Predicting disease-related fatigue based on physiological signals: application to post-COVID-19 and cancer.

Neusa Martins, Nathanaël Perraudin

Master project - Fall 2023

Figure 1: Graphical abstract. The goal of this project is to investigate the feasibility of using physiological parameters monitored by wearable devices as objective indicators of disease-related fatigue.

1 Abstract

Fatigue is a multidimensional symptom associated with different physical and psychological conditions. It is one of the most common and debilitating symptoms of the post-COVID-19 condition and the most common complaint among cancer survivors. Disease-related fatigue remains underdiagnosed and undertreated in clinical settings as detection and monitoring rely on self-reporting. This project investigates using different physiological markers as objective predictors of disease-related fatigue. For that purpose, wearable devices are used during weekly monitoring phases (7 days, monitored during 24h) to monitor post-COVID-19 patients and patients after successful cancer treatment during their stay in a rehabilitation clinic. The dataset
recorded during weekly monitoring phases comprises time-series data, including self-reported fatigue (4 samples/day) and continuous physiological markers (256 Hz to 1 Hz, depending on the parameter). The project aims to establish a new monitoring tool for the continuous and objective assessment of fatigue status in patients.

2 Details of the project

The primary objective of this master thesis is to establish correlations between physiological markers and reported fatigue scores. The project consists of several stages, each building upon the previous one to gain deeper insights.

In the initial phase, the student will preprocess raw time-series data to create a relevant and organized dataset. This step involves extracting key features from the time series and resampling them appropriately to improve usability. Next, the student will explore the dataset and uncover simple correlations using statistical tools like Pandas. While statistics are helpful, they have limitations. Thus, in the third stage, the student will endeavor to predict the fatigue score using basic Machine Learning models such as Support Vector Machines (SVM) and Random Forests. The aim is to determine which input features have the most significant impact on the fatigue score. To achieve this, the learned model will be explained using attribution techniques (see [2, 1]). In the final phase, the student will advance to training a small recurrent network to predict the fatigue score. This model’s advantage lies in its ability to leverage past measures, allowing it to estimate the fatigue level by considering the patient’s history.

Following this systematic approach, the project aims to gain valuable insights into the relationship between physiological markers and reported fatigue scores. The combination of statistical analysis, machine learning, and the use of a recurrent network will contribute to a comprehensive understanding of the factors influencing fatigue levels in the patients of interest.

3 Student Tasks

Stage 1: formalizing the problem
- Assemble all the time series in a convenient dataset
- Handle the different sampling rates, by extracting meaningful features and resampling all time-series
- Build input-output pairs

Stage 2: Correlation analysis
- Simple statistical data exploration using Pandas
- Produce correlation plots for the most relevant features

Stage 3: Simple ML models
- Predict the current fatigue score based on historical data of features (e.g. considering the previous 6 hours of physiological features)
- Use interpretability/explainability techniques to determine the importance of each features.
Stage 4: Recurrent models

- Models the fatigue score as a time-series
- Implement a simple recurrent network

Stage 5: Reporting

- Write a report
- Final presentation

4 Additional information

What will you gain?  Handling of clinical data; Experiences with time series analysis and development of models at different complexities; Critical evaluation of modeling output and its medical relevance; Potential impact on an important clinical issue (quantification of fatigue condition).

Requirements:  Motivation to work with real medical data, machine learning fundamentals, linear algebra, Python, experience with Git,

Supervisor:  Neusa Martins (Neusa.Martins@empa.ch), Nathanaël Perraudin (nathanael.perraudin@sdc.ethz.ch)

Is a publication possible?  Yes, we intended to publish the thesis results.

References
