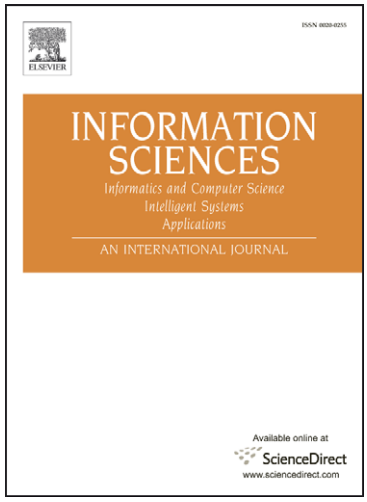

A multilayer perceptron neural network-based approach for the identification of responsiveness to interferon therapy in multiple sclerosis patients

Giuseppe Calcagno, Antonino Staiano, Giuliana Fortunato, Vincenzo Brescia-Morra, Elena Salvatore, Rosario Liguori, Silvana Capone, Alessandro Filla, Giuseppe Longo, Lucia Sacchetti

PII: S0020-0255(10)00322-1
DOI: 10.1016/j.ins.2010.07.004
Reference: INS 8744

To appear in: *Information Sciences*

Received Date: 16 November 2009
Revised Date: 13 May 2010
Accepted Date: 13 July 2010

Please cite this article as: G. Calcagno, A. Staiano, G. Fortunato, V. Brescia-Morra, E. Salvatore, R. Liguori, S. Capone, A. Filla, G. Longo, L. Sacchetti, A multilayer perceptron neural network-based approach for the identification of responsiveness to interferon therapy in multiple sclerosis patients, *Information Sciences* (2010), doi: 10.1016/j.ins.2010.07.004

A multilayer perceptron neural network-based approach for the identification of responsiveness to interferon therapy in multiple sclerosis patients☆

Giuseppe Calcagno[a,b], Antonino Staiano[*,c], Giuliana Fortunato[a,d], Vincenzo Brescia-Morra[e], Elena Salvatore[e], Rosario Liguori[a,d], Silvana Capone[a,d], Alessandro Filla[e], Giuseppe Longo[f], Lucia Sacchetti[**,a,d,1]

[a] *CEINGE Biotecnologie Avanzate, Napoli, Italy*
[b] *Dipartimento di Scienze per la Salute, Università degli Studi del Molise, Campobasso, Italy*
[c] *Dipartimento di Scienze Applicate, Università degli Studi di Napoli "Parthenope", Italy*
[d] *Dipartimento di Biochimica e Biotecnologie Mediche, Università degli Studi di Napoli "Federico II", Italy*
[e] *Dipartimento di Scienze Neurologiche, Università degli Studi di Napoli "Federico II", Italy*
[f] *Dipartimento di Scienze Fisiche, Università degli Studi di Napoli "Federico II", Italy*

Abstract

Multiple sclerosis is an idiopathic inflammatory disease characterized by multiple focal lesions in the white matter of the central nervous system. Multiple sclerosis patients are usually treated with interferon$-\beta$, but disease activity decrease in only $30\% - 40\%$ of patients. In the attempt to differentiate be-

☆Author contribution. Designed the research: GC, LS. Performed the genetic analysis: GC, GF, RL, SC. Designed the statistical approach: AS, GL. Performed statistical and machine learning analysis: AS, GF. Recruited patients and monitored IFN therapy: VB, ES, AF. Wrote the paper: GC, GF, AS, LS. Coordinated the research and Supported Financially the study: LS.

*Corresponding author: Centro Direzionale, Isola C4 - 80143 Napoli, Italy. Phone: +390815476520, Fax: +390815476514.

**Principal corresponding author: Via Pansini, 5 - 80131 Napoli, Italy. Phone: +390817463541, Fax: +390813737808.

Email addresses: antonino.staiano@uniparthenope.it (Antonino Staiano), sacchett@unina.it (Lucia Sacchetti)

[1]Work supported by grants from Regione Campania (DGRC 2362/07) and MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca) P.S. 35-126/IND.

tween responders and non responders, we screened the main genes involved in the interferon signaling pathway for 38 single nucleotide polymorphisms (SNPs) in a multiple sclerosis Caucasian population from South Italy. We then analyzed the data using a multilayer perceptron neural network-based approach, in which we evaluated the global weight of a set of SNPs localized in different genes and their association with response to interferon therapy through a feature selection procedure (a combination of automatic relevance determination and backward elimination). The neural approach appears to be a useful tool in identifying gene polymorphisms involved in the response of patients to interferon therapy: two out of five genes were identified as containing 4 out of 38 significant single nucleotide polymorphisms, with a global accuracy of 70% in predicting responder and non responder patients.

Key words: Multilayer Perceptron, Automatic Relevance Determination, Multiple Sclerosis, Gene Polymorphisms, Interpheron-β

1. Introduction

Multiple sclerosis (MS) is an inflammatory, autoimmune disease, characterized by multiple focal lesions (plaques) in the white matter that lead to axon demyelization of the central nervous system neurons. Disease susceptibility probably results from interaction of genetic background with environmental factors [15, 26]. Cytokines play a fundamental role in the pathogenesis of the disease: macrophages within the plaques secrete tumor necrosis factor-alpha (TNF$-\alpha$, which is toxic for oligodendroglia at elevated concentrations) and interleukin$-1-$beta (IL$-1-\beta$, which increases the proliferation rate of astrocytes) [26]. MS patients are usually treated with interferon$-\beta$ (IFN$-\beta$), which decreases the number of clinical relapses, slows progression of disability and reduces magnetic resonance imaging activity [11, 16, 13]. Clinical and instrumental criteria are not completely satisfactory in predicting response to treatment [12]. The immune mechanisms underlying the clinical effects of IFN$-\beta$ in MS are poorly understood, and the therapeutic effect of IFN could be related to a shift in cytokine secretion from a Th1 to a Th2 pattern [15]. Key steps in the IFN$-\beta$ signaling pathway are phosphorylation of IFN receptors (IFNAR-1 and IFNAR-2 subunits) by the Janus kinases JAK1 and TYK2 followed by recruitment, activation and release in the cytosol of STAT-1 and STAT-2 proteins, which, together with nuclear p48/IRF-9, form an active transcription factor that translocates into the

nucleus, where it promotes induction of such genes as IRF−1 [6]. Experimental evidence suggests that genetic variations in the IFN signaling pathway could be involved in MS susceptibility [9, 18], but studies of a link between these variations and IFN therapy outcome have yielded inconclusive results [18, 32, 25, 33]. The aim of our study was to try to identify gene polymorphisms in the main genes involved in IFN's mechanism of action that might distinguish between responder and non responder MS patients. We screened the IFNAR−1, IFNAR−2, STAT−1, STAT−2 and, IRF−1 genes for 38 Single Nucleotide Polymorphisms (SNPs) in a MS Caucasian population from Southern Italy. Data are usually analyzed through a multimarker analysis in which haplotype block structures in responder and non responder patients are compared. However, computational intelligence techniques are now widely used to investigate the involvement of gene polymorphisms in several classes of diseases [29, 19, 31]. We analyzed our data using a Multilayer perceptron (MLP) neural network as a classification tool, in conjunction with two common feature selection methods namely, Automatic Relevance Determination (ARD) and Backward Elimination (BE), to identify a subset of SNPs that have a predictive power in distinguishing responding from non responding MS patients to IFN therapy. We also compared the neural approach with multimarker analysis, which revealed an interesting point of intersection between the respective results, and with logistic regression that is a standard classification approach for the analysis of binary outputs. In addition to MLP, we also tested a support vector machine[2] and decision trees[8]. However, the latter two methods were not effective in distinguishing responder from non responder patients. Therefore, herein we report only the results obtained with the MLP approach.

2. Materials and Methods

2.1. Patients

182 unrelated MS patients, 110 women and 72 men (mean age 47 years ±8.6 SD), 140 with relapsing/remitting MS and 42 with secondary progressive MS according to Poser's criteria [24], came from the MS Center of the Neurologic Clinic of the University of Naples Federico II. Patients underwent IFN therapy in agreement with recommendations of the Quality Standards Subcommittee of the American Academy of Neurology (1994 and subsequent update). Each patient was examined every 3 months by the same neurologist for at least 24 months after therapy onset. The response to treatment was

evaluated according to the following clinical endpoints: 1) number of relapses in the 24 months during treatment; and 2) progression of disability, defined as an increase in Kurtzke Expanded Disability Status Scale (EDSS) of at least one point, sustained at least over 3 months. During the 24 months of follow-up, patients with no evidence of disability progression and who were relapse-free or had one relapse, were classified "responders" Patients with two or more relapses and/or an increase of EDSS of at least one point were classified "non responders". A follow-up of 24 months was decided for two reasons: (1) most IFN efficacy studies are performed for this length of time; and (2) to limit the clinical impact of developing neutralizing antibodies (NAbs) to IFN$-\beta$, which generally occurs between 6 and 24 months after treatment onset, and mostly 48 months after. Accordingly, 136 patients were classified as responders and 46 patients as non responders. The study conformed to the ethical guidelines of the Helsinki II Declaration, and each subject gave their informed consent to the study.

2.2. Haplotype analysis

A haplotype is a set of SNPs that are statistically associated. It is thought that these associations, and the identification of a few alleles[2] of a haplotype block, can unambiguously identify all other polymorphic sites in its region. This information is necessary for investigations of the genetics underlying common diseases. The term *haplotype block* is used to refer to an individual collection of polymorphisms (SNPs in our case), allele mutations, within a genetic segment. Allele frequencies were calculated by allele counting and departure from Hardy-Weinberg expectation was evaluated by χ^2 analysis. Associations of the SNPs with categorical variables were evaluated with the χ^2 test. The Haploview 3.2 software[3] was used to examine haplotype block structures and to generate haplotypes in these blocks. The resultant block structure was determined according to the algorithm of Gabriel et al. [10], where a block is created if it is 95% of informative pair-wise SNP comparisons show a strong linkage disequilibrium with D' (a normalized measure of allelic association) equal to or greater than 0.8. Linkage disequilibrium [17] is the non-random association of alleles or genes at two or more specific location of a chromosome (loci). Linkage disequilibrium describes a situation

[2] one of a series of different forms of a gene

[3] www.broad.mit.edu/mpg/haploview/download.php

in which some combinations of alleles or genetic markers occur more or less frequently in a population than would be expected from a random formation of haplotypes from alleles based on their frequencies. We carried out a permutation test (the haplotypes of responder and non responder patients were permuted 100.000 times) to detect differences in haplotype distribution between groups. The permutation test checks the null hypothesis, i.e. case and control haplotypes are a random sample from a single set of haplotype frequencies, versus cases are more similar to each other than to controls.

2.3. Logistic Regression

Logistic Regression (LR) is a statistical classification model well suited (and thus frequently used) for medical applications in which a two class decision problem needs to be addressed. This method could be extended to multiclass problems. In this study, an LR model was fitted to the data by using the iterative reweighted least squares (IRLS) algorithm [2] to set the weights and the intercept.

2.4. MLP-based approach

2.4.1. MLP neural network

An MLP consists of two layers of adaptive weights with full connectivity between inputs and hidden units, and between hidden units and outputs (see Figure 1). If we denote by $x_i, i = 1, \ldots, d$ the input values given to the network, where d is the number of SNPs for each patient, the first layer forms M linear combinations of x_i's which give rise to the intermediate activation variables a_j

$$a_j = \sum_{i=1}^{d} w_{ji} x_i + w_{j0}, j = 1, \ldots M$$

with one variable a_j associated with each hidden unit. w_{ji} represents the elements of the first layer weight matrix and w_{j0} are the bias parameters associated with the hidden units j. The variables a_j are then transformed by the nonlinear activation functions of the hidden layer. By using the $tanh$ activation function, the outputs of the hidden units are then given by $z_j = tanh(a_j), j = 1, \ldots, M$.

The z_j are then transformed by the second layer of weights and biases to give second-layer activation values a_k

$$a_k = \sum_{j=1}^{M} w_{kj} z_j + w_{k0}, k = 1, \ldots c,$$

where c is the total number of outputs. Finally, these values are passed through the output unit activation function to give output values y_k where $k = 1, \ldots c$ (in our case we just use a single output neuron for two classes). In classification problems with multiple mutually exclusive classes, a logistic sigmoidal activation function applied to each of the network output is considered, so that

$$y_k = \frac{1}{1 + \exp(-a_k)}, k = 1, \ldots, c.$$

The network is trained by the backpropagation algorithm considering the cross-entropy error function over N input patterns:

$$E = -\sum_{n=1}^{N} \{t^n \ln y^n + (1 - t^n) \ln(1 - y^n)\}$$

when comparing the network output activations y^n with the targets t^n, $n = 1, \ldots, N$ (i.e. the class labels of the patients: 1 for responder and 0 for non responder).

The network performance is evaluated by counting the number of correct patient class labels predicted by the network. Precision (P) and recall (R) were also computed as:

$$P = \frac{Num_{TR}}{Num_{TR} + Num_{FR}} \times 100$$

$$R = \frac{Num_{TR}}{Num_{TR} + Num_{FNR}} \times 100$$

where Num_{TR}, Num_{FR} and Num_{FNR} are the numbers of true responders, false responders and false non responders, respectively.

2.4.2. MLP neural network data preprocessing

SNP calls at each site were converted into numeric values assigned according to control patient frequencies: 1 for homozygous major allele, 2 for heterozygous and 3 for homozygous minor allele. It should be noted that for the MLP, which treats SNPs as continuous variables, this representation assumes that heterozygous alleles are half-way between the homozygous alleles and the two alleles are not treated symmetrically. From a machine learning perspective, our patients and the corresponding SNPs form a data matrix consisting of 182 rows and 38 columns.

2.4.3. Feature selection

Automatic relevance determination

Inferring the relative importance of features has attracted much attention in the machine learning and statistics research communities [14]. Automatic relevance determination (ARD) is a Bayesian method used to assess the importance of features. It can be applied to standard feed-forward neural networks [20, 22, 28] and has many applications [27, 30, 3]. This approach optimizes model evidence (marginal likelihood), the classic criterion for Bayesian modeling, and generates hyperparameters that represent the relevance of different input features. A separate hyperparameter α_i, $i = 1, \ldots, d$, is associated with each i-th input variable of an MLP. Each of these α_i represents the inverse variance of the distribution of the weights fanning out from a particular input. These hyperparameters are modified during training. The initial prior distribution $P(w_i)$ of the weights w_i is assumed to have a Gaussian distribution with zero mean which is in the form

$$P(w_i) = \frac{1}{Z_w(\alpha_i)} \exp(-\frac{\alpha_i}{2} \|w_i\|^2),$$

where w_i are the weights fanning out from the i-th input, $Z_w(\alpha_i)$ is the normalization constant, and α_i is the hyperparameter of weight w_i. As training of the MLP progresses, each hyperparameter α_i is re-estimated to a new value $\alpha_i^{'}$ using [20]

$$\alpha_i^{'} = \frac{\gamma_i}{\mu_i^2},$$

where μ_i is the posterior mean of the weights that corresponds to an input i, and $\gamma_i \in [0, 1]$ is a measure of how "well determined" its corresponding

parameter w_i is by the data [20]. The quantity γ is calculated from the eigenvalues λ_i of the matrix W of weights and biases in the network using

$$\gamma = \sum_i \frac{\lambda_i}{\lambda_i + \alpha}.$$

The quantity γ effectively captures the influence of the likelihood and the prior (i.e., when $\gamma_i \approx 1$, α_i is small and w_i is highly constrained by the prior). On completion of training, a small α_i means the corresponding input (e.g., the SNP associated to input i) is important in discriminating MS responder and non responder patients. Conversely, a large α_i indicates that the i-th SNP is less important to discriminate between responder and non responder MS patients.

Backward Elimination

Backward elimination (BE)[2] is a greedy strategy in which one starts with the set of all the variables (i.e, the features) and progressively eliminates the least promising one, while evaluating the performance of the learning system each time with the new subset of variables. This method yields nested subset of variables. In BE, we start with the full feature-subset, i.e., the feature-subset at the onset has all the d input variables in it. Now, each of the features is dropped one by one, and d models are learnt on subsets that contain $d-1$ features each. This requires pre-setting a learning algorithm (and its associated parameters for model training, MLP and LR in our case). We obtain d models at this point, and the model that performs the best is now chosen. Since each of these models has $d-1$ features in them, by choosing the best model, we have selected the best feature-subset with $d-1$ features and thus, we have eliminated the worst feature (out of d) for modeling the given property. The feature-subset now contains $d-1$ features (as chosen in the previous step). These iterations are continued until either a pre-specified target size (desired number of descriptors) is reached or the desired classification accuracy is obtained.

2.4.4. Leave-one-out cross-validation

Because of the small number of patients in this study, the MLP and LR were trained and evaluated using leave-one-out cross-validation. Cross-validation is a way to make the best use of a data set for both learning and validation. If the data set consists of n patients, the leave-one-out cross-validation involves using a single patient from the original sample as the

validation data, and the remaining n-1 patients as the training data. This is repeated such that each patient in the sample is used once as the validation data. The aggregate test results from all the n phases of the cross-validation would be used to obtain a final estimate of the prediction accuracy. Before the model training, we had to sample further the data set of patients. In fact, the classes of responder and non responder patients were very similar (in terms of SNP site values) and heavily overlapped. This led, as several preliminary experiments proved, to the overfit of the class of responder patients (in fact, the fractions of responder and non responder patients are unbalanced, i.e. 136 responders and 46 non responders). In practice, the MLP (and LR) provided "responders" as output in all the cases whatever the class of patients given as input. Therefore, in order to obtain prediction results not biased by the unbalanced size of our classes, we randomly sampled from the set of patients labeled as responders in order to build a new data set with an equal number of responders and non responders patients to IFN therapy. The random sampling was repeated 200 times, and each time the MLP and the LR were trained on the resulting data set in combination with leave-one-out cross-validation.
LR, MLP, ARD and BE algorithms are written in Matlab using the Netlab Toolbox [21].

3. Results

3.1. Identification of SNPs with haplotype analysis

Tables 1, 2 and 3 show the allele and genotype association for each polymorphism investigated, and the response to IFN treatment in responder and non responder MS patients. There were no differences in genotype frequencies of the 38 SNPs apart from rs1547550 in the STAT−1 gene ($p = 0.016$). In this case, the GG genotype was more frequent in non responder (65.9%) than in responder (42.2%) MS patients. MS patient bearing the GG genotype, compared with CC+CG genotype, had a more than double risk (O.R.= 2.64, 95% CI: $1.30 - 5.37$, $p = 0.005$) of not responding to IFN treatment. The haplotype analysis of the 38 SNPs revealed an extensive linkage disequilibrium across all the five genes (data not shown). The haplotype analysis between responder and non responder patients identified a strong linkage disequilibrium only in $IRF - 1$ gene. The $IRF - 1$ selected haplotype included 7 SNPs, one in intron 1, one in intron 3, four in intron 9 and one in exon

10. As shown in Table 4, the CTATTGA haplotype was present only in non responder MS patients ($p = 0.00$, permuted 100.000 times).

3.2. Identification of SNPs with automatic relevance determination and backward elimination

ARD assigned to each of the 38 input features (i.e. the total number of SNPs in our patients), a separate regularization coefficient (called "hyperparameter"and denoted by α_i). The ARD model was run 5 times because result can vary depending on the initial values of hyperparameters. Based on the resulting α_i, the inputs were ranked according to their relevance. The median of the 5 ranks was used as the main criterion to select which input to drop. Figure 2 shows the distribution of the hyperparameter values after ARD processing. Because ARD doesn't indicate the threshold for significant SNPs, we selected the 10 best ranked SNPs (i.e., those with hyperparameter values below 25), namely, IFNAR−2 rs2834154, rs2284549, rs2236756, rs2236757, STAT−1 rs1547550, rs2066803, IRF−1 rs2070723, rs2070731, IFNAR−1 rs2243590, and rs2252931. Choosing this threshold is highly subjective. Usually, in the absence of prior knowledge about the importance of input features, one may choose a value above which there is a gap in the distribution of hyperparameter values. However, it is advisable not to drop a large number of features. Based on our experimental results, we selected a threshold of 25 because it allowed us to retain a meaningful set of SNPs. Next, we trained the MLP using BE to compute the minimal set of discriminating SNPs starting from the ones resulting from ARD. The most discriminant minimal subset of SNPs contained 4 elements: STAT−1 rs1547550, rs2066803, IRF−1 rs2070723, and rs2070731. It is noteworthy that simultaneous use of multiple SNPs had a significantly better predictive power than any one SNP alone.

3.3. Identification of responder and non responder MS patients with MLP

The MLP was trained with all the 38 SNPs, the 10 best ranked by ARD and the minimal subset identified by BE. The network was provided, as usual, with the patients (i.e., the corresponding 38 SNPs associated to each patient) as input, and the corresponding target class label (1 for responders and 0 for non responders). The network we used had two layers of adaptive weights with 14 hidden neurons and one output neuron. With a network output greater than 0.5, the class label predicted by the network would be 1 (responder patients) whereas, with a network output less than 0.5, the class

label predicted by the network would be 0(non responder patients). This is because we used cross-entropy as error criterion in network learning and the logistic sigmoid activation function in the output neuron, so that the output of the network represents posterior class probabilities. As shown in Table 5, the MLP model correctly predicted a positive response to therapy with an accuracy above 70%. These results suggest that, according to the MLP, the STAT−1 rs1547550, rs2066803, and IRF−1 rs2070723, rs2070731 SNPs are collectively correlated to the response of patients to IFN therapy. Our estimated error rate of about 30% of the total number of patients is, preliminarily, acceptable given the small size of the sample with which to train the network, and considering that no prior information was used to aid the selection of the most significant SNPs for neural network learning.

3.4. LR classification of MS patients

As with the MLP neural network, the LR model was fitted through leave-one-out cross-validation to the data set by using 38 SNPs, 10 SNPs best ranked by ARD and the minimal subset of SNPs identified by BE. The LR results are depicted in Table 6. Figure 3 shows the ROC curves comparison of MLP and LR, clearly indicating that MLP outperforms LR. Nonetheless, as Table 6 suggests, the LR performance agrees with the MLP performance in terms of SNPs correlations since the best prediction accuracy is obtained on the STAT−1 rs1547550, rs2066803, and IRF−1 rs2070723, rs2070731 SNPs selected by the ARD and BE feature selection procedure.

4. Discussion

The MLP predicted responder and non responder patients with an acceptable accuracy on the basis of 4 SNPs of the 38 investigated by ARD and BE procedures. The selected SNPs localized in the STAT−1 and IRF−1 genes that are associated with IFN response in MS patients. The STAT−1 gene is involved in immunological self tolerance. In fact, STAT−1 gene-deficient animals have increased susceptibility to autoimmune disease [15]. Moreover, virally mediated inhibition of STAT−1 function has been associated with cellular resistance to IFNs [34]. In our MS population, two of the selected STAT−1 intron polymorphisms (rs1547550, rs2066803) could alter STAT−1 transcript levels by interfering with pre-mRNA processing [7]. Pathogenetic sequence variations that do not cause protein changes but result in abnormal splicing have been described previously [7]. In particular, STAT−1 rs2066803

is localized at position 98 from the c-terminal of exon 24 and it could take part in this mechanism. Regarding the IRF−1 gene, we previously reported an association between seven SNPs (six intronic and one in the 3'UTR region) and MS susceptibility [9]. The present study confirms and extends the role of the IRF−1 gene in the response to IFN−β in MS disease. To our knowledge, this is the first study to suggest an association between IRF−1 genetic variations and response to IFN treatment. Recently, it was demonstrated that IRF−2, a functional antagonist of IRF−1, IRF−4, IRF−6 and IRF−8 expression levels, differs between responder and non responder MS patients, thereby implicating IRF transcription factors in drug response [1].

5. Conclusions

Human genome analysis and high-throughput techniques have produced a mass of complex biological data often resulting in an analytic bottleneck. Traditional methods of statistical analysis could often benefit from cooperation with machine learning algorithms to cope with this flood of information. These algorithms are designed to clean up a variety of patterns, both linear and nonlinear, from large, noisy, and complex data sets that may also contain a great deal of irrelevant information [29, 19, 31]. We have studied the relationship between response to IFN therapy in MS using 38 SNPs in 5 genes. A novelty of our study is the use of a neural network to investigate the involvement of gene polymorphisms in the response of MS to INF treatment in addition to a traditional haplotype approach. The two methods identified the SNPs that characterize responders and non responder patients. The MLP procedure provided a global prediction accuracy of 70%, which is satisfactory given the wide inter-individual heterogeneity in the response to IFN therapy. In this context, the neural network approach may be used to identify meaningful SNPs and to classify a given population. The MLP neural network allows the simultaneous use of multiple SNPs which, as our experiments proved, has significantly better predictive power than any one SNP alone. This neural network approach, in which multifactorial SNPs that form a single biological mechanism can be combined, is an important step away from the traditional method of looking at single SNP associations. The neural network seems to be a useful tool in identifying gene polymorphisms involved in the response of MS patients to IFN treatment. This approach allows us to determine, without prior domain knowledge, the global contribution of several SNPs located in different genes, and to highlight a hitherto

unknown association between STAT−1 and IRF−1 gene polymorphisms and response to IFN therapy. The patient data available to us is, of course, of limited size and therefore no definitive and significant statistical conclusions could be outlined at this stage. Nonetheless, our results indicate that using a larger number of patients, this approach could be useful, in its present form or by combining it with soft computing techniques [23], such as rough sets or fuzzy sets (giving the MLP more flexibility to represent discrete data as in our study), to discover and demonstrate new associations and genes for better customized IFN−β treatment in MS patients.

Acknowledgments

We are grateful to Jean Ann Gilder for editing the text.

Figure 1: **The general architecture of an MLP (Multi Layer Perceptron).** The MLP standard structure consists of two layers of adaptive weights, i.e. the input-to-hidden weights and the hidden-to-output weights, and a number M of hidden neurons and c output neurons. The bias parameters in the first layer are shown as weights from an extra input having a fixed value of $x_0 = 1$. Similarly, the bias parameters in the second layer are shown as weights from an extra hidden unit, with activation fixed at $z_0 = 1$.

Figure 2: **ARD (Automatic Relevance Determination) SNPs selection procedure.** ARD assigns a value, called "hyperparameter" and denoted by α, for each SNP. Thus, each of the 38 SNPs had a corresponding α_i, for $i = 1, \ldots, 38$. A low hyperparameter value means that its corresponding SNP is one of the most significant among those investigated. The histogram shows the number of SNPs (y axis:counts) whose hyperparameters have a given value (x axis). For example, 10 SNPs have a value of the corresponding hyperparameters less or equal to 25 (the first two bars). The values of the hyperparameters on the x axis are computed as the median over 5 runs of the procedure.

Figure 3: MLP and LR comparison: ROC curves. The dotted curves represent the non responder patients class, while the continuous curve corresponds to the responder patients class.

Table 1: **Allele and genotype association between patients with multiple sclerosis treated with interferon (non responders=N.R.; responders=R.) and the 25 SNPs in the IFNAR−1 and IFNAR−2 genes.**

GENE	SNP ID	N. of patients		Genotype	Genotype Freq.		Count (Freq.) Major allele		P-value
		N.R.	R.		N.R.	R.	N.R.	R.	
IFNAR−1				TT	23.8	21.5			
	rs2243590	42	130	TC	64.3	44.6	**T** 47 (0.560)	**C** 146 (0.562)	.0532
				CC	11.9	33.8			
				GG	63.6	51.1			
	rs2252931	44	133	GA	36.4	40.6	**G** 72 (0.818)	**G** 190 (0.714)	.0541
				AA	-	8.3			
				GG	60.9	58.5			
	rs2243600	46	135	GT	32.6	37.0	**G** 71 (0.772)	**G** 208 (0.770)	.9785
				TT	6.5	4.4			
IFNAR−2				AA	11.6	10.8			
	rs2300370	43	120	AG	48.8	46.7	**G** 55 (0.640)	**G** 158 (0.658)	.7533
				GG	39.5	42.5			
				AA	70.7	76.0			
	rs2248412	41	121	AG	19.5	20.7	**A** 66 (0.805)	**A** 209 (0.864)	.1993
				GG	9.8	3.3			
				AA	34.8	40.3			
	rs2834154	46	134	AC	54.3	48.5	**A** 57 (0.620)	**A** 173 (0.646)	.6547
				CC	10.9	11.2			
				CC	46.7	34.8			
	rs2154430	45	135	CT	33.3	51.9	**C** 57 (0.633)	**C** 164 (0.607)	.6617
				TT	20.0	13.3			
				AA	38.6	40.7			
	rs2236756	44	135	AC	50.0	47.4	**A** 56 (0.636)	**A** 174 (0.644)	.8907
				CC	11.4	11.9			
				AA	4.7	11.4			
	rs2284549	43	114	AT	53.5	41.2	**T** 59 (0.686)	**T** 155 (0.680)	.9160
				TT	41.9	47.4			
				AA	6.5	10.4			
	rs2284551	46	135	AG	54.3	42.2	**G** 61 (0.663)	**G** 185 (0.685)	.6943
				GG	39.1	47.4			
				AA	37.8	40.7			
	rs2834163	45	135	AG	51.1	46.7	**A** 57 (0.633)	**A** 173 (0.641)	.8992
				GG	11.1	12.6			
				AA	7.5	9.5			
	rs2236757	40	126	AG	50.0	41.3	**G** 54 (0.675)	**G** 176 (0.698)	.6925
				GG	42.5	49.2			
				AA	10.9	15.9			
	rs2236758	46	132	AG	52.2	47.7	**G** 58 (0.630)	**G** 159 (0.602)	.6335
				GG	37.0	36.4			

Table 2: **Allele and genotype association between patients with multiple sclerosis treated with interferon (non responders=N.R.; responders=R.) and the 25 SNPs in the STAT−1 gene. ∗ Statistically significant at χ^2 test.**

GENE	SNP ID	N. of patients		Genotype Freq.			Count (Freq.) Major allele		*P*-value
		N.R.	R.		N.R.	R.	N.R.	R.	
STAT−1	rs2066802	44	134	AA	90.9	81.3	**A** 84 (0.955)	**A** 243 (0.907)	.1547
				AG	9.1	18.7			
				GG	-	-			
	rs2066794	46	133	GG	100	99.2	**G** 92 (1.000)	**G** 265 (0.996)	.5559
				TG	-	0.8			
				TT	-	-			
	rs2066805	46	134	AA	89.1	94.8	**A** 87 (0.946)	**A** 261 (0.974)	.1931
				AG	10.9	5.2			
				GG	-	-			
	rs2066800	44	134	AA	97.7	97.0	**A** 87 (0.989)	**A** 264 (0.985)	.8054
				AG	2.3	3.0			
				GG	-	-			
	rs2066797	46	133	AA	97.8	89.5	**A** 91 (0.989)	**A** 252 (0.947)	.0848
				AG	2.2	10.5			
				GG	-	-			
	rs2066795	46	135	CC	60.9	64.4	**C** 74 (0.804)	**C** 215 (0.796)	.8680
				CT	39.1	30.4			
				TT	-	5.2			
	rs2066799	44	133	AA	95.5	91.9	**A** 86 (0.977)	**A** 255 (0.959)	.4206
				AG	4.5	8.1			
				GG	-	-			
	rs2066801	46	136	CC	100	100	**C** 92 (1.000)	**C** 272 (1.000)	—
				AC	-	-			
				AA	-	-			
	rs1547550	44	135	CC	6.8	10.4	**G** 70 (0.795)	**G** 178 (0.659)	.0162*
				CG	27.3	47.4			
				GG	65.9	42.2			
	rs2066793	46	135	CC	67.4	71.1	**C** 74 (0.804)	**C** 224 (0.830)	.5830
				CT	26.1	23.7			
				TT	6.5	5.2			
	rs2066803	46	123	GG	91.3	94.3	**G** 88 (0.957)	**G** 238 (0.967)	.6280
				GT	8.7	4.9			
				TT	-	0.8			
	rs2066818	46	134	GG	93.5	95.5	**G** 89 (0.967)	**G** 262 (0.978)	.5880
				TG	6.5	4.5			
				TT	-	-			

Table 3: **Allele and genotype association between patients with multiple sclerosis treated with interferon (non responders=N.R.; responders=R.) and the 13 SNPs in STAT−2 and IRF−1 genes.**

GENE	SNP ID	N. of patients N.R.	N. of patients R.	Genotype Freq.	Genotype Freq. N.R.	Genotype Freq. R.	Count (Freq.) Major allele N.R.	Count (Freq.) Major allele R.	P value
STAT−2	rs2066819	46	135	GG	93.5	95.6	**G** 89 (0.967)	**G** 264 (0.978)	.5806
				AG	6.5	4.4			
				AA	-	-			
	rs2020854	46	134	AA	93.5	94.0	**A** 89 (0.967)	**A** 260 (0.970)	.8945
				AG	6.5	6.0			
				GG	-	-			
	rs2066811	46	134	AA	100	99.3	**A** 92 (1.000)	**A** 267 (0.996)	.5574
				GA	-	0.7			
				GG	-	-			
	rs2066807	46	131	CC	93.5	95.4	**C** 89 (0.967)	**C** 256 (0.977)	.6108
				GC	6.5	4.6			
				GG	-	-			
	rs2066808	46	136	TT	93.5	94.1	**T** 89 (0.967)	**T** 264 (0.971)	.8769
				CT	6.5	5.9			
				CC	-	-			
IRF−1	rs2070721	44	135	AA	4.9	34.1	**A** 56 (0.636)	**A** 154 (0.570)	.2749
				AC	45.5	45.9			
				CC	13.6	20.0			
	rs2070723	37	133	TT	56.8	43.6	**T** 57 (0.770)	**T** 177 (0.665)	.0850
				TC	40.5	45.9			
				CC	2.7	10.5			
	rs2070728	43	132	GG	53.5	45.5	**G** 63 (0.733)	**G** 178 (0.674)	.3105
				GA	39.5	43.9			
				AA	7.0	10.6			
	rs2070729	45	135	TT	13.3	20.0	**G** 58 (0.644)	**G** 154 (0.570)	.2161
				TG	44.4	45.9			
				GG	42.2	34.1			
	rs2070730	45	131	CC	47.8	43.5	**C** 65 (0.707)	**C** 174 (0.664)	.4550
				CT	45.7	45.8			
				TT	6.5	10.7			
	rs2070731	45	121	AA	48.9	44.6	**A** 65 (0.722)	**A** 162 (0.669)	.3577
				AG	46.7	44.6			
				GG	4.4	10.7			
	rs839	46	136	GG	50.0	43.4	**G** 66 (0.717)	**G** 179 (0.658)	.2945
				GA	43.5	44.9			
				AA	6.5	11.8			
	rs2070727	46	136	GG	60.9	57.4	**G** 56 (0.608)	**G** 156 (0.573)	.175
				TG	39.1	42.6			
				TT	-	-			

Table 4: **The IRF−1 haplotype structure obtained in the responder and non responder multiple sclerosis patients investigated by the SNPs analysis. ∗ Statistically significant at χ^2 test.**

IRF−1 hapl.	rs2070721 A/C	rs2070723 C/T	rs2070728 G/A	rs2070729 G/T	rs2070730 C/T	rs2070731 A/G	rs839 G/A	Total Freq.	Relative Freq. N.R	Relative Freq. R.	P-value	Perm. P-value
1	A	T	G	G	C	A	G	56.4	61.9	54.6	.218	.999
2	C	C	A	T	T	G	A	23.9	16.3	26.4	.048*	.738
3	C	T	G	T	C	A	G	8.5	7.6	8.8	.727	1.000
4	C	C	A	T	T	A	A	3.3	1.1	4.0	.176	.999
5	**C**	**T**	**A**	**T**	**T**	**G**	**A**	1.9	7.6	0.0	.000*	.000*
6	C	T	G	T	T	G	A	1.1	2.2	0.7	.246	1.000

Table 5: **MLP performance. Prediction accuracy, precision and recall are shown as mean (with standard deviation in parentheses) of 200 runs of the experiment. The MLP was trained with all the 38 SNPs and, 10 SNPs selected by the ARD procedure and with the 4 SNPs determined by BE.**

	38 SNPs	***10 SNPs***	***4 SNPs***
Prediction accuracy (%)	61.68 (5.63)	69.944 (4.17)	70.78 (5.46)
Precision (%)	64.213 (1.1)	71.741 (0.44)	74.517 (0.7)
Recall (%)	49.67 (1.21)	62.372 (0.67)	60.674 (0.89)

Table 6: **LR performance. Prediction accuracy, precision and recall are shown as mean (with standard deviation in parentheses) of 200 runs of the experiment. The LR was trained with all the 38 SNPs and, 10 SNPs selected by the ARD procedure and with the 4 SNPs determined by BE.**

	38 SNPs	***10 SNPs***	***4 SNPs***
Prediction accuracy (%)	58.642 (4.53)	65.32 (3.39)	67.483 (5.04)
Precision (%)	61.21 (1.0)	69.841 (0.35)	72.22 (0.9)
Recall (%)	49.88 (1.28)	60.11 (0.65)	58.84 (0.71)

References

References

[1] S. E. Baranzini, P. Mousavi, J. Rio, S. J. Caillier, A. Stillman, P. Villoslada, M. M. Wyatt, M. Comabella, L. D. Greller, R. Somogyi, X. Montalban, J. R. Oksenber, Transcription-based prediction of response to ifnbeta using supervised computational methods, PLoS Biol 3 (2005) 166–176.

[2] C. M. Bishop, Pattern Recognition and Machine Learning, Springer-Verlag, New York, 2006.

[3] A. Browne, A. Jakary, S. Vinogradov, Y. Fu, R. F. Deicken, Automatic relevance determination for identifying thalamic regions implicated in schizophrenia, IEEE Trans. Neural Networks 19 (2008) 1101–1107.

[4] E. Byun, S. J. Caillier, X. Montalban, P. Villoslada, O. Fernandez, D. Brassat, M. Comabella, J. Wang, L. F. Barcellos, S. E. Baranzini, J. R. Oksenberg, Genome-wide pharmacogenomic analysis of the response to interferon beta therapy in multiple sclerosis, Arch Neurol 65 (2008) 337–344.

[5] S. Cunningham, C. Graham, M. Hutchinson, A. Droogan, K. O'Rourke, C. Patterson, G. McDonnell, S. Hawkins, K. Vandenbroeck, Pharmacogenomics of responsiveness to interferon ifn-beta treatment in multiple sclerosis: A genetic screen of 100 type i interferon-inducible genes, Clin. Pharmacol. Ther. 78 (2005) 635–646.

[6] T. Decker, S. Stockinger, M. Karaghiosoff, M. Muller, P. Kovarik, Ifns and stats in innate immunity to microorganisms, J Clin Invest 109 (2002) 1271–1277.

[7] J. Duan, M. S. Wainwright, J. M. Comeron, N. Saitou, A. R. Sanders, J. Gelernter, P. V. Gejman, Synonymous mutations in the human dopamine receptor d2 (drd2) affect mrna stability and synthesis of the receptor, Hum Mol Genet 12 (2003) 205–216.

[8] R. O. Duda, P. E. Hart, D. G. Stork, Pattern Classification, 2nd Edition, Wiley, 2001.

[9] G. Fortunato, G. Calcagno, V. Bresciamorra, E. Salvatore, A. Filla, S. Capone, R. Liguori, S. Borelli, I. Gentile, F. Borrelli, G. Borgia, L. Sacchetti, Multiple sclerosis and hepatitis c virus infection are associated with single nucleotide polymorphisms in interferon pathway genes, J Interferon Cytokine Res 28 (2008) 141–152.

[10] S. B. Gabriel, S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S. N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E. S. Lander, M. J. Daly, D. Altshuler, The structure of haplotype blocks in the humane genome, Science 296 (2002) 2225–29.

[11] I. M. S. Group, Interferon beta-1b is effective in relapsing-remitting multiple sclerosis: I. clinical results of a multicenter, randomized, double-blind, placebo-controlled trial, Neurology 43 (1993) 655–661.

[12] M. S. C. Group, H. Wiendl, K. V. Toyka, P. Rieckmann, R. Gold, H. P. Hartung, H. Hohlfeld, Basic and escalating immunomodulatory treatments in multiple sclerosis: Current therapeutic recommendations, Journal of Neurology 255 (2008) 1449–1463.

[13] P. S. Group, Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis, Lancet 352 (1998) 1498–1504.

[14] I. Guyon, A. Elissee, An introduction to variable and feature selection, The Journal of Machine Learning Research, Special issue on variable and feature selection (2003) 1157–1182.

[15] D. A. Hafler, J. M. Slavik, D. E. Anderson, K. C. O'Connor, P. De Jager, C. Baecher-Allan, Multiple sclerosis, Immunol Rev 204 (2005) 208–231.

[16] L. D. Jacobs, D. L. Cookfair, R. A. Rudick, R. M. Herndon, J. R. RIchert, A. M. Salazar, J. S. Fischer, D. E. Goodkin, C. V. Granger, J. H. Simon, J. J. Alam, D. M. Bartoszak, D. N. Bourdette, J. Braiman, C. M. Brownscheidle, M. E. Coats, S. L. Cohan, D. S. Dougherty, R. P. Kinkel, M. K. Mass, F. E. Munschauer, R. L. Priore, P. M. Pullicino, B. J. Scherokman, R. H. e. a. Whitham, Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis, Ann Neurol 39 (1996) 285–294.

[17] H-Y. Jung, Y-J. Park, Y-J. Kim, J-S. Park, K. Kimm, I. Koh, New methods for imputation of missing genotype using linkage disequilibrium and haplotype information, Information Sciences 177 (2007) 804–814.

[18] L. Leyva, O. Fernandez, M. Fedetz, E. Blanco, V. E. Fernandez, B. Oliver, A. Leon, M. J. Pinto-Medel, C. Mayorga, M. Guerrero, G. Luque, A. Alcina, F. Matesanz, Ifnar1 and ifnar2 polymorphisms confer susceptibility to multiple sclerosis but not to interferon-beta treatment response, J of Neuroimmunology 163 (2005) 165–171.

[19] J. Listgarten, S. Damaraju, B. Poulin, L. Cook, J. Dufour, A. Driga, J. Mackey, D. Wishart, R. Greiner, B. Zanke, Predictive models for breast cancer susceptibility from multiple single nucleotide polymorphisms, Clinical Cancer Research 10 (2004) 2725–2737.

[20] D. J. C. MacKay, Bayesian methods for backpropagation networks, in: Domany, van Hemmen, Shulten (Eds.), Models of Neural Networks III, Springer-Verlag, New York, 1994.

[21] I. T. Nabney, NETLAB: algorithms for pattern recognition, Springer-Verlag, New York, 2002.

[22] R. M. Neal, Bayesian learning for neural networks, Ph.D. thesis, University of Toronto, Department of Statistics (1994).

[23] S. K. Pal, Soft data mining, computational theory of perceptions, and rough-fuzzy approach Information Sciences, Vol. 163, Issues 1-3, (2004) 5–12.

[24] C. M. Poser, D. W. Paty, L. Scheinberg, W. I. McDonald, F. A. Davis, G. C. Ebers, K. P. Johnson, W. A. Sibley, D. H. Silberberg, W. W. Tourtellotte, New diagnostic criteria for multiple sclerosis: guidelines for research protocols, Ann Neurol 13 (1983) 227–31.

[25] U. Sriram, L. Barcellos, P. Villoslada, J. Rio, S. Baranzini, S. Caillier, A. Stillman, S. Hauser, X. Montalban, J. Oksenberg, Pharmacogenomic analysis of interferon receptor polymorphisms in multiple sclerosis, Genes and Immunity 4 (2003) 147–152.

[26] A. Svejgaard, The immunogenetics of mutiple sclerosis, Immunogenetics 60 (2008) 275–286.

[27] H. H. Thodberg, A review of bayesian neural networks with an application to near infrared spectroscopy, IEEE Trans Neural Networks 7 (1996) 56–72.

[28] M. E. Tipping, Bayesian inference: an introduction to principles and practice in machine learning, in: O. Bousquet, U. von Luxburg, G. Rätsch (Eds.), Advanced Lectures on Machine Learning, Springer, 2004, pp. 41–62.

[29] Y. Tomita, S. Tomida, Y. Hasegawa, Y. Suzuki, T. Shirakawa, T. Kobayashi, H. Honda, Artificial neural network approach for selection of susceptible single nucleotide polymorphisms and construction of prediction model on childhood allergic asthma, BMC Bioinformatics 5 (2004) 120–132.

[30] I. Ulusoy, C. M. Bishop, Automatic relevance determination for the estimation of relevant features for object recognition, in: Proceedings of the 14th Conference on Signal Processing and Communication Applications, 2006, pp. 1–4.

[31] P. Unneberg, M. Stromberg, F. Sterky, Snp discovery using advanced algorithms and neural networks, Bioinformatics 21 (2005) 2528–2530.

[32] K. Vandenbroeck, C. Hardt, J. Louage, P. Fiten, S. Jackel, I. Ronsse, J. T. Epplen, L. M. E. Grimaldi, T. Olsson, M. G. Marrosu, A. Billiau, G. Opdenakker, Lack of association between the interferon regulatory factor-1 (irf1) locus at 5q31.1 and multiple sclerosis in germany, northern italy, sardinia and sweden, Genes and Immunity 1 (2000) 290–292.

[33] B. Weinstock-Guttman, D. Badgett, K. Patrick, L. Hartrich, D. Hall, M. Baier, J. Feichter, M. Ramanathan, Genomic effects of interferon-beta in multiple sclerosis patients, Journal of Immunology 171 (2003) 2694–2702.

[34] L. H. Wong, H. Sim, M. Chatterjee-Kishore, I. Hatzinisiriou, R. J. Devenish, G. Stark, S. J. Ralph, Isolation and characterization of a human stat1 gene regulatory element. inducibility by interferon (ifn) types i and ii and role of ifn regulatory factor-1, J Biol Chem 31 (2002) 19408–19417.