




Article

# Combining Clinical and Genetic Data to Predict Response to Fingolimod Treatment in Relapsing Remitting Multiple Sclerosis Patients: A Precision Medicine Approach

Laura Ferrè <sup>1,2,3</sup>, Ferdinando Clarelli <sup>2</sup>, Beatrice Pignolet <sup>4,5</sup>, Elisabetta Mascia <sup>2</sup>, Marco Frasca <sup>6,7,8</sup> ,  
Silvia Santoro <sup>2</sup>, Melissa Sorosina <sup>2</sup> , Florence Bucciarelli <sup>4,5</sup>, Lucia Moiola <sup>1</sup>, Vittorio Martinelli <sup>1</sup>,  
Giancarlo Comi <sup>3</sup>, Roland Liblau <sup>5,9</sup>, Massimo Filippi <sup>1,3,10,11</sup>, Giorgio Valentini <sup>6,7,8</sup>  and Federica Esposito <sup>1,2,\*</sup>

- <sup>1</sup> Neurology and Neurorehabilitation Unit, IRCCS San Raffaele Hospital, 20132 Milan, Italy  
<sup>2</sup> Laboratory of Human Genetics of Neurological Disorders, IRCCS San Raffaele Hospital, 20132 Milan, Italy  
<sup>3</sup> Vita-Salute San Raffaele University, 20132 Milan, Italy  
<sup>4</sup> Centre Hospitalier Universitaire de Toulouse, CEDEX 9, 31059 Toulouse, France  
<sup>5</sup> Institut Toulousain des Maladies Infectieuses et Inflammatoires (Infinity), INSERM UMR1291–CNRS UMR5051—Université Toulouse III, CEDEX 3, 31024 Toulouse, France  
<sup>6</sup> AnacletoLab, Dipartimento di Informatica, Università degli Studi di Milano, 20133 Milan, Italy  
<sup>7</sup> Data Science Research Center, Università degli Studi di Milano, 20133 Milan, Italy  
<sup>8</sup> Infolife National Lab, CINI, 00185 Rome, Italy  
<sup>9</sup> Department of Immunology, Toulouse University Hospitals, CEDEX 3, 31024 Toulouse, France  
<sup>10</sup> Neuroimaging Research Unit, IRCCS San Raffaele Hospital, 20132 Milan, Italy  
<sup>11</sup> Neurophysiology Unit, IRCCS San Raffaele Hospital, 20132 Milan, Italy  
\* Correspondence: [esposito.federica@hsr.it](mailto:esposito.federica@hsr.it)



**Citation:** Ferrè, L.; Clarelli, F.; Pignolet, B.; Mascia, E.; Frasca, M.; Santoro, S.; Sorosina, M.; Bucciarelli, F.; Moiola, L.; Martinelli, V.; et al. Combining Clinical and Genetic Data to Predict Response to Fingolimod Treatment in Relapsing Remitting Multiple Sclerosis Patients: A Precision Medicine Approach. *J. Pers. Med.* **2023**, *13*, 122. <https://doi.org/10.3390/jpm13010122>

Academic Editor: Niels Bergsland

Received: 14 November 2022

Revised: 30 December 2022

Accepted: 30 December 2022

Published: 6 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** A personalized approach is strongly advocated for treatment selection in Multiple Sclerosis patients due to the high number of available drugs. Machine learning methods proved to be valuable tools in the context of precision medicine. In the present work, we applied machine learning methods to identify a combined clinical and genetic signature of response to fingolimod that could support the prediction of drug response. Two cohorts of fingolimod-treated patients from Italy and France were enrolled and divided into training, validation, and test set. Random forest training and robust feature selection were performed in the first two sets respectively, and the independent test set was used to evaluate model performance. A genetic-only model and a combined clinical-genetic model were obtained. Overall, 381 patients were classified according to the NEDA-3 criterion at 2 years; we identified a genetic model, including 123 SNPs, that was able to predict fingolimod response with an AUROC= 0.65 in the independent test set. When combining clinical data, the model accuracy increased to an AUROC= 0.71. Integrating clinical and genetic data by means of machine learning methods can help in the prediction of response to fingolimod, even though further studies are required to definitely extend this approach to clinical applications

**Keywords:** multiple sclerosis; fingolimod; machine learning; predictive model; genetic markers; precision medicine

## 1. Introduction

### 1.1. Background

Multiple Sclerosis (MS) is a chronic inflammatory disease of the central nervous system with a complex etiology and a high heterogeneity in terms of clinical presentation and treatment response [1]. In the last decade, the therapeutic opportunities for Relapsing Remitting MS (RRMS) have dramatically expanded, increasing the complexity in disease management and prompting the need for a more individualized approach that takes into account patients' specific features [2].

### 1.2. Rationale

The existence of genetic factors predisposing to MS has been demonstrated by large genome-wide association studies [3], whereas fewer, smaller studies suggested a possible influence of genetic factors in determining disease severity [4–6] and modulating response to first-line treatments like interferon and glatiramer acetate [7–13]. Fewer data are available on genetic markers associated with response to second-line drugs [14,15] and, to the best of our knowledge, no studies have investigated genetic factors associated with response to fingolimod.

Fingolimod (FTY) is a highly effective second-line drug approved for RRMS; nonetheless, some subjects show persistent disease activity during FTY treatment and the early detection of these non-responder individuals is essential to promptly address them to more effective therapies [16], given the increasing number of drugs currently available.

Machine learning (ML) algorithms, which can better accommodate the complexities of the relations among variables than the traditional regression methods, are becoming increasingly valuable tools for precision medicine [17,18].

### 1.3. Objective

In the present study, we applied ML methods to investigate the presence of a genetic signature of response to FTY that, in combination with clinical and demographic characteristics, could support the prediction of drug response.

## 2. Materials and Methods

We collected two cohorts of RRMS patients treated with FTY for whom genetic and clinical data during therapy were available. The first cohort (OSR) included 364 patients who started FTY at San Raffaele Hospital MS center in Milan, Italy; the second cohort (CHUT) included 108 FTY-treated RRMS patients from the Centre Hospitalier Universitaire de Toulouse, France.

Baseline and follow-up clinical data during the first 2 years of FTY therapy were obtained; specifically, the following variables were collected: gender, age at onset (AAO), disease duration at treatment start (DD), annualized relapse rate (ARR) in the 2 years prior to FTY start, previous disease modifying treatment, EDSS score at treatment start and at 2 years, number of relapses, and presence of new/enlarging T2-lesions and Gd-enhancing lesions at brain MRI during therapy. Patients treated with natalizumab (NTZ) in the 9 months prior to FTY were excluded due to the known possible “rebound” effect occurring after NTZ discontinuation [19,20] in order to avoid potential misclassification.

Treatment response was assessed at 2 years using the NEDA-3 criterion (No evidence of disease activity [21]), defined as absence of relapses, active MRI lesions, and disease progression. The Ethical Committee at San Raffaele Hospital, Milan, Italy approved the study and all patients signed the informed consent.

### 2.1. Genotyping and Quality Checks

Genetic data for OSR and CHUT cohorts were obtained with the Illumina®Human-OmniExpress Kit. Standard quality control (QC) steps were applied using Plink v1.9beta [22]: per-SNPs QC included removing variants with minor allele frequency (MAF) < 0.01, genotyping rate < 0.97, and deviation from Hardy-Weinberg equilibrium ( $p < 10^{-4}$ ). Per-sample QC was performed removing subjects with call rate < 0.95, excess relatedness, sex mismatch, or that were ancestry outliers according to a multidimensional scaling analysis.

The two datasets were merged and SNPs were pruned, discarding those in linkage disequilibrium (LD) ( $r^2 > 0.2$ ), in order to remove redundancies in the set of markers, thus reducing the dimensionality of the analyses. Genotypes were coded in discrete additive dosages of minor alleles for ML modeling.

## 2.2. Predictive Models

### 2.2.1. Genetic Model

In order to avoid selection bias [23] and model overfitting, after combining the two cohorts, we created three subsets called training set (TRset), validation set (Vset), and test set (TEset), including 40%, 40%, and 20% of studied subjects, respectively. The three groups were generated maintaining the proportion between OSR and CHUT cohorts, as well as the EDA/NEDA ratio within each group.

Our preliminary experiments showed that Random Forests (RFs) achieved more stable and accurate results compared to other models (Support Vector Machines, K-nearest-neighbours, decision trees, bagging and boosting) confirming previous ML results on genetic data [24,25]. For this reason, we used RFs to predict response to FTY based on genetic and clinical features.

The TRset was used to train the RFs, whereas the Vset was used to select the group of SNPs (signature) to be included in the predictive model, starting from a dataset of about 113,000 LD-pruned SNPs. Feature selection was performed using a robust cross-validated selection method, based on the stability of the SNPs to be selected. We designed a feature selection algorithm to detect the SNPs that most steadily correlate, according to Pearson coefficient, with the response status across multiple samples of the data. Specifically, we repeated a cross-validation procedure 100 times, storing at each repetition the k top-ranked SNPs in each fold of the cross-validation. Then, the algorithm chose only the SNPs selected with at least a relative frequency f across the repeated cross-validations.

Of note, we used the Vset to select the SNP signature, distinct from both the TRset and TEset, thus avoiding the selection bias and reducing the overfitting of the trained model [23]. Finally, the TEset was used to test the models trained using the selected SNP signature.

We performed an extensive model selection by comparing the results of about 2000 models resulting from different combinations of the learning parameters of the RFs by varying the number of decision trees of the ensemble between 10 and 100, the maximum number of nodes for each decision tree from 1 to 100, the minimum number of examples stored in each leaf node from 1 to 10, and of the robust cross-validated feature selection method by varying the minimum relative frequency f of each feature between 0.05 and 1, and the number k of top features to be considered at each iteration of the cross-validation between 50 and 1000. It is easy to see that for a fixed k, large values of f (close to 1) lead to small sets f of selected SNPs, whereas small values lead to larger SNP signatures.

The best models were selected and evaluated on the TEset as measured by the area under the receiver operating characteristic curve (AUROC); the Area Under the Precision-Recall Curve (AUPRC), F-score, and accuracy were also calculated to evaluate models performance.

The analyses were performed using R statistical environment, version 3.6.3; specifically, the “randomForest”, “caret” and “precrec” R packages were used, as well as a set of functions designed specifically for the present analyses that are available upon request.

### 2.2.2. Combined Clinical and Genetic Model

As for the genetic model, we trained the RFs separately on the clinical data and selected four signatures through which we achieved the best results on the Vset, starting from 17 clinical parameters (gender, AAO, age at treatment start, DD, ARR in the 2 years before FTY, baseline EDSS, presence of new T2-lesions and Gd+-lesions at baseline MRI plus other nine binary features derived from the categorization of previous treatments). We ran about 2000 different models and selected the top four models according to their performance.

Finally, we combined the genetic and clinical models by means of multi-view random forests, where each decision tree was separately trained on a different bootstrap sample of either genetic or clinical data. The predictions of the resulting decision trees were finally combined to obtain the consensus multi-view prediction of the RF ensemble [26].

### 2.2.3. Predictive Performance of the Genetic Model in Patients Treated with Other Immunomodulatory Drugs

In order to assess whether the set of SNPs predicted as associated to a better response were rather markers of a mild disease activity, we considered two independent cohorts of patients treated with glatiramer acetate (GA,  $n = 273$ ) and beta-interferons (IFN,  $n = 304$ ), enrolled at San Raffaele Hospital MS center, with available genetic data. For these subjects, the same clinical information previously mentioned was collected during treatment, and response was assessed at 2 years using the NEDA-3 criterion. The genetic model selected in FTY-treated patients was applied to GA and IFN-treated subjects, and the AUROC was calculated to assess whether it was specific for FTY response or if it was more generally associated to disease activity.

### 2.2.4. Classification Performance of the Combined Model

In order to test the ability of the selected model to identify patients that respond to FTY, we considered the predicted probability of non-response to FTY as calculated on the independent TEset by the selected clinical and genetic model, and we divided these values into tertiles; the highest tertile included the patients that the model predicted with higher likelihood as being non-responders to FTY (PrNR), whereas the lowest tertile grouped patients more likely to respond to treatment (PrR). We then compared the lowest and highest tertiles in terms of disease activity during FTY treatment; specifically, we compared the number of new and/or enlarging T2 lesions and the number of relapses between the two groups by means of the non parametric Mann-Whitney test; we then compared the proportion of patients free from MRI activity and free from clinical activity and the proportion of patients achieving the NEDA outcome by means of a chi-square test.

## 3. Results

### 3.1. Summary of Results

In the present study we identified a genetic signature of 123 SNPs that was able to predict FTY response with an area under the receiver operating characteristic (AUROC) = 0.65 and the model accuracy further increased when considering also clinical data (AUROC = 0.71). Even though the model predictivity is not enough for implementation in clinical practice, our findings suggest that the combination of clinical and genetic data by means of machine learning methods can support in the prediction of response to FTY.

### 3.2. Detailed Results

We applied the genetic and the combined clinical-genetic model to the prediction of the EDA/NEDA status using the OSR/CHUT combined cohort according to the experimental set-up described in Section 2.

More precisely, 342 FTY-treated RRMS patients from the OSR cohort and 78 from the CHUT dataset were considered for the following steps, after removal of patients treated with NTZ in the previous nine months and genetic QCs.

Among them, 17 and 22 patients in the CHUT and OSR cohort, respectively, were excluded because we were not able to correctly classify them according to the NEDA-3 criterion due to missing clinical data.

The final analysis was performed on 381 patients, of whom 197 showed evidence of disease activity (EDA) and 184 were NEDA during the two-year follow-up. Among them, 152 entered the TRset, 152 the Vset, and 77 the independent TEset. Clinical and demographic characteristics of the included patients, stratified according to the three sets, are described in Table 1.

**Table 1.** Clinical and demographic characteristics of included patients, stratified according to TRset, Vset, and TEsset.

	Whole Cohort (n: 381)	TRset (n: 152)	Vset (n: 152)	TEset (n: 77)	p-Value
F:M ratio	270:111	101:51	119:33	50:27	0.03
Age at disease onset, mean ± SD	29 ± 9.5	28.3 ± 9.2	29.1 ± 9.7	30.2 ± 9.6	n.s.
Age at FTY start, mean ± SD	39.5 ± 9.5	38.8 ± 9.1	39.6 ± 9.7	40.4 ± 9.8	n.s.
Disease duration (yrs), mean ± SD	10.5 ± 7.6	10.5 ± 7.1	10.5 ± 8.2	10.2 ± 7.1	n.s.
ARR in the 2 years prior FTY, mean ±SD	0.82 ± 0.84	0.74 ± 0.83	0.93 ± 0.90	0.78 ± 0.73	n.s.
Previous DMT					
Naïve	30 (7.9%)	12 (7.9%)	13 (8.6%)	5 (6.5%)	
No therapy	26 (6.8%)	12 (7.9%)	10 (6.6%)	4 (5.2%)	
IFN	149 (39.1%)	59 (38.8%)	58 (38.2%)	32 (41.5%)	
GA	104 (27.3%)	40 (26.3%)	40 (26.3%)	24 (31.2%)	n.s.
DMF	13 (3.4%)	4 (2.6%)	6 (3.9%)	3 (3.9%)	
Teriflunomide	11 (2.9%)	4 (2.6%)	3 (2%)	4 (5.2%)	
Immunosuppressants	17 (4.5%)	6 (4.0%)	9 (5.8%)	2 (2.6%)	
Natalizumab	29 (7.6%)	13 (8.6%)	13 (8.6%)	3 (3.9%)	
Other	2 (0.5%)	2 (1.3%)	0 (0%)	0 (0%)	
EDSS at FTY start, median (range)	2.0 (0–7.0)	2 (0–6.0)	2 (0–7.0)	2 (0–6.0)	n.s.
<Patients with Gd+ lesions at baseline brain MRI scan	33.5%	35.1%	31.7%	33.8%	n.s.
Patients with new/enlarged T2 lesions at baseline brain MRI scan	49.1%	45.7%	54.5%	44.9%	n.s.

FTY: fingolimod; ARR: annualized relapse rate; DMT: disease modifying treatment; IFN: interferon; GA: glatiramer acetate; DMF: dimethyl fumarate; Gd+ lesions: gadolinium enhancing lesions. P-value refers to the comparison between the 3 sets.

### 3.3. Genetic Model

We applied the robust cross-validation procedure introduced in Section 2.2.1 to select the genotypic signatures predictive of the EDA/NEDA status. Using the training dataset only to avoid the selection bias, we selected the *top-f* SNPs most steadily correlated with the response status across ten-fold cross-validation repeated 100 times (Table 2). For instance, in Table 2, the g2 model includes a signature with 123 SNPs (*sign*) selected using the first 500 top-ranked features (*top-f*) with a frequency (*min-fr*) equal to 0.1 (i.e., SNPs selected at least 10% of times across the repeated cross-validation procedure). Random forests were trained on the training set using the SNP signatures, and their parameters were selected by evaluating their AUROC, AUPRC, and F-score performance on the independent validation set to limit overfitting (see Table 2 for details). Finally, the generalization performance of the best models was evaluated on the independent test set (Table 2).

**Table 2.** Best random forest models trained on the genotypic data.

Model	Top-f	Min-fr	Sign	Ntree	Nodesize	Maxn	TR AUROC	TR AUPRC	TR F	TR Acc	TE AUROC	TE AUPRC	TE F	TE Acc
g1	500	0.05	1022	20	10	10	0.8493	0.8476	0.7973	0.796	0.65	0.6483	0.6837	0.5194
g2	500	0.1	123	10	10	30	0.8438	0.8496	0.7861	0.7565	0.6446	0.663	0.7142	0.5844
g3	500	0.05	1022	10	1	15	0.838	0.8671	0.7567	0.7631	0.5801	0.5907	0.745	0.6623
g4	500	0.15	8	10	2	60	0.858	0.8712	0.8	0.7828	0.6135	0.6176	0.7102	0.5974

The table shows the four best genetic models, named g1, g2, g3, and g4. Parameters of the feature selection method: top-f stands for the number of the top-ranked selected features; min-fr is the minimum frequency of the feature to be selected, sign is the number of selected features. Parameters of the random forests: ntree is the number of trees; nodesize is the minimum size of each leaf of the tree; maxn is the maximum number of nodes for each tree. TR\_AUROC, TR\_AUPRC, TR\_F, TR\_acc, TE\_AUROC, TE\_AUPRC, TE\_F, TE\_acc stand for AUROC, AUPRC, F-score, and Accuracy, respectively, on the training and test set.

The AUROC calculated on the independent TEset was close to 0.65, and F-scores and AUPRC were even higher, suggesting that the NEDA classification can be learnt from genetic data, although the accuracy of the prediction was not enough for application in everyday clinical practice.

Due to the smaller number of variants included in the second model (g2)—that makes it more appealing for a potential clinical application—and the very similar predictive performance, we selected it for further analyses; the list of the 123 SNPs prioritized by the analysis is reported in Supplementary Table S1.

In order to gain some biological insight from the detected signature associated with the FTY response, we performed an enrichment analysis to identify the pathways in which the selected variants were involved. First, we annotated the variants using the Ensembl Variant Effect Predictor tool (VeP), then, we selected genes for which at least a moderate impact was predicted and belonging to the biotypes “protein coding”, “processed transcript”, and “retained intron”. We then performed an over representation analysis with the tool Webgestalt [27] using the KEGG pathway database as reference. Of the 73 annotated genes, 30 were annotated to functional categories present in the database and, though no pathway survived multiple testing correction, among the top enriched pathways (Table 3) we found “Sphingolipid signaling pathway” (*p*-value: 0.008), “Sphingolipid metabolism” (*p*-value: 0.011), “Cell adhesion molecules (CAMs)” (*p*-value: 0.013), and “Inflammatory bowel disease (IBD)” (*p*-value: 0.021). These results suggest that our algorithm is indeed able to identify genes that play a role in modulating FTY mechanism of action and therapeutic response.

**Table 3.** Over-representation analysis of the genes selected by the ML model.

Description	Size	Expected	Enrichment	<i>p</i> Value	FDR
Renin secretion	65	0.23	13.26	0.001	0.42
Calcium signaling pathway	183	0.64	6.28	0.003	0.51
Sphingolipid signaling pathway	118	0.41	7.30	0.008	0.65
Sphingolipid metabolism	47	0.16	12.22	0.011	0.65
Cholesterol metabolism	50	0.17	11.49	0.013	0.65
Cell adhesion molecules (CAMs)	144	0.50	5.98	0.013	0.65
cGMP-PKG signaling pathway	163	0.57	5.29	0.018	0.67
Cortisol synthesis and secretion	64	0.22	8.98	0.021	0.67
Inflammatory bowel disease (IBD)	65	0.23	8.84	0.021	0.67
Long-term potentiation	67	0.23	8.58	0.022	0.67

Size: gene set size; Expected: n° of predicted overlap; Enrichment ratio: ratio between the observed and expected overlap; FDR: false discovery rate.

### 3.4. Combined Clinical and Genetic Model

We then applied the same algorithm to the clinical data: interestingly, among the clinical features steadily selected for inclusion in the model, there were age at treatment start, ARR in the 2 years before FTY start, and presence of new T2 and Gd+ lesions at baseline, which are already known to be associated with disease activity upon FTY treatment [19].

Table 4 reports the four best models trained on the clinical data that performed slightly better in terms of AUROC (0.69) and AUPRC (0.67), compared to the genetic model.

**Table 4.** Best Random forest models trained on the clinical data.

Model	Top-f	Min-fr	Sign	Mtry	N Tree	Node Size	Maxn	TR AUROC	TR AUPRC	TR F	TR acc	TE AUROC	TE AUPRC	TE F	TE acc
c1	8	0.5	9	10	100	2	100	1	1	1	1	0.6895	0.6709	0.7339	0.6494
c2	8	0.5	9	4	20	2	100	0.9785	0.9803	0.9255	0.9211	0.6405	0.7320	0.7091	0.6364
c3	8	0.5	9	10	20	2	30	0.9152	0.9281	0.8434	0.8289	0.623	0.6422	0.7379	0.6494
c4	2	0.05	14	3	10	1	100	0.9971	0.9974	0.9684	1	0.6895	0.6709	0.7339	0.6494

The table shows the four best clinical models, named c1, c2, c3 and c4. Parameters of the feature selection method: top-f stands for the number of the k top-ranked selected features; min-fr is the minimum frequency f of the features to be selected. Sign is the cardinality of the clinical signature. Parameters of the random forests: mtry is the number of the randomly selected features at each node of the decision trees, ntree is the number of trees; nodesize is the minimum size of each leaf of the tree; maxn is the maximum number of nodes for each tree. TR\_AUROC, TR\_AUPRC, TR\_F, TR\_acc, TE\_AUROC, TE\_AUPRC, TE\_F, TE\_acc stand for AUROC, AUPRC, F-score, and Accuracy, respectively, on the training and test set.

We then combined the selected genotypic and clinical signatures using multi-view random forests and obtained an increase in the predictive accuracy of the model, with an AUROC of 0.71 and an AUPRC of 0.73 on the TEset (Table 5).

**Table 5.** Best multi-view random forest model trained on combined clinical and genetic data.

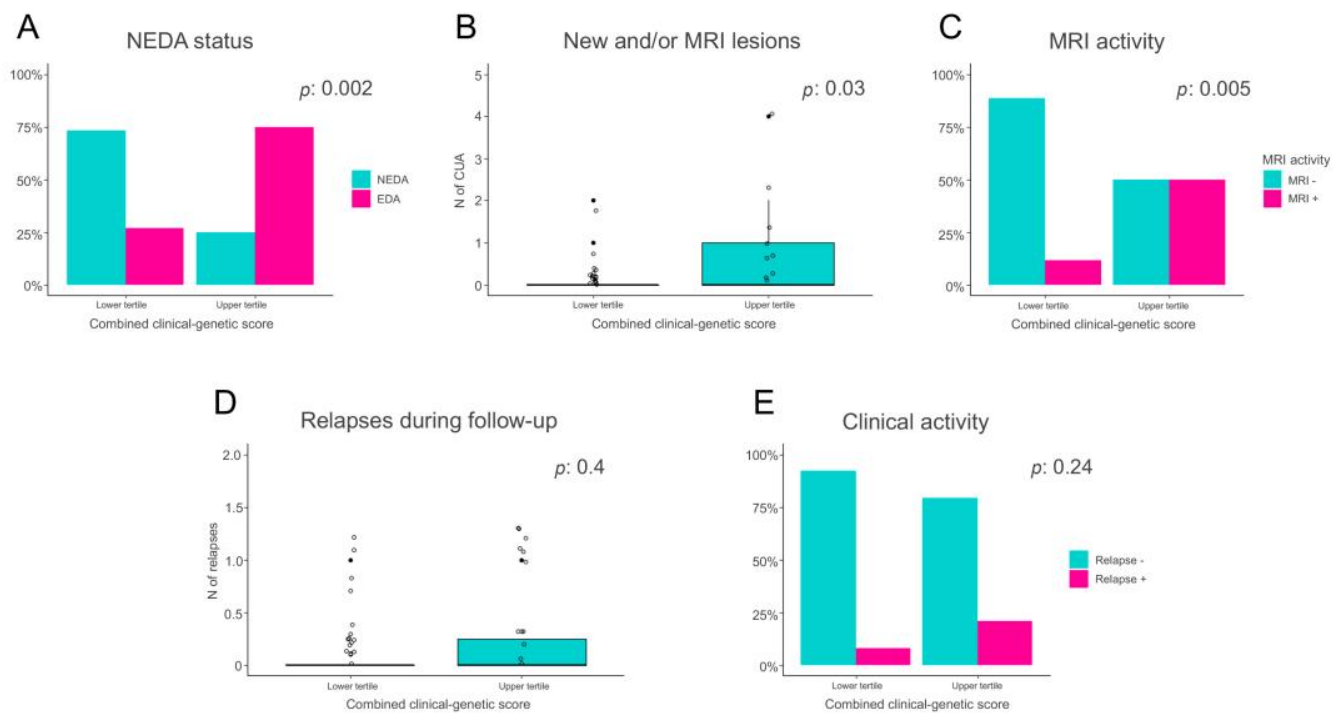
Model	TR_AUROC	TR_AUPRC	TR_F	TR_Acc	TE_AUROC	TE_AUPRC	TE_F	TE_acc
g2-c1	0.9997	0.9997	0.9933	0.9934	0.7095	0.7328	0.7328	0.6623

The model name g2-c1 refers to the single-layer models it derives from. TR\_AUROC, TR\_AUPRC, TR\_F, TR\_acc, TE\_AUROC, TE\_AUPRC, TE\_F, TE\_Acc stand for AUROC, AUPRC, F-score, and Accuracy, respectively, on the training and test set.

These results suggest that combining clinical and genetic data can help in predicting response to FTY.

To further confirm this evidence, we considered the predicted probability of non-response to FTY as calculated on the independent TEset by the combined model g2-c1, and we considered the patients that the model predicted with higher likelihood as being non-responders to FTY (PrNR) and those deemed more likely to respond to treatment (PrR).

As expected, a significantly greater proportion of patients predicted with higher likelihood as being non-responders had evidence of disease activity during the 2-year follow-up, compared to the patients predicted with higher likelihood as being responders (75% vs. 27%,  $p$ : 0.0019, Figure 1A). Moreover, PrNR patients also showed a higher neuroradiological activity during FTY treatment, both in terms of number of active lesions (2.1 vs. 0.61 on average in the PrNR and PrR group, respectively,  $p$ : 0.034 Figure 1B) and of proportion of individuals with an active MRI scan (50% vs. 11.5%  $p$ : 0.005, Figure 1C). The same direction of effect was also found when considering the level of clinical activity, PrNR patients showing a higher number of relapses (0.25 vs. 0.15  $p$ : 0.40, Figure 1D) and a larger proportion of patients with a clinical reactivation (20.8% vs. 7.7%  $p$ : 0.24, Figure 1E) compared to PrR, although not statistically significant.



**Figure 1.** Comparison of disease activity levels between patients predicted to be non-responders to FTY (PrNR) and patients likely to respond to treatment (PrR). (A) Proportion of patients with No Evidence of Disease Activity (NEDA) and Evidence of Disease Activity (EDA) in the PrR and PrNR groups. (B) Number of new and/or active lesions in the PrR and PrNR groups. (C) Proportion of patients with or without MRI activity in the PrR and PrNR groups. (D) Number of clinical relapses in the PrR and PrNR groups. (E) Proportion of patients with or without clinical activity in the PrR and PrNR groups.

### 3.5. Evaluation of the Model in Independent Cohorts of Patients Treated with First-Line Drugs

Finally, to test whether our genetic model was specific for predicting FTY treatment response or was a prognostic algorithm that more broadly predicts MS disease activity, we tested it in patients treated with IFN and GA. In the IFN cohort, 110 patients had NEDA and 189 EDA, whereas in the GA cohort, 93 had NEDA and 170 EDA. We then tested the model trained on the FTY cohort on these 2 datasets and obtained an AUROC of 0.55 and 0.51, respectively, suggesting that the model holds specificity for FTY therapy and is not able to predict response to IFN and GA. We did not test the clinical model on these cohorts, given that the identified clinical predictors are mainly prognostic factors already known to be associated to response to first-line treatment, so we do not expect the clinical model to be specific to FTY-treated patients.

## 4. Discussion

MS is a complex disorder, with substantial heterogeneity in terms of response to treatments, that would greatly benefit from a more individualized management; indeed, predicting the response of MS patients to therapies has been an open problem for many years [28]. Recently, due to the advancements and the spreading of artificial intelligence (AI) methodologies, several ML algorithms have been applied in the MS field; in particular, AI has been applied to MRI data to perform lesion and tissue segmentation [29,30] and to aid in the differential diagnosis with MS mimics [31–33]. A few studies also applied AI to clinical data in order to predict disease course and progression [34,35] but, to our knowledge, a single study applied ML methods to genetic data in order to predict treatment response in patients treated with a first-line drug [36].



In the present study, we applied a supervised ML algorithm to clinical and genetic data in a cohort of MS patients treated with FTY in order to develop a predictive model of response to the drug, and we identified a genetic signature of 123 SNPs that was able to predict treatment response in an independent cohort, even though its predictive accuracy is not enough for its use in clinical practice (AUROC = 0.65, AUPRC = 0.66). When considering clinical data only, a similar performance was obtained (AUROC = 0.69, AUPRC = 0.67), suggesting that clinical information can possibly add to the predictive value of the model. Interestingly, among the clinical parameters that were retained by the robust feature selection algorithm were the age at treatment start, the ARR in the 2 years before FTY, and the presence of new T2 and Gd+ lesions at baseline MRI, variables that have already been associated with disease activity and response to FTY treatment [19]. These results indicate that, even if the predictive power of the model is limited, our ML approach is able to identify clinically significant predictors.

Similarly, when applying an over-representation analysis to the genetic hits selected by the algorithm, we found that there was an enrichment of genes implicated in sphingolipid metabolism and in cell-adhesion pathways, further supporting the ability of the model to identify biologically meaningful signals implicated in response to FTY.

As expected, the highest accuracies were obtained when considering both clinical and genetic data: in fact, the combined model yielded an AUC of 0.71 and an AUPRC of 0.73. Such predictive performance, although insufficient to guide decision making in clinical practice, suggests that ML methods have the potential to stratify patients for whom FTY is a treatment option. Moreover, our results are in line, if not better, than those obtained by a previous study on glatiramer acetate [36].

When analyzing only patients whose response state is predicted with higher or lower likelihood as being non-responder to FTY, the algorithm was able to identify two categories of patients that significantly differ in terms of disease activity during FTY treatment: only 25% of PrR patients showed disease reactivation in the 2 years after treatment start compared to 75% of PrNR. Similarly, PrNR patients have a significantly greater rate of MRI activity and showed a modest trend towards more relapses compared to PrR.

Among the strengths of our work is the detailed clinical characterization that allowed the improvement of the genetic-only model. Most importantly, our study design included a validation step in an independent cohort that was completely blind to the model training and to the features selection process; compared to internal cross-validation procedures, the availability of an external independent population allows for an unbiased and not overly-optimistic estimation of the generalizability of the model and reduces the risk of model overfitting, thus increasing the reliability of our findings [23,37,38].

Moreover, our robust procedure to select informative SNP signatures, combined with an unbiased evaluation of the experimental results using an integrated ML approach, offers a methodological framework that can be applied to MS or also other complex human traits.

On the other hand, we are aware of some limitations of the present study, mainly due to the modest sample size, especially in relation to the high dimensionality of genetic data [39], with an independent test set of only 77 patients, and the relatively short follow-up.

## 5. Conclusions

Our study shows that ML methodologies hold potential to be applied to clinical and genetic data towards a more personalized approach in MS and advocates for further studies addressing the issue of predicting treatment response. In perspective, the availability of larger cohorts and an effective combination of clinical, omics, and other types of data (e.g., imaging data) through an integrated ML approach [40] (such as our proposed multi-view RF ensembles) could lead to improved response predictions potentially applicable in clinical practice.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/jpm13010122/s1>, Supplementary Table S1: Genetic signature prioritized by the machine learning analysis.

**Author Contributions:** Conceptualization: F.E., L.F. and G.V.; methodology: F.C. and G.V.; software: M.F. (Marco Frasca) and G.V.; validation: M.F. (Marco Frasca) and G.V.; formal analysis: L.F., M.F. (Marco Frasca) and G.V.; investigation: F.E., L.F. and B.P.; resources: L.M., V.M., G.C., R.L. and M.F. (Massimo Filippi); data curation: B.P., E.M., S.S., M.S. and F.B.; writing—original draft preparation: L.F. and F.E.; writing—review and editing: L.F., F.C., B.P., E.M., M.F. (Marco Frasca), S.S., M.S., F.B., L.M., V.M., G.C., R.L., M.F. (Massimo Filippi), G.V. and F.E.; visualization: L.F. and F.E.; supervision: R.L., M.F. (Massimo Filippi) and F.E.; project administration: F.E.; funding acquisition: F.E. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by “Fondazione Italiana Sclerosi Multipla” [project 2013/R/13].

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Ospedale San Raffaele (FINGO-MS, 15/05/2014).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Patient-level data presented in this study are available on request to the corresponding author. Data regarding genetic variants selected by the model are included in Supplementary Table S1.

**Conflicts of Interest:** Ferdinando Clarelli, Beatrice Pignolet, Elisabetta Mascia, Marco Frasca, Santoro Silvia, Sorosina Melissa, Florence Bucciarelli and Giorgio Valentini have nothing to disclose. Laura Ferrè received compensation for speaking activities from Novartis. Lucia Moiola received compensation for consulting services, travel grants, and/or speaking activities from Biogen, Serono, Sanofi, Teva, Roche, and Novartis. Vittorio Martinelli has received compensations for consulting services and/or speaking activities from Novartis, Genzyme, Almirall, TEVA, Biogen, and Merck-Serono. Giancarlo Comi has received personal compensation for consulting and speaking activities from F. Hoffmann-La Roche Ltd., Roche SpA, Novartis, Teva Pharmaceutical Industries Ltd, Teva Italia Srl, Sanofi Genzyme, Genzyme Corporation, Genzyme Europe, Merck KGaA, Merck Serono SpA, Celgene Group, Biogen Idec, Biogen Italia Srl, Almirall SpA, Forward Pharma, Medday, and Excemed. Roland Liblau received grant support from Pierre Fabre, GlaxoSmithKline, and BMS. He received speaker or scientific board honoraria from Biogen, Servier, Novartis, and Sanofi-Genzyme. Massimo Filippi is Editor-in-Chief of the Journal of Neurology; received compensation for consulting services and/or speaking activities from Bayer, Biogen Idec, Merck-Serono, Novartis, Roche, Sanofi Genzyme, Takeda, and Teva Pharmaceutical Industries; and receives research support from Biogen Idec, Merck-Serono, Novartis, Roche, Teva Pharmaceutical Industries, Italian Ministry of Health, Fondazione Italiana Sclerosi Multipla, and ARiSLA (Fondazione Italiana di Ricerca per la SLA). Federica Esposito has received compensation for consulting services and/or speaking activities from Novartis, Sanofi Genzyme, Almirall, TEVA, and Merck-Serono.

## References

1. Thompson, A.J.; Baranzini, S.E.; Geurts, J.; Hemmer, B.; Ciccarelli, O. Multiple Sclerosis. *Lancet* **2018**, *391*, 1622–1636. [[CrossRef](#)] [[PubMed](#)]
2. Montalban, X.; Gold, R.; Thompson, A.J.; Otero-Romero, S.; Amato, M.P.; Chandraratna, D.; Clanet, M.; Comi, G.; Derfuss, T.; Fazekas, F.; et al.ECTRIMS/EAN Guideline on the Pharmacological Treatment of People with Multiple Sclerosis. *Eur. J. Neurol.* **2018**, *25*, 215–237. [[CrossRef](#)] [[PubMed](#)]
3. Patsopoulos, N.A.; Baranzini, S.E.; Santaniello, A.; Shoostari, P.; Cotsapas, C.; Wong, G.; Beecham, A.H.; James, T.; Replogle, J.; Vlachos, I.S.; et al. Multiple Sclerosis Genomic Map Implicates Peripheral Immune Cells & Microglia in Susceptibility. *Science* **2019**, *365*, 50. [[CrossRef](#)]
4. Hauser, S.L.; Oksenberg, J.R.; Lincoln, R.; Garovoy, J.; Beck, R.W.; Cole, S.R.; Moke, P.S.; Kip, K.E.; Gal, R.L.; Long, D.T. Interaction between HLA-DR2 and Abnormal Brain MRI in Optic Neuritis and Early MS. *Neurology* **2000**, *54*, 1859–1861. [[CrossRef](#)]
5. Barcellos, L.F.; Oksenberg, J.R.; Begovich, A.B.; Martin, E.R.; Schmidt, S.; Vittinghoff, E.; Goodin, D.S.; Pelletier, D.; Lincoln, R.R.; Bucher, P.; et al. HLA-DR2 Dose Effect on Susceptibility to Multiple Sclerosis and Influence on Disease Course. *Am. J. Hum. Genet.* **2003**, *72*, 710–716. [[CrossRef](#)]

6. Briggs, F.B.S.; Shao, X.; Goldstein, B.A.; Oksenberg, J.R.; Barcellos, L.F.; De Jager, P.L. Genome-Wide Association Study of Severity in Multiple Sclerosis. *Genes Immun.* **2011**, *12*, 615–625. [[CrossRef](#)]
7. Comabella, M.; Craig, D.W.; Morcillo-Suárez, C.; Río, J.; Navarro, A.; Fernández, M.; Martín, R.; Montalban, X. Genome-Wide Scan of 500 000 Single-Nucleotide Polymorphisms Among Responders and Nonresponders to Interferon Beta Therapy in Multiple Sclerosis. *Arch. Neurol.* **2009**, *66*, 972–978. [[CrossRef](#)]
8. Esposito, F.; Sorosina, M.; Ottoboni, L.; Lim, E.T.; Replogle, J.M.; Raj, T.; Brambilla, P.; Liberatore, G.; Guaschino, C.; Romeo, M.; et al. A Pharmacogenetic Study Implicates SLC9a9 in Multiple Sclerosis Disease Activity. *Ann. Neurol.* **2015**, *78*, 115–127. [[CrossRef](#)]
9. Clarelli, F.; Liberatore, G.; Sorosina, M.; Osiceanu, A.M.; Esposito, F.; Mascia, E.; Santoro, S.; Pavan, G.; Colombo, B.; Moiola, L.; et al. Pharmacogenetic Study of Long-Term Response to Interferon- $\beta$  Treatment in Multiple Sclerosis. *Pharm. J.* **2017**, *17*, 84–91. [[CrossRef](#)]
10. Mahurkar, S.; Moldovan, M.; Suppiah, V.; Sorosina, M.; Clarelli, F.; Liberatore, G.; Malhotra, S.; Montalban, X.; Antigüedad, A.; Krupa, M.; et al. Response to Interferon-Beta Treatment in Multiple Sclerosis Patients: A Genome-Wide Association Study. *Pharm. J.* **2017**, *17*, 312–318. [[CrossRef](#)]
11. Grossman, I.; Avidan, N.; Singer, C.; Goldstaub, D.; Hayardeny, L.; Eyal, E.; Ben-Asher, E.; Paperna, T.; Pe'er, I.; Lancet, D.; et al. Pharmacogenetics of Glatiramer Acetate Therapy for Multiple Sclerosis Reveals Drug-Response Markers. *Pharmacogenet. Genom.* **2007**, *17*, 657–666. [[CrossRef](#)]
12. Kulakova, O.; Bashinskaya, V.; Kiselev, I.; Baulina, N.; Tsareva, E.; Nikolaev, R.; Kozin, M.; Shchur, S.; Favorov, A.; Boyko, A.; et al. Pharmacogenetics of Glatiramer Acetate Therapy for Multiple Sclerosis: The Impact of Genome-Wide Association Studies Identified Disease Risk Loci. *Pharmacogenomics* **2017**, *18*, 1563–1574. [[CrossRef](#)]
13. Tsareva, E.Y.; Kulakova, O.G.; Boyko, A.N.; Shchur, S.G.; Lvovs, D.; Favorov, A.V.; Gusev, E.I.; Vandenbroeck, K.; Favorova, O.O. Allelic Combinations of Immune-Response Genes Associated with Glatiramer Acetate Treatment Response in Russian Multiple Sclerosis Patients. *Pharmacogenomics* **2012**, *13*, 43–53. [[CrossRef](#)]
14. Dominguez-Mozo, M.I.; Perez-Perez, S.; Villar, L.M.; Oliver-Martos, B.; Villarrubia, N.; Matesanz, F.; Costa-Frossard, L.; Pinto-Medel, M.J.; García-Sánchez, M.I.; Ortega-Madueño, I.; et al. Predictive Factors and Early Biomarkers of Response in Multiple Sclerosis Patients Treated with Natalizumab. *Sci. Rep.* **2020**, *10*, 14244. [[CrossRef](#)] [[PubMed](#)]
15. Alvarez-Lafuente, R.; Blanco-Kelly, F.; Garcia-Montojo, M.; Martínez, A.; De Las Heras, V.; Dominguez-Mozo, M.I.; Bartolome, M.; Garcia-Martinez, A.; De la Concha, E.G.; Urcelay, E.; et al. CD46 in a Spanish Cohort of Multiple Sclerosis Patients: Genetics, mRNA Expression and Response to Interferon-Beta Treatment. *Mult. Scler.* **2011**, *17*, 513–520. [[CrossRef](#)] [[PubMed](#)]
16. Ferrè, L.; Mogavero, A.; Clarelli, F.; Moiola, L.; Sangalli, F.; Colombo, B.; Martinelli, V.; Comi, G.; Filippi, M.; Esposito, F. Early Evidence of Disease Activity during Fingolimod Predicts Medium-Term Inefficacy in Relapsing-Remitting Multiple Sclerosis. *Mult. Scler.* **2021**, *27*, 1374–1383. [[CrossRef](#)] [[PubMed](#)]
17. Rajkomar, A.; Dean, J.; Kohane, I. Machine Learning in Medicine. *N. Engl. J. Med.* **2019**, *380*, 1347–1358. [[CrossRef](#)] [[PubMed](#)]
18. Ho, D.S.W.; Schierding, W.; Wake, M.; Saffery, R.; O'Sullivan, J. Machine Learning SNP Based Prediction for Precision Medicine. *Front. Genet.* **2019**, *10*, 267. [[CrossRef](#)]
19. Esposito, F.; Ferrè, L.; Clarelli, F.; Rocca, M.A.; Sferruzza, G.; Storelli, L.; Radaelli, M.; Sangalli, F.; Moiola, L.; Colombo, B.; et al. Effectiveness and Baseline Factors Associated to Fingolimod Response in a Real-World Study on Multiple Sclerosis Patients. *J. Neurol.* **2018**, *265*, 896–905. [[CrossRef](#)]
20. Jokubaitis, V.G.; Li, V.; Kalincik, T.; Izquierdo, G.; Hodgkinson, S.; Alroughani, R.; Lechner-Scott, J.; Lugaresi, A.; Duquette, P.; Girard, M.; et al. Fingolimod after Natalizumab and the Risk of Short-Term Relapse. *Neurology* **2014**, *82*, 1204–1211. [[CrossRef](#)]
21. Rotstein, D.L.; Healy, B.C.; Malik, M.T.; Chitnis, T.; Weiner, H.L. Evaluation of No Evidence of Disease Activity in a 7-Year Longitudinal Multiple Sclerosis Cohort. *JAMA Neurol.* **2015**, *72*, 152. [[CrossRef](#)]
22. Chang, C.C.; Chow, C.C.; Tellier, L.C.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *Gigascience* **2015**, *4*, 7. [[CrossRef](#)] [[PubMed](#)]
23. Ambroise, C.; McLachlan, G.J. Selection Bias in Gene Extraction on the Basis of Microarray Gene-Expression Data. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6562–6566. [[CrossRef](#)]
24. Goldstein, B.A.; Polley, E.C.; Briggs, F.B.S. Random Forests for Genetic Association Studies. *Stat. Appl. Genet. Mol. Biol.* **2011**, *10*, 32. [[CrossRef](#)]
25. Zhuang, X.; Ye, R.; So, M.T.; Lam, W.Y.; Karim, A.; Yu, M.; Ngo, N.D.; Cherny, S.S.; Tam, P.K.H.; Garcia-Barcelo, M.M.; et al. A Random Forest-Based Framework for Genotyping and Accuracy Assessment of Copy Number Variations. *NAR Genom. Bioinforma.* **2020**, *2*, lqaa071. [[CrossRef](#)] [[PubMed](#)]
26. Gonzalez, A.; Vazquez, D.; Lopez, A.M.; Amores, J. On-Board Object Detection: Multicue, Multimodal, and Multiview Random Forest of Local Experts. *IEEE Trans. Cybern.* **2017**, *47*, 3980–3990. [[CrossRef](#)]
27. Liao, Y.; Wang, J.; Jaehnig, E.J.; Shi, Z.; Zhang, B. WebGestalt 2019: Gene Set Analysis Toolkit with Revamped UIs and APIs. *Nucleic Acids Res.* **2019**, *47*, W199–W205. [[CrossRef](#)]
28. Río, J.; Comabella, M.; Montalban, X. Predicting Responders to Therapies for Multiple Sclerosis. *Nat. Rev. Neurol.* **2009**, *5*, 553–560. [[CrossRef](#)]
29. Danelakis, A.; Theoharis, T.; Verganelakis, D.A. Survey of Automated Multiple Sclerosis Lesion Segmentation Techniques on Magnetic Resonance Imaging. *Comput. Med. Imaging Graph.* **2018**, *70*, 83–100. [[CrossRef](#)] [[PubMed](#)]

30. Gabr, R.E.; Coronado, I.; Robinson, M.; Sujit, S.J.; Datta, S.; Sun, X.; Allen, W.J.; Lublin, F.D.; Wolinsky, J.S.; Narayana, P.A. Brain and Lesion Segmentation in Multiple Sclerosis Using Fully Convolutional Neural Networks: A Large-Scale Study. *Mult. Scler.* **2020**, *26*, 1217–1226. [[CrossRef](#)]
31. Eshaghi, A.; Riyahi-Alam, S.; Saeedi, R.; Roostaei, T.; Nazeri, A.; Aghsaei, A.; Doosti, R.; Ganjgahi, H.; Bodini, B.; Shakourirad, A.; et al. Classification Algorithms with Multi-Modal Data Fusion Could Accurately Distinguish Neuromyelitis Optica from Multiple Sclerosis. *NeuroImage. Clin.* **2015**, *7*, 306–314. [[CrossRef](#)] [[PubMed](#)]
32. Eshaghi, A.; Wottschel, V.; Cortese, R.; Calabrese, M.; Sahraian, M.A.; Thompson, A.J.; Alexander, D.C.; Ciccarelli, O. Gray Matter MRI Differentiates Neuromyelitis Optica from Multiple Sclerosis Using Random Forest. *Neurology* **2016**, *87*, 2463–2470. [[CrossRef](#)] [[PubMed](#)]
33. Kim, H.; Lee, Y.; Kim, Y.H.; Lim, Y.M.; Lee, J.S.; Woo, J.; Jang, S.K.; Oh, Y.J.; Kim, H.W.; Lee, E.J.; et al. Deep Learning-Based Method to Differentiate Neuromyelitis Optica Spectrum Disorder From Multiple Sclerosis. *Front. Neurol.* **2020**, *11*, 599042. [[CrossRef](#)] [[PubMed](#)]
34. Pinto, M.F.; Oliveira, H.; Batista, S.; Cruz, L.; Pinto, M.; Correia, I.; Martins, P.; Teixeira, C. Prediction of Disease Progression and Outcomes in Multiple Sclerosis with Machine Learning. *Sci. Rep.* **2020**, *10*, 21038. [[CrossRef](#)]
35. De Brouwer, E.; Becker, T.; Moreau, Y.; Havrdova, E.K.; Trojano, M.; Eichau, S.; Ozakbas, S.; Onofrij, M.; Grammond, P.; Kuhle, J.; et al. Longitudinal Machine Learning Modeling of MS Patient Trajectories Improves Predictions of Disability Progression. *Comput. Methods Programs Biomed.* **2021**, *208*, 106180. [[CrossRef](#)]
36. Ross, C.J.; Towfic, F.; Shankar, J.; Laifenfeld, D.; Thoma, M.; Davis, M.; Weiner, B.; Kusko, R.; Zeskind, B.; Knappertz, V.; et al. A Pharmacogenetic Signature of High Response to Copaxone in Late-Phase Clinical-Trial Cohorts of Multiple Sclerosis. *Genome Med.* **2017**, *9*, 50. [[CrossRef](#)]
37. Bin Rafiq, R.; Modave, F.; Guha, S.; Albert, M.V. Validation Methods to Promote Real-World Applicability of Machine Learning in Medicine. In Proceedings of the 2020 3rd International Conference on Digital Medicine and Image Processing, Kyoto, Japan, 6–9 November 2020; pp. 13–19. [[CrossRef](#)]
38. Singh, V.; Pencina, M.; Einstein, A.J.; Liang, J.X.; Berman, D.S.; Slomka, P. Impact of Train/Test Sample Regimen on Performance Estimate Stability of Machine Learning in Cardiovascular Imaging. *Sci. Rep.* **2021**, *11*, 14490. [[CrossRef](#)]
39. Berisha, V.; Krantsevich, C.; Hahn, P.R.; Hahn, S.; Dasarathy, G.; Turaga, P.; Liss, J. Digital Medicine and the Curse of Dimensionality. *NPJ Digit. Med.* **2021**, *4*, 153. [[CrossRef](#)]
40. Gliozzo, J.; Mesiti, M.; Notaro, M.; Petrini, A.; Patak, A.; Puertas-Gallardo, A.; Paccanaro, A.; Valentini, G.; Casiraghi, E. Heterogeneous Data Integration Methods for Patient Similarity Networks. *Brief. Bioinform.* **2022**, *23*, bbac207. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.