TU WIEN Informatics

Predicting and Communicating Outcome of COVID-19 Hospitalizations with Medical Images and Clinical Data

Oliver Stritzel
Data Science

TU Wien Informatics
Institute of Visual Computing and Human-Centered Technology
Research Unit of Computer Graphics
Supervisor: Assistant Prof. Dr. Renata Raidou

Problem & Motivation

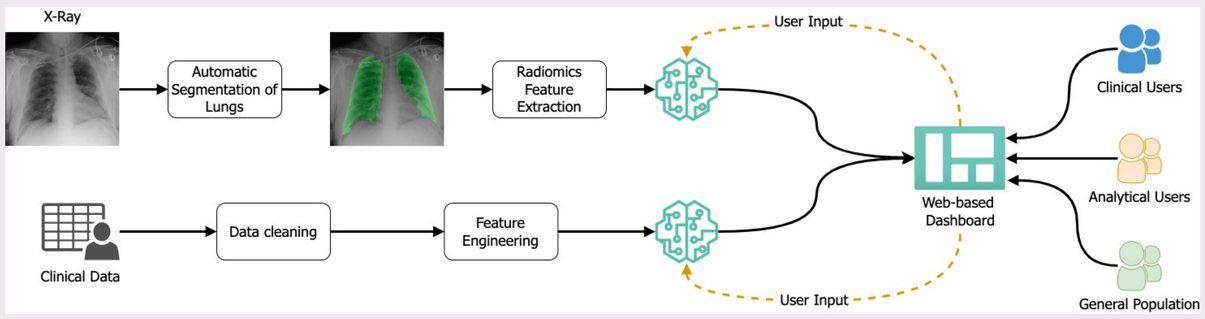
The COVID-19 pandemic put pressure on several public institutions, mainly the health-care system. Besides the danger of the actual illness, there also has been a rise of fake news and alternative facts. It is of increased importance to communicate novel scientific insights to numerous possible user groups, including laypeople as well as medical and analytical experts in an effort to **increase risk perception** and provide **decision making support**.

Goals and Research Questions

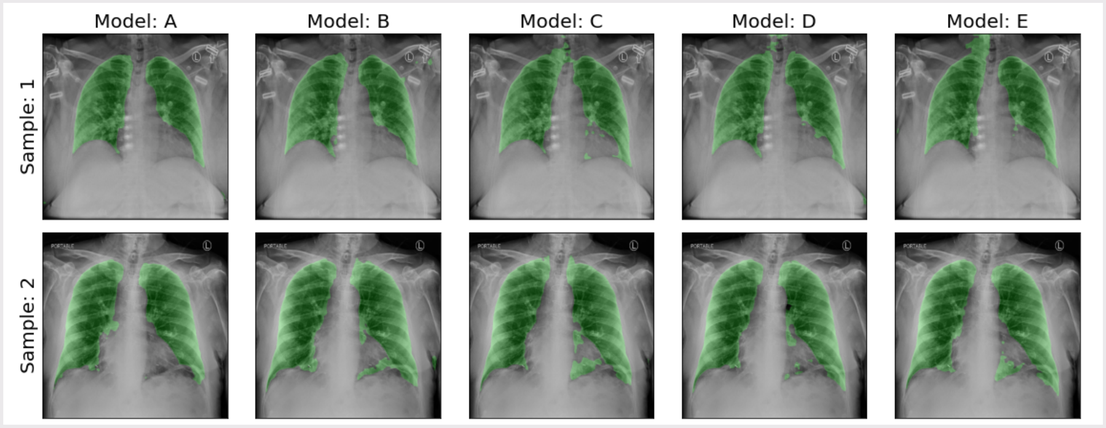
We propose a novel **visual analytics application** which communicates insights into COVID-19 hospitalization outcome prediction using the publicly available Stony Brooks COVID-19 dataset. The dataset includes electronic health data as well as medical image data in form of patients chest X-rays. We face several challenges that define our research questions:

1. **Automatic Lung Segmentation** using transfer learning to enable radiomics feature extraction.
2. **Prediction of COVID-19 hospitalized patients outcome**
3. Development of a **visual analytics** solution to support understanding into the prediction of COVID-19 outcomes for different user groups.

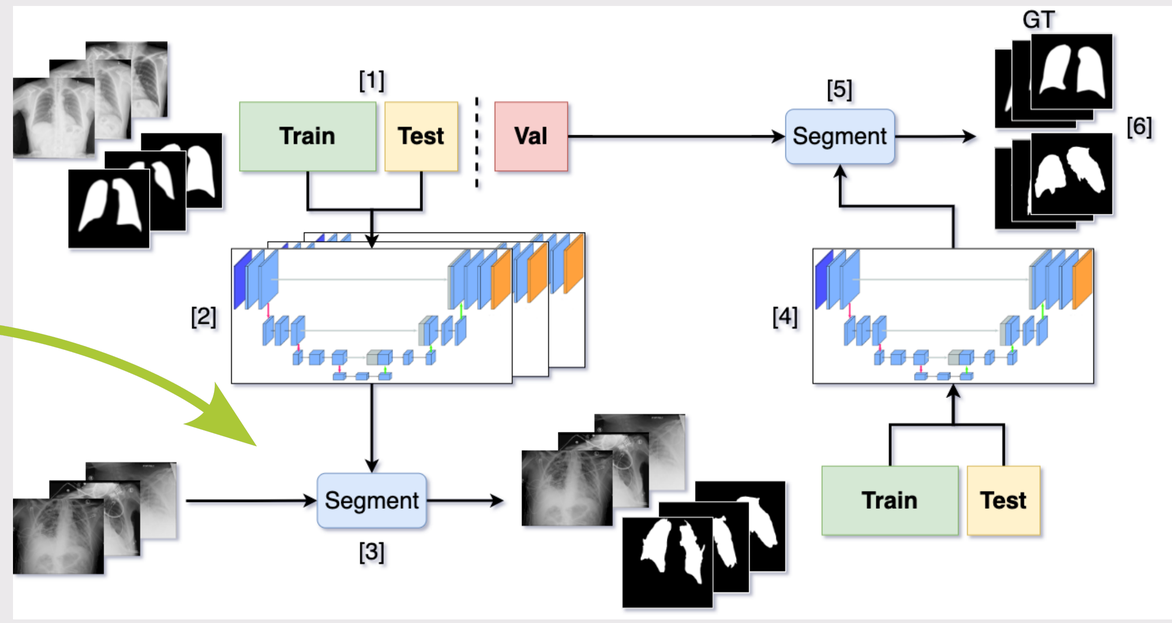
Approach



1. Automatic Lung Segmentation



Qualitative evaluation of segmentation results from multiple U-Net-like **deep learning models** using **transfer learning** and different **data augmentation** and image **pre-processing** techniques.



Quantitative evaluation of segmentation without available ground truth (GT) using a novel reverse transfer learning approach, so that popular segmentation metrics are applicable again.

2. Hospitalization Outcome Prediction & Data Analysis

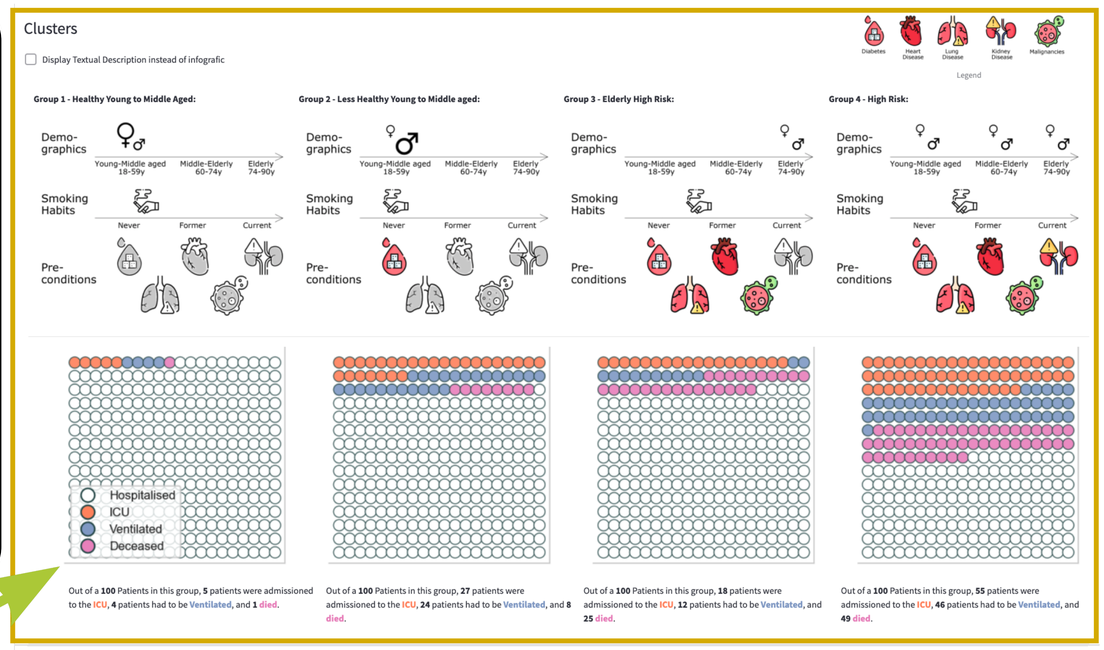
- Multi-target prediction including combinations of targets: patient ventilated, admissioned to ICU, deceased
- Data preprocessing: imputation, scaling, dimensionality reduction, data re-sampling for imbalanced data
- Clustering patients into different groups based on electronic health data: **Healthy Young to Middle** aged, **Elderly High Risk** (old patients with certain less serious pre-conditions), **Less Healthy Young to Middle aged** (certain early pre-conditions like diabetes, increased blood pressure), **High Risk** (cancer patients, diabetes, heart issues, chronic kidney disease)

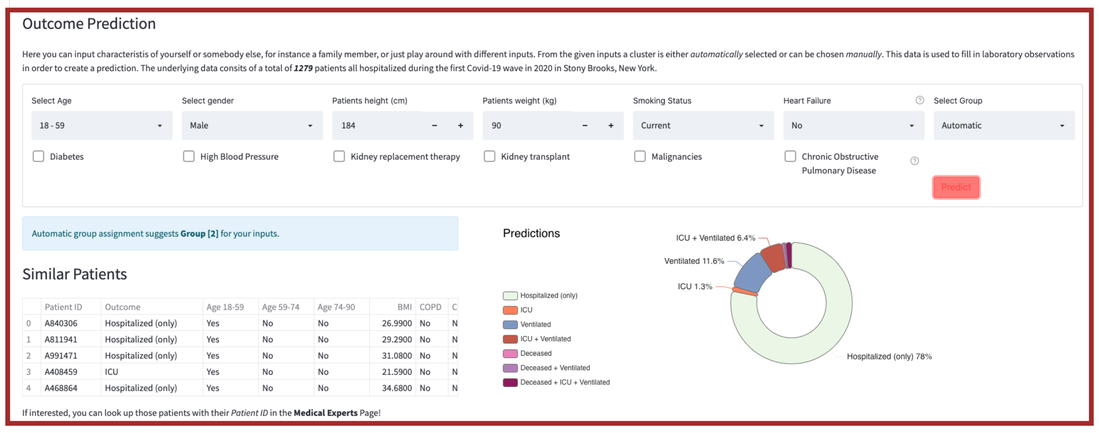
3. Visual Analytics

- **Requirement analysis** following multi-level typology of abstract visualization (Brehmer 2013)
- **Implementation** of interactive web-based dashboard
- **Evaluation** of design choices by conducting **case-studies** with laymen users and **usage scenarios** for remaining groups

Contributions

- Novel strategy to **evaluate segmentation** performance **without GT**
- Indications that merging image and electronic health data increases outcome prediction performance slightly (0.78 → 0.785 ROC)
- Visual Analytics dashboard, evaluating design choices with a small user group conducting case studies with lay users





	Patient ID	Outcome	Age 18-59	Age 59-74	Age 74-90	BMI	COPD	C
0	A840306	Hospitalized (only)	Yes	No	No	26.9900	No	N
1	A811941	Hospitalized (only)	Yes	No	No	29.2900	No	N
2	A991471	Hospitalized (only)	Yes	No	No	31.0800	No	N
3	A408459	ICU	Yes	No	No	21.5900	No	N
4	A468864	Hospitalized (only)	Yes	No	No	34.6800	No	N

Figure: Dashboad for laypeople users including overview of risk groups and outcome prediction enabled by interactive user input