IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

SMURF: Systematic Methodology for Unveiling Relevant Factors in retrospective data on chronic disease treatments

FRANKLIN PARRALES BRAVO [1,6], ALBERTO A. DEL BARRIO GARCÍA [1],(Senior Member, IEEE), ANA BEATRIZ GAGO VEIGA [2], MARÍA MERCEDES GALLEGO [2], MARINA RUIZ [3], ANGEL GUERRERO PERAL [3], SASO DZEROSKI [5], JOSÉ L. AYALA [1,4],(Member, IEEE)

[1]Dpt. of Computer Architecture and Automation, Complutense University of Madrid, 28040 Madrid, Spain (e-mail: (F.P.) fparrale@ucm.es, (A.B.) abarriog@ucm.es, (J.A.) jayala@ucm.es)
[2]Neurology Department, "La Princesa" University Hospital, Calle Diego de Leon, 62, 28006 Madrid, Spain (e-mail: (M.G.) mariamercedes.gallegosacristana@salud.madrid.org, (A.G.) anabeatriz.gago@salud.madrid.org
[3]Headache Unit, Department of Neurology, Hospital Clínico Universitario de Valladolid, Av. Ramón y Cajal, 3, 47003 Valladolid,Spain (e-mail: (M.R.) ruiz.marina@gmail.com, (A.G.) gueneurol@gmail.com)
[4]CCS-Center for Computational Simulation, Campus de Montegancedo UPM, 28660 Boadilla del Monte, Spain
[5]Department of Knowledge Technologies, Jozef Stefan Institute, Ljubljana,Slovenia (e-mail: saso.dzeroski@ijs.si)
[6]Carrera de Ing. en Sistemas Computacionales, Facultad Ciencias Matemáticas y Física, Universidad de Guayaquil, Guayaquil,Ecuador (e-mail: franklin.parralesb@ug.edu.ec)

Corresponding author: Franklin Parrales Bravo (e-mail: fparrale@ucm.es).

"This paper has been supported by the Spanish MINECO and CM under grants S2018/TCS-4423 and TIN 2015-65277-R. The project was co-financed by the Ministry of Education, Science, Technology and Innovation (SENESCYT) of the Government of the Republic of Ecuador (8905-AR5G-2016) and the HIPEAC collaboration grant (H2020-ICT-2015-687689)."

ABSTRACT

Deciding on continuous treatment of chronic diseases is vital in terms of economy, quality of life and time. We present a holistic data mining approach that addresses the prediction of the therapeutic response in a panoramic and feedback way, while unveiling relevant medical factors. Panoramic prediction makes it possible to decide whether the treatment will be beneficial without using previous knowledge and without involving unnecessary treatments. Feedback prediction can be more accurate prediction since it considers the results of previous stages of the treatment. A novel label encoding called *Simulated Annealing and Rounding* (SAR) encoding is also proposed to help improve the accuracy of prediction in both approaches. To unveil the medical factors that make treatment effective for patients, various techniques are applied to the prediction models found through the proposed approaches. Finally, this methodology is applied in the realistic scenario of analyzing electronic medical records of migraineurs under BoNT-A treatment.

The results show a significant improvement in accuracy due to the use of SAR encoding, from close to 60% (baseline) to 75% with panoramic prediction, and up to around 90% when using feedback prediction. Furthermore, the following factors have been found to be relevant when predicting the migraine treatment responses: migraine time evolution, unilateral pain, analgesic abuse, headache days and the retroocular component. According to doctors, these factors are also medically relevant and in alignment with the medical literature.

INDEX TERMS multi-target prediction, classification algorithms, data mining, simulated annealing

I. INTRODUCTION

One of the biggest problems associated with chronic diseases (CDs) is the continuous treatment required to mitigate or eliminate their symptoms. This must be considered when deciding whether a continuous treatment may be beneficial for a specific patient. In addition, people with chronic conditions typically have more health needs at any age, so the associated costs become disproportionately high [1]. For example, patients suffering from Parkinson's disease usually discontinue the treatment due to its ineffectiveness when

mitigating the pain [2], which thus involves wasting money. In order to avoid this, cost-benefit analyses have been applied, such as those for patients with hepatitis C [3], [4]. The conclusions drawn from these studies are diverse. For some cases, the doctors conclude that it is better to employ the treatment in short periods than in early phases of the CD [4]. However, earlier treatment has also been associated with a faster recovery [5]. Therefore, it is important to establish a prediction model of the response to customize the treatment for each patient.

Predictive models of response to any CD treatment can leverage the use of electronic medical records (EMRs). The rapid growth of EMRs requires that the traditional analysis of data collected manually by medical experts be combined with computational methods to help in the decision making process of a specific treatment. In this respect, the use of data mining techniques has made it possible to tackle the analysis of medical data and the construction of prediction models [6], [7]. Furthermore, some machine learning techniques have proved to be better suited for the analysis of medical databases because of the derivation of symbolic rules, the use of background knowledge, pattern-recognition and interpretation of time-ordered data. Hence, it is essential to study these techniques and select the most appropriate ones to provide an accurate prediction. Nevertheless, dealing with medical data is not an easy task, as some problems such as heterogeneity [8] or simply the lack of values (missing values) [9] typically arise in EMRs.

As regards the techniques mentioned above, there are different methodologies that consider EMRs for the prediction of the therapeutic response to certain CD treatments or responses to continuous oncological treatments. These methodologies predict the therapeutic response after several stages of the treatment, but they do not consider the prediction of responses to several stages together. One example is the work presented by Kurosaki et al. [10], who propose the use of decision trees to model the prediction of the final outcome of the treatment of chronic hepatitis C after 48 weeks of PEG-IFN/RBV therapy treatment. Another methodology is presented by Lambin et al. [11]. This work considers the prediction of the prognosis and the response to an oncological treatment based on radiation through the use of multifactorial decision support systems. Both methodologies discretize and normalize the data to avoid sensitivity to different orders of data scales. They also deal with the missing values, replacing them with calculated estimates.

In addition, there are some methodologies designed to reveal medical factors that influence the effectiveness of treatment. For example, Armañanzas et al. [12] use the Feature Subset Selection (FSS) technique to reveal the most important factors in predicting the severity of symptoms in a patient with Parkinson's disease [13] using non-motor symptoms. In [14], the authors propose the extraction of the most important attributes within a continuous CD treatment using a consensus tree. Hence, on the basis of the aforementioned works, it is desirable to incorporate the features of all these techniques into our methodology in order to be able to predict the response to several treatment stages as well as to reveal the reasons that make them effective.

To the best of our knowledge, there is no methodology in literature that predicts the responses to the different single stages of a continuous treatment as well as unveiling the main factors that contribute to such responses. Our main contributions can thus be summarized as follows:

- We consider a fine-grained solution through the panoramic prediction approach when no session has yet been made.
- Once the treatment has begun and the results of some stages are known, feedback prediction is proposed, which is more accurate. This technique leverages the treatment responses in previous stages of the treatment to predict the next stage.
- We bridge the gap between the biomedical community and the data mining community thanks to the extraction of medical factors that make the treatment effective or not.
- Our methodology considers missing values and personalizes the model to the features of the patient's EMR by means of a hierarchy of models.
- We apply our integrated approach to a real scenario with migraine patients treated with OnabotulinumtoxinA. The space of parameters that leads to an optimal solution has been efficiently explored by means of our novel *Simulated Annealing and Rounding* (SAR) encoding approach, which surpasses the results expected by the use of the unmodified algorithms. In this way, the classification algorithms used in the experiments have benefited significantly from the use of SAR, achieving mean accuracies higher than 75% and 90% through the application of panoramic and feedback prediction, respectively.
- The medical features that influence the effectiveness of the treatment are extracted and discussed by doctors. The results also validate the effectiveness of the techniques employed.

The remainder of the paper is organized as follows. Section II describes the work related with the techniques applied to CDs for predicting the treatment response. In section III, our methodology for predicting treatment results is explained in detail. Section IV describes the experiments and unveils the medical factors that make the treatment effective in the case of migraine. Finally, our conclusions and future lines of work are given in Section V.

II. RELATED WORK

Nowadays, modern hospitals possess a wide variety of monitoring and data collection devices that provide relatively inexpensive means of collecting and storing data in inter and intrahospital information systems. Hence, we can benefit from the availability of these EMRs to build our prediction models for a given treatment. The use of EMRs makes it possible to efficiently manage the medical diagnosis and

health in patients, thus improving the patient care [15]. One of the benefits is an improvement in the availability of graphics, the organization of data and the readability of patients' clinical information [16]. Diabetes or chronic obstructive pulmonary diseases are instances of CDs that have leveraged the use of EMRs to manage patient care [17], [18]. Nevertheless, although it has been shown that the data extracted from EMRs can be used to create predictive models, it is important to improve the extraction of these data and to include more variables so that these models can be more clinically useful and serve to better understand the trends of a given disease [19]. In this respect, there are several works that extract medical characteristics that influence a certain disease or treatment. For this purpose, several computational techniques can be found in the literature, and they employ either consensus models [14], or feature subset selection [12]. In each of these works, the doctors analyze the factors found by the algorithm to provide a clinical meaning. It is therefore important to consider the extraction of these factors in our methodology.

One of the most common problems involved in the collection of medical data is the presence of missing values [9]. In fact, it is a common situation when extracting data and inferring the models in a medical scenario. In order to mitigate these problems, Lambin et al. [11] propose replacing missing values with calculated estimates. However, the absence of data can have a value in itself. In fact, Pagan et al. create a set of models to tackle the lack of information due to the malfunction of any medical sensor [20]. This "lack of clinical information" needs to be considered in our methodology because it can provide useful information to build a set of prediction models in order to adapt the prediction to the missing values appearing in the EMRs.

Another problem to consider is the heterogeneity of medical data. The data can come in the form of images (X-rays, magnetic resonance scans, etc), interviews with the patients, laboratory data, as well as the doctor's observations and interpretations [16]. The homogeneity of the information can be addressed by simplifying and categorizing the data. For instance, this can be carried out through the transformation of heterogeneous clinical data to labels [8]. For this purpose, the labels must be defined and agreed by the experts in the disease to be analyzed to achieve an adequate representation of the medical information [21]. Hence, as pointed out in the aforementioned works, it is desirable that our methodology deals with heterogeneous data, leveraging the labelling provided by medical experts.

Nowadays different data mining techniques are available for processing the EMRs and performing the generation of predictive models for a given CD treatment. There are methods that allow us to obtain models to carry out the prediction of a single target variable or the simultaneous prediction of multiple target variables [22]. In our prior work [14], the use of several classification algorithms for a single target variable prediction has been explored, considering one model for each stage of a CD treatment such as that for migraine. However, given the multiple stages associated with CD treatments, we can make use of the simultaneous prediction of treatment responses for all the stages. With this purpose, we can make use of existing multi-target algorithms in the literature, such as predictive clustering trees, binary relevance classifiers and the hierarchy of multilabel classifiers [23].

Another computational technique called *fuzzy set theory* [24] has been used in different medical applications that involve uncertainty and ambiguity of criteria [25], such as developing a knowledge-based system for breast cancer classification [26], a fuzzy decision support system for the diagnosis of heart disease [27], among others. This technique is useful for our scenario, where it is necessary to address ambiguity when considering the membership of each record to a specific prediction model out of a set of models.

The methodology presented in this article brings together all the important aspects highlighted in this section. The proposal offers a holistic approach to predict the therapeutic response to several stages of treatment. On the one hand, it considers panoramic prediction based on predictive clustering to give a general idea of how the patient will evolve after receiving a certain medication. On the other hand, feedback prediction based on fuzzy logic and a hierarchy of models is also presented to predict the response to a treatment session, given the previous responses to the prior sessions. Furthermore, our methodology allows dealing with missing values and heterogeneous data after a preprocessing stage. In addition, we integrate several methods to extract the medical factors that determine the effectiveness of the treatment.

Finally, the proposed methodology is put into practice in the study of migraine, as the asssociated treatment requires several sessions to improve the quality of life of patients.

III. METHODOLOGY

Our methodology utilizes medical data for the prediction of the therapeutic response to a given treatment. Figure 1 presents the framework on which this paper is based. In this way it is possible to personalize the medical treatment depending on the prediction provided on the response to the treatment.

As an initial step, the severity indexes are established to measure the effectiveness of the treatment. In CDs there is an extensive literature on established and validated severity indexes. Many chronic diseases typically have more than one index available and validated to evaluate the patient's degree of impairment or unavailability.

Afterwards, the preprocessing step deals with the heterogeneity of medical data, applying a categorization method based on the mean and the standard deviation.

Next, two ways of predictive modeling have been considered. On the one hand, an approach based on multiobjective prediction is applied. This technique leverages *panoramic prediction*. This type of prediction provides the medical specialists with an overview of the responses to multiple sessions of the treatment without waiting for the response to a session to predict the next one. This method performs

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2927429, IEEE Access

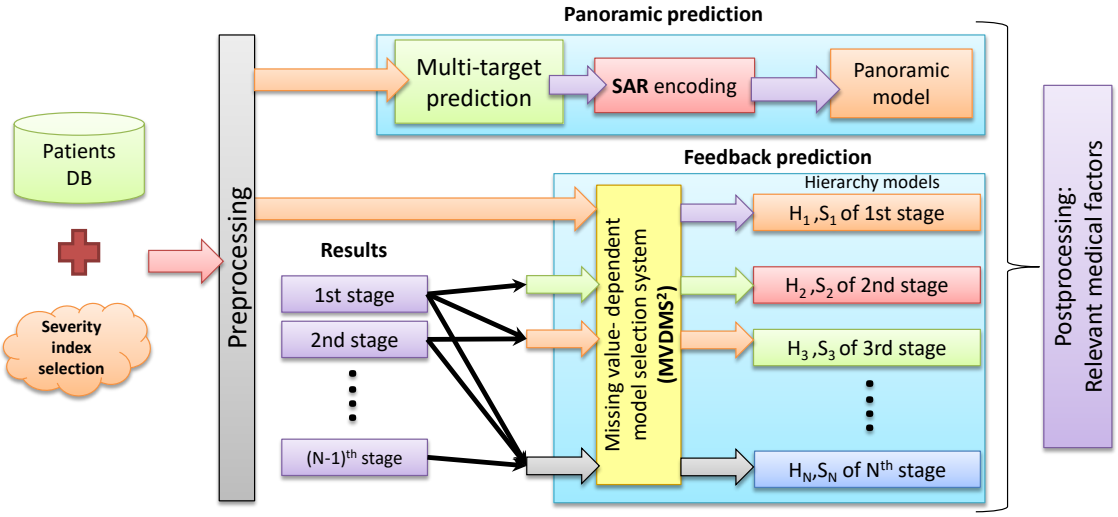

FIGURE 1: Methodology workflow

an accuracy optimization through the use of our proposed *Simulated Annealing and Rounding* (SAR) encoding over the data previously categorized by doctors.

On the other hand, *feedback prediction* is considered as a second predictive modeling approach. In this type of prediction, all the medical factors previously categorized by doctors are considered for the first session of the treatment. This information then becomes the input for the *Missing Value-Dependent Model Selection System* (MVDMS^2). This model selection system is responsible for grouping the medical records according to the number of incomplete or missing data, something that commonly arises in the medical field. This method generates a hierarchy H_i of prediction models and a fuzzy selector S_i based on the number of missing values for each stage of the treatment, considering the results of the previous stages of the treatment. For example, it is necessary to consider the results of the 1st, 2nd, ..., $(N-1)^{th}$ stages in order to predict the subsequent N^{th} response. In this way, a more accurate prediction is achieved.

Both approaches generate different predictive models for several treatment sessions. These models are then studied in order to reveal the medical factors that make the treatment effective or not. The details associated with every block shown in Figure 1 will be described in this section.

A. SEVERITY INDEX SELECTION

In order to estimate the goodness of the solutions, it is necessary to define a metric, namely the *severity index*, that indicates how efficient the treatment session has been. For this purpose, there are many works in the medical literature on various CDs in which different severity indexes are presented. For example, in the case of Parkinson's Disease (PD), the HY scale [28] is a long-established instrument used to categorize patients according to PD stages. Another metric is CISI-PD [29], which extends the evaluated motor symptoms criteria to more complex aspects such as the patients' cognitive state. For migraine, the severity is typically evaluated using the HIT6 value [30], the intensity, the duration, the frequency of attacks [31] or headache days [32]. Hence, the severity index is something inherent to the CD under consideration, but also depends on the number of EMRs containing the index.

B. PRE-PROCESSING

In this first stage of the proposed methodology, the data retrieved from the EMRs should be previously coded by medical experts in the disease to reduce their heterogeneity. However, the heterogeneity of data may still persist in medical factors previously categorized by doctors. In order to tackle this situation, a categorization based on the mean and the standard deviation of the values is applied. Given a minimum (V_{min}) and a maximum (V_{max}) values, this type of categorization centers the intervals around the mean (μ), and defines subsequent intervals by adding/subtracting the standard deviation (σ). The intervals $[V_{min}, \mu-\sigma]$, $(\mu-\sigma, \mu+\sigma]$ and $(\mu+\sigma, V_{max}]$ are defined for our categorization.

C. PANORAMIC PREDICTION

The suitability of this method can be appreciated from studying the advisability of a certain drug for every stage involved in a CD treatment [5]. Predicting the response to every session of the treatment corresponds to what personalized medicine is seeking: to allow a cost-benefit analysis by the doctors and thus a decision as to whether the cost and other details inherent to the medication can be taken on by the patient [33]. Nevertheless, this first prediction approach aims to analyze a medical scenario involving the therapeutic response to a given treatment without any prior knowledge or *feedback*. This implies that the results of the first or subsequent treatment applications are not yet known. Thus,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2927429, IEEE Access

the responses cannot be used at different treatment stages to perfect the prediction model.

In order to achieve this goal two techniques are employed. First, due to the need to know in advance the treatment responses to all the stages, a multi-target prediction approach is implemented. Second, the SAR encoding, which is a coding technique based on the Simulated Annealing (SA) algorithm and rounding, seeks to better represent the information coded by the doctors and then improve the prediction of the therapeutic responses. The details of both steps are explained in the following subsections.

1) Multi-target prediction

In contrast to the traditional prediction based on a single target variable, this approach [22] allows us to simultaneously predict a vector y of z responses for a vector x with p features, using a function $h(x)$:

$$h(x) : x = (x_1, x_2, \ldots, x_p) \xrightarrow{h(x)} y = (y_1, y_2, \ldots, y_z) \tag{1}$$

where $h(x)$ is the model to learn with the use of multi-target classifier algorithms.

In our scenario, it is possible to take advantage of the features that this technique provides. Our objective is to present the prediction for treatment response in subsequent stages, which can be translated into a multi-target problem. This implies an important advantage because it is not necessary to obtain clinical data after a session of treatment to guess the response after the following session. The input for this technique will include the data numerically labelled following the label encoding explained in Section III-B.

Methods such as predictive clustering trees (PCT), binary relevance (BR) and the hierarchy of multilabel classifiers (HOMER) have been recommended to carry out the learning of multi-target prediction models [23].

In *predictive clustering trees* (PCT), decision trees partition the set of examples into subsets in which the examples have similar values of the target variables, while clustering produces subsets in which the examples have similar values of the descriptive variables [34].

In the *binary relevance* [35] (BR) method, the transformation of the multi-target prediction into p binary classification problems is considered. A prediction model is learnt for every target variable $(y_1, \ldots, y_z)$ independently. After that, all the results are combined to determine the predicted class set.

In the *hierarchy of multi-label classifiers* [36] (HOMER), a hierarchy of multiple labels is built and a classifier is obtained for the label sets of each node of the hierarchy.

These techniques are implemented in two java frameworks: The *MEKA* [37] for multi-label classification, and *CLUS* [38] for predictive clustering. The MEKA framework implements the BR and HOMER methods, while PCTs are implemented in CLUS.

2) SAR encoding

At this point of the methodology, all the values have been categorized in the preprocessing stage explained in Section III-B. Although everything is categorized, there are data with non-numeric labels given by doctors. With the purpose of homogenizing the medical factor labels and working with numerical optimization algorithms as well as allowing a better representation of the labels with respect to the models, the nominal labels established by the doctors should be converted to numeric labels. This can be done by using consecutive natural numbers different from 0 for each label. Although this basic encoding method has the advantage of being simple, the numerical values can be misunderstood when applying and obtaining prediction models with the data mining algorithms. For example, a variable that identifies the sex of the patient will take values of M for men and W for women. When these nominal values are converted to numerics they can be transformed into 1 and 2, respectively. However, M, which takes the value of 1, is not greater than or lower than W, which has taken the value of 2.

Another encoding approach is called one-hot encoding [39]. The strategy of this method is to convert the different column labels of the original dataset into columns of a new dataset, defining a column for each different value. Then, the cell values of the new dataset will be filled with 1's or 0's (true/false) in the corresponding column according to the label value of the original dataset. This has the benefit of not weighting a value improperly. However, its drawback consists of adding many new columns to the data set. Hence, this approach is not good for processing medical data, since what we pursue is to reduce the number of medical factors to be considered by the data mining algorithms.

The SAR method is applied to find a better representation of the numeric labels. This proposed technique improves the numeric labels as it obtains high prediction accuracies without adding more columns to the dataset. We call it SAR because this technique uses the SA algorithm [40] and a rounding operation to perform small numeric label perturbations for each column of the medical dataset. The inputs of this method are the dataset to be analyzed, the number of decimal digits to consider (D), and the classification algorithm to be used to generate the predictive model. The outputs are the set of optimal column weights (W_{opt}), the optimal number of fractional digits for rounding (d_{opt}), and the optimized classification model (M_{opt}). The symbology used is presented in Table 1. The different steps in SAR are shown in Figure 2 and explained in the following lines:

- The input of the algorithm is an initial dataset O containing m clinical records, each described over the same set of n medical factors (columns) $c_0, c_1, c_2, \ldots, c_{n-1}$. The conversion of nominal labels to numbers is performed following a consecutive order of integers begining with 1. This is done for the n columns of the dataset. The modified dataset will be called O', with $c'_0, c'_1, c'_2, \ldots, c'_{n-1}$ as modified columns.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2927429, IEEE Access

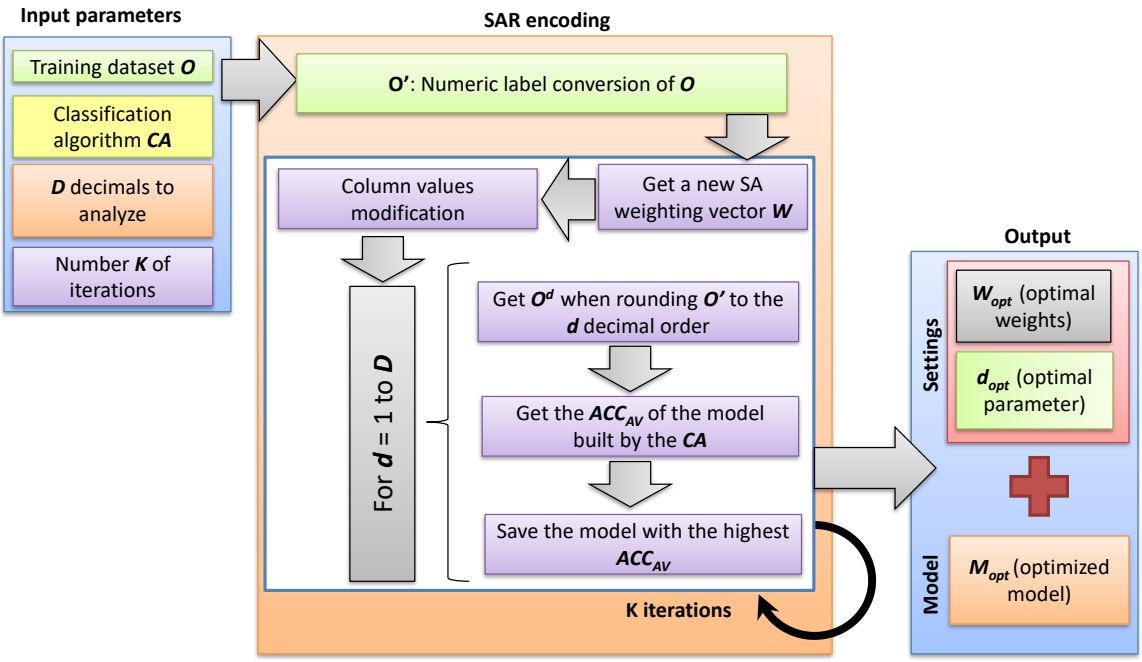

FIGURE 2: SAR encoding diagram

TABLE 1: Description of variables employed in the SAR encoding

Name	Description
O	Training dataset.
O'	Modified dataset.
m	Number of records of the initial dataset.
r	Number of records of the training dataset.
n	Number of columns of the dataset.
c_i	Medical factor (column) of a dataset.
T	Threshold, number of columns that will be taken into account for grouping the dataset records.
W	Set of weights of the columns of a dataset.
w_i	Weight of the i^{th} column of a dataset.
$o_{i,j}$	Cell value of the i^{th} record and j^{th} column of a dataset O.
D	Number of decimal positions to analyze.
d	Decimal position.
O^d	O' modified dataset and rounded to the d decimal position.
c_j^d	j^{th} column of the O^d dataset.
$o_{i,j}^d$	Cell value of the i^{th} record and j^{th} column of a dataset O^d.
ACC_i^d	Accuracy of the O^d dataset in the i^{th} stage.
ACC_{AV}^d	Average accuracy of the O^d dataset.

- Once the O' dataset is generated, the run to algorithm [40] is run to perform the feature weighting task. The SA algorithm will attempt to assign different weights w_j, $0 \leq j < n$, i.e. one for each column $c'_j \in O'$. The weights w_j will reflect the degree of relevance of a column c'_j for a problem to solve, where $w_j \in IR\{0, 1\}$. The values of every cell $o'_{i,j} \in O'$ are multiplied by the corresponding weight w_j through the $o'_{i,j} \times w_j$ operation, $\forall i, 0 \leq i < m$ and $\forall j, 0 \leq j < n$. This multiplication is illustrated in Figure 3.
- The O' dataset is rounded to the nearest tenth, hundredth, thousandth, and other decimals in order to generate small perturbations among the different numeric labels in each column [41], [42]. These rounded labels will generate modifications in the prediction models learnt by the classification algorithms that work with real numbers. The number of decimals to be considered when rounding is defined by the parameter D, generating different datasets O^d with columns c_j^d and cells $o_{i,j}^d$, where $1 \leq d \leq D$, $0 \leq i < m$ and $0 \leq j < n$. For example, if D is equal to 3, three modified datasets O^1, O^2 and O^3 will be generated to train the SAR optimization, rounding to the nearest tenth, hundredth and thousandth for the first, second and third modified

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2927429, IEEE Access

W (weights):

w_1	w_2	w_3
0.795269560373731	0.18469775	0.767716221

O' (dataset):

c'_1	c'_2	c'_3	Class A_1	Class A_2
1	2	3	low	high
2	3	1	low	low
3	1	2	high	high

feature weighting $c'_1 \times w_1$, $c'_2 \times w_2$, $c'_3 \times w_3$

O' :

c'_1	c'_2	c'_3	Class A_1	Class A_2
0.79526956	0.369395501	2.303148663	low	high
1.590539121	0.554093251	0.767716221	low	low
2.385808681	0.18469775	1.535432442	high	high

O^2 dataset

c^2_1	c^2_2	c^2_3	Class A_1	Class A_2
0.80	0.37	2.30	low	high
1.59	0.55	0.77	low	low
2.39	0.18	1.54	high	high

FIGURE 3: Weighting dataset and rounding to hundredths

datasets, respectively. An example of this step is shown in Figure 3 when rounding O' to the hundredth (O^2).

- The prediction models are learnt by the classification algorithm when training it with each of the modified datasets O^d. The accuracy for each of the stages, ACC_i^d, $0 \leq i < s$, of any modified dataset O^d by d, $1 \leq d \leq D$, is obtained through Equation 2. True positives (TP^d) and true negatives (TN^d) refer to the correct prediction of positive and negative responses to treatment of the i class attribute, respectively. False positives (FP^d) and false negatives (FN^d) refer to the wrong prediction of positive and negative responses to treatment of the i class attribute, respectively.

$$ACC_i^d = \frac{TP_i^d + TN_i^d}{TP_i + FP_i^d + TN_i^d + FN_i^d} . \quad (2)$$

- The average accuracy of all s stages, ACC_{AV}^d (Equation 3), associated with the modified dataset O^d will be the value to be improved. In this respect, we define $(1 - ACC_{AV}^d)$ as the fitness value to be diminished by the SA algorithm. A number K of iterations is defined as an input parameter in order to limit the execution of the SA algorithm.

$$ACC_{AV}^d = \frac{1}{s}\left(\sum_{i=1}^{s} ACC_i^d\right) \quad (3)$$

- The outputs of the SAR encoding are the transformation settings to apply in the initial dataset O and the optimized model M_{opt}. These settings are composed of the set of weights W_{opt} to apply to the columns in the initial dataset and the number of fractional digits d_{opt} that achieved the best accuracy. These outputs will be used to transform the data and to apply the model M_{opt} in order to obtain the panoramic prediction of the treatment response.

D. FEEDBACK PREDICTION

This methodology seeks to improve the prediction of the therapeutic response by creating a predictive model for each treatment stage. Figure 4 shows the diagram of the missing value-dependent model selection system (MVDMS^2). The purpose of the system is to generate a hierarchy of predictive models by taking into account the missing values in the medical records. To do this, the preproccesed dataset of m records is split into two groups of records. A total of 75% of these records will be used as the O dataset (with r records) to train the system. The rest of the data ($m - r$ records) will be used for testing purposes. Then, the records of the training dataset are clustered first by their NA values. Afterwards, for each of the resulting groups, the initial predictive models are obtained using the classifier algorithm specified as input parameter. These models are optimized through the application of the SAR encoding. To take into account those patient records that do not meet the membership rule of the groups, a fuzzy selector is trained. This selector establishes the membership rules for each record according to its number of NAs. The final product of this system is the hierarchy of models whose membership rules are governed by the fuzzy selector. In order to obtain the accuracy of the system, the test dataset is used to apply the fuzzy selector and the hierarchy models. For the second and following sessions of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2927429, IEEE Access

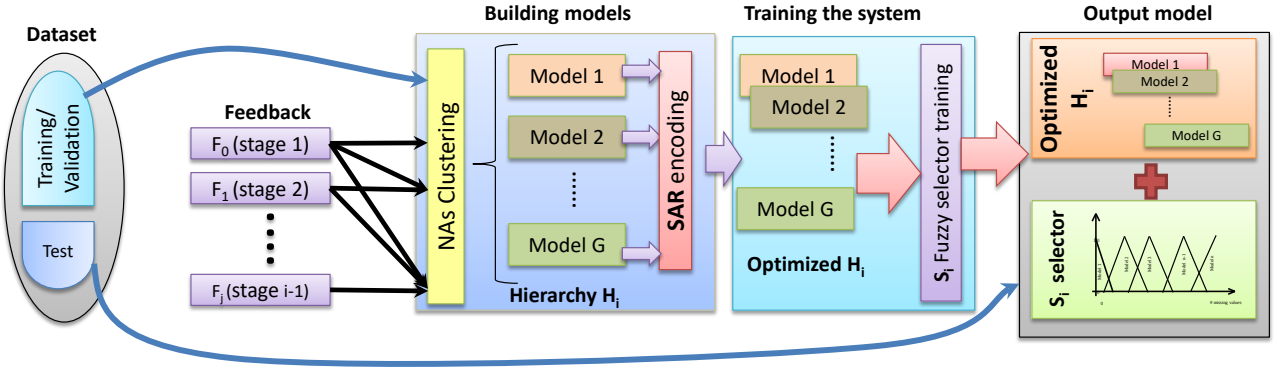

FIGURE 4: Missing value-dependent model selection system (MVDMS^2) for the i^{th} stage

Initial dataset

Register	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	.....	Feature N
1	2	2	3	1	NA		NA
2	3	NA	1	2	NA		NA
3	1	1	2	3	1		NA
4	1	1	2	2	NA		1
5	2	NA	1	NA	1		1
Total NA	0	2	0	1	3		2

Table A

Columns	# NA
Feature 5	3
Feature 2	2
Feature N	2
Feature 4	1
.....	.....

Table B

Register	Feature 5	Feature 2	Feature N	Feature 4
1	0	1	0	1
2	0	0	0	1
3	1	1	0	1
4	0	1	1	1
5	1	0	1	0

.....

FIGURE 5: Data structure to analyze missing values found in medical records

the treatment, the previous responses are used as feedback to predict the response to the next treatment session. Every process shown in Figure 4 will be described in detail in the following subsections.

1) Clustering of missing values

The preprocessed data is the system input. Their missing values are represented by the label NA in the EMRs. Figure 5 shows an example of the medical dataset provided with many EMRs. O is the training/validation dataset composed of r records and n medical factors (columns), where r is 75% of the m records from the initial dataset. The NA values of every cell $o_{i,j} \in O$, $0 \leq i < r$ and $0 \leq j < n$, are accumulated in the "Total NA" row (Figure 5) for each column of the dataset. Afterwards, the n columns are sorted in descending order according to the number of NAs in Table A. Then, only the first T features will be considered to create Table B. The value of this threshold T should be defined by doctors depending on the number of features they wish to take into account. The cells $b_{i,j}$ of Table B, $0 \leq i < r$ and $0 \leq j < T$, will be filled with 0's and 1's according to Equation 4.

$$b_{i,j} = \begin{cases} 1, & \text{if } o_{i,j} \neq \text{NA} \\ 0, & otherwise \end{cases} \tag{4}$$

Table B provides the information about the NA values of every record. The registers contained in Table B are then grouped using the *k*-medians clustering [43] algorithm. The *k*-nearest neighbor algorithm (*k*-NN) has not been used for this task since *k*-medians is less sensitive to outliers. It should be noted that given a number of record groups G, the parameter k will be set to this value. Afterwards, it is important to generate rules for defining the membership of the medical records belonging to each group.

2) Initial set of models and SAR encoding

Once the groups of medical records have been found, each group is treated as a different dataset. Each of them is trained with a classification algorithm. In order to improve the accuracy, each model is optimized by using the SAR encoding proposed in Section III-C2. The output of this training phase will be a hierarchy of optimized models.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2927429, IEEE Access

3) Fuzzy model selector

In Section III-D1, different groups of medical records and their respective membership rules were established. However, some medical records do not belong to any given group because they do not satisfy the group membership rules. To solve this issue, we can benefit from the use of fuzzy logic to soften the membership rules, selecting the most suitable model for each medical record.

For this purpose, we have defined a mapping table (T_{map}) composed of r records that considers their number of NA cells and also the cluster assigned in Section III-D1. Then we proceed to use this table for training-testing the fuzzy classifier and obtaining the fuzzy model selector (fuzzy rules that select the model for each record). Afterwards, these rules will be applied to the $m - r$ remaining records (testing set), taking into account their number of NAs. After that, missing values of every record are replaced considering the other records of its cluster.

The Fuzzy Unordered Rule Induction Algorithm (FURIA) [44] has been selected as the algorithm to train our model selector because it combines the advantages of the RIPPER algorithm [45] with fuzzy logic, generating simple and compact sets of fuzzy classification rules.

4) Prediction models for other stages

In order to generate the i^{th} hierarchy of prediction models (H_i), the proposed flow requires the original dataset O as well as the feedbacks F_j, $0 \leq j < i$, corresponding to the $(i-1)$ previous stages of the treatment. These will be the inputs for the MVDMS^2 system that will generate as output the hierarchy of classifiers based on fuzzy membership rules for the specific session to predict.

E. UNVEILING RELEVANT MEDICAL FACTORS

Once the models of panoramic prediction and feedback prediction have been found, it is important to clarify which are the medical factors that allow the treatment to be effective in each group of patients. Since some predictive models generated by the weighted feature vectors can achieve the same accuracy, it is necessary to study the medical factors that appear the most in those models. For this purpose, we consider a consensus model to reveal these factors. The idea is not to build a consensus predictor model, but to understand the most relevant attributes that exist in the majority of the induced prediction models of the best classifier.

The consensus model technique presented in [14] has been considered for the aforementioned purpose. Using this technique, many decision trees will be induced using a resampling method (k-fold cross validation with k=10). Q models will be induced using the weighted feature vectors as well as the parameter d that achieves the best accuracies for each model. The number of levels to consider in the consensus model (L_{max}) is set to 3 in order to achieve a good degree of comprehension. The most important medical factors in the feedback prediction are those that appear in each stage at least once in one of their consensus models. In the case of panoramic prediction, the most important medical factors are simply those that appear in the consensus model, since it is constructed from multi-target predictive models. Afterwards, the relevant features will be discussed by doctors.

IV. ANALYSIS CASE

We will describe how our study can be applied in a real case involving a chronic disease, namely the prediction of the treatment response when treating migraines with OnabotulinumtoxinA. This case study has been selected due to the complexity of the problem in terms of modeling and the selection of variables, as well as its socio-economic impact.

Chronic migraine is defined as a headache occurring on 15 or more days per month for more than 3 months, and which has the features of a migraine headache on at least 8 days per month [46]. Globally, approximately 2% of the population experiences chronic migraine [47]. In addition to the increased use of analgesic medication, visits to doctors, and visits to the emergency services, chronic migraine has a high socioeconomic cost, with higher direct and indirect costs. Furthermore, chronic migraine sufferers are more prone to anxiety, depression, other chronic diseases (respiratory, heart or circulatory) and more chronic pain, all of this associated with significant personal, societal, and economic burdens [48].

OnabotulinumtoxinA (BoNT-A) has been a widely used treatment for chronic migraine since its approval in 2010 by the Food and Drug Administration in the United States (FDA), having also shown a more sustained effect and better tolerability than topiramate in the few comparative studies performed [49], [50]. In clinical practice, about 20-30% of chronic migraineurs do not respond to BoNT-A [51]. One of the most debated aspects in recent years has been the possible relationship between the clinical phenotype of migraine attacks and the response to BoNT-A. As has been mentioned in [52], it would be very useful to know beforehand which patients will respond and which will not. Knowing the phenotype-response relationship may help in the development of new treatments for the 20-30% of patients that do not respond to the treatment. Besides the cost, it would prevent the patients from suffering the pain associated with the treatment.

A. PATIENTS

The sample consists of patients with chronic migraine and under previous or current treatment with BoNT-A, with follow-up at the headache unit of two tertiary-level hospitals. The dataset was collected retrospectively from the review of their EMRs. A total of 173 patients were included (116 from *Hospital Clínico Universitario* in Valladolid and 57 from *Hospital Universitario de La Princesa*, in Madrid). Sixty-two baseline variables were categorized. These variables were related to the following points: clinical pain features, demographic features of patients, comorbidities, tested and concomitant preventive drugs, pain impact measures, and all available analytical parameters. The latter were obtained

from blood tests recorded in the medical history in the 3 months prior to, or 3 months after, the first infiltration, and included hemogram and liver, renal, thyroid, serum iron, vitamin B12, folic acid and vitamin D profiles. The efficacy of BoNT-A was evaluated by comparing the baseline situation (before the first infiltration) and the situation after 12-16 weeks following each of the infiltrations, through the following parameters: number of days of pain per month, percentage reduction in days with pain, subjective intensity of pain, number of days of disability due to pain per month, drug consumption for pain and adverse effects of infiltration. Since this was a retrospective study, not all the data could be obtained for each patient in a systematic way.

Some patients are *non-respondent*, while others respond after the i^{th} session. In addition, some problems are encountered while examining these data. For example, a small set of patients with many attributes is typically present in our medical databases. Furthermore, the incompleteness of data is another problem that must be dealt with. Some attributes are given as continuous numeric values while other attributes are categorical. All in all, medical records need to be preproccesed before inferring models for predicting the outcome of the treatment.

B. SEVERITY INDEX

The reduction *(R)* and the adverse *(A)* effects, which are more frequently found in the databases, have been selected to define the severity index to predict (class attribute). *R* and *A* are measurable values from an objective point of view based on definitions. *R* is a clinical objective value categorized from 1 to 4 according to the percentage of reduction in days of migraine, with a value of 1 when the percentage reduction in days of migraine is less than or equal to 25%, 2 for the interval between 25% and 49%, 3 for the interval between 50% and 74%, and 4 when the percentage is greater than or equal to 75%. *A* is equal to 1 when there are no adverse effects, 2 when there are mild adverse effects (easily tolerated), 3 when there are moderate adverse effects (interfere with usual activities and may require suspension of treatment) and 4 when there are serious adverse effects (incapacitate or disable usual activities, and require suspension of treatment as well as medical intervention).

In order to obtain a directly proportional variable, the index (N_{AC}) has been determined by dividing R and A. A similar approach to the one based on [51] (two response categories: low and high) has been considered for severity index categorization. Lower responses are labeled when the N_{AC} value falls into the *(V_{min}, cut-off point)* interval, while high response labels are used for those values falling within the *(cut-off point, V_{max})* interval. In this case, $V_{min} = 0.25$ occurs when $R = 1$ and $A = 4$, while $V_{max} = 4$ occurs when $R = 4$ and $A = 1$. We select a cut-off point of 1.40. The reason to use this value is the fact of trying to emulate the criterion of a 30% decrease, which was used in [51] and considered as an effective response to the treatment. In this way, values lower than 1.40 represent 30% of the values that N_{AC} can take. Thus, the low and high categories are defined with the intervals (0.25, 1.40) and (1.40, 4), respectively.

C. SETTINGS FOR EXPERIMENTS

The number of SAR iterations (K parameter) was defined as two million, considering this as a sufficient number of iterations for the algorithm to converge. When training-testing the predictions over class attributes, the *k*-fold cross validation (*k*=10) was applied by [53]. The results presented in this section are based on the measured accuracy of the *k*-fold cross validation. This method has been used to avoid reporting overoptimistic results of classifier algorithms because of overfitting. Moreover, sensitivity and specificity values of induced models need to be considered because those measures are more frequently used and more important than the classification accuracy in medical applications [54]. The sensitivity measures the fraction of positive cases that are classified as positive, while the specificity measures the fraction of negative cases classified as negative. In our case, the positive values will be the patients who have a good therapeutic response (labelled as “high") to the treatment, while the negative cases will be the ones that obtain a bad response (labelled as “low").

D. PANORAMIC PREDICTION

The parameter D was equal to 3 because three decimal magnitude orders have been considered. PCT, BR and HOMER have been employed, selecting *Random Tree* (RT) as the classifier algorithm due to the accurate results achieved in [14]. More details about the parameters of the classifiers used in the experiments are presented in Table 2. The baseline results (without any rounding) are shown in Table 3. Also, feature subset selection (FSS) has been used for evaluation, applying the CFS method [55] with the following parameters: 1 as the number of threads to use, 1 as the size of the thread pool (number of cores in the CPU), and best-first as the search method.

With the purpose of comparing the performance of the methods presented in Table 3, we use Nemenyi’s test procedure to conduct all pairwise comparisons in a multiple comparison analysis. Following García and Herrera [56], we have selected Nemenyi’s test since we do not know a priori what distribution the data follows. Moreover, the interpretation it gives us is easier to understand than the f-value of ANOVA, for instance. The adjusted p-values are compared against a significance level of α = 0.05 to reject or accept the null hypothesis that a pair of methods perform equally. By observing the p-values of the tests from Tables 4, 5 and 6, the conclusions are: (1) The use of SAR encoding with $d = 1$ produces a significant improvement in the accuracy values of the PCT, BR and HOMER methods without FSS (baseline), and with it for the first, second and third stage with the exception of BR with FSS. (2) The use of SAR encoding in PCT with $d = 1$ produces significant improvements in the baseline and FSS values of the BR and HOMER methods for the first, second and third stage. (3) The use of SAR encoding with

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2927429, IEEE Access

TABLE 2: Description of the classifiers parameters used in experiments

Classification algorithm	Parameters
Predictive clustering tree (PCT)	classifier=RandomTree, Heuristic = Gain ratio
Binary relevance (BR)	classifier=RandomTree
HOMER	type=Random, classifier=RandomTree, Multi-label learner=BR

TABLE 3: Estimated performance metrics (mean ± std deviation) of panoramic prediction with $D = 3$ using 10-fold cross validation. The best results are highlighted in bold.

Method	Setting	First stage			Second stage			Third stage		
		Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
PCT	baseline	61.63%±3.75%	0.65±0.02	0.25±0.02	63.95%±2.74%	0.75±0.05	0.57±0.03	62.79%±4.16%	0.65±0.02	0.61±0.02
	FSS	63.07%±2.41%	0.61±0.03	0.66±0.04	64.61%±2.13%	0.56±0.02	0.72±0.02	60.76%±2.81%	0.52±0.05	0.69±0.02
	$d = 1$	**73.84%±0.91%**	0.72±0.02	0.75±0.01	**75.38%±1.24%**	0.73±0.03	0.76±0.02	**74.61%±2.14%**	0.73±0.03	0.75±0.02
	$d = 2$	70.93%±1.27%	0.74±0.02	0.61±0.02	66.92%±0.95%	0.68±0.03	0.65±0.03	68.46%±1.66%	0.71±0.03	0.67±0.02
	$d = 3$	65.12%±1.05%	0.74±0.03	0.48±0.02	66.28%±1.11%	0.65±0.01	0.67±0.02	67.69%±2.42%	0.60±0.02	0.54±0.04
BR	baseline	62.79%±2.37%	0.75±0.05	0.56±0.04	54.65%±4.18%	0.60±0.03	0.54±0.03	67.44%±3.56%	0.75±0.04	0.63±0.03
	FSS	63.95%±2.17%	0.75±0.03	0.57±0.02	68.60%±3.42%	0.75±0.04	0.64±0.03	66.27%±3.75%	0.68±0.02	0.61±0.03
	$d = 1$	70.93%±1.01%	0.74±0.02	0.61±0.02	72.09%±1.32%	0.75±0.03	0.70±0.02	73.84%±2.41%	0.72±0.02	0.75±0.03
	$d = 2$	67.44%±0.95%	0.68±0.03	0.66±0.04	69.76%±2.07%	0.68±0.03	0.70±0.03	72.09%±2.71%	0.72±0.02	0.71±0.03
	$d = 3$	63.95%±1.26%	0.79±0.02	0.56±0.04	65.12%±1.62%	0.74±0.03	0.48±0.04	68.60%±2.23%	0.75±0.02	0.64±0.02
HOMER	baseline	55.81%±2.14%	0.55±0.03	0.56±0.04	56.97%±3.86%	0.55±0.04	0.57±0.03	61.62%±4.13%	0.62±0.02	0.61±0.03
	FSS	56.97%±1.81%	0.55±0.03	0.57±0.03	58.13%±2.15%	0.58±0.02	0.57±0.03	60.46%±3.84%	0.58±0.04	0.61±0.02
	$d = 1$	68.60%±1.04%	0.72±0.02	0.66±0.03	72.09%±2.36%	0.72±0.03	0.71±0.03	73.25%±3.24%	0.72±0.02	0.73±0.02
	$d = 2$	63.95%±0.72%	0.65±0.02	0.63±0.01	66.27%±1.96%	0.65±0.02	0.66±0.02	67.44%±2.26%	0.72±0.03	0.64±0.04
	$d = 3$	59.30%±0.87%	0.58±0.02	0.59±0.02	60.46%±1.71%	0.58±0.02	0.61±0.03	65.11%±2.34%	0.65±0.03	0.64±0.03

TABLE 4: Nemenyi-test p-values over the 10-fold cross validation accuracy values of methods used in Table 3 for the first stage

		PCT					BR					HOMER			
		Baseline	FSS	d=1	d=2	d=3	Baseline	FSS	d=1	d=2	d=3	Baseline	FSS	d=1	d=2
PCT	FSS	0.99999	-	-	-	-	-	-	-	-	-	-	-	-	-
	d=1	$2.20{\cdot}10^{-5}$	0.00077	-	-	-	-	-	-	-	-	-	-	-	-
	d=2	0.0008	0.0154	0.99999	-	-	-	-	-	-	-	-	-	-	-
	d=3	0.68523	0.98252	0.14702	0.60169	-	-	-	-	-	-	-	-	-	-
BR	Baseline	1	1	0.00033	0.00782	0.95201	-	-	-	-	-	-	-	-	-
	FSS	0.98661	1	0.01324	0.13567	0.99999	0.99993	-	-	-	-	-	-	-	-
	d=1	0.00094	0.01755	0.99999	1	0.62872	0.00898	0.149	-	-	-	-	-	-	-
	d=2	0.08485	0.44299	0.81558	0.99623	0.99941	0.31935	0.8961	0.99722	-	-	-	-	-	-
	d=3	0.98953	1	0.01158	0.12329	0.99998	0.99996	1	0.13567	0.88079	-	-	-	-	-
HOMER	Baseline	0.89397	0.41693	$2.70{\cdot}10^{-10}$	$3.80{\cdot}10^{-8}$	0.00601	0.55099	0.0836	$4.80{\cdot}10^{-8}$	$5.00{\cdot}10^{-5}$	0.0927	-	-	-	-
	FSS	0.96316	0.59003	$1.40{\cdot}10^{-9}$	$1.60{\cdot}10^{-7}$	0.01456	0.72134	0.157	$2.00{\cdot}10^{-7}$	0.00016	0.1718	1	-	-	-
	d=1	0.02662	0.21445	0.95786	0.99993	0.98661	0.13752	0.6852	0.99996	1	0.6592	$7.10{\cdot}10^{-6}$	$2.50{\cdot}10^{-5}$	-	-
	d=2	0.9922	1	0.00991	0.11024	0.99996	0.99998	1	0.1216	0.86171	1	0.10413	0.18995	0.62872	-
	d=3	0.99983	0.94559	$9.50{\cdot}10^{-8}$	$7.10{\cdot}10^{-6}$	0.115	0.9795	0.5549	$8.70{\cdot}10^{-6}$	0.00301	0.5822	0.99993	1	0.00061	0.61331

TABLE 5: Nemenyi-test p-values over the 10-fold cross validation accuracy values of methods used in Table 3 for the second stage

		PCT					BR					HOMER			
		Baseline	FSS	d=1	d=2	d=3	Baseline	FSS	d=1	d=2	d=3	Baseline	FSS	d=1	d=2
PCT	FSS	1	-	-	-	-	-	-	-	-	-	-	-	-	-
	d=1	0.00101	0.00276	-	-	-	-	-	-	-	-	-	-	-	-
	d=2	0.97305	0.99429	0.20441	-	-	-	-	-	-	-	-	-	-	-
	d=3	0.99941	0.99997	0.055	1	-	-	-	-	-	-	-	-	-	-
BR	Baseline	0.5081	0.34259	$9.90{\cdot}10^{-10}$	0.00736	0.04094	-	-	-	-	-	-	-	-	-
	FSS	0.78511	0.90222	0.52756	1	0.99972	0.0009	-	-	-	-	-	-	-	-
	d=1	0.03456	0.07183	0.9999	0.80968	0.46954	$3.50{\cdot}10^{-7}$	0.9789	-	-	-	-	-	-	-
	d=2	0.30646	0.46572	0.93095	0.99789	0.94692	$3.40{\cdot}10^{-5}$	1	0.99997	-	-	-	-	-	-
	d=3	1	1	0.00565	0.99876	1	0.23826	0.9556	0.11826	0.59781	-	-	-	-	-
HOMER	Baseline	0.78826	0.62872	$1.40{\cdot}10^{-8}$	0.03063	0.13026	1	0.0048	$3.50{\cdot}10^{-6}$	0.00024	0.4965	-	-	-	-
	FSS	0.86907	0.73534	$3.70{\cdot}10^{-8}$	0.04991	0.18995	1	0.0086	$8.00{\cdot}10^{-6}$	0.00049	0.6094	1	-	-	-
	d=1	0.03456	0.07183	0.9999	0.80968	0.46954	$3.50{\cdot}10^{-7}$	0.9789	1	0.99997	0.1183	$3.50{\cdot}10^{-6}$	$8.00{\cdot}10^{-6}$	-	-
	d=2	0.99826	0.99986	0.07636	1	1	0.02856	0.9999	0.55099	0.96972	1	0.09688	0.14508	0.55099	-
	d=3	0.99357	0.97058	$1.20{\cdot}10^{-6}$	0.23014	0.55099	0.99894	0.0605	0.00015	0.00554	0.9294	0.99999	1	0.00015	0.46954

TABLE 6: Nemenyi-test p-values over the 10-fold cross validation accuracy values of methods used in Table 3 for the third stage

		PCT					BR					HOMER			
		Baseline	FSS	d=1	d=2	d=3	Baseline	FSS	d=1	d=2	d=3	Baseline	FSS	d=1	d=2
PCT	FSS	0.99998	-	-	-	-	-	-	-	-	-	-	-	-	-
	d=1	$8.30{\cdot}10^{-5}$	$1.10{\cdot}10^{-6}$	-	-	-	-	-	-	-	-	-	-	-	-
	d=2	0.51198	0.09004	0.42062	-	-	-	-	-	-	-	-	-	-	-
	d=3	0.81264	0.2637	0.17182	1	-	-	-	-	-	-	-	-	-	-
BR	Baseline	0.83547	0.28767	0.15497	1	1	-	-	-	-	-	-	-	-	-
	FSS	0.9899	0.69257	0.02856	0.99947	1	1	-	-	-	-	-	-	-	-
	d=1	0.00026	$4.10{\cdot}10^{-6}$	1	0.59392	0.29076	0.26663	0.0605	-	-	-	-	-	-	-
	d=2	0.00614	0.00018	0.9998	0.96416	0.79758	0.77229	0.3632	0.99999	-	-	-	-	-	-
	d=3	0.42062	0.06245	0.51198	1	1	1	0.9982	0.68523	0.98252	-	-	-	-	-
HOMER	Baseline	1	1	$1.30{\cdot}10^{-5}$	0.26958	0.56663	0.59781	0.9261	$4.40{\cdot}10^{-5}$	0.00139	0.2044	-	-	-	-
	FSS	0.99999	1	$1.30{\cdot}10^{-6}$	0.09688	0.27854	0.30328	0.7107	$4.80{\cdot}10^{-6}$	0.00021	0.0675	1	-	-	-
	d=1	0.00108	$2.20{\cdot}10^{-5}$	1	0.80669	0.50422	0.47337	0.1451	1	1	0.8715	0.00021	$2.60{\cdot}10^{-5}$	-	-
	d=2	0.87384	0.33586	0.12674	1	1	1	1	0.22483	0.72134	1	0.65538	0.3528	0.41693	-
	d=3	0.99999	0.9746	0.00237	0.93858	0.99571	0.99696	1	0.00614	0.07521	0.8961	0.99918	0.97819	0.0196	0.99849

$d = 2$ and $d = 3$ does not produce significant improvements in the accuracy values of PCT, BR and HOMER baseline and FSS for the first, second and third stage. According to these results, the highest accuracies are obtained when there are more perturbations in the numerical labels, e.g. rounding to the tenth ($d = 1$) instead of the hundredth ($d = 2$). We can infer that SAR encoding then becomes an important factor, as it helps to optimize the feature weighting task, moving the solution within the search space to avoid being caught in a local minimum.

The aforementioned tables show that 73.26%, 75.58% and 74.61% are the best mean values of accuracy for the first, second and third stages of treatment, and they were obtained when performing PCT with $d = 1$. In addition, 0.72, 0.73 and 0.73 are the mean sensitivity values obtained when $d = 1$ for the first, second and third stages of treatment, respectively, indicating a good detection of patients who respond positively to treatment. Moreover, the model obtained with PCT obtains mean specificity values of 0.75, 0.76 and 0.75 for the first, second and third stage of the treatment, which indicates that this model is good when detecting patients who respond negatively to all stages of treatment. We can conclude that this panoramic prediction allows the doctor to provide an insightful preliminary criterion for the response to the treatment and make the respective medical decisions.

E. FEEDBACK PREDICTION

In this phase of the methodology, only rounding to the tenth has been considered, given the good results obtained in the previous experiment. To do this, the parameter D has been set to 1. Additionally, we have considered setting the number of groups G to 3 in order to categorize the records according to their low, medium and high level of NA cells. The dataset has been split using 75% for training and 25% for testing the hierarchical model. The training dataset has been split into training and validation when using the k-fold validation approach. The table B described in Section III-D1 is generated from the training dataset. This table is used for clustering the records by their NA values when applying the k-medians clustering with $k = G$. A hierarchy of models based on the number of NAs in each record is obtained, with 1, 4 and 12 being the number of NAs found in the centroids of the first, second and third groups, respectively. The groups obtained are trained using the RT+SAR and RT+FSS combinations. CFS is the FSS method used in this experiment and it is set in the same way as in Section IV-D. RT has been selected for the classification task, given the high accuracies obtained in [14]. The accuracies achieved by the three hierarchical models (models 1, 2 and 3) generated when considering the NA number are presented in Table 7.

With the purpose of building a fuzzy selector that considers the number of NAs in new records when assigning the correspondent model, the FURIA algorithm has been applied to the T_{map} table described in Section III-D3 with the following parameters: 3 folds for pruning (the rest for growing the rules), 2 as the number of optimization runs, 2 as the minimum total weight of the instances in a rule and 2 as the number of decimal places to be used for the output of numbers in the model. One rule per model with an accuracy of 85.52% has been obtained with the FURIA algorithm. Regarding accuracy, we need to clarify that the purpose of these rules is not to classify treatment responses, but to build a fuzzy selector that assigns the corresponding model. The rules R1, R2 and R3 are defined according to the number of missing values in the inteval [0,14], where:

- R1: If the number of NAs falls in the region defined by the trapezoidal membership function with [0, 0, 3, 4], then the selected model will be "model1" with a CF of 0.83
- R2: If the number of NAs falls in the region defined by the trapezoidal membership function with [3, 4, 14, 14], then the selected model will be "model2" with a CF of 0.85
- R3: If the number of NAs falls in the region defined by the trapezoidal membership function with [0, 0, 11, 12], then the selected model will be "model2" with a CF of 0.78

where 14 is the maximum number of NAs found in the medical registers, CF is the certainty factor and [a,b,c,d] represents the boundaries of the trapezoidal region [44]. These functions are graphically represented in Figure 6. If the number of NAs falls in the middle of three regions as

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2927429, IEEE Access

in the case of NAs = 3, the selected model will be the model with the highest CF value. After that, the missing values are replaced with the values obtained from multiple imputation within their group. These rules have been applied to the records, obtaining the best results when performing the RT+SAR combination instead of RT+FSS, with a mean accuracy of 90.17%, 87.86% and 85.54%, as shown in the row labelled "Hierarchy" in Table 7. Moreover, the high values of sensitivity and specificity indicate the goodness of the hierarchy model when predicting the "high" and "low" responses to treatment.

In order to check whether the improvement in classification due to the SAR method is statistically significant, the Mann-Whitney U (non-parametric) test was carried out between the accuracy values of the hierarchical model of RT+SAR and the RT+FSS methods using 10-fold cross validation. The results are presented in Table 7. The results of models 1, 2 and 3 are not taken into account in the statistical validation since they are part of the final hierarchical model built in the MVDMS^2 process presented in Section IV-E. The adjusted p-values are compared against a significance level of $\alpha = 0.05$ to reject or accept the null hypothesis that a pair of methods perform equally. We have obtained the p-values of $1.083{\cdot}10^{-5}$, $1.013{\cdot}10^{-5}$ and 0.002089 for the first, second and third stages of the treatment prediction, respectively. These values, being less than 0.05, guarantee that there is a significant difference in the distributions of values between the two methods.

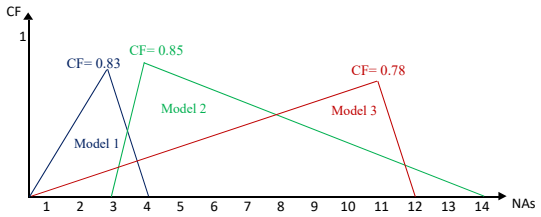

FIGURE 6: Membership functions of the fuzzy model selector

F. RELEVANT MEDICAL FACTORS

After obtaining the prediction models presented in Tables 3 and 7, we proceeded to extract the most important factors for each model. For this purpose, the steps described in Section III-E were applied to each model. The methodology applied in [14] has also been considered for the relevant factors analysis for the first, second and third stages of treatment. For the case of PCT with $d = 1$ (panoramic prediction model), the medical factors are the same in the three stages, given that a consensus model is obtained from the multi-target prediction models. All relevant medical factors from the 13 consensus models are presented in Table 8. The selected features when applying FSS in Section IV-D are also presented in Table 8. The "GON", "1st grade family with migraine" and "Drugs tested before toxin" medical factors have been selected for the consensus models 9, 8 and 6 times, respectively. Moreover, "1st grade family with migraine" and "Chronic migraine time evolution" have also been selected for the majority of the first and second stage consensus models, respectively. Furthermore, "Headache days per month", "Unilateral pain" and "Migraine days per month" are present in the majority of the third stage consensus models. Finally, "GON", "Drugs tested before toxin" and "Chronic migraine time evolution" have also been selected by FSS, which was described in Section IV-D.

To summarize, the medical factors that appear at least once for each stage in the consensus models of the feedback prediction or in the panoramic consensus model are: "Hemoglobin", "Analgesic abuse", "Serum iron", "1st grade family with migraine", "Chronic migraine time evolution", "GOT", "Headache days per month", "Unilateral pain", "Platelets", "Anxiety" and "Onset age of toxin treatment". These are the most important medical factors among all the relevant factors from the consensus models presented in Table 8.

G. MEDICAL DISCUSSION

In agreement with current publications, we find some predictors of response to treatment with BoNT-A, namely: migraine time evolution [57], [58], unilateral pain [58], [59], analgesic abuse [60], days of headache [58] and the retroocular component [61]. Moreover, these articles continue supporting the approach of not delaying treatment with BoNT-A in those patients who have a diagnosis of chronic migraine, who will improve more than those with a shorter evolution time and with a profile of lesser severity of the migraine. Following this line of thought, it is not strange to find that the presence of status, the number of triptans per month or the number of previous tested drugs, are also predictors of response.

A fact not assessed so far that we find very interesting is the predictive nature of the response in patients who take concomitant oral preventive treatment. Although it is not described in the literature, it is possible that the variables such as relatives in the first degree, the catamenial component and the presence of sensory alterations such as sono or photophobia, are predictive, either because they really assure us that we are dealing with a chronic migraine, a fact whose diagnosis is not always easy when a patient presents daily headaches and the semiological profile is no longer so pure.

We could not find a clinical relation with the analytical parameters (liver profile, iron, platelets, creatinine, hemoglobin) and associated pathologies such as dermopathy, gastropathy, dyslipidemia, hypertension and lung disease. But these points open up future lines of research with more targeted prospective studies. We also agree with the literature that neither gender nor nausea or vomiting [62] have been predictive. Although, in contrast to these, we do predict psychic state and concomitant sensory symptoms (photo / sono and osmophobia). The latter is because of what has already been explained, and the state of mind is due to the fact that it could be related to a more serious patient profile.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2927429, IEEE Access

TABLE 7: Estimated performance metrics (mean $\pm$ std deviation) of hierarchy models with $D = 1$ and $G = 3$ using 10-fold cross validation. The best results are highlighted in bold.

Algorithms	Model	First stage			Second stage			Third stage		
		Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
RT+SAR	Model 1	89.28%±2.14	0.87±0.02	0.91±0.03	92.00%±1.16	0.94±0.02	0.87±0.02	93.33%±0.92	0.85±0.02	0.94±0.01
	Model 2	88.88%±2.38	0.91±0.02	0.85±0.03	89.47%±1.91	0.90±0.01	0.88±0.02	84.61%±2.42	0.86±0.02	0.83±0.03
	Model 3	91.30%±1.02	0.93±0.01	0.85±0.02	90.40%±1.14	0.87±0.03	0.92±0.01	92.85%±0.97	0.88±0.02	0.95±0.01
	Hierarchy	**90.17%±1.31**	0.90±0.01	0.91±0.01	**87.86%±2.04**	0.86±0.02	0.90±0.01	**85.54%±1.92**	0.92±0.01	0.84±0.02
RT+FSS	Model 1	80.01%±0.71	0.85±0.01	0.75±0.01	73.33%±0.84	0.71±0.02	0.75±0.01	86.66%±1.35	0.85±0.01	0.87±0.03
	Model 2	76.92%±2.14	0.71±0.03	0.83±0.02	84.61%±1.93	0.86±0.02	0.83±0.03	69.23%±3.81	0.71±0.03	0.66±0.04
	Model 3	80.01%±0.95	0.71±0.02	0.87±0.01	66.66%±2.26	0.62±0.03	0.71±0.03	73.33%±2.12	0.71±0.02	0.75±0.02
	Hierarchy	83.23%±1.83	0.84±0.01	0.83±0.02	78.61%±1.76	0.80±0.03	0.77±0.02	82.17%±1.84	0.87±0.03	0.81±0.02

TABLE 8: Relevant factors for models of RT, PCT of Table 3, RT+SAR of Table 7 and FSS

	Model	First stage	Second stage	Third stage
Consensus models	RT	Hemoglobin, Analgesic abuse, GPT, Serum iron, GON, 1st grade family with migraine, Calcium antagonists, Chronic migraine time evolution, GOT, Retroocular component	Migraine days per month, Drugs tested before toxin, Dermopathy, Analgesics days per month, Pneumopathy, Serum iron, Concomitant treatment with statins, GPT, Triptans per month, Analgesic abuse	Concomitant treatment with statins, Drugs tested before toxin, Tricyclic antidepressants, Retroocular component, Migraine days per month, Serum iron, Headache days per month
	RT+SAR Model 1	Platelets, Concomitant treatment with statins, Unilateral pain, GON, 1st grade family with migraine, Onset age of toxin treatment, Creatinine	Analgesics days per month, Preventive oral treatment at time of infiltration, Neuromodulator, Concomitant antidepressant treatment, GON, Chronic migraine time evolution, Chronic migraine	Hemoglobin, Platelets, History of migraine status, Headache days per month, 1st grade family with migraine, GON, Preventive oral treatment at time of infiltration, Unilateral pain, Pneumopathy, Catamenial, Depression
	RT+SAR Model 2	Concomitant antihypertensive treatment, Analgesics days per month, Migrant status history, GON, Onset age of toxin treatment, 1st grade family with migraine	GON, Chronic migraine time evolution, Headache days per month, Hemoglobin, Catamenial, Concomitant antihypertensive treatment, Creatinine, Onset age of toxin treatment, Anxiety, Depression	GOT, GON, Preventive oral treatment at time of infiltration, Drugs tested before toxin, GGT, Onset age of toxin treatment, Migraine days per month
	RT+SAR Model 3	Retroocular component, GGT, Migraine days per month, Drugs tested before toxin, Neuromodulator, Concomitant antihypertensive treatment, Enolism, Analgesics days per month, 1st grade family with migraine	Unilateral pain, GON, Drugs tested before toxin, Chronic migraine time evolution, Chronic migraine, Anxiety,1st grade family with migraine, Analgesics days per month, Platelets	Concomitant treatment with statins, Headache days per month, Unilateral pain, Migraine days per month, GON, Chronic migraine time evolution, Onset age of toxin treatment, 1st grade family with migraine, Platelets, Anxiety
	PCT ($d = 1$)	GOT, Drugs tested before toxin, Chronic migraine, Unilateral pain, Analgesic abuse, Headache days per month, 1st grade family with migraine, Anxiety		
FSS (Section IV-D)		Sex, Chronic migraine, Chronic migraine time evolution, GON, Drugs tested before toxin, Preventive oral treatment, catamenial, Concomitant treatment with statins, Gastropathy, Pneumopathy, Headache days per month, Analgesics days per month		

To conclude, several of the medical factors relevant to the treatment of migraine under BoNT-A are coherent with the medical literature. These are: migraine time evolution, unilateral pain, analgesic abuse, days of headache and the retroocular component. Other medical factors revealed as relevant by our methodology have no medical explanation yet. Therefore, they should be studied in the future with more specific prospective studies.

V. CONCLUSIONS

This study presents a methodology for obtaining a treatment response prediction to various stages of a treatment. Knowing in advance the answers to the continuous treatment is vital in terms of economy, quality of life and time.

Two data mining approaches have been considered to solve this problem. The first is based on panoramic prediction, generating a predictive model that allows us to know in advance the response to the treatment over all its stages. The second approach offers a different predictive model for each stage of the treatment, taking into account for each stage of the treatment the medical results of the previous stage, except for the first one. Additionally, consensus models have been applied to extract the relevant medical factors that influence the response to the treatment. In this way, it is possible to overcome the existing gap between the biomedical community and the data mining community, allowing the former to analyze the medical factors that enable an effective treatment from a clinical point of view.

With the purpose of verifying the effectiveness of the techniques described, they have been applied to a real scenario, specifically to the treatment of migraine, given its socioeconomic impact. Accuracies close to 75% for all stages have been achieved using the panoramic approach. In addition, an important improvement has been obtained with the second approach, obtaining accuracies close to 90% for each stage. As a consequence of the extraction of the most relevant medical factors from the predictive models, the clinical opinion has been corroborated, on the basis of factors such as: the duration of the migraine, unilateral pain, analgesic abuse, retroocular component and days with headache. This knowledge allows us to conclude that the features considered in our prediction models are coherent with respect to the medical literature. Moreover, new medical factors are revealed for further studies, such as: oral concomitant preventive treatment, iron, platelets, creatinine and hemoglobin.

Future work should address the inclusion of medical criteria in predictive models, which would make it possible to obtain decision support systems to select the most appropriate continuous treatment for each patient.

REFERENCES

[1] J. C. Kvedar, A. L. Fogel, E. Elenko, and D. Zohar, "Digital medicine's march on chronic disease," Nature biotechnology, vol. 34, no. 3, p. 239, 2016.

[2] M. Sensi, G. Cossu, F. Mancini, M. Pilleri, M. Zibetti, N. Modugno, R. Quatrale, F. Tamma, A. Antonini, M. Aguggia et al., "Which patients discontinue? issues on levodopa/carbidopa intestinal gel treatment: Italian multicentre survey of 905 patients with long-term follow-up," Parkinsonism & related disorders, vol. 38, pp. 90–92, 2017.

[3] A. J. Leidner, H. W. Chesson, F. Xu, J. W. Ward, P. R. Spradling, and S. D. Holmberg, "Cost-effectiveness of hepatitis c treatment for patients in early stages of liver disease," Hepatology, vol. 61, no. 6, pp. 1860–1869, 2015.

[4] D. B. Rein, J. S. Wittenborn, B. D. Smith, D. K. Liffmann, and J. W. Ward, "The cost-effectiveness, health benefits, and financial costs of new antiviral treatments for hepatitis c virus," Clinical infectious diseases, vol. 61, no. 2, pp. 157–168, 2015.

[5] W. H. Herman, W. Ye, S. J. Griffin, R. K. Simmons, M. J. Davies, K. Khunti, G. E. Rutten, A. Sandbaek, T. Lauritzen, K. Borch-Johnsen et al., "Early detection and treatment of type 2 diabetes reduce cardiovascular morbidity and mortality: a simulation of the results of the anglo-danish-dutch study of intensive treatment in people with screen-detected diabetes in primary care (addition-europe)," Diabetes care, vol. 38, no. 8, pp. 1449–1455, 2015.

[6] H. Wu, S. Yang, Z. Huang, J. He, and X. Wang, "Type 2 diabetes mellitus prediction model based on data mining," Informatics in Medicine Unlocked, vol. 10, pp. 100–107, 2018.

[7] J. Chen, K. Li, Z. Tang, K. Bilal, and K. Li, "A parallel patient treatment time prediction algorithm and its applications in hospital queuing-recommendation in a big data environment," IEEE Access, vol. 4, pp. 1767–1783, 2016.

[8] V. Huddar, B. K. Desiraju, V. Rajan, S. Bhattacharya, S. Roy, and C. K. Reddy, "Predicting complications in critical care using heterogeneous clinical data," IEEE Access, vol. 4, pp. 7988–8001, 2016.

[9] A. Peterkova, M. Nemeth, and A. Bohm, "Computing missing values using neural networks in medical field," in 2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES). IEEE, 2018, pp. 000 151–000 156.

[10] M. Kurosaki, Y. Tanaka, N. Nishida, N. Sakamoto, N. Enomoto, M. Honda, M. Sugiyama, K. Matsuura, F. Sugauchi, Y. Asahina et al., "Pre-treatment prediction of response to pegylated-interferon plus ribavirin for chronic hepatitis c using genetic polymorphism in il28b and viral factors," Journal of hepatology, vol. 54, no. 3, pp. 439–448, 2011.

[11] P. Lambin, R. G. Van Stiphout, M. H. Starmans, E. Rios-Velazquez, G. Nalbantov, H. J. Aerts, E. Roelofs, W. Van Elmpt, P. C. Boutros, P. Granone et al., "Predicting outcomes in radiation oncology—multifactorial decision support systems," Nature reviews Clinical oncology, vol. 10, no. 1, p. 27, 2013.

[12] R. Armañanzas, C. Bielza, K. R. Chaudhuri, P. Martinez-Martin, and P. Larrañaga, "Unveiling relevant non-motor parkinson's disease severity symptoms using a machine learning approach," Artificial intelligence in medicine, vol. 58, no. 3, pp. 195–202, 2013.

[13] F. Parrales Bravo, A. A. Del Barrio García, M. Gallego de la Sacristana, L. López Manzanares, J. Vivancos, and J. L. Ayala Rodrigo, "Support system to improve reading activity in parkinson's disease and essential tremor patients," Sensors, vol. 17, no. 5, p. 1006, 2017.

[14] F. B. Parrales, A. B. G. Del, M. Gallego, A. V. Gago, M. Ruiz, A. P. Guerrero, J. Ayala et al., "Prediction of patient's response to onabotulinumtoxina treatment for migraine." Heliyon, vol. 5, no. 2, pp. e01 043–e01 043, 2019.

[15] M. Gil, R. El Sherif, M. Pluye, B. C. Fung, R. Grad, and P. Pluye, "Towards a knowledge-based recommender system for linking electronic patient records with continuing medical education information at the point of care," IEEE Access, 2019.

[16] G. Manogaran, C. Thota, D. Lopez, V. Vijayakumar, K. M. Abbas, and R. Sundarsekar, "Big data knowledge system in healthcare," in Internet of things and big data technologies for next generation healthcare. Springer, 2017, pp. 133–157.

[17] S. S. Wu, K. S. Chan, J. Bae, and E. W. Ford, "Electronic clinical reminder and quality of primary diabetes care," Primary care diabetes, vol. 13, no. 2, pp. 150–157, 2019.

[18] N. D. Bakerly, A. Woodcock, S. Collier, D. A. Leather, J. P. New, J. Crawford, C. Harvey, J. Vestbo, and I. Boucot, "Benefit and safety of fluticasone furoate/vilanterol in the salford lung study in chronic obstructive pulmonary disease (sls copd) according to baseline patient characteristics and treatment subgroups," Respiratory medicine, vol. 147, pp. 58–65, 2019.

[19] A. Sollie, R. H. Sijmons, C. Helsper, and M. E. Numans, "Reusability of coded data in the primary care electronic medical record: A dynamic cohort study concerning cancer diagnoses," International journal of medical informatics, vol. 99, pp. 45–52, 2017.

[20] J. Pagán, D. Orbe, M. Irene, A. Gago, M. Sobrado, J. L. Risco-Martín, J. V. Mora, J. M. Moya, and J. L. Ayala, "Robust and accurate modeling approaches for migraine per-patient prediction from ambulatory data," Sensors, vol. 15, no. 7, pp. 15 419–15 442, 2015.

[21] M. F. Collen and M. J. Ball, The history of medical informatics in the United States. Springer, 2015.

[22] W. Waegeman, K. Dembczyński, and E. Hüllermeier, "Multi-target prediction: a unifying view on problems and methods," Data Mining and Knowledge Discovery, vol. 33, no. 2, pp. 293–324, 2019.

[23] G. Madjarov, D. Kocev, D. Gjorgjevikj, and S. Džeroski, "An extensive experimental comparison of methods for multi-label learning," Pattern recognition, vol. 45, no. 9, pp. 3084–3104, 2012.

[24] L. A. Zadeh, "Information and control," Fuzzy sets, vol. 8, no. 3, pp. 338–353, 1965.

[25] H. Jemal, Z. Kechaou, and M. B. Ayed, "Towards a medical intensive care unit decision support system based on intuitionistic fuzzy logic," in International Conference on Intelligent Systems Design and Applications. Springer, 2016, pp. 602–611.

[26] M. Nilashi, O. Ibrahim, H. Ahmadi, and L. Shahmoradi, "A knowledge-based system for breast cancer classification using fuzzy logic method," Telematics and Informatics, vol. 34, no. 4, pp. 133–144, 2017.

[27] A. K. Paul, P. C. Shill, M. R. I. Rabin, and M. Akhand, "Genetic algorithm based fuzzy decision support system for the diagnosis of heart disease," in 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV). IEEE, 2016, pp. 145–150.

[28] M. M. Hoehn, M. D. Yahr et al., "Parkinsonism: onset, progression, and mortality," Neurology, vol. 50, no. 2, pp. 318–318, 1998.

[29] P. Martínez-Martín, M. J. Forjaz, E. Cubo, B. Frades, J. de Pedro Cuesta, and E. P. Members, "Global versus factor-related impression of severity in parkinson's disease: a new clinimetric index (cisi-pd)," Movement Disorders, vol. 21, no. 2, pp. 208–214, 2006.

[30] M. Yang, R. Rendas-Baum, S. F. Varon, and M. Kosinski, "Validation of the headache impact test (hit-6™) across episodic and chronic migraine," Cephalalgia, vol. 31, no. 3, pp. 357–367, 2011.

[31] A. Gasbarrini, A. L. De, G. Fiore, M. Gambrielli, F. Franceschi, V. Ojetti, E. Torre, G. Gasbarrini, P. Pola, and M. Giacovazzo, "Beneficial effects of helicobacter pylori eradication on migraine." Hepato-gastroenterology, vol. 45, no. 21, pp. 765–770, 1998.

[32] J. Schoenen, J. Jacquy, and M. Lenaerts, "Effectiveness of high-dose riboflavin in migraine prophylaxis a randomized controlled trial," Neurology, vol. 50, no. 2, pp. 466–470, 1998.

[33] W. Rogowski, K. Payne, P. Schnell-Inderst, A. Manca, U. Rochau, B. Jahn, O. Alagoz, R. Leidl, and U. Siebert, "Concepts of 'personalization'in personalized medicine: implications for economic evaluation," Pharmacoeconomics, vol. 33, no. 1, pp. 49–59, 2015.

[34] H. Blockeel and L. De Raedt, "Top-down induction of first-order logical decision trees," Artificial intelligence, vol. 101, no. 1-2, pp. 285–297, 1998.

[35] M.-L. Zhang and Z.-H. Zhou, "Ml-knn: A lazy learning approach to multi-label learning," Pattern recognition, vol. 40, no. 7, pp. 2038–2048, 2007.

[36] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Effective and efficient multilabel classification in domains with large number of labels," in Proc. ECML/PKDD 2008 Workshop on Mining Multidimensional Data (MMD'08), vol. 21. sn, 2008, pp. 53–59.

[37] J. Read, P. Reutemann, B. Pfahringer, and G. Holmes, "Meka: a multi-label/multi-target extension to weka," The Journal of Machine Learning Research, vol. 17, no. 1, pp. 667–671, 2016.

[38] "Clus: Framework for predictive clustering," http://clus.sourceforge.net/doku.php, accessed: 2018-04-17.

[39] Z. Yu, Z. Niu, W. Tang, and Q. Wu, "Deep learning for daily peak load forecasting—a novel gated recurrent neural network combining dynamic time warping," IEEE Access, 2019.

[40] E. Aarts and J. Korst, "Simulated annealing and boltzmann machines," 1988.

[41] N. J. Higham, Accuracy and stability of numerical algorithms. Siam, 2002, vol. 80, p. 54.

[42] M. I. García Planas and S. Tarragona Romero, "Perturbación de los valores propios simples de matrices de polinomios dependientes diferenciablemente de parámetros," in 2nd Meeting on Linear Algebra Matrix analysis and applications. Servicio de publicaciones de la UPV, 2010, pp. 1–7.

[43] A. K. Jain and R. C. Dubes, "Algorithms for clustering data," 1988.

[44] J. Hühn and E. Hüllermeier, "Furia: an algorithm for unordered fuzzy rule induction," Data Mining and Knowledge Discovery, vol. 19, no. 3, pp. 293–319, 2009.
[45] W. W. Cohen, "Fast effective rule induction," in Machine Learning Proceedings 1995. Elsevier, 1995, pp. 115–123.
[46] H. C. C. of the International Headache Society (IHS et al., "The international classification of headache disorders, (beta version)," Cephalalgia, 2013.
[47] J. Natoli, A. Manack, B. Dean, Q. Butler, C. Turkel, L. Stovner, and R. Lipton, "Global prevalence of chronic migraine: a systematic review," Cephalalgia, 2009.
[48] D. Buse, A. Manack, D. Serrano, C. Turkel, and R. Lipton, "Sociodemographic and comorbidity profiles of chronic migraine and episodic migraine sufferers," Journal of Neurology, Neurosurgery & Psychiatry, vol. 81, no. 4, pp. 428–432, 2010.
[49] N. T. Mathew and S. F. A. Jaffri, "A double-blind comparison of onabotulinumtoxina (botox) and topiramate (topamax) for the prophylactic treatment of chronic migraine: A pilot study," Headache: The Journal of Head and Face Pain, vol. 49, no. 10, pp. 1466–1478, 2009.
[50] R. K. Cady, C. P. Schreiber, J. A. Porter, A. M. Blumenfeld, and K. U. Farmer, "A multi-center double-blind pilot comparison of onabotulinumtoxina and topiramate for the prophylactic treatment of chronic migraine," Headache: The Journal of Head and Face Pain, vol. 51, no. 1, pp. 21–32, 2011.
[51] S. D. Silberstein, D. W. Dodick, S. K. Aurora, H.-C. Diener, R. E. DeGryse, R. B. Lipton, and C. C. Turkel, "Per cent of patients with chronic migraine who responded per onabotulinumtoxina treatment cycle: Preempt," J Neurol Neurosurg Psychiatry, vol. 86, no. 9, pp. 996–1001, 2015.
[52] C. Lovati and L. Giani, "Action mechanisms of onabotulinum toxin-a: hints for selection of eligible patients," Neurological Sciences, vol. 38, no. 1, pp. 131–140, 2017.
[53] N. Sabahat, A. A. Malik, and F. Azam, "A size estimation model for board-based desktop games," IEEE Access, vol. 5, pp. 4980–4990, 2017.
[54] S. H. Park and K. Han, "Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction," Radiology, vol. 286, no. 3, pp. 800–809, 2018.
[55] M. A. Hall, "Correlation-based feature selection for machine learning," 1999.
[56] S. Garcia and F. Herrera, "An extension on"statistical comparisons of classifiers over multiple data sets"for all pairwise comparisons," Journal of Machine Learning Research, vol. 9, no. Dec, pp. 2677–2694, 2008.
[57] E. J. Eross, J. P. Gladstone, S. Lewis, R. Rogers, and D. W. Dodick, "Duration of migraine is a predictor for response to botulinum toxin type a," Headache: The Journal of Head and Face Pain, vol. 45, no. 4, pp. 308–314, 2005.
[58] C. Domínguez, P. Pozo-Rosich, M. Torres-Ferrús, N. Hernández-Beltrán, C. Jurado-Cobo, C. González-Oria, S. Santos, M. Monzón, G. Latorre, L. Álvaro et al., "Onabotulinumtoxina in chronic migraine: predictors of response. a prospective multicentre descriptive study," European journal of neurology, vol. 25, no. 2, pp. 411–416, 2018.
[59] N. T. Mathew, J. Kailasam, and L. Meadors, "Botulinum toxin type a for the treatment of nummular headache: four case studies," Headache: The Journal of Head and Face Pain, vol. 48, no. 3, pp. 442–447, 2008.
[60] F. G. Freitag, "Importance of botulinum toxin for prevention of migraine," Expert review of neurotherapeutics, vol. 10, no. 3, pp. 339–340, 2010.
[61] K.-H. Lin, S.-P. Chen, J.-L. Fuh, Y.-F. Wang, and S.-J. Wang, "Efficacy, safety, and predictors of response to botulinum toxin type a in refractory chronic migraine: A retrospective study," Journal of the Chinese Medical Association, vol. 77, no. 1, pp. 10–15, 2014.
[62] M. Jakubowski, P. J. McAllister, Z. H. Bajwa, T. N. Ward, P. Smith, and R. Burstein, "Exploding vs. imploding headache in migraine prophylaxis with botulinum toxin a," Pain, vol. 125, no. 3, pp. 286–295, 2006.