Topic: Predicting Disease Progression in Multiple Sclerosis using Machine Learning

Multiple sclerosis (MS) is a chronic disease of the central nervous systems (CNS) in which the immune system mistakenly attacks the myelin sheath around nerve fibers causing interruption of signals from CNS to different body parts. MS is considered the main cause of disability in young people. More than 700,000 people were diagnosed with MS in Europe at the middle of their working life (commonly between the age of 20 to 50 years) [1] leading to reduced quality of life in addition to economic and social costs [2]. Relapsing-remitting MS (RRMS) is the most common type of MS that affects around 80% of patients. Medications and treatments can help control the disease course or reduce number of relapses in order to prevent further worsening of disease symptoms. Among different treatments, fingolimod (Gilenya, Novartis Pharma AG, Basel, Switzerland) was reported to be effective and efficient in the treatment of people with RRMS [3]. However, better understanding of treatment effects is still needed to optimize and personalize treatment and intervention strategies.

For fingolimod in particular, Novartis has conducted a long-term non-interventional study called PANGAEA [4]. Over a 60-month period, clinical data of around 4,000 patients was collected from routine interventions at intervals of typically 3 months. Clinical data includes both continuous (e.g. vital signs and lab values) and categorical signals (e.g. assessments and questionnaires). Relevant output parameters include relapse density, disability status and brain lesions, collectively known as the NEDA3 score [5].

It is well known that not every patient reacts equally to the treatment. However, little is known about why some patients respond better than others. An early predictor for outcome would be essential for better targeting treatments. Firstly, the longitudinal effect of treatment on different patients can be investigated using mixed effect model that accounts for the correlation between repeated measurements within the same patient [6]. Mixed model estimates the longitudinal effect on both group level and patient level by allowing each participant to have his/her own random intercept that may differ from the mean intercept of the whole population. Moreover, it was reported to be a robust statistical model for analyzing longitudinal data taking into consideration the inter-individual differences [7]. In addition, machine learning has been shown to be effective for outcome prediction on other medical datasets, such as PhysioNet 2012 [8], MIMIC-III [9] and eICU [10]. Models derived from Recurrent Neural Networks (RNNs) have been developed to cope with non-regular sampling typically present in hospitals [11], some of which describe internal dynamics by ordinary differential equations [12]. In addition, some tabular models have been effective at handling missing values typically present in real-world clinical data and enabling interpretability using built-in attention mechanisms [13].

Therefore, this thesis aims to evaluate disease progression of RRMS patients under fingolimod in terms of number of relapses, disease disability, and MS lesions appearing on MRI scans (NEDA3 score) using the PANGAEA data set.

The proposed work consists of the following parts:

- Literature and patent research of relevant work resulting in a comprehensive list of existing work on machine learning for clinical studies, with a focus on longitudinal methods

- Preliminary data analysis of the disease progression of RRMS under fingolimod medication (e.g., by means of linear mixed-effects models), both at population and individual level.

- Implementation of at least one predictive deep learning model (e.g., time-aware RNNs) that predicts the NEDA3 score at future time points, typically 6, 9 and 12 months in advance

- Evaluation of predictive performance using accuracy, sensitivity and specificity

The thesis must contain a detailed description of all developed and used algorithms as well as a profound result evaluation and discussion. The implemented code has to be documented and provided. An extended research on literature, existing patents and related work in the corresponding areas has to be performed.

Advisors:	Alzhraa A. Ibrahim, M. Sc.; Thomas Altstidl, M. Sc.; Dr. Dario Zanca; and Prof. Dr. Bjoern Eskofier
Student:	Student Name
Start − End:	t.b.d.

References

[1] Gitto L.: *Living with Multiple Sclerosis in Europe: Pharmacological Treatments, Cost of Illness and Health Related Quality of Life Across Countries.* Exon Publications, 17-37, 2017.

[2] Höer A., et al.: *Multiple sclerosis in Germany: data analysis of administrative prevalence and healthcare delivery in the statutory health system.* BMC health services research, 14(1), 1-7, 2014.

[3] Ziemssen T., et al.: *Long-term real-world effectiveness and safety of fingolimod over 5 years in Germany.* Journal of neurology, 1-10, 2022.

[4] Ziemssen T., Kern R., and Cornelissen C.: *The PANGAEA study design - a prospective, multicenter, non-interventional, long-term study on fingolimod for the treatment of multiple sclerosis in daily practice.* BMC Neurology, 15(93), 2015.

[5] Lu, G., et al.: *The evolution of No Evidence of Disease Activity in multiple sclerosis.* Multiple sclerosis and related disorders 20, 231-238 (2018).

[6] Cnaan, A., Laird, N. M., Slasor, P. : *Mixed models: using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data.* Tutorials in Biostatistics: Statistical Modelling of Complex Medical Data, 2, 127-158, 2004.

[7] Van Dongen, Hans PA, et al. : *Mixed-model regression analysis and dealing with interindividual differences.* Methods in enzymology, 384, 139-171, 2004.

[8] Silva I., et al.: *Predicting In-Hospital Mortality of ICU Patients: The PhysioNet/Computing in Cardiology Challenge 2012.* Computing in Cardiology, 39, 245-248, 2012.

[9] Johnson A., Pollard T., Shen L., Lehman L., Feng M., Ghassemi M., Moody B., Szolovits P., Celi L., and Mark R.: *MIMIC-III, a freely accessible critical care database.* Scientific Data, 3(1), 1-9, 2016.

[10] Pollard T., Johnson A., Raffa J., Celi L., Mark R., and Badawi O.: *The eICU Collaborative Research Database, a freely available multi-center database for critical care research.* Scientific Data, 5(1), 1-13, 2018.

[11] Weerakody P., Wong K., Wang G., and Ela W.: *A review of irregular time series data handling with gated recurrent neural networks.* Neurocomputing, 441, 161-178, 2021.

[12] De Brouwer E., Simm J., Arany A., and Moreau Y.: *GRU-ODE-Bayes: Continuous Modeling of Sporadically-Observed Time Series.* Adv. in Neural Information Processing Systems, 32, 2019.

[13] Arik S., and Pfister T.: *Tabnet: Attentive interpretable tabular learning.* AAAI, 35, 6679-6687, 2021.