



**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**

## 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.

The screenshot shows the Hugging Face Datasets page for the 'beans' dataset. The top section includes filters for Tasks (Image Classification), Sub-tasks (multi-class-image-classification), Languages (English), Multilinguality (monolingual), and Size Categories (1K<n<10K). Below these are filters for Language Creators (expert-generated), Annotations Creators (expert-generated), Source Datasets (original), Tags (Croissant), and License (mit).

The 'Dataset card' section shows the dataset name 'beans' with 22 likes. The 'Dataset Viewer' section shows a table of data rows. The table has columns for 'image\_file\_path', 'image', and 'labels'. The 'image\_file\_path' column shows the path to the dataset files. The 'image' column shows small thumbnail images of beans. The 'labels' column shows the class labels, which are 'angular\_leaf\_spot'.

The right side of the page shows download statistics: 'Downloads last month' is 11,591. There are buttons for 'Use in Datasets library' and 'Edit dataset card'. Below these are links for 'Homepage', 'Repository', and 'Paper'. The 'Homepage' is 'Beans Homepage', the 'Repository' is 'AI-Lab-Makerere/ibean', and the 'Paper' is 'N/A'. There are also links for 'Leaderboard' and 'Point of Contact', both of which are 'N/A'. At the bottom, there are statistics for 'Size of downloaded dataset files' (180 MB) and 'Size of the auto-converted Parquet files' (180 MB), along with the 'Number of rows' (1,295).

### 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).

beans > data				
Name	Date modified	Type	Size	
test	3/24/2024 12:42 PM	Compressed (zipped)...	17,294 KB	
test-00000-of-00001.parquet	3/24/2024 12:42 PM	PARQUET File	17,292 KB	
train	3/24/2024 12:42 PM	Compressed (zipped)...	140,442 KB	
train-00000-of-00001.parquet	3/24/2024 12:42 PM	PARQUET File	140,404 KB	
validation	3/24/2024 12:42 PM	Compressed (zipped)...	18,071 KB	
validation-00000-of-00001.parquet	3/24/2024 12:42 PM	PARQUET File	18,067 KB	

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.



IMAGE DATA DETECTED!

**Are you building an object  
detection project?**

You can change this choice under "Dashboard > Labeling method".

Yes

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)

Upload mode

☐ Select individual files ⓘ

☒ Select a folder ⓘ

Select files

345 files

Upload into category

☐ Automatically split between training and testing ⓘ

☒ Training

☐ Testing

Label

☐ Infer from filename ⓘ

☐ Leave data unlabeled ⓘ

☒ Enter label:

Angular Leaf Spot

## Impulse Design

Select each values as shown in the figure:

The screenshot shows the Impulse ML Studio interface with four tabs:

- Image data** (Red tab):
  - Input axes: image
  - Image w...: 96
  - Image h...: 96
  - Resize mode: F (Fixed)
- Image** (White tab):
  - Name: Image
  - Input axes (1): ☒ image
- Transfer Learning (Images)** (Purple tab):
  - Name: Transfer learning
  - Input features: ☒ Image
  - Output features: 3 (Angular Leaf Spot, Bean Rust, Healthy)
- Output features** (Green tab):
  - 3 (Angular Leaf Spot, Bean Rust, Healthy)

A green **Save Impulse** button is located below the Output features tab.

Next, Generate the features in the Image tab:

The screenshot shows the Training set and Feature explorer sections:

- Training set**:
  - Data in training set: 1,034 items
  - Classes: 3 (Angular Leaf Spot, Bean Rust, Healthy)
  - Generate features** button
- Feature generation output**:
  - Job started
  - Creating windows from files...
  - [ 1/1034] Creating windows from files...
  - [ 155/1034] Creating windows from files...
  - [ 426/1034] Creating windows from files...
  - [ 709/1034] Creating windows from files...
  - [ 965/1034] Creating windows from files...
- Feature explorer**:
  - Scatter plot showing data points for Angular Leaf Spot (blue), Bean Rust (orange), and Healthy (green).

## 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.)

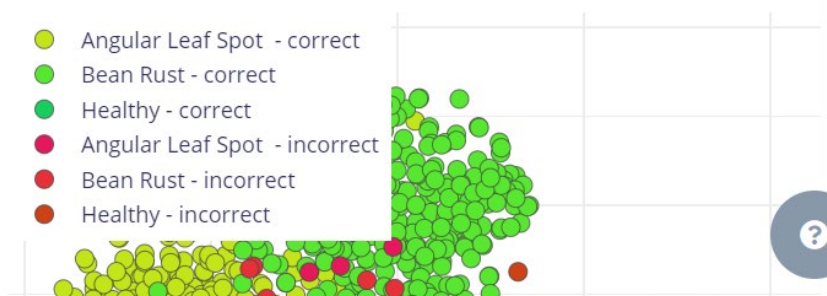
### Last training performance (validation set)



### Confusion matrix (validation set)

	ANGULAR LEAF SPOT	BEAN RUST	HEALTHY
ANGULAR LEAF SPOT	94.1%	5.9%	0%
BEAN RUST	16.7%	77.8%	5.6%
HEALTHY	0%	5.9%	94.1%
F1 SCORE	0.89	0.82	0.94

### Data explorer (full training set) ?



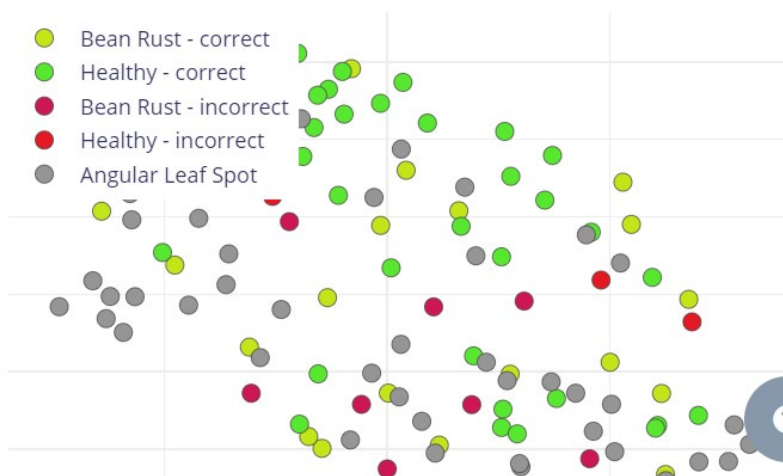
## Testing the trained model:

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

 **ACCURACY**  
**84.71%**

	ANGULAR LE	BEAN RUST	HEALTHY	UNCERTAIN
ANGULAR LE	-	-	-	-
BEAN RUST	11.6%	79.1%	7.0%	2.3%
HEALTHY	0%	4.8%	90.5%	4.8%
F1 SCORE	0.00	0.86	0.92	

### Feature explorer



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.