Machine Learning Report

Daniel Zügner (6832915) Konstantin Kobs (6414943)
Sazan Hoti (6832974)

Exercise 3

Features

We engineered five different features that can be separated into two different groups of features:

1. **Color means**: We calculated the means for all three color channels of the images. This led to three features. We chose these features because we could see blue and purple color tones in images with a Chagas parasite.
2. **Gradient features**: We calculated the gradients for every pixel in the image. We then chose the maximum gradient as well as the standard deviation as features.

Model

We used — as desired — the Gaussian Discriminant Analysis and fitted two Multinomial Gaussian curves to each of the two classes. The means for the two classes were calculated and are as follows:

class	max_gradient	std_gradient	red	green	blue
Chagas parasite present	317.8683	53.0011	176.1726	139.5242	142.9174
Parasite not present	55.7774	10.4742	163.7130	126.8280	127.3415

It is easy to see that the means for the gradient features are really different which means that these features are a good indication finding Chagas parasites. We used a scatter plot to visualize these two features and found a really good separation, which can be seen in Figure 1.

With this model, we were able to predict the correct class for all examples, however these examples were used in training the model as well, so it is more likely that they are correctly classified.

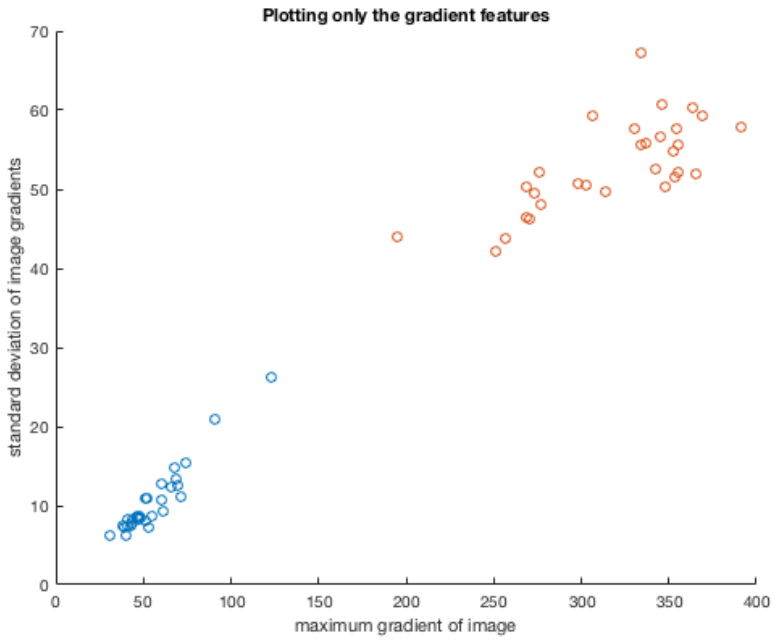

Figure 1: Two of the five features plotted. It can be seen that these features are very good choices for predicting the presence of the Chagas parasite.