AI for Biotechnology
Exercise 3
Prof. Dr. Dominik Grimm

Bioinformatics Research Lab

TUM Campus Straubing for Biotechnology and Sustainability

Exercise E3.1

We have developed a novel diagnostic test in our newly founded startup. This test screens for genetic mutations to then predict if a certain patient suffers from a certain rare disease. The following measures are recorded by our scientific team in a clinical trail study:

	Disease	Healthy	Total
Prediction Disease	190	210	400
Prediction Healthy	10	3590	3600
Total	200	3800	4000

a) What is the accuracy, precision, recall, F1-score, specificity, false positive rate and the Matthew's Correlation Coefficient? (Compute the numbers of paper)

Accuracy	0.945
Precision	0.475
Recall	0.950
F1-Score	0.633
Specificity	0.945
FPR	0.055
MCC	0.65

b) The PPV (precision) is defined as $\frac{TP}{TP+FP}$. Think about how you would compute the Negative Predictive Value (NPV) and compute this value (on paper).

The Negative Predictive Value (NPV) is defined as $\frac{TN}{TN+FN}$. Here $NPV = 99.7\%$.

c) Interpret your results. Is the developed diagnosis test a good test? What are the strength and weaknesses?

Based on the metrics computed in (a) it is clear that the test is a poor diagnosis tool. It only has a precision of $\sim 47.5\%$, which means that if the test outcome of the test is positive we do not know if the patient really suffers from the disease or if the patient is health. However, the test might be good as a screening tool, because if the test is negative we can be highly certain that the patient will not suffer from the disease as indicated by the $NPV = 99.7\%$.

Exercise E3.2

Download the Jupyter Notebook `Exercise3.ipynb` and solve the exercises in this Jupyter Notebook.

Solution can be found in `Exercise3Solution.ipynb`