PROTOTYPING

Crop Disease Detection

End of internship presentation

Rania Khemiri (she/her)

Prototyping Architect Intern
AWS

Agenda

Personal background

Business objectives

Key features

Demo

Technical Approach

Next Steps

Mentors

Personal Background

Degree

- Computer Networks and Telecommunications Engineering degree
- No prior ML knowledge
- No prior AWS Cloud knowledge

Internship

- Start date : 3rd April 2023
- End date : 18th August 2023

School

- Location: Toulouse, France

Business Objectives

- Crop monitoring is a labor-intensive task. Regular inspection for disease takes **10% to 30%** of the overall time dedicated on a field.
- Crop loss due to plant disease costs **$220 billion** annually (global economy)
- This application targets **Agronomists** who operate on multiple fileds

Business Objectives

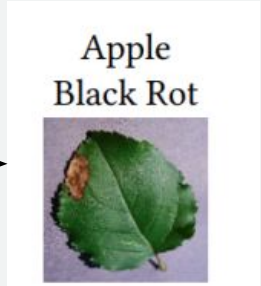
Problem Statement

Currently no viable ML solution is implemented for crop disease detection

Why?

Poor quality or lack of crop dataset

in a lab setting →



on the field →

Objectives and requirements

Quality of data

Leverage agronomists' knowledge of crops

Scalability

Needs to encompass multiple crop types

Usability

Easy enough interface to be used by non-tech users

Key features

- Handle image upload
- Detect **disease** and **crop type**
- Allow data labeling (bounding boxes and class)
- Save image and corresponding annotations
- Enhance initial dataset with new images and annotations
- Monitor model performance and use enhanced dataset to re-train model

Demo

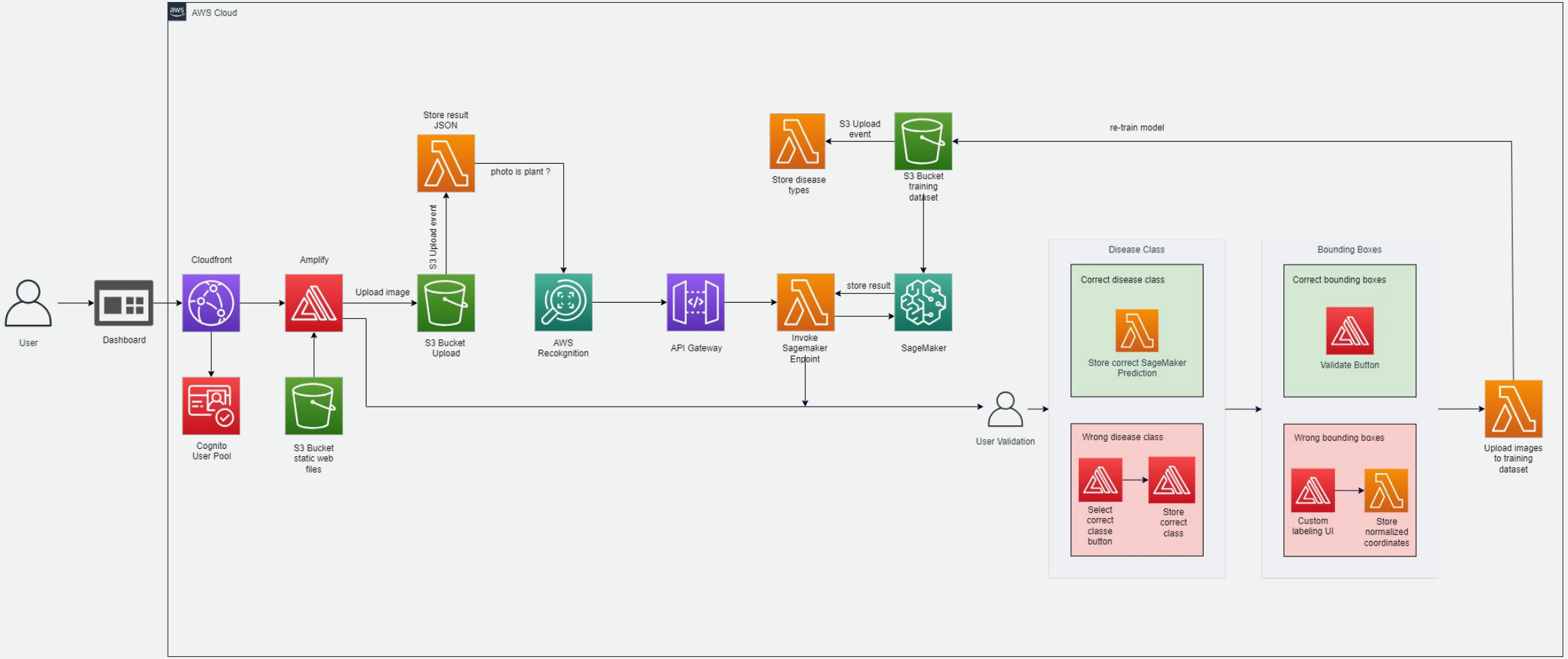
Technical Approach

Preparation Phase

Identifying the problem and customer

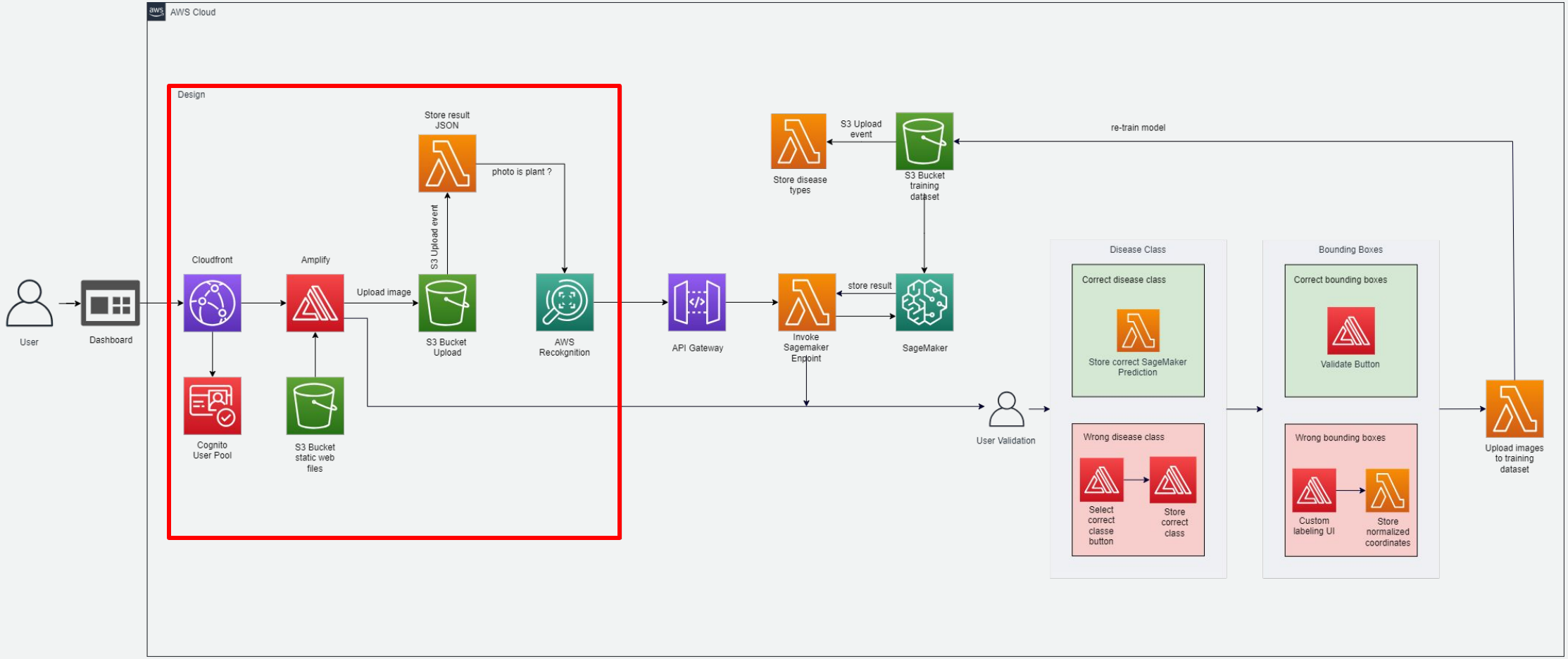
- TFC Input
- Choosing the model
- Defining Architecture

Architecture



Design Phase

Design Phase



Design Phase

Amplify / React

- Front end development
- Back end integration with AWS Services

Cognito

Authentication and authorization

S3

- Storing result
- Storing images

Rekognition

Checking that image is a leaf

CloudFront

Deployment

Image Upload

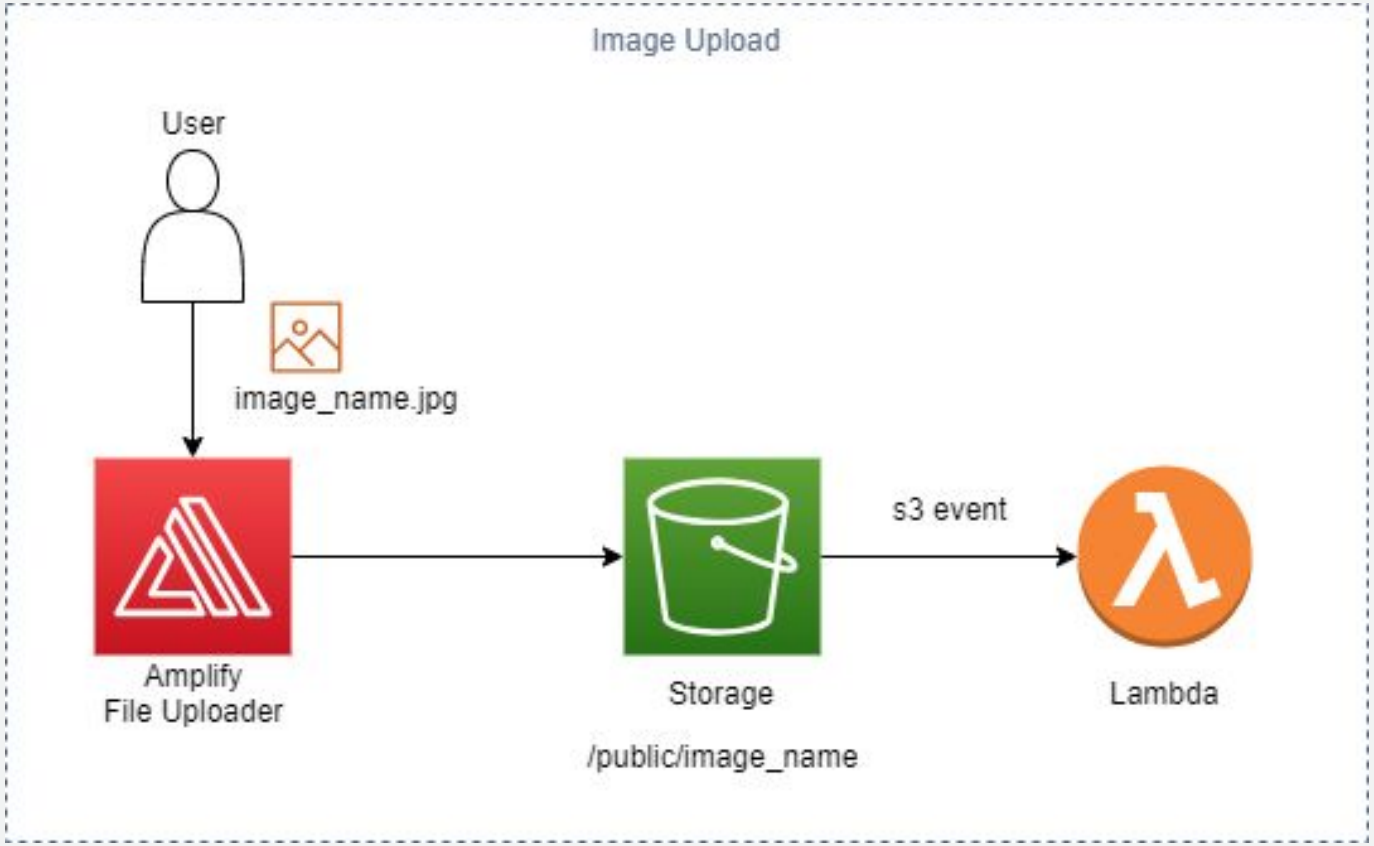
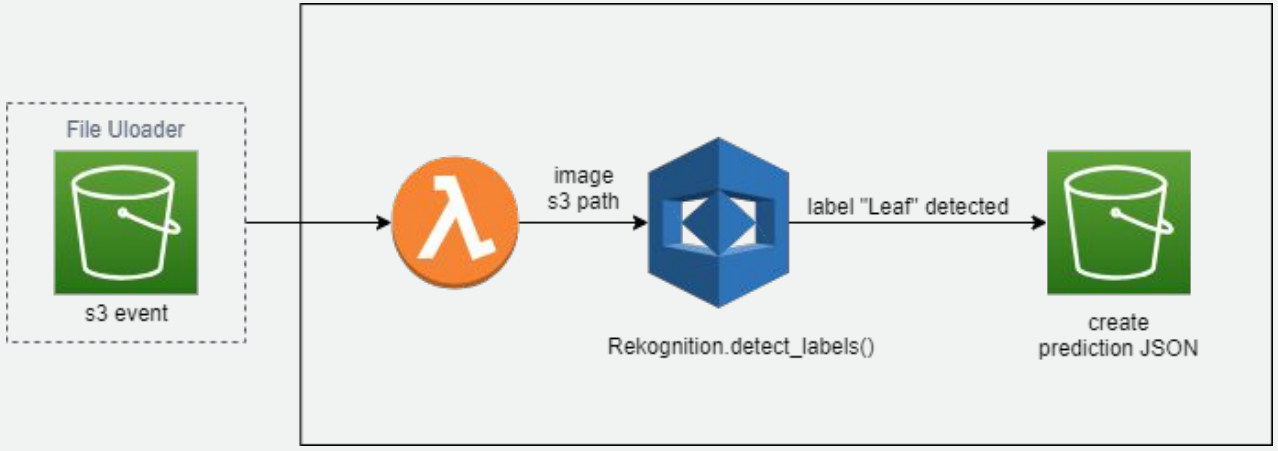
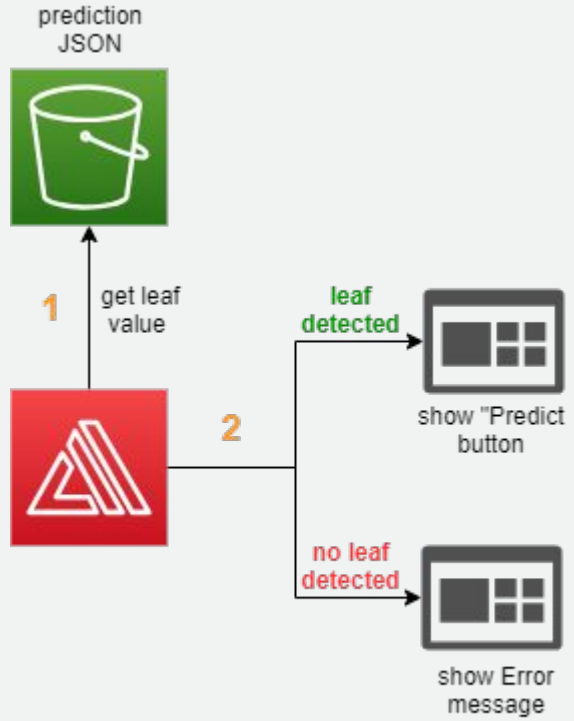


Image Analysis

Step 1

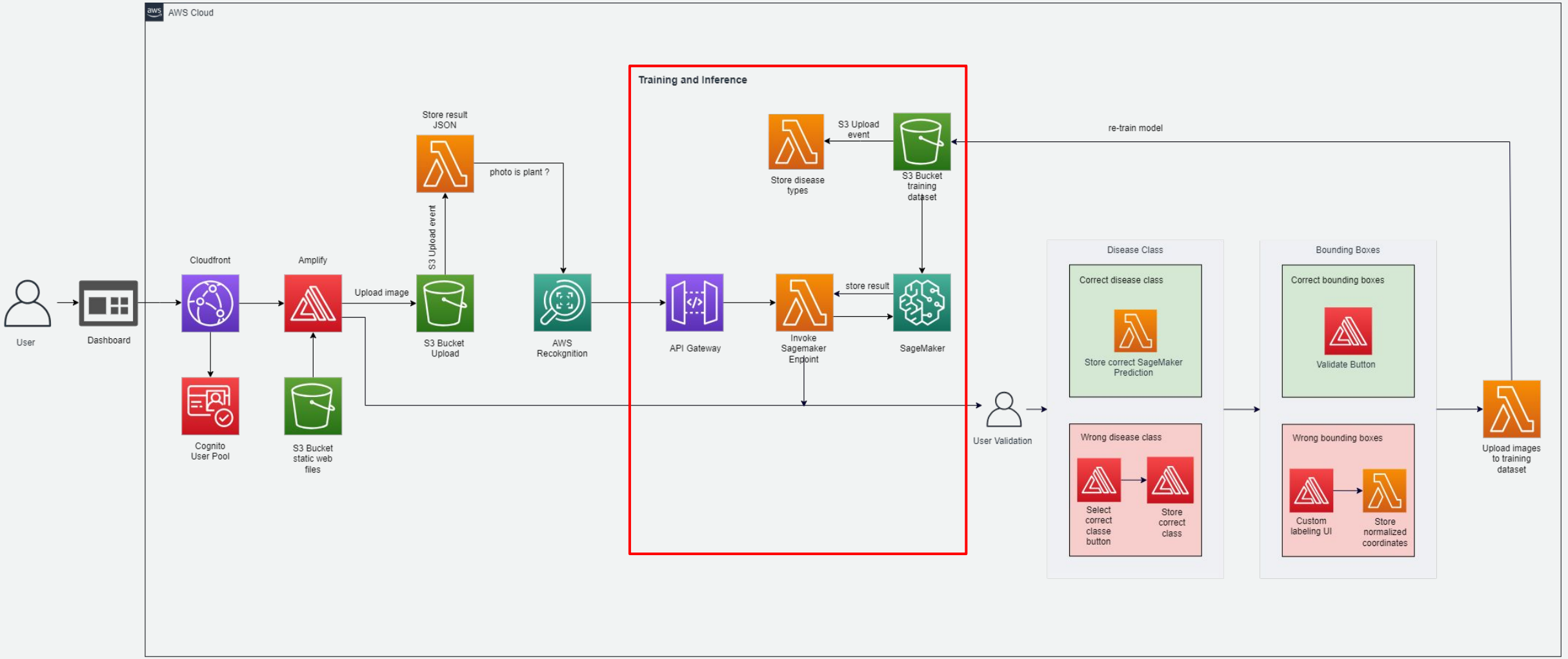


Step 2

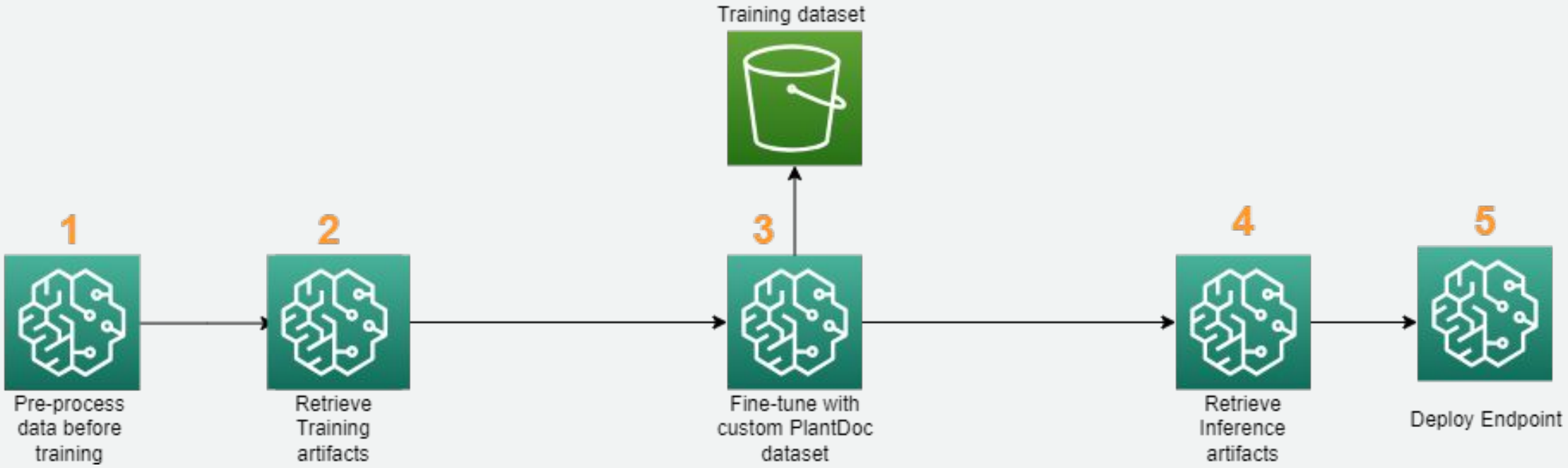


ML Phase

ML Phase



ML pipeline



Data Analysis

Pre-trained Model

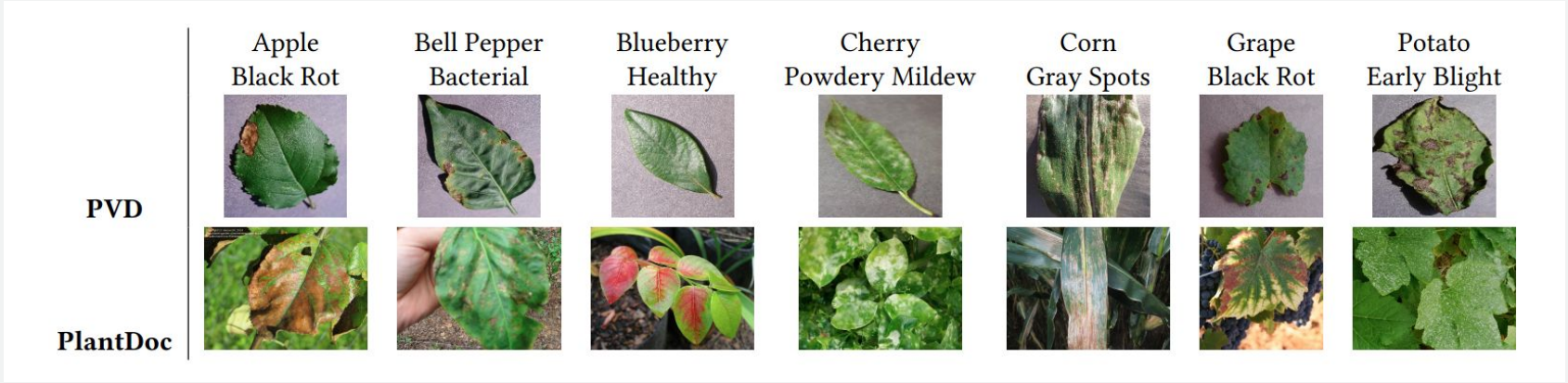
Tensorflow Resnet50

Dataset

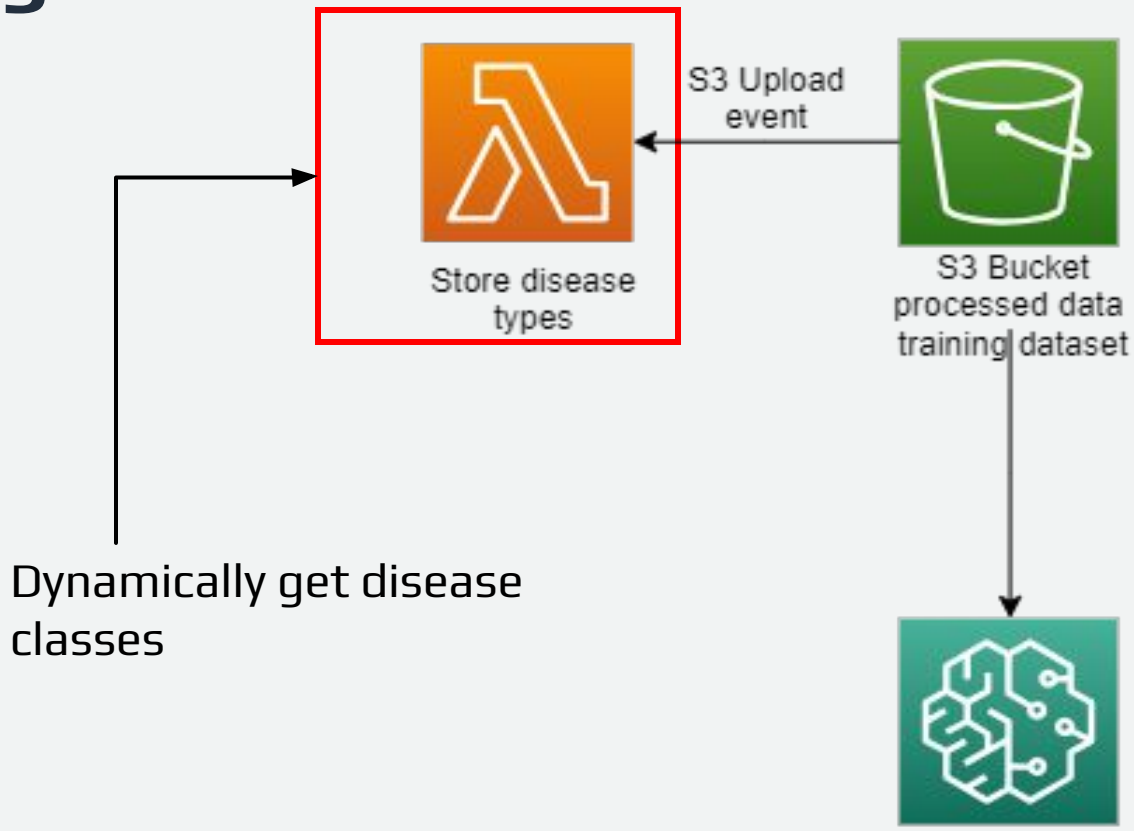
Plantdoc dataset

Pre-processing

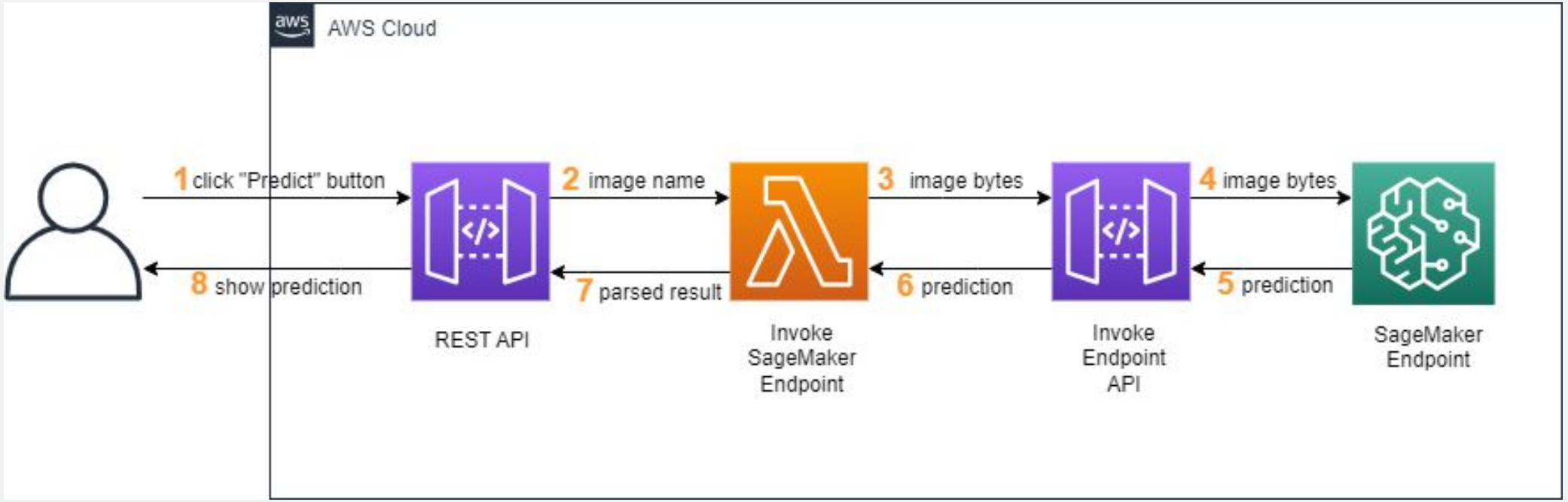
- Image size
- Color channels



Training

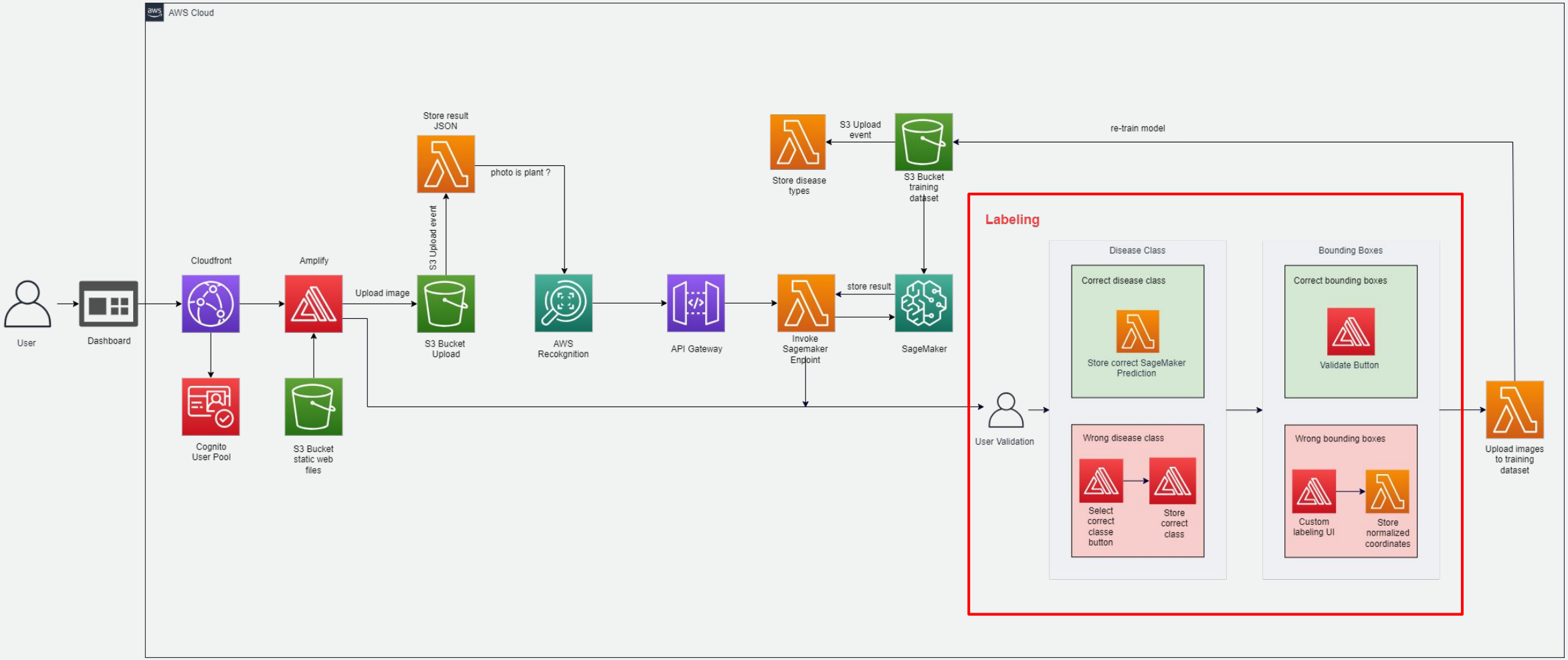


Inference



Data Labeling Phase

Data Labeling Phase



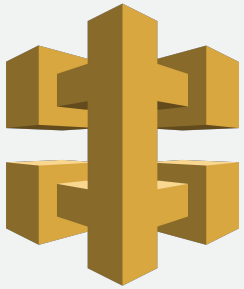
Ground Truth vs Custom Labeling UI

Ground Truth

- Impossible to integrate with existing user pool
- Limited in defining disease classes

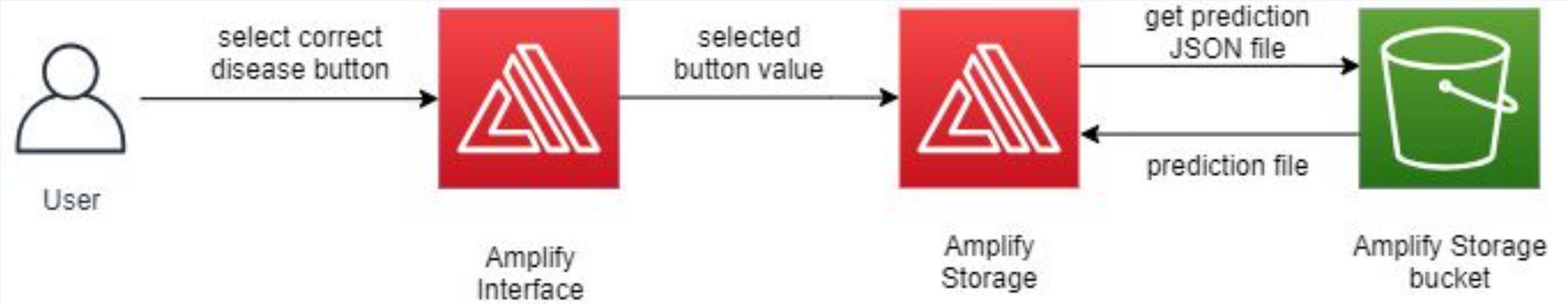
Custom Labeling UI

- Custom UI using Typescript and AWS services

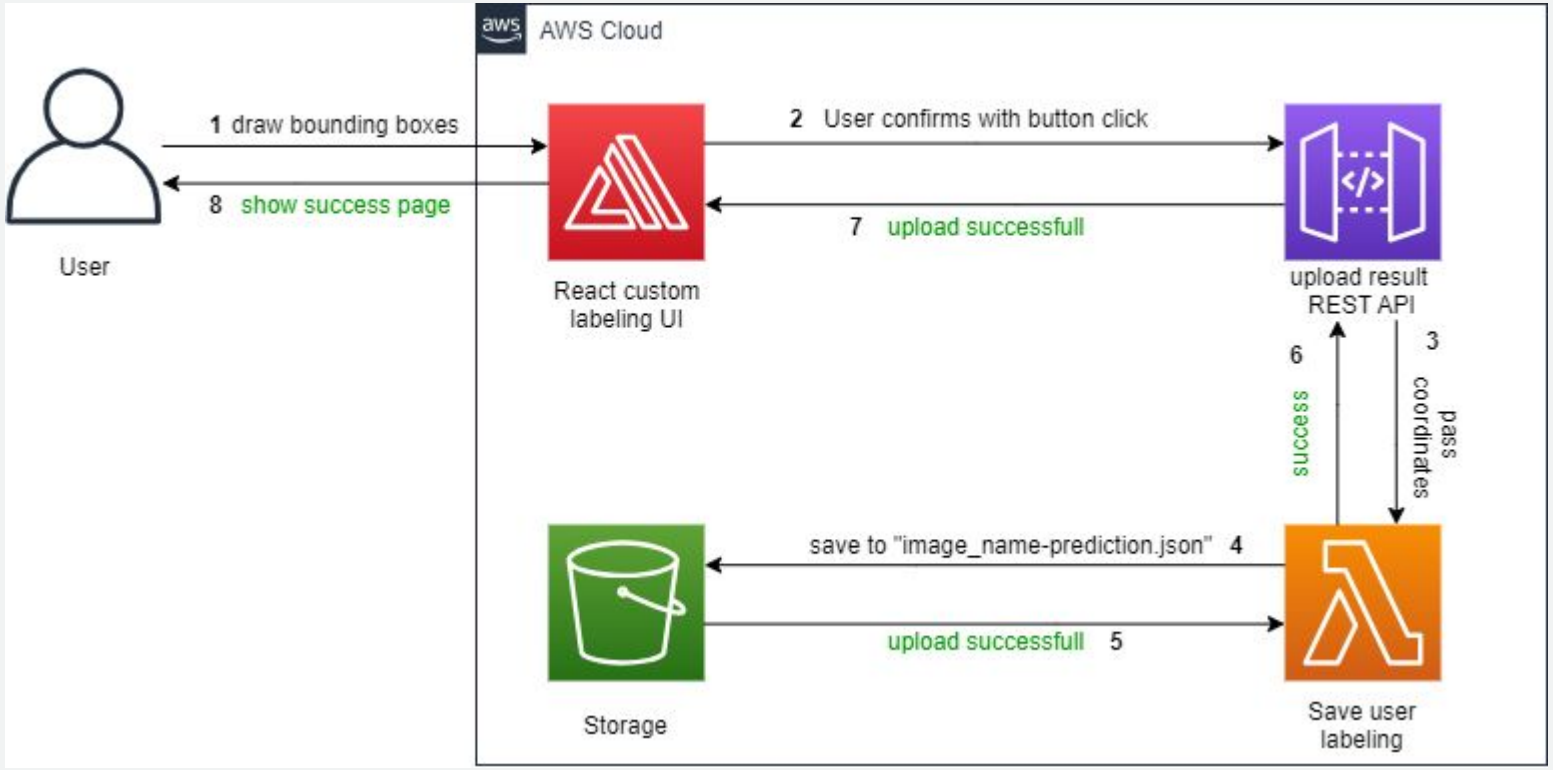


- Leverage AWS Amplify and S3 integration

Correct disease class



Correct bounding boxes



Model Monitoring and retraining

Steps to monitor and retrain model

Accuracy Goal

- Need to define a goal accuracy
- What parameters to implement? New photo numbers per category? Other parameter?
- Periodic retraining?

Sagemaker Pipeline

- Set up a Sagemaker pipeline
- Move the annotated images from the upload bucket to the dataset bucket

Oversampling issue

- Need to deal with inequality of sample size per disease

Alarm to track error percentage

- Track the percentage of user rectifications versus the number of inference
- Set up an alarm to trigger re-training for over 30% of rectifications

Next Steps

	Suggestion	Advancement
Smart Crops	Possible add-on to the smart crops demonstrator	Process in progress
Open source	Open source the project on AWS	Process in progress
TFC Demo	Handover the project to TFC	Process in progress
Handover	Handover to a team member (Paul Devillers)	Process planned

Next Steps : road to optimization

WHAT TO IMPROVE

- Add Infrastructure As Code (Terraform)
- Add other crop types to asses model performance

FEEDBACK FROM TFC

- Keep track of image metadata
- Localize images with metada on a map
- Image classification for crop types before detecting the disease

Fun Experiment : Synthetic Data

Text to image : « Apple rust disease leaf »

Image to image :

Mentors

Paul Devillers

Overall mentorship in different phases

Mohamed Ali Jamaoui

Data Analysis and ML pipeline assistance

Florian Clanet

ML inference and training assistance

Bishesh Adhikari

Data Analysis and Preparation assitance

Ion Kleopas

Labeling Phase assitance

Stephen Hibbert

Labeling Phase assistance

Additional resources

https://gitlab.aws.dev/rkhemiri/crops-disease-detection
Gitlab repository (Internal only)

Thank you!

Rania Khemiri
rkhemiri@amazon.fr