BugBot: Discussion of Different Approaches

Shirley Fong, Keegan Veazey, Le Ju

April 8, 2025

In this report, we discuss the various aspects of our project, BugBot, that could have been approached differently. Based on our experiences with the project, there are a few key changes and directions that could have been implemented. The workflow of BugBot is depicted in Figure 1, showing all the various aspects and how they interact with each other.

The project began with data collection, which spanned approximately 1-2 weeks. Each team member responsible for gathering images for 3-4 insects, totaling 11 classes, comprising of 160 images per class. Once the project proposal was completed, the team immediately initiated the data collection process, which was a combination of web scraping and manual selection. This phase proved to be time consuming, particularly trying to collect 160 images per insect. In hindsight, However, we recognize that allocating more time to data collection could have been beneficial. The initial deadline of 1-2 weeks felt rushed, and given the importance of larger datasets for improving model generalization, a more extended collection period would have likely enhanced our results.

The second phase in our workflow was data preprocessing, which consisted of image hashing, applying data augmentation techniques, splitting the data, etc. A challenge that we encountered during the project was data leakage, which we identified by detecting duplicate images across the training, validation, and testing sets. This experience highlighted the importance of applying image hashing as the first step of the preprocessing pipeline, as visual inspection alone was insufficient for identifying duplicates. Fortunately, we were able to resolve the issue quickly, and it did not significantly impact our project.

From the beginning, our processes (e.g., scripts and pipelines) were automated and modularized. While we successfully structured our workflow, having the final coding rubric earlier would have provided clearer guidance and helped streamline our coding practices. As a result, we needed to refine our scripts at the end of the project, which, while valuable, was a bit time-consuming.

Overall, our project was successful, resulting in a model capable of identifying insects to a certain extent. However, reflecting on our experience, we recognize some areas for improvement. If we were to undertake this project again, we would refine our approach by making strategic adjustments to enhance efficiency, data quality, and overall model performance.

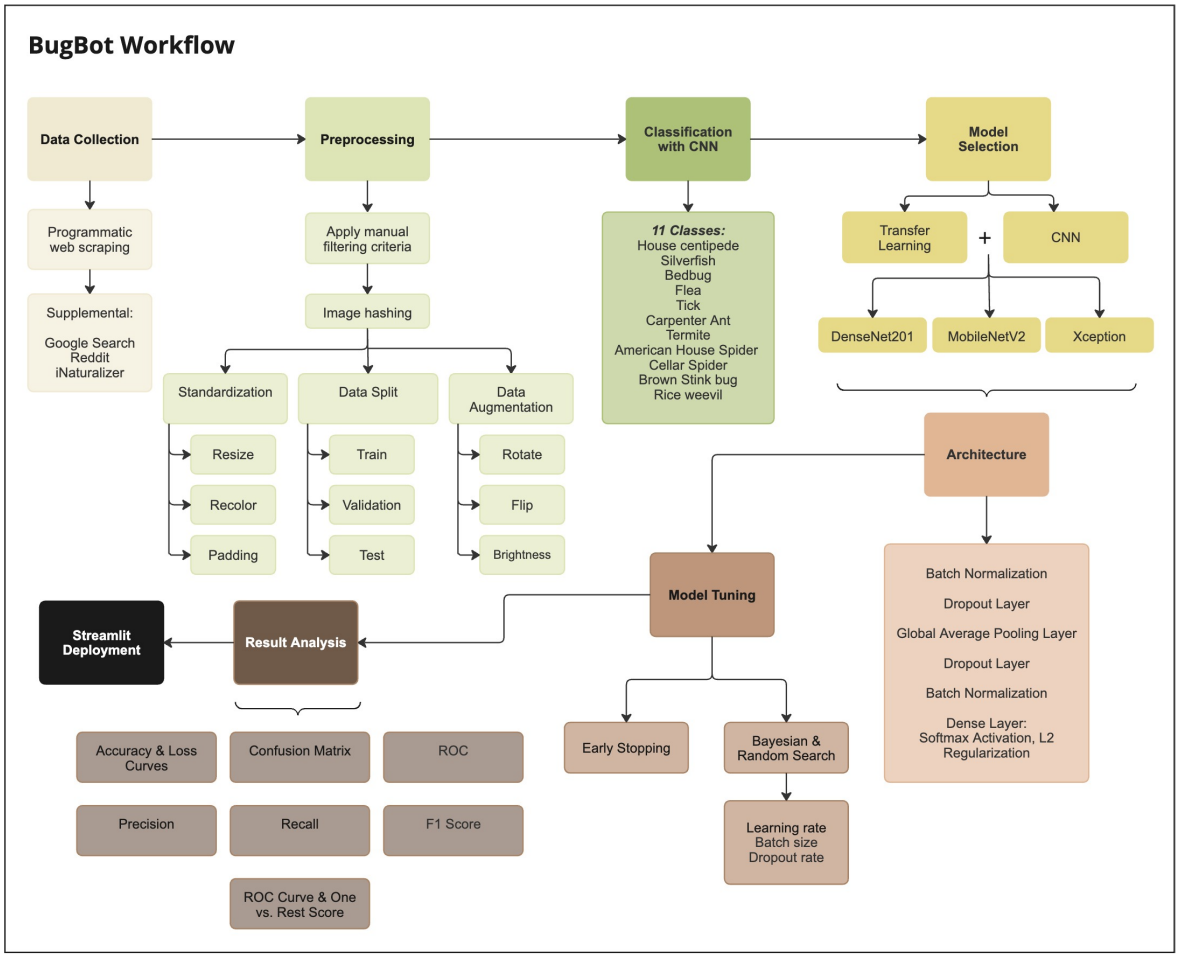

Figure 1: BugBot Workflow