Project No.	Project Title	
2022-01-106	Temporal Modeling and Clustering of Deterioration Patterns in ALS	
Academic Advisor		**Co-Advisor**
Prof. Boaz Lerner		
Team Members		
Matan Gilboa		
gilboama@post.bgu.ac.il		

Abstract

Amyotrophic lateral sclerosis (ALS) is an incurable neurodegenerative disease. ALS incidence is approximately 1:50,000 patients. Most of the causes (~90%) of the disease and its mechanisms are unknown to the medicine. A person with ALS will live 2-5 years on average after their diagnosis. Since there is no real treatment to ALS, earlier diagnosis can allow patients to take a proactive approach and to adjust their lifestyle to cope with the changes expected, use the extra time for problem-solving, and to avoid the stress of being stuck without options, struggling to get by when they have suddenly lost an ability.

Most of the health databases as those of ALS are characterized by time-series data (TSD). One of the main challenges when dealing with TSD is how to represent it. On the one hand, we would like to use it with all timestamps and features in order to perceive the underlying mechanism within the data. On the other hand, classic ML models (e.g., random forest) cannot exploit the richness within the temporal data and some aggregation actions are required. Another major challenge with TSD is how to measure similarity between sequences (say of clinic visits of patients) since they can have different length and interval between measurements. Therefore, a similarity measurement method should take these issues into consideration, which is what the dynamic time warping (DTW) algorithm does.

In this project, we applied the methodology of Halbersberg and Lerner (2019) to deal with both the TSD and heterogeneity of the ALS disease challenges. It consists of two main phases: The first phase is clustering similar patients by their deterioration similarity, aiming to find similar deterioration-based groups within patients. The second phase is deterioration pattern mining. It discovers patterns that can be hidden in large sequential databases and that are interpretable by humans, and therefore useful for understanding the data.

We applied the methodology to a database obtained from Meuhedet HMO. The database includes data of 162 ALS patients and 648 control subjects matched by their gender, age, and area of residence. Several pre-selected laboratory tests were used, and three methods were tested: a basic approach, the pattern mining approach (second phase), and a combined approach containing both clustering and pattern mining (both phases). The learning task was the detection of ALS, and the evaluation was made using the F1 measure. The approach that yielded the highest F1value was the combined approach.

Keywords: ALS, machine learning, clustering, pattern mining, disease deterioration