

THE DEVELOPMENT AND VALIDATION OF A DISABILITY AND OUTCOME PREDICTION ALGORITHM IN MULTIPLE SCLEROSIS PATIENTS

SIMONA OPREA¹, ANDREI VĂLEANU^{2*}, SIMONA NEGREȘ²

¹Bucharest Emergency University Hospital, 169 Splaiul Independenței, 050098 Bucharest, Romania

²“Carol Davila” University of Medicine and Pharmacy, Faculty of Pharmacy, Department of Pharmacology and Clinical Pharmacy, 6 Traian Vuia Street, 020956 Bucharest, Romania

*corresponding author: andrei_valeanu@yahoo.com

Manuscript received: January 2020

Abstract

At the moment, multiple sclerosis (MS) is considered one of the major disability factors among the young population. Considering the high prevalence and severity of this chronic disease, the aim of this study was to develop a disability and outcome prediction algorithm in MS patients. The data from two MS patient groups was analysed – Group A (151 patients with the following drug therapies: interferon beta-1a, glatiramer acetate, teriflunomide, natalizumab) and Group B (58 patients treated with natalizumab). Considering certain demographical and clinical predictive variables, as well as different disability threshold values, several prediction models were developed and validated, which are able to estimate the disability (Expanded Disability Status Scale, EDSS) and outcome probabilities. The prediction model validation on estimating the disability and outcome probabilities yielded a maximum Area Under the Receiver Operating Characteristic Curve (ROC AUC) Score of 80 - 82%. The overall results indicate that for Group A disability prediction based on EDSS is appropriate, while for Group B the outcome might be a better measure for probability estimation. Hence, the results suggest that the outcome is a more appropriate measure for monitoring natalizumab treated MS patients, since it depends on other clinical characteristics as opposed to the sole disability estimation through the EDSS score. A disability and outcome prediction model in MS patients was developed and validated. Despite the fact that the obtained results were satisfactory given the small dataset, by embedding more predictive demographical and clinical variables in the algorithm, as well as including more patients, the model could be used through an online platform in MS patients' monitoring and prioritization, in order to succeed in improving the quality of life and reducing disability progression.

Rezumat

La momentul actual, scleroza multiplă (SM) este considerată un factor major de dizabilitate în rândul populației tinere. Prin urmare, considerând prevalența și severitatea ridicată a acestei patologii, obiectivul prezentului studiu a fost reprezentat de dezvoltarea unui model de predicție a dizabilității și prognosticului pacienților cu scleroză multiplă. Au fost analizate datele provenite de la 2 grupuri de pacienți: grupul A (151 de pacienți cu următoarele terapii: interferon beta-1a, glatiramer acetat, teriflunomidă, natalizumab) și grupul B (58 de pacienți tratați cu natalizumab). Pe baza unor variabile predictive specifice (demografice și clinice), precum și prin considerarea mai multor valori prag de dizabilitate, au fost dezvoltați și validați mai mulți algoritmi de predicție, care au estimat dizabilitatea (scorul EDSS – Expanded Disability Status Scale) și prognosticul prin intermediul unui model probabilistic. Validarea modelelor de predicție din punctul de vedere al estimării dizabilității și prognosticului a condus la obținerea unui scor *ROC AUC* (Area Under the Receiver Operating Characteristic Curve) maxim de 80 - 82%. Rezultatele obținute au indicat faptul că pentru grupul A o predicție a dizabilității pe baza EDSS-lui este potrivită, în schimb ce în cazul grupului B prognosticul ar putea constitui o măsură mai bună pentru estimarea probabilității. Astfel, rezultatele sugerează că în cazul pacienților tratați cu natalizumab prognosticul este un parametru mai potrivit pentru monitorizarea pacienților suferind de SM, acesta depinzând de mai multe caracteristici clinice spre deosebire de simpla estimare a dizabilității prin intermediul scorului EDSS. În studiul prezentat, a fost dezvoltat și validat un model de predicție a dizabilității și prognosticului pacienților cu SM. În ciuda faptului că rezultatele au fost satisfăcătoare ținând cont de dimensiunea redusă a seturilor de date, pe viitor, prin includerea mai multor parametri demografici și clinici, precum și prin adăugarea mai multor pacienți în cadrul modelului de predicție, algoritmul ar putea fi utilizat la scară largă prin intermediul unei platforme online în vederea monitorizării și prioritizării pacienților cu SM. Aceste aspecte ar putea conduce pe termen lung la limitarea dizabilității și îmbunătățirea calității vieții pacienților suferind de această patologie.

Keywords: multiple sclerosis; disability scale; outcome; prediction model

Introduction

At the moment, multiple sclerosis (MS) is considered one of the major disability factors among the young

population. MS prevalence varies considerably, with a high frequency in Europe and North America (more than 100 cases *per* 100,000 inhabitants) and a lower

one in other regions [2, 10]. Due to the risk factors which contribute to the onset and exacerbation of MS, additional patient monitoring techniques are required, regardless of the MS type or of the recommended drug therapy [7, 13].

Machine learning techniques have been gaining more insight over the last 10 years, due to the fast development of artificial intelligence, as well as the increased prevalence of certain chronic diseases. These methods are based on algorithms of high complexity (such as neural networks or random forests), applied on a regular basis for disease diagnosis. Well known examples include the diagnosis of diabetic neuropathy and Alzheimer's disease by using image analysis [3, 5, 10, 14]. Regarding multiple sclerosis diagnosis, monitoring and management, such artificial intelligence techniques are mainly based on multiple sclerosis disease course classification based on central nervous system lesions [7, 21]. In addition, other studies used machine learning or statistical approaches for predicting the treatment response or the EDSS (Expanded Disability Status Scale) score evolution [17, 22], while others computed a more sensitive and complex disability scale, namely the Combinatorial Weight-Adjusted Disability Score, in order to perform relevant simulations of MS disease progression [20]. Therefore, considering the high prevalence and severity of this chronic disease, as well as the fact that until now only few machine learning methods for MS patients monitoring were developed, the aim of this study was to develop a disability and outcome prediction algorithm in MS patients.

Materials and Methods

Study design

In order to develop a clinically relevant prediction model, the patients which needed to be included in the study had to be carefully selected. Hence, two patient groups were selected and the data were independently analysed for the two groups. The patients were monitored in the Neurology Department of the Bucharest Emergency University Hospital (BEUH), Romania according to the updated guidelines regarding MS diagnosis and treatment [12, 16]. The study was validated by the Hospital Ethics Local Committee, and an informed consent was obtained from all study participants. The procedures were in accordance with the ethical standards of the responsible committee on human experimentation and with the Declaration of Helsinki.

Patients with different types of therapy (Group A). 151 patients with either primary progressive or relapsing-remitting MS and with at least 2 neurological evaluations in the last 2 years were included in this group. The primary progressive MS form was not separated from the relapsing-remitting MS form due to the fact that the MS disease type was not mentioned for all the

patients who were included in the study. In addition, the small number of patients which were selected (151) would have made the separate training of the model on two different subgroups (patients with primary progressive and relapsing-remitting MS form) a difficult task and due to the missing data, clinically irrelevant. For each patient, several demographical and clinical variables were considered: gender, age at diagnosis, age, Expanded Disability Status Scale (EDSS) score at disease onset ($EDSS_{onset}$), disease duration, number of treatments and the current EDSS score ($EDSS_{present}$). The group consisted of 61.59% females and 38.41% males. The mean age at diagnosis was 30.07 years (standard deviation (SD) = 9.3) and the mean age among the patients was 40.27 years (SD = 10.68). The mean disease duration was 10.2 years. The patients received the following therapies: interferon beta-1a (46 patients – 30.46%), glatiramer acetate (45 patients – 29.8%), teriflunomide (3 patients – 1.99%), natalizumab (56 patients – 37.09%), unknown (1 patient – 0.66%).

Patients treated with natalizumab (Group B). The second group of patients which were included in the study consisted of all patients treated with natalizumab at BEUH and part of the Multiple Sclerosis National Programme. There were 58 patients included in the study, 58.62% females and 41.38% males. The mean age at diagnosis was 26.38 years (standard deviation (SD) = 7.78) and the mean age among the patients was 35.33 years (SD = 9.21). The analysed variables were gender, age at diagnosis, age, $EDSS_{onset}$, natalizumab treatment duration, number of treatments before natalizumab, $EDSS_{present}$ and outcome (disease prognosis – marked by the neurologist as good or bad based on multiple criteria). The mean natalizumab treatment duration was 3.38 years.

The development of the disability and outcome prediction model

All the specified variables, both for Group A and Group B were used for the development of the disability and outcome prediction model. There were several models which were independently built for each group. The algorithm development and validation were performed by using Python programming language, version 3.5.2. Logistic regression was used as machine learning algorithm, due to its ability of performing well on small datasets and generating good calibrated probabilities. Numpy, Scipy and Scikit-learn libraries were used for data preprocessing and machine learning development [15]. For the patients belonging to Group B, both a disability and an outcome prediction model were built (based on $EDSS_{present}$ and Outcome), while for the other group, only a disability prediction model was developed (based on $EDSS_{present}$). The difference between the two patient groups regarding the prediction algorithms was due to the fact that in Group B all patients were treated with the same drug (natalizumab), which allowed the neurologist to perform a relevant and uniform analysis related to the disease prognosis.

All the predictive models were based on a probability estimation.

In order to properly build the disability prediction model, for each group several classes were created, based on certain EDSS_{onset} threshold values (EDSS_{threshold}). Several threshold values were set for Group A and Group B. The patients with EDSS_{onset} ≤ EDSS_{threshold} were included in class 1 and the patients with EDSS_{onset} > EDSS_{threshold} belonged to class 0. Hence, a disability prediction model was developed for each patient group and threshold value and estimated the probability that a certain patient will keep his EDSS below a certain value after a number of years since diagnosis (the probability of belonging to class 1 (of maintaining EDSS_{present} below EDSS_{threshold})). Therefore,

the higher the probability is, the higher the chances are that the MS patient will have a good evolution in time. It is worth mentioning that the algorithm predicts a non-disability probability. The EDSS threshold values were 2.5, 3.5, 4, 4.5 and 5 for Group A and 2.5, 4 and 5 for Group B.

The outcome prediction model was developed in a similar manner, only that the two classes (0 and 1) were created based on the disease prognosis set by the neurologist (class 0 corresponded to a bad prognosis, while class 1 corresponded to a good prognosis).

Table I presents the input (predictive) variables (which were used by the logistic regression algorithm for estimating the probabilities), as well as the output (predicted) variables for both patient groups.

Table I

The input and predicted variables used by the logistic regression algorithm

Group A (151 patients)	
Input (predictive) variables	Output
Gender	<ul style="list-style-type: none"> The probability of maintaining EDSS_{present} under the threshold value (EDSS_{threshold} = 2.5, 3.5, 4, 4.5 and 5) (the probability of belonging to class 1)
Age at diagnosis	
Age	
EDSS _{onset}	
Disease duration (years)	
Number of treatments	
Group B (58 natalizumab treated patients)	
Input (predictive) variables	Output
Gender	<ul style="list-style-type: none"> The probability of maintaining EDSS_{present} under the threshold value (EDSS_{threshold} = 2.5, 4 and 5) The probability of a good outcome (disease prognosis) (the probability of belonging to class 1 in terms of outcome)
Age at diagnosis	
Age	
EDSS _{onset}	
Natalizumab treatment duration (years)	
Number of treatments before natalizumab	

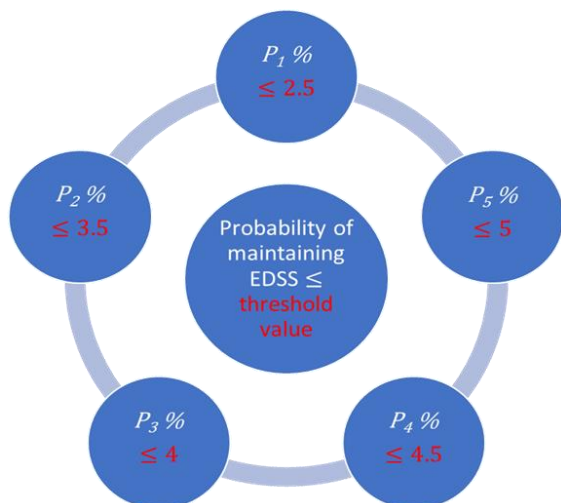


Figure 1.

The estimated probabilities of maintaining patient EDSS below a threshold value (Group A – 151 patients) Usually, $P_1 < P_2 < P_3 < P_4 < P_5$, since maintaining EDSS below smaller threshold values is more difficult (for example, 4.5 as opposed to 5)

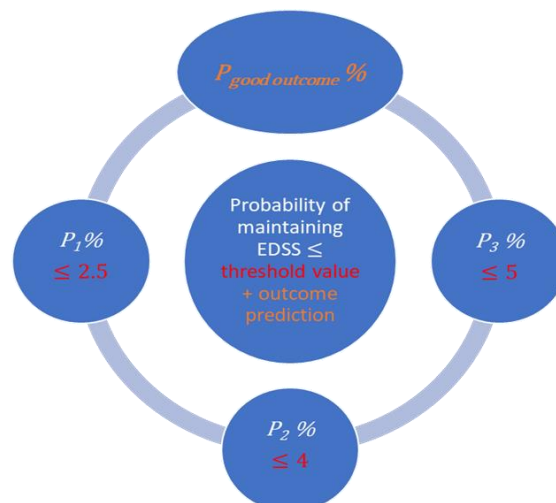


Figure 2.

The estimated probabilities of maintaining patient EDSS below a threshold value (Group B – 58 patients) Usually, $P_1 < P_2 < P_3$. $P_{good\ outcome}$ cannot be included in this relationship, since it reflects other clinical characteristics of the patient and additional considerations of the neurologist beside the EDSS score

Figures 1 and 2 offer a simplified visualization of the estimated probabilities for all the algorithms (Group A – Figure 1, Group B – Figure 2). The estimated probabilities can be regarded as MS patients’ disability and outcome profiles.

Prediction model validation

A very important phase in the study was the regression model validation and was performed in order to evaluate the predictive ability of the algorithm. The validation was performed in a similar manner for both patient groups. The dataset was split (for each patient group and EDSS threshold value) in a training set and a test set by using a Monte Carlo cross validation technique. The algorithm was built based on the training set and its predictive ability was afterwards evaluated on the test set (not available data). The ratio between the training and the test set was of 14:1 for Group A and of 9:1 for Group B, due to the smaller number of patients belonging to the latter group [6].

For a complete evaluation, both the predictive ability of including the patients in one of the classes (0 or 1)

(the classification task) and of estimating the probability of belonging to class 1 were tested. Class 0 was considered the negative class and class 1 the positive one. The classification was evaluated based on accuracy, precision, recall and F1 score and the probability estimation was tested by computing the Brier Score and the Area Under the Receiver Characteristic Operating Curve (ROC AUC) Score. The classification statistical parameters evaluated the percentage of correctly classified patients (accuracy), the false positives (FP) (precision), or the false negatives (FN) (recall). The most important validation measure was the ROC AUC Score, since it quantified the ability of the model of estimating well calibrated probabilities, but also its ability of discriminating between the two classes [4, 8, 15, 18]. Table II presents the details regarding the computation of the statistical measures used for the regression logistic validation. Figure 3 presents a scheme of the phases involved in the validation process.

Table II

Computation of the validation measures and their optimal values [15]

Validation measure	Formula	Range	Optimal value
Accuracy	$\frac{\text{Number of correctly classified}}{N}$	0 - 100%	100%
Precision	$\frac{TP}{TP + FP}$	0 - 100%	100%
Recall	$\frac{TP}{TP + FN}$	0 - 100%	100%
F1 score	$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	0 - 100%	100%
Brier score	$\frac{1}{N} \times \sum_1^N (\text{Pr}(1) - \text{Class})^2*$	0 - 1	0
ROC AUC score	Based on the TP rate (recall) vs. FP rate graphical representation	0 - 100%	100%

*N = the total number of patients from the test set; Pr (1) = the predicted probability of belonging to class 1; Class = the true class (0 or 1); TP = true positives; FP = false positives; FN = false negatives.

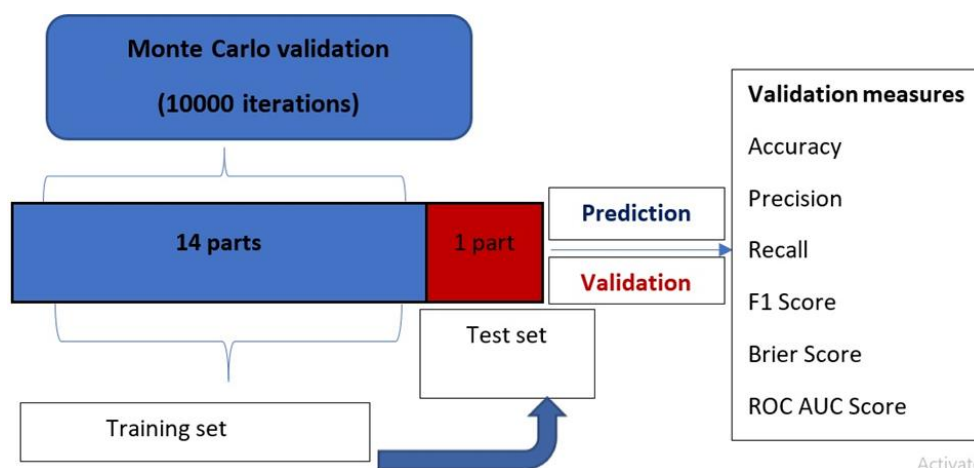


Figure 3.

The scheme of the validation process for Group A. The validation was performed for Group B in a similar manner, with two differences: a 9:1 ratio between the training set and the test set and 5000 iterations in the Monte Carlo cross validation

Results and Discussion

Validation results

Tables III and IV present the results obtained during the prediction model validation process. The statistical measures are given for each pair patient group – EDSS_{threshold}.

The output data generated by the prediction models – examples on real patients

For a better understanding of the clinical relevance of the prediction models, Tables V and VI present the results regarding the disability and outcome prediction for two patients belonging to the analysed groups. In order to perform the disability and outcome estimation, the regression logistic algorithm was trained on the rest of the patients for each case and was afterwards applied for the probability estimation on the selected patients. In order to allow a relevant comparison, two patients with similar characteristics regarding age, gender and drug therapy were selected.

Regarding the prediction models which were developed in this study, Tables III and IV present the validation

process results. Due to the slightly unbalanced dataset (the positive class being found in a higher proportion in the majority of the cases), the most useful statistical measure for estimating the classification ability of the model was the F1 Score, a weighted average between precision and recall, who took into account both the FPs and the FNs [15]. The F1 Score values varied in the range of 78.06% and 84.48% for Group A and in the range of 58.55% and 70.31% for Group B. In both patient groups, the F1 Score reached its maximum value for the disability prediction with an EDSS_{threshold} = 5 (84.48% – Group A, 70.31% – Group B). The enhanced results obtained for the classification task for the EDSS threshold values of 4.5 and 5 as opposed to 3.5 and 4 can be explained by the fact that in this situations a better differentiation between the two classes can be outlined (the patients with an EDSS in the range of 3.5 and 4.5 could be quite similar in terms of clinical characteristics to those with an EDSS below 3.5, which makes a prediction for threshold values of 3.5 and 4 quite difficult) [19].

Table III

The validation results of the disability prediction model for Group A (151 patients)

Validation measure	EDSS ≤ 2.5	EDSS ≤ 3.5	EDSS ≤ 4	EDSS ≤ 4.5	EDSS ≤ 5
Accuracy	76.62%	72.25%	73.77%	77%	76.74%
Precision	81.45%	85.38%	89.76%	92.98%	97.71%
Recall	77.75%	74.17%	75.59%	77.71%	75.86%
F1 Score	78.06%	78.21%	80.96%	83.73%	84.48%
Brier Score	0.1754	0.2016	0.2029	0.1847	0.1838
ROC AUC Score	82.21%	77.3%	76.93%	80.9%	80.22%

Table IV

The validation results of the disability and outcome prediction model for Group A (58 natalizumab treated patients)

Validation measure	EDSS ≤ 2.5	EDSS ≤ 4	EDSS ≤ 5	Outcome prediction
Accuracy ¹	63.67%	55%	61.51%	72.53%
Precision ¹	62.75%	80.12%	84.71%	66.95%
Recall ¹	60.08%	57.03%	63.57%	69.3%
F1 Score ¹	58.55%	64.22%	70.31%	64.88%
Brier Score ²	0.2392	0.2681	0.2516	0.1884
ROC AUC Score* ²	66.67%	54.89%	59.18%	80.29%

* Most important validation measure – it evaluated the model’s ability of estimating well calibrated probabilities [15];

¹ Classification validation measures; ² Probability prediction validation measures

Table V

The disability prediction results for a patient treated with glatiramer acetate (first line therapy) (Group A)

Patient characteristics	
Gender	Female
Age at diagnosis	33
Age	47
EDSS _{onset}	1.5
EDSS _{present}	2.5
Disease duration (years)	14
Number of treatments	1
Predicted results (Output data)	
Probability (EDSS ≤ 2.5)	59.69%
Probability (EDSS ≤ 3.5)	68.62%
Probability (EDSS ≤ 4)	75.97%
Probability (EDSS ≤ 4.5)	83%
Probability (EDSS ≤ 5)	86.63%

Table VI

The disability and outcome prediction results for a patient treated with natalizumab (second line therapy) (Group B)

Patient characteristics	
Gender	Female
Age at diagnosis	31
Age	46
EDSS _{onset}	1.5
EDSS _{present}	2
Natalizumab treatment duration (years)	9
Number of treatments before natalizumab	0
Good outcome?	NO
Predicted results (Output data)	
Probability (EDSS ≤ 2.5)	72.41%
Probability (EDSS ≤ 4)	78.54%
Probability (EDSS ≤ 5)	79.55%
Probability – good outcome	86.8% (incorrect prediction)

However, the most important validation measure of the prediction models was the ROC AUC Score, which quantified the ability of the logistic regression algorithms of estimating well calibrated probabilities. More specific, the higher the ROC AUC Score is, the more precise the predicted probabilities are (personalized for each patient depending of his/her characteristics) [18]. The ROC AUC Score values varied in the range of 77.3 and 82.21% for Group A (151 patients) and between 59.18 and 80.29% for Group B (58 natalizumab treated patients). In both cases, the maximum value of the ROC AUC Score in terms of disability estimation was obtained for $EDSS_{\text{threshold}} = 2.5$ (82.21% – Group A, 66.67% – Group B), very probable due to the relatively equal distribution of the dataset between the positive and negative class which was obtained for this threshold value. Nevertheless, in terms of outcome prediction (realized only for Group B), the ROC AUC Score reached higher values (80.29%) than those obtained for the disability prediction in the patients treated with natalizumab (59.18% - 66.67%). The optimized results which were obtained for the outcome prediction suggest that the outcome might represent a more appropriate measure for probability estimation than the EDSS threshold value, being also a more appropriate evaluation measure for natalizumab treated MS patients as opposed to the quantification of the disability degree by using the EDSS [19].

Tables V and VI reflect the clinical relevance of the algorithms, by outlining the results obtained on two real MS patients included in the present study. Both patients had a good evolution in terms of disability (EDSS), which was also positively predicted by the logistic regression algorithm. The model accurately predicted the probabilities of maintaining the EDSS score below the selected threshold values (2.5, 3.5, 4, 4.5 and 5 for the first patient and 2.5, 4 and 5 for the second patient). More exactly, for the glatiramer treated patient, the predicted probabilities varied in the range of 59.69% ($EDSS_{\text{threshold}} = 2.5$) and 86.62% ($EDSS_{\text{threshold}} = 5$) and were considered accurate,

since the patient's EDSS score evolved from 1.5 at diagnosis ($EDSS_{\text{onset}}$) to 2.5 at the present moment ($EDSS_{\text{present}}$). Secondly, the differences between the 5 predicted probabilities were also relevant: the higher the threshold value was, the higher the predicted probability was. These differences were due to the fact that maintaining the EDSS score below a lower value (such as 2.5 - 59.69%) is more difficult than maintaining it below a higher value (such as 3.5 - 68.62%). In addition, the differences are consistent with the fact MS is a progressive, debilitating disease, the disability degree of the patients usually evolving over time [1]. These results prove the ability of the model of performing a good differentiation depending on different EDSS threshold values, which might allow not only a good estimation of non-disability probabilities, but also a relevant stratification and prioritization of MS patients. For example, while comparing two patients, the one with an 11% probability of maintaining the EDSS below 4 will have a higher disability risk than the one with a 15% probability of maintaining the EDSS score below the same value. Hence, the patient with a higher risk will have a higher priority during the monitoring process, in order to succeed in improving the quality of life and reducing the disability progression in MS patients.

For the natalizumab treated patient, similar results were obtained in terms of disability prediction. These results are consistent with the fact that the two patients had similar characteristics, both being females and being diagnosed with MS in their early 30s. Neither patient had any prior treatment before the current treatment, glatiramer acetate and natalizumab respectively. However, the natalizumab treated patient received mitoxantrone during the therapy, very probable due to a high risk of developing primary progressive MS [10, 11]. For this patient, the algorithm falsely predicted a good outcome (86.8%), despite of the fact that the neurologist did not set a good patient prognostic during the examination. These aspects are of utmost importance and reflect the need of including a complete list of

patient clinical characteristics and of embedding in the prediction model all related therapies which might influence the disability and outcome of MS patients. The most similar study with the one which was developed was conducted by Zhao and embedded as predictive variables demographical data, clinical characteristics and image analysis in order to predict MS patients' disability depending on EDSS [22]. Despite the high complexity of the predictive variables and the consistent number of patients which were included in the study (1693), the mean accuracy and recall varied in the range of 68% and 71%, as opposed to a mean accuracy of 75.3% and a mean recall of 76.22% in the present study, both obtained for Group A (151 MS patients with different drug therapies) by considering all 5 cases according to the EDSS threshold value.

However, the main elements of differentiation and originality which were present in the current study were the good ROC AUC Score (a mean ROC AUC Score of 79.51% for the disability prediction in Group A and 80.29% for the outcome prediction in natalizumab treated patients). As opposed to the current study, the one developed by Zhao did not develop a probability estimation algorithm, nor did it aim at building an outcome prediction model [22].

Another study with a reasonable amount of similarity to the present one was conducted by Weideman, which proved the advantages of developing a novel MS severity score, with an important predictive value regarding future relapses and disability accumulation [20]. The main advantage of this score over the classical EDSS based MS Severity Score was the substitution of age with disease duration, which as a consequence leads to a proper quantification of disease onset. However, even though the web application which was developed provides by means of complex statistical models different correlation analyses and simulations which might aid in therapeutic decisions, it was not based on a specific machine learning technique, which might be considered as a drawback as opposed to the present study. Therefore, it should be mentioned that that the final model which is used in the computations is a statistical model embedded in a disease severity scale, and not a machine learning model [20].

By contrast, a study conducted by Stühler aimed at developing a proper machine learning algorithm (based on Bayesian generalized linear models) which estimated the treatment response in MS patients [17]. The validation results were considered clinically relevant for both the relapse and the confirmed disability progression prediction. However, the statistical measures used for validation (mean squared error, negative log-likelihood and C-Index) were different than the ones used in the present study and hence, a relevant comparison cannot be performed from this point of view, even though both studies were based on a probabilistic approach in terms of prediction. The

authors therefore conclude that the main advantage of the present work over the one conducted by Stühler was the EDSS threshold probability prediction, which might lead to a more accurate quantification of MS patients evolution as opposed to the sole estimation of disability progression [17].

Therefore, despite the limitations and the disadvantages which were highlighted in detail, the developed algorithms present a consistent degree of clinical relevance, being able of retrieving patient-tailored disability and outcome profiles.

In addition, future studies aim at developing a simple and intuitive online platform, which could be used by the physician for patient stratification and prioritization in terms of outcome probability estimation. Consequently, this option might lead on the long term to an important improvement in the quality of life of MS patients.

Conclusions

A prediction model of the disability and outcome of MS patients was developed and validated. In order to improve the clinical relevance of the algorithm, future studies must include more patients, select more relevant predictive variables and create an online platform. After these additional steps are undertaken, the prediction algorithms might be used for routine patient monitoring and prioritization in hospitals, for an early estimation of the most likely disability degree and outcome of a certain patient, both depending on the treatment duration and the complete disease modifying therapy.

Conflict of interest

The authors declare no conflict of interest.

References

1. Cree BAC, Gourraud PA, Oksenberg JR, Bevan C, Crabtree-Hartman E, Gelfand JM, Goodin DS, Graves J, Green AJ, Mowry E, Okuda DT, Pelletier D, von Büdingen HC, Zamvil SS, Agrawal A, Caillier S, Ciocca C, Gomez R, Kanner R, Lincoln R, Lizee A, Qualley P, Santaniello A, Suleiman L, Bucci M, Panara V, Papinutto N, Stern WA, Zhu AH, Cutter GR, Baranzini S, Henry RG, Hauser SL, Long-term evolution of multiple sclerosis disability in the treatment era. *Ann Neurol.*, 2016; 80(4): 499-510.
2. Cutter GR, Zimmerman J, Salter AR, Knappertz V, Suarez G, Waterbor J, Howard VJ, Ruth AM, Causes of death among persons with multiple sclerosis. *Mult Scler Relat Disord.*, 2015; 4(5): 484-490.
3. Désir C, Bernard S, Petitjean C, Heutte L, A Random Forest Based Approach for One Class Classification in Medical Imaging. In Wang F, Shen D, Yan P, Suzuki K, (editors) MLMI: International Workshop on Machine Learning in Medical Imaging, 2012; vol. 7588: 234-241.
4. Drotar P, Smekal Z, Comparative study of machine learning techniques for supervised classification of biomedical data. *Acta Electrotechnica et Informatica.* 2014; 14(3): 5-10.

5. DuBrava S, Mardekian J, Sadosky A, Bienen EJ, Parsons B, Hopps M, Markman J, Using Random Forest Models to Identify Correlates of a Diabetic Peripheral Neuropathy Diagnosis from Electronic Health Record Data. *Pain Medicine*, 2017; 18(1): 107-115.
6. Fonseca-Delgado R, Gomez-Gil P, An assessment of ten-fold and Monte Carlo cross validations for time series forecasting. In: 10th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE) Mexico City, Mexic, 2013; 215-220.
7. Ion-Mărgineanu A, Kocevar G, Stamile C, Sima DM, Durand-Dubief F, Van Huffel S, Sappey-Marinier D, Machine Learning Approach for Classifying Multiple Sclerosis Courses by Combining Clinical Data with Lesion Loads and Magnetic Resonance Metabolic Features. *Front Neurosci.*, 2017; 11: Art. 398: 1-11.
8. Jiao Y, Du P, Performance measures in evaluating machine learning based bioinformatics predictors for classifications. *Quantit Biol.*, 2016; 4(4): 320-330.
9. Kamal KC, Alexandru DO, Rogoveanu OC, Kamal D, Buteică AS, Bălșeanu TA, Pănuș C, Maria DT, Trăistaru MR, Results of alendronate treatment in patients diagnosed with aseptic osteonecrosis of the femoral head. *Farmacia*, 2019; 67(2): 318-322.
10. Leray E, Moreau T, Fromont A, Edan G, Epidemiology of multiple sclerosis. *Revue Neurologique*, 2016; 172(1): 3-13.
11. Martinelli Boneschi F, Vacchi L, Rovaris M, Capra R, Comi G, Mitoxantrone for multiple sclerosis (Review). *Cochrane Database of Systematic Reviews*, 2013; 31(5): 1-42.
12. Montalban X, Gold R, Thompson AJ, Otero-Romero S, Pia Amato M, Chandraratna D, Clanet M, Comi G, Derfuss T, Fazekas F, Hartung HP, Havrdova E, Hemmer B, Kappos L, Liblau R, Lubetzki C, Marcus E, Miller DH, Olsson T, Pilling S, Selmaj K, Siva A, Sorensen PS, Sormani MP, Thalheim C, Wiendl H,ECTRIMS/EAN Guideline on the pharmacological treatment of people with multiple sclerosis. *Multi Scler.*, 2018; 24(2): 96-120.
13. Riaz MK, Potential drug-drug interactions and strategies for their detection and prevention, *Farmacia*, 2019; 67(4): 572-579.
14. Sarica A, Cerasa A, Quattrone A, Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. *Front Aging Neurosci.*, 2017; 9: Art 329: 1-12.
15. Scikit learn developers, Documentation of scikit-learn 0.21.2, <http://scikit-learn.org/>.
16. Spelman T, Butzkueven H, Kappos L, Trojano M, Wiendl H, Chen Y, Campbell N, Licata S, Ho PR, In treatment-naive patients with relapsing-remitting multiple sclerosis (RRMS), initiating natalizumab earlier is associated with greater disability improvement than delaying treatment: real-world results from the TYSABRI® Observational Program (TOP) (P6.350). *Neurology*, 2017; 88(16 Suppl.).
17. Stühler E, Braune S, Lionetto F, Heer Y, Jules E, Westermann C, Bergmann A, van Hövell P, Neuro Trans Data Study Group, Framework for personalized prediction of treatment response in relapsing remitting multiple sclerosis. *BMC Med Res Methodol.*, 2020; 20: Art. 24: 1-15.
18. Tape TG, Interpreting Diagnostic Tests. The Area Under an ROC Curve, <http://gim.unmc.edu/>.
19. Van Munster CE, Outcome measures in clinical trials for multiple sclerosis. *CNS Drugs*, 2017; 31(3): 217-236.
20. Weideman AM, Barbour C, Tapia-Maltos MA, Tran T, Jackson K, Kosa P, Komori M, Wichman A, Johnson K, Greenwood M, Bielekova B, New Multiple Sclerosis Disease Severity Scale Predicts Future Accumulation of Disability. *Front Neurol.*, 2017; 8: Art. 598: 1-18.
21. Wottschel V, Alexander DC, Kwok PP, Chard DT, Stromillo ML, De Stefano N, Thompson AJ, Miller DH, Ciccarelli O, Predicting outcome in clinically isolated syndrome using machine learning. *NeuroImage: Clinical*, 2015; 7: 281-287.
22. Zhao Y, Healy BC, Rotstein D, Guttmann CR, Bakshi R, Weiner HL, Brodley CE, Chitnis T, Exploration of machine learning techniques in predicting multiple sclerosis disease course. *PLoS One*, 2017; 12(4): e0174866: 1-13.