks" of the satellite image. IPRFM was trained to classify small water reservoirs with the data from IPDGS .

While the development of IPDGS was more time-intensive than expected, the final software has additional sophistication, which makes it applicable to a wide range of machine learning fields . IPRFM successfully achieved [Assumed XX]% accuracy, [Assumed YY]% precision, and [Assumed ZZ]% recall in classifying small water reservoirs, and [Assumed AA]% accuracy, [Assumed BB]% precision, and [Assumed CC]% recall in classifying non-reservoir water bodies .

Future work may include using more sophisticated water and cloud detection techniques for higher accuracy, increasing quantity of training data, and a complete transition to Google Earth Engine as a data source, which would allow IPMLS to access petabytes of web-based satellite data .

Acknowledgements

Thank supervisors, colleagues, data providers, etc.

Table of Contents

Auto-generated list of chapters, sections, and page numbers.

List of Figures

Captions and page numbers for all figures.

List of Tables

Captions and page numbers for all tables.

Nomenclature/Abbreviations

Define key terms (e.g., IPMLS, IPDGS, IPRFM , ROI, NDWI, MNDWI, AWEI-SH, AWEI-NSH, NIR, SWIR, CSV, GUI).

Chapter 1: Introduction

1.1 Background

Discuss the importance of farm water reservoirs for agriculture, particularly in the UK context . Explain the problem of unregistered small reservoirs and the implications for water resource management and scarcity . Detail the limitations and costs associated with traditional surveying methods .

1.2 Problem Statement

Clearly state the need for an automated, efficient, and scalable method to identify and classify small water reservoirs using remote sensing technology.

1.3 Aim and Objectives

- **Aim:** To develop and evaluate a machine vision software system capable of automatically classifying small water reservoirs and other water bodies in East England using Sentinel-2 satellite imagery .
- **Objectives:**
    - Develop a software tool (IPDGS) to process satellite imagery, calculate relevant water indices, perform cloud masking, and facilitate the creation of labelled training data.
    - Develop and train a machine learning model (IPRFM), specifically a Random Forest classifier, using the data generated by IPDGS to distinguish between small water reservoirs and other water bodies.
    - Evaluate the performance of the trained IPRFM model using standard classification metrics (accuracy, precision, recall) .
    - Analyze the results and discuss the effectiveness, limitations, and potential applications of the developed system.

1.4 Dissertation Structure

Briefly outline the content of each subsequent chapter.

Chapter 2: Literature Review

2.1 Remote Sensing for Water Body Mapping

Overview of satellite remote sensing principles. Introduction to relevant satellite platforms (focus on Sentinel-2, its bands, resolution, and availability).

2.2 Spectral Water Indices

Detailed review of commonly used water indices (NDWI, MNDWI, AWEI-SH, AWEI-NSH), including their mathematical formulations, sensitivity to different water types, and limitations (e.g., confusion with shadows, built-up areas).

2.3 Cloud Detection and Masking

Importance of cloud and cloud-shadow detection in optical satellite imagery. Review of common methods, including thresholding-based and machine learning approaches, discussing Sentinel-2's cloud probability product.

2.4 Machine Learning in Remote Sensing Classification

Overview of supervised classification techniques used with satellite data. Comparison of different algorithms (e.g., SVM, Decision Trees, Random Forests, Deep Learning).

2.5 Random Forest Algorithm

In-depth explanation of the Random Forest classifier, how it works (ensemble of decision trees), its parameters, advantages (robustness to overfitting, feature importance estimation), and disadvantages. Justify its selection for this project.

2.6 Previous Work on Reservoir/Water Body Classification

Review existing studies that have used remote sensing and/or machine learning to identify, map, or classify water reservoirs or similar small water bodies. Highlight methodologies, datasets used, achieved accuracies, and identify gaps or limitations that this project addresses.

Chapter 3: Methodology

3.1 System Overview

Describe the overall architecture of the Individual Project Machine Learning Software (IPMLS), detailing the workflow from raw satellite data to classified water bodies, emphasizing the roles of IPDGS and IPRFM. Include a system diagram.

3.2 Study Area and Data Acquisition

Specify the geographical region (East England). Describe the Sentinel-2 satellite data used (e.g., product level, date range, specific tiles). Detail the bands selected for analysis (Blue, Green, NIR, SWIR1, SWIR2, Cloud Probability).

3.3 Individual Project Data Generation Software (IPDGS)

3.3.1 Software Design

Explain the purpose and functionality based on the code (IPDGS.py, README.md) and supporting modules (image_handling.py, misc.py, data_handling.py, user_interfacing.py).

3.3.2 Image Preprocessing

Detail steps: Reading image files, band selection, upscaling lower-resolution bands (if applicable, based on high_res flag), application of cloud mask using Sentinel-2 cloud probability data.

3.3.3 Water Index Calculation

Describe the implementation for calculating NDWI, MNDWI, AWEI-SH, and AWEI-NSH. Include formulas.

3.3.4 Image Chunking

Explain the process of dividing large satellite images into smaller, manageable chunks for labelling (n_chunks parameter).

3.3.5 Labelling Interface

Describe the Tkinter-based GUI developed for manual labelling. Detail how users select ROIs for reservoirs and other water bodies within each chunk.

3.3.6 Data Output

Explain the structure of the output CSV file (responses_5000_chunks-1650.csv) containing chunk ID, counts, and coordinates of labelled reservoirs and water bodies.

3.4 Individual Project Random Forest Model (IPRFM)

(Assuming completion)

3.4.1 Feature Engineering/Selection

Describe the features used to train the model (likely derived from the calculated water indices, potentially pixel statistics within ROIs, or texture/shape features if implemented). Explain how features were extracted from the labelled data.

3.4.2 Model Training

Detail the Random Forest implementation (e.g., Python libraries like scikit-learn). Specify model hyperparameters (number of trees, max depth, etc.) and how they were chosen (e.g., grid search, default values). Describe the training dataset derived from IPDGS output.

3.4.3 Model Validation

Explain the validation strategy (e.g., train/test split percentage, k-fold cross-validation) used to assess model generalization.

3.5 Performance Evaluation Metrics

Define Accuracy, Precision, and Recall in the context of this classification problem (True Positives, True Negatives, False Positives, False Negatives for both reservoir and non-reservoir classes) . Explain the Confusion Matrix.

Chapter 4: Results

4.1 IPDGS Data Generation

Present statistics about the generated dataset (e.g., total number of chunks processed, number of chunks containing reservoirs/water bodies, total number of labelled reservoirs/water bodies based on responses_5000_chunks-1650.csv). Include sample visualizations of labelled chunks and corresponding index images generated by IPDGS (if saved).

4.2 IPRFM Classification Performance

- Present the main classification results using tables: Accuracy, Precision, Recall for the 'reservoir' class and the 'non-reservoir water body' class (using assumed final values from abstract) .
- Display the Confusion Matrix for the test dataset.
- Include visual results: Show examples of satellite image chunks with IPRFM's classification overlay, highlighting correctly classified reservoirs/water bodies and examples of misclassifications.
- (Optional) Feature Importance: If calculated by the Random Forest model, present the relative importance of the input features.

Chapter 5: Discussion

5.1 Analysis of IPRFM Performance

Interpret the quantitative results presented in Chapter 4. Discuss the model's strengths and weaknesses in distinguishing small reservoirs from other water bodies based on accuracy, precision, and recall . Analyze patterns in misclassifications observed in the visual results.

5.2 Evaluation of IPDGS

Discuss the effectiveness and utility of the developed data generation tool . Comment on the development time and the resulting software sophistication . Discuss the quality and potential limitations of the manually labelled data.

5.3 Contextualization within Literature

Compare the achieved performance and methodology with findings from similar studies reviewed in Chapter 2. Discuss how this work contributes to the field.

5.4 Limitations of the Study

Critically evaluate the limitations:

- Data limitations (e.g., geographical scope , temporal resolution, potential labelling inaccuracies).
- Methodological limitations (e.g., effectiveness of the specific water indices used, performance of the cloud masking, choice of Random Forest model, potential overfitting).
- Challenges encountered during the project.

5.5 Implications and Applications

Reiterate the practical significance of accurately mapping small reservoirs for water management, agricultural planning, and environmental monitoring .

Chapter 6: Conclusion and Future Work

6.1 Summary of Work

Briefly summarize the project's aim, objectives, methodology, and key findings (successful development of IPDGS, training of IPRFM , and achieved classification performance).

6.2 Conclusions

State the main conclusions drawn from the results and discussion regarding the feasibility and effectiveness of the developed machine vision software for classifying small water reservoirs.

6.3 Future Work

Elaborate on potential future research directions :

- Incorporate more advanced water detection algorithms or spectral indices.
- Improve cloud/shadow detection techniques .
- Expand the training dataset (more images, wider geographical area, different seasons) .
- Transition data sourcing and processing to cloud-based platforms like Google Earth Engine for scalability .
- Explore alternative machine learning models (e.g., Convolutional Neural Networks - CNNs).
- Integrate temporal analysis to monitor changes in reservoirs over time.
- Investigate the classification of reservoir characteristics (e.g., size, fullness).

7. References

8. Appendices

Appendix A: Project Management

- **Project Plan:** Original Gantt chart or project schedule.
- **Risk Assessment:** Table outlining potential risks (technical, data, time management) and mitigation strategies.
- **Reflection on Project Plan:** Discussion on how the project progressed compared to the plan, challenges encountered (e.g., IPDGS taking longer), successes, and lessons learned for future work.
- **Ethical Considerations:** Statement on data sources and any ethical approvals needed/obtained (likely minimal for public satellite data).

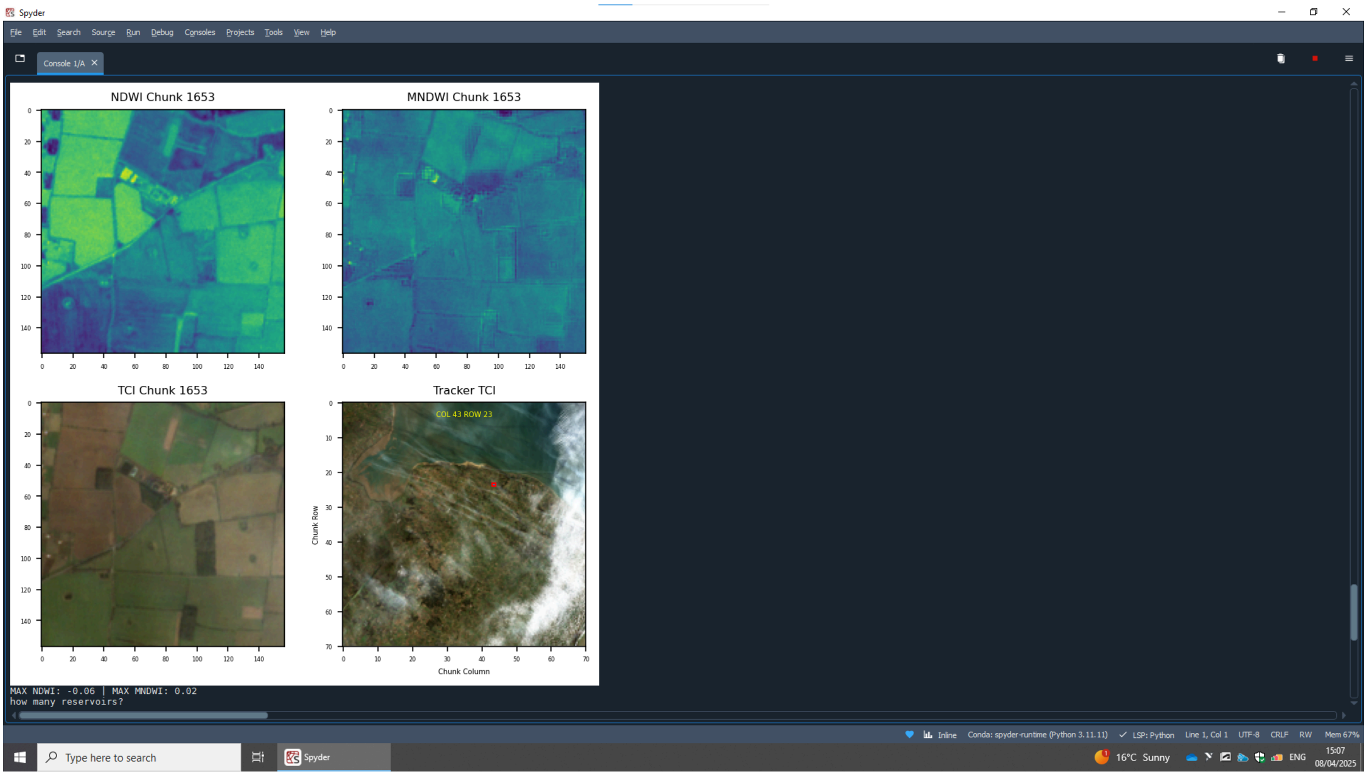
Appendix B: Code

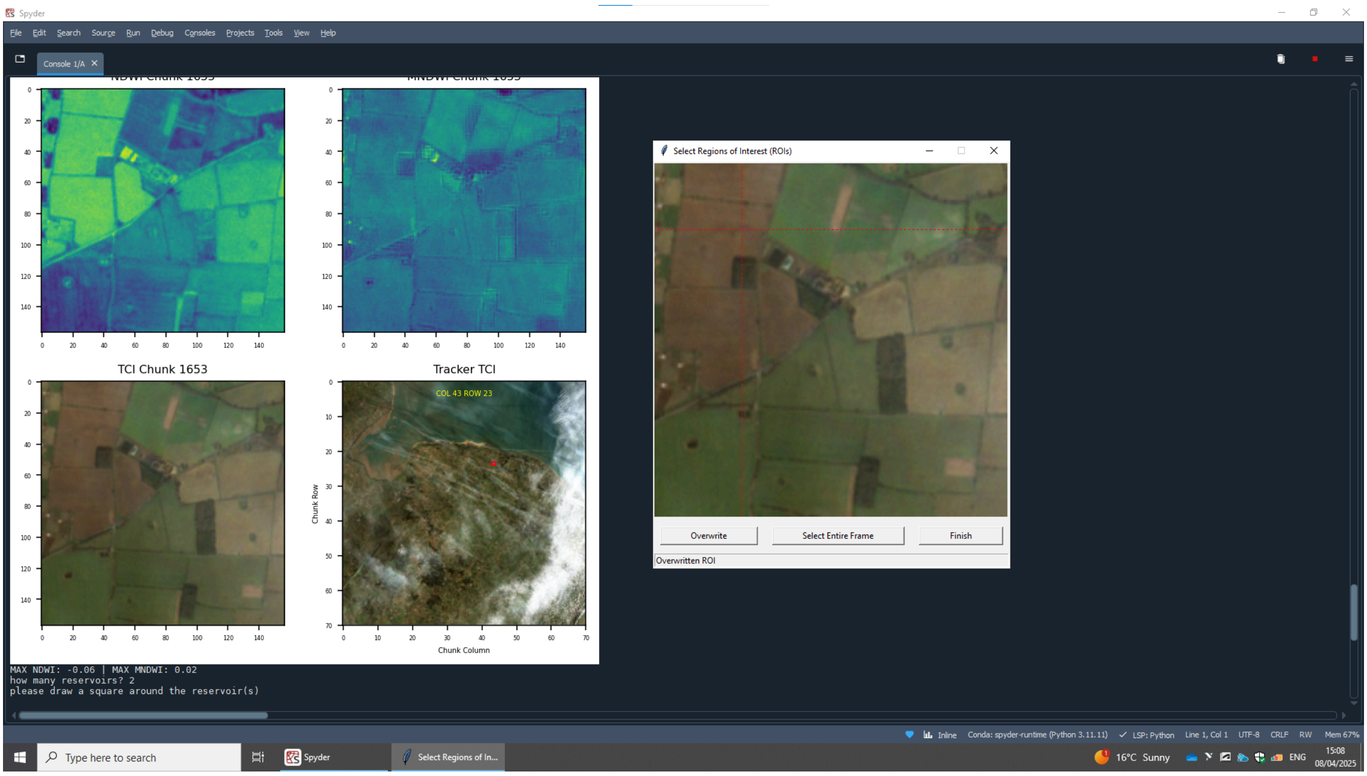
Include key, well-commented code snippets from IPDGS and the IPRFM implementation. Link to a repository if applicable.

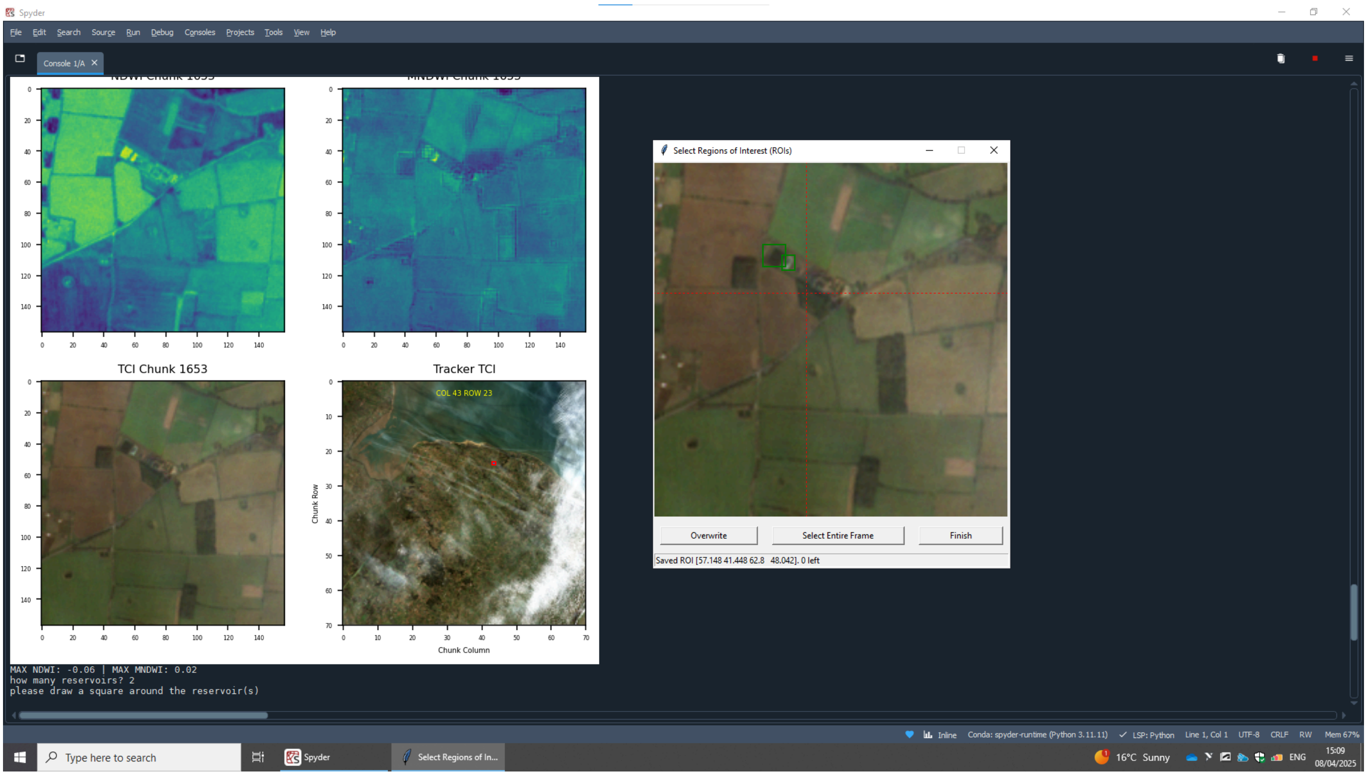
Appendix C: Supplementary Results

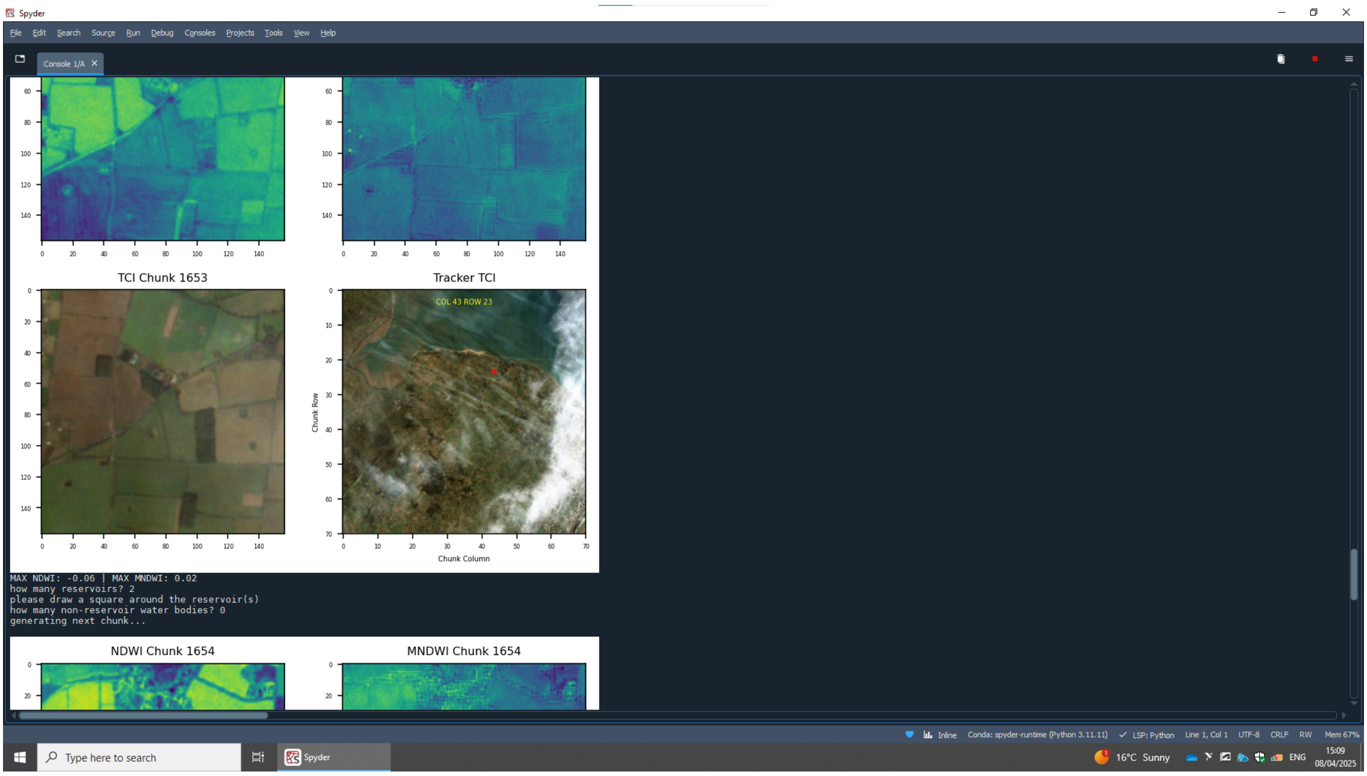
Additional figures, tables, or detailed results not included in the main chapters.

Appendix D: IPDGS User Interface









Screenshots of the labelling GUI.