Survival Analysis for Multiple Sclerosis: Predicting Risk of Disease Worsening

Notebook for the iDPP Lab on Intelligent Disease Progression Prediction at CLEF 2023

Ruben Branco*, Joana Barros Valente, Andreia S. Martins, Diogo F. Soares, Eduardo N. Castanho, Sara C. Madeira and Helena Aidos

LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

Abstract

Multiple sclerosis (MS) is a chronic neurodegenerative disease with a wide range of clinical manifestations and disease courses. Prognosis prediction is therefore an important tool for clinical decision-making and treatment administration. As proposed by the iDPP ⋰ CLEF 2023 challenge, we have explored several survival prediction models to rank MS patients according to the risk of worsening. Two definitions of disease worsening were explored, with subtask (a) using a fixed EDSS threshold and subtask (b) considering a baseline-dependent threshold. The models were trained using demographic and clinical data collected at diagnosis and available EDSS and relapse follow-up records. The temporal nature of the latter was addressed by computing feature statistics. Model performance was evaluated through c-index. In Task 1a, the Random Survival Forest model reached 0.801 (95% CI [0.678; 0.924]) in the private test set, while in Task 1b, the Fast Kernel SVM model reached 0.690 (95% CI [0.591; 0.788]).

Keywords

multiple sclerosis, disease prognosis, survival analysis

1. Introduction

Multiple sclerosis (MS) is a chronic autoimmune disease of the central nervous system, causing progressive impairment of motor and cognitive functions over several years [1]. It usually presents in young adults (20 - 40 years), is more prevalent in women, and its course, clinical manifestations, and treatment responses vary across patients [1]. Most MS patients (85%-90%) are diagnosed with relapsing-remitting MS (RRMS), a disease course characterized by the occurrence of relapses, followed by periods of total or partial recovery [2]. Eventually, these patients may shift to secondary-progressive MS (SPMS), experiencing periods of progression that may include relapses and periods of relatively stable disability [2]. Finally, the remaining 10-15% MS patients are diagnosed with primary-progressive MS (PPMS), where there is steady neurologic worsening from the onset [2]. While patients suffer from disabilities such as fatigue, urinary incontinence, vision loss, impairment of coordination, spasticity, muscle weakness, and

---

CEUR Workshop Proceedings (CEUR-WS.org)

cognitive impairment, respiratory failure is the most common cause of death in MS [3, 4]

Since there is no known cure for MS, treatments focus on retarding the natural progression of the disease and improving the patient's quality of life [5]. Several factors add to the disease heterogeneity, such as genetics [6, 7], sex, lifestyle, and environmental [8]. Machine learning techniques are used to understand disease mechanisms and assist in clinical decision-making [9]. Such methods have been applied to the identification of biological markers for both diagnosis and prognosis [9, 10], discovering patients with similar disease progression courses [11].

The iDPP ⠒ CLEF 2023 challenge[1] uses highly curated data from real MS and ALS patients, followed at clinical institutions in Lisbon, Portugal, Turin and Pavia, Italy. The focus of this challenge is the prognostic prediction in both MS and ALS [12].

The objective of task 1 is to rank MS patients based on the risk of worsening, setting the problem as a survival analysis task. Mainly, this task focuses on the worsening as measured by the Expanded Disability Status Scale (EDSS) [13], with two possible definitions of worsening, according to clinical standards: one based on the times a patient crosses a given threshold of the EDSS (task 1a), and the other based on the baseline EDSS evaluated and the occurrence of an increase of EDSS by a certain amount of points (task 1b).

In our contribution to iDPP, we relied on a methodology based on data preprocessing to handle the temporal nature of patient data. Data selection was used to filter out variables with a high percentage of missing values. From the original features, we computed statistical measures to capture the temporal progression of the disease. We then used three types of survival analysis models from the state-of-the-art literature in survival analysis: Tree-based, SVM-based, and Cox Proportional-Hazards Models.

The paper is organized as follows: Section 2 introduces related works; Section 3 describes our approach and the experimental setup; Section 4 discusses our main findings; finally, Section 5 draws some conclusions.

2. Related Work

Multiple sclerosis (MS) is a chronic, inflammatory, demyelinating, and neurodegenerative disease of the central nervous system (CNS) [1]. While the underlying causes of MS are still unknown, existing evidence suggests that the combination and interaction of some factors, such as genetics, environment, and lifestyle, contribute to disease susceptibility and development [1].

Thus, machine learning (ML) is being increasingly explored in MS research. For instance, some recent studies focused on developing diagnostic models by implementing supervised learning techniques to distinguish MS individuals from healthy ones [14] and to classify MS patients into one of the MS types [15, 16, 17, 18]. Furthermore, the application of ML models in MS extends to prognostic analyses where studies focus on disease course predictions in the short-to-medium-term [19, 20, 9], and long-term [21]. The prediction (dependent) variables analyzed include measures of disability (EDSS change [19], EDSS score [21, 22], an indicator of disability progression [19], the occurrence of relapses [19]), and transition from RRMS to SPMS [20, 9]. Regarding the predictors (independent variables), these studies considered demographic and clinical variables (e.g., gender, age, EDSS at study entry, motor function score, disease

---

[1] http://brainteaser.dei.unipd.it/challenges/idpp2023/

type, relapses) [19, 20, 9, 21, 22], MRI variables [19, 9], and neurophysiological variables (e.g., motor evoked potential scores (MEP), central motor conduction time) [19, 21]. The results found suggest that EDSS at study entry [19], MEP variables [19], changes in Timed 25-Foot Walk (T25FWT) [21], early changes in EDSS [21], age [22], carbonyl proteins [22], and inflammatory indicators [22] help to predict the EDSS score/ EDSS change, and that patients' historical clinical profiles improve the ability to predict a change from RRMS to SPMS [9].

Statistical learning (e.g., survival analysis) has also been applied to MS research to identify the time to occurrence of some combined outcomes (such as thresholded EDSS, relapse occurrence, white matter lesions, cortical lesions, cognitive status impairment) [23, 24], the transition from benign MS to no-longer benign MS [25], the time to SPMS conversion [26, 27, 28], or time to death [29]. In particular, in [29], the authors develop an immune-associated gene risk score related to overall survival using a Cox proportional hazards model. Cox proportional hazards model has also been used in [24] to predict time to disability progression over two years by relying on several clinical measures (e.g., EDSS, T25FWT), in [28] to develop a survival score to identify groups of patients with different risk of conversion to SPMS, and in [27] to predict the risk of conversion to SPMS within 10, 15, 20 years of RRMS onset. Non-proportional hazard Cox regression has been proposed to predict the time to evidence of disease activity [23], defined as a composite outcome including EDSS progression, relapse occurrence, white matter lesions, Gad+ lesions, cortical lesions, and cognitive status. In addition, in [25], a multivariate Cox model was used to develop an incremental score indicating whether a patient will evolve as benign MS or no-longer benign MS, and in [30], to obtain the probability of one, two, and three years relapse-free survival. Finally, random survival forests has been used to identify the most predictive factors associated with the risk of SPMS conversion adjusted for therapy [26].

3. Methodology

As noted in the introduction, we set out to contribute to Task 1 of the iDPP ⁖ CLEF 2023 challenge. Task 1 regards predicting the risk of disease worsening in MS patients. More formally, the objective is to build a system that, given data regarding the patient, assesses how early it is likely that the patient experiences a worsening event, framed as a survival task. The likelihood of worsening is measured from 0 to 1.

The worsening of a patient is defined using the EDSS, according to clinical standards, and following two definitions for two sub-tasks:

- **Task 1a** – the patients EDSS value crosses the threshold of 3 at least twice within a one-year interval;
- **Task 1b** – considering the patients' first recorded EDSS value as baseline, according to clinical protocols, worsening happens when:
    - baseline < 1 and an increase of 1.5 points is first observed;
    - $1 \leq$ baseline < 5.5 and an increase of 1 point is first observed;
    - baseline ≥ 5.5 and an increase of 0.5 points is first observed.

Our methodology employs trial-and-tested survival analysis methods, with a data preprocessing pipeline to handle the temporal nature of the patient data (Section 3.1).

Table 1
Selected Features

Component	Feature	Type	Computed?
Static	Sex	Binary	
	Residence	Categorical	
	Age at Onset	Integer	
	Diagnostic Delay	Float	
	Spinal Cord Symptom	Boolean	
	Brainstem Symptom	Boolean	
	Eye Symptom	Boolean	
	Supratentoral Symptom	Boolean	
	Time since Onset	Integer	
EDSS	EDSS as evaluated by clinician:	Float	
	First assessement		✓
	Last assessment		✓
	Max value of assessments		✓
	Mean value of assessments		✓
	Min value of assessments		✓
	Std values of assessments		✓
	EDSS change		
Relapses	Delta time between Time 0 and Last relapse	Float	✓
	Relapse Count	Float	✓

3.1. Data Preprocessing

The dataset made available with this challenge comprises information about MS patients and contains static and dynamic components. Static data is regarding the demographical characterization of patients, and the dynamic comprises two years and a half of health records, containing information on: relapses, EDSS scores, evoked potentials, MRIs, and MS course [12]. Data were preprocessed to have the characteristics needed by the survival models. Hence, we evaluated the shape of data (the dataset was provided in several files which were processed to meet the requirements for the models), the number of missing values, and the distributions of each feature. Therefore, we selected a set of features according to this previous analysis. From these selected features, we computed some others to include as the temporal understanding of the disease progression since we could not use the clinical records directly. The features chosen for our models are documented in Table 1. The computed features are regarding EDSS scores and Relapses frequency and are marked in the table. We use the values of the first and last, maximum, minimum, mean, and standard deviation of the EDSS assessments. Also regarding EDSS, we created a variable with the change (slope) between assessments. Regarding relapses, we computed a feature with the delta time from the time 0 and the last record relapse.

3.2. Modeling

Three types of models were used: Tree-based (Random Survival Forest and Gradient Boosting), SVM-based (Survival SVM), and Cox Proportional-Hazards Model. In total, we experimented with five models. We used the implementations provided by the Python package `sksurv` [31].
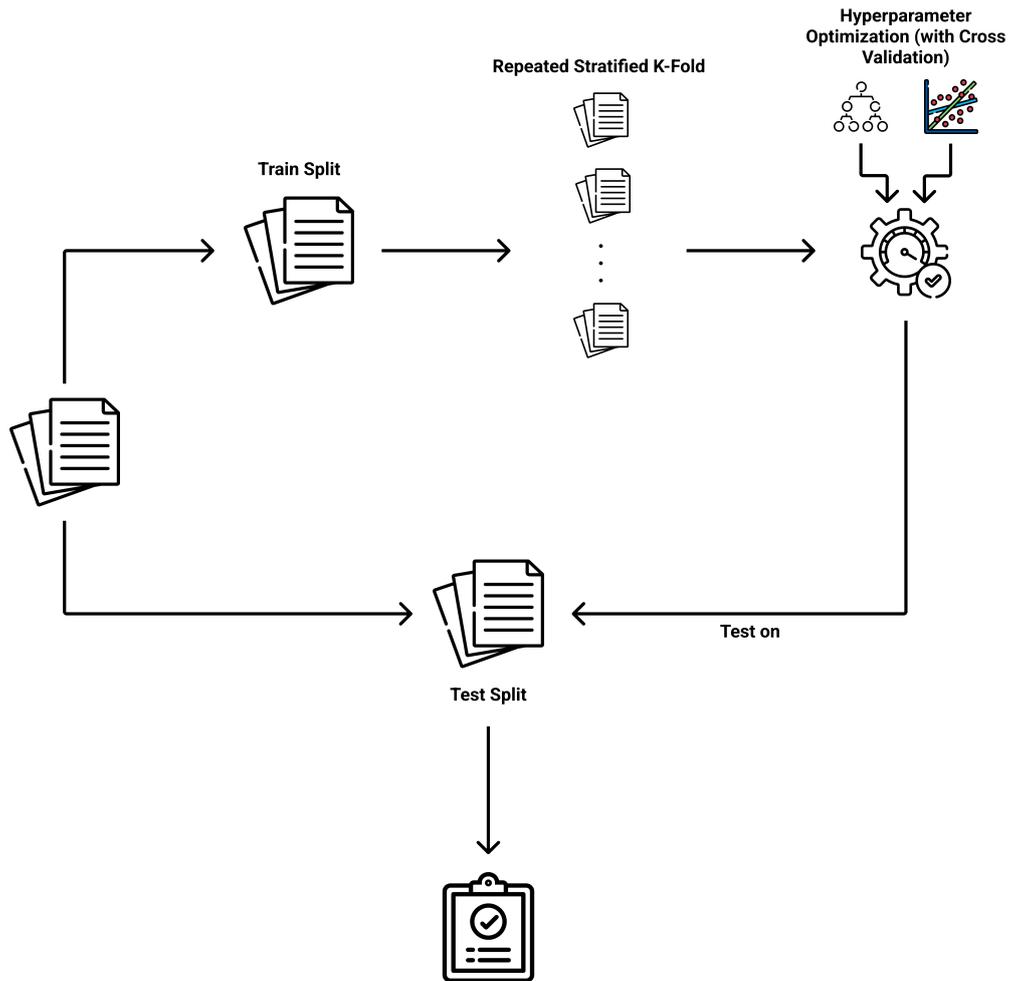
Figure 1: Pipeline for data splitting, optimization, and model selection.

Our pipeline consists of 3 stages, as pictured in Figure 1:

1. Train/test split (90% train, 10% test);
2. Hyperparameter Optimization using Repeated Stratified K-Fold Cross Validation;
3. Test the best classifier for each model type on the same test split. The resulting scores were used to select the most promising classifiers to submit to the challenge.

`Optuna` [32] was used for hyperparameter optimization, allowing the hyperparameter search to be guided through an optimization task rather than a brute-force approach. For reproducibility, Table 2 shows the hyperparameter search space defined for the experiments. The metric used for hyperparameter optimization is the same evaluation metric for the Task, Harrell's Concordance Index metric [33] (c-index).

Each model was optimized and tested on the same test set split, using the c-index metric, and the resulting scores were used to choose the appropriate models to submit to the challenge.

Table 2
Hyperparameter space for each model. Int and Float Distributions describe a search space between two integers or floating points, while CategoricalDistribution describes a search space between a set of values. Trials are the number of hyperparameter configurations to be tested.

Model	Hyperparameter	Distribution Space	Trials
CoxPHSurvivalAnalysis	alpha	FloatDistribution(0, 5)	200
RandomSurvivalForest	n_estimators	IntDistribution(10, 1000)	50
FastSurvivalSVM(max_iter=100000)	alpha	FloatDistribution(0, 5)	100
FastKernelSurvivalSVM(max_iter=100000)	alpha kernel	FloatDistribution(0, 5) CategoricalDistribution(["linear", "rbf", "poly"])	100
GradientBoostingSurvivalAnalysis	n_estimators learning_rate max_depth	IntDistribution(10, 1000) FloatDistribution(0.1, 1) IntDistribution(1, 10)	100

We conducted experiments using the datasets made available by the organization for Task 1 [34, 35]. All the experiments were run on a Desktop Computer with an AMD Ryzen 9 7950X 16-Core with 64GB of RAM and Ubuntu 22.04.2. The code was run using Python 3.10.11.

4. Results

In this section, we present the results obtained, along with a brief discussion of them.

We first ran a hyperparameter search, as described in Section 3, for each task to obtain the models we would submit to the challenge. Table 3 contains the hyperparameter settings for the models for Task 1a and 1b.

C-Index measures the discriminative ability of a given survival model, taking up values between 0 and 1, and as such, higher index values mean better predictive power. A value of 0.5 indicates a random predictive power, and if the confidence interval does not contain 0.5, then it means that with statistical significance, the model's predictive power is not random.

Table 4 contains the c-index scores for Task 1a, computed with a private test set by the iDPP organizers and in a subset of the training dataset. As can be observed, the confidence intervals for all the models are above 0.5, meaning all the models do have predictive capabilities. Furthermore, the c-index values obtained with the subset of the training data are slightly lower than the private test set provided by the iDPP organizers for all models, except for Gradient Boosting Survival Analysis method, which is quite lower in the training subset compared to the private test set. Overall, Random Survival Forest method is the best performing model in our subset of the training dataset, with a c-index of 0.847, and also in the private test set, obtaining 0.801 c-index, while for the remaining models the c-index range from 0.760 to 0.792.

Task 1b revealed to be a more complex challenge, as expected, due to the fact that the worsening event has three different definitions, which may introduce some difficulty for the

Table 3
Hyperparameter values for each model used in Task 1a (Value A column) and Task 1b (Value B column), found through hyperparameter optimization. *: indicates a floating point with additional decimal places.

Model	Hyperparameter	Value A	Value B
CoxPHSurvivalAnalysis	alpha	0.358*	0.006*
RandomSurvivalForest	n_estimators	353	995
FastSurvivalSVM(max_iter=100000)	alpha	0.012*	0.259*
FastKernelSurvivalSVM(max_iter=100000)	alpha	0.024*	1.393*
	kernel	poly	linear
GradientBoostingSurvivalAnalysis	n_estimators	796	503
	learning_rate	0.028*	0.940*
	max_depth	6	5

Table 4
Task 1a challenge results. C-Index score for each model is provided, along with a 95% confidence interval, in square brackets, for the test set. A higher C-Index score is better. The Train C-Index is obtained from a subset of the training dataset, while the Test C-Index is calculated from a private test set.

Model	Train C-Index	Test C-Index
CoxPHSurvivalAnalysis	0.775	0.790 [0.640; 0.941]
RandomSurvivalForest	**0.847**	**0.801** [0.678; 0.924]
FastSurvivalSVM	0.766	0.777 [0.626; 0.929]
FastKernelSVM	0.730	0.792 [0.649; 0.935]
GradientBoostingSurvivalAnalysis	0.604	0.760 [0.605; 0.916]

models to establish the relationship between the covariates (features) and event occurrence and time to the occurrence.

Table 5 contains the c-index scores for Task 1b. As can be seen, the c-index values are significantly lower when compared with Task 1a. Contrary to task 1a, the c-index values in the subset of the training dataset are higher than the private test set, with Random Survival Forest having the highest difference. Furthermore, Random Survival Forest and Gradient Boosting Survival Analysis contains 0.5 within their 95% confidence interval, meaning that their true predictive power could be, at its worse, random choice (despite obtaining 0.599 and 0.601, respectively). Overall, the Fast Kernel SVM is the best performing model in the subset of the training dataset, with a c-index of 0.727, and in the private test set, with a c-index of 0.69.

5. Conclusions

As no known cure for MS exists, the primary objective of treatments is to impede the progression of the disease to ensure a high standard of living for the patient. The objective of iDPP ⠕ CLEF

Table 5

Task 1b challenge results. C-Index score for each model is provided, along with a 95% confidence interval, in square brackets, for the test set. A higher C-Index score is better. The Train C-Index is obtained from a subset of the training dataset, while the Test C-Index is calculated from a private test set.

Model	Train C-Index	Test C-Index
CoxPHSurvivalAnalysis	0.713	0.683 [0.581; 0.786]
RandomSurvivalForest	0.697	0.599 [0.494; 0.704]
FastSurvivalSVM	0.716	0.677 [0.575; 0.780]
FastKernelSVM	**0.727**	**0.690** [0.591; 0.788]
GradientBoostingSurvivalAnalysis	0.590	0.601 [0.489; 0.713]

2023 challenge is to rank MS patients based on the risk of worsening, setting the problem as a survival analysis task.

Our methodology involved data preprocessing to handle the temporal nature of patient data by selecting relevant features and computing additional variables to capture the temporal progression of the disease. We used three types of survival analysis models: Tree-based, SVM-based, and Cox Proportional-Hazards Models.

For Task 1a, which aimed to identify patients whose EDSS value crosses a threshold within a one-year interval, all models exhibited predictive capabilities, with the Random Survival Forest achieving the highest c-index.

When considering Task 1b, which considered different definitions of disease worsening based on the patient's baseline EDSS value, Fast Kernel SVM model performed best. However, the c-index scores for Task 1b were lower compared to Task 1a. Task 1b introduced three definitions of worsening; therefore, it is expected for the algorithms to show difficulties modeling the relationship between the features, event occurrence, and time to the occurrence.

Acknowledgments

References

[1] M. Filippi, A. Bar-Or, F. Piehl, P. Preziosa, A. Solari, S. Vukusic, M. A. Rocca, Multiple sclerosis, Nature Reviews Disease Primers 4 (2018) 10–1038.

[2] S. Klineova, F. D. Lublin, Clinical course of multiple sclerosis, Cold Spring Harbor perspectives in medicine 8 (2018) a028928.

[3] K. Harding, F. Zhu, M. Alotaibi, T. Duggan, H. Tremlett, E. Kingwell, Multiple cause of death analysis in multiple sclerosis: A population-based study, Neurology 94 (2020) e820–e829.

[4] P. Ragonese, P. Aridon, G. Salemi, M. D'Amelio, G. Savettieri, Mortality in multiple sclerosis: a review, Eur. J. Neurol. 15 (2008) 123–127.

[5] H. Li, F. Hu, Y. Zhang, K. Li, Comparative efficacy and acceptability of disease-modifying therapies in patients with relapsing–remitting multiple sclerosis: a systematic review and network meta-analysis, Journal of Neurology 267 (2019) 3489–3498. doi:10.1007/s00415-019-09395-w.

[6] I. M. S. G. C. (IMSGC), Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis, Nat. Genet. 45 (2013) 1353–1360.

[7] International Multiple Sclerosis Genetics Consortium, Network-based multiple sclerosis pathway analysis with GWAS data from 15,000 cases and 30,000 controls, Am. J. Hum. Genet. 92 (2013) 854–865.

[8] T. Olsson, L. F. Barcellos, L. Alfredsson, Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis, Nature Reviews Neurology 13 (2017) 25–36. doi:10.1038/nrneurol.2016.187.

[9] R. Seccia, D. Gammelli, F. Dominici, S. Romano, A. C. Landi, M. Salvetti, A. Tacchella, A. Zaccaria, A. Crisanti, F. Grassi, L. Palagi, Considering patient clinical history impacts performance of machine learning models in predicting course of multiple sclerosis, PLOS ONE 15 (2020) e0230219. doi:10.1371/journal.pone.0230219.

[10] N. Aslam, I. U. Khan, A. Bashamakh, F. A. Alghool, M. Aboulnour, N. M. Alsuwayan, R. K. Alturaif, S. Brahimi, S. S. Aljameel, K. Al Ghamdi, Multiple sclerosis diagnosis using machine learning and deep learning: Challenges and opportunities, Sensors 22 (2022). doi:10.3390/s22207856.

[11] E. Kotelnikova, N. A. Kiani, E. Abad, E. H. Martinez-Lapiscina, M. Andorra, I. Zubizarreta, I. Pulido-Valdeolivas, I. Pertsovskaya, L. G. Alexopoulos, T. Olsson, R. Martin, F. Paul, J. Tegnér, J. Garcia-Ojalvo, P. Villoslada, Dynamics and heterogeneity of brain damage in multiple sclerosis, PLOS Computational Biology 13 (2017) e1005757. doi:10.1371/journal.pcbi.1005757.

[12] H. Aidos, R. Bergamaschi, P. Cavalla, A. Chiò, A. Dagliati, B. Di Camillo, M. A. de Carvalho, N. Ferro, P. Fariselli, J. M. G. Dominguez, S. C. Madeira, E. Tavazzi, IDPP@CLEF 2023: The intelligent disease progression prediction challenge, in: Lecture Notes in Computer Science, Lecture notes in computer science, Springer Nature Switzerland, Cham, 2023, pp. 491–498.

[13] S. Meyer-Moock, Y.-S. Feng, M. Maeurer, F.-W. Dippel, T. Kohlmann, Systematic literature review and validity evaluation of the expanded disability status scale (EDSS) and the multiple sclerosis functional composite (MSFC) in patients with multiple sclerosis, BMC Neurology 14 (2014). doi:10.1186/1471-2377-14-58.

[14] L. Mezzaroba, A. N. C. Simão, S. R. Oliveira, T. Flauzino, D. F. Alfieri, W. L. de Carvalho Jennings Pereira, A. P. Kallaur, M. A. B. Lozovoy, D. R. Kaimen-Maciel, M. Maes, et al., Antioxidant and anti-inflammatory diagnostic biomarkers in multiple sclerosis: a machine learning study, Molecular Neurobiology 57 (2020) 2167–2178.

[15] S. Fiorini, A. Verri, A. Tacchino, M. Ponzio, G. Brichetto, A. Barla, A machine learning pipeline for multiple sclerosis course detection from clinical scales and patient reported outcomes, in: 2015 37th Annual International Conference of the IEEE engineering in medicine and biology society (EMBC), IEEE, 2015, pp. 4443–4446.

[16] A. Marzullo, G. Kocevar, C. Stamile, F. Durand-Dubief, G. Terracina, F. Calimeri, D. Sappey-Marinier, Classification of multiple sclerosis clinical profiles via graph convolutional neural networks, Frontiers in neuroscience 13 (2019) 594.

[17] A. Ion-Mărgineanu, G. Kocevar, C. Stamile, D. M. Sima, F. Durand-Dubief, S. Van Huffel, D. Sappey-Marinier, Machine learning approach for classifying multiple sclerosis courses by combining clinical data with lesion loads and magnetic resonance metabolic features, Frontiers in neuroscience 11 (2017) 398.

[18] Y. Karaca, C. Cattani, M. Moonis, Comparison of deep learning and support vector machine learning for subgroups of multiple sclerosis, in: Computational Science and Its Applications–ICCSA 2017: 17th International Conference, Trieste, Italy, July 3-6, 2017, Proceedings, Part II, Springer, 2017, pp. 142–153.

[19] B. Bejarano, M. Bianco, D. Gonzalez-Moron, J. Sepulcre, J. Goñi, J. Arcocha, O. Soto, U. D. Carro, G. Comi, L. Leocani, et al., Computational classifiers for predicting the short-term course of multiple sclerosis, BMC neurology 11 (2011) 1–9.

[20] A. Tacchella, S. Romano, M. Ferraldeschi, M. Salvetti, A. Zaccaria, A. Crisanti, F. Grassi, Collaboration between a human group and artificial intelligence can improve prediction of multiple sclerosis course: a proof-of-principle study, F1000Research 6 (2018) 2172.

[21] L. V. Bosma, J. J. Kragt, D. L. Knol, C. H. Polman, B. M. Uitdehaag, Clinical scales in progressive ms: predicting long-term disability, Multiple Sclerosis Journal 18 (2012) 345–350.

[22] T. Flauzino, A. N. C. Simão, W. L. de Carvalho Jennings Pereira, D. F. Alfieri, S. R. Oliveira, A. P. Kallaur, M. A. B. Lozovoy, D. R. Kaimen-Maciel, M. Maes, E. M. V. Reiche, Disability in multiple sclerosis is associated with age and inflammatory, metabolic and oxidative/nitrosative stress biomarkers: results of multivariate and machine learning procedures, Metabolic brain disease 34 (2019) 1401–1413.

[23] R. Magliozzi, A. Scalfari, A. I. Pisani, S. Ziccardi, D. Marastoni, F. B. Pizzini, A. Bajrami, A. Tamanti, M. Guandalini, S. Bonomi, et al., The csf profile linked to cortical damage predicts multiple sclerosis activity, Annals of neurology 88 (2020) 562–573.

[24] F. Pellegrini, M. Copetti, M. P. Sormani, F. Bovis, C. de Moor, T. P. Debray, B. C. Kieseier, Predicting disability progression in multiple sclerosis: Insights from advanced statistical modeling, Multiple Sclerosis Journal 26 (2020) 1828–1836.

[25] L. Razzolini, E. Portaccio, M. L. Stromillo, B. Goretti, C. Niccolai, L. Pastò, I. Righini, E. Prestipino, M. Battaglini, A. Giorgio, et al., The dilemma of benign multiple sclerosis: Can we predict the risk of losing the "benign status"? a 12-year follow-up study, Multiple sclerosis and related disorders 26 (2018) 71–73.

[26] A. I. Pisani, A. Scalfari, F. Crescenzo, C. Romualdi, M. Calabrese, A novel prognostic score to assess the risk of progression in relapsing- remitting multiple sclerosis patients, European journal of neurology 28 (2021) 2503–2512.

[27] A. Manouchehrinia, F. Zhu, D. Piani-Meier, M. Lange, D. G. Silva, R. Carruthers, A. Glaser, E. Kingwell, H. Tremlett, J. Hillert, Predicting risk of secondary progression in multiple sclerosis: a nomogram, Multiple Sclerosis Journal 25 (2019) 1102–1112.

[28] C. Vasconcelos, J. Aurenção, R. Alvarenga, L. Thuler, Long-term ms secondary progression: Derivation and validation of a clinical risk score, Clinical Neurology and Neurosurgery 194 (2020) 105792.

[29] F. Ye, X. Wu, T. Wang, J. Liang, J. Li, Y. Dai, K. Lan, W. Sheng, Identification of immune-associated gene signature and immune cell infiltration related to overall survival in progressive multiple sclerosis, Multiple Sclerosis and Related Disorders 55 (2021) 103188.

[30] F. Ye, J. Liang, J. Li, H. Li, W. Sheng, Development and validation of a five-gene signature to predict relapse-free survival in multiple sclerosis, Frontiers in Neurology 11 (2020) 579683.

[31] S. Pölsterl, scikit-survival: A library for time-to-event analysis built on top of scikit-learn, Journal of Machine Learning Research 21 (2020) 1–6.

[32] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019.

[33] F. E. Harrell, R. M. Califf, D. B. Pryor, K. L. Lee, R. A. Rosati, Evaluating the yield of medical tests, Jama 247 (1982) 2543–2546.

[34] G. Faggioli, A. Guazzo, S. Marchesin, L. Menotti, I. Trescato, H. Aidos, R. Bergamaschi, G. Birolo, P. Cavalla, A. Chiò, A. Dagliati, M. de Carvalho, G. M. Di Nunzio, P. Fariselli, J. M. García Dominguez, M. Gromicho, E. Longato, S. C. Madeira, U. Manera, G. Silvello, E. Tavazzi, E. Tavazzi, M. Vettoretti, B. Di Camillo, N. Ferro, Intelligent Disease Progression Prediction: Overview of iDPP@CLEF 2023, in: A. Arampatzis, E. Kanoulas, T. Tsikrika, S. Vrochidis, A. Giachanou, D. Li, A. Aliannejadi, M. Vlachos, G. Faggioli, N. Ferro (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), Springer, Heidelberg, Germany, 2023.

[35] G. Faggioli, A. Guazzo, S. Marchesin, L. Menotti, I. Trescato, H. Aidos, R. Bergamaschi, G. Birolo, P. Cavalla, A. Chiò, A. Dagliati, M. de Carvalho, G. M. Di Nunzio, P. Fariselli, J. M. García Dominguez, M. Gromicho, E. Longato, S. C. Madeira, U. Manera, G. Silvello, E. Tavazzi, E. Tavazzi, M. Vettoretti, B. Di Camillo, N. Ferro, Overview of iDPP@CLEF 2023: The Intelligent Disease Progression Prediction Challenge, in: M. Aliannejadi, G. Faggioli, N. Ferro, M. Vlachos (Eds.), CLEF 2023 Working Notes, CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073., 2023.