Title Pending

Using Artificial Intelligence to predict test result

By

Lukas Schießer

Submitted in partial fulfillment of the requirements for the degree of
Bachelor of Science in Cognitive Science
to the
Institute of Cognitive Science at the Osnabrück University
November 24, 2020

Thesis Supervisor:
Johannes
Thesis Supervisor:

Abstract

Acknowledgements

I want to thank...

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Introduction for this thesis by explaining motivation for this topic and giving a short overview.
Motivation: overwhelmed health care systems, shortages of testing supplies, AI as supplementary technique to help.

Chapter 2

Methods

2.1 Data

The data used to train the classifiers was provided by Brinati et al. [1]. It was collected between the end of February 2020 and mid of March 2020 from patients admitted to the *IRCSS Ospedale San Raffaele* and consists of 279 individuals who were selected randomly. For each individual, the data set provides their age, gender, results of a routine blood screening, and the result of a PCR test for Sars-CoV-2. Any other datapoint that could possibly identify the patient, i.e., date of admittance or date of PCR test were not recorded. Table 2.1 provides common satistics for the numerical features of the data set.

Feature	Unit	Mean	Std	Median
Age	Years	61.78	17.81	64
White Blood Cell Count (WBC)	10^9/L	8.55	4.86	7.10
Platelets	10^9/L	226.5	101.2	205.00
Neutrophils	10^9/L	6.20	4.17	5.10
Lymphocytes	10^9/L	1.19	0.80	1.00
Monocytes	10^9/L	0.61	0.41	0.50
Eosinophils	10^9/L	0.06	0.13	0.00
Basophils	10^9/L	0.01	0.04	0.00
C-reactive protein (CRP)	mg/L	90.89	94.42	54.20
Aspartate Aminotransferase (AST)	U/L	54.20	57.61	36.00
Alanine Aminotransferase (ALT)	U/L	44.92	45.50	31.00
Alkaline Phosphatase (ALP)	U/L	89.89	89.09	71.00
Gamma Glutamil Transferasi (GGT)	U/L	82.48	132.70	41.00
Lactate dehydrogenase (LDH)	U/L	380.45	193.98	328.00

Table 2.1: Descriptive statistics for numerical features in data set

2.2 Multivariate Imputation by Chained Equations

Describe MICE algorithm (keine Bewertung, findet in Discussion statt)
Maybe include PMM?

2.3 K-fold nested cross validation

2.4 Classifiers used

2.4.1 Random Forest

2.4.2 Logistic regression

Chapter 3

Results

3.1 Results of own implementation

3.2 Comparison with original paper

Chapter 4

Discussion

4.1 Discussion of Results

4.2 Scenarios for real-world validation

Declaration

I declare that..

Appendix A

Appendix

Bibliography

[1] D. Brinati et al. "Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study". In: *J Med Syst* 44.8 (2020), p. 135. ISSN: 1573-689X (Electronic) 0148-5598 (Linking). DOI: 10.1007/s10916-020-01597-4. URL: https://www.ncbi.nlm.nih.gov/pubmed/32607737,https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7326624/pdf/10916_2020_Article_1597.pdf,https://zenodo.org/record/3886927#.X3xhPO1CRPY.