

Explaining Artificial Intelligence Predictions of Disease Progression with Semantic Similarity

Susana Nunes[†], Rita T. Sousa[†], Filipa Serrano[†], Ruben Branco, Diogo F. Soares, Andreia S. Martins, Eleonora Auletta, Eduardo N. Castanho, Sara C. Madeira, Helena Aidos and Catia Pesquita*

LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

Abstract

The complexity of neurodegenerative diseases has motivated the development of artificial intelligence approaches to predicting risk of impairment and disease progression. However, and despite the success of these approaches, their mostly black-box nature hinders their adoption for disease management. Explainable artificial intelligence holds the promise to bridge this gap by producing explanations of models and their predictions that promote understanding and trust by users. In the biomedical domain, given its complexity, explainable artificial intelligence approaches have much to benefit from being able to link models to representations of domain knowledge – ontologies. Ontologies afford more explainable features because they are semantically enriched and contextualized and as such can be better understood by end users; and they also model existing knowledge, and thus support inquiry into how a given artificial intelligence model outcome fits with existing scientific knowledge. We propose an explainability approach that leverages on the rich panorama of biomedical ontologies to build semantic similarity-based explanations that contextualize patient data and artificial intelligence predictions. These explanations mirror a fundamental human explanatory mechanism - similarity - while tackling the challenges of data complexity, heterogeneity and size.

Keywords

Semantic Similarity, Ontology, Amyotrophic Lateral Sclerosis, Explainable AI

1. Introduction

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease targeting the upper and lower motor neurons, leading to progressive and diffuse paralysis and eventual respiratory death. The underlying causes of ALS are not fully understood, and there is a heterogeneity of clinical symptoms and a highly fluctuating life expectancy. Therefore, multiple approaches have been proposed to predict disease progression to improve patient-personalized treatment [1, 2].

The iDPP :robot: CLEF 2022 [3, 4] is focused on the evaluation of Artificial Intelligence (AI)

CLEF 2022: Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy

*Corresponding author.

[†]These authors contributed equally.

✉ scnunes@ciencias.ulisboa.pt (S. Nunes); risousa@ciencias.ulisboa.pt (R. T. Sousa); fserrano@lasige.di.fc.ul.pt (F. Serrano); rmbranco@ciencias.ulisboa.pt (R. Branco); dfsoares@ciencias.ulisboa.pt (D. F. Soares); amartins@lasige.di.fc.ul.pt (A. S. Martins); eauletta@ciencias.ulisboa.pt (E. Auletta); ejcastanho@ciencias.ulisboa.pt (E. N. Castanho); sacmadeira@ciencias.ulisboa.pt (S. C. Madeira); haidos@ciencias.ulisboa.pt (H. Aidos); clpesquita@ciencias.ulisboa.pt (C. Pesquita)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

algorithms to predict ALS progression. Task 2, in particular, is dedicated to predicting time of impairment targeting different events: non-invasive ventilation (NIV), percutaneous endoscopic gastrostomy (PEG), and death. It includes highly-curated datasets based on real ALS patients, followed at clinical institutions in Lisbon, Portugal, and Turin, Italy.

Machine learning (ML) approaches are increasingly successful in predicting the progression of ALS, including the prognostic prediction of the need for non-invasive ventilation [5, 6] and patient profiling [7]. The complexity of the underlying domain and data, coupled with the black-box nature of many of the algorithms employed hinder their widespread adoption on the management of these diseases. To tackle this, one of the challenge's goals is also to propose new approaches to make those prediction algorithms explainable.

Explainable artificial intelligence (XAI) is focused on developing approaches that ensure algorithmic fairness, identify potential bias/problems in the training data, ensure that the algorithms perform as expected, and bridge the gap between the ML community and other scientific disciplines [8]. The lack of trust is the main barrier in the adoption of AI in clinical practice [9]. XAI approaches can be widely classified into two categories: transparent models, which are interpretable by design and include decision trees and linear models [10] and post-hoc explainability [11, 12], where models are explained through external techniques. This last category can also be divided into model-agnostic techniques, designed to be applied to any ML model, and model-specific techniques tailored to explain a particular ML model. Recently, post-hoc methods have been applied to explain recurrent neural networks that predict the need for non-invasive ventilation [13].

In the biomedical domain, given its complexity, XAI approaches have much to benefit from being able to link ML models to representations of domain knowledge [14, 15]. These *Explainable Knowledge-enabled Systems* include a representation of the domain knowledge and support explanation approaches that are context-aware and provenance enabled [16]. This knowledge integration is crucial, since well-grounded explanations should be domain-dependent, which means that they should be set within a context that depends on the task, background knowledge and expectations of the user [17].

Semantic web technologies such as ontologies and knowledge graphs represent an unparalleled solution to the problem of knowledge-enabled explanations since they provide the needed semantic context [18]. Ontologies are semantic models that encode the knowledge of a domain, in which each element is precisely defined, and the relationships between elements are parameterized or constrained [19].

Ontologies add two very relevant dimensions to explainability: one the one hand, they afford explainable features, i.e., features that are semantically enriched and contextualized and as such can be better understood by end users; and on the other, they model existing knowledge, and thus support inquiry into how a given AI model outcome fits with extant knowledge. These two aspects are key for explainability and trust.

This work builds on the recently developed Brainteaser Ontology [20] and explores how this ontology coupled with the rich panorama of more general biomedical ontologies can support similarity-based explanations for patient end-stage event predictions that build upon the contextualization of patient data and AI predictions.

2. Proposal

Similarity assessment is a natural explanatory mechanism [21] since the identification of similar features to group similar objects is a basic cognitive ability. This is a frequent strategy in clinical settings, where similarity between patients can be used by clinicians to help decide on the best course of action through analogical reasoning [22]. Measuring similarity is also fundamental to many ML algorithms, making similarity both a representation that can be explored by ML and explanation approaches. The basis of a similarity-based explanation is that a prediction for a given entity is formulated based on its similarity to other relevant entities with known outcome.

While comparing the numerical values assigned to patient features such as weight or blood pressure can be achieved directly, the comparison of more complex concepts, such as symptoms, is not straightforward. However, when ontologies are used to describe data, the semantic annotations of data objects can be used to compute their semantic similarity [23]. A semantic annotation assigns real-world entities in a domain to their semantic description (i.e. class of an ontology) [24], and the ontology data model can be applied to a set of individual entities to create a knowledge graph [25]. Ontologies thus provide the scaffolding for comparing entities at a higher level of complexity by comparing their meaning. A semantic similarity measure is then a function that, given two entities described in the knowledge graph, returns a numerical score reflecting the closeness in meaning between them.

Figure 1 presents a subgraph of the Brainteaser Ontology (BO) enriched with the National Cancer Institute Thesaurus (NCIt) ontology [26], where patients are described according to the symptoms that characterize them. Symptoms are organized into a hierarchical structure given by the *is_a* links. Patients B and C can be considered more similar because they share the annotations "Spasticity" and "Feeding difficulties and mismanagement". Patient A is also somewhat similar to Patient B since they are both annotated with "Muscle symptom". Most semantic similarity measures are taxonomic in nature, they explore the hierarchical relations between classes in an ontology. For instance, Patient B is directly annotated with "Muscle twitching" whereas Patient A exhibits "Muscle symptom". By exploring the hierarchical relations modeled in the ontology, both patients share an annotation to "Muscle symptom". Consequently, ontologies with a rich taxonomic backbone typically allow more precise measures of semantic similarity.

Semantic similarity-based explanations are based on understanding why particular objects, in this case, patients, are considered similar or different and how that relates to the AI predictions. Classical semantic similarity returns a single score, which on the one hand is easy to understand and condenses the representation of a potential very large feature set into a single score, but on the other affords little insight. To improve on the level of detail supported by similarity-based explanations, we propose to compute several semantic similarity scores that describe similarity in meaning under different semantic aspects encoded by the ontologies. A semantic aspect (SA) corresponds to a subgraph of the ontology rooted at a class of interest. For example, an explanation may be grounded on three different semantic similarity scores: co-morbidities similarity, lifestyle similarity and symptom similarity.

These semantic similarity-based explanations are considered post-hoc, since they provide explanations for the models' predictions, as opposed to approaches that explain the models' workings. They are also considered local, since they focus on explaining specific predictions

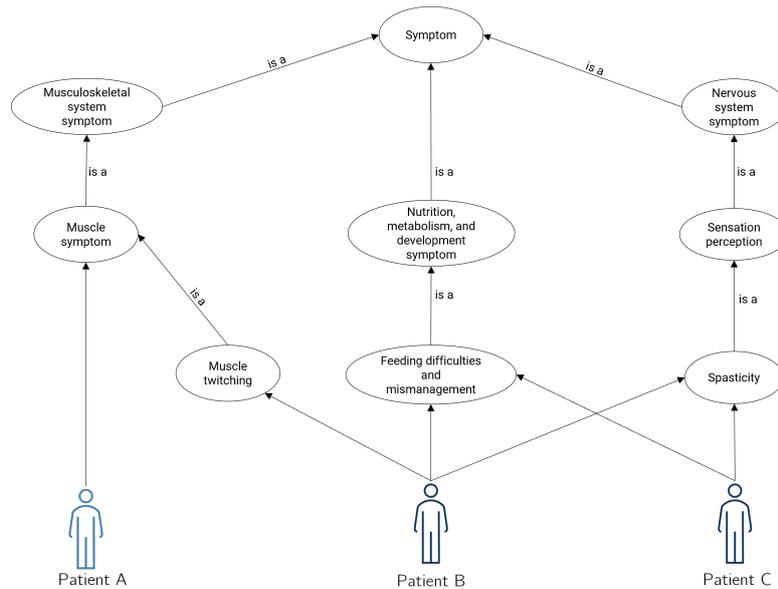


Figure 1: Semantic annotation of patients using a subgraph of Brainteaser Ontology enriched with the NCI ontology.

made by the model, instead of providing global explanations that apply to all predictions [27]. In particular, semantic similarity explanations focus on how a certain prediction fits with the patient features and their interpretation according to existing scientific knowledge encoded in the ontologies.

3. Methodology

In the context of the iDPP $\ddot{\cdot}$ CLEF 2022, we propose a novel approach that generates semantic similarity-based explanations for patient-level predictions. The underlying idea is that we can explain the prediction for one patient by considering aspect-oriented semantic similarity with other relevant patients based on the most important features used by ML approaches or selected by users.

To build rich and easy to understand semantic-similarity based explanations, our approach requires five steps (see Figure 2) : (1) the enrichment of the Brainteaser Ontology through integration of other biomedical ontologies; (2) the semantic annotation of patients (if not already available); (3) the similarity calculation between patients; (4) selection of the set of patients to explain a specific prediction; and (5) the visualization of the generated similarity-based explanations.

3.1. Ontology Integration

To ensure a rich and comprehensive semantic annotation of the data, we enrich the Brainteaser Ontology with links to other ontologies. The BO models the data collected to describe disease

progression in ALS and MS. It reuses single classes from other biomedical ontologies listed in Table 1. However it does not include import statements which hinders its ability to support a more complete semantic similarity calculation since it is missing the context of the reused classes. By explicitly linking BO to the other 9 biomedical ontologies and controlled vocabularies it reuses through import statements we can establish a rich semantic landscape with approximately 770,680 classes in total and covering a variety of domains.

Table 1

Main ontologies and number of classes used to create the Brainteaser Ontology.

Ontology	Classes
National Cancer Institute thesaurus (NCIt)	173,001
Ontology of Genes and Genomes (OGG)	69,689
Uberon	20,849
Medical Action Ontology (MAxO)	15,086
Ontology for MIRNA Target (OMIT)	90,916
SNOMED Clinical Terms (SNOMED CT)	358,483
Anatomical Therapeutic Chemical Classification (ATCC)	6,567
International Standard Classification of Occupations (ISCO)	619
Experimental Factor Ontology (EFO)	35,470

Figure 3 represents the impact of this enrichment via integration. On the left side the subgraph of the neighborhood of the classes "Rheumatoid Arthritis", "Pulmonary Sarcoidosis" and "Heart Disorder" as modeled in the BO is shown. Using the unenriched BO ontology, would result in patients exhibiting these diseases or disorders being all considered equally similar: the shared meaning of these ontology classes is simply that they are all subclasses of "Disease or Disorder". The right side illustrates the enriched subgraph including the full NCIt hierarchy, which encodes considerably more contextual information, and supports semantic similarity measures in correctly identifying that "Rheumatoid Arthritis" and "Pulmonary Sarcoidosis" are more similar, since they are both Autoimmune Diseases.

3.2. Patient Semantic Annotation

Semantic annotation needs to be performed only when patient data is not annotated with the ontology. Semantic annotations can be performed manually, but there are a number of existing tools that facilitate this step such as the well known NCBO Annotator [28, 29]. An annotation assigns an ontology class to describe an entity. In our case an annotation is a tuple $a = \langle \textit{patient}, \textit{ontology class}, \textit{time point} \rangle$ that defines an ontology class to describe a patient at a given point in time (clinical visit). A patient can then be represented by a vector of all their annotations, i.e. the ontology classes that describe them and the time points where the assignment was registered.

$$P = \{a_1, \dots, a_n\} \quad (1)$$

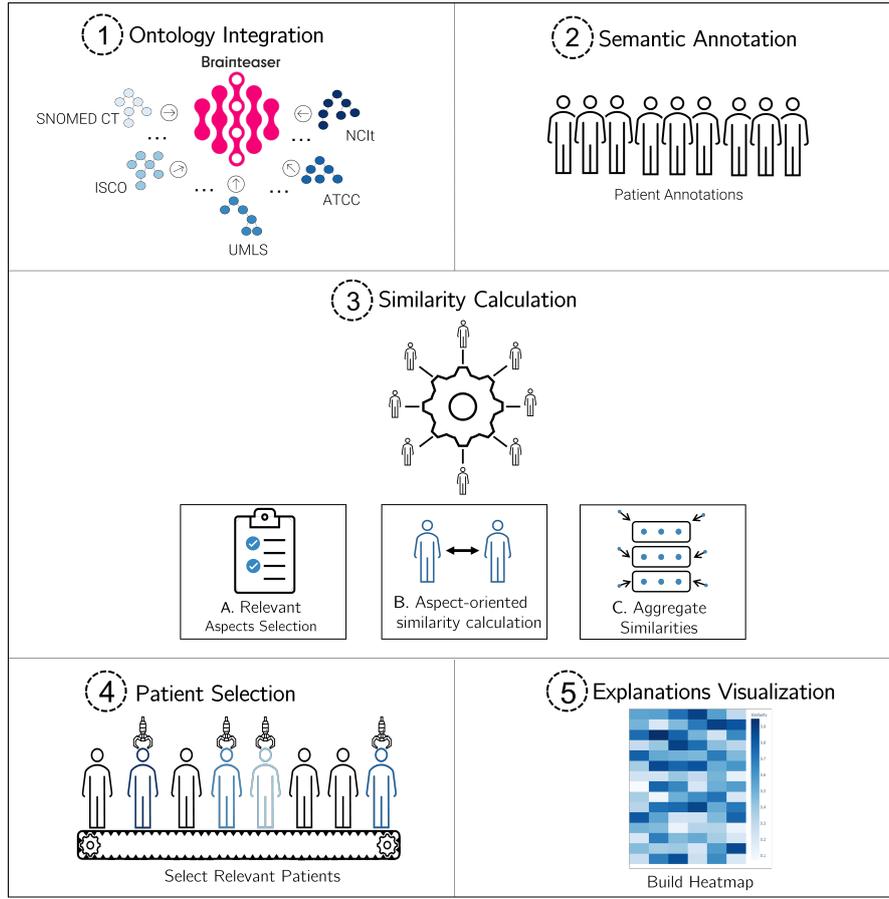


Figure 2: Overview of the semantic-similarity explanations methodology

3.3. Semantic Similarities between Patients

We take as input the most important features employed by ML methods to predict disease progression and/or other features considered relevant by users, and use them to define the different semantic aspects for which semantic similarity is calculated.

More formally, the similarity between two patients corresponds to a vector of semantic similarity scores corresponding to each relevant aspect. This similarity may be computed irrespective of time of annotation (considering all annotations for all time points) or by time point.

$$sim(P_1, P_2) = \{sim_{SA_1}(P_1, P_2), \dots, sim_{SA_n}(P_1, P_2)\} \quad (2)$$

Several well-established works explore the taxonomical (hierarchical) component of the ontologies to measure the shared meaning between two entities in the ontology [30].

A popular semantic similarity measure is ResnikBMA based on the class-based measure proposed by Resnik [31] in which the similarity between two classes corresponds to the In-

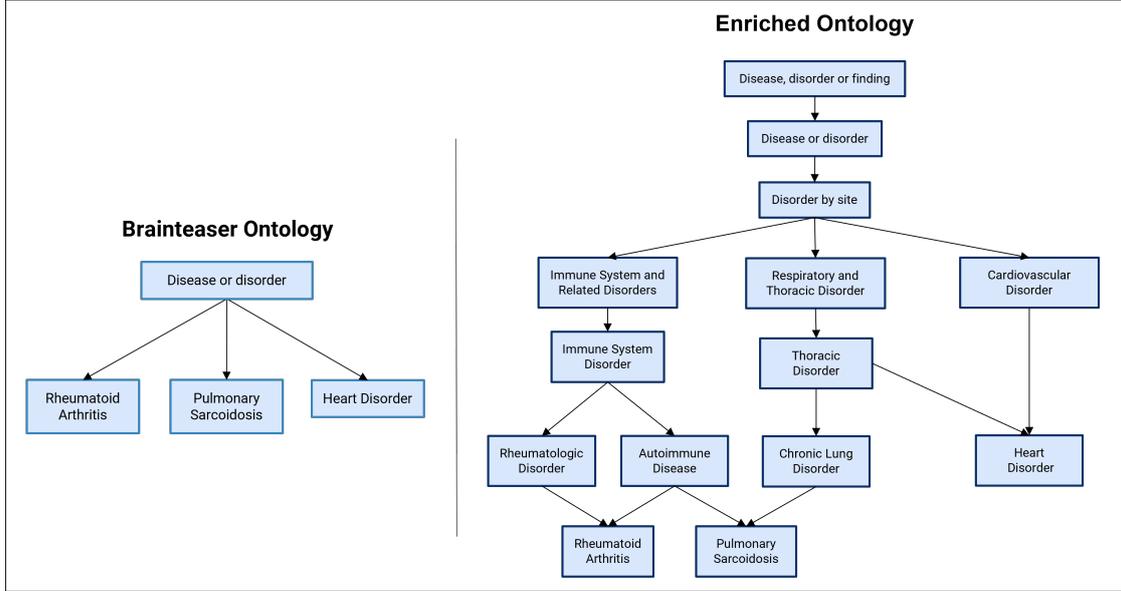


Figure 3: Impact of the Brainteaser Ontology enrichment through the integration of several relations between domains.

formation Content (IC) of their most informative common ancestor. The IC is a measure that reflects the specificity of a class in the ontology.

$$sim(c_1, c_2) = \max\{IC(c) : c \in \{A(c_1) \cap A(c_2)\}\} \quad (3)$$

where c is a class in $A(c_i)$, the set of ancestors of c_i .

ResnikBMA then uses the Best-Match Average to consider the best scoring pairs of classes from each entity.

$$BMA(e_1, e_2) = \frac{\sum_{c_1 \in A(e_1)} sim(c_1, c_2)}{2|A(e_1)|} + \frac{\sum_{c_2 \in A(e_2)} sim(c_1, c_2)}{2|A(e_2)|} \quad (4)$$

where $A(e_i)$ is the number of annotations for entity e_i and $sim(e_1, e_2)$ is the semantic similarity between the class c_1 and class c_2 .

Recently, more sophisticated approaches based on knowledge graph embeddings allow the representation of each entity with a vector that approximates the similarity properties of the graph and can then be used to compute similarity using operations such as cosine similarity [32].

3.4. Building Semantic Similarity Explanations

The models proposed by our team for the task of predicting the time of impairment for ALS of the iDPP :CLEF 2022 addressed the prediction of the event separately from predicting the time window to the event. For event prediction, all time points were considered but the first three proved to be the most relevant for the models. These were then also used in predicting time to event. Likewise, to develop explanations better suited for this approach, we also developed two types of explanations: event explanation and time window to event explanation.

A key aspect is selecting the most relevant patients to explain an event prediction. We define three types of explanatory patients: (1) The N most similar patients exhibiting the same outcome; (2) The N least similar patients with the same outcome; (3) The N most similar patients exhibiting a different outcome. Since we compute several similarities in the previous step, to select these explanatory patients, we propose an aggregated similarity where different features can be assigned weights. Candidate explanatory patients are then ranked by their semantic similarity to the target patient and the final set of explanatory patients is selected.

A semantic similarity explanation of a given target patient can then be defined as a two-dimensional tensor where the first dimension represents patients and the second dimension the semantic similarities according to different aspects:

$$SSE(P_{target}) = \{sim(P_{target}, P_i), \dots, sim(P_{target}, P_n)\} \quad (5)$$

where $P_{i..n}$ correspond to patients sampled from the three types.

Regarding the prediction of a time window to a specific event, we follow a similar approach: (1) The N most similar patients exhibiting the same event in the same time window; (2) The N least similar patients exhibiting the same event in the same time window; (3) The N most similar patients exhibiting the same event in a different time window. N can be defined by the end-user, and may take a different value for each patient type.

The explanations for the time window to event predictions take the form of a three-dimensional tensor, where the first dimension is the explanatory patients, the second dimension corresponds to the semantic similarities according to different aspects and the third dimension represents the time points.

3.5. Explanation Visualization

Finally, one fundamental aspect of explainability is communication to the end user. Our proposal orchestrates semantic similarity explanations into a visualization that combines global and aspect-oriented similarity for different sets of relevant patients.

We chose to use heatmaps to visualize the similarity explanations. For event prediction explanations the representation of the two-dimensional tensor is straightforward, the y-axis contains the patients and the x-axis the similarity scores according to the different aspects. Each cell in the map is a colour-coded representation of the similarity of each patient to the patient we want to explain for each feature, where a darker colour corresponds to higher similarity. The visualization also includes the global similarity employed to select the different types of patients. For time to event explanations, the third dimension is encapsulated within the x-axis with three cells (one per time point).

4. Semantic Similarity Explanations

To showcase our proposal for semantic similarity explanations, we simulated patients with annotations to the BO. One of them is the target patient for whom an event is predicted and an explanation is built. The remaining patients are patients for whom the outcome is known and they are used to explain the target patient.

To explain an event prediction, we set $N=2$ for the three types of explanatory patients. To explain time window prediction we employed $N=1,5,1$ for each explanatory patient type respectively. The heatmaps were built using the following selection of features: lifestyle, onset, co-morbidities, symptoms and pharmacological substances. Figure 4 depicts the event prediction explanation, while Figure 5 illustrates the time window to event prediction explanation. Patients in green correspond to the most similar patients to the target patient (P1) and whose outcome event or time to event is the same. Patients in white correspond to the most similar patients similar to P1, but with a different prediction. Finally, patients in pink correspond to the least similar patients to the target patient, but who share the same prediction.

Figure 4: Semantic similarity explanation for the event prediction.

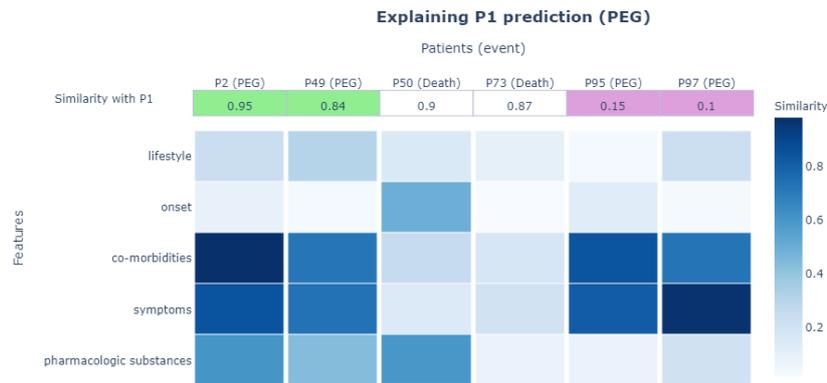
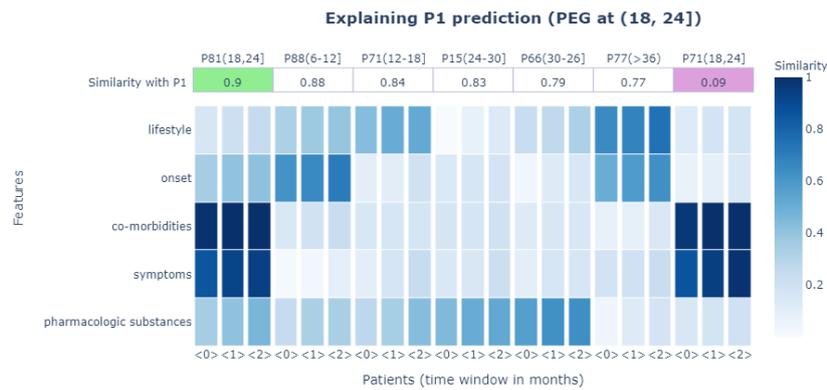


Figure 5: Semantic similarity explanation for the time window to the event prediction.



5. Conclusions

Similarity is a natural explanatory approach, but computing similarity for complex data is challenging. Semantic similarity provides an opportunity for explainability in the context of the iDPP CLEF 2022 since it allows more complex patient comparisons supported by the scientific context encoded in ontologies. Our proposal is based on first enriching the Brainteaser Ontology with explicit imports of reused ontologies to support a more granular computation of patient semantic similarity. The semantic similarity explanations are based on calculating specific semantic similarity values according to relevant aspects either identified as relevant for the ML predictions or selected by the end users. The end result is heatmap-based visualizations that allow the comparison of the target patient whose predictions we want to understand with other relevant patients with known outcomes. This method can be integrated with other types of similarities, for instance, comparisons of Revised Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS-R) scores.

A challenge for this explainability method is that it requires patient data to be sufficiently detailed to ensure high quality annotations. The data currently available for the challenge can be used to support these methods, but it does not yet afford a complete patient representation under the BO ontology. As data becomes richer and more detailed, the potential value of semantic similarity explanations increases. However, their true value must be measured in user studies, where predictions and explanations are shown to experts and their ability to help users understand a prediction is assessed.

Acknowledgements

The authors are funded by the FCT through LASIGE Research Unit (ref. UIDB/00408/2020 and ref. UIDP/00408/2020), AIpALS (PTDC/CCI-CIF/4613/2020), PhD research scholarships to RTS (SFRH/BD/145377/2019), DFS (2020.05100.BD), ENC (2021.07810.BD); and by the BRAINTEASER project and KATY project that have received funding from the European Union's Horizon 2020 research and innovation programme under the grants agreements No 101017598 and No 101017453, respectively.

References

- [1] V. Grollemund, P.-F. Pradat, G. Querin, F. Delbot, G. Le Chat, J.-F. Pradat-Peyre, P. Bede, Machine learning in amyotrophic lateral sclerosis: achievements, pitfalls, and future directions, *Frontiers in neuroscience* 13 (2019) 135.
- [2] S. Pires, M. Gromicho, S. Pinto, M. d. Carvalho, S. C. Madeira, Patient stratification using clinical and patient profiles: Targeting personalized prognostic prediction in als, in: *International Work-Conference on Bioinformatics and Biomedical Engineering*, Springer, 2020, pp. 529–541.
- [3] A. Guazzo, I. Trescato, E. Longato, E. Hazizaj, D. Dosso, G. Faggioli, G. M. Di Nunzio, G. Silvello, M. Vettoretti, E. Tavazzi, C. Roversi, P. Fariselli, S. C. Madeira, M. de Carvalho, M. Gromicho, A. Chiò, U. Manera, A. Dagliati, G. Birolo, H. Aidos, B. Di Camillo,

- N. Ferro, Intelligent Disease Progression Prediction: Overview of iDPP@CLEF 2022, in: A. Barrón-Cedeño, G. Da San Martino, M. Degli Esposti, F. Sebastiani, C. Macdonald, G. Pasi, A. Hanbury, M. Potthast, G. Faggioli, N. Ferro (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.
- [4] A. Guazzo, I. Trescato, E. Longato, E. Hazizaj, D. Dosso, G. Faggioli, G. M. Di Nunzio, G. Silvello, M. Vettoretti, E. Tavazzi, C. Roversi, P. Fariselli, S. C. Madeira, M. de Carvalho, M. Gromicho, A. Chiò, U. Manera, A. Dagliati, G. Birolo, H. Aidos, B. Di Camillo, N. Ferro, Overview of iDPP@CLEF 2022: The Intelligent Disease Progression Prediction Challenge, in: G. Faggioli, N. Ferro, A. Hanbury, M. Potthast (Eds.), *CLEF 2022 Working Notes, CEUR Workshop Proceedings (CEUR-WS.org)*, ISSN 1613-0073, 2022.
- [5] D. Soares, R. Henriques, M. Gromicho, S. Pinto, M. d. Carvalho, S. C. Madeira, Towards triclustering-based classification of three-way clinical data: A case study on predicting non-invasive ventilation in als, in: *International Conference on Practical Applications of Computational Biology & Bioinformatics*, Springer, 2020, pp. 112–122.
- [6] A. S. Martins, M. Gromicho, S. Pinto, M. de Carvalho, S. C. Madeira, Learning prognostic models using diseaseprogression patterns: Predicting the need for non-invasive ventilation in amyotrophic lateral sclerosis, *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (2021).
- [7] T. Leão, S. C. Madeira, M. Gromicho, M. de Carvalho, A. M. Carvalho, Learning dynamic bayesian networks from time-dependent and time-independent data: Unraveling disease progression in amyotrophic lateral sclerosis, *Journal of Biomedical Informatics* 117 (2021) 103730.
- [8] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, L. Kagal, Explaining explanations: An overview of interpretability of machine learning, in: *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, IEEE, 2018, pp. 80–89.
- [9] E. Glikson, A. W. Woolley, Human trust in artificial intelligence: Review of empirical research, *Academy of Management Annals* 14 (2020) 627–660.
- [10] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, *Nature Machine Intelligence* 1 (2019) 206–215.
- [11] M. T. Ribeiro, S. Singh, C. Guestrin, "Why should i trust you?": explaining the predictions of any classifier, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, Association for Computing Machinery, New York, NY, USA, 2016, p. 1135–1144. doi:10.1145/2939672.2939778.
- [12] R. Guidotti, A. Monreale, S. Ruggieri, D. Pedreschi, F. Turini, F. Giannotti, Local rule-based explanations of black box decision systems, *arXiv preprint arXiv:1805.10820* (2018).
- [13] M. Müller, M. Gromicho, M. de Carvalho, S. C. Madeira, Explainable models of disease progression in als: Learning from longitudinal clinical data with recurrent neural networks and deep model explanation, *Computer Methods and Programs in Biomedicine Update* 1 (2021) 100018.
- [14] A. Holzinger, C. Biemann, C. S. Pattichis, D. B. Kell, What do we need to build explainable ai systems for the medical domain?, *arXiv preprint arXiv:1712.09923* (2017).
- [15] B. Wollschlaeger, E. Eichenberg, K. Kabitzsch, Explain yourself: A semantic annotation

- framework to facilitate tagging of semantic information in health smart homes., in: HEALTHINF, 2020, pp. 133–144.
- [16] I. Tiddi, et al., Directions for explainable knowledge-enabled systems, Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges 47 (2020) 245.
- [17] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, G.-Z. Yang, XAI - explainable artificial intelligence, Science Robotics 4 (2019) eaay7120. doi:10.1126/scirobotics.aay7120.
- [18] F. Lécué, On the role of knowledge graphs in explainable ai, Semantic Web 11 (2019) 41–51.
- [19] S. Staab, R. Studer, Handbook on ontologies, Springer-Verlag, Berlin Heidelberg, 2010. doi:10.1007/978-3-540-92673-3.
- [20] M. Bettin, G. M. Di Nunzio, D. Dosso, G. Faggioli, N. Ferro, N. Marchetti, G. Silvello, Deliverable 9.1 – Project ontology and terminology, including data mapper and RDF graph builder, BRAINTEASER, EU Horizon 2020, Contract N. GA101017598. <https://brainteaser.health/>, 2021.
- [21] D. Wang, Q. Yang, A. Abdul, B. Y. Lim, Designing theory-driven user-centric explainable ai, in: Proceedings of the 2019 CHI conference on human factors in computing systems, 2019, pp. 1–15.
- [22] D. Dumas, P. A. Alexander, L. M. Baker, S. Jablansky, K. N. Dunbar, Relational reasoning in medical education: Patterns in discourse and diagnosis., Journal of Educational Psychology 106 (2014) 1021.
- [23] C. Pesquita, Towards semantic integration for explainable artificial intelligence in the biomedical domain., in: HEALTHINF, 2021, pp. 747–753.
- [24] J. Jovanović, E. Bagheri, Semantic annotation in biomedicine: the current landscape, Journal of biomedical semantics 8 (2017) 1–18.
- [25] L. Ehrlinger, W. Wöß, Towards a definition of knowledge graphs., SEMANTiCS (Posters, Demos, SuCCESS) 48 (2016) 1–4.
- [26] S. de Coronado, L. W. Wright, G. Fragoso, M. W. Haber, E. A. Hahn-Dantona, F. W. Hartel, S. L. Quan, T. Safran, N. Thomas, L. Whiteman, The nci thesaurus quality assurance life cycle, Journal of biomedical informatics 42 (2009) 530–539.
- [27] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models, ACM computing surveys (CSUR) 51 (2018) 1–42.
- [28] C. Jonquet, N. Shah, C. Youn, M. Musen, C. Callendar, M.-A. Storey, Ncbo annotator: Semantic annotation of biomedical data, ISWC (2009).
- [29] A. Tchechmedjiev, A. Abdaoui, V. Emonet, S. Melzi, J. Jonnagaddala, C. Jonquet, Enhanced functionalities for annotating and indexing clinical text with the NCBO Annotator+, Bioinformatics 34 (2018) 1962–1965. URL: <https://doi.org/10.1093/bioinformatics/bty009>. doi:10.1093/bioinformatics/bty009.
- [30] C. Pesquita, D. Faria, A. O. Falcão, P. Lord, F. M. Couto, Semantic similarity in biomedical ontologies, PLOS Computational Biology 5 (2009) 1–12. doi:10.1371/journal.pcbi.1000443.
- [31] P. Resnik, Using information content to evaluate semantic similarity in a taxonomy, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1, IJCAI'95, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995, p. 448–453.

- [32] Q. Wang, Z. Mao, B. Wang, L. Guo, Knowledge graph embedding: A survey of approaches and applications, *IEEE Transactions on Knowledge and Data Engineering* 29 (2017) 2724–2743. doi:10.1109/TKDE.2017.2754499.