iDPP@CLEF 2023: The Intelligent Disease Progression Prediction Challenge

Helena Aidos[1], Roberto Bergamaschi[2], Paola Cavalla[3], Adriano Chiò[4], Arianna Dagliati[2], Barbara Di Camillo[5], Mamede Alves de Carvalho[1], Nicola Ferro[5], Piero Fariselli[4], Jose Manuel García Dominguez[6], Sara C. Madeira[1], and Eleonora Tavazzi[7]

[1] University of Lisbon, Portugal
{haidos, sacmadeira}@fc.ul.pt, mamedemg@mail.telepac.pt
[2] University of Pavia, Italy
roberto.bergamaschi@mondino.it, arianna.dagliati@unipv.it
[3] "Città della Salute e della Scienza", Turin, Italy
paola.cavalla@unito.it
[4] University of Turin, Italy
{adriano.chio, piero.fariselli}@unito.it
[5] University of Padua, Italy
{barbara.dicamillo, nicola.ferro}@unipd.it
[6] Gregorio Marañon Hospital in Madrid, Spain
jgarciadominguez@salud.madrid.org
[7] IRCCS Foundation C. Mondino in Pavia, Italy
eleonoratavazzi@gmail.com

Abstract. *Amyotrophic Lateral Sclerosis (ALS)* and *Multiple Sclerosis (MS)* are chronic diseases characterized by progressive or alternate impairment of neurological functions (motor, sensory, visual, cognitive). Patients have to manage alternated periods in hospital with care at home, experiencing a constant uncertainty regarding the timing of the disease acute phases and facing a considerable psychological and economic burden that also involves their caregivers. Clinicians, on the other hand, need tools able to support them in all the phases of the patient treatment, suggest personalized therapeutic decisions, indicate urgently needed interventions.

The goal of iDPP@CLEF is to design and develop an evaluation infrastructure for AI algorithms able to:

1. better describe disease mechanisms;
2. stratify patients according to their phenotype assessed all over the disease evolution;
3. predict disease progression in a probabilistic, time dependent fashion.

iDPP@CLEF run as a pilot lab in CLEF 2022, offering tasks on the prediction of ALS progression and a position paper task on explainability of AI algorithms for prediction; 5 groups submitted a total of 120 runs and 2 groups submitted position papers.

iDPP@CLEF will continue in CLEF 2023, focusing on the prediction of MS progression and exploring whether pollution and environmental data can improve the prediction of ALS progression.

1 Introduction

Amyotrophic Lateral Sclerosis (ALS) and *Multiple Sclerosis (MS)* are severe chronic diseases characterized by a progressive but variable impairment of neurological functions, characterized by high heterogeneity both in presentation features and rate of disease progression. As a consequence patients' needs are different, challenging both caregivers and clinicians. Indeed, the time of relevant events is variable, which is associated with uncertainty regarding the opportunity of critical interventions, like non-invasive ventilation and gastrostomy in the case of ALS, with implications on the quality of life of patients and their caregivers. For this reason, clinicians need tools able to support their decision in all phases of disease progression and underscore personalized therapeutic decisions. Indeed, this heterogeneity is partly responsible for the lack of effective prognostic tools in medical practice, as well as for the current absence of a therapy able to effectively slow down or reverse the disease course. On the one hand, patients need support for facing the psychological and economic burdens deriving from the uncertainty of how the disease will progress; on the other, clinicians require tools that may assist them throughout the patient's care, recommending tailored therapeutic decisions and providing alerts for urgently needed actions.
We need to design and develop *Artificial Intelligence (AI)* algorithms to:

- stratify patients according to their phenotype all over the disease evolution;
- predict the progression of the disease in a probabilistic, time dependent way;
- better describe disease mechanisms.

The *Intelligent Disease Progression Prediction at CLEF (iDPP@CLEF)* lab[8] aims to deliver an evaluation infrastructure for driving the development of such AI algorithms. Indeed, in this context, it is fundamental, even if not so common yet, to develop shared approaches, promote the use of common benchmarks, foster the comparability and replicability of the experiments. Differently from previous challenges in the field, iDPP@CLEF addresses in a systematic way some issues related to the application of AI in clinical practice in ALS and MS. In addition to defining risk scores based on the probability of occurrence of an event in the short or long term period, iDPP@CLEF also addresses the issue of providing information in a more structured and understandable way to clinicians.

The paper is organized as follows: Section 2 discusses related works; Section 3 presents what has been done in iDPP@CLEF 2022 while Section 4 introduces the plans for iDPP@CLEF 2023; finally, Section 5 draws some conclusions.

2 Related Works

Within CLEF, there have been no other labs on this or similar topics before.

Outside CLEF, there have been a recent challenge on Kaggle[9] in 2021 and some older ones, the DREAM 7 ALS Prediction challenge[10] in 2012 and the

[8] https://brainteaser.health/open-evaluation-challenges/
[9] https://www.kaggle.com/alsgroup/end-als
[10] https://dreamchallenges.org/dream-7-phil-bowen-als-prediction-prize4life/

DREAM ALS Stratification challenge[11] in 2015. The Kaggle challenge used a mix of clinical and genomic data to seek for insights about the mechanisms of ALS and difference between people with ALS who progress faster versus those who develop it more slowly. The DREAM 7 ALS Prediction challenge [7] asked to use 3 months of ALS clinical trial information (months 0–3) to predict the future progression of the disease (months 3–12), expressed as the slope of change in *ALS Functional Rating Scale Revisited (ALSFRS-R)* [3], a functional scale that ranges between 0 and 40. The DREAM ALS Stratification challenge asked participants to stratify ALS patients into meaningful subgroups, to enable better understanding of patient profiles and application of personalized ALS treatments. Differently from these previous challenges, iDPP@CLEF focuses on explainable AI and on temporal progression of the disease.

Finally, when it comes to MS, studies are mostly conducted on closed and proprietary datasets and iDPP@CLEF represents one of the first attempts to create a public and shared dataset.

3 iDPP@CLEF 2022

iDPP@CLEF run as a pilot lab for the first time in CLEF 2022[12] [5, 6] and focused on pilot activities aimed both at an initial exploration of ALS progression prediction and at understanding of the challenges and limitations to refine and tune the labs itself for future iterations.

Tasks iDPP@CLEF 2022 consisted of the following tasks:

- **Pilot Task 1 - Ranking Risk of Impairment**: it focused on ranking of patients based on the risk of impairment in specific domains. More in detail, we used the ALSFRS-R scale to monitor speech, swallowing, handwriting, dressing/hygiene, walking and respiratory ability in time and asked participants to rank patients based on time to event risk of experiencing impairment in each specific domain.
- **Pilot Task 2 - Predicting Time of Impairment**: it refined Task 1 by asking participants to predict when specific impairments will occur (i.e. in the correct time-window). In this regard, we assessed model calibration in terms of the ability of the proposed algorithms to estimate a probability of an event close to the true probability within a specified time-window.
- **Position Paper Task 3 - Explainability of AI algorithms**: we evaluated proposals of different (visualization) frameworks able to explain the multivariate nature of the data and the model predictions.

Participation 43 participants registered for iDPP@CLEF 2022 and 5 participants successfully submitted a total of 120 runs for Task 1 and Task 2; moreover,

[11] https://dx.doi.org/10.7303/syn2873386.
[12] https://brainteaser.health/open-evaluation-challenges/idpp-2022/

Table 1: Break-down of the runs submitted by participants for each task and sub-task. Participation in Task 3 does not involve submission of runs and it is marked just with a tick.

Team Name	Total	Task 1	Task 2	Task 3	Paper
BioHIT	18	9	9	–	–
CompBioMed	40	22	18	–	Pancotti et al. [11]
FCOOL	15	–	15	✓	Branco et al. [1] and Nunes et al. [10]
LIG GETALP	23	12	11	–	Mannion et al. [9]
SBB	24	12	12	–	Trescato et al. [12]
UNIPV	–	–	–	✓	Buonocore et al. [2]
Total	**120**	**55**	**65**		

2 position papers were submitted for the explainability task, as detailed in Table 1. Submission of participants are openly available in git repositories[13].

Datasets iDPP@CLEF 2022 created 3 datasets, for the prediction of specific events related to ALS, consisting of fully anonymized data from 2,250 real patients from medical institutions in Turin, Italy, and Lisbon, Portugal. The datasets contain both static data about patients, e.g. age, onset date, gender, ... and event data, i.e. 18,512 ALSFRS-R questionnaires and 4,015 spyrometries.

Table 2 summarizes the main features and variables available in the dataset. The following data are available for both the training and the test sets:

- the first available ALSFRS-R questionnaire at `Time 0` (both single question scores and total score).
  Thus, for example, time-of-onset and time-of-diagnosis are expressed as relative delta with respect to `Time 0` in months (also fractions);
- the slope of the ALSFRS-R score between time-of-onset and `Time 0` as:

$$slope = \frac{48 - \text{ALSFRS-R-score}\,(\texttt{Time 0})}{\texttt{Time 0} - \texttt{TimeOnset}}$$

- all the other static data, with a complete list available at http://brainteaser.dei.unipd.it/challenges/idpp2022/assets/other/static-vars.txt
- visits , containing either other ALSFRS-R questionnaires or Spirometry, i.e. *Forced Vital Capacity (FVC)*. The complete list of variables for each visit is available at http://brainteaser.dei.unipd.it/challenges/idpp2022/assets/other/visits.txt.

Measures iDPP@CLEF adopted several state-of-the-art evaluation measures to assess the performance of the prediction algorithms, among which:

- *ROC curve and/or the precision-recall curve (and area under the curve)* to show the trade-off between clinical sensitivity and specificity for every possible cut-off of the risk scores;

[13] https://bitbucket.org/brainteaser-health/

Table 2: Main features of the iDPP@CLEF dataset.

Section	Sub-section	Variables
Baseline	Patient	Sex, Date of Birth
	ALS Onset	Date, Site
	Diagnosis	Date, Regions affected, Diagnostic Delay, FVC, BMI at diagnosis
Follow-up	Progression scores	ALSFRS-R, Rate of disease progression
	Tests	Hematologic tests, Muscle strength assessed by manual testing, Respiratory function tests
	Therapy	ALS treatments
	Other	Regions affected, Upper and lower motor neuron signs, Cognitive and neurophysiological changes
Clinical Events	History	BMI premorbid, Family history, Comorbidities, Previous surgery and trauma
	Interventions	Date of NIV, Date of PEG, Date of Tracheostomy
	Survival	Date of death
Lifestyle	Lifestyle	Working activity, Physical activity, History of smoking, Marital status, Education level

– *Concordance Index (C-index)* to summarize how well a predicted risk score describes an observed sequence of events.
– *E/O ratio and Brier Score* to assess whether or not the observed event rates match expected event rates in subgroups of the model population.
– *Specificity and recall* to assess, for each interval, the ability of the models of correctly identify true positives and true negatives.
– *Distance* to assess how far the predicted time interval was from the true time interval.

To ease the computation and reproducibility of the results, scripts for computing the measures are available in the following repository: https://bitbucket.org/brainteaser-health/idpp2022-performance-computation.

4 iDPP@CLEF 2023

iDPP@CLEF 2023[14] will organize the following activities:

– **Task 1 – Predicting Risk of Disease Worsening (MS)**: it will focus on ranking subjects based on the risk of worsening, setting the problem as a survival analysis task. More specifically the risk of worsening predicted by the algorithm should reflect how early a patient experience the event "worsening". Worsening is defined based on EDSS *Expanded Disability Status Scale (EDSS)* [8] accordingly to clinical standards. In particular we will consider two different definitions in subtasks 1a and 1b, respectively.
  • Subtask 1a: the patient crosses the threshold EDSS ≥ 3 at least twice within one year interval
  • Subtask 1b: the second definition of worsening depends on the first recorded value accordingly to current clinical protocols. If Baseline EDSS

[14] https://brainteaser.health/open-evaluation-challenges/idpp-2023/

< 1, worsening event occurs when and increase of EDSS by 1.5 points is first observed; if $1 \leq$ Baseline EDSS < 5.5, worsening event occurs when and increase of EDSS by 1 point is first observed; if baseline EDSS $\geq$ 5.5, worsening event occurs when and increase of EDSS by 0.5 points is first observed.

In both cases the occurrence of the worsening event and the time of occurrence will be pre-computed by the challenge organizers.

- **Task 2 – Predicting Probability of Worsening (MS)**: it will refine Task 1 asking participants to explicitly assign a probability of worsening at different time windows (e.g. between years 4 and 6, 6 and 8, 8 and 10 etc.). Worsening will be defined in two different ways in subtasks 2a and 2b as for Task1.
- **Position Paper Task 3 – Impact of Exposition to Pollutants (ALS)**: we will evaluate proposals of different approaches to assess if exposure to different pollutants is a useful variable to predict time to *Percutaneous Endoscopic Gastrostomy (PEG)*, *Non-Invasive Ventilation (NIV)* and death in ALS patients.

We will provide retrospective, fully anonymized MS and ALS clinical data including demographic and clinical characteristics, coming from clinical institutions in Italy, Portugal, and Spain.

For Task 1 and Task 2 we will release a brand new dataset with MS data consisting of about 1,800 patients. Accordingly to the survival analysis settings, for each subject in the training we will provide a label 0 or 1 indicating if the subject experienced the event "worsening" (label 1) or not (label 0) and the time-of-event, which indicates the time of the event for subjects experiencing it, or the time of censoring if the patient has not experience the event yet at the time of data dump.

For Position Paper Task 3 we will re-use the ALS dataset developed in iDPP@CLEF 2022, consisting of about 2,250 patients, and will extend it with environmental and pollution data.

5 Conclusions

iDPP@CLEF is a new shared tasks focusing on predicting the temporal progression of ALS and MS and on the explainability of the AI algorithms for such prediction. The first edition focused on ALS progression prediction and participation was satisfactory, hinting at the interest of the community concerning the task. More so, the solutions identified by participants range over several different techniques and provided valid input to such a highly relevant domain as the prediction of the ALS progression.

For the second iteration, iDPP@CLEF 2023 we plan to investigate MS progression prediction and how to exploit pollution and environmental data to improve progression prediction of ALS.

References

1. Branco, R., Soares, D., Martins, A.S., Auletta, E., Castanho, E.N., Nunes, S., Serrano, F., Sousa, R.T., Pesquita, C., Madeira, S.C., Aidos, H.: Hierarchical Modelling for ALS Prognosis: Predicting the Progression Towards Critical Events. In: [4], pp. 1211–1227
2. Buonocore, T.M., Nicora, G., Dagliati, A., Parimbelli, E.: Evaluation of XAI on ALS 6-months mortality prediction. In: [4], pp. 1236–1255
3. Cedarbaum, J.M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., Nakanishi, A.: The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. Journal of the Neurological Sciences **169**(1–2), 13–21 (October 1999)
4. Faggioli, G., Ferro, N., Hanbury, A., Potthast, M. (eds.): CLEF 2022 Working Notes, CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073. http://ceur-ws.org/Vol-3180/ (2022)
5. Guazzo, A., Trescato, I., Longato, E., Hazizaj, E., Dosso, D., Faggioli, G., Di Nunzio, G.M., Silvello, G., Vettoretti, M., Tavazzi, E., Roversi, C., Fariselli, P., Madeira, S.C., de Carvalho, M., Gromicho, M., Chiò, A., Manera, U., Dagliati, A., Birolo, G., Aidos, H., Di Camillo, B., Ferro, N.: Intelligent Disease Progression Prediction: Overview of iDPP@CLEF 2022. In: Barrón-Cedeño, A., Da San Martino, G., Degli Esposti, M., Sebastiani, F., Macdonald, C., Pasi, G., Hanbury, A., Potthast, M., Faggioli, G., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Thirteenth International Conference of the CLEF Association (CLEF 2022), Lecture Notes in Computer Science (LNCS) 13390, Springer, Heidelberg, Germany (2022)
6. Guazzo, A., Trescato, I., Longato, E., Hazizaj, E., Dosso, D., Faggioli, G., Di Nunzio, G.M., Silvello, G., Vettoretti, M., Tavazzi, E., Roversi, C., Fariselli, P., Madeira, S.C., de Carvalho, M., Gromicho, M., Chiò, A., Manera, U., Dagliati, A., Birolo, G., Aidos, H., Di Camillo, B., Ferro, N.: Overview of iDPP@CLEF 2022: The Intelligent Disease Progression Prediction Challenge. In: [4], pp. 1130–1210
7. Küffner, R., Zach, N., Norel, R., Hawe, J., Schoenfeld, D., Wang, L., Li, G., Fang, L., Mackey, L., Hardiman, O., Cudkowicz, M., Sherman, A., Ertaylan, G., Grosse-Wentrup, M., Hothorn, T., van Ligtenberg, J., Macke, J.H., Meyer, T., Schölkopf, B., Tran, L., Vaughan, R., Stolovitzky, G., Leitner, M.L.: Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. Nature Biotechnology **33**(1), 51–57 (January 2015)
8. Kurtzke, J.F.: Rating neurologic impairment in multiple sclerosis. an expanded disability status scale (edss). Neurology **33**(11), 1444–1452 (November 1983)
9. Mannion, A., Chevalier, T., Schwab, D., Goeuriot, L.: Predicting the Risk of & Time to Impairment for ALS patients. In: [4]
10. Nunes, S., Sousa, R.T., Serrano, F., Branco, R., Soares, D.F., Martins, A.S., Auletta, E., Castanho, E.N., Madeira, S.C., Aidos, H., Pesquita, C.: Explain-

ing Artificial Intelligence Predictions of Disease Progression with Semantic Similarity. In: [4], pp. 1256–1268
11. Pancotti, C., Birolo, G., Sanavia, T., Rollo, C., Fariselli, P.: Multi-Event Survival Prediction for Amyotrophic Lateral Sclerosis. In: [4], pp. 1269–1276
12. Trescato, I., Guazzo, A., Longato, E., Hazizaj, E., Roversi, C., Tavazzi, E., Vettoretti, M., Di Camillo, B.: Baseline Machine Learning Approaches To Predict Amyotrophic Lateral Sclerosis Disease Progression. In: [4], pp. 1277–1293