# Certificate in Scientific Computation and Data Sciences Research Contract

# Addendum (Milestones and Project Roadmap)

This project involves the development and implementation of a machine learning model focused on the classification and detection of various characteristics of Atta leafcutter-ant images sourced from iNaturalist. Presented here is a detailed roadmap encompassing all stages of the project, from initial dataset preparation to the final stages of research analysis and paper writing, along with a carefully structured timeline designed to guide the project towards its successful completion. The student is expected to adhere strictly to the proposed deadlines, ensuring systematic progress and timely completion of both the modeling aspects and the research paper. This document will serve as an addendum to the existing research project agreement between the student and the faculty member, outlining the responsibilities and expectations for this endeavor.

## 1. January 16 - February 5 (3 weeks): Dataset Preparation

1. Data Collection
    a. Download the images using the URLs provided in the Excel worksheet.
    b. Collect additional images if necessary to balance the dataset.
2. Data Annotation
    a. Manually label a subset of images with bounding boxes around the ants for object detection.
    b. Annotate the images with tags for classification (e.g., alate, worker, male, female, alive, dead).
    c. Use semi-automated tools to accelerate the annotation process if possible.
3. Data Augmentation
    a. Apply data augmentation techniques to increase dataset size and variability (e.g., rotations, flipping, scaling, color variations).
4. Data Preprocessing
    a. Resize images to a uniform size required by the chosen models.
    b. Normalize pixel values for model input.
    c. Split the dataset into training, validation, and test sets.

## February 6 - February 19 (2 weeks): Initial Literature Review and Model Selection

5. Initial Literature Review
    a. Conduct a preliminary review of existing literature on image recognition in entomology, focusing on similar studies.
    b. Document methods and findings from relevant studies for the Introduction and Background sections of the paper.
6. Model Selection
    a. Pre-trained Model Evaluation
        i. Review and select appropriate pre-trained models for object detection (e.g., Faster R-CNN, YOLO, SSD).
        ii. Select pre-trained models for image classification (e.g., ResNet, Inception, EfficientNet).
    b. Baseline Model Setup
        i. Set up a simple model to establish baseline performance.
        ii. Evaluate the performance and adjust the model selection if necessary.
    c. Documentation for Paper
        i. Start drafting the Methodology section, describing the criteria for model selection.

# 7. February 20 - March 12 (3 weeks): Model Training

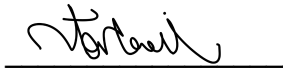
    a. Object Detection
        i. Fine-tune the chosen object detection model on the annotated subset to locate ants in images.
        ii. Validate and adjust the model based on performance metrics (e.g., mAP, IoU).
    b. Image Classification
        i. Fine-tune the chosen classification model on the labeled dataset to identify the types and conditions of ants.
        ii. Implement multi-class classification strategies to handle multiple labels per image (e.g., sex, life stage).
    c. Model Optimization
        i. Experiment with hyperparameters to optimize model performance.
        ii. Utilize techniques like cross-validation for more robust model evaluation.
    d. Methodology Documentation
        i. Continuously document the training process, parameter tuning, and model architectures for the Methodology section.

# 8. March 13 - March 26 (2 weeks): Model Evaluation and Methodology Documentation

    a. Performance Metrics
        i. Use appropriate metrics to evaluate object detection (e.g., precision, recall) and classification accuracy.
        ii. Perform error analysis to understand and improve model shortcomings.

b. Iterative Improvement
    i. Refine annotations and retrain models as necessary based on performance evaluations.
    ii. Explore advanced techniques like ensemble methods or custom architectures for improved results.
c. Results Compilation
    i. Record and compile results, including tables and charts, for the Results section.
    ii. Perform statistical analysis as necessary to validate findings.

9. March 27 - April 9 (2 weeks): Research Analysis and Drafting Paper
    a. Analysis and Discussion
        i. Analyze the results in the context of the objectives.
        ii. Discuss the significance, implications, and limitations of the findings.
        iii. Compare results with existing literature.
    b. Drafting Paper Sections
        i. Introduction: Cover background information and the problem statement.
        ii. Literature Review: Expand on the initial review with findings from the project.
        iii. Methodology: Detail the dataset, models used, and the training process.
        iv. Results: Present findings with data visualizations.
        v. Discussion: Interpret results, compare with existing studies, and discuss implications.
        vi. Conclusion: Summarize key findings and propose future research directions.
    c. Peer Review and Feedback
        i. Seek feedback from peers or advisors on the draft.
        ii. Revise the paper based on feedback.
10. April 10 - April 20 (1.5 weeks): Finalization and Submission
    a. Final Draft Preparation
        i. Perform thorough proofreading and formatting according to journal guidelines.
        ii. Ensure all citations and references are correctly formatted.

Faculty Supervisor Signature 

Date 12/04/2023

Student Signature 

Date 12/07/2023