



## Predicting seminal quality with artificial intelligence methods

David Gil<sup>a,\*</sup>, Jose Luis Girela<sup>b,1</sup>, Joaquin De Juan<sup>b</sup>, M. Jose Gomez-Torres<sup>b</sup>, Magnus Johnsson<sup>c</sup>

<sup>a</sup> Computing Technology and Data Processing, University of Alicante, Spain

<sup>b</sup> Department of Biotechnology, University of Alicante, Spain

<sup>c</sup> Lund University Cognitive Science, Sweden

### ARTICLE INFO

#### Keywords:

Artificial neural network  
Support Vector Machines  
Decision trees  
Diagnosis  
Decision support system  
Expert system  
Semen quality  
Male fertility potential

### ABSTRACT

Fertility rates have dramatically decreased in the last two decades, especially in men. It has been described that environmental factors, as well as life habits, may affect semen quality. Artificial intelligence techniques are now an emerging methodology as decision support systems in medicine.

In this paper we compare three artificial intelligence techniques, decision trees, Multilayer Perceptron and Support Vector Machines, in order to evaluate their performance in the prediction of the seminal quality from the data of the environmental factors and lifestyle.

To do that we collect data by a normalized questionnaire from young healthy volunteers and then, we use the results of a semen analysis to assess the accuracy in the prediction of the three classification methods mentioned above.

The results show that **Multilayer Perceptron and Support Vector Machines show the highest accuracy, with prediction accuracy values of 86% for some of the seminal parameters.** In contrast decision trees provide a visual and illustrative approach that can compensate the slightly lower accuracy obtained.

In conclusion artificial intelligence methods are a useful tool in order to predict the seminal profile of an individual from the environmental factors and life habits. From the studied methods, Multilayer Perceptron and Support Vector Machines are the most accurate in the prediction. Therefore these tools, together with the visual help that decision trees offer, are the suggested methods to be included in the evaluation of the infertile patient.

© 2012 Elsevier Ltd. All rights reserved.

### 1. Introduction

After the publication of a meta-analysis directed by Elisabeth Carlsen (Carlsen, Giwercman, Keiding, & Skakkebaek, 1992), there is a debate about a possible decline in semen quality. Numerous studies show a decrease in semen parameters during the past two decades (Auger, Kunstmann, Czyglik, & Jouannet, 1995; Splingart et al., 2011; Swan, Elkin, & Fenster, 1997; Swan, Elkin, & Fenster, 2000), affecting the male fertility potential. Between the factors considered to explain this decline, there is an increase in the incidence of male reproductive diseases (Irvine, 2000; Splingart et al., 2011), but also has been suggested the effect of environmental or occupational factors (Giwercman & Giwercman, 2011; Wong, Zielhuis, Thomas, Merkus, & Steegers-Theunissen, 2003), to a certain lifestyle (Martini et al., 2004; Agarwal, Desai, Ruffoli, & Carpi, 2008).

To evaluate the male partner, clinicians use the data obtained from semen analysis (Kolettis, 2003), and they compare the obtained results with the corresponding reference value estab-

lished by World Health Organization (WHO, 1999). Semen analysis is a good predictor of the male fertility potential (Bonde et al., 1998; Guzik et al., 2001; Slama et al., 2002; Zinaman, Brown, Selevan, & Clegg, 2000), and is also necessary to evaluate candidates to become semen donors (Barratt et al., 1998; Carrell, Cartmill, Jones, Hatasaka, & Peterson, 2002; Ecochard, Cottinet, Mathieu, Rabilloud, & Czyba, 1999; Society British Andrology, 1999). There is a high variability in the testicular function of an individual (Keel, 2006), so **it is recommended to interpret the results taking in account certain factors (i.e., fever, toxic exposure) that can modify the semen parameters (Rowe & Comhaire, 2000).**

In this paper, assuming the influence of environmental factors and life habits in semen quality, we compare the prediction accuracy of three different artificial intelligence (AI) methods, Multilayer Perceptron (MLP), Support Vector Machines (SVM) and Decision Trees (DT), to determine the best Decision Support Systems (DSS) that can help in the evaluation of male fertility potential.

The current advances and progresses in the field of AI have led to the emergence of expert systems and DSSs for economics, linguistics, management science, mathematical modelling, psychology, etc.

There are good classifiers in the AI such as artificial neural network (ANN), DT (Polat & Gunes, 2009b), SVM (Conforti & Guido,

\* Corresponding author.

E-mail address: [dgil@dtic.ua.es](mailto:dgil@dtic.ua.es) (D. Gil).

<sup>1</sup> These authors equally contributed to this work.

2010) or even hybrid methods that combine ANNs and fuzzy logic into fuzzy neural networks (FNN) (Kahramanli & Allahverdi, 2008) which are widely used to aid medical diagnosis by means of decision support systems construction.

Moreover, the use of AI has also become widely accepted in medical applications. These methods include advantages as (a) Ease of optimisation, resulting in cost-effective and flexible non-linear modelling of large data sets; (b) Accuracy for predictive inference, with potential to support clinical decision making; (c) These models can make knowledge dissemination easier by providing explanation, for instance, using rule extraction or sensitivity analysis (Lisboa & Taktak, 2006).

The amount of information used in order to predict the male fertility potential makes really useful the application with AI methods not only for improving the accuracy but also to select the best features as, very often, the number of feature to deal with is huge (Subashini, Ramalingam, & Palanivel, 2009; Polat & Gunes, 2009a; Gil, Johnsson, Garcia Chamizo, Paya, & Fernandez, 2009; Gil & Johnsson, 2010a, 2010b, 2011; Gil, Johnsson, Garcia Chamizo, Paya, & Fernandez, 2011). This will lead to the knowledge discovery in databases, data mining or the process of extracting patterns from large data sets.

The main objective of this paper is to compare a number of classification methods applied to male fertility data sets. The approach chosen to address the problem is by the use of different AI methods, in particular **DT, MLP and SVM**. A comparative study of those will give us insight into the merits of the different methods when used on this problem.

The remaining part of the paper is organized as follows: First we start defining the materials and methods of the study (study population and the variables of the questionnaire). Then we continue with a brief description of the AI methods used in this paper: MLP, SVM and DT. Then we proceed by describing the design of our proposal and the experiments carried out in the result section; the available data as well as a detailed explanation of the different values of our database; Then we continue by describing the subsequent testing carried out in order to analyse the results; Finally we draw the relevant conclusions.

## 2. Materials and methods

### 2.1. Study population

The study was performed with young healthy volunteers among the students of the **University of Alicante, 100 volunteers' between 18 and 36 years old participated in the study**. After being informed they were asked to provide a semen sample after 3 to 6 days of sexual abstinence, and a semen analysis according to World Health Organization (WHO) guidance was performed. Those with previous known reproductive alterations (i.e., Varicocele) were excluded of the statistical analysis.

### 2.2. Variables of the questionnaire

On the day of the analysis, the volunteers were asked to fulfill a questionnaire about life habits and health status. **Table 1** shows the name, a description and the range of values of the variables used in the study as well as the values normalized. **Table 1** shows the treatment we have done to the different database input fields. We have converted the input data into a range of normalization according to the follow rules:

- (a) Numerical variables such as age, Year of analysis, the number of exposures or Average number of cigarettes per day are normalized onto the interval (0, 1). For instance

the average number of cigarettes per day is indicated in the second column (Values) as a range between the minimum 0 and the maximum 25. This means that the persons that consumes **25 cigarettes is normalized to the value 25/25 = 1** whereas an individual that consumes 14 cigarettes per day is normalized to the value  $14/25 = 0.56$

- (b) The variables with only two independent attributes are prearranged with binary values (0, 1). We have many of those representations.
- (c) The variables with three independent attributes, such as "Vaccines received", "High fevers in the last year" and "Smoking habit" are prearranged using the ternary values (−1, 0, 1). For example, the last one "Smoking habit" will take −1 for never, 0 represents occasional and 1 daily.
- (d) The variables with four independent attributes, such as "Season in which the analysis was performed" or "Marital status" are prearranged using the four different and equal distance values (−1, −0.33, 0.33, 1). For instance, "Marital status" will take −1 unmarried and without a partner, −0.33 represents unmarried with partner, 0.33 is the value married and 4 for something else.

The variables with a range of more than four independent attributes (i.e., some of them have 10 different values) are considered numerical variables as in the option (a).

### 2.3. Decision trees

A Decision tree is a classifier in the form of a tree structure (see Fig. 1), where each node is either a leaf node, which indicates the value of the target attribute (class) of examples, or a decision node, which specifies some test to be carried out on a single attribute-value, with one branch and sub-tree for each possible outcome of the test.

Decision trees are powerful and popular tools for classification and prediction. In contrast to other AI methods, such as MLP and SVM (that both will be explained later), the advantages of decision trees is due to the fact that they represent rules. Rules can readily be expressed so that persons can understand them or even directly used in a database.

In some areas of applications, the accuracy of a classification or prediction is the only thing that matters. In such situations we do not necessarily concern about how the model works. However, in other situations, the ability to explain the reason for a decision, is crucial. For instance, in medicine in general, and to predict the male fertility potential in particular, one has describe the the factors involved, so that other physicians can utilize this knowledge in launching a successful prediction. This domain experts must recognize and approve this discovered knowledge, and for this we need good descriptions.

**Modeled after the structure of a tree, decision trees (DTs) are able to provide a good explanation applicable to the life habits of the populations and therefore interpret problems very much according to the principles of mathematical and statistical principles (Risso & Brida, 2010).**

A decision tree may be used to classify a given example (see Fig. 1); one begins at the root and follows the path provided by the answers to the questions in the internal nodes until a leaf is reached. The DT algorithm family includes classical algorithms, such as ID3 (Quinlan, 1986), C4.5 (Quinlan, 1993), and CART (Breiman, 1984).

DT is a fairly mature technique which includes models such as C5.0, C&RT, CHAID, QUEST and ID3, the difference of which is in the derivations of formula, such as entropy and information gain, to determine the attributes that result in splitting. **However, most decision trees consist of many nodes, which, under certain**

**Table 1**

List of variables with their descriptions and their values range.

| Variable description   | Values (min–max)   | Normalized           |
|--|--|----------------------|
| Season in which the analysis was performed.                  | (1) winter, (2) spring, (3) Summer, (4) fall   | (−1, −0.33, 0.33, 1) |
| Age at the time of analysis.                                 | 18–36  | (0, 1)               |
| Body mass index (BMI)  | 17–34  | (0, 1)               |
| Marital status   | (1) Unmarried and without a partner, (2) unmarried with partner, (3) married, (4) other.   | (−1, −0.33, 0.33, 1) |
| Children   | (1) Yes, (2) no  | (0, 1)               |
| Birth city   | (1) Alacanti (2) marina baixa, (3) marina alta, (4) the Comtat, (5) l'Alcoy, (6) l'alt Vinalopo, (7) el vinalopo Mitj, (8) baix Vinalopo, (9) la vega baja, (10) other cities.   | (0, 1)               |
| City of the family home                                      | (1) Alacanti (2) marina baixa, (3) marina alta, (4) the Comtat, (5) l'Alcoy, (6) l'alt Vinalopo, (7) el vinalopo Mitj, (8) baix Vinalopo, (9) la vega baja, (10) other cities.   | (0, 1)               |
| City of residence during the course                          | (1) Alacanti (2) marina baixa, (3) marina alta, (4) the Comtat, (5) l'Alcoy, (6) l'alt Vinalopo, (7) el vinalopo Mitj, (8) baix Vinalopo, (9) la vega baja, (10) other counties. | (0, 1)               |
| Number of siblings.  | 0–6  | (0, 1)               |
| Birth defect, genetic disorder or any hereditary disease     | (1) Yes, (2) no  | (0, 1)               |
| Childish diseases (i.e., chicken pox, measles, mumps, polio) | (1) Yes, (2) no  | (0, 1)               |
| Vaccines received  | (1) Children calendar, (2) i Children calendar and tetanus, (3) only tetanus.  | (−1, 0, 1)           |
| Allergy  | (1) Yes, (2) no  | (0, 1)               |
| Accident or serious trauma                                   | (1) Yes, (2) no  | (0, 1)               |
| Surgical intervention  | (1) Yes, (2) no  | (0, 1)               |
| Significant diseases   | (1) Yes, (2) no  | (0, 1)               |
| High fevers in the last year                                 | (1) less than three months ago, (2) more than three months ago, (3) no   | (−1, 0, 1)           |
| Chemotherapy   | (1) Yes, (2) no  | (0, 1)               |
| Radiations (i.e., X-rays)                                    | (1) Yes, (2) no  | (0, 1)               |
| Time since radiation exposure                                | (1) Less than 3 months, (2) More than 3 months.  | (0, 1)               |
| Number of exposures  | 0–20   | (0, 1)               |
| Frequency of alcohol consumption                             | (1) several times a day, (2) every day, (3) several times a week, (4) once a week, (5) hardly ever or never  | (0, 1)               |
| Alcohol consumption  | Grams per week (0–312)   | (0, 1)               |
| Smoking habit  | (1) never, (2) occasional (3) daily.   | (−1, 0, 1)           |
| Average number of cigarettes per day                         | 0–25   | (0, 1)               |
| Years of exposure to cigarettes                              | 0–19   | (0, 1)               |
| Coffee consumption   | (1) none, (2) 1 or 2 coffees a day, (3) 3 to 5 coffees a day, (4) more than 5 coffees a day.   | (−1, −0.33, 0.33, 1) |
| Drugs consumption  | (1) Yes, (2) no  | (0, 1)               |
| Number of hours spent sitting per day                        | 1–16   | (0, 1)               |
| Average hours sleep per day                                  | 6–12   | (0, 1)               |
| Wearing tight clothes  | (1) Yes, (2) no  | (0, 1)               |
| Warm baths   | (1) Yes, (2) no  | (0, 1)               |
| Average hours per week playing sports                        | 0–40   | (0, 1)               |
| Average ejaculations per week                                | 1–14   | (0, 1)               |

circumstances, hinder analysis or interpretation of information (Aitkenhead, 2008; Tokatli, Kurt, & Ture, 2009).

The classification and regression trees (CART or C&RT) method of Breiman (1984) generates binary decision trees. In the real world, the chances of biased binary outcomes are few, but the binary method allows for easy interpretation and analysis (Prasher et al., 2003; Tokatli et al., 2009). Thus, the study uses the binary method in the DT to interpret the life habits in prediction of male fertility, and attempts to deduce a correlated prediction factor.

#### 2.4. ANN – multilayer perceptron

An MLP (Ripley, 1996; Haykin, 1998; Bishop, 2005) consists of multiple layers of neurons, generally three (where each layer fully connected to the next one): an input layer that receives external inputs, one hidden layer, and an output layer which generates the classification results (see Fig. 2).

Except for the input layer every neuron in the other layers is a computational element with a nonlinear activation function. The principle of the network is that when data are presented at the input layer, the network neurons run calculations in the consecutive layers until an output value is obtained at each of the output neurons. This output will indicate the appropriate class for the input data.

Each neuron (see Fig. 3) in the input and the hidden layers is connected to all neurons in the next layer by weighted connections. The neurons of the hidden layers (see Fig. 3) compute weighted sums of their inputs and adds a threshold. The resulting sums are used to calculate the activity of the neurons by applying a sigmoid activation function.

This process is defined as follows:

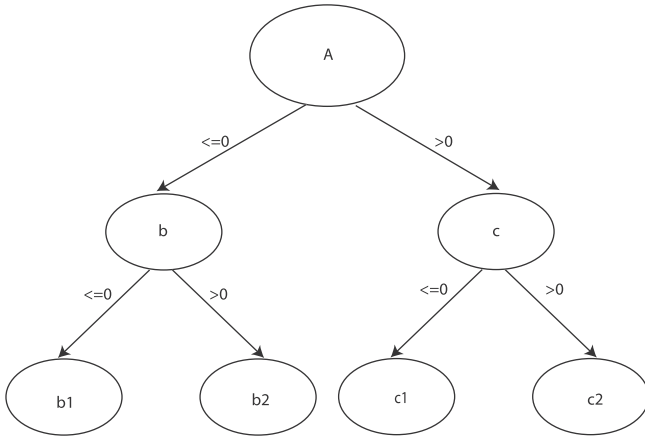
$$v_j = \sum_{i=1}^p w_{ji} x_i + \theta_j, \quad y_j = f_j(v_j), \quad (1)$$

where  $v_j$  is the linear combination of inputs  $x_1, x_2, \dots, x_p$ , and the threshold  $\theta_j$ ,  $w_{ji}$  is the connection weight between the input  $x_i$  and the neuron  $j$ , and  $f_j$  is the activation function of the  $j_{th}$  neuron, and  $y_j$  is the output. The sigmoid function is a common choice of activation function. It is defined as:

$$f(t) = \frac{1}{1 + e^{-t}}. \quad (2)$$

A single neuron in the MLP is able to linearly separate its input space into two subspaces by a hyperplane defined by the weights and the threshold. The weights define the direction of this hyperplane whereas the threshold term  $\theta_j$  offsets it from origo.

In order to train the MLP we use a supervised learning method denominated backpropagation (Rumelhart, Hinton, & Williams, 1986), which is a gradient descent method, for the adaptation of



**Fig. 1.** General scheme of a decision tree. This technique is very appreciated for the doctors as it follows similar rules of procedure.

the weights. This means the network is presented with input examples as well as the corresponding desired output. This algorithm runs as follows:

- The weight vectors  $w_i$  are updated by

$$w(t+1) = w(t) + \Delta w(t). \quad (3)$$

- Where

$$\Delta w(t) = -\eta \partial E(t) / \partial w. \quad (4)$$

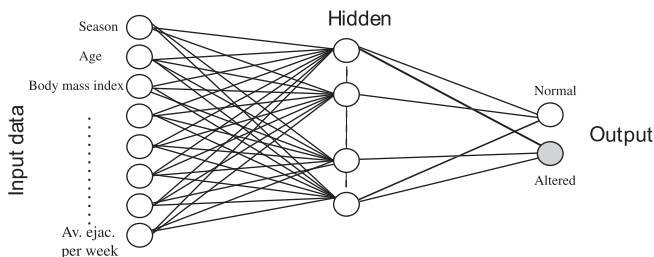
- Compute the error  $E(t+1)$ , where  $t$  is the iteration number,  $w$  is the weight vector, and  $\eta$  is the learning rate.

The backpropagation algorithm adapts the weights and the thresholds of the neurons in a way that minimizes the error function  $E$

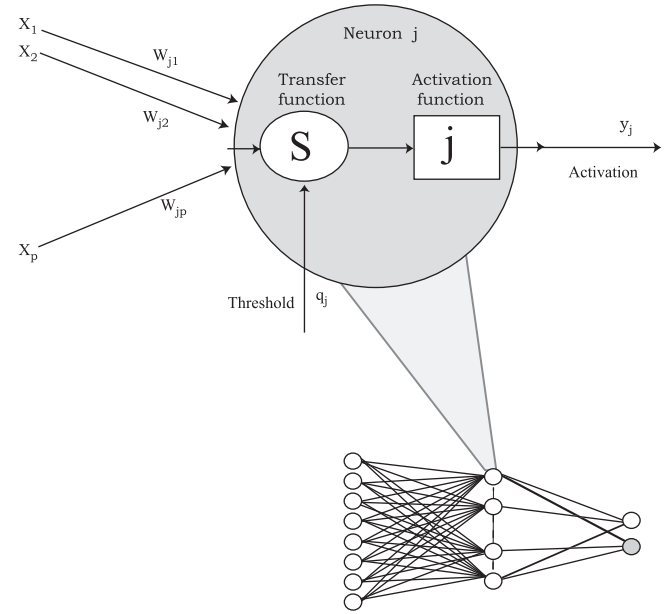
$$E = \frac{1}{2} \sum_{p=1}^n (d_p - y_p)^2, \quad (5)$$

where  $y_p$  is the actual output and  $d_p$  the desired output for input pattern  $p$ .

The minimization of  $E$  can be accomplished by gradient descent, i.e., the weights are adjusted to change the value of  $E$  in the direc-



**Fig. 2.** The architecture of the MLP network (input layer, hidden layer and output layer). The input layer represents the input data (the input data set is described in Section 3.1). The usage of a hidden layer enables the representation of data sets that are not linearly separable. The output layer represents the classification result and it contains as many outputs as the problem has classes, although here only one neuron is shown. The weights and the threshold of the MLP are calculated during an adaptation process.



**Fig. 3.** A neuron in the hidden or the output layer in the MLP. In the experimentation section the number of hidden neurons of the MLP will be established.

tion of its negative gradient. The exact updating rules can be calculated by applying derivatives and the chain rule (for the weights between the input and the hidden layer).

## 2.5. Support vector machines

In this section, the basic concept of SVM will be briefly described. More thorough descriptions can be found in (Burges, 1998; Hsu, Chang, & Lin, 2003; Theodoridis & Koutroumbas, 2008). A typical two class problem as Fig. 4 shows is similar to the problem of diagnosing urological patients as either ill or healthy.

For a two class classification problem, it is necessary to first try to estimate a function  $f: \mathfrak{R}^N \rightarrow \{\pm 1\}$  using training data, which are  $l$   $N$ -dimensional patterns  $x_i$  and class labels  $y_i$ , where

$$(x_1, y_1), \dots, (x_l, y_l) \in \mathfrak{R}^N \times \{\pm 1\} \quad (6)$$

such that  $f$  will classify new samples  $(x, y)$  correctly.

Given this classification problem the SVM classifier, as described by Vapnik (1995), Guyon, Boser, and Vapnik (1992) and Cortes and Vapnik (1995), satisfies the following conditions:

$$\begin{cases} \mathbf{w}^T \varphi(x_i) + b \geq +1 & \text{if } y_i = +1, \\ \mathbf{w}^T \varphi(x_i) + b \leq -1 & \text{if } y_i = -1, \end{cases} \quad (7)$$

which is equivalent to

$$y_i [\mathbf{w}^T \varphi(x_i) + b] \geq 1, \quad i = 1, 2, \dots, l. \quad (8)$$

Here training vectors  $x_i$  are mapped into a higher dimensional space by the function  $\varphi$ . The equations of (2) construct a hyperplane  $\mathbf{w}^T \varphi(x_i) + b = 0$  in this higher dimensional space that discriminates between the two classes showed in Fig. 4. Each of the two half-spaces defined by this hyperplane corresponds to one class,  $H_1$  for  $y_i = +1$  and  $H_2$  for  $y_i = -1$ . Therefore the SVM classifier corresponds to decision functions:

$$y(x) = \text{sign}[\mathbf{w}^T \varphi(x_i) + b] \quad (9)$$

Thus the SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. The margin of a linear classifier is the minimal distance of any training point to the hyperplane which is the distance between the dotted lines  $H_1$  and  $H_2$  and



the solid line showed in Fig. 4. The points  $x$  which lie on the solid line satisfy  $\mathbf{w}^T \varphi(x_i) + b = 0$ , where  $\mathbf{w}$  is normal to the hyperplane,  $|b|/\|\mathbf{w}\|$  is the perpendicular distance from the hyperplane to the origin, and  $\|\mathbf{w}\|$  is the Euclidean norm of  $\mathbf{w}$ .  $1/\|\mathbf{w}\|$  is the shortest distance from the separating hyperplane to the closest positive (negative) example. Therefore, the margin of a separating hyperplane will be  $1/\|\mathbf{w}\| + 1/\|\mathbf{w}\|$ . To calculate the optimal separating plane is equivalent to maximizing the separation margin or distance between the two dotted lines  $H_1$  and  $H_2$ .

It has to be considered that  $H_1: \mathbf{w}^T \varphi(x_i) + b = 1$  and  $H_2: \mathbf{w}^T \varphi(x_i) + b = -1$  are parallel since they have the same normal  $\mathbf{w}$  and perpendicular distance from the origin,  $|1 - b|/\|\mathbf{w}\|$  for  $H_1$  and  $|-1 - b|/\|\mathbf{w}\|$  for  $H_2$ , and that no training points fall between them. Thus we expect the solution for a typical two dimensional problem to have the form shown in Fig. 4. Those training points which gives equality in (3) are lying on one of the hyperplanes  $H_1$  and  $H_2$  are called support vectors and they are indicated in Fig. 4 by means of a gray color.

In practice, a separating hyperplane may not exist, i.e., if a high noise level causes a large overlap of the classes. To allow for the possibility of examples violating the edges of the margin (when perfect separation is not possible) (Martin, Fowlkes, & Malik, 2003), one introduces slack variables (Scholkopf, Smola, Williamson, & Bartlett, 2000; Cortes & Vapnik, 1995; Vapnik, 1995; Bennett & Mangasarian, 1992).

$$\xi_i \geq 0, \quad i = 1, 2, \dots, l, \quad (10)$$

in order to relax the constraints to

$$y_i(\mathbf{w}^T \varphi(x_i) + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, l. \quad (11)$$

Fig. 4 shows the basic idea of SVM, which is to map the data into some other dot product space (called the feature space)  $F$  via a non-linear map

$$\Phi: \mathcal{R}^N \rightarrow F \quad (12)$$

and perform the above linear algorithm in  $F$ . As can be seen, this only requires the evaluation of dot products.

$$K(x, x') = (\Phi(x) \cdot \Phi(x')). \quad (13)$$

Clearly, if  $F$  is high-dimensional, this approach is not feasible and we have to find a computationally cheaper way. The key observation then is to substitute it for a simple kernel  $K$  that can be evaluated efficiently. Kernel substitution is a method for using a linear classifier algorithm to solve a non-linear problem by mapping the original non-linear observations into a higher-dimensional space (as it is indicated in Fig. 4), where the linear classifier is subsequently used; this makes a linear classification in the new space equivalent to non-linear classification in the original space. That transformation is

done using Mercer's theorem (Williamson, Smola, & Scholkopf, 2001), which states that any continuous, symmetric, positive semi-definite kernel function  $K(x, x')$  can be expressed as a dot product in a high-dimensional space.

In that point the tools to construct nonlinear classifiers are defined. We substitute  $\Phi(x_i)$  for each training example  $x_i$ , and perform the optimal hyperplane algorithm in  $F$ . Because we are using kernels, we will thus end up with nonlinear decision function of the form

$$y(x) = \text{sign} \left[ \sum_{i=1}^l y_i K(x, x') + b \right]. \quad (14)$$

where  $K(x, x')$  is called the Kernel function, whose value is equal to the inner product of two vectors  $x$  and  $x'$  in the feature space; namely,  $K(x, x') = \varphi(x)^T \varphi(x')$ . Through the Kernel function, all computations are performed in the low-dimensional input space. Any function that satisfies the Mercers condition (Courant & Hilbert, 1953) can be used as the Kernel function.

There are many possibilities to define a function to be used as a Kernel. However, typical examples of kernels used in SVM, which have been successfully applied to a wide variety of applications, are linear, polynomials, radial basic functions and hyperbolic tangent:

$$\text{Linear kernel: } k(\mathbf{x}, \mathbf{x}') = \mathbf{x} \cdot \mathbf{x}' \quad (15)$$

$$\text{Polynomial kernel: } k(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}' + c)^d \quad (16)$$

$$\text{RBF kernel: } k(\mathbf{x}, \mathbf{x}') = \exp \left( -\frac{\gamma(\mathbf{x} - \mathbf{x}')^2}{\sigma} \right) \quad (17)$$

$$\text{Sigmoid kernel: } k(\mathbf{x}, \mathbf{x}') = \tanh(\gamma \mathbf{x} \cdot \mathbf{x}' + c) \quad (18)$$

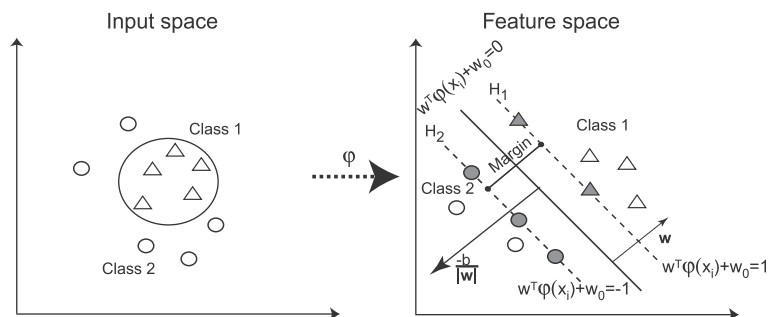
### 3. Experimentation

#### 3.1. Semen parameters

The semen analysis procedures are standardized by a WHO publication (WHO, 1999). This publication establishes the reference value for the different semen parameters and the nomenclature (Eliasson et al., 1970) that describes the deviations from these reference values.

This nomenclature includes:

- Normozoospermia: all values are over the lower limit.
- Asthenozoospermia: percentage of progressively motile spermatozoa below the lower reference limit.
- Oligozoospermia: total number of spermatozoa below the lower reference limit.



**Fig. 4.** The mapping between input space and feature space in a two class problem with the SVM. Mapping the training data non-linearly into a higher dimensional feature space via function  $\varphi$ .  $H_1$  and  $H_2$  are parallel since they have the same normal  $\mathbf{w}$  and perpendicular distance from the origin,  $|\pm 1 - b|/\|\mathbf{w}\|$ , and that no training points fall between them. The support vectors are the gray triangles and circles respectively located on  $H_1$  and  $H_2$ . The distance from  $\mathbf{w}$  to these support vectors is  $1/\|\mathbf{w}\|$  and the margin is simply  $2/\|\mathbf{w}\|$ .

- Teratozoospermia: percentage of morphologically normal spermatozoa below the lower reference limit.

Of all men surveyed, almost 38% had some alteration in sperm parameters. By studying in detail the individuals with altered parameters the most frequently found seminal alteration was asthenozoospermia in 47.37% of cases, followed by oligozoospermia in 21.05%. The combined changes astenoteratozoospermia and oligoasthenozoospermia shown a frequency of 10.53% in both cases. 7.89% of individuals with changes in semen parameters showed a oligoasthenoteratozoospermia. Finally, the less frequent alteration was teratozoospermia, present in 2.63% of the cases alterations. In this study we use these laboratory results of the seminal analysis as a control to evaluate the accuracy in the prediction of the diagnosis of the ANN developed. This ANN was only based in the answers to the questionnaire of the individuals who participate.

### 3.2. Development of the AI methods

#### 3.2.1. Decision tree

In this section we explain the algorithm C4.5, which is the one we have decided to use for the performance of the decision tree. C4.5 is an extension of the basic ID3 algorithm designed by Quinlan (Quinlan, 1986; Quinlan, 1993) to address some issues not dealt with by ID3, such as avoiding overfitting the data, handling training data with missing attribute values and improving computational efficiency among other less important.

The ID3 algorithm can be summarized as follows:

1. Take all unused attributes and count their entropy concerning test samples
2. Choose attribute for which entropy is minimum (or, equivalently, information gain is maximum)
3. Make node containing that attribute

The decision of choosing DT as the first method is due on one hand that it is the most useful technique for physicians that find on it a similar way that they use in their usual procedure. Furthermore, DT will help us to reduce the number of input data.

Research indicates that application of data dimensionality reduction as a pre-step to the classification procedure does improve the classification accuracy (Raymer, Punch, Goodman, Kuhn, & Jain, 2000; Fu & Wang, 2003). Furthermore, feature selection can also provide a better understanding of the underlying process that generated the data (Guyon & Elisseeff, 2003). It is beneficial, and also demanded in many cases, to limit the number of input features when building classification systems in order to have a good predictive and less computationally intensive model (Zhang, 2000).

Fig. 5 shows the execution carried out with a decision tree. This figure could help to establish what are the fields more important for the output. In fact, once the tree is obtained we have carried out other tests where the input data are reduced to only the fields that appear in the tree of the figure, that in our case are five.

#### 3.2.2. MLP

For the construction of the MLP architecture we have proceeded as follows:

- (a) Layer 1 corresponds directly to the input vector, that is, all the parameters/fields of the input record.
- (b) Layer 2 (the hidden layer). The number of hidden neurons for this layer is the most elaborated question in the network's architecture. This number represents a trade off between performance and the risk of over fitting. In fact, the number of neurons in a hidden layer will significantly

influence the ability of the network to generalize from the training data to unknown examples (Pal & University of Nottingham, 2002). In order to keep a good generalization and consequently a high accuracy, without getting over fitting, the experiments carried out showed that a low number of neurons for this layer lead to a poor performance for both training and test sets. At the opposite edge, a high number of neurons performs very well for training and test sets, although the risk of over fitting is high. The various tests carried out between these extreme options leads to meet the optimal solution for this layer with 21 neurons.

- (c) Layer 3 (the output layer) (Classification). It has two outputs for classification/prediction: normal and altered.

The learning algorithm used is backpropagation with adaptive learning rate, constant momentum and an optimized algorithm based on the gradient descent method. The backpropagation training parameters are shown in Table 2.

#### 3.2.3. SVM

In this study, Radial Basis Functions Kernel (RBF) function has been adopted because we believe that it is a suitable choice for our problem. The RBF kernel nonlinearly maps samples into a higher dimensional space. Thus, unlike the linear kernel, it can handle the case when the relation between class labels and attributes is nonlinear.

Furthermore, the linear kernel is a special case of RBF as (Keerthi & Lin, 2003) shows that the linear kernel with a penalty parameter  $C$  has the same performance as the RBF kernel with some parameters  $(C, \gamma)$ . In addition, the sigmoid kernel behaves like RBF for certain parameters (Lin & Lin, 2003). Moreover, the number of hyperparameters influences the complexity of model selection. The polynomial kernel has more hyperparameters than the RBF kernel. Finally, the RBF kernel has less numerical difficulties. One key point is  $0 < K_{ij} \leq 1$  in contrast to polynomial kernels of which kernel values may go to infinity ( $\gamma x_i^T x_j + r > 1$ ) or zero ( $\gamma x_i^T x_j + r < 1$ ) while the degree is large.

## 4. Results

### 4.1. Performance

Frequently, the complete data set is divided into two subsets: the training set and the test set. Here, the training set is used to determine the system parameters, and the test set is used to evaluate the diagnosis accuracy and the network generalization. Cross-validation has been widely used to assess the generalization of a network. The cross-validation estimate of accuracy determining by the overall number of correct classifications divided by the total number of examples in the dataset.

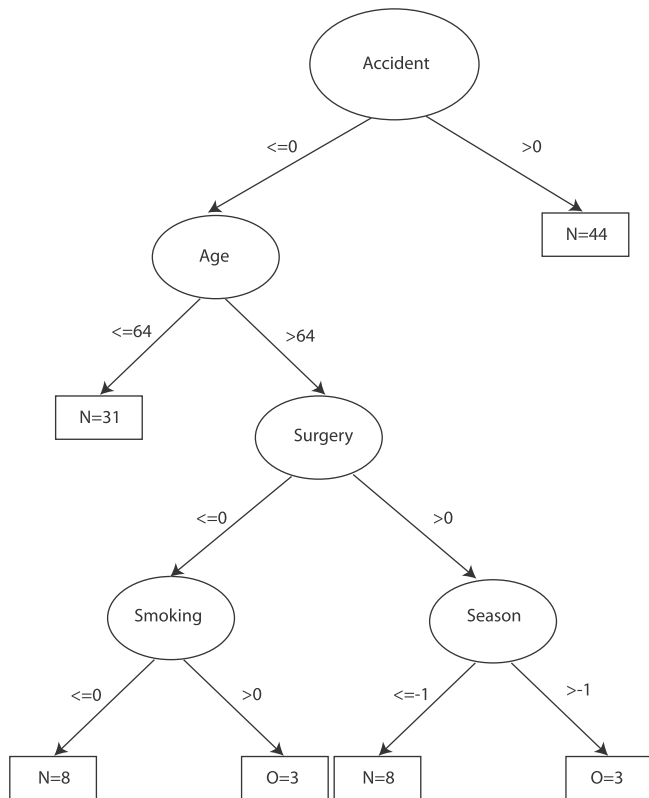
$$Acc_{cv} = \frac{1}{n} \sum_{x_i \in S} \delta(I(S_i, x_i), y_i), \quad (19)$$

where  $n$  is the size of the dataset  $S$ ,  $x_i$  is the example of  $S$ ,  $y_i$  is the target of  $x_i$ , and  $S_i$  is the probable target of  $x_i$  by the classifier function  $I$ . Therefore:

$$\delta(i, j) = \begin{cases} 1 & \text{if } i \in N_c(t), \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

Specifically, for this study we have applied a ten-fold cross-validation method for the performance assessment of every network. The data has been divided in ten sets ( $S_1, S_2, \dots, S_{10}$ ) and the ten experiments performed are shown in Fig. 6.

The test data are chosen randomly from the initial data and the remaining data form the training data. The method is called



**Fig. 5.** This figure shows the execution carried out with a decision tree. This figure could help to establish what are the fields more important for the output.

10-fold cross validation since this process has been performed ten times. The function approximation fits a function using the training set only. Then the function approximation is asked to predict the output values for the data in the testing set. The errors it makes are accumulated to provide the mean absolute test set error, which is used to evaluate the model.

The method to evaluate these four AI methods is to obtain some measures as classification accuracy, sensitivity, specificity, positive predictive value, negative predictive value and a confusion matrix. A confusion matrix (Kohavi & Provost, 1998) contains information about actual and predicted classifications done by a classification system.

The experimentation has been done three times in order to analyze the predictive value of each one of the seminal parameter studied. The reason to do that is that Sperm Concentration, Motility and Morphology are pathologies which have a different origin. Therefore, the individuals with those different alterations can not be grouped together. This will be exposed in the following three sections.

#### 4.1.1. Sperm concentration

Table 3 shows the confusion matrix provided for our classifiers, the MLP, SVM and DT running a two classes problem: Assisted reproduction database.

**Table 2**  
Backpropagation training parameters.

| Parameters                   | Value               |
|------------------------------|---------------------|
| Learning rate                | 0.01                |
| Adaptive learning rate       | 0.1                 |
| Constant momentum            | 0.2                 |
| Epochs                       | 100–1000–10000      |
| Minimum performance gradient | $\frac{1}{e^{-10}}$ |

In this case represents the values Normal and Altered, where altered means the Sperm concentration parameter. Classification accuracy, sensitivity, specificity, positive predictive value and negative predictive value (represented by the Table 4) can be defined by using the elements of the confusion matrix (Table 3). Every equation has three values according to the classifiers MLP, SVM and DT, respectively.

Sperm concentration is the seminal parameter that shows the highest prediction accuracy of the three parameter.

#### 4.1.2. Percentage of motile sperm

Table 5 shows the values Normal and Altered, where altered means the percentage of motile sperm parameter. Classification accuracy, sensitivity, specificity, positive predictive value and negative predictive value (represented by the Table 6) can be defined by using the elements of the confusion matrix (Table 5).

Although the accuracy of motility is slightly lower than concentration we can also predict it with a significant accuracy.

#### 4.1.3. Normal morphology

Table 7 shows the values Normal and Altered, where altered means the Normal morphology parameter. Classification accuracy, sensitivity, specificity, positive predictive value and negative predictive value (represented by the Table 8) can be defined by using the elements of the confusion matrix (Table 7).

Although the accuracy of motility is slightly lower than concentration we can also predict it with a significant accuracy.

Morphology seems to be less associated with environmental factors and life habits than the previous two seminal parameters.

In general, the values for specificity and Positive predictive value present lower accuracy. This happens even when the classification and the confusion matrix are good. The explanation is in the unbalanced classes. The altered output has always much less registers than the normal ones.

The excellent results provided in Table 3 by means of the confusion matrix as well as the equations of classification accuracy, sensitivity, specificity, positive predictive value and negative predictive value lead us to a convergence which is proposed in Fig. 7. This figure represents two options of diagnosis: laboratory and AI approach. This former one is easier (and especially more comfortable and cheaper) since it is only necessary to fill a questionnaire, besides it could be used as a first step to decide whether or not it is useful to do the laboratory approach.

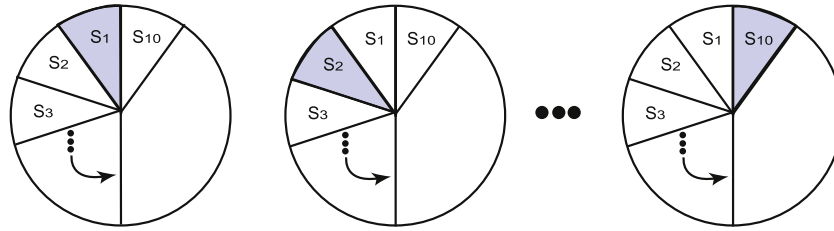
## 5. Discussion

In this paper, we have evaluated the performance of three AI methods, DT, MLP and SVM, and its application in the prediction of the male fertility potential. Based on the data obtained from 100 volunteers between 18 and 36 years we show the relationship of life habits and environmental factors with semen parameters.

The results have established that all these techniques achieve a high accuracy regarding the different measurement parameters. Specificity and positive predictive value present lower percentages, even when the classification and the confusion matrix are good.

The possible explanation of this is that the study population shows an imbalanced distribution (Sun, Kamel, Wong, & Wang, 2007), and there are a bigger number of individuals with normal semen parameters than the ones with alterations.

In general MLP and SVM are the methods that achieve the highest accuracy, being SVM the one with the higher Sensitivity whereas MLP obtains superior Specificity values. Therefore, these two methods seem to be useful tools for the predictions of the seminal parameters from environmental factors and life habits, especially because of their generalization potential with new



**Fig. 6.** Cross validation method. The cycle simulates the behavior of a wheel turning each of the portion at a time. In other words, for the first experiment we have S2 up to S10 as the training parts and S1 is the test part. This procedure will continue for 10 times; each of them it will permute the 10 different parts.

**Table 3**

Definition of the confusion matrix (for the Assisted reproduction dataset) with the values for every measure of the MLP, SVM and DT classifiers. The values of the confusion matrix are normal and altered, where altered means the Sperm concentration parameter.

| Actual | MLP<br>Predicted |   | SVM<br>Predicted |   | DT<br>Predicted |   |
|--------|------------------|---|------------------|---|-----------------|---|
|        | P                | N | P                | N | P               | N |
| P      | 80               | 5 | 83               | 2 | 82              | 3 |
| N      | 9                | 6 | 12               | 3 | 13              | 2 |

**Table 4**

Equations according to the MLP, SVM and DT classifiers.

|                               |  | MLP<br>(%) | SVM<br>(%) | DT<br>(%) |
|-------------------------------|--|------------|------------|-----------|
| Classification accuracy (%)   | $\frac{TP+TN}{TP+FP+FN+TN} \times 100 =$ | 86         | 86         | 84        |
| Sensitivity (%)               | $\frac{TP}{TP+FN} \times 100 =$          | 94.1       | 97.7       | 96.5      |
| Specificity (%)               | $\frac{TN}{FP+TN} \times 100 =$          | 40         | 20         | 13.3      |
| Positive predictive value (%) | $\frac{TP}{TP+FP} \times 100 =$          | 89.9       | 87.4       | 86.3      |
| Negative predictive value (%) | $\frac{TN}{FN+TN} \times 100 =$          | 54.5       | 60         | 40        |

registers. However, the slightly lower accuracy of DT is compensated by some advantages, such as to be simple (in a visual and illustrative way) to understand and interpret. In general, people are able to understand decision tree models after a brief explanation. It requires little data preparation. Other techniques often require data normalization. In addition, in this paper DT has been used also as a first step in data preprocessing in order to reduce the input data. After this phase, MLP as well as SVM have carried out tests with a lower number of input data.

ANN has been used, in the field of reproductive biology, to predict the results of IVF/ICSI (Kaufmann, Eastaugh, Snowden, Smye, & Sharma, 1997; Kshirsagar et al., 2005; Wald et al., 2005), to assess sperm morphology (Linneberg, Salamon, Svarer, Hansen, & Meyrowitsch, 1994) and to predict the presence of sperm in testes of men with nonobstructive azoospermia (Ma, Chen, Wang, Hu, & Huang, 2011; Samli & Dogan, 2004). Other previous studies, trying to identify the relationship between environmental factors and semen quality have shown several lacks and limitations. Most of those problems are related to the complexity of the information treated due to its diversity and heterogeneous data types. In our opinion, the problem is that previous works (Li et al., 2011) have approached the study using linear statistical techniques. In this paper, we develop the use of non linear techniques that may allow a better approach of the complexity of the problem (Almeida, 2002; Cleophas & Cleophas, 2009).

Due to the high accuracy obtained with the different classification methods, they seems to be an alternative to more expensive laboratory tests, at least during the initial moment of the fertility

**Table 5**

Definition of the confusion matrix (for the Assisted reproduction dataset) with the values for every measure of the MLP, SVM and DT classifiers. The values of the confusion matrix are normal and altered, where altered means percentage of motile sperm parameter.

| Actual | MLP<br>Predicted |    | SVM<br>Predicted |   | DT<br>Predicted |    |
|--------|------------------|----|------------------|---|-----------------|----|
|        | P                | N  | P                | N | P               | N  |
| P      | 60               | 11 | 70               | 1 | 58              | 13 |
| N      | 16               | 13 | 23               | 6 | 17              | 12 |

**Table 6**

Equations according to the MLP, SVM and DT classifiers.

|                               |  | MLP<br>(%) | SVM<br>(%) | DT<br>(%) |
|-------------------------------|--|------------|------------|-----------|
| Classification accuracy (%)   | $\frac{TP+TN}{TP+FP+FN+TN} \times 100 =$ | 73         | 76         | 70        |
| Sensitivity (%)               | $\frac{TP}{TP+FN} \times 100 =$          | 84.5       | 98.6       | 81.7      |
| Specificity (%)               | $\frac{TN}{FP+TN} \times 100 =$          | 44.8       | 20.7       | 41.4      |
| Positive predictive value (%) | $\frac{TP}{TP+FP} \times 100 =$          | 79         | 75.3       | 77.3      |
| Negative predictive value (%) | $\frac{TN}{FN+TN} \times 100 =$          | 54.2       | 85.7       | 48        |

**Table 7**

Definition of the confusion matrix (for the Assisted reproduction dataset) with the values for every measure of the MLP, SVM, and DT classifiers. The values of the confusion matrix are normal and altered, where altered means Normal morphology parameter.

| Actual | MLP<br>Predicted |    | SVM<br>Predicted |    | DT<br>Predicted |    |
|--------|------------------|----|------------------|----|-----------------|----|
|        | P                | N  | P                | N  | P               | N  |
| P      | 67               | 25 | 68               | 24 | 66              | 26 |
| N      | 6                | 2  | 7                | 1  | 7               | 1  |

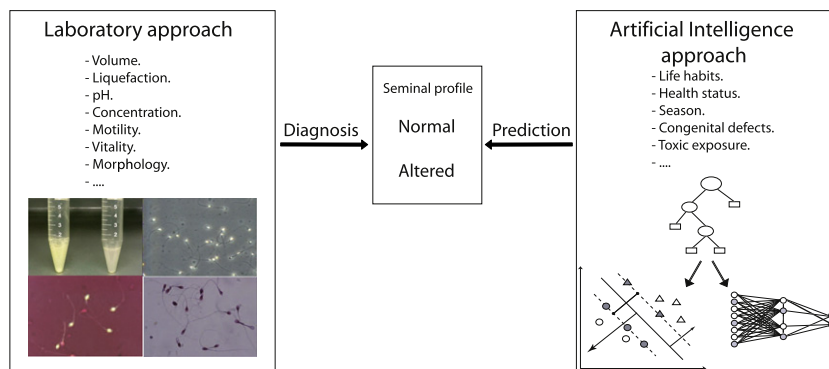
**Table 8**

Equations according to the MLP, SVM and DT classifiers.

|                               |  | MLP<br>(%) | SVM<br>(%) | DT<br>(%) |
|-------------------------------|--|------------|------------|-----------|
| Classification accuracy (%)   | $\frac{TP+TN}{TP+FP+FN+TN} \times 100 =$ | 69         | 69         | 67        |
| Sensitivity (%)               | $\frac{TP}{TP+FN} \times 100 =$          | 72.8       | 73.9       | 71.7      |
| Specificity (%)               | $\frac{TN}{FP+TN} \times 100 =$          | 25         | 12.5       | 12.5      |
| Positive predictive value (%) | $\frac{TP}{TP+FP} \times 100 =$          | 91.8       | 90.7       | 90.4      |
| Negative predictive value (%) | $\frac{TN}{FN+TN} \times 100 =$          | 7.4        | 4          | 3.7       |

study of the couple, as well as in the selection of semen donors. It is also important to highlight that the study population of this work presents close characteristics to regular candidates to become semen donors (i.e., young male, university students) (Thorn,





**Fig. 7.** This figure represents how it is possible to reach a diagnosis either with a laboratory approach, which include expensive tests, sometime uncomfortable for the patients, or with an AI approach, which is easier since it is only necessary to fill a questionnaire. Sometimes, this former approach could be a first step to decide whether or not is useful to do the laboratory approach.

et al., 2008). These AI methods can also help to make decision in the next steps of the infertility treatment, focusing or not on the male partner and avoiding painful and expensive examinations on the female.

The authors have experience in the field of urology by applying several AI techniques in classification tasks with very good results (Gil et al., 2009; Gil & Johnsson, 2010b; Gil et al., 2011). This makes it possible to adapt the artificial intelligence to meet the needs of the data analyzed in this study.

In future studies, we will investigate the effect of imbalanced classes on classification performance when developing AI methods, like MLP, SVM and DT, for decision support systems for medical diagnosis (Mazurowski et al., 2008). Moreover, we suggest the use of machine learning methods, as well as data mining, which may provide a more precise correlation between seminal data and the information collected by means of the questionnaires. The authors in association with a Reproductive Fertility Clinic are collecting more amount of data. For this purpose we will create data warehousing where will be easier to merge historical and current data. The benefit of using data warehousing is twofold. On one hand, it will improve the accuracy of our systems. On the other hand, the it will supports performance improvement and it will tend to equilibrate classes.

In conclusion, to our knowledge, this is the first time that MLP, SVM and DT have been compared addressing the problem of the relationship between life habits and semen quality. MLP and SVM show the highest prediction accuracy, with an accuracy of 86% for the sperm concentration and an accuracy between and 73% 76% for the motility. Although morphology seems to be less associated with environmental factors and life habits and probably with a major link to genetic alterations (Yatsenko et al., 2011) it also gets better accuracy with MLP and SVM. The clarity, reduction and simplification of the problem provided by DT must also be highlighted. We also suggest further studies using this methodology to identify those factors that affect specifically to some of the seminal parameters, to provide some recommendations that can improve the male fertility potential.

## Acknowledgment

We want to express our acknowledgement to the Ministry of Science and Innovation (Ministerio de Ciencia e Innovación - MIC-INN) through the “José Castillejo” program from Government of Spain as well as the Vicerrectorado de Innovación, University of Alicante, Spain (Vigrob-137) and to the Swedish Research Council through the Swedish Linnaeus project Cognition, Communication and Learning (CCL) as funders of the work exhibited in this paper.

## References

- Agarwal, A., Desai, N. R., Ruffoli, R., & Carpi, A. (2008).
- Aitkenhead, M. J. (2008). A co-evolving decision tree classification method. *Expert Systems with Applications*, 34(1), 18–25.
- Almeida, J. S. (2002). Predictive non-linear modeling of complex data by artificial neural networks. *Current Opinion in Biotechnology*, 13(1), 72–76.
- Auger, J., Kunstmann, J. M., Czyglik, F., & Jouannet, P. (1995). Decline in semen quality among fertile men in paris during the past 20 years. *New England Journal of Medicine*, 332(5), 281–285.
- Barratt, C. L., Clements, S., & Kessopoulou, E. (1998). Semen characteristics and fertility tests required for storage of spermatozoa. *Human Reproduction (Oxford, England)*, 13(Suppl. 2), 1–7. discussion 8–11.
- Bennett, K. P., & Mangasarian, O. L. (1992). Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, 1(1), 23–34.
- Bishop, C. M. (2005). *Neural networks for pattern recognition*. Oxford Univ Pr.
- Bonde, J. P., Ernst, E., Jensen, T. K., Hjollund, N. H., Kolstad, H., Henriksen, T. B., et al. (1998). Relation between semen quality and fertility: A population-based study of 430 first-pregnancy planners. *Lancet*, 352(9135), 1172–1177.
- Breiman, L. (1984). *Classification and regression trees*. Wadsworth International Group, 1984. ISBN 9780534980535.
- Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2), 121–167.
- Carlsen, E., Giwercman, A., Keiding, N., & Skakkebaek, N. E. (1992). Evidence for decreasing quality of semen during past 50 years. *BMJ*, 305(6854), 609–613.
- Carrell, D. T., Cartmill, D., Jones, K. P., Hatasaka, H. H., & Peterson, C. M. (2002). Prospective, randomized, blinded evaluation of donor semen quality provided by seven commercial sperm banks. *Fertility and Sterility*, 78(1), 16–21.
- Cleophas, T. J., & Cleophas, T. F. (2009). Artificial intelligence for diagnostic purposes: Principles, procedures and limitations. *Clinical Chemistry and Laboratory Medicine*, 48(2), 159–165.
- Conforti, D., & Guido, R. (2010). Kernel based support vector machine via semidefinite programming: Application to medical diagnosis. *Computers & Operations Research*, 37(8), 1389–1394.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- Courant, R., & Hilbert, D. (1953). *Methods of mathematical physics* (Vol. 1). Interscience Publishers.
- Ecochard, R., Cottinet, D., Mathieu, C., Rabilloud, M., & Czyba, J. C. (1999). The mean of sperm parameters in semen donations from the same donor. An important prognostic factor in insemination. *International Journal of Andrology*, 22(3), 163–172.
- Eliasson, R., Helinga, F., Lbcke, F., Meyhfer, W., Niermann, H., Steeno, O., et al. (1970). Empfehlungen zur nomenklatur in der andrologie. *Andrologia*, 2, 1257.
- Fu, X., & Wang, L. (2003). Data dimensionality reduction with application to simplifying RBF network structure and improving classification performance. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 33(3), 399–409.
- Gil, D., & Johnsson, M. (2010a). Supervised som based architecture versus multilayer perceptron and rbf networks (pp. 15–24).
- Gil, D., & Johnsson, M. (2010b). Using support vector machines in diagnoses of urological dysfunctions. *Expert Systems with Applications*, 37(6), 4713–4718.
- Gil, D., & Johnsson, M. (2011). Chapter 5 – Support vector machines in medical classification tasks (pp. 1–22).
- Gil, D., Johnsson, M., Garcia Chamizo, J. M., Paya, A. S., & Fernandez, D. R. (2009). Application of artificial neural networks in the diagnosis of urological dysfunctions. *Expert Systems with Applications*, 36(3P2), 5754–5760.
- Gil, D., Johnsson, M., Garcia Chamizo, J. M., Paya, A. S., & Fernandez, D. R. (2011). Review article: Modelling of urological dysfunctions with neurological etiology by means of their centres involved. *Applied Soft Computing*, 11(8), 4448–4457.

- Giwerzman, A., & Giwerzman, Y. L. (2011). Environmental factors and testicular function. *Best Practice & Research. Clinical Endocrinology & Metabolism*, 25(2), 391–402.
- Guyon, I., Boser, B., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proc. of the 5th annual workshop of computational learning theory* (pp. 144–152). ACM.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3, 1157–1182.
- Guzick, D. S., Overstreet, J. W., Factor-Litvak, P., Brazil, C. K., Nakajima, S. T., Coutifaris, C., et al. (2001). Sperm morphology, motility, and concentration in fertile and infertile men. *New England Journal of Medicine*, 345(19), 1388–1393.
- Haykin, S. (1998). *Neural networks: A comprehensive foundation*, Englewoods Cliffs. NJ: Prentice-Hall.
- Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification. Taipei, Taiwan: Department of Computer Science and Information Engineering, National Taiwan University, Taiwan.
- Irvine, D. S. (2000). Male reproductive health: Cause for concern? *Andrologia*, 32(4–5), 195–208.
- Kahramanli, H., & Allahverd, N. (2008). Design of a hybrid system for the diabetes and heart diseases. *Expert Systems with Applications*, 35(1–2), 82–89.
- Kaufmann, S. J., Eastaugh, J. L., Snowden, S., Smye, S. W., & Sharma, V. (1997). The application of neural networks in predicting the outcome of in-vitro fertilization. *Human Reproduction*, 12(7), 1454.
- Keel, B. A. (2006). Within-and between-subject variation in semen parameters in infertile men and normal semen donors. *Fertility and Sterility*, 85(1), 128–134.
- Keerthi, S. S., & Lin, C. J. (2003). Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation*, 15(7), 1667–1689.
- Kohavi, R., & Provost, F. (1998). Glossary of terms. *Machine Learning*, 30(2/3), 271–274.
- Kolettis, P. N. (2003). Evaluation of the subfertile man. *American Family Physician*, 67(10), 2165–2172.
- Kshirsagar, A. V., Murthy, L., Chelu, L., Lamb, D., Ross, L., & Niederberger, C. (2005). Predicting outcomes for intracytoplasmic sperm injection. *Fertility and Sterility*, 84, S274–S274.
- Li, Y., Lin, H., Li, Y., & Cao, J. (2011). Association between socio-psycho-behavioral factors and male semen quality: systematic review and meta-analyses. *Fertility and sterility*, 95(1), 116–123.
- Lin, H. T. & Lin, C. J. (2003). *A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods*. Taipei: Department of Computer Science and Information Engineering, National Taiwan University.
- Linneberg, C., Salamon, P., Svarer, C., Hansen, L. K., & Meyrowitsch, J. (1994). Towards semen quality assessment using neural networks. In *Neural networks for signal processing [1994] IV. Proceedings of the 1994 IEEE workshop* (pp. 509–517). IEEE.
- Lisboa, P. J., & Taktak, A. F. G. (2006). The use of artificial neural networks in decision support in cancer: A systematic review. *Neural Networks*, 19(4), 408–415.
- Ma, Y., Chen, B., Wang, H. X., Hu, K., & Huang, Y. R. (2011). Prediction of sperm retrieval in men with non-obstructive azoospermia using artificial neural networks: leptin is a good assistant diagnostic marker. *Human Reproduction*, 26(2), 294.
- Martin, D. R., Fowlkes, C. C., & Malik, J. (2003). Learning to detect natural image boundaries using brightness and texture. *Advances in neural information processing systems*, 1279–1286.
- Martini, A. C., Molina, R. I., Estofan, D., Senestrari, D., Fiore de Cuneo, M., & Ruiz, R. D. (2004). Effects of alcohol and cigarette consumption on human seminal quality. *Fertility and Sterility*, 82(2), 374–377.
- Mazurowski, M. A., Habas, P. A., Zurada, J. M., Lo, J. Y., Baker, J. A., & Tourassi, G. D. (2008). Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural Networks*, 21(2), 427–436.
- Pal, M., & University of Nottingham GB. (2002). *Factors influencing the accuracy of remote sensing classification: A comparative study*. University of Nottingham.
- Polat, K., & Gnes, S. (2009a). A new feature selection method on classification of medical datasets: Kernel f-score feature selection. *Expert Systems with Applications*, 36(7), 10367–10373.
- Polat, K., & Gnes, S. (2009b). A novel hybrid intelligent method based on c4.5 decision tree classifier and one-against-all approach for multi-class classification problems. *Expert Systems with Applications*, 36(2, Part 1), 1587–1592.
- Prasher, S. O., Enright, P., Madramootoo, C., Burgess, M., Goel, P. K., Callum, I., et al. (2003). Application of decision tree technology for image classification using remote sensing data. *Agricultural Systems*, 76(3), 1101–1117.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufman.
- Raymer, M. L., Punch, W. F., Goodman, E. D., Kuhn, L. A., & Jain, A. K. (2000). Dimensionality reduction using genetic algorithms. *IEEE Transactions on Evolutionary Computation*, 4(2), 164–171.
- Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge University Press.
- Risso, W. A., & Brida, J. G. (2010). Hierarchical structure of the german stock market. *Expert Systems with Applications*, 37(5), 3846–3852.
- Rowe, P. J., & Comhaire, F. H. (2000). *WHO manual for the standardized investigation, diagnosis and management of the infertile male*. Cambridge University Press.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.
- Samli, M. M., & Dogan, I. (2004). An artificial neural network for predicting the presence of spermatozoa in the testes of men with nonobstructive azoospermia. *The Journal of Urology*, 171(6), 2354–2357.
- Scholkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New support vector algorithms. *Neural Computation*, 12(5), 1207–1245.
- Slama, R., Eustache, F., Ducot, B., Jensen, T. K., Jorgensen, N., Horte, A., et al. (2002). Time to pregnancy and semen parameters: A cross-sectional study among fertile couples from four European cities. *Human Reproduction*, 17(2), 503–515.
- British Andrology Society guidelines for the screening of semen donors for donor insemination (1999). *Human Reproduction (Oxford, England)*, 14(7), 1823–1826.
- Springart, C., Frapsauce, C., Veau, S., Barthelemy, C., Royere, D., & Guerif, F. (2011). Semen variation in a population of fertile donors: Evaluation in a French centre over a 34-year period. *International Journal of Andrology*.
- Subashini, T. S., Ramalingam, V., & Palanivel, S. (2009). Breast mass classification based on cytological patterns using rbfn and svm. *Expert Systems with Applications*, 36(3, Part 1), 5284–5290.
- Sun, Y., Kamel, M. S., Wong, A. K. C., & Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition*, 40(12), 3358–3378.
- Swan, S. H., Elkin, E. P., & Fenster, L. (1997). Have sperm densities declined? A reanalysis of global trend data. *Environmental Health Perspectives*, 105(11), 1228–1232.
- Swan, S. H., Elkin, E. P., & Fenster, L. (2000). The question of declining sperm density revisited: An analysis of 101 studies published 1934–1996. *Environmental Health Perspectives*, 108(10), 961–966.
- Theodoridis, S., & Koutroumbas, K. (2008). *Pattern recognition*. 9781597492720. Elsevier/Academic Press.
- Tokatli, F., Kurt, I., & Ture, M. (2009). Using Kaplan–Meier analysis together with decision tree methods (c&rt, chaid, quest, c4.5 and id3) in determining recurrence-free survival of breast cancer patients. *Expert Systems with Applications*, 36(2, Part 1), 2017–2026.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York: Springer-Verlag.
- Wald, M., Sparks, A. E. T., Sandlow, J., Van-Voorhis, B., Syrop, C. H., & Niederberger, C. S. (2005). Computational models for prediction of ivf/icsi outcomes with surgically retrieved spermatozoa. *Reproductive Biomedicine Online*, 11(3), 325–331.
- WHO. (1999). *WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction* (4th ed.). Published on behalf of the World Health Organization by Cambridge University Press, Cambridge, UK.
- Williamson, R. C., Smola, A. J., & Scholkopf, B. (2001). Generalization performance of regularization networks and support vector machines via entropy numbers of compact operators. *IEEE Transactions on Information Theory*, 47(6), 2516–2532.
- Wong, W. Y., Zielhuis, G. A., Thomas, C. M., Merkus, H. M., & Steegers-Theunissen, R. P. (2003). New evidence of the influence of exogenous and endogenous factors on sperm count in man. *European Journal of Obstetrics Gynecology and Reproductive Biology*, 110(1), 49–54.
- Yatsenko, A. N., O'Neil, D. S., Roy, A., Arias-Mendoza, P. A., Chen, R., Murthy, L. J., et al. (2011). Association of mutations in the zona pellucida binding protein 1 (zbp1) gene with abnormal sperm head morphology in infertile men. *Molecular Human Reproduction*.
- Zhang, G. P. (2000). Neural networks for classification: A survey. *IEEE Transactions on Systems, Man and Cybernetics, Part C*, 30(4), 451–462.
- Zinaman, M. J., Brown, C. C., Selevan, S. G., & Clegg, E. D. (2000). Semen quality and human fertility: A prospective study with healthy couples. *Journal of Andrology*, 21(1), 145–153.