

Computer Vision Assignment 03

xAI

Melchor Lafuente
Iker Jauregui

Contents

1	Introduction	2
2	LIME	2
3	Best Model's Performance	3
4	Best Model's Interpretation	4
4.1	Best Classified Samples	4
4.1.1	Cassava Bacterial Blight (CBB)	4
4.1.2	Cassava Brown Streak Disease (CBSD)	4
4.1.3	Cassava Green Mottle (CGM)	5
4.1.4	Cassava Mosaic Disease (CMD)	5
4.1.5	Healthy	6
4.2	Worst Classified Samples	6
4.2.1	Cassava Bacterial Blight (CBB)	6
4.2.2	Cassava Brown Streak Disease (CBSD)	7
4.2.3	Cassava Green Mottle (CGM)	7
4.2.4	Cassava Mosaic Disease (CMD)	8
4.2.5	Healthy	8
5	Conclusions	9
6	Future Work	9

1 Introduction

The aim of this project is to explain and interpret the best cassava disease classification model that we have trained so far. To do so, we use the LIME technique.

2 LIME

LIME stands for Local Interpretable Model-agnostic Explanations, and it's basically a clever way to peek inside black-box models and understand why they make specific predictions. The key idea is pretty straightforward: instead of trying to explain the entire model (which can be impossibly complex), LIME focuses on explaining individual predictions locally.

Here's how it works. Say we want to understand why our model classified a particular cassava leaf image as having a specific disease. LIME takes that image and creates a bunch of similar versions by randomly "turning off" parts of it (using superpixels). It then feeds all these perturbed versions to our model and sees how the predictions change. With this dataset of variations and their predictions, LIME fits a simple, interpretable model (usually a linear one) that approximates the complex model's behavior just around that specific image.

The beauty of LIME is that it's model-agnostic, so it doesn't care if we're using a random forest, a neural network, or, like in this case, a ResNet18. It just treats the model as a black box that spits out predictions. The output is a small set of features (in our case, image regions) that contributed most to the prediction, either positively or negatively. So we might see, for example, that the model predicted "Cassava Brown Streak Disease" because of specific discolored patches on the leaf, which makes intuitive sense and helps us trust (or distrust) the model's decision.

3 Best Model's Performance

The best model we trained achieved a F1 score of **0.7050** (Figure 1) with standard inference and a F1 score of **0.7261** (Figure 2) with TTA (test-time augmentation).

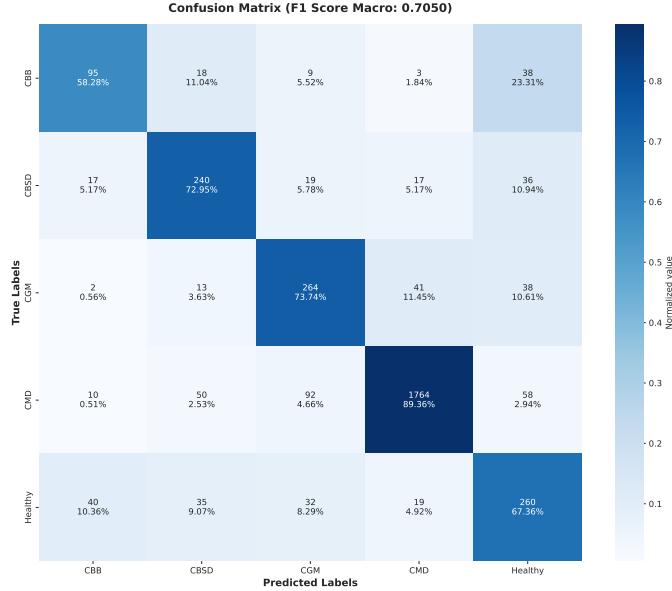


Figure 1: Confusion matrix over Test set for the best model with standard inference.

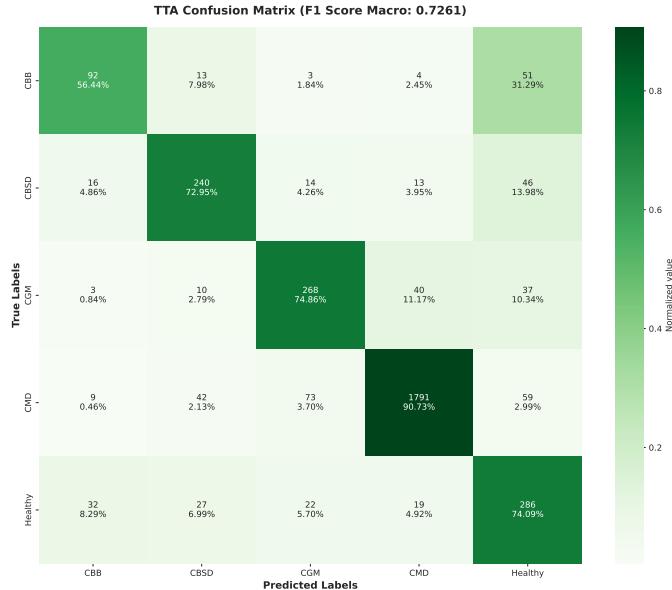


Figure 2: Confusion matrix over Test set for the best model with TTA.

4 Best Model's Interpretation

In this section, we will try to understand the behavior of the best model trained so far. To do so, for each class, we will select the best and worst classified samples and we will apply LIME technique to see on which part of the images is the model focusing.

4.1 Best Classified Samples

4.1.1 Cassava Bacterial Blight (CBB)

It seems that the model is focusing on the brown wilted areas of the leaf (Figure 3). This kind of defect is known to be characteristic of the CBB disease.



Figure 3: Best classified sample from CBB disease.

4.1.2 Cassava Brown Streak Disease (CBSD)

The image (Figure 4) seems to show something like the root of the plant and the CBSD disease attacks this part of the plant, so we can say that, at least, the model has related roots with this disease.

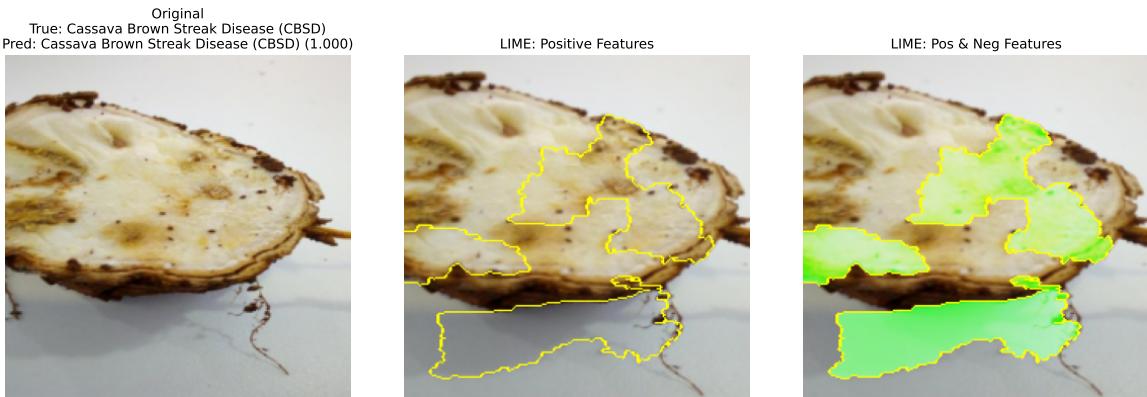


Figure 4: Best classified sample from CBSD disease.

4.1.3 Cassava Green Mottle (CGM)

This disease is characterized by a light spotted pattern present on the leaves. As we can see on LIME's segmentation (Figure 5), the model seems to be focusing on the right areas.

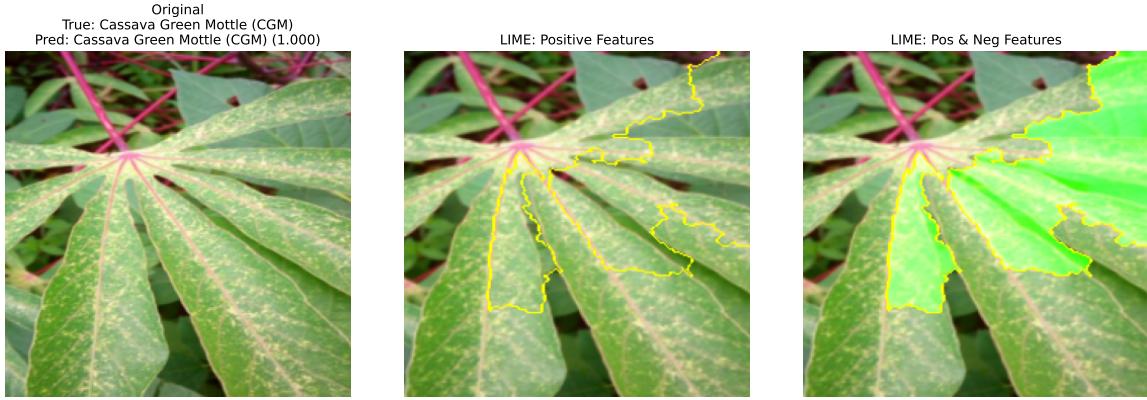


Figure 5: Best classified sample from CGM disease.

4.1.4 Cassava Mosaic Disease (CMD)

The CMD disease's main characteristic is a yellow mosaic pattern on the leaves. Also, it produces a deformation of the leaves. If we look to the segmentation of LIME (Figure 6), we can see how the model is focusing on both features.

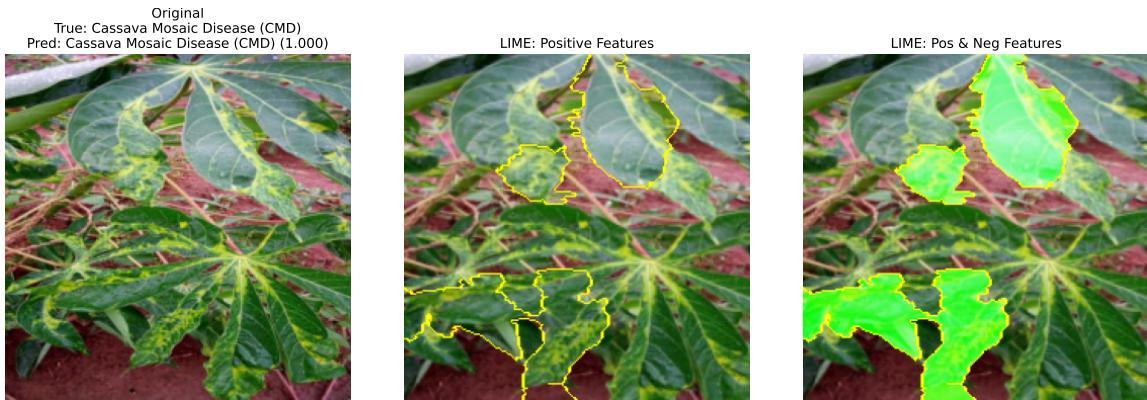


Figure 6: Best classified sample from CMD disease.

4.1.5 Healthy

Healthy plants don't show any deformation or color patterns. We can see how the model is focusing mostly on the leaf in order to predict this class (Figure 7).

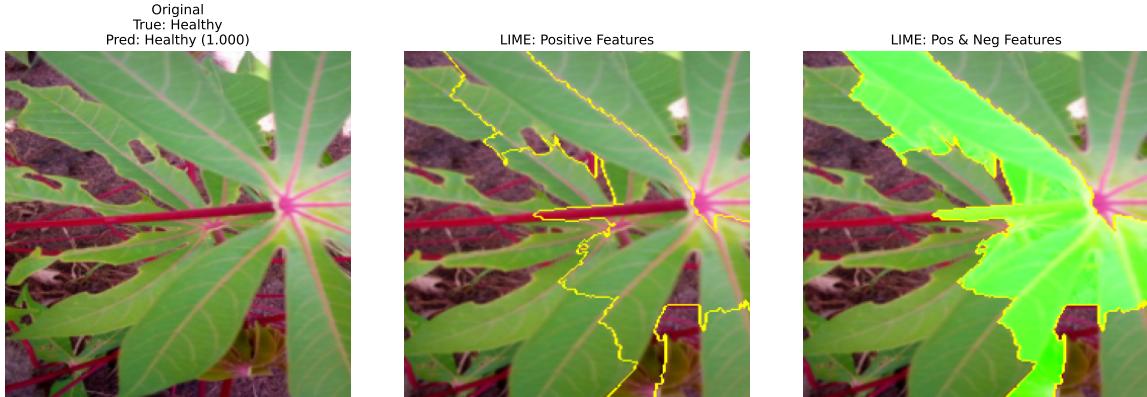


Figure 7: Best classified sample from Healthy class.

4.2 Worst Classified Samples

4.2.1 Cassava Bacterial Blight (CBB)

If we look closely to both the image and the LIME segmentation (Figure 8), we can see why the model misclassified this sample. The model is focusing on the yellow stripes of the leaves instead of the brown wilted areas. That could be the reason of why the model predicted CMD instead of CBB.

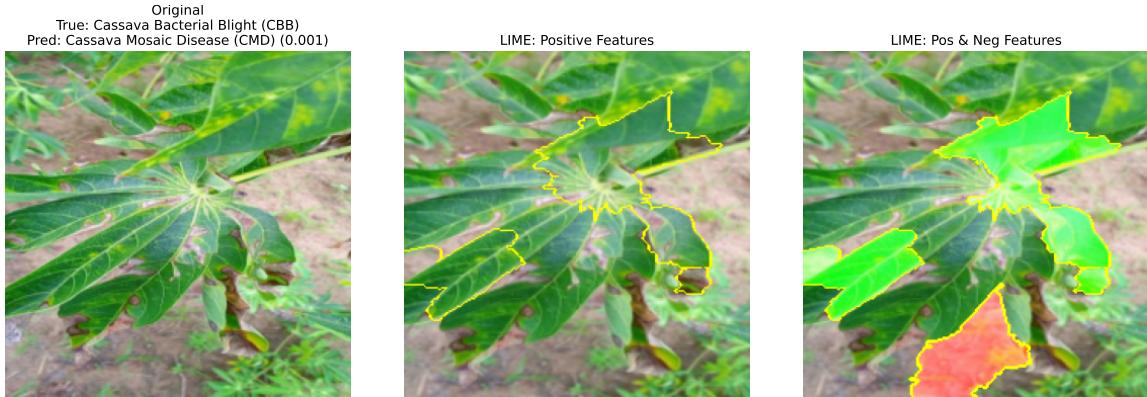


Figure 8: Worst classified sample from CBB disease.

4.2.2 Cassava Brown Streak Disease (CBSD)

Looking to the sample (Figure 9), we can't identify any disease symptoms over the plant. As it is estimated that the 8% of the data is mislabeled, we could think that our model's guess was right when predicting this image as Healthy.

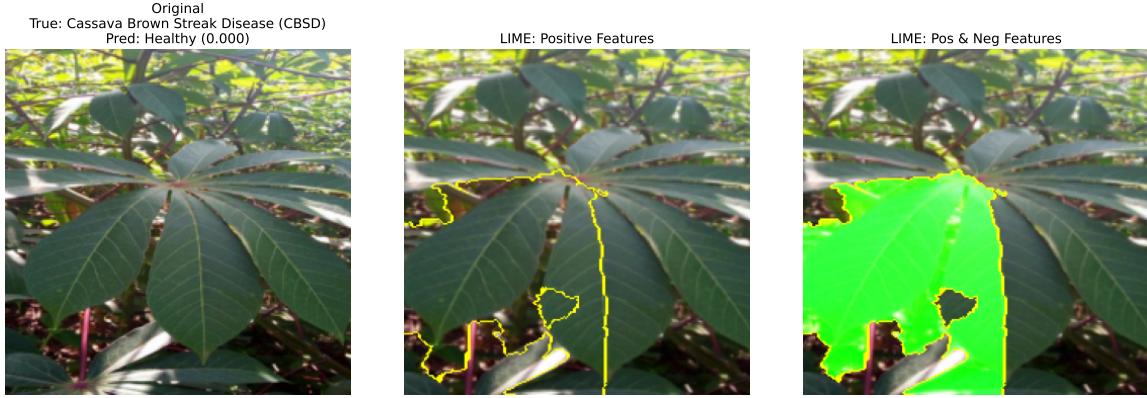


Figure 9: Worst classified sample from CBSD disease.

4.2.3 Cassava Green Mottle (CGM)

As we can see in the image (Figure 10), the upper leaves don't show any symptoms and the light spotted pattern of the lower leaves are hardly visible. So, we can consider that the prediction made by our model (Healthy) isn't that bad.

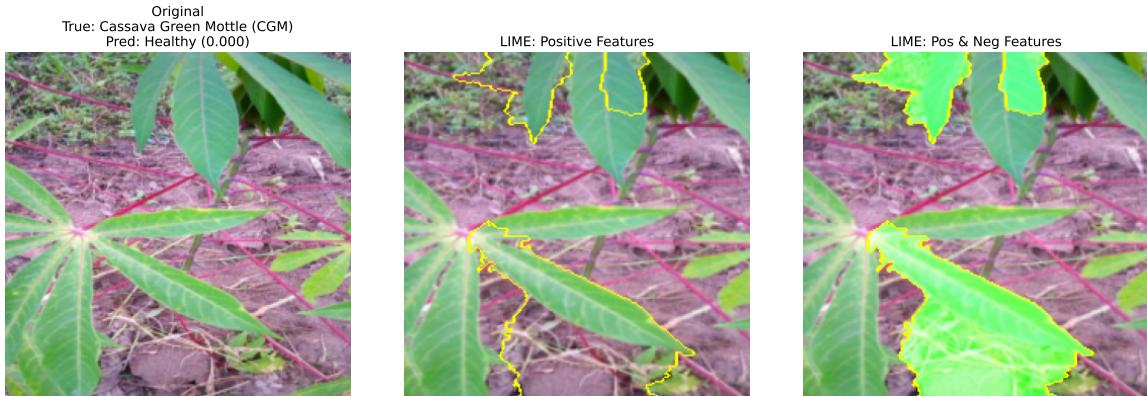


Figure 10: Worst classified sample from CGM disease.

4.2.4 Cassava Mosaic Disease (CMD)

Apart from the necrotic factor of the roots, the CBSD also appears as a yellow coloration (called *chlorosis*) of the leaves. If we look at the misclassified sample (Figure 11), we can see how it meets with the previous description. So, maybe our model didn't fail at predicting this sample as CBSD instead of CMD.

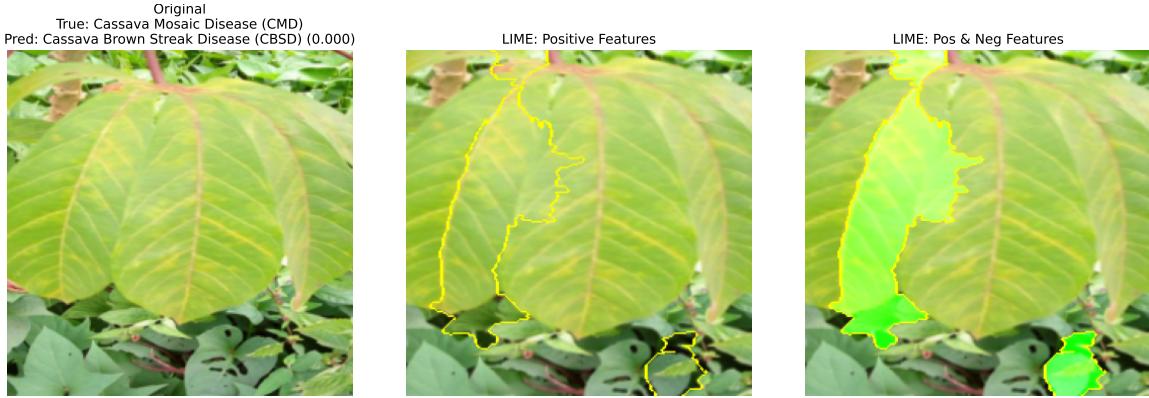


Figure 11: Worst classified sample from CMD disease.

4.2.5 Healthy

Once again, we can see how the sample is clearly mislabeled, as there are obvious yellow mosaic patterns on the leaves (Figure 12). So, it makes perfect sense for our model to predict CMD instead of Healthy.



Figure 12: Worst classified sample from Healthy class.

5 Conclusions

The LIME interpretability analysis revealed that our model is learning meaningful disease-specific features. For best classified samples, the model correctly focused on characteristic symptoms: brown wilted areas for CBB, root structures for CBSD, light spotted patterns for CGM, yellow mosaics for CMD, and uniform green leaves for Healthy plants.

Moreover, the analysis of worst classified samples uncovered potential mislabeling issues in the dataset. Several misclassifications appeared to be reasonable predictions given the visual evidence, suggesting that the reported 8% mislabeling rate may be affecting model performance.

6 Future Work

Given the evidence of mislabeling, in order to improve the performance over the *Cassava Disease Classification* problem, a data cleaning process guided by model predictions and expert validation would be highly recommended.