iDPP@CLEF 2024: 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.

iDPP@CLEF run in CLEF 2022 and 2023, offering tasks on the prediction of ALS and MS progression, using retrospective patient clinical data complemented with environmental data.

iDPP@CLEF 2024 will focus on prospective patient data for ALS collected via a dedicated app developed by the BRAINTEASER project and sensor data in the context of clinical trials in Turin, Pavia, Lisbon, and Madrid. For MS, iDPP@CLEF 2024 will rely on retrospective patient data complemented with environmental and pollution data from clinical institutions in Pavia and Turin.

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; moreover, 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] is organized by the BRAINTEASER project and 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: Sections 2 and 3 present what has been done in iDPP@CLEF 2022 and 2023; Section 4 introduces the datasets made available by iDPP@CLEF; while Section 5 discusses the plans for iDPP@CLEF 2024; finally, Section 6 draws some conclusions.

2 iDPP@CLEF 2022

iDPP@CLEF [6, 7] run as a pilot lab for the first time in CLEF 2022 and focused on pilot activities aimed both at an initial exploration of ALS progression

[8] https://brainteaser.health/open-evaluation-challenges/

prediction and at understanding the challenges and limitations to refine and tune the lab itself for future iterations.

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 *ALS Functional Rating Scale Revisited (ALSFRS-R)* scale [1] 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 frameworks able to explain the multivariate nature of the data and the model predictions.

43 participants registered for iDPP@CLEF 2022 and 5 participants successfully submitted a total of 120 runs for Task 1 and Task 2; moreover, 2 position papers were submitted for the explainability task.

Submission of participants are openly available in git repositories[9] and all the participant papers and slides are available through the iDPP@CLEF 2022 web site[10].

3 iDPP@CLEF 2023

iDPP@CLEF 2023 [3, 4] was the second iteration of the lab, expanding its scope to include both ALS and MS in the study of disease progression. The activities in iDPP@CLEF 2023 focus on two objectives: exploring the prediction of MS worsening and conducting a more in-depth analysis of ALS compared to iDPP@CLEF 2022, with the addition of environmental data.

iDPP@CLEF 2023[11] will organize the following activities:

- **Task 1 – Predicting Risk of Disease Worsening (MS)**: it focused 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 experiences the "worsening" event, and should range between 0 and 1. Worsening is defined on the basis of the *Expanded Disability Status Scale (EDSS)* [8], accordingly to clinical standards. In particular, we considered two different definitions of worsening corresponding to two different sub-tasks:

[9] https://bitbucket.org/brainteaser-health/
[10] https://brainteaser.health/open-evaluation-challenges/idpp-2022/
[11] https://brainteaser.health/open-evaluation-challenges/idpp-2023/

  - *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 < 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 ≥ 5.5, worsening event occurs when and increase of EDSS by 0.5 points is first observed.

  For each sub-task, participants were given a dataset containing 2.5 years of visits, with the occurrence of the worsening event and the time of occurrence pre-computed by the challenge organizers.
- **Task 2 – Predicting Cumulative Probability of Worsening (MS)**: it refined Task 1 by asking participants to explicitly assign the cumulative probability of worsening at different time windows, i.e., between years 0 and 2, 0 and 4, 0 and 6, 0 and 8, 0 and 10.
  In particular, we considered two different definitions of worsening corresponding to two different sub-tasks – *subtask 2a* and *subtask 2b* – defined according to the same rules as in Task 1.
- **Position Paper Task 3 – Impact of Exposition to Pollutants (ALS)**: we evaluated 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. This task was based on the same data and the same design as Task 1 in iDPP@CLEF 2022. The difference with respect to the previous year task is that we complemented those data with environmental data to investigate the impact of exposition to pollutants on prediction of disease progression. Since both training and test data were immediately available, we considered these submissions as position papers.

45 participants registered for iDPP@CLEF 2023 and 10 participants successfully submitted a total of 163 runs for Task 1, Task 2, and , Task 3.

Submission of participants are openly available in git repositories[12] and all the participant papers and slides are available through the iDPP@CLEF 2023 web site[13].

4 Datasets

iDPP@CLEF 2022 created a dataset, for the prediction of specific events related to ALS, consisting of fully anonymized data from 2,204 real patients from medical institutions in Turin, Italy, and Lisbon, Portugal. The dataset contains 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.

[12] https://bitbucket.org/brainteaser-health/
[13] https://brainteaser.health/open-evaluation-challenges/idpp-2023/

iDPP@CLEF 2023 created a dataset, for the prediction of worsening of MS, consisting of fully anonymized data from 1,792 real patients from medical institutions in Pavia, Italy, and Turin, Italy. The dataset contains both static data about patients, e.g. age, gender, ... and 2.5 years of visits. (EDSS scores, evoked potentials, relapses, MRIs).

All the datasets are highly curated and they are produced from the *BrainTeaser Ontology (BTO)*[14] [5] which ensures the consistency of the data represented. Moreover, several checks have been performed to ensure that all the instances are clean, contain proper values in the expected ranges, and do not have contradictions.

All the datasets are available for further research on Zenodo [2] and access to them is regulated by tthe BRAINTEASER data sharing policy. Researchers have to submit a brief proposal, specifying the research questions, the adopted methods and algorithms, how datasets will be used, what are the expected outcomes. Before granting the access to the datasets, a committee constituted by members of the BRAINTEASER project assesses the proposal in order to ensure a proper use of the datasets and the quality and appropriateness, also from a medical point of view, of the expected claims and inferences derived from the datasets.

5 iDPP@CLEF 2024

While the previous editions focused on retrospective patient data, for ALS iDPP@CLEF 2024 will focus on prospective patient data collected via a dedicated app developed by the BRAINTEASER project and sensor data in the context of clinical trials in Turin, Pavia, Lisbon, and Madrid.

For MS, iDPP@CLEF 2024 will rely on retrospective patient data prepared in iDPP@CLEF 2023 complemented with environmental and pollution data from clinical institutions in Pavia and Turin.

In particular, we will organize the following activities:

- **Task 1 - Predicting ALSFRS-R score from sensor data (ALS)**: it will focus on predicting the ALSFRS-R score, assigned by medical doctors roughly every three months, from the sensor data collected via the app. The ALSFRS-R score is a somehow "subjective" evaluation performed by a medical doctor and this task will help in answering a currently open question in the research community, i.e. whether it could be derived from objective factors.
- **Task 2 - Predicting patient self-assessment score from sensor (ALS)**: it will focus on predicting the self-assessment score assigned by patients from the sensor data collected via the app. If the self-assessment performed by patients, more frequently than the assessment performed by medical doctors every three months or so, can be reliably predicted by sensor and app data, we can imagine a proactive application which, monitoring the sensor data, alerts the patient if an assessment is needed.

[14] https://w3id.org/brainteaser/ontology

– **Task 3 - Predicting relapses from EDSS sub-scores and environmental data (MS)**: it will focus on predicting a relapse using environmental data and EDSS sub-scores. This task will allow us to assess if exposure to different pollutants is a useful variable in predicting a relapse.

We will provide prospective, fully anonymized MS and ALS clinical data including demographic and clinical characteristics as well as environmental and sensor data, coming from clinical trials currently running at institutions in Italy, Portugal, and Spain.

For Task 1 and Task 2, we will release a brand-new dataset consisting of 100 ALS patients that were followed for up to 18 months and whose progression was tracked by regular clinical evaluations. Note that even if the absolute number of patients might not seem high, this is a very rich dataset due to the daily collection of sensor data, with tens of thousands of data points in total.

For Task 3, we will re-use part of the MS dataset developed in iDPP@CLEF 2023 and we will extend it with environmental and pollution data.

6 Conclusions

iDPP@CLEF is a 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, iDPP@CLEF 2022, focused on ALS progression prediction and participation was satisfactory, hinting at the interest of the community concerning the task.

The second iteration, iDPP@CLEF 2023, we investigated MS progression prediction and how to exploit pollution and environmental data to improve progression prediction of ALS. The participation sensibly increased with respect to the previous edition, indicating the strong interest of the community on these topics.

As a further outcome of the first two iterations of iDPP@CLEF, we produced two large datasets containing retrospective clinical data about real patients and made them freely available through Zenodo.

The third iteration, iDPP@CLEF 2024, will shift the focus to prospective data about ALS patients, collected via a mobile app and dedicated sensors made available by the BRAINTEASER project, and on the prediction of ALSFRS-R scores from sensor data. For MS, we will focus on understanding the impact of pollution and environmental data on the prediction of the progression of the disease.

Acknowledgements

The work reported in this paper is partially supported by the European Union Horizon 2020 BRAINTEASER project[15], grant agreement number GA101017598.

[15] https://brainteaser.health/

References

1. 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)
2. Faggioli, G., Guazzo, A., Marchesin, S., Menotti, L., Trescato, I., Aidos, H., Bergamaschi, R., Birolo, G., Cavalla, P., Chiò, A., Dagliati, A., de Carvalho, M., Di Nunzio, G.M., Fariselli, P., García Dominguez, J.M., Gromicho, M., Longato, E., Madeira, S.C., Manera, U., Silvello, G., Tavazzi, E., Tavazzi, E., Vettoretti, M., Di Camillo, B., Ferro, N.: BRAINTEASER ALS and MS Datasets. Zenodo, https://doi.org/10.5281/zenodo.8083180 (June 2023)
3. Faggioli, G., Guazzo, A., Marchesin, S., Menotti, L., Trescato, I., Aidos, H., Bergamaschi, R., Birolo, G., Cavalla, P., Chiò, A., Dagliati, A., de Carvalho, M., Di Nunzio, G.M., Fariselli, P., García Dominguez, J.M., Gromicho, M., Longato, E., Madeira, S.C., Manera, U., Silvello, G., Tavazzi, E., Tavazzi, E., Vettoretti, M., Di Camillo, B., Ferro, N.: Intelligent Disease Progression Prediction: Overview of iDPP@CLEF 2023. In: Arampatzis, A., Kanoulas, E., Tsikrika, T., Vrochidis, S., Giachanou, A., Li, D., Aliannejadi, A., Vlachos, M., Faggioli, G., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fourteenth International Conference of the CLEF Association (CLEF 2023), Lecture Notes in Computer Science (LNCS) 14163, Springer, Heidelberg, Germany (2023)
4. Faggioli, G., Guazzo, A., Marchesin, S., Menotti, L., Trescato, I., Aidos, H., Bergamaschi, R., Birolo, G., Cavalla, P., Chiò, A., Dagliati, A., de Carvalho, M., Di Nunzio, G.M., Fariselli, P., García Dominguez, J.M., Gromicho, M., Longato, E., Madeira, S.C., Manera, U., Silvello, G., Tavazzi, E., Tavazzi, E., Vettoretti, M., Di Camillo, B., Ferro, N.: Overview of iDPP@CLEF 2023: The Intelligent Disease Progression Prediction Challenge. In: Aliannejadi, M., Faggioli, G., Ferro, N., Vlachos, M. (eds.) CLEF 2023 Working Notes, pp. 1123–1164, CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, https://ceur-ws.org/Vol-3497/. (2023)
5. Faggioli, G., Marchesin, S., Menotti, L., Di Nunzio, G.M., Silvello, G., Ferro, N.: The BrainTeaser Ontology. Zenodo, https://doi.org/10.5281/zenodo.7886997 (May 2023)
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.: 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), pp. 395–422, Lecture Notes in Computer Science (LNCS) 13390, Springer, Heidelberg, Germany (2022)

7. 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: Faggioli, G., Ferro, N., Hanbury, A., Potthast, M. (eds.) CLEF 2022 Working Notes, pp. 1130–1210, CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073. http://ceur-ws.org/Vol-3180/ (2022)
8. Kurtzke, J.F.: Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). Neurology **33**(11), 1444–1452 (November 1983)