Exercise 2 – April 30th
Accuracy criteria

A *supervision classification program has been run on a training set of 100,000 medical images in order to train an algorithm for identifying the images corresponding to a cancer diagnostic. The result of the classification is as follows*

Actual	Predicted: Negative	Predicted: Positive
Negative	94,000	500
Positive	3,500	2,000

1. What is the accuracy of the trained algorithm on the Training Set?

The accuracy is the proportion of proper predictions:

$$\boldsymbol{Accuracy = \frac{94,000 + 2,000}{100,000} = 96\%}$$

2. What is the recall of the trained algorithm on the Training Set?

Recall is the proportion of actual positive which are properly predicted:

$$\boldsymbol{Recall = \frac{2,000}{5,500} = 36\%}$$

3. What is the precision of the trained algorithm on the Training Set?

Precision is the proportion of predicted positive which are properly predicted:

$$\boldsymbol{Recall = \frac{2,000}{2,500} = 80\%}$$

4. What is the F1-Scoreof the trained algorithm on the Training Set?

F1-Score is the harmonic average of Precision and Recall:

$$\boldsymbol{F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 53\%}$$

5. Are you satisfied with the results? Why?

   The Recall value is not satisfactory it is too low: we do not want to miss 64% of the Positive cancer cases.