Electrical and Computer Engineering
Erik Jonsson School of Engineering & Computer Science
The University of Texas at Dallas
Dr. Tooraj Nikoubin

Lab 7 – Part2:
TinyML – Image Classification

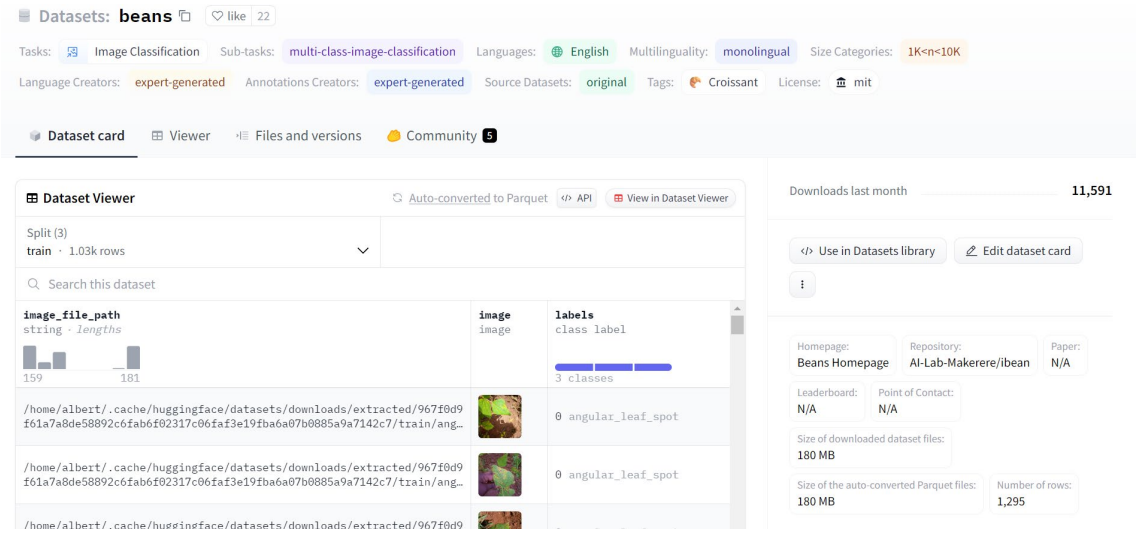
EEDG/CE 6302 Embedded Systems

Image classification using large dataset from Hugging Face

Intro to Hugging Face

The Hugging Face Hub is a large machine learning community and platform with over 18,000 open-source and publicly available datasets, over 120,000 created models, and applications (called Spaces) that leverage AI to perform a task. This open and community-centric approach allows people to easily collaborate and build ML projects together. The Hub is a central place where anyone can explore, experiment, and work together to build projects with machine learning.

Hugging Face Datasets are a library of high quality datasets curated by ML researchers and professionals. We will be using the beans dataset from Hugging Face Datasets in this project, and we'll upload it to Edge Impulse to use for model training and then for deployment on an edge device.



Working with the Dataset

First, open the dataset on the Hugging Face website. Next, download it to your local computer by following the instructions shown:

1. Open a terminal or command line on your machine, and navigate using “cd” to the directory that you want to save the dataset and run the below code:
   (If git is not installed in your system, please install it before proceeding. You can follow the below links for the guide to install git:

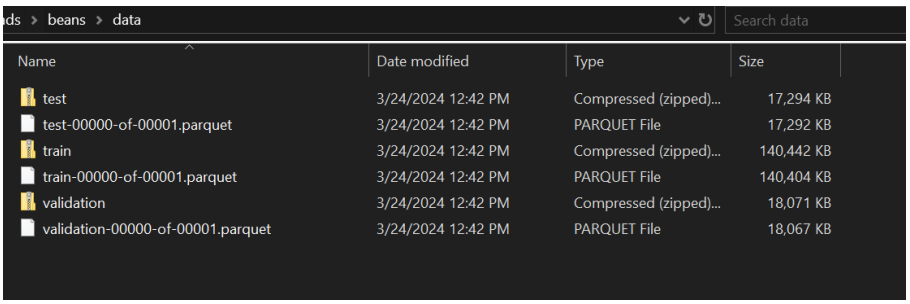
1. https://github.com/git-guides/install-git
2. https://phoenixnap.com/kb/how-to-install-git-windows
)

git lfs install
git clone https://huggingface.co/datasets/beans

```
git lfs install
git clone https://huggingface.co/datasets/beans
```

```
C:\Windows\system32>cd..

C:\Windows>cd..

C:\>cd Users

C:\Users>cd axt210111

C:\Users\AXT210111>cd Downloads

C:\Users\AXT210111\Downloads>git lfs install
Updated Git hooks.
Git LFS initialized.

C:\Users\AXT210111\Downloads>git clone https://huggingface.co/datasets/beans
Cloning into 'beans'...
remote: Enumerating objects: 54, done.
remote: Total 54 (delta 0), reused 0 (delta 0), pack-reused 54
Unpacking objects: 100% (54/54), 10.78 KiB | 36.00 KiB/s, done.
Updating files: 100% (8/8), done.
Filtering content: 100% (6/6), 343.33 MiB | 22.20 MiB/s, done.
```

Navigate to the folder where you performed the git clone, and you'll have a .zip file there. Unzip that file, and you will then have a series of folders. Inside the data folder, you will have 3 more folders, where the images are located for Training, Testing, and Validation. (We don't need the Validation set of images for this project).

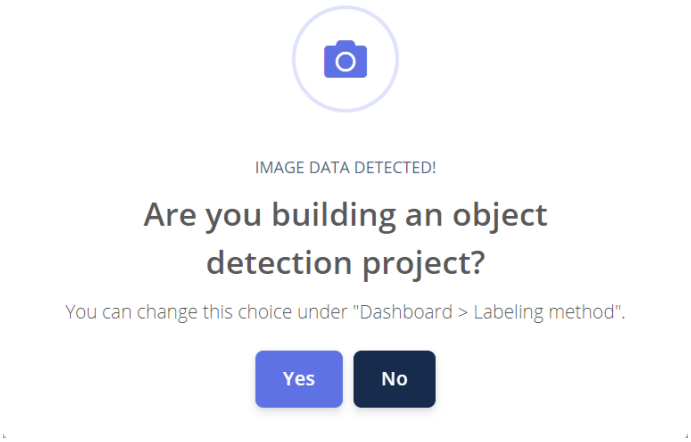


At this point we're ready to upload this data to Edge Impulse for our model

creation.

Uploading the Dataset to Edge Impulse

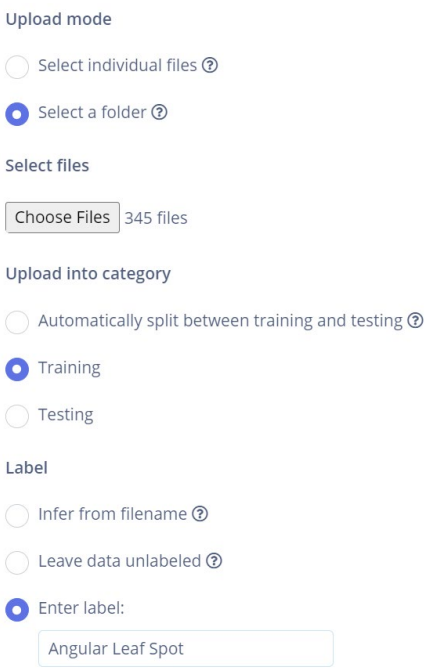
First make sure that you have selected one label per data item as labeling method from the Edge Impulse dashboard. If the below dialog box pops up when uploading data, select No.



Open Edge Impulse and navigate to the project that you want to do the Image classification. We are doing disease classification of the Plant Beans.

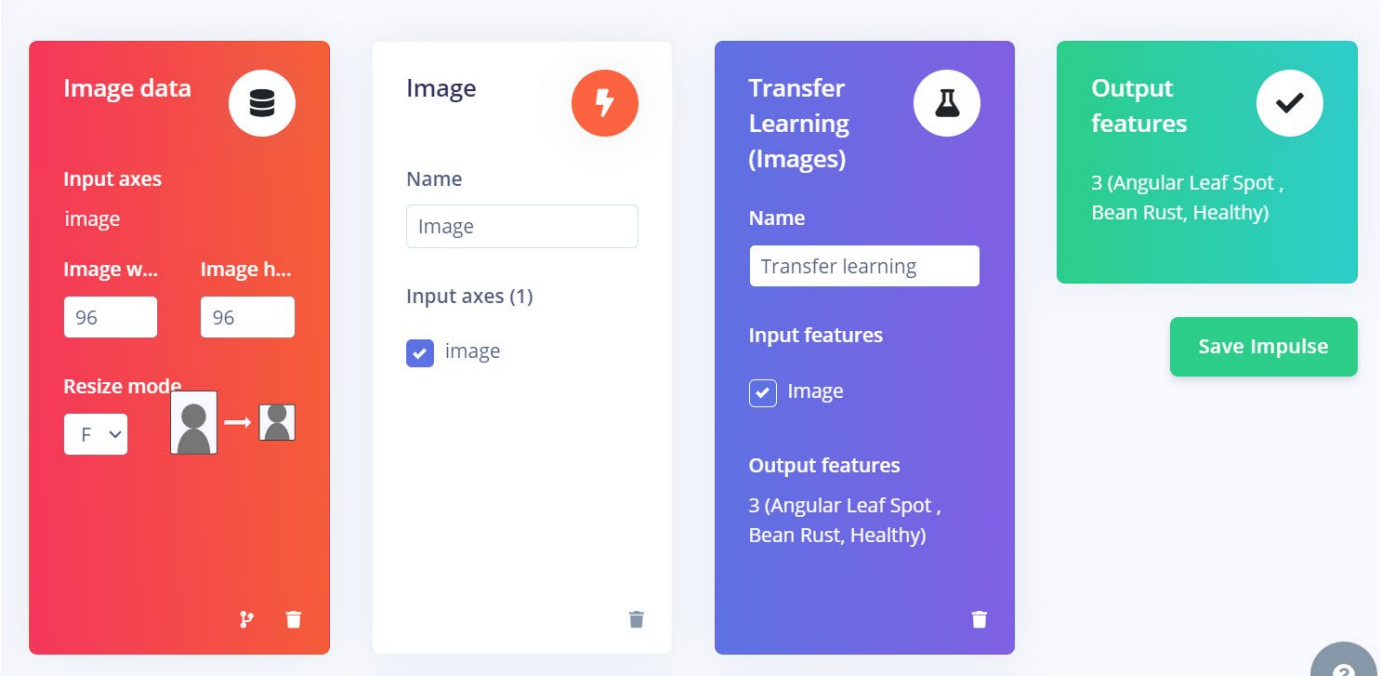
In Data acquisition tab, upload each dataset :

1. Training (Healthy, angular_leaf_spot, bean_rust)
2. Testing (Healthy, angular_leaf_spot, bean_rust)

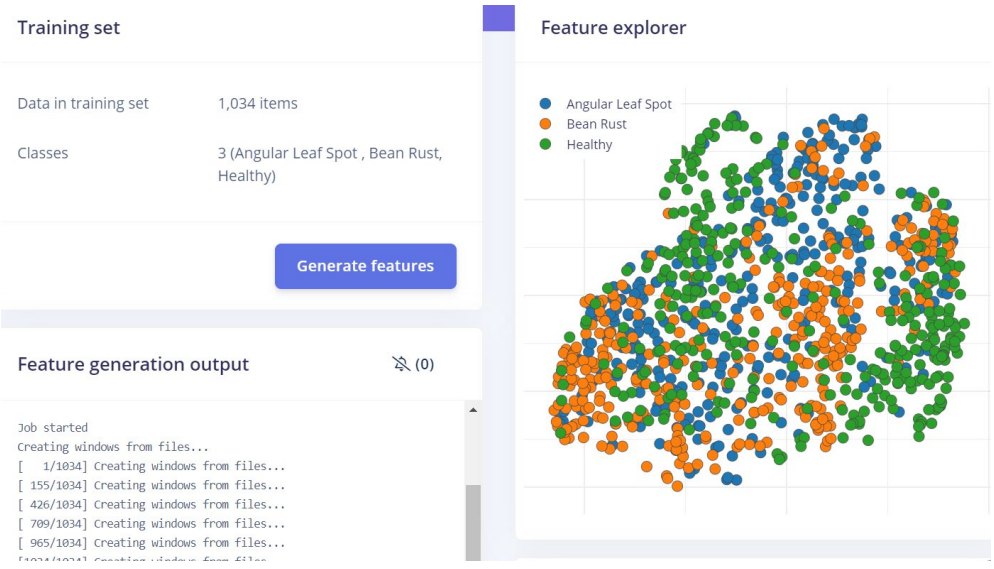


Impulse Design

Select each values as shown in the figure:

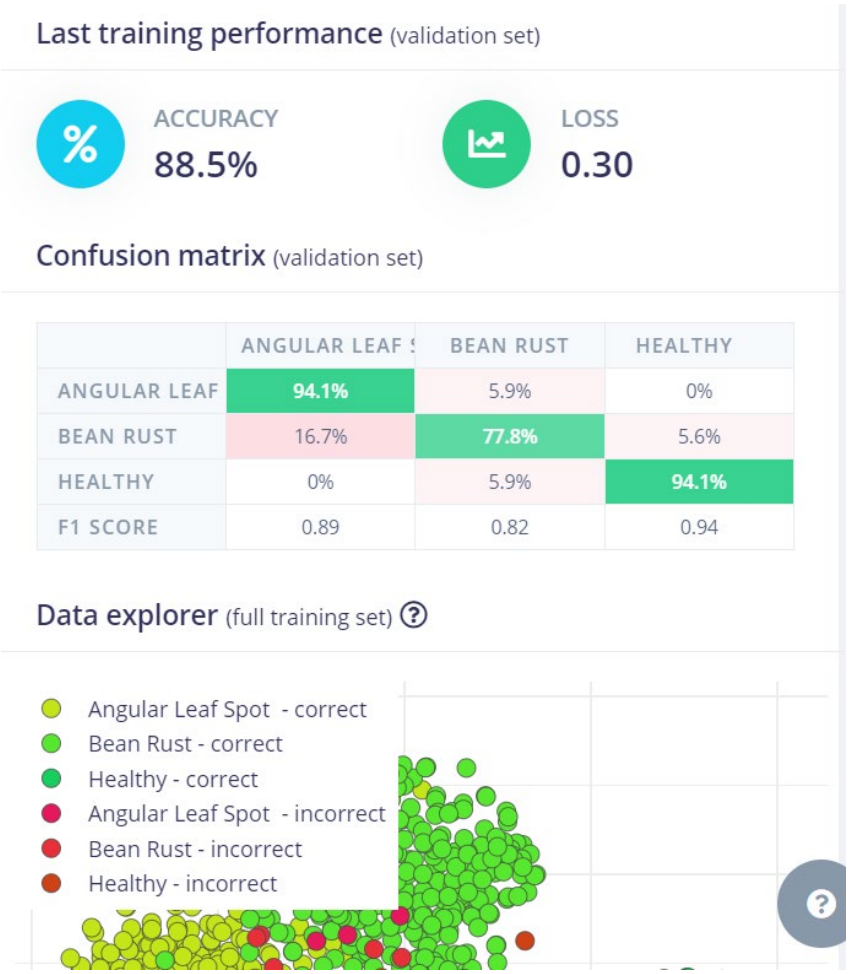


Next, Generate the features in the Image tab:



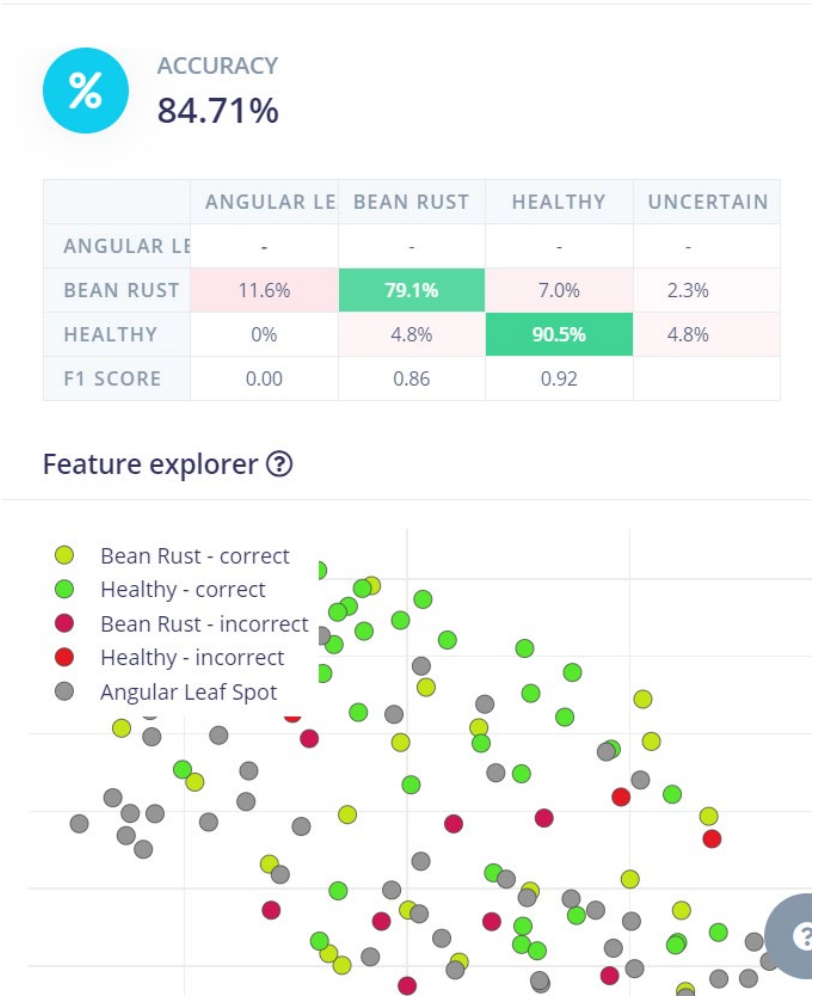
Creating and Testing the Model

With the data uploaded to the Edge Impulse Studio, we can start training our model.
Algorithm is MobileNetV2 96x96 0.35 (final layer: 16 neurons, 0.1 dropout) with a training cycle of 120 epochs, Validation set size of 10%, and a learning rate of 0.0005. (You can change these parameter values to find the higher accuracy for testing and training data.)



Testing the trained model:

Go to Model Testing and click on classify all to do testing.



Following this same methodology, you can make use of any Hugging Face dataset to help build and train a machine learning model with Edge Impulse.