




## Article

# A Linear Predictor Based on FTIR Spectral Biomarkers Improves Disease Diagnosis Classification: An Application to Multiple Sclerosis

Francesca Condino <sup>1</sup>, Maria Caterina Crocco <sup>2,3</sup>, Domenico Pirritano <sup>4,5</sup>, Alfredo Petrone <sup>5</sup>,  
Francesco Del Giudice <sup>5,6</sup> and Rita Guzzi <sup>2,7,\*</sup>

- <sup>1</sup> Department of Economics, Statistics and Finance "Giovanni Anania", University of Calabria, 87036 Rende, Italy; francesca.condino@unical.it
- <sup>2</sup> STAR Research Infrastructure, University of Calabria, 87036 Rende, Italy; mariacaterina.crocco@unical.it
- <sup>3</sup> Department of Physics, Molecular Biophysics Laboratory, University of Calabria, 87036 Rende, Italy
- <sup>4</sup> SOC Neurologia, Azienda Ospedaliero-Universitaria Renato Dulbecco, 88100 Catanzaro, Italy; pirritanodomenico@gmail.com
- <sup>5</sup> UOC Neurologia, Azienda Ospedaliera dell'Annunziata, 87100 Cosenza, Italy; a.petrone@aocs.it (A.P.); fra.delgiu@libero.it (F.D.G.)
- <sup>6</sup> SOC Neurologia, Ospedale Jazzolino, Azienda Ospedaliera Provinciale, 89900 Vibo Valentia, Italy
- <sup>7</sup> CNR-NANOTEC, Department of Physics, University of Calabria, 87036 Rende, Italy
- \* Correspondence: rita.guzzi@fis.unical.it

**Abstract:** Multiple sclerosis (MS) is a neurodegenerative disease of the central nervous system that can lead to long-term disability. The diagnosis of MS is not simple and requires many instrumental and clinical tests. Sampling easily collected biofluids using spectroscopic approaches is becoming of increasing interest in the medical field to integrate and improve diagnostic procedures. Here we present a statistical approach where we combine a number of spectral biomarkers derived from the ATR-FTIR spectra of blood plasma samples of healthy control subjects and MS patients, to obtain a linear predictor useful for discriminating between the two groups of individuals. This predictor provides a simple tool in which the contribution of different molecular components is summarized and, as a result, the sensitivity (80%) and specificity (93%) of the identification are significantly improved compared to those obtained with typical classification algorithms. The strategy proposed can be very helpful when applied to the diagnosis of diseases whose presence is reflected in a minimal way in the analyzed biofluids (blood and its derivatives), as it is for MS as well as for other neurological disorders.

**Keywords:** multiple sclerosis; plasma; ATR-FTIR; multivariate analysis; discrimination; logistic regression



**Citation:** Condino, F.; Crocco, M.C.; Pirritano, D.; Petrone, A.; Del Giudice, F.; Guzzi, R. A Linear Predictor Based on FTIR Spectral Biomarkers Improves Disease Diagnosis Classification: An Application to Multiple Sclerosis. *J. Pers. Med.* **2023**, *13*, 1596. <https://doi.org/10.3390/jpm13111596>

Academic Editor: Francis L. Martin

Received: 27 September 2023

Revised: 2 November 2023

Accepted: 9 November 2023

Published: 11 November 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/>).

## 1. Introduction

Multiple sclerosis (MS) is an immune-mediated demyelinating and inflammation disorder of the central nervous system (CNS) and is the major cause of non-traumatic neurological disability in young adults. The latest update on the prevalence of the disease reports that there are about 2.9 million people living with MS worldwide, with a higher prevalence in higher income areas [1,2]. Risk factors are not well defined and involve interplay between genetic, lifestyle and environmental factors [3]. There is a gender prevalence in MS, and women are more affected than men in a 3 to 1 ratio. An important feature of this disease is that the clinical course can be highly variable and heterogeneous in the type of clinical manifestations, course of the disease, and degree of disability [4]. The disability level is generally quantified via the expanded disability status scale (EDSS), a numeric index which ranges from 0 (normal status) to 10, with a variation step of 0.5. It

is assigned by a neurologist using an empirical assessment, and evaluates neurological findings in eight functional systems, as well as the walking ability of the patient [5].

MS diagnosis is not very easy, as the initial symptoms of the disease may be generic and nonspecific, including changes in vision, mobility, and cognitive abilities [6]. There is no single test for the diagnosis, but diagnostic criteria have been validated (namely, the McDonald criteria revised in 2017), based on clinical observation of the patient, as well as laboratory and instrumental tests such as magnetic resonance imaging and oligoclonal bands in the cerebrospinal fluid (CSF) [7]. In particular, evidence of neurological damage disseminated in time and space in multiple regions of the CNS should be seen. Blood tests are generally used to rule out diseases that resemble MS and might be confused with it. The overall protocol requires long time for the tests and evaluations to take place, which are certainly not pleasant for the patients. The ultimate goal is to make an early and accurate diagnosis in order to provide effective drug treatment to control the disease, which currently remains incurable.

Recently, the application of analytical and spectroscopic methods to analyze human biofluids (e.g., plasma, serum, saliva, urine, bile), searching for biomarkers or even aiming at early stage disease detection, is becoming an important issue in the biomedical research field. Such methods include nuclear magnetic resonance (NMR) [8,9], mass spectrometry [10], differential scanning calorimetry [11–13], and Raman and infrared spectroscopy [14–16]. These experimental methods are often supported by a number of multivariate statistical approaches in the analysis of their data. Vibrational spectroscopy can be applied in different modalities and has shown great potential to classify, either as healthy or pathological, a number of biological materials, such as cells, tissues, blood, saliva, and for many different diseases [17]. The samples' preparation requires minimal or no treatment, and experiments are rapid and label-free. These advantages have contributed to a wide range of applications, and studies for diagnostic purposes have been reported for different forms of cancer [18,19], COVID-19 [20–22], systemic amyloidosis [23], as well as neurodegenerative diseases [24,25]. Despite infrared spectroscopy having demonstrated its potential as a diagnostic tool, the translation of these experimental and computational techniques into clinical practice does not seem so straightforward. Standardized protocols should be implemented and established at different steps, including sample acquisition and storage, condition measurements, data correction, and analysis [17,26,27].

In a previous study, we applied attenuated total reflection Fourier transform IR (ATR-FTIR) spectroscopy to the plasma of 85 subjects: 45 MS patients with different clinical records in terms of disability level and time duration of the disease and 40 healthy controls (HC) with no inflammatory or autoimmune disease diagnosis [28]. Although CSF may be the ideal source of biomarkers for MS, it is not an easy biofluid to be collected, and the availability of healthy samples for comparison is also a limiting factor. In contrast, blood can be routinely collected; it perfuses all body organs, and also exchanges molecules (potential biomarkers) with the CSF.

The analysis of our experimental data showed that the spectral features of absorbance spectra for HC and MS are similar to each other, with the latter having a slightly higher variability [28]. Using univariate and multivariate statistical approaches on the ATR-FTIR spectra, it was found that: i) MS patients were characterized by a higher lipid/protein ratio compared to HC individuals; and ii) a random forest algorithm was able to select regions for discriminating the disease occurrence with a good degree of accuracy (sensitivity 78%; specificity 83%) with respect to other two classification algorithms (PCA-LDA and PLS-DA).

In this work, we have proposed a new data analysis based on a linear combination of a number of spectral biomarkers, derived from those experimental spectra, including some previously selected. We showed that the predictive model obtained with this approach significantly improves the performance of the classification, especially for the specificity index. The advantage of using this model is that it combines the most important spectral indicators that jointly contribute to the samples' classification, through a simple expression

giving a single score value, and can be easily used to establish whether the subject may likely belong to the class of healthy or diseased individuals.

## 2. Materials and Methods

### 2.1. Biological Samples and Spectral Acquisition

Blood samples of 45 MS patients with a confirmed diagnosis according to the McDonald criteria [7] were collected at the Annunziata Hospital in Cosenza (Italy). The criteria of inclusion and the degree of disease progression were established by the neurologist. A more detailed description of the clinical and personal information of the patients can be found in ref. [28]. Briefly, the data included: age (average 42.7, range 22–69), gender (31 females, 14 males), disease phenotype/form (38 relapsing-remitting multiple sclerosis, RRMS; 7 secondary-progressive multiple sclerosis, SPMS), disease duration (from 1 to 47 years), disability degree expressed in terms of EDSS (range 0.5–7.0). Most of the MS patients (32/45, 71%) had a mild score ( $EDSS \leq 3.0$ ) and the others (13/45, 29%) had a moderate/severe level of disease ( $EDSS \geq 3.5$ ).

HC subjects (n. 40) were recruited at the same Annunziata Hospital (n. 21) and among the blood donors at the Health Center of the University of Calabria (n. 19). The gender distribution was 21 females and 19 males, with an average age of 37.3 years (range 24–60 years, one outlier of 71 years). For all of them there was no evidence of neurological or inflammatory disease.

All the participants were fully informed about the study and provided their written consent. The study was approved by the Ethics Committee of the northern area of the Calabria region (protocol code n. 50, 14 February 2017).

ATR-FTIR experiments were performed as reported in ref. [28]. For each plasma sample, up to five replicate spectra were acquired and then averaged to obtain the absorbance spectrum of each individual. The regions selected for the statistical data analysis were the fingerprint region ( $1800\text{--}900\text{ cm}^{-1}$ ) and high region ( $3050\text{--}2800\text{ cm}^{-1}$ ).

### 2.2. Computational Analysis

Statistical differences in the vector normalized absorbance spectra of MS and HC groups were assessed using a two-tailed parametric *t*-test. A *p*-value  $\leq 0.05$  was considered significant in all statistical tests.

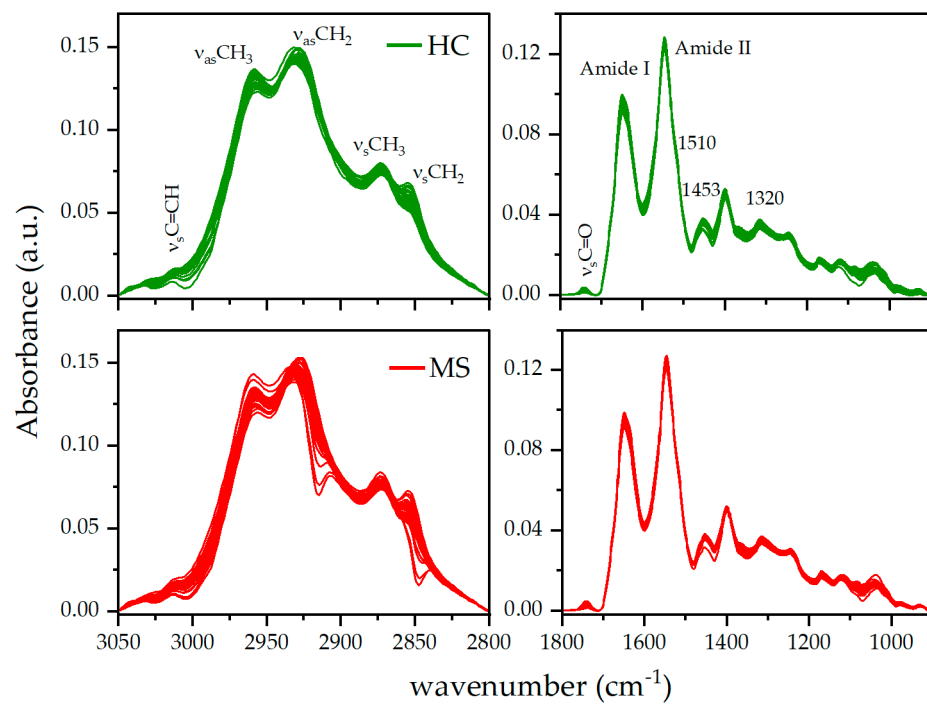
A number of spectral indicators were selected from the absorbance spectra based on their biological significance in this specific study, and also considering those commonly identified in the FTIR spectra of biofluids [25,28–31]. The considered spectral biomarkers are listed in Table 1, and are used to construct the final linear predictor for distinguishing between healthy controls and patients. To this end, a generalized linear model (see, for example, the data analysis of Agresti [32]) is considered to relate the *p* spectral biomarkers  $\mathbf{x} = (x_1, \dots, x_p)$  to the probability  $\pi(\mathbf{x})$  of having the disease. In particular, a logit link function is considered, as the response is a binary variable, and the resulting logistic regression model can be expressed as follows:

$$\pi(\mathbf{x}) = \frac{\exp\left(\beta_0 + \sum_{j=1}^p \beta_j x_j\right)}{1 + \exp\left(\beta_0 + \sum_{j=1}^p \beta_j x_j\right)} \quad (1)$$

or, alternatively:

$$\text{logit}[\pi(\mathbf{x})] = \beta_0 + \sum_{j=1}^p \beta_j x_j \quad (2)$$

where  $\beta_0$  represents the intercept of the model.



**Figure 1.** ATR-FTIR spectra of plasma of HC and MS patients in the high region (**left panels**) and fingerprint region (**right panels**). Pre-processing spectral analysis includes cut, rubberband baseline subtraction, and vector normalization. The stretching vibration of specific functional groups and peaks of interest are also shown (see text for details).

**Table 1.** Spectral indicators derived from data in Figure 1, indicating the area ratios of the lipid/protein components, intensity ratio of specific peaks; bandwidth of the ester C=O at 1739 cm<sup>-1</sup> and of the amide I peak; intensity values in the fingerprint region (1320 and 1510 cm<sup>-1</sup>) and in the high region (2860 and 3016 cm<sup>-1</sup>). Values are given as average ± standard deviation. The *p*-value is obtained via unpaired *t*-test. The last column refers to studies with a similar choice of the indicated spectral regions and wavenumber of functional groups.

	HC	MS	<i>p</i> -Value	Refs
$A_{HR}/A_{amide\ I + amide\ II}$	1.0552 ± 0.0107	1.0621 ± 0.0114	0.005	[25,28,31]
$A_{C=O}/A_{HR}$	0.0036 ± 0.0009	0.0042 ± 0.0013	0.028	[29]
$A_{C=O}/A_{1453}$	0.0341 ± 0.0080	0.0388 ± 0.0120	0.038	[28]
$A_{HR+C=O}/A_{amide\ I + amide\ II}$	1.0591 ± 0.0112	1.0666 ± 0.0122	0.004	[29]
$A_{CH2asym}/A_{CH2sym + CH2asym}$	0.8750 ± 0.0068	0.8744 ± 0.0101	0.776	[31]
$A_{CH2asym}/A_{HR}$	0.4423 ± 0.0059	0.4402 ± 0.0097	0.230	[29]
$A_{Olefinic}/A_{HR}$	0.0110 ± 0.0012	0.0115 ± 0.0015	0.106	[29]
$A_{amide\ I}/A_{amide\ II}$	0.7250 ± 0.0165	0.7115 ± 0.0136	<0.001	[29]
$I_{1453}/I_{1650}$	0.3719 ± 0.0166	0.3834 ± 0.0148	0.001	[25,28]
$I_{1739}/I_{1468}$	0.0702 ± 0.0158	0.0796 ± 0.0229	0.034	[28]
Bandwidth <sub>C=O</sub>	16.0404 ± 1.1338	15.9344 ± 1.0337	0.653	[28]
Bandwidth <sub>amide\ I</sub>	36.5558 ± 0.7052	36.3905 ± 0.7795	0.310	[28]
$I_{1320}$	0.0342 ± 0.0012	0.0349 ± 0.0010	0.003	[28]
$I_{1510}$	0.0559 ± 0.0010	0.0563 ± 0.0008	0.024	[25,28]
$I_{2860}$	0.0597 ± 0.0030	0.0616 ± 0.0031	0.004	[28]
$I_{3016}$	0.0120 ± 0.0013	0.0128 ± 0.0016	0.016	[28]

To select the relevant covariates, a backward stepwise selection approach was used for removing the weakest related variables, based on the probability of the Wald statistic. The maximum likelihood estimation method was considered to obtain the regression coefficients,  $\beta_j$  ( $j = 0, \dots, p$ ), and the Wald test was used to assess their significance.

The obtained estimated linear predictor can be used to classify each subject as diseased or healthy, according to whether it assumes a positive or a negative value or, equivalently, whether the corresponding estimated probability is higher or lower than the cutoff of 0.5.

The performance of the final model in predicting MS diagnosis was evaluated through the usual metrics, namely sensitivity, specificity, and accuracy. Furthermore, the receiver operating characteristic (ROC) curve was obtained and the area under the curve (AUC) is considered a measure of the predictive power of the method. Finally, a bootstrap method was used to test the difference in the AUC of the ROC curves among patients with different EDSS scores or disease duration.

### 3. Results

#### 3.1. ATR-FTIR Spectra of MS and HC Plasma Samples: Identification of the Spectral Biomarkers

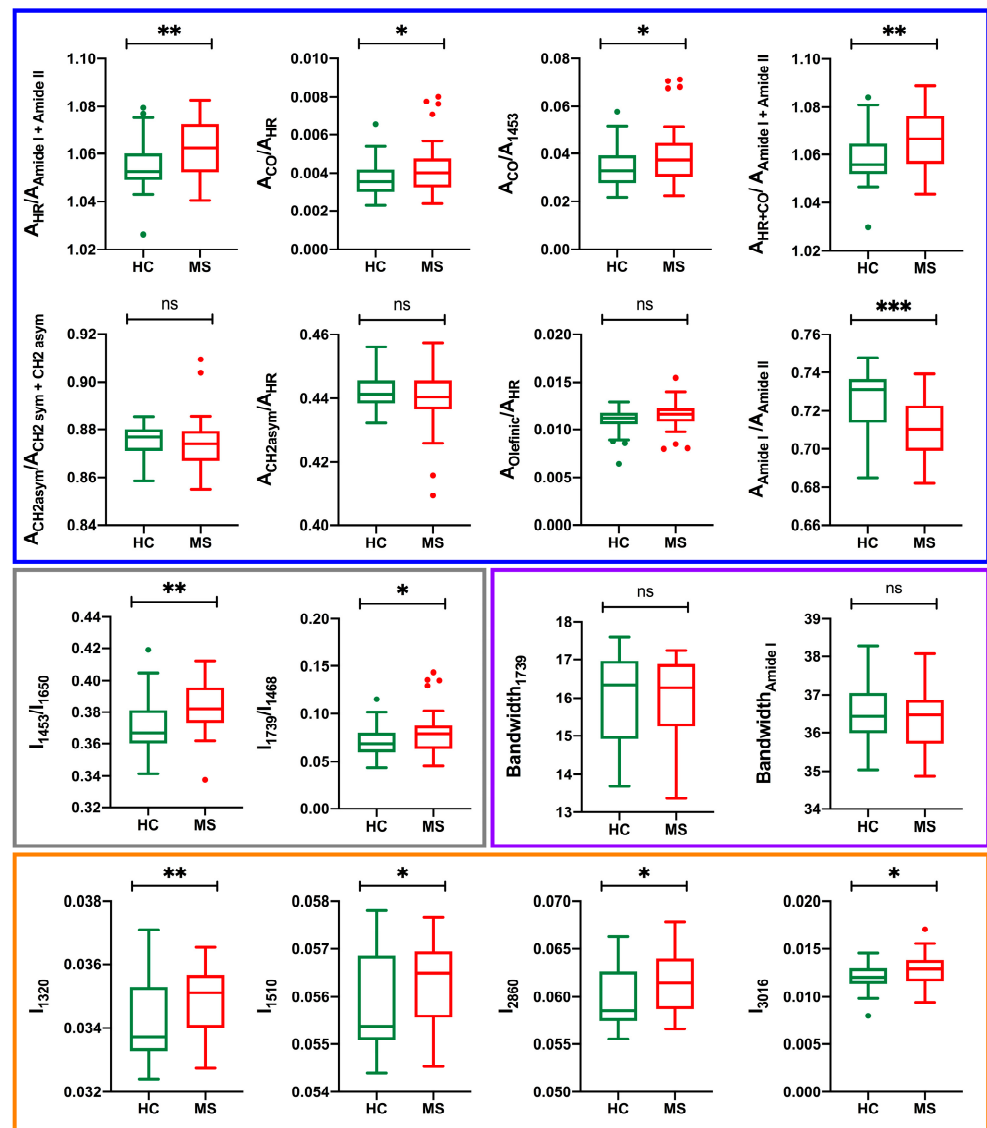
Figure 1 shows the ATR-FTIR spectra of MS patients and HC in the two most significant regions: 3050–2800  $\text{cm}^{-1}$ , containing the symmetric and asymmetric stretching vibrations of  $\text{CH}_2$  and  $\text{CH}_3$ , and the fingerprint region in the 1800–900  $\text{cm}^{-1}$  range, where the most prominent peaks are the amide I and amide II peaks that are due to the absorption band of proteins.

The visual inspection of the spectra does not provide any information about key specific features of MS disease. To reveal the differences between the two groups of individuals, we can determine a number of spectral indicators correlated both with the molecular components present in the plasma and with specific peaks associated with molecular functional groups. To this end, in the present analysis, we consider several spectral indicators (see Table 1), some of them previously derived [28] and others determined in the current analysis using the same experimental spectra. Similar indicators were considered in previous studies (see the references in Table 1) and describe (1) the area ratios of molecular components such as lipid/proteins determined as the area,  $A_{\text{HR}}$ , corresponding to the region between 3050 and 2800  $\text{cm}^{-1}$  and the area corresponding to the amide I and amide II peak,  $A_{\text{HR}}/A_{\text{amideI+amideII}}$ , the area of the olefinic  $\text{C}=\text{CH}$  peak over  $A_{\text{HR}}$ ,  $A_{\text{olefinic}}/A_{\text{HR}}$ , the area of the asymmetric stretching vibration peak of  $\text{CH}_2$ ,  $A_{\text{CH}_2\text{as}}/A_{\text{HR}}$ , and  $A_{\text{amide I}}/A_{\text{amide II}}$ ; (2) the intensity ratio of specific peaks of functional groups,  $I_{1453}/I_{1650}$ ,  $I_{1739}/I_{1468}$ ; (3) the bandwidth of amide I and of the ester  $\text{C}=\text{O}$  peak at 1739  $\text{cm}^{-1}$ ; and (4) the intensity at 1320, 1510, 2860, and 3016  $\text{cm}^{-1}$ , selected as important wavenumbers for the discrimination of MS from HC according to the random forest classification algorithm. Bandwidth values were determined at 75% of the height of the maximum peak intensity from the baseline.

As a preliminary statistical evaluation, we determined the average and standard deviation of these spectral indicators, and assessed the mean differences between MS and HC using two-tailed *t*-test. The *p*-values reported in Table 1 show that most of the spectral parameters have a statistical significance ( $p < 0.05$ ), but for five of them the difference between the two groups of subjects is not significant. In Figure 2 the box-plots of each individual parameter are reported.

#### 3.2. Linear Predictor for Classification

As a further step, our goal was to simultaneously select and combine the most significant spectral indicators described in the above section into a concise predictive tool for diagnostic purposes. In this respect, as described in Section 2.2, we started from the full model containing all the spectral indicators reported in Table 1, and obtained the final model via a backward stepwise selection to remove the irrelevant indicators. The remaining variables, which mainly contributed to the classification of the two groups of subjects, are the six indicators reported in Table 2. In the same table, the fitted coefficients of the linear combination described in Equation (2) are given, together with their standard error and the corresponding results from the Wald test.



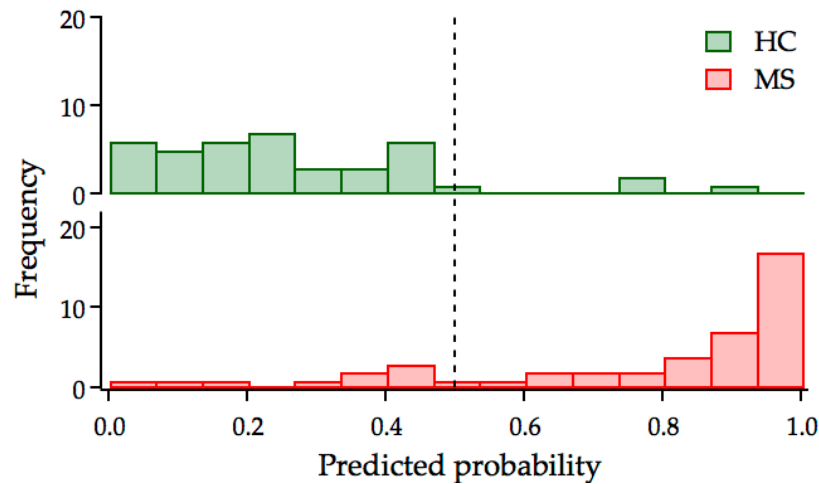
**Figure 2.** Box-plot of the spectral parameters derived from the analysis of the ATR-FTIR absorption spectra, illustrating the distribution of their values in the HC and MS groups. The *p*-value is indicated as \* (*p* < 0.05); \*\* (*p* < 0.01); \*\*\* (*p* < 0.001).

**Table 2.** Fitted coefficients  $\hat{\beta}_j$  for the indicators in the final model, corresponding standard errors, and results from the Wald test.

	$\hat{\beta}_j$	S.E.	z	<i>p</i> -Value
$A_{CH2asym}/A_{CH2sym + CH2asym}$	659.218	177.363	13.814	<0.001
$A_{amide I}/A_{amide II}$	−200.233	53.352	14.086	<0.001
$I_{1453}/I_{1650}$	58.460	24.677	5.612	0.018
$I_{1320}$	−1897.473	710.464	7.133	0.008
$I_{1510}$	−2128.135	670.065	9.821	0.002
$I_{2860}$	1618.604	498.808	19.530	0.001
Constant	−367.473	139.994	6.890	0.009

Considering the obtained values for the selected indicators, a classification for each subject can be obtained and compared with their real health status, to assess the capability of the model in distinguishing between patients and controls. A comparison of the predicted probability distributions within the HC and MS groups is shown in Figure 3. The plot

clearly shows a very good performance of the model, with the clustering of the HC subjects on the left and the MS on the right of the cutoff dashed line at 0.5. The misclassified subjects in both groups are only a few samples and these are located, respectively, to the right and to the left of the threshold value.



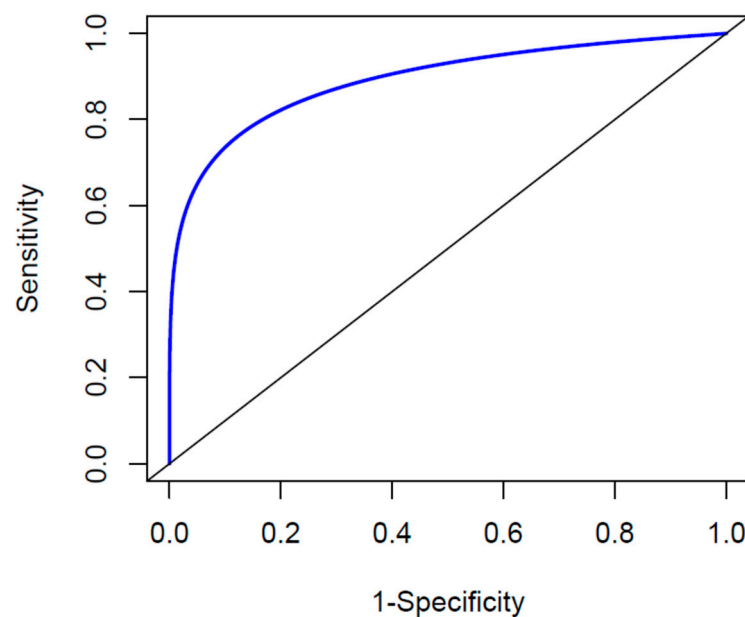
**Figure 3.** Distribution of the predicted probability for HC (upper panel) and MS (bottom panel) individuals. The dashed vertical line indicates the 0.5 score threshold. Score values  $\leq 0.5$  indicate subjects classified as healthy, score values  $\geq 0.5$  indicate subjects classified as diseased.

In Table 3, the values for sensitivity, specificity, accuracy, and AUC are reported for both the whole MS group and separated in subgroups according to EDSS values or disease duration. In all cases, the indicators show the good capability of the fitted model to distinguish between MS and HC, with an accuracy and an AUC higher, respectively, than 85% and 87%. The accuracy of the model is also reflected in the global ROC curve, which is characterized by an AUC value of 0.9 (Figure 4). Moreover, no difference in the AUC of the ROC curves is found when patients with different EDSS scores ( $p$ -value = 0.833) or disease duration ( $p$  = 0.265) are compared.

**Table 3.** Performance of the fitted linear predictor in distinguishing HC from MS, considered either as a whole group or divided in subgroups according to disability level and disease duration.

		Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
		80.00	92.50	85.88	0.90
EDSS	0.5–3.0	81.25	92.50	87.50	0.90
	3.5–7.0	76.90	92.50	88.70	0.89
Disease Duration	$\leq 10$ y	81.50	92.50	88.10	0.87
	$> 10$ y	77.70	92.50	87.90	0.94

Tests of the statistical significance of the classification score values were performed in the HC and MS groups according to age and gender. No significant correlation was found between the linear predictor values and age in the MS ( $p$ -value = 0.521) or control group ( $p$ -value = 0.610), and no difference in the scores’ distribution was observed between males and females (MS:  $p$ -value = 0.983; HC:  $p$ -value = 0.529).



**Figure 4.** Predictive power of the obtained model in distinguishing between HC and MS, as described by a smoothed ROC curve.

#### 4. Discussion

The diagnosis of MS, and in general of neurodegenerative diseases, may be very complex, as there is no clinically validated single test available. The need for new diagnostic tools based on advanced experimental and computational techniques could be of considerable support to neurologists, and allow patients to be treated early with therapies to control these currently incurable diseases. In particular, biofluid spectroscopy is relevant in the field of liquid biopsy for searching and identifying diagnostic biomarkers, with the additional support of multivariate analysis [17,33].

For our investigation, we used ATR-FTIR spectroscopic data obtained from plasma samples collected during the routine clinical observation of MS patients. The FTIR absorbance spectra recorded for the diseased subjects showed similar spectral features compared to those registered for the healthy subjects, both in the high region ( $3100\text{--}2800\text{ cm}^{-1}$ ) and in the fingerprint region ( $1800\text{--}900\text{ cm}^{-1}$ ) of the spectra. This similarity has been found in other FTIR studies on plasma [19,20,30], and accounts for the similarity of the molecular components. To highlight subtle spectral differences between the two classes of subjects, several indicators related to the two most representative molecular components, lipids and proteins, were selected. Almost all of these bioindicators (Table 1) show statistical significance, reveal specific biochemical changes, and, therefore, are good candidates for sample differentiation. In particular, the lipid/protein ratio in the MS group is higher than the HC group's. The increase of the lipid level is observed in the plasma of MS patients and, in parallel, the protein level decreases, as it can be evidenced, for example, by the  $A_{\text{amide I}}/A_{\text{amide II}}$  ratio. The trend of these molecular components is an important signature for MS disease and depends on the biofluids interrogated. In fact, our results agree very well with previous FTIR studies on the serum of MS patients, where a total protein content decrease and a lipid content increase, with respect to controls, was also observed [30]. On the contrary, in the CSF this ratio is reversed [29]. The altered lipid content is a direct consequence of the demyelination process that is a hallmark of MS. Within the lipid analysis it can be also noted that the intensity of the C=O band increases in MS, possibly due to the presence of oxidized lipids produced by free radicals whose concentration increases during inflammatory states [34,35]. Similarly, concerning the protein content in plasma, where albumin and immunoglobulin G (IgG) are the two most abundant proteins, the observed decrease may result from the extravasation of plasma components through damaged vessels. Interestingly, the albumin and IgG quotients, related to their concentration in the



serum and CSF, are a measure of the blood–CSF dysfunction and have been used as an indicator of blood–brain barrier (BBB) integrity [36,37].

In addition to the univariate analysis of the spectral parameters that can potentially be linked to disease-induced biochemical changes, we combined these indicators to construct a new diagnostic index. The result of this procedure is a linear predictor in which only the spectral parameters that jointly contribute most to the sample differentiation of the two classes, HC and MS, are included (Table 2). The six spectral biomarkers that we established in the final model belong to both the infrared regions considered, the high region and the fingerprint region. It is interesting to note that the regression analysis reveals the significant role of the  $A_{\text{CH}_2\text{asym}}/A_{\text{CH}_2\text{sym}+\text{CH}_2\text{asym}}$  ratio in predicting the healthy status, even if the mean difference between HC and MS patients appears not to be significant in the univariate analysis. This is due to the correlation between the considered covariates that, even if it does not pose a multicollinearity problem, acts in determining a suppressor effect (see, for example, [38,39]). This simple predictive tool provides, for our cohort of samples, a better classification in terms of sensitivity, specificity, and accuracy compared to the previous analysis based on random forest, PCA-LDA, and PLS-DA algorithms [28]. In fact, sensitivity increased from 78 to 80% whereas specificity increased from 83 to 92.5%, and the AUC value was 0.9. Compared with the research data on the spectroscopic analysis of biological samples for MS and other neurological diseases, the performance values obtained for our model are, as far as we know, the most favorable. In fact, in MS studies, an AUC value of 0.86 was obtained in a FTIR study on CSF aiming to separate patients with clinically isolated syndrome from RRMS patients [29], whereas slightly lower AUC values (0.82–0.83) were obtained from NMR data [40]. Finally, by using the random forest model on FTIR fingerprint data, an overall precision of 83.3% was found when distinguishing healthy from pathological (MS and amyotrophic lateral sclerosis) serum samples [41]. Considering other neurological diseases, such as Alzheimer's, Paraskevaidi et al. [25] reported specificity and sensitivity values of 70% from the FTIR blood analysis of the fingerprint region. These values increased to 86% when other clinical data were included in the classification model. A high level of performance ( $\geq 90\%$ ) for classification models have been generally reported for the spectroscopic analysis of biological fluids (plasma, serum, saliva) for cancers [14,18,19] and COVID-19 [20,22] detection. Obviously, when the biofluid used is in contact with the pathological region, it is more sensitive in reporting disease-related information. However, peripheral blood and its constituents seem to be the most convenient specimens for diagnostic and biomarker detection. In fact, the challenge with liquid biopsy is the possibility of using readily available biofluids with little or no invasiveness for the patients.

## 5. Conclusions

The results presented in this study have demonstrated that ATR-FTIR spectroscopy applied to human plasma, supported by statistical regression methods, provides valuable spectral biomarkers that can be used for further analysis to improve sample classification in MS disease. A set of parameters was derived from univariate and multivariate statistical analysis of the experimental infrared spectra representing the biochemical fingerprint of the biofluid of HC and MS patients. These spectral biomarkers, characterized by a different statistical significance value, were combined into a linear predictor representing a single, simple diagnostic index whose value varied between 0 and 1. Through a backward stepwise procedure, the irrelevant variables were removed, and only the six most important ones were considered in the final model. Using this model, the predicted probability calculated for our samples provides a better performance than that previously obtained by PCA-LDA, PLS-DA, and RF classification algorithms. It is important to point out that the two approaches are not alternative and, in fact, this new data analysis used spectral biomarker data partially derived in previous work.

To create a more general validity, the logistic regression model should be applied to different infrared spectra for sampling other diseases, in which a sub-optimal level of

classification has been obtained. Moreover, it should be also made clear that the starting set of parameters may be different from the ones we have used here and should be related to the data set available and to the possible biochemical differences of the disease.

Finally, the study we presented, consisting of human plasma analysis with ATR-FTIR spectroscopy, combined with suitable data analysis, is a very promising diagnostic tool for MS, demonstrating that even for neurological diseases it is possible to obtain very effective discrimination results.

**Author Contributions:** Conceptualization, R.G. and F.C.; sample acquisition and clinical evaluation, D.P., F.D.G. and A.P.; methodology, R.G., F.C. and M.C.C.; validation, R.G. and F.C.; visualization, F.C., M.C.C. and R.G.; writing—original draft preparation, R.G. and F.C.; writing—review and editing, R.G., F.C. and M.C.C.; supervision, R.G.; project administration, R.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** This study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the northern area of the Calabria region (protocol n. 50 of 14 February 2017). All methods were performed in accordance with the relevant guidelines and regulations.

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

**Data Availability Statement:** Additional raw data are available from the corresponding author upon reasonable request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. GBD 2016 Multiple Sclerosis Collaborators. Global, Regional, and National Burden of Multiple Sclerosis 1990–2016: A Systematic Analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **2019**, *18*, 269–285. [[CrossRef](#)]
2. Walton, C.; King, R.; Rechtman, L.; Kaye, W.; Leray, E.; Marrie, R.A.; Robertson, N.; La Rocca, N.; Uitdehaag, B.; van der Mei, I.; et al. Rising Prevalence of Multiple Sclerosis Worldwide: Insights from the Atlas of MS, Third Edition. *Mult. Scler.* **2020**, *26*, 1816–1821. [[CrossRef](#)] [[PubMed](#)]
3. Olsson, T.; Barcellos, L.F.; Alfredsson, L. Interactions between Genetic, Lifestyle and Environmental Risk Factors for Multiple Sclerosis. *Nat. Rev. Neurol.* **2017**, *13*, 25–36. [[CrossRef](#)]
4. Reich, D.S.; Lucchinetti, C.F.; Calabresi, P.A. Multiple Sclerosis. *N. Engl. J. Med.* **2018**, *378*, 169–180. [[CrossRef](#)]
5. Macaron, G.; Ontaneda, D. Diagnosis and Management of Progressive Multiple Sclerosis. *Biomedicines* **2019**, *7*, 56. [[CrossRef](#)] [[PubMed](#)]
6. Milo, R. Effectiveness of Multiple Sclerosis Treatment with Current Immunomodulatory Drugs. *Expert Opin. Pharmacother.* **2015**, *16*, 659–673. [[CrossRef](#)] [[PubMed](#)]
7. Thompson, A.J.; Banwell, B.L.; Barkhof, F.; Carroll, W.M.; Coetzee, T.; Comi, G.; Correale, J.; Fazekas, F.; Filippi, M.; Freedman, M.S.; et al. Diagnosis of Multiple Sclerosis: 2017 Revisions of the McDonald Criteria. *Lancet Neurol.* **2018**, *17*, 162–173. [[CrossRef](#)]
8. Gebregiworgis, T.; Powers, R. Application of NMR Metabolomics to Search for Human Disease Biomarkers. *Comb. Chem. High Throughput Screen* **2012**, *15*, 595–610. [[CrossRef](#)]
9. Emwas, A.-H.; Roy, R.; McKay, R.T.; Tenori, L.; Saccenti, E.; Gowda, G.A.N.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; et al. NMR Spectroscopy for Metabolomics Research. *Metabolites* **2019**, *9*, 123. [[CrossRef](#)]
10. Crutchfield, C.A.; Thomas, S.N.; Sokoll, L.J.; Chan, D.W. Advances in Mass Spectrometry-Based Clinical Biomarker Discovery. *Clin. Proteom.* **2016**, *13*, 1. [[CrossRef](#)]
11. Rodrigo, A.; Ojeda, J.L.; Vega, S.; Sanchez-Gracia, O.; Lanás, A.; Isla, D.; Velazquez-Campoy, A.; Abian, O. Thermal Liquid Biopsy (TLB): A Predictive Score Derived from Serum Thermograms as a Clinical Tool for Screening Lung Cancer Patients. *Cancers* **2019**, *11*, 1012. [[CrossRef](#)] [[PubMed](#)]
12. Anesi, F.; Hermoso-Durán, S.; Rizzuti, B.; Bruno, R.; Pirritano, D.; Petrone, A.; Del Giudice, F.; Ojeda, J.; Vega, S.; Sanchez-Gracia, O.; et al. Thermal Liquid Biopsy (TLB) of Blood Plasma as a Potential Tool to Help in the Early Diagnosis of Multiple Sclerosis. *J. Pers. Med.* **2021**, *11*, 295. [[CrossRef](#)]
13. Garbett, N.C.; Miller, J.J.; Jenson, A.B.; Chaires, J.B. Calorimetry Outside the Box: A New Window into the Plasma Proteome. *Biophys. J.* **2008**, *94*, 1377–1383. [[CrossRef](#)] [[PubMed](#)]
14. Baker, M.J.; Hussain, S.R.; Lovergne, L.; Untereiner, V.; Hughes, C.; Lukaszewski, R.A.; Thiéfin, G.; Sockalingum, G.D. Developing and Understanding Biofluid Vibrational Spectroscopy: A Critical Review. *Chem. Soc. Rev.* **2016**, *45*, 1803–1818. [[CrossRef](#)] [[PubMed](#)]

15. Baker, M.J.; Trevisan, J.; Bassan, P.; Bhargava, R.; Butler, H.J.; Dorling, K.M.; Fielden, P.R.; Fogarty, S.W.; Fullwood, N.J.; Heys, K.A.; et al. Using Fourier Transform IR Spectroscopy to Analyze Biological Materials. *Nat. Protoc.* **2014**, *9*, 1771–1791. [[CrossRef](#)] [[PubMed](#)]
16. Naumann, D. Ft-Infrared and Ft-Raman Spectroscopy in Biomedical Research. *Appl. Spectrosc. Rev.* **2001**, *36*, 239–298. [[CrossRef](#)]
17. Theakstone, A.G.; Rinaldi, C.; Butler, H.J.; Cameron, J.M.; Confield, L.R.; Rutherford, S.H.; Sala, A.; Sangamnerkar, S.; Baker, M.J. Fourier-Transform Infrared Spectroscopy of Biofluids: A Practical Approach. *Transl. Biophotonics* **2021**, *3*, e202000025. [[CrossRef](#)]
18. Sala, A.; Anderson, D.J.; Brennan, P.M.; Butler, H.J.; Cameron, J.M.; Jenkinson, M.D.; Rinaldi, C.; Theakstone, A.G.; Baker, M.J. Biofluid Diagnostics by FTIR Spectroscopy: A Platform Technology for Cancer Detection. *Cancer Lett.* **2020**, *477*, 122–130. [[CrossRef](#)]
19. Giamougiannis, P.; Morais, C.L.M.; Rodriguez, B.; Wood, N.J.; Martin-Hirsch, P.L.; Martin, F.L. Detection of Ovarian Cancer ( $\pm$  Neo-Adjuvant Chemotherapy Effects) via ATR-FTIR Spectroscopy: Comparative Analysis of Blood and Urine Biofluids in a Large Patient Cohort. *Anal. Bioanal. Chem.* **2021**, *413*, 5095–5107. [[CrossRef](#)]
20. Banerjee, A.; Gokhale, A.; Bankar, R.; Palanivel, V.; Salkar, A.; Robinson, H.; Shastri, J.S.; Agrawal, S.; Hartel, G.; Hill, M.M.; et al. Rapid Classification of COVID-19 Severity by ATR-FTIR Spectroscopy of Plasma Samples. *Anal. Chem.* **2021**, *93*, 10391–10396. [[CrossRef](#)] [[PubMed](#)]
21. Martinez-Cuazitl, A.; Vazquez-Zapien, G.J.; Sanchez-Brito, M.; Limon-Pacheco, J.H.; Guerrero-Ruiz, M.; Garibay-Gonzalez, F.; Delgado-Macuil, R.J.; de Jesus, M.G.G.; Corona-Perezgrovas, M.A.; Pereyra-Talamantes, A.; et al. ATR-FTIR Spectrum Analysis of Saliva Samples from COVID-19 Positive Patients. *Sci. Rep.* **2021**, *11*, 19980. [[CrossRef](#)]
22. Barauna, V.G.; Singh, M.N.; Barbosa, L.L.; Marcarini, W.D.; Vassallo, P.F.; Mill, J.G.; Ribeiro-Rodrigues, R.; Campos, L.C.G.; Warnke, P.H.; Martin, F.L. Ultrarapid On-Site Detection of SARS-CoV-2 Infection Using Simple ATR-FTIR Spectroscopy and an Analysis Algorithm: High Sensitivity and Specificity. *Anal. Chem.* **2021**, *93*, 2950–2958. [[CrossRef](#)] [[PubMed](#)]
23. Ami, D.; Lavatelli, F.; Rognoni, P.; Palladini, G.; Raimondi, S.; Giorgetti, S.; Monti, L.; Doglia, S.M.; Natalello, A.; Merlini, G. In Situ Characterization of Protein Aggregates in Human Tissues Affected by Light Chain Amyloidosis: A FTIR Microspectroscopy Study. *Sci. Rep.* **2016**, *6*, 29096. [[CrossRef](#)] [[PubMed](#)]
24. Mordechai, S.; Shufan, E.; Katz, B.S.P.; Salman, A. Early Diagnosis of Alzheimer’s Disease Using Infrared Spectroscopy of Isolated Blood Samples Followed by Multivariate Analyses. *Analyst* **2017**, *142*, 1276–1284. [[CrossRef](#)]
25. Paraskevaidi, M.; Morais, C.L.M.; Lima, K.M.G.; Snowden, J.S.; Saxon, J.A.; Richardson, A.M.T.; Jones, M.; Mann, D.M.A.; Allsop, D.; Martin-Hirsch, P.L.; et al. Differential Diagnosis of Alzheimer’s Disease Using Spectrochemical Analysis of Blood. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E7929–E7938. [[CrossRef](#)] [[PubMed](#)]
26. Cameron, J.M.; Butler, H.J.; Anderson, D.J.; Christie, L.; Confield, L.; Spalding, K.E.; Finlayson, D.; Murray, S.; Panni, Z.; Rinaldi, C.; et al. Exploring Pre-Analytical Factors for the Optimisation of Serum Diagnostics: Progressing the Clinical Utility of ATR-FTIR Spectroscopy. *Vib. Spectrosc.* **2020**, *109*, 103092. [[CrossRef](#)]
27. Butler, H.J.; Smith, B.R.; Fritzsche, R.; Radhakrishnan, P.; Palmer, D.S.; Baker, M.J. Optimised Spectral Pre-Processing for Discrimination of Biofluids via ATR-FTIR Spectroscopy. *Analyst* **2018**, *143*, 6121–6134. [[CrossRef](#)] [[PubMed](#)]
28. Crocco, M.C.; Moyano, M.F.H.; Annesi, F.; Bruno, R.; Pirritano, D.; Del Giudice, F.; Petrone, A.; Condino, F.; Guzzi, R. ATR-FTIR Spectroscopy of Plasma Supported by Multivariate Analysis Discriminates Multiple Sclerosis Disease. *Sci. Rep.* **2023**, *13*, 2565. [[CrossRef](#)]
29. Yonar, D.; Ocek, L.; Tiftikcioglu, B.I.; Zorlu, Y.; Severcan, F. Relapsing-Remitting Multiple Sclerosis Diagnosis from Cerebrospinal Fluids via Fourier Transform Infrared Spectroscopy Coupled with Multivariate Analysis. *Sci. Rep.* **2018**, *8*, 1025. [[CrossRef](#)]
30. Kołodziej, M.; Chrabaszcz, K.; Pięta, E.; Pięgieś, N.; Rudnicka-Czerwiec, J.; Bartosik-Psujek, H.; Paluszkiwicz, C.; Cholewa, M.; Kwiatek, W.M. Spectral Signature of Multiple Sclerosis. Preliminary Studies of Blood Fraction by ATR FTIR Technique. *Biochem. Biophys. Res. Commun.* **2022**, *593*, 40–45. [[CrossRef](#)]
31. Ozek, N.S.; Tuna, S.; Erson-Bensan, A.E.; Severcan, F. Characterization of microRNA-125b Expression in MCF7 Breast Cancer Cells by ATR-FTIR Spectroscopy. *Analyst* **2010**, *135*, 3094–3102. [[CrossRef](#)]
32. Agresti, A. *Categorical Data Analysis; Wiley Series in Probability and Statistics*, 1st ed.; Wiley: Hoboken, NJ, USA, 2002; ISBN 978-0-471-36093-3.
33. Freitas, D.L.D.; Câmara, I.M.; Silva, P.P.; Wanderley, N.R.S.; Alves, M.B.C.; Morais, C.L.M.; Martin, F.L.; Lajus, T.B.P.; Lima, K.M.G. Spectrochemical Analysis of Liquid Biopsy Harnessed to Multivariate Analysis towards Breast Cancer Screening. *Sci. Rep.* **2020**, *10*, 12818. [[CrossRef](#)]
34. van Horssen, J.; Witte, M.E.; Schreiberlt, G.; de Vries, H.E. Radical Changes in Multiple Sclerosis Pathogenesis. *Biochim. Et Biophys. Acta. (BBA)-Mol. Basis Dis.* **2011**, *1812*, 141–150. [[CrossRef](#)] [[PubMed](#)]
35. Hollen, C.; Neilson, L.E.; Barajas, R.F.; Greenhouse, I.; Spain, R.I. Oxidative Stress in Multiple Sclerosis—Emerging Imaging Techniques. *Front. Neurol.* **2023**, *13*, 1025659. [[CrossRef](#)] [[PubMed](#)]
36. LeVine, S.M. Albumin and Multiple Sclerosis. *BMC Neurol.* **2016**, *16*, 47. [[CrossRef](#)]
37. Puthenparampil, M.; Tomas-Ojer, P.; Hornemann, T.; Lutterotti, A.; Jelcic, I.; Ziegler, M.; Hülsmeier, A.J.; Cruciani, C.; Faigle, W.; Martin, R.; et al. Altered CSF Albumin Quotient Links Peripheral Inflammation and Brain Damage in MS. *Neurol.-Neuroimmunol. Neuroinflamm.* **2021**, *8*, 1–14. [[CrossRef](#)] [[PubMed](#)]
38. Pedhazur, E.J. *Multiple Regression in Behavioral Research: Explanation and Prediction*, 3rd ed.; Belmont: Wadsworth, CA, USA, 1997; ISBN 978-0-03-072831-0.

39. Agresti, A. *Statistical Methods for the Social Sciences/Alan Agresti and Barbara Finlay*, 4th ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2009; ISBN 978-0-13-713150-1.
40. Blystad, I.; Håkansson, I.; Tisell, A.; Ernerudh, J.; Smedby, Ö.; Lundberg, P.; Larsson, E.-M. Quantitative MRI for Analysis of Active Multiple Sclerosis Lesions without Gadolinium-Based Contrast Agent. *AJNR Am. J. Neuroradiol.* **2016**, *37*, 94–100. [[CrossRef](#)]
41. Khoury, Y.E.; Collongues, N.; Sèze, J.D.; Gulsari, V.; Patte-Mensah, C.; Marcou, G.; Varnek, A.; Mensah-Nyagan, A.G.; Hellwig, P. Serum-Based Differentiation between Multiple Sclerosis and Amyotrophic Lateral Sclerosis by Random Forest Classification of FTIR Spectra. *Analyst* **2019**, *144*, 4647–4652. [[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.