

Master's Project 2020/21
Digital Health & Machine Learning
Prof. Dr. Christoph Lippert, *HPI*, Prof. Dr. Friedrich C. Luft, *Charité*,
Prof. Dr. Pao-Hwa Lin, *Duke University*

Digitizing and Analysing historic patient files using Deep Learning
Or: Why was Dr. Kempner's rice diet so effective against hypertension?

In this project you will travel back in history to 1950, when the German doctor Walter Kempner introduced a dietary treatment for severe (malignant) hypertension at Duke University, which at that time had an average survival time of <6 months. His diet, which is low in sodium(chloride), proteins, and simple carbohydrates, was dramatically successful and he received the highest American Heart Association award for his work. However, which component was primarily responsible remains unknown. Kempner published case reports, but studied 18,000 patients at Duke University between 1950 until 1985. His team conducted detailed examinations with repetitive *daily* observations over months that were written on "yellow sheets" (see below).



Anonymized example of an observation record

Kempner's long-term observational studies could shed light on whether Na^+/Cl^- solute, calories, protein, or other factors are responsible for blood-pressure lowering, weight loss, and cardiovascular risk reduction. A systematic statistical analysis of these data has never been performed. The aim of this project is to plan, develop and test a system to digitize the handwritten historic patient files by means of modern deep learning based optical character recognition (OCR). Initially, over 3.000 files from Duke University are available for data analysis,

model development and evaluation. Subsequently, a critical data-driven statistical analysis shall give insights as to why Kempner's rice diet has been so successful.



Besides acting as hypertension treatment, Kempner's strict diet led to positive effects on the overall health of his patients such as weight loss, improvements in diabetic retinopathy and increased lung capacity

A study including 18,000 persons on a defined, dietary regimen for months, will never again be performed. We believe that Kempner's data could contribute to the still unanswered question: "Why does the blood pressure go up?" Your contributions in this project are thus highly relevant to cardiovascular health.

What you will do

- Identify key attributes from handwritten paper sheets
- Identify and test existing OCR solutions
- Plan, design and implement a deep learning based OCR tool
- Implement UIs and Annotation tools to generate labeled training data
- Develop means to measure model uncertainty and include human feedback
- Identify and conduct statistical tests based on the extracted data
- Present your work to researchers and medical professionals

What you will learn

- The fascinating history of Dr. Kempner's Rice Diet
- How to plan and conduct a complex machine learning (ML) project on digital health data
- How to leverage user-centered design principles to design interactive ML pipelines
- How ML works in the field and how to overcome its challenges
- Development processes in a team with outside collaborators

What you should bring with you

- Software engineering skills
- Hands-on experience in ML technologies and an interest in automated image processing
- Project management and soft skills
- High motivation and commitment
- Enthusiasm for working alongside medical researchers and professionals