Carpal tunnel syndrome automatic classification: electromyography vs. ultrasound imaging

Maurizio Maravalle[1], Federica Ricca[2], Bruno Simeone[2], Vincenzo Spinelli[3]

[1]Università de L'Aquila, maurizio.maravalle@cc.univaq.it
[2]Sapienza, Università di Roma, federica.ricca@uniroma1.it
[3]Istat - Istituto Nazionale di Statistica vispinel@istat.it

Abstract

We investigate automatic classification procedures for the diagnosis of the Carpal Tunnel Syndrome, a disease frequently observed in occupational medicine. We apply different classification techniques to a medical data set of patients reporting the typical symptoms of this syndrome and exploit the predictive power of such data to classify subjects as "sick" or "healthy", according to the information obtained from the electromyography and the ultrasound imaging tests. Particular attention is paid to the "Box-Clustering" methodology which, among the tested techniques, is the most recent one. We show that all the automatic classification methods have a comparable diagnostic performance but, in some cases, Box-Clustering performs better than the others. Even if for the diagnosis of Carpal Tunnel Syndrome electromyography cannot be completely replaced by ultrasound imaging, our results show that ultrasound scan can be a valuable screening tool to detect the pathology.

Keywords: Automatic classification Box-Clustering Carpal Tunnel Syndrome Ultrasound imaging Electromyography.

1 Introduction

In recent years, the widespread use of computers and the fast improvement of technologies allowed the creation and the analysis of large data sets for problems in a variety of application areas, such as in industry, business, and many other real-life decision contexts. In particular, in medical applications, the increasing importance of data analysis is leading to the use of statistical and mathematical methods for the automatic diagnosis of a given syndrome. In the medical field, the main interest is in the study of the cause-effect relationships between a set of symptoms observed on a group of patients and a specific syndrome. The classical methods used for data analysis range from decision trees, neural networks to the more recent technique known as Logical Analysis of Data (LAD) which showed a very good performance when adopted for medical diagnosis applications [1, 2, 6, 9, 22]). Following the general principles of LAD - that was originally developed for the case of binary data - a new methodology called Box-Clustering (BC) was introduced in [15] based on the notion of "homogeneous boxes". In spite of its recent appearance, the experimental results obtained with BC have shown that it is a very competitive classification method (see, e.g., [29, 30]).

In this paper, we study the Carpal Tunnel Syndrome (CTS), for which the diagnosis is particularly important in occupational medicine [17, 20], since it is generally due to the overuse of the arms (especially the right one) and it affects muscles, tendons and nerves [10, 21, 23, 27]. CTS produces an external compression of the median nerve at the wrist level, the consequence being a reduction of the

nerve conduction capability. The clinical picture of the patient is usually taken as the "gold standard" for the diagnosis in the medical context (*clinical diagnosis*) [5, 31]. However, some relevant tests can be carried out to help the medical doctor in providing a correct diagnosis, such as electromyography which measures the nerve conduction directly, and ultrasound scan of the wrist which visualizes the modified configuration of the median nerve [10]. Traditionally, in the medical literature, the possibility of an automatic diagnosis for CTS always falls back upon the use electromyography, while the diagnostic power of the ultrasound test is not universally accepted. However, in the last decade, a renewed interest for the ultrasound scan arose in studies published both in radiological and neuromuscular journals, providing a revaluation of the utility of such test in the CTS diagnosis (for a detailed critical review, see [5]). In spite of the abundance of papers on this topic, the characteristics of the corresponding study designs - which may be related to the composition of the group of patients, the execution of the diagnostic tests, the adopted gold standard, the criteria to consider a test positive, etc. - are so different that a comparison of the results is quite difficult. Actually, in [5] it is pointed out that in the majority of such papers the study design is affected by serious methodological shortcomings and, thus, the authors propose a clear description of a firm methodology for the study of the diagnostic capability of ultrasound scan for CTS.

Following the indication given in [5], the objective of the present paper is two-fold. First of all, we want to investigate the effectiveness of the diagnosis of CTS based on the ultrasound scan versus the one performed via the electromyography [19]. The advantage here is to replace an invasive diagnostic tool with an easy, non-invasive and low cost one [24]. This point will be evaluated by applying different data analysis techniques, but special attention will be paid to BC. In fact, our second aim is to show that BC - which is employed here for the first time in an actual medical application - is a particularly valuable tool for medical diagnosis, since, besides its classification capability, it is also able to produce information about typical profiles characterizing patients affected by the syndrome, thus providing additional support for the medical doctors in their decisions. In fact, we will describe how the outcome of a BC procedure can be easily red and interpreted by a medical doctor, since the configuration of the "boxes" - which are the actual output of the procedure - has a natural interpretation as the intersection of intervals of values related to the symptoms variables. In this sense, BC merges together the advantages of a sophisticated combinatorial technique and the ease of interpretation of the results required by the non-technical experts.

The rest of the paper is organized as follows. In Section 2 we present the medical data set on which our experiments are performed. In Section 3 some details are provided about the BC methodology, while in Section 4 we illustrate the BC classification technique based on a given "system of boxes". In Section 5 we introduce our experimental plan and in Section 6 we apply a selection of classical classification techniques to predict the presence or absence of CTS in our set of subjects; the same data set is then processed by BC and its performance is compared with the previous experimental results. Besides this comparison, in our experiments we focus the attention on the analysis of the variables related to the ultrasound imaging and study the possibility of obtaining an efficient method for the correct classification of CTS patients relying only on the ultrasound variables. Finally, in Section 7, some conclusions are drawn and an emphasis is given to the particular advantages of using BC methodology in a medical context.

2 Input data

The study is based on a set of data collected in the period ranging from December 2003 to September 2004 on a sample of 102 subjects who were examined at their right wrist, performing in the same day both electromyography and ultrasound imaging. The analysis focuses on the right wrist injury which is the most frequently observed one. Thus, we consider the one-side classification variable called "Right Carpal Tunnel Syndrome" (and denoted by RCTS) for which a value equal to 0 corresponds to the ab-

sence of the injury at the right wrist, while it is equal to 1 when the syndrome is present at such wrist. The data set has 64 records with no (right) injury and 38 records with (right) injury.
Electromyography was performed twice on each patient[1] measuring two different variables, namely, the *Distal Latency of the Motor nerve* (or *Distal Motor Latency*, DML) and the *Nerve Conduction Velocity* (NCV). For these variables, a value is considered "at risk" if it is greater than or equal to 4.4 msec (time delay measured in milliseconds) for the former, and less than or equal to 41.5 msec for the latter. The electromyography variables (EMG) were labeled by NCVR and DMLR, respectively, where the "R" denotes that they refer to the right wrist.

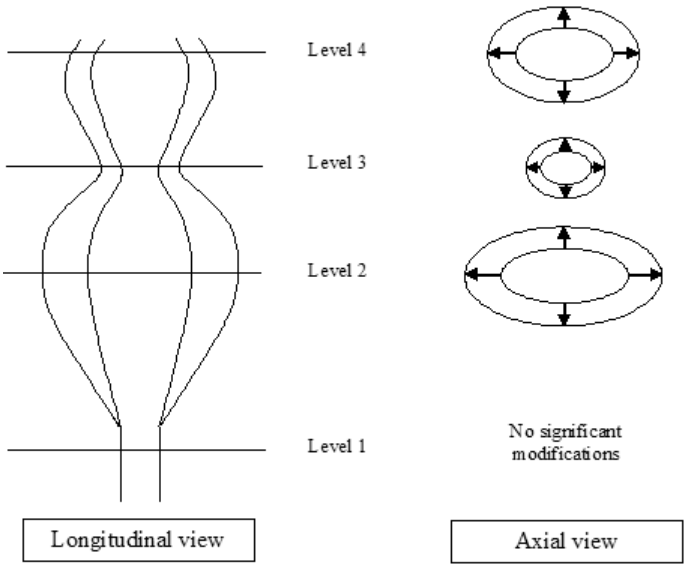

Figure 1: Oval shaped nerve sections at the four different levels.

The ultrasound graphic imaging was carried on, independently of the other examinations[2]: through a longitudinal scanning of the wrist (that locates the the track of the median nerve) it was possible to take the graphical imaging of the nerve at four specific axial levels. As shown in Figure 1, levels 4, 3 and 2 well represent the head, the neck and the body of the typical "hourglass" configuration of a nerve affected by CTS. However, the hourglass configuration alone is not generally considered a certificate of CTS, since, according to the medical literature, this configuration may be observed on a subject either if he/she suffers from CTS or if he/she does not [23, 25]. Hence, additional information was obtained related to the size of the hourglass nerve. Through a scanning orthogonal to the nerve axis it is possible to measure the sections of the median nerve. In particular, we computed the area of an oval shaped section at the four axial levels obtaining the values, denoted by R1, R2, R3 and R4, of the ultrasound scan variables (US) reported in Table 1 (see Figure 1).

On the basis of the values measured at levels 2 and 3, an *index of compression* (called CR2R3) was also computed as the percentage of the median nerve compression at level 3 with respect to the largest section

[1]The two examinations were performed independently by two different doctors working in different sites in Italy, one at the Department of Neurophysiopathology of the "*San Salvatore*" Hospital in the city of L'Aquila, and the other at the INAIL-Abruzzi Regional Polydiagnostic Center.

[2]It was performed by a third medical doctor under the protocol that he was neither informed about the diagnosis based on the other examination, nor about the specific clinics of the subject under study.

of the median nerve (the one at level 2):

$$CR2R3 = \frac{R2 - R3}{R2} \times 100. \tag{1}$$

Table 1: Full description of US and EMG variables and their coding.

US	Labels of the US variables
Level 1	R1
Level 2	R2
Level 3	R3
Level 4	R4
Compression Index	CR2R3
EMG	**Labels of the EMG variables**
Nerve Conduction Velocity	NCVR
Distal Motor Latency	DMLR

This means that we have two different data sets, one related to the two EMG variables, and another one given by the five US variables reported in Table 1.

3 Box-Clustering

In Logic Mining n observations are given in the form of d-dimensional vectors. Here we assume that the vectors are not constrained in type, so that they can be either binary or real, or of both types. In addition, a binary outcome variable, defining two separated classes, is associated to each of these vectors and, according to its value, the corresponding observation is termed *positive* (or *true*, or *sick* - in the medical terminology), or *negative* (or *false*, or *healthy*) [7, 8, 14]. A *box* in $\mathbb{R}^d$ is a multi-dimensional interval delimited by two *bounding points* $L = (l_1, \ldots, l_d), U = (u_1, \ldots, u_d)$ in $\mathbb{R}^d$:

$$I(L,U) = \{x \in \mathbb{R}^d : \ l_i \leq x_i \leq u_i, \ i = 1, \ldots, d\}. \tag{2}$$

A box is called *positive* (or *negative*) if it includes only positive (respectively, negative) observations. Positive and negative boxes are also called *homogeneous boxes*. For any finite set of points $S \in \mathbb{R}^d$ we define its *box-closure* $[S]$ as the intersection of all boxes containing the points in S. For two sets of points $S, T \in \mathbb{R}^d$, and their corresponding box-closures $B_S = [S]$ and $B_T = [T]$, the *join* $B_S \vee B_T$ is the box $[S \cup T]$.

The BC problem consists of generating a system of boxes satisfying a certain set of conditions related to some desired geometrical properties of the boxes. The system can be generated by formulating and solving a suitable optimization problem in which both the constraints and the objective function are defined on the basis of criteria related to the above properties (for details, see [29]). Let us denote by $\mathscr{B} = \{B_1, \ldots, B_m\}$ a system of m boxes in $\mathbb{R}^d$. In the following we recall the main geometrical properties for a system of boxes.

Given a set of observations P (for example the set of positive points in the training set), a system of boxes is a *coverage* of P if every observation of P is included in at least one box of $\mathscr{B}$ and every box in $\mathscr{B}$ includes at least one observation of P. In other words, $P \subseteq \bigcup_{i=1}^{m} B_i$ and $P \cap B_i \neq \emptyset$ for every $i = 1, \ldots, m$.

A system $\mathscr{B}$ is *homogeneous* if all the boxes are homogeneous, i.e., all the points inside a box are from the same class. It is *spanning* if every box is exactly the box-closure of the set of points included in it.

An additional criterion for evaluating a box system is to measure *overlapping*, that is, possible intersections between boxes in the system. In principle, overlapping boxes should be forbidden, but, in order to favor the other criteria, overlaps are generally allowed, provided that no point of the training set belongs to two boxes simultaneously. Finally, we define a system of boxes $\mathscr{B}$ *saturated* if the join of any two positive (negative) boxes of $\mathscr{B}$ contains some negative (positive) point.

From a learning point of view, given, for example the set P of positive points in the training set, we want to find the minimum number of homogeneous (positive) boxes covering all points in P. Since this is a hard problem [15], one falls back upon searching for a *spanning and saturated system of homogeneous boxes covering the points in the training set*. Even if it is not guaranteed that the number of boxes in the generated system is minimum, one can be sure that every training point is inside a box of its class. In [15] the above problem was solved by a simple agglomerative approach which generates boxes sequentially without any special rule for the best choice of the pair of boxes to be joined at each iteration. Here we adopt a clustering method based on a new class of graphs called "Incompatibility Graphs" (see, [28]). Given a supervised classification problem (P,N), where P and N are two finite and disjoint sets of positive and negative points in $\mathbb{R}^d$, respectively, we can define a graph G with vertex set $V(G) = P$ and such that two vertices $u,v \in P$ are connected by an edge iff $[\{u,v\}] \cap N \neq \emptyset$. G is called the *Incompatibility Graph* (IG) of (P,N) and it is denoted by $G_{P,N}$. Important relations exist between systems of positive boxes for (P,N) and stable sets in the corresponding IG $G_{P,N}$. As known, for any given graph G, a *stable set* is a subsets of pairwise non-adjacent vertices of G. In $\mathbb{R}^d$ the points of a homogeneous box correspond to a stable set in the related IG, but the reverse is not necessarily true. In fact, a stable set of the given $G_{P,N}$ may not correspond to a homogeneous box when $d > 2$ (see, [28]). Given a set of homogeneous boxes, one can extend the definition of IG to the more general case in which the vertices of the graph correspond to boxes and an edge exists between two boxes B_S and B_T iff $[S \cup T] \cap N \neq \emptyset$. In this case, the IG related to the set of boxes $\mathscr{B}$ is denoted by $G_{\mathscr{B},N}$. Notice that, $G_{\mathscr{B},N}$ coincides with $G_{P,N}$ in the particular case of boxes corresponding to single positive points (singletons). In Figure 2 we report the main steps of our clustering algorithm (ALGORITHM: Clustering) which, starting from a given set of boxes $\mathscr{B}$, tries to find a new one (with a smaller number of boxes) by merging together boxes corresponding to the vertices of a stable set of $G_{\mathscr{B},N}$. At step 4 the algorithm searches for a large cardinality stable set of the current $G_{\mathscr{B},N}$ denoted by $\hat{\mathscr{B}}$. Then, steps 7-12 try to join as many boxes as possible in $\hat{\mathscr{B}}$. During this loop, joined boxes are deleted from the system $\mathscr{B}$, which is then updated by adding the new obtained box B. In principle, one would search for a maximum cardinality stable set, but, due to the computational complexity for solving this problem, at step 4 a simple greedy heuristic procedure can be applied, such as, for example, the so called GMIN algorithm which consists of selecting a vertex of minimum degree, removing it and its neighbors from the graph, and iterating the process until no vertices remain (see, for example, [26]). this is a fast and simple method to get a maximal (but not necessarily maximum) stable set which usually delivers quite close to the maximum.

If the input box set $\mathscr{B}$ is saturated, the clustering algorithm returns the same system of boxes $\mathscr{B}$, otherwise it always provides a new updated system with a smaller number of boxes. The test is performed by checking if the current graph $G_{\mathscr{B},N}$ is complete, since, in this case, no further join is possible for the boxes in $\mathscr{B}$.

In Figure 3 we report the high level algorithm for the generation of the system of boxes based on incompatibility graphs (ALGORITHM: BC). The algorithm starts from the set of positive points, so that the first system of boxes is given by the singletons in P. The system of boxes is then updated by repeatedly calling (at step 5) the subroutine Clustering until the current box system is saturated.

The maximum number of iterations for the search loop is not greater than the number of points in the training set, and the last system generated by the above procedure is a saturated system of homogeneous (positive) boxes. The complete description of the algorithm would require details about the rule adopted for the selection of the two boxes to join at each iteration. This depends on the generation procedure

of the possible joins and from the criterion for ranking them and choose one. For details about this implementation, the reader is referred to [15, 29, 30].

ALGORITHM: Clustering$(\mathscr{B}, P, N)$

```
1  begin
2    Construct G_{B,N} related to the current set of boxes B
3    if G_{B,N} is not a complete graph
4      heuristically find a maximum stable set of G_{B,N}
5      let B̂ = {B̂_i} be such stable set
6      B := B̂_1
7      for i = (2, ..., |B̂|)
8        if (B ∨ B̂_i ∩ N = ∅)
9          B := B ∨ B̂_i
10         B := B − B̂_i
11       endif
12     endfor
13     B := B ∪ B
14   endif
15   return(B)
16 end
```

Figure 2: Clustering algorithm based on $G_{\mathscr{B},N}$.

ALGORITHM: BC(P, N)

```
1  begin
2    B_1 := P
3    do
4      B^+ := B_1
5      B_1 = Clustering(B^+, P, N)
6    while(B_1 ≠ B^+)
7    return(B^+)
8  end
```

Figure 3: Box Clustering algorithm.

It must be pointed out that for the same classification problem (P,N), the above algorithms can be adopted also to obtain a saturated system of negative boxes. Actually, in this case, the same analysis applies but the clustering procedure must rely on the incompatibility graph $G_{\mathscr{B},P}$ instead of on $G_{\mathscr{B},N}$.

4 Box-Clustering Classifier

The algorithm described in Section 3 provides solutions to the geometrical problem of clustering positive (or negative) points into a saturated system of homogeneous boxes, corresponding to a system $\mathscr{B}^+$

of positive boxes and a system $\mathscr{B}^-$ of negative ones. In this section we define our approach for the classification problem based on such system of boxes.

A BC-based classifier requires the following three inputs: (a) a set of positive boxes $\mathscr{B}^+$; (b) a set of negative boxes $\mathscr{B}^-$; (c) the point p to classify. A function $w(B,p)$ measuring the *attraction intensity* of the box B with respect to the point p must be also defined. The output of the classifier is the estimated class for the point p. For a given point p, the classifier computes all the weights $w_i^+ = w(B_i,p)$ and $w_j^+ = w(B_j,p)$ w.r.t. to all the positive boxes $B_i \in \mathscr{B}^+$ and negative boxes $B_j \in \mathscr{B}^-$, respectively, and assigns p to the class of the box corresponding to the minimum of such weights.

Figure 4 reports the steps of the classification procedure based on $\mathscr{B}^+$ and $\mathscr{B}^-$ (ALGORITHM: BC-classifier). If one of the tests performed at step 4 and 5 is successful, the point p can be univocally classified, but, if the minimum weight of p w.r.t. a positive box is equal to the one w.r.t a negative box (that is, $w^+ = w^- = \bar{w}$), an additional test is required and the classifier computes the number of positive and negative boxes with minimum weight w.r.t. p (denoted by n^+ and n^-, respectively). A tie for this check definitively means that p cannot be classified.

ALGORITHM: BC-classifier$(\mathscr{B}^+, \mathscr{B}^-, p)$

```
1  begin
2    let w+ = min{w(B,p) | B ∈ B+}
3    let w- = min{w(B,p) | B ∈ B-}
4    if(w+ < w-) return: p is positive
5    if(w+ > w-) return: p is negative
6    if(w+ = w- = w̄)
7      let n+ = |{B ∈ B+ | w(B,p) = w̄}|
8      let n- = |{B ∈ B- | w(B,p) = w̄}|
9      if(n+ > n-) return: p is positive
10     if(n+ < n-) return: p is negative
11   endif
12   return: p cannot be classified (classification failure)
13 end
```

Figure 4: A BC-based classifier related to box systems $\mathscr{B}^+$ and $\mathscr{B}^-$.

Different measures can be adopted to compute the weight $w(B,p)$ for a box B and a point p. If the box systems $\mathscr{B}^+$ and $\mathscr{B}^-$ provide a coverage of the whole observation space (both training and testing sets), one can naturally define $w(B,p)$ as the following characteristic function:

$$w(B,p) = \begin{cases} 0 & p \in B \\ 1 & otherwise \end{cases} . \tag{3}$$

If, on the contrary, there are some points in the observation space that are not covered by any box in $\mathscr{B}^+$ or in $\mathscr{B}^-$, a *distance* measure is required to define the attraction intensity between a non-covered point p and a box B (see, e.g., [22]). In our BC-classifier we considered a box as a continuous set of points and computed the distance between a box B and a point p as the Manhattan distance between p and the point in B closest to p.

5 The experimental plan

In order to empirically evaluate the efficiency of BC and its use in data analysis, we applied it to our CTS data set, and we compared its results with those provided by a wide range of other frequently-used classification methods. We considered several classes of methods, as defined in Weka software [13]: decision trees (*j48*), neural networks (*multilayer*), SVM (*smo*), bayes classifiers (*naive bayes*, *bayes net* and locally weighted naive bayes, denoted by *lwl*), regression (*regression*), and multinomial logistic regression (*logistic*). In addition, we implemented BC independently, since it is not still available in Weka or in other data analysis software packages.
The use of the simple test accuracy for the evaluation of the predictive power of a classification method is not a straight-forward choice. The literature proposes alternative and possibly more meaningful methods to evaluate the performance of a classifier. An example is the method based on the *ROC* (*Receiver Operating Characteristic*) plots which are frequently used in clinical medicine [32]; another one is the *AUC* (*Area Under Curve*) method, which is widely used to measure the model performance in binary classification problems [11, 16]. In particular, to evaluate a medical test, the concepts of "sensitivity" and "specificity" are often used; more generally, these concepts are readily usable for the evaluation of any binary classifier. They depend on the prevalence of the disease in the population of interest. To understand the utility of clinical tests, patients can be grouped into the following four classes, according to their clinical situation and their predictive value given by the test:

- *true positive* (A):
  the patient has the disease and the test is positive;
- *false positive* (B):
  the patient does not have the disease but the test is positive;
- *false negative* (C):
  the patient has the disease but the test is negative;
- *true negative* (D):
  the patient does not have the disease and the test is negative.

The *sensitivity* of a clinical test refers to the ability of the test to correctly identify those patients with the disease, while the *specificity* refers to the ability of the test to correctly identify those patients without the disease. Let n_A, n_B, n_C, and n_D denote the number of patients from the above defined classes, respectively. In order to measure sensitivity, one can compute the index TPR (*True Positive Rate*):

$$TPR = \frac{n_A}{n_A + n_C} \tag{4}$$

Similarly, specificity of a classifier can be measured through the index TNR (*True Negative Rate*):

$$TNR = \frac{n_D}{n_D + n_B}. \tag{5}$$

The relationship between sensitivity and specificity, as well as the performance of the classifier, can be visualized and studied using the graphical plot of the ROC space (π_{ROC}), where the *x* and *y* axes refer to FPR = (1 - TNR) and TPR , respectively. By definition, the points in π_{ROC} are inside the unit box $[0,1]^2$, and the four corner points have a special meaning:

- $P_{best} = (0,1)$ is the perfect classifier: it classifies all positive cases and negative cases correctly (TNR=1, TPR=1).

- $(0,0)$ represents a classifier that predicts all cases to be negative (TNR=1, TPR=0), while $(1,1)$ corresponds to a classifier that predicts every case to be positive (TNR=0, TPR=1).
- $P_{worst} = (1,0)$ is the classifier that provides an incorrect prediction for all classifications (TNR=0, TPR=0).

Our analysis is based on the assumption that all methods are required to have a 100% performance and no error is admitted. We point out here that, an alternative analysis could be performed if one would accept a percentage of errors different from 0. Actually, by accepting increments in the per cent number of errors, for example for the FPR index, for each given method one could obtain a set of performance points (corresponding to increasing percentages of tolerated errors) and plot a curve in π_{ROC} approximating the trend of such points; the method's performance could be then evaluated by measuring the area under this curve.

For every method we consider the corresponding point (FPR,TPR) and its distance to P_{best}, so that the "best methods" are those which minimize this distance. Notice that in our CTS application we do not have any information about possible different costs for errors of different type, such as errors of type B and C listed above. For this reason, we suppose here that the cost is given by a constant, but it must be pointed out that the distances in Table 2 and in Table 3 could be computed with a more general measure taking into account also a weighting function related to the cost.

In the following section, we exploit this graphical analysis to discuss our experimental results, and to present a synthetic comparison between the tested classifier based on the two data sets for CTS provided by the US and EMG variables.

6 Results

Cross-validation is a widespread strategy to perform model selection, because of its simplicity and its apparent universality, see [3]. Many results exist on the model selection performance of cross-validation procedures, see [12]. In our study, for the two sets of variables - US and EMG - we performed a 10-fold cross validation for each method mentioned in Section 5, and for BC as well. Tables 2 and 3 show the best results obtained for the two sets of variables.

Table 2: Performance of the classification methods based on the EMG variables.

classifier	n_A	n_B	n_C	n_D	FPR	TPR	distance to P_{best}	
j48	37	1	1	63	0.0156	0.9737	0.0306049	
multilayer	37	2	1	62	0.0312	0.9737	0.0408544	
smo	34	1	4	63	0.0156	0.8947	0.1064170	
lwl	37	0	1	64	0.0000	0.9737	0.0263158	$*$
naive bayes	36	3	2	61	0.0469	0.9474	0.0704794	
bayes net	37	0	1	64	0.0000	0.9737	0.0263158	$*$
regression	37	0	1	64	0.0000	0.9737	0.0263158	$*$
logistic	36	2	2	62	0.0312	0.9474	0.0612098	
BC	37	0	1	64	0.0000	0.9737	0.0263158	$*$

The methods that provide the best results are marked by an asterisk in the last column of the table. A general good performance can be observed for all methods with both data sets, since basically only few observations are misclassified. The performance is particularly good when the EMG variables are considered: in this case, the values of n_B and n_C are very small and several 0 and 1 are observed. It must be also noticed, however, that the number of false negative predictions - which is the most serious error in medical applications - is greater than the number of false positive ones. Even if the behavior of the tested classification methods is similar, a general increase of the number of this type of errors is observed when the US variables are adopted instead of the EMG ones. This result confirms for CTS the already known superior diagnostic power of the nerve conduction study over the ultrasound imaging.

Table 3: Performance of the classification methods based on the US variables.

classifier	n_A	n_B	n_C	n_D	FPR	TPR	distance to P_{best}	
j48	34	5	4	59	0.0781	0.8947	0.1310870	
multilayer	34	4	4	60	0.0625	0.8947	0.1224200	
smo	35	4	3	60	0.0625	0.9211	0.1006920	
lwl	37	10	1	54	0.1562	0.9737	0.1584510	
naive bayes	35	5	3	59	0.0781	0.9211	0.1110680	
bayes net	29	5	9	59	0.0781	0.7632	0.2493950	
regression	34	4	4	60	0.0625	0.8947	0.1224200	
logistic	35	3	3	61	0.0469	0.9211	0.0918148	
BC	38	1	0	63	0.0156	1.0000	0.0156000	$*$

Nevertheless, it must be pointed out that in our experiments the behavior of BC seems to be not affected by which set of variables is used. By looking at the simple numbers n_B and n_C - which for BC are very small in both tables - one may check that BC turns out to be as good as the other methods with the EMG variables and clearly better with the US ones. Thus, since n_B and n_C are monotone decreasing w.r.t. the increasing performance of a classification method, no matter which evaluation measure one uses, for our CTS data set BC would always turn out to be the one which works better.

According to these promising results, BC seems to be the right learning method for the CTS automatic diagnosis, since it showed to be able to exploit well the information lying in the EMG variables, as well as the one in the US variables. From our results, one may even guess that the US variables carry as much information as EMG ones, thus suggesting a completely different and new approach in the specific medical context of automatic diagnosis of CTS.

All the analyzed classification methods are comparable also under a computational viewpoint: all of them showed to be really fast for the real-life data set under study and they all delivered the final output almost instantly. In fact, the running times are within few seconds for all the methods provided by Weka (Java-based), while BC takes just a fraction of a second for each run.

To have a synthetic picture of the performance of all the tested methods, in Figure 5 and Figure 6 we show the points of coordinates FPR and TPR corresponding to each method drawn in π_{ROC}. The pictures show the location of each classifier-point in the plane and its distance to the reference point P_{best}, emphasizing the "*good position*" of BC w.r.t. the other methods. In Figure 5 many methods (including BC) provide the same best result, i.e., they correspond to the point $(0, 0.9737)$ and, therefore, all of them are represented together by a single grey point. The method *lwl* provided very poor results with the US variables, and, thus, we did not plot it in Figure 6.

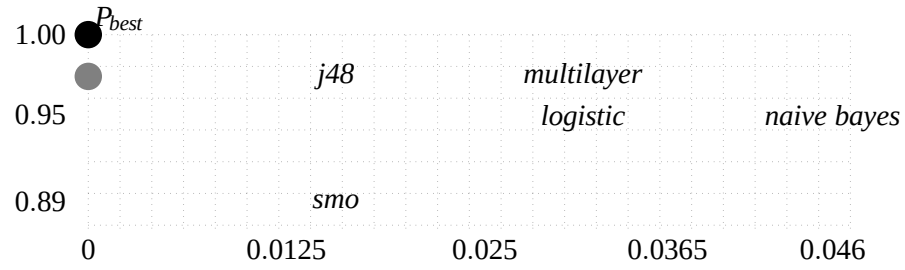

Figure 5: Representation in π_{ROC} of the classification methods' performances: EMG results. The methods *lwl*, *bayes net*, *regression*, and BC have the same performance and, thus, they are all represented by a common grey point.

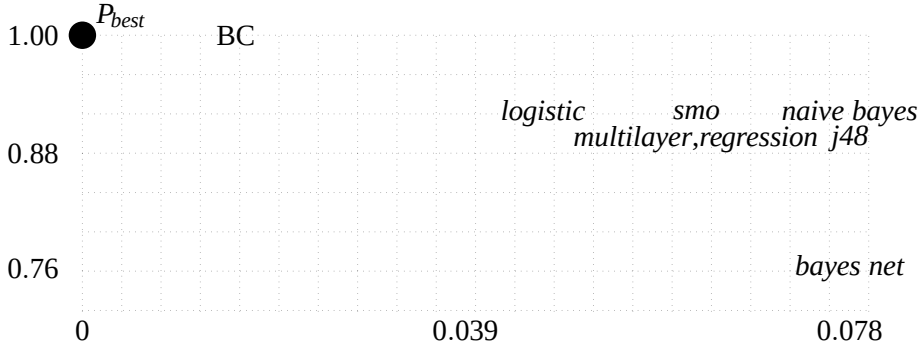

Figure 6: Representation in π_{ROC} of the classification methods' performances: US results. The performance of *lwl* was too bad w.r.t. FPR to be plotted.

In Table 4 we can visualize the box system that we used for the classification problem with the EMG approach: B_1^{EMG} is the negative box (healthy patients) and B_2^{EMG} is the positive one (sick patients). For the US analysis we adopted the box set shown in Table 5, where B_1^{US} is a negative box, while B_2^{US} and B_3^{US} are the positive ones.

Table 4: Box system based on the EMG variables: B_1^{EMG} is a negative box, B_2^{EMG} is positive.

	B_1^{EMG}		B_2^{EMG}	
NCVR	39.50	68.00	39.60	54.30
DMLR	2.30	4.20	4.40	6.80

Table 5: Box system based on the US variables: B_1^{US} is a negative box, while B_2^{US} and B_3^{US} are positive.

	B_1^{US}		B_2^{US}		B_3^{US}	
R1	1.03	2.50	1.50	2.78	2.03	2.51
R2	2.35	14.10	4.06	16.70	5.93	10.99
R3	2.76	15.19	2.43	9.87	4.52	9.39
R4	1.93	7.08	3.62	15.32	5.77	9.74
CR2R3	−19.80	31.80	32.00	56.80	14.50	23.80

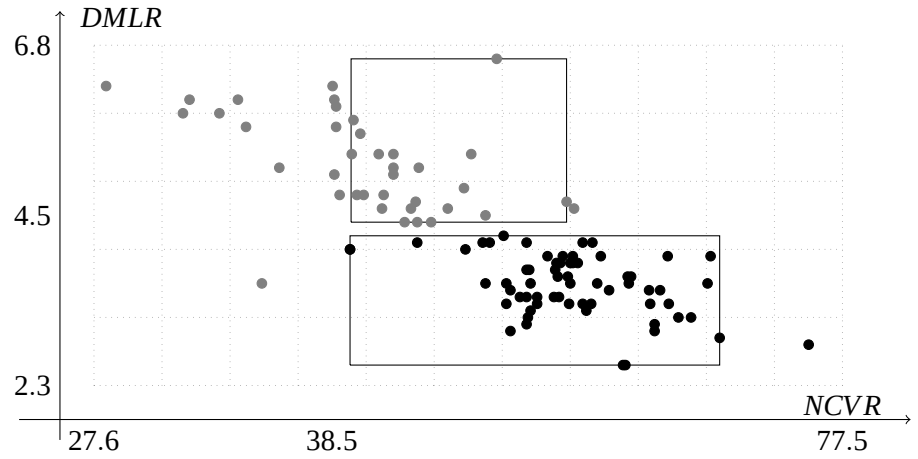

Figure 7: Box system for EMG variables: negative points are reported in black, positive ones in grey.

Under a methodological viewpoint, our results seem to be very promising. First of all, it must be pointed out that all points in our test set were classified, meaning that our implementation of the BC-classifier never returns a failure. Moreover, only one error was made in each application of the method (see, tables 2 and 3). In addition, both box systems showed in Table 4 and Table 5 have a very low number of boxes (two and three, respectively), thus suggesting that our heuristic solution for the BC geometrical problem is close to the optimum. In particular, for the analysis based on EMG variables we have exactly two

boxes, one positive and one negative and this is the optimal number of boxes that one can get. For this case, since we have only two variables, we can provide an explicit geometrical representation in $\mathbb{R}^2$ of the two boxes B_1^{EMG} and B_2^{EMG}, together with the positive and negative points of our data set. In Figure 7 black and grey points correspond to negative and positive points, respectively, and it is easy to see that the two boxes of the system are well separated. The picture also shows that - as we already pointed out - when the EMG variables are adopted, a "natural" separation of the points of the data set arises.

The experimental results obtained with BC become more relevant if one considers the advantages of this methodology for the readability and interpretation of the results by a medical doctor. If, on the one hand, BC is rigidly constrained by the geometrical shape of the boxes, on the other hand, the intervals of values provided by BC for the variables are easily interpretable by the clinician in terms of cut-off values. They offer viable explanations of the syndrome presence or absence, according to the specific profiles observed for subjects that are classified in the same box. In this sense, each box provides a specific set of useful indications in the diagnostic process. The number of boxes, which is also provided by the procedure, gives additional information in the recognition of the pathology profiles. The EMG system reported in Figure 7 provides a clear example of the typical geometrical structure of a "good" system of boxes. The plot refers to the system described in Table 4 and includes all points in the data set. The points in the training set are obviously included in their own homogeneous box; there are some points of the test set which are inside a box and some others which are not, but their classification is always unambiguous, since it is clear which box is the closest for each of them. From the picture one can also see the ease of interpretation of the boxes. In fact, in this simple case, one can realize that the DMLR value is crucial for the diagnosis of CTS, while NCVR alone could be not sufficient for this task. Actually, the combination of high values of DMLR with low values of NCVR provides the typical profile of a sick patient, while the opposite (low values of DMLR with high values of NCVR) seems to correspond to healthy patients.

We do not go beyond this kind of analysis for our BC results, rather we leave the final judgement to medical doctors who are the only ones who may evaluate wether the good technical performance of BC can be confirmed also under a medical point of view.

7 Conclusions

In the present paper we dealt with the automatic diagnosis of CTS with the two-fold aim of showing the high performance of statistical techniques for this specific type of diagnosis and assessing that ultrasound imaging is a useful tool in managing the CTS. Even if for the diagnosis of CTS electromyography cannot be completely replaced by ultrasound imaging, our results show that ultrasound scan can be a valuable screening tool to detect the pathology. All the applied methodologies showed a good and comparable prediction power, but BC seems to provide the best performance. As expected, our results confirmed the high reliability of EMG testing, but they also showed that US can be very informative, especially if the BC-based classifier is used. This result is particularly important for the CTS diagnosis at the early "irritative" stage of the pathology when often even neurophysiological tests are unable to detect the presence of the syndrome.

Given that US can be a valuable diagnostic tool for CTS, at least in a screening stage, one should not underestimate its many advantages, namely,

- low cost;
- ease of repetition;

- portability on the workplace;
- pain-free and non-invasive nature (implying easy acceptance by the patients);
- detection of possible underlying anomalies;
- when CTS is excluded, detection of other possible causes of hand pain.

The above aspects become much more relevant if one considers the ultrasound imaging as a screening tool for the diagnosis of CTS in occupational medicine, which was the original motivation of our work. For example, in large firms, it is recommended to repeat the test periodically on all the employees, in order to detect the pathology when it is still at an early stage. Then, on the basis of the result of this test, the - more expensive and more invasive - electromyographic examination can be performed only on those who are already suspected of suffering from CTS. The last two features are particularly important, since they provide additional information which can be useful in view of possible later surgical operations for CTS, and in case of a different diagnosis, respectively.

The promising results obtained in this experimental work for the real-life application related to CTS suggest that BC could be a valuable tool for the medical diagnosis and encourage the study of this methodology: further developments of the technique and additional empirical applications to medical data sets will be crucial for a definitive assessment of this methodology in the future.

References

[1] G. Alexe, S. Alexe, D.E. Axelrod, T.O. Bonates, I. Lozina, M. Reiss, P.L. Hammer, Breast cancer prognosis by combinatorial analysis of gene expression data, *Breast Cancer Res.*, 8: 1–20 (2006).

[2] S. Alexe, E.H. Blackstone, P.L. Hammer, H. Ishwaran, M.S. Lauer, C.E.P. Snader, Coronary risk prediction by Logical Analysis of Data, *Annals of Operations Research*, 119: 15–42 (2003).

[3] S. Arlot and A. Celisse, A survey of cross-validation procedures for model selection, *Statistics Surveys*, 4: 40–79, (2010).

[4] E. Balas, C.S. Yu, On graphs with polynomially solvable maximal-weight clique problem, *Networks*, 19: 247–253 (1989).

[5] R. Beekman, L.H. Visser, Sonography in the diagnosis of carpal tunnel syndrome: a critical review of the literature, *Muscle Nerve* 27: 26–33 (2003).

[6] T. Bonates, P.L. Hammer, Logical Analysis of Data: from combinatorial optimization to medical applications, *Annals of Operations Research*, 148: 203–225 (2006).

[7] E. Boros, P.L. Hammer, T. Ibaraki, A. Kogan, E. Mayoraz, and I. Muchnik, An implementation of logical analysis of data, *IEEE Transactions on Knowledge and Data Engineering*, 12: 292–306 (2000).

[8] E. Boros, T. Ibaraki, L. Shi, M. Yagiura, Generating all good patterns in polynomial expected time, *Lecture at the 6th International Symposium on Artificial Intelligence and Mathematics*, Ft. Lauderdale, Florida, January 2000.

[9] E. Boros, P.L. Hammer, T. Ibaraki, A. Kogan, A logical analysis of numerical data, *Mathematical Programming*, 79: 163 – –190 (1997).

[10] F.P. Cantatore, F. Dell'Accio, G. Lapadula, Carpal tunnel syndrome: a review, *Clinical Rheumatology*, 16: 596–603 (1997).

[11] S. Wu, P. Flach, A scored AUC Metric for Classifier Evaluation and Selection, *Second Workshop on ROC Analysis in ML*, Bonn, Germany, August 11, 2005.

[12] R. Kohavi, A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, IJCAI 95, Montrèal, Quèbec, Canada, August 20-25, 1995.

[13] M. Hall, F. Eibe, G. Holmes, B. Pfahringer, P. Reutemann and I.H. Witten. The WEKA Data Mining Software: An Update, *SIGKDD Explorations*, 11 (2009).

[14] P.L. Hammer, A. Kogan, B. Simeone, S. Szedmák, *Pareto-optimal patterns in logical analysis of data*, RUTCOR Research Report, RRR 7-2001.

[15] P.L. Hammer, Y. Liu, B. Simeone, S. Szedmák, Saturated systems of homogeneous boxes and the logical analysis of numerical data. *Discrete Applied Mathematics*, 144: 103–109 (2004).

[16] J. Huang and C.X. Ling, Using AUC and Accuracy in Evaluating Learning Algorithms, *IEEE Transactions on Knowledge and Data Engineering*, 17: 299–310, (2005).

[17] L. Isolani, R. Bonfiglioli, G.B. Raffi, F.S. Violante, Different case definitions to describe the prevalence of occupational carpal tunnel syndrome in meat industry workers, *International Archives of Occupational and Environmental Health*, 75: 229–234 (2002).

[18] D.S. Johnson, M. Yannakakis, C.H. Papadimitriou, On generating all maximal independent sets, *Information Processing Letters*, 27: 119–123 (1988).

[19] D. Lee, M.T. van Holsbeeck, P.K. Janevski, D.L. Ganos, D.M. Ditmars, V.B. Darian, Diagnosis of carpal tunnel syndrome: ultrasound versus electromyography, *Radiologic Clinics of North America*, 37: 859–872 (1999).

[20] C.T. Leffler, S.N. Gozani, Z.Q. Nguyen, D. Cros, An automated electrodiagnostic technique for detection of carpal tunnel syndrome, *Neurology and Clinical Neurophysiology*, 3: 2–10 (2000).

[21] R.T. Manktelow, P. Binhammer, L.R. Tomat, V. Bril, J.P. Szalai, Carpal tunnel syndrome: cross-sectional and outcome study in Ontario workers, *Journal of Hand Surgery*, 29: 307–317 (2004).

[22] T. Mitchell (1997), *Machine learning, McGraw-Hill*, NewYork (1997).

[23] C.A. Pacek, J. Tang, R.J. Goitz, R.A. Kaufmann, Zong-Ming Li, Morphological analysis of the carpal tunnel, *Hand*, 5: 77–81 (2010).

[24] D.H. Palmer, L.P. Hanrahan, Social and economic costs of carpal tunnel surgery, *Instr, Course Lect.* 44: 167–172 (1995).

[25] G.D. Papanicolau, S.J. McCabe, J. Firrell, The prevalence and characteristics of nerve compression symptoms in the general population, *Journal of Hand Surgery*, 12: 712–717 (1987).

[26] S. Sakai, M. Togasaki, K. Yamazaki, A note on greedy algorithms for the maximum independent set problem, *Discrete Applied mathematics*, 126: 313–322 (2003).

[27] B.A. Silverstein, L.J. Fine, T.J. Armstrong, Occupational factors and carpal tunnel syndrome, *Americal Journal of Industrial Medicine*, 11: 343–358 (1987).

[28] B. Simeone, E. Boros, F. Ricca, V. Spinelli Incompatibility graphs in data mining, *Technical Report*, Department of Statistical sciences, n. 10/2011.

[29] B. Simeone, V. Spinelli, The optimization problem framework for box clustering approach in logic mining, *Proceedings of Euro XXII* - 22*nd European Conference on Operational Research, Prague, July* 2007.

[30] B. Simeone, G. Felici, and V. Spinelli, A graph coloring approach for box clustering techniques in logic mining, *Proceedings of Euro XXII* - 22*nd European Conference on Operational Research, Prague, July* 2007.

[31] W.A. Swen, J.W. Jacobs, F.E. Bussemaker, J.W. de Waard, W. Bijlsma, Carpal tunnel sonography by the rheumatologist versus nerve conduction study by the neurologist, *Journal of Rheumatology*, 28: 62–69, (2001).

[32] M.H. Zweig, G. Campbell, Receiver-Operating Characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, *Clinical Chemistry*, 39: 561–577, (1993).