SP-104-AIML Active Learning Management

CS 4850 – Fall 2025

August 26, 2025

|  |  |  |  |
| --- | --- | --- | --- |
| A person in a checkered shirt  AI-generated content may be incorrect.  Josh Smith  Team Leader  Developer | Elijah Merrill  Developer | Noah Lane  Documentation | Matthew Hall  Documentation |

**Team Members:**

|  |  |  |
| --- | --- | --- |
| Name | Role | Cell / Alt Email |
| Josh Smith (Team Lead) | Developer | 706-414-2827  [joshuasmith0515@gmail.com](mailto:joshuasmith0515@gmail.com) |
| Elijah Merrill | Developer | 478-283-0811  [elijahmerrill04@gmail.com](mailto:elijahmerrill04@gmail.com) |
| Matthew Hall | Documentation | 678-873-8542  [thematthewhall7@gmail.com](mailto:thematthewhall7@gmail.com) |
| Noah Lane | Documentation | 706-591-2312  [noahlane142@gmail.com](mailto:noahlane142@gmail.com) |
| Sharon Perry | Advisor | 770-329-3895  [sperry46@kennesaw.edu](mailto:sperry46@kennesaw.edu) |

**Abstract**:

We aim to create a program that is intelligent enough to be an image labeling system using an active learning loop to efficiently interpret satellite imagery. The core objective is to minimize human labeling efforts while trying to achieve decent classification accuracy. By integrating an active learning system with uncertain sampling, the system selectively acquires the most informative images for labeling. The system will use tools like Python, mainly for the use of its libraries to help build an active learning framework, Kaggle for datasets, and Azure for testing large taxing runs due to high input of images.

**Platform**: Active learning system for labeling satellite images using Python, Kaggle, & Azure

The project focuses on utilizing Python and open-sourced machine learning libraries. Python is a computer programming language that has gained a lot of traction as a go-to language for programming web applications to full-on automation. To achieve an active learning environment, the framework we’ll be using is modAL. ModAL is an active learning framework model that is built on top of the scikit-learn library. Other libraries like Pandas and NumPy will also be incorporated to help manage and visualize the data. The satellite imagery datasets will be obtained via Kaggle, a platform for obtaining open datasets and resources to help solve real-world problems, and Eurodatacube, a cloud platform with access to a variety of observation data related to Earth to be used for analysis.

Hardware requirements are fairly minimal, requiring a standard desktop setup or laptop with sufficient storage for image datasets along with processing power in order to train our classification models should suffice. For larger experiments, it is possible we may seek GPU support through cloud computing platforms like Azure to speed up model training and evaluation.

**Collaboration Tools:**

Microsoft Teams: For status meetings and weekly/bi-weekly group meetings

GroupMe, Discord, and/or Cellphones (via Call/Text): For more general conversation, simple questions, or conceptual brainstorming on ideas with the project.

**STATEMENT OF PARTICIPATION**

By submitting this assignment, I acknowledge that I will participate in all meetings, communications, deliverables and other tasks necessary to complete the project. If I do not participate, I understand that Professor Perry will meet with me to remedy the situation.