Article

Not peer-reviewed version

Artificial Intelligence Algorithm Based on Genetics to Predict the Response to Interferon Beta Treatment in Multiple Sclerosis Patients

Edgar Rafael Ponce de León-Sánchez *, Jorge Domingo Mendiola-Santibañez, Omar Arturo Domínguez-Ramírez *, Ana Marcela Herrera-Navarro, Alberto Vazquez-Cervantes, Hugo Jiménez-Hernández, Jose Alfredo Acuña-García, Rafael Duarte-Pérez, José Manuel Álvarez-Alvarado

Posted Date: 23 March 2026

doi: 10.20944/preprints202603.1838.v1

Keywords: multiple sclerosis; machine learning; biomarkers; fuzzy system; genetic algorithm

 doi:10.20944/preprints202603.1838.v1

Article

Artificial Intelligence Algorithm Based on Genetics to Predict the Response to Interferon Beta Treatment in Multiple Sclerosis Patients

Edgar Rafael Ponce de León-Sánchez [1,*], Jorge Domingo Mendiola-Santibañez [2], Omar Arturo Domínguez-Ramírez [3,*], Ana Marcela Herrera-Navarro [1], Alberto Vazquez-Cervantes [4], Hugo Jiménez-Hernández [1], Jose Alfredo Acuña-García [1], Rafael Duarte-Pérez [1] and José Manuel Álvarez-Alvarado [2]

[1] Facultad de Informática, Universidad Autónoma de Querétaro, Querétaro 76230, México
[2] Facultad de Ingeniería, Universidad Autónoma de Querétaro, Querétaro 76010, México
[3] Centro de Investigación en Tecnologías de Información y Sistemas, Universidad Autónoma del Estado de Hidalgo, Pachuca 42039, México
[4] Centro de Ingeniería y Desarrollo Industrial, Querétaro 76125, Mexico
* Correspondence: eponcedeleon13@alumnos.uaq.mx (E.R.P.d.L.-S.); omar@uaeh.edu.mx (O.A.D.-R.)

Abstract

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system (CNS) affecting nearly 3 million people worldwide. Although its etiology and pathogenesis are not clearly established, it is likely the result of complex interactions between genetic and environmental factors. The MS autoimmune etiology has been the goal of therapeutic approaches for patients. IFN-β is one of the most prescribed disease-modifying therapies in MS patients. However, IFN-β is partially effective, and a significant proportion of patients partially respond or do not respond to this treatment. Machine learning (ML) is a subset of artificial intelligence (AI) focusing on developing data-driven computational models to enhance specific tasks. In recent years, unsupervised learning methods, such as hierarchical clustering and K-Means, have been applied in the MS study for classification tasks. Although these methods are relatively easy to implement, they rarely provide the best solution due to the large number of arbitrary decisions, and they perform only when the dataset contains clusters of similar sizes and no notable outliers. Therefore, in this paper, we propose an AI algorithm including a fuzzy expert system, based on the opinion of a neurology expert, to improve the efficiency in the assignment of unknown class labels associated to the response to IFN-β in MS patients, and a genetic algorithm (GA) to optimize the hyperparameter tuning of a deep learning (DL) model trained with genetic biomarkers to estimate whether the patients are potential candidates to be treated with this therapy. The experimental results show the fuzzy system achieved a 80% of classification efficiency against 64% of a conventional hierarchical clustering method. Furthermore, the artificial neural network (ANN) model, whose hyperparameters were optimized by the GA, achieved an 1.0 accuracy against 0.8 of a multi-layer perceptron (MLP), whose hyperparameters were adjusted by a conventional tuning method.

Keywords: multiple sclerosis; machine learning; biomarkers; fuzzy system; genetic algorithm

1. Introduction

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system (CNS), pathologically characterized by several degrees of demyelination, axonal loss, and gliosis, and clinically by a variety of neurological signs and symptoms [1]. Generally, the most common initial symptoms are sensory disturbances, motor alterations, and brainstem dysfunction [2]. Although MS can take several different forms, the most common type is relapsing-remitting multiple sclerosis (RRMS),

 doi:10.20944/preprints202603.1838.v1

characterized by alternating periods of remission and exacerbation of symptoms. Primary progressive multiple sclerosis (PPMS) is characterized by progressive damage of symptoms without periods of remission. Secondary progressive multiple sclerosis (SPMS) begins as PPMS, and symptoms gradually worsen [3]. Although the etiology and pathogenesis of MS are not well established, they likely result from complex interactions between genetic and environmental factors [4]. Environmental factors include viral infections, vitamin D deficiency, smoking, and others [5–7]. Some of these factors activate the human immune system, leading to immune dysregulation and an immune attack on the myelin sheaths of the CNS. Immune dysregulation manifests as infiltration of immune cells into the CNS, leading to myelin reduction, axonal loss, gliosis, and neuronal degeneration [8]. Several studies have shown that susceptibility to MS is genetically dependent, but specific genetic factors remain largely unknown.

MS treatment is challenging and involves some drugs acting through different mechanisms. Indication depends fundamentally on the clinical course and the form of the disease. MS treatment can be divided into three categories: symptom treatment, relapse treatment, and disease-modifying therapy. The primary goal of MS treatment is to slow disease progression. Glatiramer acetate and interferon beta (IFN-β) are the first-line therapies for MS. In patients who do not respond to treatment, several second-line therapies are available, such as natalizumab and fingolimod [9]. IFN-β is one of the most prescribed disease-modifying therapies for RRMS patients. IFN-β has multiple pathways of action on the immune system: it can inhibit activated T-cell proliferation, prevent migration of activated immune cells across the blood-brain barrier (BBB), inhibit the production of pro-inflammatory cytokines (e.g., IL-2, IL-12, IFN-γ), induce the increase of anti-inflammatory cytokines (e.g., IL-4, IL-5, IL-10, and TGF-β), and promote remyelination in the CNS [10,11]. IFN-β can also prevent the differentiation of Th1/Th17 inflammatory cells and shift the Th cell phenotype from Th1 inflammatory to Th2 anti-inflammatory [12]. Some studies have analyzed the IFN-β treatment effects on cytokine patterns in RRMS patients [13,14]. Other studies have shown IFN-β reduces disease activity [15,16]. Despite its high efficacy, some MS patients do not respond or partially respond to this treatment [17].

2. Background

In recent decades, technological advances have enabled large-scale studies of genes, proteins, and metabolites, leading to the creation of omics disciplines such as genomics, proteomics, and metabolomics, among others. Each of these areas has contributed to a better understanding of the causes of some diseases. In genomics, large amounts of data can be analyzed more efficiently, a task that, until a few years ago, was very difficult with conventional statistical analysis techniques [18]. The human genome contains around 20,000 genes; while this volume of data is complex and multidimensional, there is a need to use new artificial intelligence (AI) methods—such as expert systems, machine learning (ML) and deep learning (DL) algorithms, and evolutionary or metaheuristic algorithms—that can process large amounts of data more efficiently [19].

To date, the field of expert systems has been one of the most active in different research areas, such as disease diagnosis. An expert system is a computer program that emulates the reasoning process of a human expert and can efficiently manage uncertain and imprecise information [20]. A fuzzy expert system incorporates fuzzy sets and fuzzy logic into its reasoning process and knowledge base. Some studies have applied fuzzy systems to analyze neurological diseases [21]. However, most of these studies have only focused on MS diagnosis [22–25]. The study of response to MS treatments remains a challenge.

ML and DL algorithms are based on mathematical models that find natural patterns in data and are emerging as useful tools in the field of bioinformatics [26–28]. Over the past decade, the application of ML algorithms has increased across different medical fields, including cardiology, oncology, and neurology [29,30]. ML is classified into two techniques: supervised learning, which trains a model with known input and output data to predict future outcomes, and unsupervised learning, which identifies hidden patterns in input data without labeled outputs. These learning models, trained with patients'

 doi:10.20944/preprints202603.1838.v1

genetic characteristics, can help improve the diagnosis of some diseases, such as early MS [31–33], and estimate the potential response to MS treatments, such as natalizumab and fingolimod [34,35].

Hierarchical clustering is an unsupervised learning algorithm that groups data into a tree of nested sets. Some studies have applied hierarchical and non-hierarchical clustering methods to classify treatment responses in MS patients [36,37]. Other studies have analyzed the genome to identify genetic factors associated with the response to IFN-β treatment. In [38], Gurevich *et. al.* classified 50 SPMS patients (20 treated with IFNβ-1a and 30 untreated) using hierarchical clustering according to the IFN-inducible gene expression profile identified in clinically responsive RRMS patients. The gene expression profile was determined by searching for differentially expressed genes (DEGs) between IFN-treated (n = 10) and untreated (n = 25) RRMS patients. As a result, 104 DEGs, enriched by the IFN signaling pathway (p=7.4e-08), were identified in RRMS patients treated with IFN. In [39], Coelewij *et. al.* identified early immunogenicity biomarkers associated with the future development of antidrug antibodies (ADAs) in RRMS patients prior to their first IFN-β treatment. DEGs in SPMS patients versus RRMS ADAneg and ADApos versus RRMS ADAneg patients were used to classify patients using hierarchical clustering and principal component analysis (PCA). The results showed these biomarkers could aid future treatment decisions in RRMS patients.

K-Means is another unsupervised iterative clustering algorithm that uses centroid calculations to divide a dataset into similar groups based on the distance between their centroids. Some studies have implemented this method for classification tasks. In [40], Mesidor *et. al.* proposed a group-based trajectory modeling (GBTM) approach to identify patient groups with homogeneous dosing patterns for the drugs IFN-β1a and amitriptyline. Characteristics differences between groups were identified using chi-square and analysis of variance, both weighted by the posterior probability of group membership. As a result, seven dosing patterns were identified for IFN-β1a and five for amitriptyline. Whereas in [41], Kular *et. al.* analyzed deoxyribonucleic acid (DNA) methylation in the core of non-neuronal cells isolated from 38 samples of normal-appearing white matter (NAWM) from MS patients compared to control individuals. a total of 1226 significant differentially methylated positions (DMP) with adjusted p-values across the genome ($p - adj < 0.05$) were identified between MS patients and healthy controls. Genes involved in biological processes were classified using K-Means clustering into five groups. The findings suggest that NAWM glial cells are highly altered, revealing relevant molecular changes underlying MS neuropathology. Although unsupervised learning techniques are relatively easy to implement, some limitations are identified: 1) Hierarchical clustering rarely provides the best solution due to the large number of arbitrary decisions. 2) K-Means works effectively only when the dataset contains clusters of similar size, and there are no notable outliers or density variations.

In [42], Tao *et. al.* implemented a feature selection method to identify differentially correlated gene (DCE) pairs associated with IFN-β1b treatment response in RRMS patients. As a result, 22 statistically significant DCE ($p < 0.01$) were identified, involving 41 unique genes. Using these genes as predictors, a support vector machine (SVM) model was trained, achieving a predictive accuracy of 0.8095. However, some limitations are apparent: 1) The training dataset does not contain output labels, and the method used to determine the IFN-β1b responder and non-responder categories is not clearly described; 2) The training dataset has only 25 samples in total, and to delete potentially ambiguous results caused by patients whose first relapse occurred after 5 years, 7 patients were excluded from the analysis, resulting in a loss of potentially relevant information; 3) Parameter-based learning models, such as the SVM model, are trained by tuning their hyperparameters. These are configurations that are usually tuned arbitrarily before the training process begins to optimize model performance. However, arbitrary hyperparameter tuning does not always yield the best performance [43].

In contrast, genetic algorithms (GAs) are metaheuristic methods that search for optimal solutions in the search space and can be useful for optimizing the hyperparameters of learning models to achieve higher performance [44]. Therefore, this paper attempts to address some limitations of previous studies and proposes an AI algorithm that includes:

 doi:10.20944/preprints202603.1838.v1

- A method for validating the normal distribution of a dataset containing demographic, clinical, and genetic variables from 25 RRMS patients treated with IFN-β1b for two years.
- A fuzzy expert system, based on the opinion of a neurology expert, to improve the efficiency of assigning unknown class labels from the dataset, using demographic and clinical features as input variables.
- A GA to optimize the configuration hyperparameter tuning of an ANN model trained with genetic variables associated with the response to IFN-β1b, with the target to estimate whether new patients are potential candidates to be treated with this drug and achieve higher predictive performance than previous studies.

The clinic implementation of this algorithm could support specialists' decision-making in estimating the response to MS treatments, preventing ineffective therapies, and reducing healthcare costs.

3. Materials and Methods

The implemented research strategy is described in the flowchart in Figure 1.

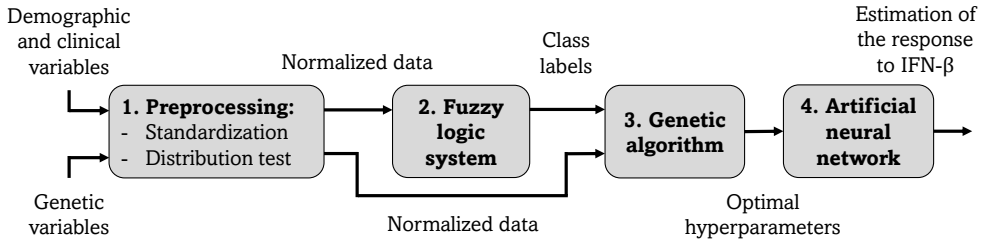

Figure 1. Proposed methodology. Demographic, clinical, and genetic characteristics are collected from the database. Next, they are preprocessed using standardization and distribution test techniques. Then, the demographic and clinical variables are fed into a fuzzy logic system to get the unknown class labels. Finally, the genetic variables with their labeled outputs are fed into a GA to find the optimal hyperparameter configuration for training an ANN model to perform predictions.

3.1. Database

The training and validation dataset was collected from the GSE24427 array expression profiling experiment, available in the Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/geo/), accessed on 15 Oct 2025. The database is the same as the one used in [42], although the selected biomarkers differ. Using the GPL96 [HG-U133A] microarray platform (Affymetrix Human Genome U133A Array), serum gene expression profiles were obtained from 25 German RRMS patients who were treated with subcutaneous IFN-β1b (Bayer Healthcare) at a dose of 250 μg per day for two years. Blood samples were drawn from patients before the first and second IFN-β1b injections, and at 1, 12, and 24 months. Expression values were analyzed using GEO2R, an interactive web tool that allows visualization of specific gene expression values through the Profile graph function. The GPL96 platform provides demographic and clinical characteristics of RRMS patients, which are presented in Table 1.

 doi:10.20944/preprints202603.1838.v1

Table 1. Demographic and clinical characteristics of RRMS patients treated with IFN-β1b for two years.

Sample	Age	Initial EDSS [1]	EDSS [1] after a year	EDSS [1] after two years
1	63	4	4.5	5.5
2	45	1	1	1
3	25	1	1	1
4	27	4	4	3.5
5	51	3	2.5	2.5
6	41	2	2	4.5
7	44	4	3	3
8	30	1.5	2	2
9	26	4	3.5	3.5
10	42	1	1	1
11	29	2	2	2.5
12	28	1.5	1	2.5
13	48	1	1	1
14	47	3.5	3.5	3
15	42	2	2.5	3
16	50	3.5	3.5	3.5
17	37	1.5	3.0	4.5
18	43	2	1.5	2
19	54	3	2	2
20	40	1	1	1
21	48	2	2	2
22	38	2	3.5	3
23	18	1.5	1.5	2
24	24	1	1	1
25	38	1	1	1

[1] Expanded Disability Status Scale (EDSS).

On the other hand, the expression values of 13 biomarkers associated with the response to IFN-β (IL-2, IFN-γ, TNF-α, IL-4, IL-10, TGF-β, CD46, CD58, FHIT, IRF5, GAPVD1, GRM3, GRIK2) were collected and integrated into an Excel spreadsheet to train the proposed ANN model. For example, the expression values of the IL-4 cytokine are shown in Figure 2.

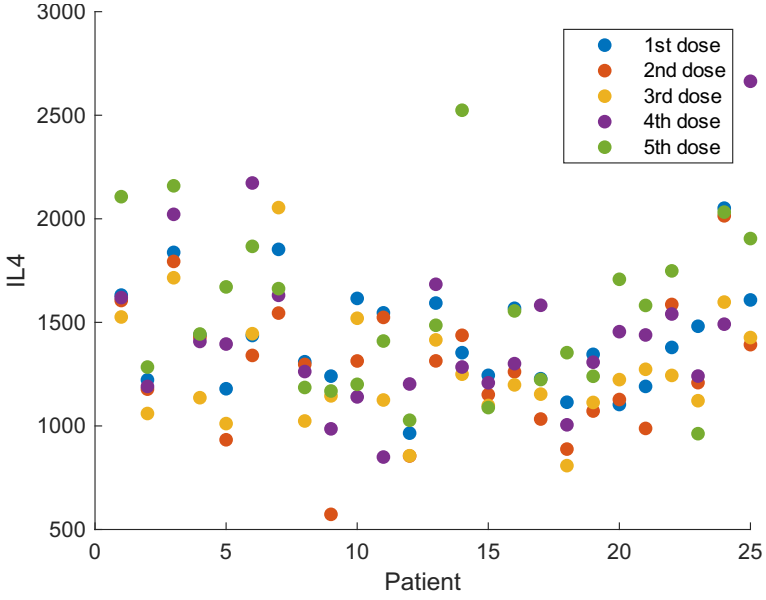

Figure 2. IL-4 cytokine. Genetic expression values of 25 RRMS patients corresponding to five doses: before the first and second IFN-β1b injections, and at 1, 12, and 24 months.

 doi:10.20944/preprints202603.1838.v1

3.2. Data Standardization

To preprocess the input database, standardization is first applied to normalize the demographic, clinical, and genetic variables. This method normalizes the characteristics by removing the mean μ, and scales them by dividing by their standard deviation s [33]. The standard score of a sample x is calculated as follows,

$$z = \frac{x - \mu}{s}. \tag{1}$$

3.3. Data Distribution Test

Before entering the standardized data into the fuzzy system and GA, it is advisable to assess whether the input variables are normally distributed. The Kolmogorov-Smirnov (KS) test is a nonparametric test used to assess whether a data set follows a normal distribution. The KS procedure for a sample compares the observed cumulative distribution function of a variable with a given theoretical distribution, such as the normal, uniform, Poisson, or exponential distributions. The KS output is computed from the largest difference (in absolute value) between the theoretical and observed cumulative distribution functions. This test of fit assesses whether the observations are consistent with the specified distribution. Equation 2 describes the empirical distribution function F_n for n ordered, independent, and identically distributed observations X_i.

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,x]}X_i \tag{2}$$

where $1_{(-\infty,x]}X_i$ is the indicator function, which is equal to 1 if $X_i \leq x$ or equal to zero otherwise.

The KS statistic for a given cumulative distribution function $F(x)$ is

$$D_n = sup_x|F_n(x) - F(x)| \tag{3}$$

where sup_x is the supremum of the set of distances. Intuitively, the statistic takes the largest absolute difference between the two distribution functions for all values of x. The KS test can also be used to test whether two underlying one-dimensional probability distributions differ. In this case, the KS statistic is

$$D_{n,m} = sup_x|F_{1,n}(x) - F_{2,m}(x)| \tag{4}$$

where $F_{1,n}(x)$ and $F_{2,m}(x)$ are the empirical distribution functions of the first and second samples respectively.

3.4. Assignment of Class Labels

Since the training dataset does not contain output information, a modified fuzzy system similar to that proposed in [46] is used to assign the unknown class labels, i.e., to classify RRMS patients as high, medium, and low responders to IFN-β1b treatment, based on their demographic and clinical characteristics. Additionally, a hierarchical clustering algorithm is trained to classify the same patients for comparison purposes.

3.4.1. Fuzzy Logic System

The main motivation for introducing fuzzy set theory was the need to model real-world phenomena that are inherently vague and ambiguous. Human knowledge of complex problems can be effectively represented using imprecise natural-language terms [47]. Fuzzy systems are structures based on fuzzy set theory and fuzzy logic for the efficient processing of imprecise information [48]. Their main property is the symbolic representation of knowledge as fuzzy conditional rules (if-then statements). The typical structure of a fuzzy system consists of three stages: Fuzzifier, Approximate Reasoning, and Defuzzifier.

 doi:10.20944/preprints202603.1838.v1

The fuzzifier transforms the values of the input variable into a fuzzy set A of linguistic values, using approximate reasoning (inference engine) and expert knowledge, which is represented as a set of fuzzy conditional rules (knowledge base). The fuzzifier can be defined as the membership function $\mu_A(x)$ of the fuzzy set A. The demographic and clinical characteristics of RRMS patients are used as input variables for the fuzzifier. The result of the approximate reasoning is a fuzzy set $B(y)$. The defuzzifier calculates a representative numerical output from the result of the fuzzy set $B(y)$. The numerical output y_0 is computed by the center of gravity (COG) method,

$$y_0 = \frac{\sum_{i=1}^{n} y_i \mu_B(y_i)}{\sum_{i=1}^{n} \mu_B(y_i)}, \tag{5}$$

where $\mu_B(y)$ represents the membership function of the fuzzy set $B(y)$ [49]. For the implementation of the fuzzy system, the input linguistic variables are defined as shown in Table 2. The sets of possible linguistic values are collections of different labels describing the variables: age, initial EDSS, EDSS after one year, EDSS after two years, and the output defining the response to IFN-β1b.

Table 2. Description of the linguistic variables (demographic, clinical characteristics, and the response to IFN-β1b).

Membership Function	Fuzzy Set	Universe of Discourse	Parameters
μ_{A_1}(Age)	A_1	$\mathbb{X}_1$: [0 a 100] years	Child, Adult, Elderly
μ_{A_2} (Initial EDSS)	A_2	$\mathbb{X}_2$: [0 a 10] units	Low, Medium, High
μ_{A_3} (EDSS after a year)	A_3	$\mathbb{X}_3$: [0 a 10] units	Low, Medium, High
μ_{A_4} (EDSS after two years)	A_4	$\mathbb{X}_4$: [0 a 10] units	Low, Medium, High
μ_B (Response to IFN-β1b)	$B(y)$	$\mathbb{Y}$: [0 a 1] units	Low, Medium, High

The fuzzy logic system is designed using the *Fuzzy Logic Designer* application in MATLAB R2023a. The structure of the fuzzy logic system is based on the Mamdani-Assilian Fuzzy Logic System (MAFS) [50], which comprises a set of fuzzy rules (knowledge base). Table 3 shows some of the 33 defined conditional rules based on the opinion of a neurology expert. The complete fuzzy rules are presented in Tables A1 and A2.

Table 3. Definition of specific fuzzy rules to determine the influence of input variables on the response to IFN-β1b treatment.

#	Fuzzy rule
1	If the age is Child, and the initial EDSS is Low, and the EDSS after one year is Low, and the EDSS after two years is Low, then the answer to IFN-β1b is Medium
2	If the age is Child and the initial EDSS is Medium and the EDSS after one year is Medium and the EDSS after two years is Medium, then the answer to IFN-β1b es Medium
3	If the age is Child, and the initial EDSS is High, and the EDSS after one year is High, and the EDSS after two years is High, then the answer to IFN-β1b es Medium
4	If the age is Child and the initial EDSS is Medium and the EDSS after one year is Low and the EDSS after two years is Low, then the answer to IFN-β1b is High

3.4.2. Hierarchical Clustering

Hierarchical clustering is a general family of unsupervised machine learning algorithms that construct nested clusters by successively merging or splitting them. This clustering hierarchy is represented as a tree (or dendrogram). The root of the tree is the single cluster containing all samples, with the leaves being the clusters containing only one sample. These algorithms perform hierarchical clustering using a bottom-up approach: each observation starts in its own cluster, and the clusters are successively merged [51]. The linking criteria determine the metric used for the merging strategy.

 doi:10.20944/preprints202603.1838.v1

1. *Ward*: It minimizes the sum of squared differences within all groups. It is an approach that minimizes variance and, in this sense, is similar to the K-Means objective function, but is applied hierarchically.
$$\frac{|\mathcal{A}| \cdot |\mathcal{B}|}{|\mathcal{A} \bigcup \mathcal{B}|} \, ||\mu_A - \mu_B||^2 \tag{6}$$
2. The maximum or complete link minimizes the maximum distance between elements of each group.
$$max\{d(x,y) : x \in \mathcal{A}, y \in \mathcal{B}\} \tag{7}$$
3. The average link minimizes the average distance between elements in each group.
$$\frac{1}{|\mathcal{A}| \cdot |\mathcal{B}|} \sum_{x \in \mathcal{A}} \sum_{y \in \mathcal{B}} d(x,y) \tag{8}$$
4. The unique link minimizes the minimum distance between the elements of each group.
$$min\{d(x,y) : x \in \mathcal{A}, y \in \mathcal{B}\} \tag{9}$$

Hierarchical clustering can also scale to a large number of samples when used with a connectivity matrix, but it is computationally expensive when no connectivity constraints are imposed between samples: it considers all possible merges at each step.

3.5. ML and DL Algorithms

Once the unknown class labels of the dataset have been assigned, a classical ML model such as the multi-layer perceptron (MLP) [52] and an ANN model [53] are trained with the genetic characteristics of RRMS patients to estimate the likely response to IFN-β1b and compare the performance of the learning models. For this purpose, the software Anaconda3 2021.05 (64-bit Python 3.8.8) and the open-source web application Jupyter Notebook are used to generate the pseudocode that runs on a 2.4 GHz Intel Core i5-1135G7 processor.

DL models have proven superior to classic ML models [54–57] due to their ability to learn highly complex relationships between input and output data. ANN models are at the core of DL, which are mathematical entities capable of representing complex functions as compositions of simpler functions. The basic component of these complex functions is the neuron, which performs a linear transformation of the input (for example, multiplying the input by a number [the weight] and adding a constant [the bias]), followed by the application of a fixed nonlinear function or activation function [58].

Mathematically, the output can be written as $o = f(w * x + b)$, with x as the input, w as the weight or scaling factor, and b as the bias or offset. f is the activation function, commonly set to the hyperbolic tangent ($tanh$). In general, x and therefore o can be simple scalars or vectors of scalars; and w can be a single scalar or a matrix, while b is a scalar or a vector (the dimensionality of the inputs and weights must match). In the latter case, the above expression is known as a layer of neurons, since it represents many neurons through multidimensional weights and biases. When the number of such layers exceeds two, the neural system is called a deep neural network. A deep or multilayer neural network is a composition of nonlinear functions,

$$x_1 = f(w_0 * x + b_0) \tag{10}$$
$$x_2 = f(w_1 * x_1 + b_1) \tag{11}$$
$$... \tag{12}$$
$$y = f(w_n * x_n + b_n) \tag{13}$$

 doi:10.20944/preprints202603.1838.v1

where the output of one layer of neurons is used as the input for the next layer, w_0 is a matrix and x is a vector. Using a vector allows w_0 to contain an entire layer of neurons, not just a single weight.

Overfitting is a common problem in ML and DL, where a model performs well with training data but not with validation data, i. e., the model is too complex with high variance. To avoid overfitting, the genetic input data (1,625 samples = 25 patients $\times$ 13 biomarkers $\times$ 5 doses) are divided into 80% for training (X_{train}) and 20% for validation (X_{test}), based on Pareto analysis [59]. Furthermore, the output labels are equally divided into 80% for training (y_{train}) and 20% for validation (y_{test}).

Furthermore, it is important to determine how effective learning models are by computing some commonly used metrics to evaluate their predictive performance [52]. Accuracy is the metric used to measure the performance of the ANN and MLP models. This metric computes the fraction or the count of correct predictions. If $\hat{y}_i$ is the predicted value of the i-th sample and y_i is the corresponding true value, then the fraction of correct predictions over $n_{samples}$ is defined as,

$$Acc(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\hat{y}_i - y_i) \tag{14}$$

where $1(x)$ is the indicator function.

Also, categorical cross-entropy ($L(y, \hat{y})$) is the loss function used to measure the performance of the ANN model. This function computes the difference between the true labels and the predicted probabilities of the model,

$$L(y, \hat{y}) = -\sum_{i=1}^{C} y_i \, log(\hat{y}_i) \tag{15}$$

where y_i is the true label for class i, $\hat{y}$ the predicted probability for class i, and C the number of classes.

3.6. Hyperparameter Optimization

Like any other hyperparameter-based ML model, a DL model is trained by tuning its hyperparameters. To train a neural network, an optimization scheme must be adopted. Hyperparameters are configurations of a learning model that can be arbitrarily tuned before starting the training process to optimize model performance. For example, in random forest (RF)-based algorithms, the number of estimators (decision trees) and the impurity criterion or measure. In contrast, model parameters, such as neural network weights w, are learned during training [60].

In this paper, a GA is proposed to find the optimal hyperparameters defining the architecture of an ANN trained with genetic variables associated with the response to IFN-β-1b. GA is a population-based metaheuristic algorithm that combines natural and genetic selection. GA searches for the optimal solution in the search space, including randomness [61]. GA is an iterative algorithm that converges in a finite number of iterations (generations), and candidate solutions evolve toward greater fitness or quality in each iteration. The algorithm ends when the optimal solution is reached, the maximum number of iterations is reached, or a stopping condition is met. Important features of a GA include initial population selection, a goal or fitness function, genetic operators, and termination criteria [62]. The algorithm begins by defining the problem and its objective function, $f(X)$, where X is a multidimensional vector with a typical dimension d. Figure 3 shows the schematic diagram of how a GA works.

 doi:10.20944/preprints202603.1838.v1

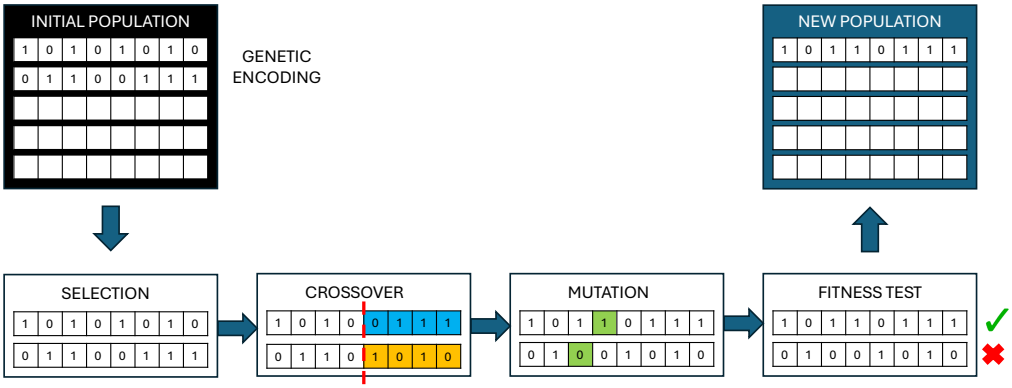

Figure 3. The initial population is randomly selected from the search space, and its members are encoded as chromosomes, represented as strings of characters. The chromosome is composed of genes. The values that genes take are alleles, which can be binary over the binary alphabet $\{0, 1\}$. The genetic operations of selection, crossover, and mutation are repeatedly applied to the population until the termination criterion is met. At the end of each iteration, the population's fitness values are computed. The members with the highest fitness are selected for mating and reproduction. Finally, the member with the highest fitness value is the optimal solution to the problem.

Genetic operators provide the basic search mechanism of GA and are used to create new solutions based on existing ones in the population. There are two basic types of operators: crossover and mutation. Crossover takes two individuals and produces two new ones, while mutation alters an individual to produce a single new solution. For real m-dimensional vectors $\bar{X}$ and $\bar{Y}$ representing chromosomes, the following operators are defined: uniform mutation and simple crossover. Let a_i and b_i be the lower and upper bounds, respectively, for each variable i. Uniform mutation randomly selects a variable, j, and sets it equal to a uniform random number bounded by the terms a_i and b_i:

$$x_i^{'} = \begin{Bmatrix} U(a_i, b_i) & si\ i = j \\ x_i & otherwise. \end{Bmatrix} \tag{16}$$

The simple real-valued crossover generates a random number r from a uniform distribution of 1 a m and creates two new individuals ($X^{'}$ and $Y^{'}$) according to the equation 17.

$$x_i^{'}(y_i^{'}) = \begin{Bmatrix} x_i(y_i) & si\ i < r \\ y_i(x_i) & otherwise. \end{Bmatrix} \tag{17}$$

The proposed GA begins by defining the problem (optimal hyperparameters) and its objective function $f(X)$ (maximum validation accuracy). In this case, X represents the following hyperparameter network options:

1. num_hidden_neurons: [64, 128, 256, 512, 768, 1024],
2. num_hidden_layers: [1, 2, 3, 4],
3. activation_function: ['relu', 'elu', 'tanh', 'sigmoid'],
4. optimizer: ['rmsprop', 'adam', 'sgd', 'adagrad','adadelta', 'adamax', 'nadam'].

Table 4 presents the pseudocode for the GA implementation. The chromosome (chosen parameters) with the highest fitness represents the global optimal solution. For comparison purposes, hyperparameters tuning of a MLP model (trained with the same genetic characteristics) is also performed using the conventional *GridSearchCV* tool from Python's sklearn library, which executes an exhaustive search for specific hyperparameter values for an estimator, considering all combinations. The hyperparameter network options for the MLP model are:

1. num_hidden_neurons: [64, 128, 256, 512, 768, 1024],
2. num_hidden_layers: [1, 2, 3, 4],

 doi:10.20944/preprints202603.1838.v1

3. activation_function: ['identity', 'logistic', 'tanh', 'relu'],
4. solver: ['lbfgs', 'sgd', 'adam'].

Table 4. Pseudocode of the implemented GA for hyperparameter optimizing.

Initialization	
	Select initial population (number of networks in each generation), N=100
	Define the objective function $f(X)$ (accuracy)
	Encode the population as chromosomes (bit strings) of length L_c
	Calculate the fitness values (accuracy) of the entire population
	Define the termination condition
	Choose the maximum number of iterations (generations or number of times to evolve the population), $Max_{Iter} = 50$
	iter = 1
while(iter $\leq Max_{Iter}$)	
	Selection: Select parents for breeding
	Crossover: Apply crossover to parents to produce offspring
	Mutation: Apply mutation to selected chromosomes
	Calculate the fitness values of the population
	Select members for the next generation based on fitness values
	If the termination condition is met, exit; otherwise, continue (iter = iter + 1)
end while	

4. Results

In this paper, demographic, clinical, and genetic characteristics of 25 RRMS patients treated with IFN-β1b treatment during two years were collected. Then, they were preprocessed by the standardization technique and the KS normal distribution test. Next, a fuzzy system to assign unknown class labels was used: high, medium, or low response to IFN-β1b. Also, a hierarchical clustering technique to classify the same patients was implemented. Subsequently, a GA to find the optimal configuration hyperparameters of an ANN model trained with the genetic variables associated with the labeled outputs was used. Also, a MLP model configured by a conventional hyperparameters tuning technique was trained with the same genetic features.

4.1. Data Distribution

After standardizing the demographic and clinical variables, and applying the KS test to verify their normal distribution ($p - value > 0.05$), the results in Table 5 were obtained. Figure 4 shows the empirical cumulative distribution functions for the demographic and clinical variables.

Table 5. Statistical results of the KS test of demographic and clinical variables.

Pairs of Variables	Statistics-D	Value-p	Distribution
Age / Initial EDSS	0.32	0.1557	Normal
Initial EDSS / EDSS after a year	0.32	0.1557	Normal
EDSS after a year / EDSS after two years	0.28	0.2850	Normal

 doi:10.20944/preprints202603.1838.v1

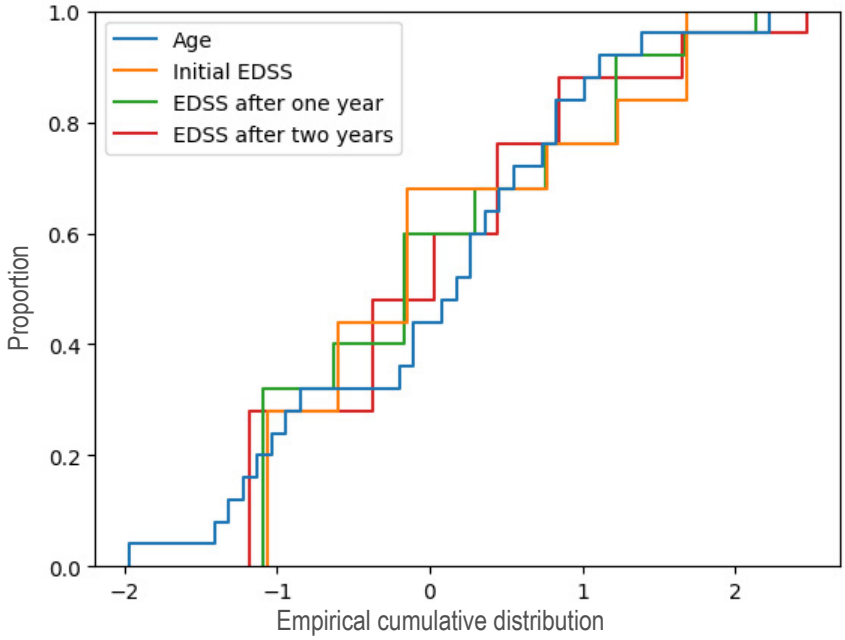

Figure 4. Graph of the empirical cumulative distribution functions. The four demographic and clinical variables: Age, initial EDSS, EDSS after a year, and EDSS after two years, follow a normal distribution.

The genetic variables were also standardized, and after applying the KS test to validate their normal distribution ($p - value > 0.05$), the results in Table 6 were obtained. Figure 5 shows the empirical cumulative distribution functions of the genetic features.

Table 6. Statistical results of the KS test on the genetic variables (first dose).

Pair of Variables	Statistic-D	Value-p	Distribution
IL-2 / IFN-γ	0.12	0.9955	Normal
IFN-γ / TNF-α	0.16	0.9149	Normal
TNF-α / IL-4	0.12	0.9955	Normal
IL-4 / IL-10	0.16	0.9149	Normal
IL-10 / TGF-β	0.12	0.9955	Normal
TGF-β / CD46	0.24	0.4754	Normal
CD46 / CD58	0.2	0.7102	Normal
CD58 / FHIT	0.16	0.9149	Normal
FHIT / IFR5	0.16	0.9149	Normal
IFR5 / GAPVD1	0.16	0.9149	Normal
GAPVD1 / GRM3	0.12	0.9955	Normal
GRM3 / GRIK2	0.16	0.9149	Normal

 doi:10.20944/preprints202603.1838.v1

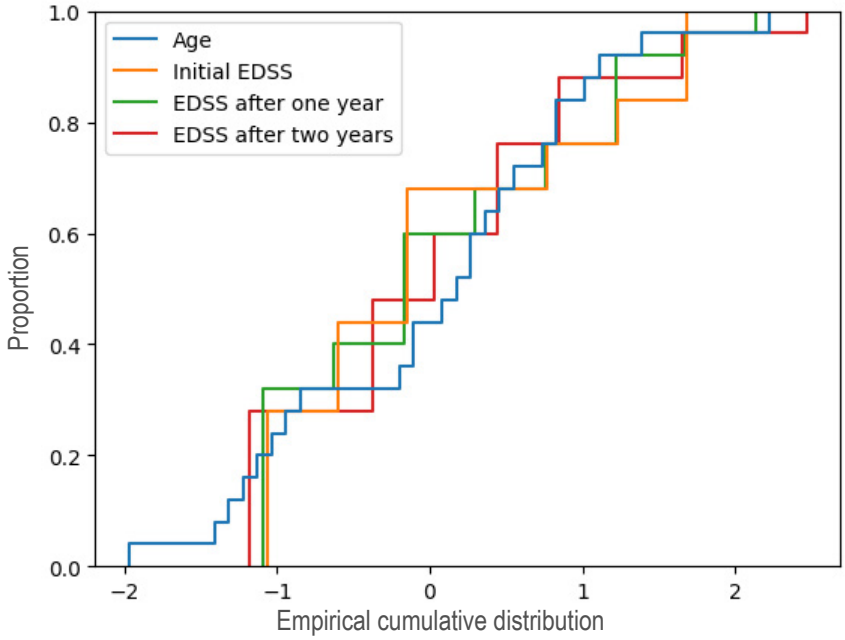

Figure 5. Graph of the empirical cumulative distribution functions. The 13 genetic variables at first dose: IL-2, IFN-γ, TNF-α, IL-4, IL-10, TGF-β, CD46, CD58, FHIT, IRF5, GAPVD1, GRM3, GRIK2, follow a normal distribution.

4.2. Fuzzy Expert System

At the fuzzyification stage, the linguistic values were computed corresponding to the numerical values of each input variable. At the approximate reasoning stage, each fuzzy rule (knowledge base) was evaluated using the membership values obtained from fuzzyification. For example, considering the input values for sample ”7” (Age: ”44”, initial EDSS: ”4”, EDSS after 1 year: ”3”, and EDSS after 2 years: ”3”), the inference engine results are shown in Table 7. In this case, only two rules had a non-zero inference result. Figure 6 shows the evaluation graph for 13 and 15 fuzzy rules.

Table 7. Inference results from 13 and 15 fuzzy rules evaluation (sample 7).

#	Rule	Inference Engine
13	If the age is Adult and the initial EDSS is Medium and the EDSS after one year is Medium and the EDSS after two years is Medium, then the IFN-β response is Medium	min(0.856, 0.75, 0.525, 0.5) = 0.5
15	If the age is Adult and the initial EDSS is Medium and the EDSS after one year is Low and the EDSS after two years is Low, then the response to IFN-β is High	min(0.856, 0.75, 0.225, 0.5) = 0.225

 doi:10.20944/preprints202603.1838.v1

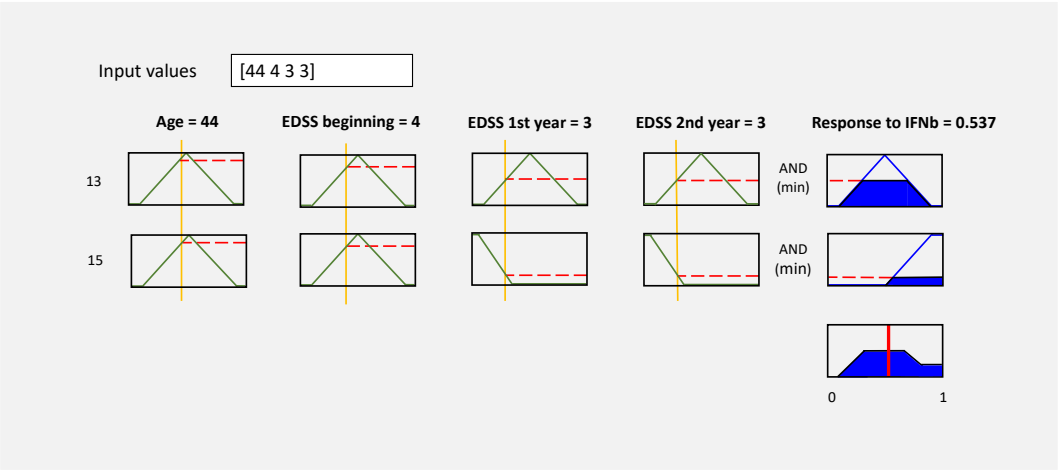

Figure 6. Rule inference graph of the *Fuzzy Logic Designer* application, describing the results of combining the inference values of the 13 and 15 fuzzy rules.

At the defuzzification stage, the numerical outputs were computed from the inference result graphs using Equation (5). For example, to obtain the numerical output for sample 7 according to the inference graph in Figure 7, the following calculation was performed:

$$y_0 = \frac{0.0 * 0.0 + 0.1 * 0.0 + 0.2 * 0.25 + 0.3 * 0.5 + \cdots + 0.8 * 0.25 + 0.9 * 0.25 + 1 * 0.25}{0.0 + 0.0 + 0.25 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.25 + 0.25 + 0.25} \tag{18}$$

$$y_0 = \frac{1.9025}{3.4} = 0.559 \approx 0.537 \tag{19}$$

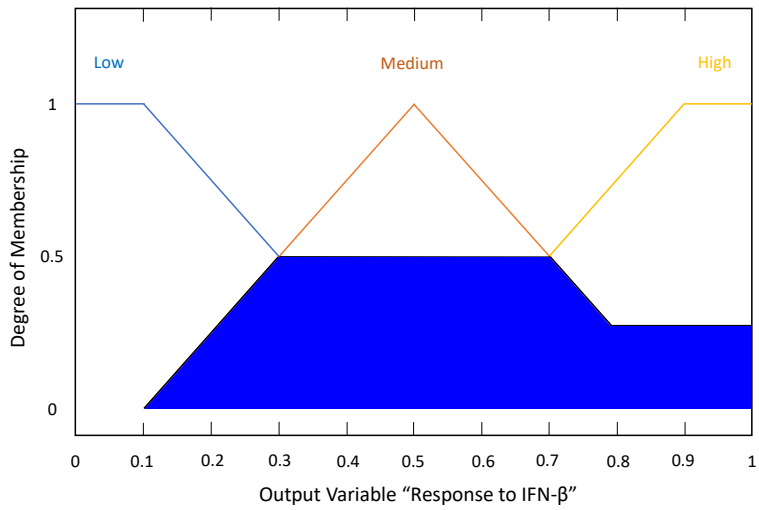

Figure 7. Graph of inference results for sample 7, which includes the values of the linguistic labels: low, medium, and high.

Finally, RRMS patients were classified as high, medium, and low responders to IFN-β1b treatment using three methods: 1) expert neurology opinion, 2) fuzzy system, and 3) conventional hierarchical clustering model. Table 8 shows the comparison results. The numerical data from Table 1 were also used to train the hierarchical clustering model ($n_{clusters} = 3$).

 doi:10.20944/preprints202603.1838.v1

Table 8. Classification of the response to IFN-$\beta 1b$. In the case of Fuzzy system, the numerical values (resulting from defuzzification) less than 0.5 are considered as low responder (LR), those equal to 0.5 as medium responder (MR), and those greater than 0.5 as high responder (HR).

Sample	Expert opinion	Fuzzy System (Defuzzification)	Hierarchical Clustering
1	LR	0.5 ⇒ MR	HR
2	MR	0.5 ⇒ MR	MR
3	MR	0.5 ⇒ MR	MR
4	MR	0.5 ⇒ MR	HR
5	MR	0.537 ⇒ HR	LR
6	LR	0.433 ⇒ LR	LR
7	HR	0.537 ⇒ HR	HR
8	MR	0.473 ⇒ LR	MR
9	MR	0.513 ⇒ HR	HR
10	MR	0.5 ⇒ MR	MR
11	MR	0.5 ⇒ MR	MR
12	LR	0.473 ⇒ LR	MR
13	MR	0.5 ⇒ MR	MR
14	MR	0.5 ⇒ MR	HR
15	LR	0.443 ⇒ LR	LR
16	MR	0.5 ⇒ MR	HR
17	LR	0.247 ⇒ LR	LR
18	MR	0.527 ⇒ HR	LR
19	HR	0.595 ⇒ HR	LR
20	MR	0.5 ⇒ MR	MR
21	MR	0.5 ⇒ MR	LR
22	LR	0.363 ⇒ LR	LR
23	MR	0.5 ⇒ MR	MR
24	MR	0.5 ⇒ MR	MR
25	MR	0.5 ⇒ MR	MR

Based on previous results, 80% of patients were correctly labeled using the fuzzy system, while 64% were correctly labeled using hierarchical clustering, regarding an expert opinion.

4.3. Hyperparameter Optimization

Table 9 presents the obtained results after implementing the proposed GA as a configuration hyperparameter optimizer of the ANN model. Table 10 shows the obtained results from the hyperparameter tuning of the MLP model by a conventional technique.

Table 9. Configuration hyperparameter optimization of the ANN model by GA (five combinations of optimal hyperparameters with the maximum accuracy).

Hidden neurons	Hidden layers	Activation function	Optimizer	Accuracy
256	3	'relu'	'adam' [1]	1.0
64	1	'sigmoid'	'nadam' [2]	1.0
64	2	'relu'	'adamax' [3]	1.0
512	1	'sigmoid'	'adamax'	1.0
128	4	'relu'	'nadam'	1.0

[1] Stochastic gradient descent method based on adaptive estimation of first and second order moments. [2] Adam variant with Nesterov moment. [3] Adam's variant based on the infinite norm is a first-order gradient-based optimization method.

 doi:10.20944/preprints202603.1838.v1

Table 10. Configuration hyperparameter tuning of the MLP model by *GridSearchCV* (four combinations of selected hyperparameters with the maximum accuracy).

Hidden Neurons	Hidden Layers	Activation Function	Solver	Accuracy
64	1	'logistic'	'sgd' [1]	0.8
128	1	'logistic'	'sgd'	0.8
256	4	'logistic'	'sgd'	0.8
512	3	'logistic'	'sgd'	0.8

[1] Stochastic gradient descent.

Figures 8 and 9 show the behavior of the accuracy performance metric and the loss function as a result of the ANN model training process over 50 epochs (generations).

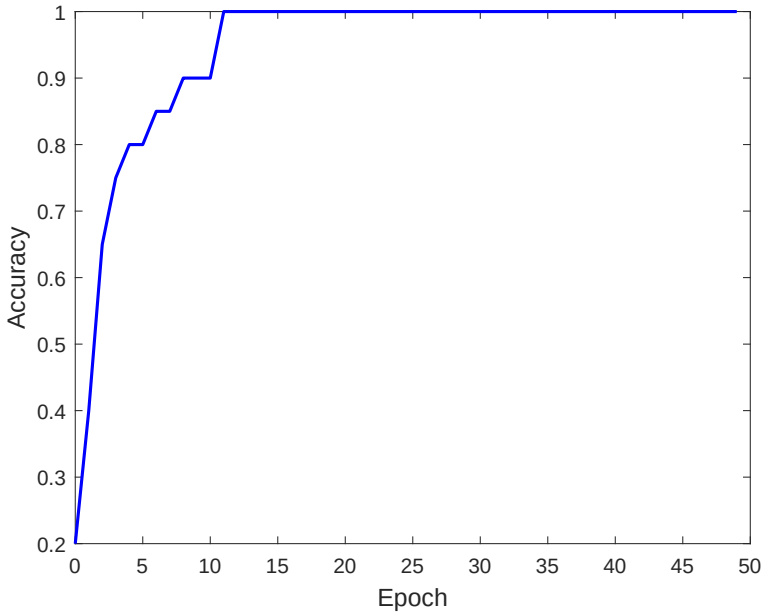

Figure 8. Behavior of the accuracy performance metric (ANN model training process). The accuracy achieved its maximum value approximately from epoch 11. Configuration hyperparameters optimized by GA: num_hidden_neurons = 64, num_hidden_layers = 2, activation_function = 'relu', optimizer = 'adamax'.

 doi:10.20944/preprints202603.1838.v1

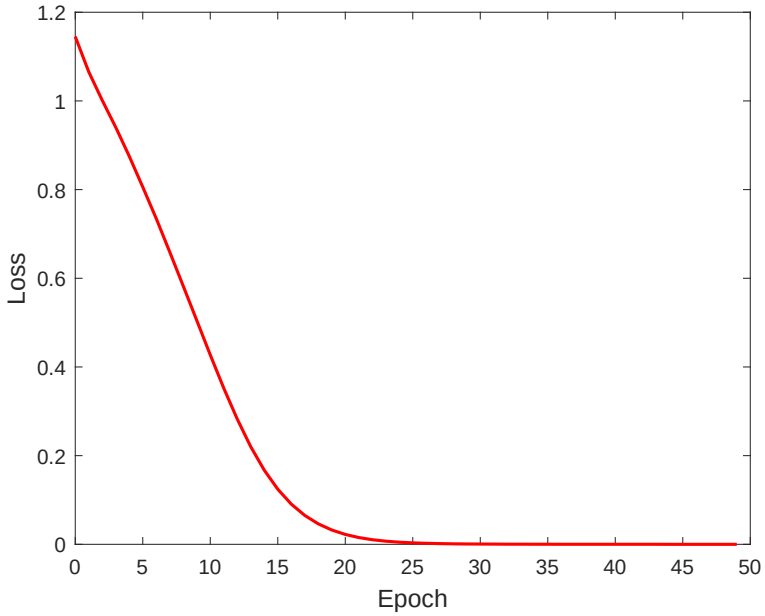

Figure 9. Loss function behavior of the ANN model training process. The loss converges to its minimum value after epoch 25.

5. Discussion

MS is an autoimmune inflammatory disease of the CNS affecting nearly 3 million people worldwide. Current disease-modifying therapies focus on delaying disease progression, treating sensitive attacks, and improving symptoms [63]. Although these therapies are effective mainly in the early phases of the disease, some MS patients partially respond or do not respond to the treatments. Hence, it is important to design new intelligent classification systems to determine the degree of response to treatments. ML and DL are subsets of AI that focus on developing computational models to improve specific tasks by making data-based predictions [27]. Over the past decade, the application of ML algorithms has increased across several medical fields [29,30].

In the field of neurology, unsupervised hierarchical clustering methods based on the genome are among the most commonly used to diagnose MS and predict the potential response to IFN-β [38,39]. Although hierarchical clustering methods are easy to implement and the results showed that some biomarkers could aid future treatment decisions, these methods rarely provide an efficient solution due to many arbitrary decisions. Also, they do not consider data outliers; i.e., they use only the distance between points to cluster them. Fuzzy logic is another subset of AI. A fuzzy expert system emulates expert reasoning and can handle uncertain and imprecise information to provide efficient solutions. Most studies that have applied fuzzy systems have focused only on the diagnosis of neurological diseases [21,24,25]. So, estimating responses to treatments remains a challenge. Hence, in this paper, we propose an AI algorithm that includes a fuzzy system to estimate the possible response to IFN-β1b treatment in RRMS patients. The fuzzy system is based on fuzzy rules defined in collaboration with a medical specialist. The modified fuzzy system addresses a limitation of the previous work [46], since the normal distribution of the input variables has not been evaluated. Therefore, the "gender" demographic variable is replaced by the clinical variable "Initial EDDS", as "gender" did not follow a normal distribution relative to the other variables. Furthermore, unlike the previous work, the IL-12 and GPC5 genetic features associated with the response to IFN-β1b were excluded from the input variables because they also did not follow a normal distribution after performing the KS test. The results of Table 8 show that the fuzzy system achieved greater efficiency in estimating the degree of response to IFN-β1b than the hierarchical clustering.

Once the fuzzy system classified the output labels of the data set, a GA was implemented to optimize the hyperparameter configuration of an ANN model, trained with genetic features associated with the response to IFN-β1b, with the target of predicting whether RRMS patients are potential

 doi:10.20944/preprints202603.1838.v1

candidates for treatment with this drug. The performance results in Tables 9 and 10 indicate that GA as a hyperparameter optimizer outperforms conventional methods of hyperparameter tuning.

6. Conclusions

The proposed AI algorithm to assess whether IFN-β is suitable for an MS patient is based on a fuzzy system developed in collaboration with a medical specialist and a metaheuristic algorithm to optimize hyperparameters. AI enables the construction of a bridge between a set of variables that may or may not be strongly correlated with the disease. According to the performed analysis, it is possible to estimate whether a patient has a high, medium, or low response to the IFN-β treatment.

The obtained results in this paper lead to a promising path to determining a better treatment for MS patients if a nonlabeled database is used. Managing uncertainty is one of the challenges of medical decision-making. AI-based predictive systems reduce uncertainty and improve assertiveness, thereby increasing precision and, consequently, confidence in decision-making about the most effective dose and combination of drugs for MS. However, several difficulties must be overcome, such as having different databases where representative variables used in the literature, including those for genetic analysis, are available, since people in different countries have different characteristics. Future work aims to design new intelligent systems to improve decision-making in the study of other diseases using different data types: genetic and clinical variables, digital images such as magnetic resonance images (MRI), etc.

Author Contributions: Conceptualization, E.R.P.d.L.-S.; methodology, E.R.P.d.L.-S., and J.M.A.-A.; software, E.R.P.d.L.-S.; validation, J.D.M.-S., and O.A.D.-R.; formal analysis, A.M.H.-N., A.V.-C., H.J.-H., J.A.A.-G., R.D.-P, and J.M.A.-A.; investigation, E.R.P.d.L.-S.; resources, J.D.M.-S., O.A.D.-R., A.M.H.-N., A.V.-C., H.J.-H., J.A.A.-G., R.D.-P, and J.M.A.-A.; writing—original draft preparation, E.R.P.d.L.-S.; writing—review and editing, E.R.P.d.L.-S. and J.D.M.-S.; supervision, O.A.D.-R.; project administration, E.R.P.d.L.-S., J.D.M.-S., and O.A.D.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by SECIHTI grant number 331870.

Data Availability Statement: Data and implemented algorithms are available upon reasonable request to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MS	multiple sclerosis
CNS	central nervous system
IFN	interferon
ML	machine learning
AI	arificial intelligence
RRMS	relapsing-remitting multiple sclerosis
GA	genetic algorithm
DL	deep learning
ANN	artificial neural network
SPMS	secondary progressive multiple sclerosis
PPMS	primary progressive multiple sclerosis
BBB	blood-brain barrier
DEG	differentially expressed genes
PCA	principal component analysis
DNA	deoxyribonucleic acid

 doi:10.20944/preprints202603.1838.v1

DCE	differentially correlated genes
ADA	anti-drug antibodies
GBTM	group-based trajectory modeling
NAWM	normal-appearing white matter
DMP	differentially methylated positions
SVM	support vector machine
DT	decision tree
RF	random forest
KN	k-neighbors
UCSC	uncorrelated shrunken centroid
RNA	ribonucleic acid
GEO	Gene Expression Omnibus
EDSS	expanded disability status scale
KS	Kolmogorov-Smirnov
COG	center of gravity
MAFS	Mamdani-Assilan fuzzy system
MLP	multilayer perceptron
SGD	stochastic gradient descent
Adam	estimation of adaptive momentum
MRI	magnetic resonance images

Appendix A

Table A1. Definition of fuzzy rules to determine the influence of input variables on the response to IFN-β1*b* treatment.

#	Fuzzy rule
1	If the age is Child, and the initial EDSS is Low, and the EDSS after one year is Low, and the EDSS after two years is Low, then the answer to IFN-β1b is Medium
2	If the age is Child and the initial EDSS is Medium and the EDSS after one year is Medium and the EDSS after two years is Medium, then the answer to IFN-β1b es Medium
3	If the age is Child, and the initial EDSS is High, and the EDSS after one year is High, and the EDSS after two years is High, then the answer to IFN-β1b es Medium
4	If the age is Child and the initial EDSS is Medium and the EDSS after one year is Low and the EDSS after two years is Low, then the answer to IFN-β1b is High
5	If the age is Child and the initial EDSS is High, and the EDSS after one year is Medium, and the EDSS after two years is Medium, then the response to the IFN-β1b is High
6	If the age is Child and the initial EDSS is High, and the EDSS after one year is Low, and the EDSS after two years is Low, then the response to the IFN-β1b is High
7	If the age is Child and the initial EDSS is High, and the EDSS after one year is Medium, and the EDSS after two years is Low, then the response to the IFN-β1b is High
8	If the age is Child and the initial EDSS is Low, and the EDSS after one year is Medium, and the EDSS after two years is Medium, then the response to the IFN-β1b is Low
9	If the age is Child and the initial EDSS is Low, and the EDSS after one year is High, and the EDSS after two years is High, then the response to IFN-β1b is Low
10	If the age is Child and the initial EDSS is Low, and the EDSS after one year is Medium, and the EDSS after two years is High, then the response to IFN-β1b is Low
11	If the age is Child and the initial EDSS is Middle and the EDSS after one year is High and the EDSS after two years is High, then the response to IFN-β1b is Low
12	If the age is Adult and the initial EDSS is Low, and the EDSS score after one year is Low, and the EDSS score after two years is Low, then the response to the IFN-β1b is Medium
13	If the age is Adult and the initial EDSS is Medium, and the EDSS score after one year is Medium, and the EDSS score after two years is Medium, then the response to the IFN-β1b is Medium
14	If the age is Adult and the initial EDSS is High, and the EDSS score after one year is High, and the EDSS score after two years is High, then the response to the IFN-β1b is Medium
15	If the age is Adult and the initial EDSS is Medium, and the EDSS score after one year is Low, and the EDSS score after two years is Low, then the response to the IFN-β1b is High
16	If the age is Adult and the initial EDSS is High and the EDSS after one year is Medium and the EDSS after two years is Medium, then the response to IFN-β1b is High
17	If the age is Adult and the initial EDSS is High, and the EDSS score after one year is Low, and the EDSS score after two years is Low, then the response to the IFN-β1b test is High

 doi:10.20944/preprints202603.1838.v1

Table A2. Definition of fuzzy rules to determine the influence of input variables on the response to IFN-β1*b* treatment.

#	Fuzzy rule
18	If the age is Adult and the initial EDSS is High, and the EDSS score after one year is Medium, and the EDSS score after two years is Low, then the response to the IFN-β1b test is High
19	If the age is Adult and the initial EDSS is Low, and the EDSS score after one year is Medium, and the EDSS score after two years is Medium, then the response to the IFN-β1b test is Low
20	If the age is Adult and the initial EDSS is Low, and the EDSS score after one year is High, and the EDSS score after two years is High, then the response to the IFN-β1b test is Low
21	If the age is Adult and the initial EDSS is Low, the EDSS score after one year is Medium, and the EDSS score after two years is High, then the response to the IFN-β1b test is Low
22	If the age is Adult and the initial EDSS is Medium, the EDSS score after one year is High, and the EDSS score after two years is High, then the response to the IFN-β1b test is Low
23	If the age is Elderly and the initial EDSS score is Low, and the EDSS score after one year is Low, and the EDSS score after two years is Low, then the response to the IFN-β1b test is Medium
24	If the age is Elderly and the initial EDSS score is Medium, and the EDSS score after one year is Medium, and the EDSS score after two years is Medium, then the response to the IFN-β1b test is Medium
25	If the age is Elderly and the initial EDSS score is High, and the EDSS score after one year is High, and the EDSS score after two years is High, then the response to the IFN-β1b test is Medium
26	If the age is Elderly and the initial EDSS score is Medium, and the EDSS score after one year is Low, and the EDSS score after two years is Low, then the response to the IFN-β1b test is High
27	If the age is Elderly and the initial EDSS score is High, and the EDSS score after one year is Medium, and the EDSS score after two years is Medium, then the response to the IFN-β1b is High
28	If the age is Elderly and the initial EDSS score is High, and the EDSS score after one year is Low, and the EDSS score after two years is Low, then the response to the IFN-β1b is High
29	If the age is Elderly and the initial EDSS score is High, and the EDSS score after one year is Medium, and the EDSS score after two years is Low, then the response to the IFN-β1b is High
30	If the age is Elderly and the initial EDSS score is Low, and the EDSS score after one year is Medium, and the EDSS score after two years is Medium, then the response to the IFN-β1b is Low
31	If the age is Elderly and the initial EDSS score is Low, and the EDSS score after one year is High, and the EDSS score after two years is High, then the response to the IFN-β1b is Low
32	If the age is Elderly and the initial EDSS score is Low, and the EDSS score after one year is Medium, and the EDSS score after two years is High, then the response to the IFN-β1b is Low
33	If the age is Elderly and the initial EDSS score is Medium, and the EDSS score after one year is High, and the EDSS score after two years is High, then the response to the IFN-β1b is Low

References

1. Aslam, N.; Khan, I.U.; Bashamakh, A.; Alghool, F.A.; Aboulnour, M.; Alsuwayan, N.M.; Alturaif, R.K.; Brahimi, S.; Aljameel, S.S.; Al Ghamdi, K. (2022). Multiple Sclerosis Diagnosis Using Machine Learning and Deep Learning: Challenges and Opportunities. *Sensors* **2022**, *22*, 7856.
2. Margarit, B. P., Monteiro, G. C., Herán, I. S., Delgado, F. R., & Izquierdo, A. Y. (2019). Esclerosis múltiple. *Elsevier* **2019**, *12(78)*, 4587–4597.
3. Martynova, E., Khaibullin, T., Salafutdinov, I., Markelova, M., Laikov, A., Lopukhov, L., & Khaiboullina, S. (2023). Seasonal changes in serum metabolites in multiple sclerosis relapse. *International Journal of Molecular Sciences* **2023**, *24(4)*, 3542.
4. Ghasemi, N., Razavi, S., & Nikzad, E. (2016). Multiple sclerosis: pathogenesis, symptoms, diagnoses and cell-based therapy. *Cell Journal (Yakhteh)* **2016**, *19(1)*, 1.
5. Tarlinton, R. E., Martynova, E., Rizvanov, A. A., Khaiboullina, S., & Verma, S. (2020). Role of viruses in the pathogenesis of multiple sclerosis. *Viruses* **2020**, *12(6)*, 643.
6. Zarghami, A., Li, Y., Claflin, S. B., van der Mei, I., & Taylor, B. V. (2021). Role of environmental factors in multiple sclerosis. *Expert Review of Neurotherapeutics* **2021**, *21(12)*, 1389-1408.
7. Dominguez-Mozo, M. I., Perez-Perez, S., Villarrubia, N., Costa-Frossard, L., Fernandez-Velasco, J. I., Ortega-Madueño, I., & Alvarez-Lafuente, R. (2021). Herpesvirus antibodies, vitamin d and short-chain fatty acids: their correlation with cell subsets in multiple sclerosis patients and healthy controls. *Cells* **2021**, *10(1)*, 119.
8. Rodríguez Murúa, S., Farez, M. F., & Quintana, F. J. (2022). The immune response in multiple sclerosis. *Annual Review of Pathology: Mechanisms of Disease* **2022**, *17(1)*, 121-139.
9. Lorscheider, J., Benkert, P., Lienert, C., Hänni, P., Derfuss, T., Kuhle, J., & Yaldizli, Ö. (2018). Comparative analysis of natalizumab versus fingolimod as second-line treatment in relapsing–remitting multiple sclerosis. *Multiple Sclerosis Journal* **2018**, *24(6)*, 777-785.
10. Mirandola, S. R., Hallal, D. E., Farias, A. S., Oliveira, E. C., Brandão, C. O., Ruocco, H. H., & Santos, L. M. (2009). Interferon-beta modifies the peripheral blood cell cytokine secretion in patients with multiple sclerosis. *International immunopharmacology* **2009**, *9(7-8)*, 824-830.

 doi:10.20944/preprints202603.1838.v1

11. Kay, M., Hojati, Z., & Dehghanian, F. (2013). The molecular study of IFNβ pleiotropic roles in MS treatment. *Iranian journal of neurology* **2013**, *12(4)*, 149.
12. Hočevar, K., Ristić, S., & Peterlin, B. (2019). Pharmacogenomics of multiple sclerosis: a systematic review. *Frontiers in neurology* **2019**, *10*, 134.
13. Ganji, A., Monfared, M. E., Shapoori, S., Nourbakhsh, P., Ghazavi, A., Ghasami, K., & Mosayebi, G. (2020). Effects of interferon and glatiramer acetate on cytokine patterns in multiple sclerosis patients. *Cytokine* **2020**, *126*, 154911.
14. Martínez-Aguilar, L., Pérez-Ramírez, C., del Mar Maldonado-Montoro, M., Carrasco-Campos, M. I., Membrive-Jiménez, C., Martínez-Martínez, F., & Jiménez-Morales, A. (2020). Effect of genetic polymorphisms on therapeutic response in multiple sclerosis relapsing-remitting patients treated with interferon-beta. *Mutation Research* **2015**, *785*, 108322.
15. Cohan, S. L., Hendin, B. A., Reder, A. T., Smoot, K., Avila, R., Mendoza, J. P., & Weinstock-Guttman, B. (2021). Interferons and multiple sclerosis: Lessons from 25 years of clinical and real-world experience with intramuscular interferon beta-1a (Avonex). *CNS drugs* **2021**, *35(7)*, 743-767.
16. Bustamante, M. F., Morcillo-Suárez, C., Malhotra, S., Rio, J., Leyva, L., Fernández, O., & Comabella, M. (2015). Pharmacogenomic study in patients with multiple sclerosis: Responders and nonresponders to IFN-β. *Neuroimmunology & Neuroinflammation* **2015**, *2(5)*, e154.
17. Río, J., Nos, C., Tintoré, M., Borrás, C., Galán, I., Comabella, M., & Montalban, X. (2002). Assessment of different treatment failure criteria in a cohort of relapsing–remitting multiple sclerosis patients treated with interferon β: Implications for clinical trials. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* **2002**, *52(4)*, 400-406.
18. Vela, M. E. F. V., & Gutiérrez-Aguilar, R. (2017). Ciencias "ómicas",¿cómo ayudan a las ciencias de la salud? *Revista digital universitaria* **2020**, *18(7)*.
19. Karlik, B. (2013). Soft computing methods in bioinformatics: A comprehensive review. *Mathematical and Computational Applications* **2013**, *18(3)*, 176-197.
20. Saibene, A., Assale, M., & Giltri, M. (2021). Expert systems: Definitions, advantages and issues in medical field applications. *Expert Systems with Applications* **2021**, *177*, 114900.
21. , Sadoughi, F., & Arani, L. A. (2022). Intelligent computer systems for multiple sclerosis diagnosis. *In Augmenting Neurological Disorder Prediction and Rehabilitation Using Artificial Intelligence* **2022**, *Academic Press*, 75–105.
22. Ayangbekun, O. J., & Jimoh Ibrahim, A. (2015). Fuzzy logic application to brain diseases diagnosis. *Journal of Emerging Trends in Computing and Information Sciences* **2015**, *6(3)*, 144-148.
23. Hosseini, A., Asadi, F., & Arani, L. A. (2020). Development of a knowledge-based clinical decision support system for multiple sclerosis diagnosis. *Journal of Medicine and Life* **2020**, *13(4)*, 612.
24. Matinfar, F., & Golpaygani, A. T. (2022). A fuzzy expert system for early diagnosis of multiple sclerosis. *Journal of Biomedical Physics & Engineering* **2022**, *12(2)*, 181.
25. Chen, J., & Gustientiedina, G. (2024). Implementation of Fuzzy Expert System to Detect Parkinson's Disease Based on Mobile. *Journal of Applied Business and Technology* **2024**, *5(2)*, 72-81.
26. Law, M. T., Traboulsee, A. L., Li, D. K., Carruthers, R. L., Freedman, M. S., Kolind, S. H., & Tam, R. (2019). Machine learning in secondary progressive multiple sclerosis: an improved predictive model for short-term disability progression. *Multiple Sclerosis Journal–Experimental, Translational and Clinical* **2019**, *5(4)*, 205521719885983.
27. Meskó, B., & Görög, M. (2020). A short guide for medical professionals in the era of artificial intelligence. *NPJ digital medicine* **2020**, *3(1)*, 126.
28. Cavaliere, C.; Vilades, E.; Alonso-Rodríguez, M.C.; Rodrigo, M.J.; Pablo, L.E.; Miguel, J.M.; López-Guillén, E.; Morla, E.M.S.; Boquete, L.; Garcia-Martin, E. (2019). Computer-Aided Diagnosis of Multiple Sclerosis Using a Support Vector Machine and Optical Coherence Tomography Features. *Sensors* **2019**, *19*, 5323.
29. Briganti, G., & Le Moine, O. (2020). Artificial intelligence in medicine: today and tomorrow. *Frontiers in medicine* **2020**, *7*, 509744.
30. Haug, C. J., & Drazen, J. M. (2023). Artificial intelligence and machine learning in clinical medicine. *New England Journal of Medicine* **2023**, *388(13)*, 1201-1208.
31. Goyal, M., Khanna, D., Rana, P. S., Khaibullin, T., Martynova, E., Rizvanov, A. A., & Baranwal, M. (2019). Computational intelligence technique for prediction of multiple sclerosis based on serum cytokines. *Frontiers in neurology* **2019**, *10*, 781.

 doi:10.20944/preprints202603.1838.v1

32. Chen, X., Hou, H., Qiao, H., Fan, H., Zhao, T., & Dong, M. (2021). Identification of blood-derived candidate gene markers and a new 7-gene diagnostic model for multiple sclerosis. *Biological Research* **2021**, *54(1)*, 12.
33. Casalino, G., Castellano, G., Consiglio, A., Nuzziello, N., & Vessio, G. (2023). MicroRNA expression classification for pediatric multiple sclerosis identification. *Journal of Ambient Intelligence and Humanized Computing* **2023**, *14(12)*, 15851-15860.
34. Fagone, P., Mazzon, E., Mammana, S., Di Marco, R., Spinasanta, F., Basile, M. S., & Mangano, K. (2019). Identification of CD4+ T cell biomarkers for predicting the response of patients with relapsing-remitting multiple sclerosis to natalizumab treatment. *Molecular medicine reports* **2019**, *20(1)*, 678-684.
35. Ebrahimkhani, S., Beadnall, H. N., Wang, C., Suter, C. M., Barnett, M. H., Buckland, M. E., & Vafaee, F. (2020). Serum exosome microRNAs predict multiple sclerosis disease activity after fingolimod treatment. *Molecular Neurobiology* **2020**, *57(2)*, 1245-1258.
36. Giordani, P., Ferraro, M. B., & Martella, F. (2020). Non-Hierarchical Clustering. *In An Introduction to Clustering with R* **2020**, *Springer Singapore*, 75-109.
37. Vichi, M., Cavicchia, C., & Groenen, P. J. (2022). Hierarchical means clustering. *Journal of Classification* **2022**, *39(3)*, 553-577.
38. Gurevich, M., Miron, G., Falb, R. Z., Magalashvili, D., Dolev, M., Stern, Y., & Achiron, A. (2015). Transcriptional response to interferon beta-1a treatment in patients with secondary progressive multiple sclerosis. *BMC neurology* **2015**, *15(1)*, 240.
39. Coelewij, L., Adriani, M., Dönnes, P., Waddington, K. E., Ciurtin, C., Havrdova, E. K., & ABIRISK Consortium. (2024). Patients with multiple sclerosis who develop immunogenicity to interferon-beta have distinct transcriptomic and proteomic signatures prior to treatment which are associated with disease severity. *Clinical Immunology* **2022**, *267*, 110339.
40. Mésidor, M., Rousseau, M. C., Duquette, P., & Sylvestre, M. P. (2021). Classification and visualization of longitudinal patterns of medication dose: an application to interferon-beta-1a and amitriptyline in patients with multiple sclerosis. *Pharmacoepidemiology and Drug Safety* **2021**, *30(9)*, 1214-1223.
41. Kular, L., Ewing, E., Needhamsen, M., Pahlevan Kakhki, M., Covacu, R., Gomez-Cabrero, D., & Jagodic, M. (2022). DNA methylation changes in glial cells of the normal-appearing white matter in Multiple Sclerosis patients. *Epigenetics* **2022**, *17(11)*, 1311-1330.
42. Tao, J., Wang, C., & Tian, S. (2020). Feature selection based on differentially correlated gene pairs reveals the mechanism of IFN-therapy for multiple sclerosis. *Bioinform. Genom* **2020**, *8*, e8812.
43. Madhiarasan, M., & Deepa, S. N. (2016). A novel criterion to select hidden neuron numbers in improved back propagation networks for wind speed forecasting. *Applied intelligence* **2020**, *44(4)*, 878-893.
44. Yang, X. S. (2020). Nature-inspired optimization algorithms. *Academic Press* **2020**.
45. Robert, C. (2021). Statistics and Analysis of Scientific Data. *Taylor & Francis* **2021**.
46. Ponce de Leon-Sanchez, E. R., Mendiola-Santibañez, J. D., Dominguez-Ramirez, O. A., Herrera-Navarro, A. M., Vazquez-Cervantes, A., Jimenez-Hernandez, H., & Senties-Madrid, H. (2023). Fuzzy logic system for classifying multiple sclerosis patients as high, medium, or low responders to interferon-beta. *Technologies* **2023**, *11(4)*, 109.
47. Czabanski, R., Jezewski, M., & Leski, J. (2017). Introduction to fuzzy systems. *Springer International Publishing* **2017**, 23-43.
48. Rutkowska, D. (2001). Neuro-fuzzy architectures and hybrid learning. *Springer Science & Business Media* **2001**, *85*.
49. Salgado-Gomes-Sagaz, F.; Zorrilla-Muñoz, V.; Garcia-Aracil, N. (2024). Rehabilitation Technologies by Integrating Exoskeletons, Aquatic Therapy, and Quantum Computing for Enhanced Patient Outcomes. *Sensors* **2024**, *24*, 7765.
50. Prokopowicz, P., Czerniak, J., Mikołajewski, D., Apiecionek, Ł., Ślezak, D. (2017). Theory and applications of ordered fuzzy numbers: a tribute to Professor Witold Kosiński. *Springer Nature* **2017**.
51. Ackermann, M. R., Blömer, J., Kuntze, D., & Sohler, C. (2014). Analysis of agglomerative clustering. *Algorithmica* **2014**, *69(1)*, 184-215.
52. Mirjalili, V., & Raschka, S. (2020). Python machine learning. *Marcombo* **2020**.
53. Jha, A. R., & Pillai, D. G. (2021). Mastering PyTorch. *Packt Publishing* **2021**, *69(1)*, 184-215.
54. Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2016). Machine learning: algorithms and applications. *Crc Press* **2016**.

 doi:10.20944/preprints202603.1838.v1

55. Montolio, A., Martín-Gallego, A., Cegoñino, J., Orduna, E., Vilades, E., Garcia-Martin, E., & Del Palomar, A. P. (2021). Machine learning in diagnosis and disability prediction of multiple sclerosis using optical coherence tomography. *Computers in Biology and Medicine* **2021**, *133*, 104416.
56. Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science* **2021**, *2(3)*, 1-21.
57. Kotsiantis, S. B. (2013). Decision trees: a recent overview. *Artificial Intelligence Review* **2013**, *39(4)*, 261-283.
58. Stevens, E., Antiga, L., & Viehmann, T. (2020). Deep Learning with PyTorch. *Simon and Schuster* **2020**.
59. Roccetti, M., Delnevo, G., Casini, L., & Mirri, S. (2021). An alternative approach to dimension reduction for pareto distributed data: a case study. *Journal of big Data* **2021**, *8(1)*, 39.
60. Villegas-Mier, C. G., Rodriguez-Resendiz, J., Álvarez-Alvarado, J. M., Jiménez-Hernández, H., & Odry, A. (2022). Optimized random forest for solar radiation prediction using sunshine hours. *Micromachines* **2022**, *13(9)*, 1406.
61. Garcia-Villegas, E.; Lopez-Garcia, A.; Lopez-Aguilera, E. (2023). Genetic Algorithm-Based Grouping Strategy for IEEE 802.11ah Networks. *Sensors* **2023**, *23*, 862.
62. Vasuki, A. (2020). Nature-inspired optimization algorithms.*CRC Press* **2020**.
63. Torkildsen, Ø., Myhr, K. M., Bø, L. Disease-modifying treatments for multiple sclerosis–a review of approved medications. *European journal of neurology* **2016**, *23*, 18-27.