Comparative evaluation of electromyography and ultrasound for carpal tunnel diagnosis via automatic classification techniques

Stefano Colucci, Maurizio Maravalle
Facoltà di Economia
Università di L'Aquila

Franco Gioia
Medico del Lavoro – Medical Plans s.r.l. Roma.

Federica Ricca, Bruno Simeone, Vincenzo Spinelli
Dip. Statistica, Probabilità e Statistiche Applicate
Sapienza, Università di Roma

In memory of Peter Hammer

Abstract

In this paper we study two methodologies for the automatic diagnosis of Carpal Tunnel Syndrome: Discriminant Analysis and Box-Clustering. The former is a traditional statistical tool, while the latter is a recent technique. The study is carried on an actual medical dataset that was collected in Italy in 2003-2004. Our first aim is to exploit the predictive power of these techniques in order to obtain a correct classification of a subject as affected by the syndrome or not, according to the data obtained from the electromyographic and the ultrasound tests. We are also interested in studying the possibility of obtaining accurate predictions by using only the information provided by the ultrasound imaging examination. The main question is to understand whether – and to what extent – ultrasound scan can replace electromyography as a diagnostic tool for Carpal Tunnel Syndrome in occupational medicine. To this purpose, automatic procedures based on electromyographic and ultrasound measurements were compared. Our study showed that both methodologies have a comparable degree of accuracy in their predictions, confirming the very good performance of both the automatic procedures as medical diagnosis tools. In particular, we noticed that, together with accurate predictions, box-clustering is also able to provide characteristic intervals of values for all the variables (boxes) that are easily interpretable by the clinician as specific profiles related to presence/absence of the syndrome. On the one hand, our results confirmed the high reliability of electromyographic testing, on the other hand, they show that, for the diagnosis of Carpal Tunnel Syndrome, this examination cannot be completely replaced by ultrasound imaging. However, they indicate that ultrasound scan can be a valuable screening tool for Carpal Tunnel Syndrome to detect the pathology when it is still at an early stage. This result is particularly relevant in the occupational medicine context: taking into account the many advantages of ultrasound scan – such as low cost, ease of repetition, non invasive nature, etc. – this technique could be adopted as a pre-test for the diagnosis of Carpal Tunnel Syndrome, while the – more expensive and more invasive – electromyographic examination could be performed in a second phase only to confirm the presence of the syndrome.

1. Materials and Methods

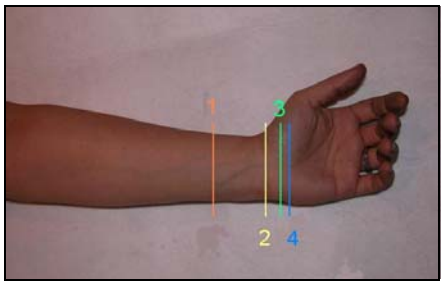
The study is based on a set of data collected in the period ranging from December 2003 to September 2004 on a sample of 102 subjects that were examined by one of the authors (Gioia), an occupational medicine specialist. All the subjects were examined at both wrists through electromyography and ultrasound imaging. In particular, electromyography was performed 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 in the INAIL-Abruzzi Regional Polydiagnostic Center.

In order to achieve the medical diagnosis for the carpal tunnel syndrome (CTS), the following evaluation criteria were adopted: the Distal Latency of the Motor nerve (or Distal Motor Latency – DML) was measured and a value greater than or equal to 4.4 msec was considered "at risk"; the

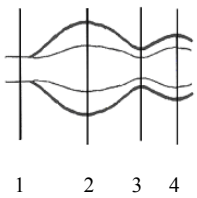
Nerve Conduction Velocity (NCV) was also considered and a value less than or equal to 41.5 m/sec, measured at the carpal tunnel level, was considered "at risk" as well. Idiopathic CTS patients, i.e., those arising spontaneously or from an obscure or unknown cause, were included in the group of the affected subjects.
The ultrasound graphic imaging was carried on, independently of the others, by a third 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. This examination was performed in the same day as the electromyography. The ultrasound machine used for this particular examination was a "PIE MEDICAL – 100 Falco" (see, Figure 1) with a linear probe with 6-8 MHz frequency.

Figure 1 – Ultrasound machine used in this study for collecting data.

The ultrasound-based diagnostic consisted into two successive phases: in the first one, a longitudinal scanning was performed in order to locate the track of the nerve; in the second one, the graphical imaging of the nerve was taken at different axial levels (the four levels shown in Figure 2). Levels 1 and 2 were located through marks on the skin surface. Level 1 corresponds to the third distal of the arm, between the wrist and the cubital fold, while level 2 is represented by the wrist fold. Level 3 was located in the center of the hook in the hamate bone, that is, the bone on the medial (ulnar) side of the distal row of the carpus. Level 4 corresponds to the distal limit of the flexor retinaculum.



a)



b)

Figure 2 – a) Different axial levels for ultrasound scan; b) "Hourglass" configuration of the median nerve.

Once levels 1-4 were located, the whole track of the nerve was visualized on the axial raw producing the confirmation that levels 4, 3 and 2 well represent the *head*, the *neck* and the *body* of the "hourglass" configuration of a nerve affected by CTS (see, Figure 2).

The hourglass configuration alone was not considered a certificate of CTS, since, according to the literature this configuration may be observed both on subjects suffering from CTS and on those who are not affected by this syndrome. Hence, the images at the four levels were obtained through a scanning orthogonal to the nerve axis, that is, the axis at which all the structures (the nerve, the flexor tendons of the hand, the muscles) are hyper echoic.
The nerve section, for the purposes of computing its area, was considered as an oval shaped section and the principal and secondary axes were measured (see, Figure 3). For both wrists, the different areas measured at different levels were classified by a label consisting of a letter (R or L, according to the "dominant" limb), followed by the number referring to the level (for the labeling of the variables see, Table 1 in Section 2).

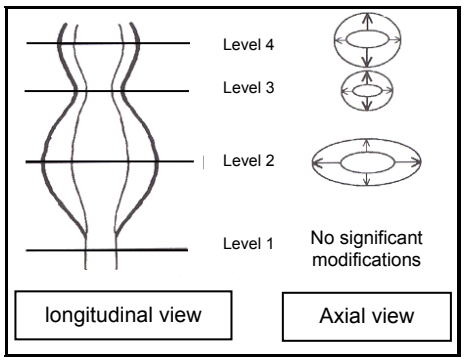


Figure 3 – Oval shaped nerve sections at the four different levels.

A simple index of compression was computed through the following formula:

$$Compression\ index = \frac{Section\ area\ at\ level\ 2 - Section\ area\ at\ level\ 3}{Section\ area\ at\ level\ 2}$$

This index measures the percentage of the median nerve compression at level 3 with respect to the largest section of the median nerve.

2. Input Data

The data consist of 102 records described by a set of explicative variables of two different types: electromyography variables (EMG) and ultrasound variables (US). All the values of the variables were measured both on the right and the left side wrist. Additional information related to the outcome of medical diagnosis was available and, according to it, the subjects were classified into four classes: no injury, injury in the right side, injury in the left side, bilateral injury. Table 1 shows a full description of the 10 ultrasound variables (5 for each side) and the 4 electromyography ones (2 for each side). There are no missing values in the data.

Exam	Longitudinal Scanning Measure (explicative variables)	
US	**Right Wrist**	**Left Wrist**
Level 1	R1	L1
Level 2	R2	L2
Level 3	R3	L3
Level 4	R4	L4
Compression Index	CR2R3	CL2L3
EMG	**Right Wrist**	**Left Wrist**
Sensory Nerve Conduction Velocity	SNCVR	SNCVL
Distal Motor Latency	DMLR	DMLL

Table 1 – Full description of the variables and their coding.

Tables 2 and 3 show some statistical indexes related to the distribution of the explicative variables (minimum, maximum, mean, standard deviation).

	US				EMG		
	R1	**R2**	**R3**	**R4**	**CR2R3**	**SNCVR**	**DMLR**
Min	1.02	2.35	2.43	1.93	-19.80	27.60	2.30
Max	3.20	17.50	16.10	15.32	56.80	77.50	6.80
Mean	1.91	7.41	5.74	6.37	17.54	49.80	4.14
StdDev	0.47	3.72	2.55	3.03	17.88	9.37	1.04

Table 2 – Statistical indexes related to the distribution of the right wrist variables.

	US				EMG		
	L1	**L2**	**L3**	**L4**	**CL2L3**	**SNCVL**	**DMLL**
Min	0.98	1.75	1.69	2.11	-27.10	30.00	2.30
Max	3.76	19.60	9.34	10.05	56.80	89.50	6.50
Mean	1.79	6.13	5.38	5.55	9.65	52.69	3.74
StdDev	0.44	2.67	1.81	1.91	13.09	8.05	0.66

Table 3 – Statistical indexes related to the distribution of the left wrist variables.

The original version of the dataset has two response variables: the first one represents the outcome of medical diagnosis, which was encoded into four possible states: 0 (no injury, 58 records), 1 (right injury, 35 records), 2 (left injury, 6 records), and 3 (bilateral injury, 3 records). The second one is a global diagnosis variable derived from the former one with values 0 and 1 for healthy (58 records) and sick (44 records) subjects, respectively.

This paper mainly focuses on the one-side diagnosis variable (Right Carpal Tunnel Syndrome – RCTS) related to injury to the right wrist. For this case our dataset has 64 records with no injury (RCTS=0) and 38 records with injury (RCTS=1). In Table 4, we show some records extracted from the dataset. For the specific purpose of our analysis, each record is completely described by the values of the seven variables measured at the right wrist. On the basis of the additional information provided by the RCTS variable, the records are divided into two groups: those affected by CTS at the right wrist (Right Injury) and those who are not (No Right Injury).

		US				EMG		
	Record	**R1**	**R2**	**R3**	**R4**	**CR2R3**	**SNCVR**	**DMLR**
Right Injury	1	1.50	4.06	2.43	3.67	40.14	34.70	3.50
	2	1.72	13.77	7.20	10.47	47.70	33.50	6.20
	3	1.88	8.06	5.48	6.40	40.30	44.50	4.40
No Right Injury	4	1.80	5.82	3.86	3.28	33.70	51.20	3.70
	5	1.38	14.75	13.50	15.16	8.40	54.60	3.50
	6	1.38	6.41	5.77	6.40	9.90	56.40	4.10

Table 4 – Example: some right side injury records extracted from the dataset.

3. The box-clustering technique

In this section we illustrate a classification technique called *box-clustering* (BC), which was first introduced in [Hammer, Liu, Simeone, Szedmák, 2004]. This technique is a branch of a more general methodology, called *Logical Analysis of Data* (LAD), pioneered in 1986 by Prof. Peter Hammer from Rutgers University, USA, and developed afterwards by Prof. Hammer himself and his School. LAD can be applied to different data analysis problems and, in particular, it showed very good performances when adopted in medical diagnosis applications where a group of subjects is followed in order to detect if they suffer from a given syndrome or not [Bonates, Hammer, 2006]. Typically, a set of variables or attributes is observed on the subjects and among the collected data there is a variable that is able to univocally identify the subjects affected by the syndrome.

Given a particular syndrome, suppose that in the group of subjects – which will be called the *training set* – we know who is affected by the syndrome and who is not. We will refer to the former ones as *positive observations* and to the latter as *negative observations*. Moreover, suppose that a set of variables or attributes has been observed and measured by appropriate indicators on all the subjects. Notice that these attributes can be numbers (such as 11.6 or 168), or ordered values with tree or more levels (such as *Low, Medium, High*), or even binary values, that is, *yes/no* variables which detect the presence or the absence of a given symptom. All the available information related to the group can be recorded in a table that collects the input data for the analysis as in Table 4. In this table each row corresponds to a subject and the rows are divided into two groups referring to the positive (RCTS=1) and to the negative (RCTS=0) subjects, respectively. The columns represent the variables: in each column we find the value of the corresponding variable observed in each subject.

Assume we have a different group of subjects for which the values of the same variables are available, but we do not know yet if they suffer from the syndrome or not. We call this group the *testing set* and we want to understand if it is possible to detect which subjects in the group risk to contract the syndrome in the short term (positive), and which do not (negative). The BC procedure 'learns from the data', in the sense that it is able to extract information from the input table of the training set and use it in order to classify each subject in the testing set either as positive or as negative. Moreover, the BC procedure is able to 'explain' *why* a subject in the testing set was classified as positive or negative with regard to the specific values of the attributes observed for him/her.

A key role in BC is played by the notion of box. A *box* is specified by an interval of values for each variable. For example, with respect to the variables measured on the right wrist, the following B_1 and B_2 are two possible boxes:

Variables	Boxes	
	B_1:	B_2:
R1	1 – 2	1 – 3.5
R2	8 – 15	10 – 15
R3	2 – 10	5 – 10
R4	2 – 12	8 – 16
CR2R3	35 – 50	8 – 35
SNCVR	40 – 50	50 – 60
DMLR	4 – 6	3 – 3.5

Table 5 – Two possible boxes for the right injury variables.

We say that a subject *belongs to* (or *falls within*) *a box* if the values of the variables observed for that subject are contained in the corresponding intervals of the box. For example, subject 3 in Table 4 belongs to B_1 but not to B_2.

We also define a box B to be *positive* if it contains at least one positive subject and no negative subject in the training set. A *negative box* can be similarly defined.

In order to solve our classification problem, we need to answer to the following two main questions:

1. how to generate a system of positive and negative boxes using the information of the input data table of a given training set;
2. how to use these boxes to classify the subjects in the testing set as positive or negative.

We will describe the procedure through a geometrical representation. To simplify the analysis, assume that we have only two variables, Sensory Nerve Conduction Velocity (SNCVR) and Distal Motor Latency (DMLR). Consider for example subject 4 in Table 4: SNCVR=51.20 and DMLR=3.70 are the corresponding values observed for the two variables. These values can be seen as the coordinates of a point P in the plane (see Figure 4). The point P is called *positive* or *negative* if the subject is positive or negative, respectively. Thus, in this case, P is negative since subject 4 is not affected by CTS at right wrist. In a similar way, we can represent a box geometrically. Consider an arbitrary box, for example the one shown in Figure 4. The box B is completely described by the rectangle with two horizontal and two vertical sides shown in the figure.

Notice that a box can be even a segment (horizontal or vertical) when one of its two intervals degenerates to a single value. It can also happen that the rectangle becomes a point when both intervals are single values.

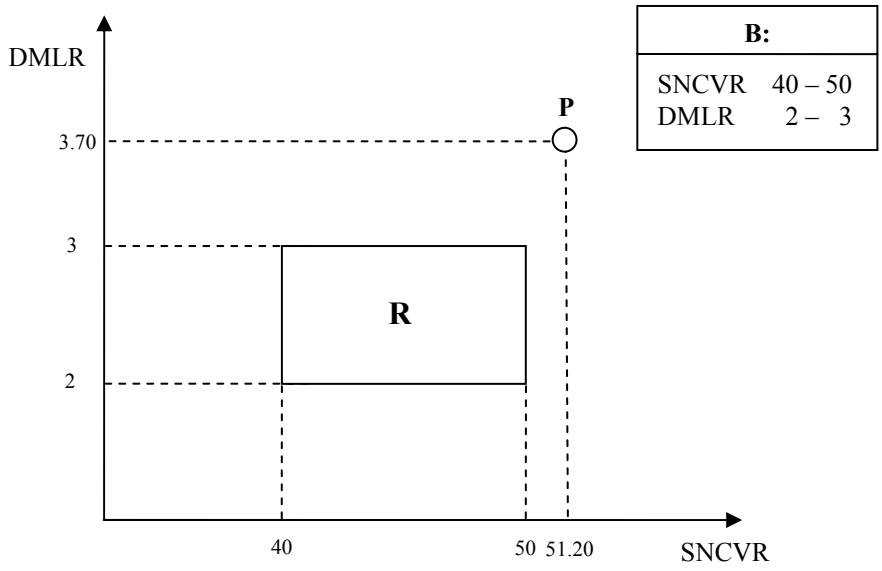


Figure 4 – Geometrical representation of a subject (point P) and a box (rectangle R).

On the basis of the above considerations, we are able to provide a simple description of the BC procedure. Starting from the positive and negative points that correspond to the subjects of the training set, the procedure is able to generate a set of rectangles (boxes) for which the following properties hold (see Figure 5):

1. a box is either positive or negative, that is, there are no boxes that contain both positive and negative points;
2. each point in the training set is contained in only one box, that is, each positive point is contained in only one positive box and each negative point is contained in only one negative box;
3. two positive (negative) boxes are always disjoint rectangles;
4. a positive box and a negative one may overlap, provided that their intersection – a rectangle itself – does not contain any point of the training set.

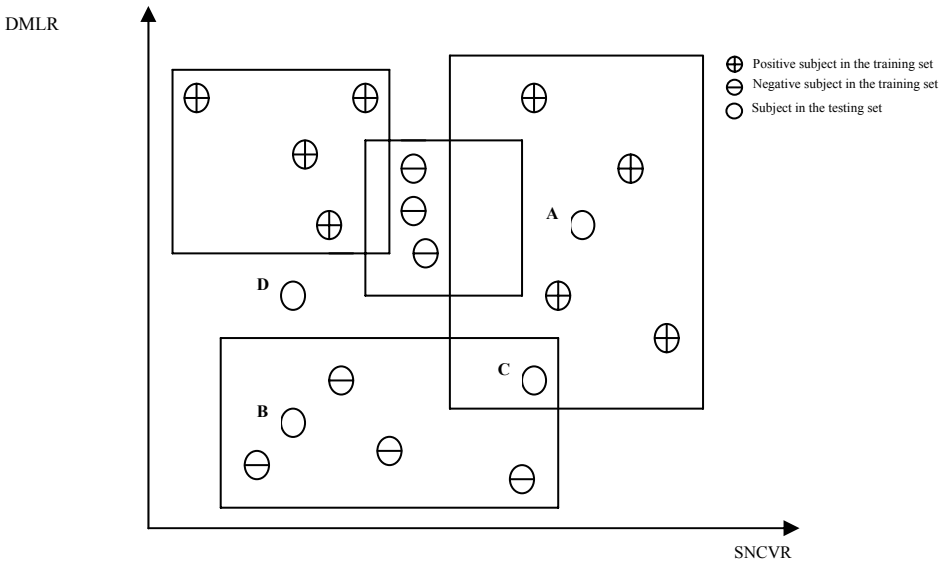


Figure 5 – An example of a set of configurations satisfying properties 1– 4.

Figure 5 also shows all possible positions of a point in the testing set. We have four possibilities:
- the point is contained in some positive box, but in no negative one (point A);
- the point is contained in some negative box, but in no positive one (point B);
- the point is contained both in some positive and in some negative box (point C);
- the point is neither contained in a positive box, nor in a negative one (point D).

According to our procedure we can state the following:
- points of type A are declared to be positive;
- points of type B are declared to be negative;
- points of type C and D are classified according to some additional criterion involving, e.g., their distances from the closest positive and negative points, or from the closest positive and negative boxes.

In order to validate the BC procedure – as any other dichotomic classification one – the traditional *k-fold cross-classification* method is adopted: given an input table having the same form as Table 4, a training set is obtained by extracting a sample of subjects from the set of rows in the table, so that both positive and negative subjects are represented in the sample. The set of the remaining subjects automatically becomes the testing set. For example, suppose that the dataset is composed only by the subjects of Table 4, then a possible choice is to include subjects 1, 3, and 4 in the training set and the remaining subjects 2, 5 and 6 in the testing set.
The set of boxes is produced on the basis of the information of the subjects in the training set and the subjects in the testing set are classified accordingly. Then, since for these subjects we already know if they suffer from the syndrome or not, for them we compare the classification observed in the input table with the classification provided by the BC procedure in order to verify if they coincide. Finally, some performance indices are computed. Among them, the most important ones are:

Accuracy index: percentage of points in the testing set for which the procedure is able to provide a correct classification.

Coverage index: percentage of points in the testing set that are 'covered' by some – positive or negative – box, that is, they are included in at least a box (points of type A, B and C in Figure 5).

Overlapping index*:* percentage of points in the testing set that are contained both in a positive and in a negative box (points of type C in Figure 5).

This procedure is repeated k times. Each time a different sample of n/k subjects for the testing set is extracted, the remaining subjects form the training set and the corresponding values for the performance indices are recorded. Clearly, a good performance of the procedure is detected by high values for accuracy and coverage and low values for overlapping.

4. Quantitative analysis

In this section we apply two automatic diagnosis methodologies for CTS. The first one is based on Discriminant Analysis (DA), a traditional classification tool in statistics; the second one on box-clustering (see Section 3 for a detailed description), a technique that appeared in the literature recently. In both cases we mainly focus our attention on the US variables and try to exploit the predictive power of our techniques in order to obtain a correct classification of a subject – as sick or not – according only to the values of the US variables observed on him/her. As we will see, the experimental results show that predictions based on EMG variables are almost certain, whatever the diagnosis technique adopted. This first result confirms the high reliability of EMG based testing.
In both the procedures, the role of the diagnosis variable RCTS is crucial. On the basis of this variable, we are able to carry out a cross validation procedure in order to test the performance of our

techniques and provide a complete analysis of the classification errors. We will also compare the predictions of CTS based on US variables with those based on EMG ones.

It must be noticed that the classification of the subjects into healthy and sick with respect to CTS is not completely objective. In fact, different schools of thought provide different classifications (even in three or four different classes) by assessing the severity of the pathology on the basis of the stage of advancement of the disease. Even if, for our specific diagnostic purposes, this study is concerned with the dichotomic classification of the subjects (healty/sick), by applying cluster analysis to our dataset, we are able to recognize five actual different classification levels for CTS.

4.1 Cluster Analysis and Discriminant analysis

We consider a two-phase analysis: in the first phase we apply cluster analysis in order to study the interrelations existing between the distribution of the ultrasound variables and the electromyographic ones in the group of subjects under study; this information is used in the second phase as input of a discriminant analysis procedure.
Cluster analysis was performed on the full dataset, that is, all the 14 variables were considered (10 ultrasound measures and 4 electromyographic ones), in order to detect the real nature of the data through a descriptive analysis. The results have pointed out that the classification into five groups may be considered optimal in a statistical sense. This agrees very well with the medical practice according to which subjects are generally divided into five groups: healthy, affected by right injury, left injury, bilateral injury, and subjects affected by a mild syndrome. In Figure 6 the five groups are represented in a Cartesian reference system obtained by Principal Component Analysis (PCA). We notice that, while offering an exhaustive view of the correlation that exists among variables, PCA also allows visualize data with the least possible distortion.
The interpretation of the groups relies on the correlation with the variables that it is possible to examine inside the circle shown in the figure. Thus, we can interpret in a statistically correct way the existing links between the ultrasound and the electromyographic measures carried out on the subjects. It is important to observe that the classification into five groups is optimal also from a statistical point of view because it maximizes the variance gain over all the possible classifications of the subjects.

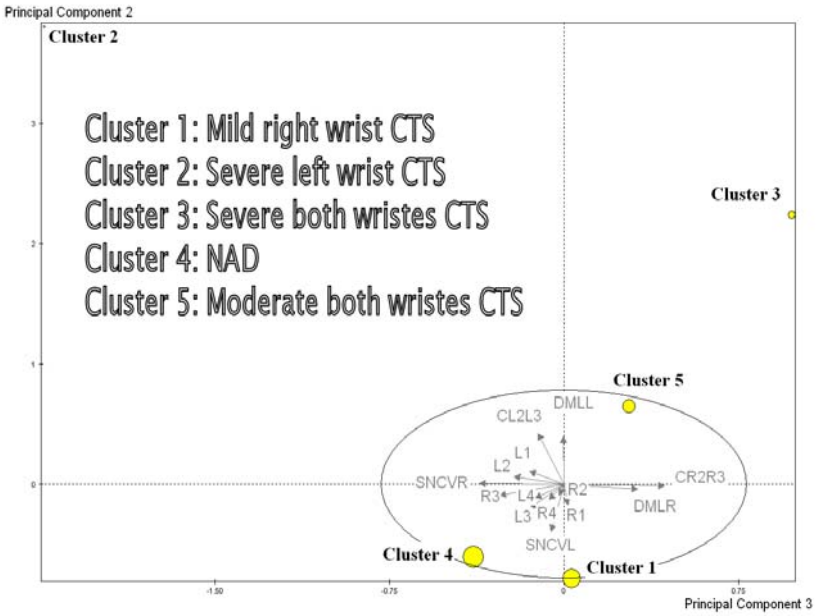


Figure 6 – Principal Component Analysis: representation of the five groups of subjects.

Discriminant analysis allows to describe the differences between two or more groups in which a population is classified, as well as, to assign a new observation to one of these groups. It is able to perform an automatic diagnosis for a new subject, but it also provides an explanation for this diagnosis that helps to understand the phenomenon under study. For this reason, the approach is traditionally accepted in medicine, since it does not behave as a "black box", but provides self explanation. In our case, the observations (subjects) have been preliminarily labelled as sick or healthy. DA has been applied to all the available data (both ultrasound and electromyographic variables were considered) in order to recognize the state of the two wrists, separately.

In the rest of the section we will discuss the results about the right wrist only based on a dataset that involves only US variables (R1, R2, R3, R4, CR2R3) and the binary diagnosis variable denoting presence/absence of CTS in the right wrist of the subject (RCTS). The statistical model is the following:

$$c_1 R_1 + c_2 R_2 + c_3 R_3 + c_4 R_4 + c_5 CR2R3 + C \qquad \begin{cases} < 0 & Sick \quad (\text{RCTS} = 1) \\ > 0 & NAD \quad (\text{RCTS} = 0) \end{cases}$$

where C is a constant and c_1, c_2, c_3, c_4, and c_5 are the coefficient associated to the five variables considered. If the function value is negative, the subject is classified as "Sick" otherwise he/she is "NAD" (not affected by injury at the right wrist).

As a first application, the analysis was performed on all the subjects, that is, the full dataset was regarded as the training set, so that the validation set was empty. In a second step, we carried out a 5-fold cross validation, with 80% of the subjects in the training set and the remaining 20% in the test one. The cross validation results confirmed the good performance already observed on the full dataset. The estimated coefficients of the discriminant function are the following:

$$c_1 = 3.03 \quad c_2 = -0.113 \quad c_3 = 0.677 \quad c_4 = -1.15 \quad c_5 = -0.24 \quad C = 15.85 .$$

Now we are going to examine the results of the analysis. We begin from the averages of each variable for the two groups of subjects with no right injury (NAD) and subjects with right injury (Sick) that are shown in Table 6. We immediately notice that the mean value of each variable for healthy subjects is below the general mean and it is above it for the sick ones.

Group	Variable	Intra-Group Mean	Total Mean
No right injury (NAD)	R1	1.707	1.905
	R2	5.551	7.407
	R3	5.129	5.743
	R4	4.833	6.367
	CR2R3	6.594	17.535
Right injury (Sick)	R1	2.239	1.905
	R2	10.533	7.407
	R3	6.777	5.743
	R4	9.949	6.367
	CR2R3	35.963	17.535

Table 6 – Intra-group and total mean for the two groups of subjects.

In Table 7 we report on the classification results obtained by DA. We define as *false positive* a "no right injury" record that was misclassified as a "right injury" one. This case corresponds to a mild error. We define as *false negative* a "right injury" record that was misclassified as a "no right injury" one. Obviously, this is a strong error, since in this case we are not able to recognize a sick patient.

Our results show only four classification errors that correspond to four false positive records, while there are no false negative ones. The total error percentage amounts to 3.92%. This percentage of error is small to the point that it can be considered acceptable.

RCTS real value	Discriminant Analysis Classification				
	Correct		Incorrect		Total
NAD	60	(93.75%)	4	(6.25%)	64
Sick	38	(100%)	0	(0%)	38
Total	98	(96.08%)	4	(3.92%)	102

Table 7 – Discriminant analysis classification.

In Figure 7 we can examine how subjects are represented by the first two factors. This analysis is useful to understand the nature of the errors. We see that two of the subjects erroneously classified as sick are 'borderline' (records 14 and 26), while the other two (records 55 and 103) are in the NAD sector. On the basis of the information contained in Table 8, the following conclusions can be drawn for the four incorrect predictions:

- for subject 14 we note that the discriminant function is just slightly below zero and this is due almost surely to the breadth of the nerve at levels 2, 3 and 4;
- for subject 26 we are almost near the threshold of 'irritative' CTS and the high compression of the nerve makes the discriminant function negative;
- for subject 55 the high compression makes the discriminant function negative;
- for subject 103 compression causes the function to be slightly negative.

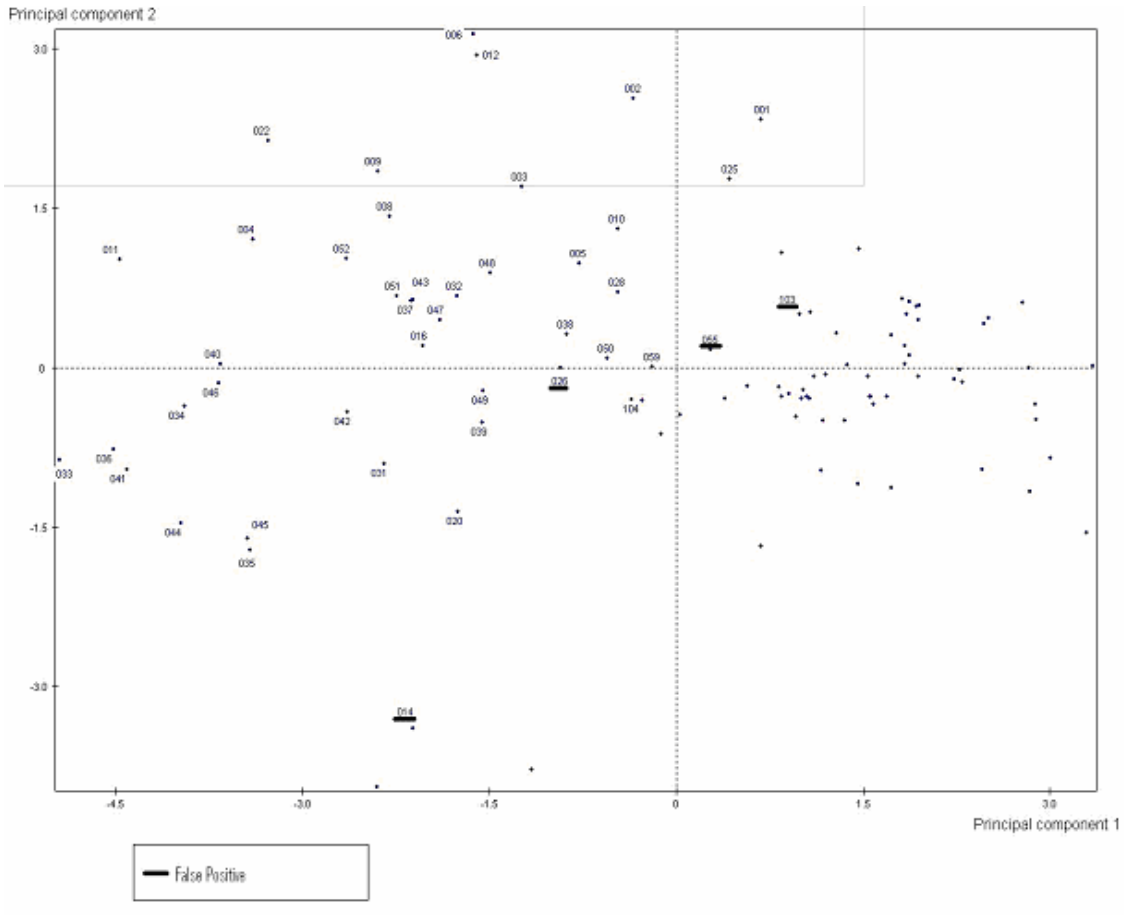


Figure 7 – Subjects profile with respect to the first two principal components.

Record	US					EMG		Discriminant Function
	R1	R2	R3	R4	CR2R3	SNCVR	DMLR	
14	1.38	14.75	13.5	15.16	8.4	54.6	3.5	-0.2935
26	2.26	8.57	6.49	7.01	24.2	43.6	4.1	-1.3981
55	2.33	7.07	4.82	4.99	31.8	51.3	3.0	-2.0590
103	1.72	5.75	4.08	5.33	29.0	51.2	2.9	-0.2872

Table 8 – Profile of the four misclassified subjects.

In order to test the performance of our discriminant function, we have carried out a cross validation procedure[1] (see Table 9). For each run (Cross) of the cross validation, and for each record in the test set, we show the predicted value of RCTS and compare it with the corresponding real value. The "Errors" column shows the number of errors of the estimated values with respect to the real ones. Since we have two different error situations, in this table we used two different typographic solutions for them:

- for false positive, that is, a "No Right Injury" record that was misclassified as a "Right Injury" one, we set a boldface 1 in the sequence.
- for false negative, that is, a "Right Injury" record that was misclassified as a "No Right Injury" one, we adopt a boldface 0 in the sequence.

The best performance is detected when k=3 since we obtain only one misclassified subject (error of 4.76%) and it corresponds to a false positive; in the case of k=1 we have the same error percentage, but the error is much relevant because we have a false negative; in the other cases we have only two subjects (error of 9.52%) and we have the worst performance when k=2.
As a global result, we can conclude that in the best case we can classify almost surely every subject, in the worst case we might be in trouble in the 9.52% of the cases, but this error can be considered acceptable from a statistical point of view.

Cross: k=1		
Variables	**Estimated values for RCTS**	**Errors**
US	1101100**0**1110000000000	1
RCTS	11011010111000000000	
Cross: k=2		
Variables	**Estimated values for RCTS**	**Errors**
US	**0**1111111**1**0111100000000	2
RCTS	11111110011110000000	
Cross: k=3		
Variables	**Estimated values for RCTS**	**Errors**
US	0100000010000000000**1**0	1
RCTS	010000001000000000000	
Cross: k=4		
Variables	**Estimated values for RCTS**	**Errors**
US	11101111111**1**000000001	2
RCTS	11101111110000000000	
Cross: k=5		
Variables	**Estimated values for RCTS**	**Errors**
US	1100**1**1110000000000001	2
RCTS	11000111000000000000	

Table 9 – DA results (RCTS estimates and errors).

[1] The level of cross validation we considered is 5-fold (i.e., k=5) as defined in Section 4.1. This means that we run the procedure 5 times and in each run we use 81 records to build the BC model (80% training set), and 21 records to test this model (20% test set).

Table 10 shows the error summary: we have a 5.7% of false positive (mild error) and only a 1.9% of false negative (strong error). The total error is only 7.6% of the cases.

Variables	False positive		False negative		Total Errors	
US	6	(5.7%)	2	(1.9%)	8	(7.6%)

Table 10 – Error summary for DA cross validation.

4.3 Box-clustering results

In the present study, the main goal of diagnostic (predictive) box-clustering modelling[2] was to test the classification potential of two alternative sets of explicative variables with respect to the classification variable RCTS. In particular, according the definitions given in Section 2, two sets of variables are considered:

- *Ultrasound* measures (US) – R1, R2, R3, R4, CR2R3;
- *Electromyography* measures (EMG) – SNCVR, DMLR.

We have derived three data sets from the original data (we will refer to them as US, EMG, and US+EMG, respectively) with the same diagnosis variable RCTS. The BC approach was applied to each dataset and the classification error evaluation was based on the same cross validation scheme adopted for DA. In Table 11 we show the results for each run in the cross validation loop and for each different set of variables.

Cross: k=1		
Variables	**Estimated values for RCTS**	**Errors**
EMG	11011010111000000000	0
US	11**1**1100**0**1110000000000	2
US+EMG	11**1**1100**0**1110000000000	2
RCTS real value	11011010111000000000	
Cross: k=2		
Variables	**Estimated values for RCTS**	**Errors**
EMG	11111110011110000000	0
US	111**0**1110**0**011100000000	2
US+EMG	11111110**0**011100000000	1
RCTS real value	11111110011110000000	
Cross: k=3		
Variables	**Estimated values for RCTS**	**Errors**
EMG	010000001000000000000	0
US	0100**1**0001000000000000	1
US+EMG	010000001000000000000	0
RCTS real value	010000001000000000000	
Cross: k=4		
Variables	**Estimated values for RCTS**	**Errors**
EMG	11101111111000000000	0
US	1110111111**1**100000000**1**	3
US+EMG	11101111111000000000	0
RCTS real value	11101111111000000000	

[2] This approach is defined as "*logic mining in supervised classification*" in the Data Mining context.

Cross: k=5		
Variables	**Estimated values for RCTS**	**Errors**
EMG	11000111000000000000	0
US	**0**1001111000000000000	2
US+EMG	1100**1**111000000000000	1
RCTS real value	11000111000000000000	

Table 11 – BC results (RCTS estimates and errors).

In Table 12 we provide an error summary for each set of variables, specifying the number of false positive, false negative, and the total. There is no doubt that we obtain the best results (no errors at all) when we use the EMG set. The US set gives raise to a non negligible error level (9.8%) especially if compared with the EMG set, while the US+EMG set ranks in the middle (3.9%).

Variables	**False positive**		**False negative**		**Total Errors**	
EMG	0	(0.0%)	0	(0.0%)	0	(0.0%)
US	4	(3.8%)	6	(5.9%)	10	(9.8%)
US+EMG	2	(1.9%)	2	(1.9%)	4	(3.9%)

Table 12 – Error summary for BC cross validation with respect to each set of variables.

Tables 13, 14, 15a and 16b provide an analytical description of the evaluated box systems corresponding to the best error performances case of Table 11, that is, the one corresponding to k=3. When a box is a singleton we use one column (see, for example, Box 3 in Tables 13 and 14). The US box system is more complex than EMG and US+EMG and, for this reason, we split it into two subsets: "RCTS=0" in Table 15a, and "RCTS=1" in Table 15b.

US+EMG	**No right injury**		**Right injury**		
	Box 1		**Box 2**		**Box 3**
R1	1.02	2.70	1.03	3.20	1.50
R2	2.35	17.50	4.33	16.70	4.06
R3	2.76	16.10	2.93	11.03	2.43
R4	1.93	15.16	3.62	15.32	3.67
CR2R3	-19.80	33.70	14.50	56.80	40.10
SNCVR	39.50	77.50	27.60	54.90	34.70
DMLR	2.30	4.10	4.40	6.80	3.50

Table 13 – Box system for US+EMG variables.

EMG	**No right injury**		**Right injury**		
	Box 1		**Box 2**		**Box 3**
SNCVR	39.50	77.50	27.60	54.90	34.70
DMLR	2.30	4.10	4.40	6.80	3.50

Table 14 – Box system for EMG variables.

US	**No right injury**				
	Box 1	**Box 2**		**Box 3**	
R1	1.80	1.38	2.70	1.02	2.50
R2	5.82	14.75	17.50	2.35	8.57
R3	3.86	13.50	16.10	2.76	7.90
R4	3.28	6.04	15.16	1.93	7.01
CR2R3	33.70	8.00	8.40	-19.80	31.80

Table 15a – Box system for US variables (RCTS=0).

US	Right injury					
	Box 1		Box 2		Box 3	
R1	1.50	2.00	1.03	2.80	2.51	3.20
R2	4.06	8.06	8.15	16.70	5.93	6.80
R3	2.43	5.48	4.08	11.03	4.52	4.64
R4	3.62	6.40	7.06	15.32	5.77	6.64
CR2R3	32.30	56.80	14.50	55.10	23.80	31.70

Table 15b – Box system for US variables (RCTS=1).

The characteristics of the box systems are strongly linked with the mutual positions of the records in the space of the chosen variables. In Figure 8, we show the distribution with respect to the EMG variables[3] and we notice the almost perfect "box-separation" of the two coloured clouds. This is an empirical justification for the very good results of BC approach when applied to EMG variables, since it provides a simple model (2-3 boxes for each class) and a null classification error (0%).
We cannot have an explicit geometrical description for the US set, but we can analyze the shape of some sub-cluster. In Figures 9 we see that for some of these sub-clusters the red and blue points are definitely not well separated. This means that we have a more complicated BC model for the US set (with 5-6 boxes for each class) and a higher classification error (9.8%) than for the EMG set.

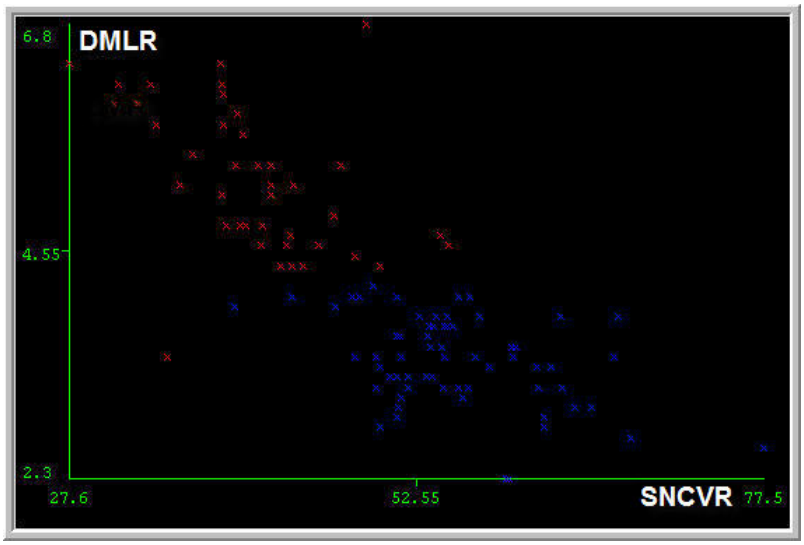


Figure 8 – Electromyography analysis.

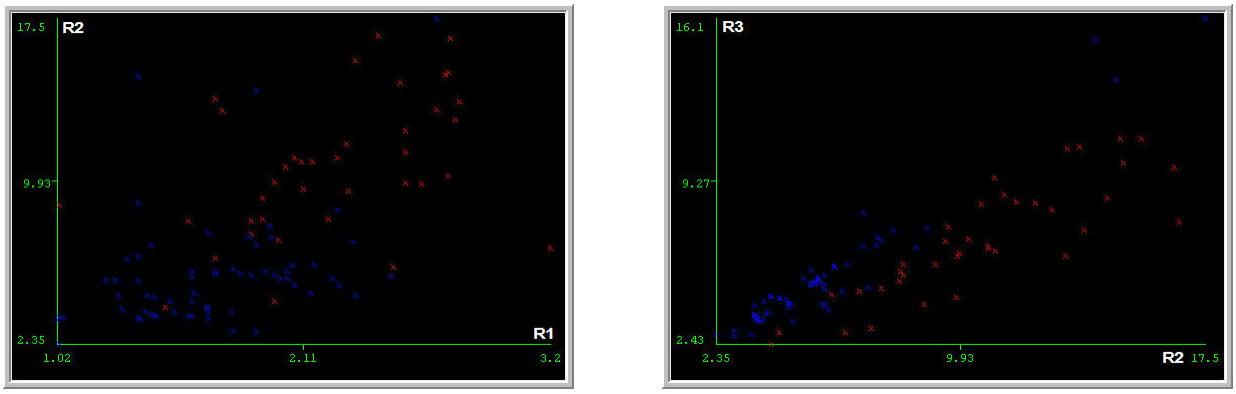


Figure 9 – Ultrasound analysis: R1 *vs* R2 and R2 *vs* R3.

[3] The negative records (RCTS=0) are represented by blue points, the positive ones (RCTS=1) by red points.

In order to analyze our results in depth, we defined a measure of the distance between two systems of boxes and we called it *normalized distance*[4]. We computed the normalized distances between every pair of the above box systems and, in particular, we found that it is equal to 0 when computed with respect to the EMG+US box system and the EMG one. This means that the boxes of these two systems contain exactly the same records: the addition of the US variables has no influence whatsoever on the classification into boxes. On the other hand, we found that the distance between the US box system and the EMG one is equal to 0.526, and this shows only a partial association between the US and the EMG box systems.

6. Conclusions

The main question that was addressed in the present work was: to what extent can Ultrasound Scan effectively replace Electromyography as a diagnostic tool for CTS in occupational medicine?

In order to answer this question, some automatic diagnosis methodologies were used. In addition to classical statistical tools, such as cluster analysis and discriminant analysis, the recent BC technique was employed here for the first time in an actual medical application. Automatic predictions of CTS based on EMG and US measurements were compared.
All the above methodologies showed a comparable degree of accuracy in their predictions. The accuracy was exceptionally good (0% errors), fairly good (9.8 % errors), and good (3.9 % errors) when relying on EMG, US, and EMG+US variables, respectively.

On the one hand, our results confirmed the high reliability of EMG testing. Admittedly, early alterations of the Nerve Velocity of Conduction are not always detected by the sole US test: as a matter of fact, neither the section areas, nor the Compression Index display a good predictive capacity of the electrophysiological parameter, which still remains pretty much in the domain of Neurology and Neurophysiopathology. It should be noticed, though, that in the early "irritative" stage of the pathology often even neurophysiological tests are unable to detect the presence of the syndrome, which is revealed only by a clinical/anamnestic exam.

On the other hand, our results indicate 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;
- non invasive nature;
- easy acceptance by the patient;
- detection of possible underlying anomalies.

The above aspects become much more relevant if one considers the ultrasound imaging as a pre-test for the diagnosis of CTS in occupational medicine. 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 suspected of suffering from CTS.

[4] We define the *normalized distance* between two systems of boxes as $(a+b)/2$, where a is the average distance between a box in the 1st system and the closest box in the 2nd one, and b is the average distance between a box in the 2nd system and the closest box in the 1st one. Here the *distance* of two boxes is defined as the percentage, among all the records that belong to either box, of those records that are not shared by both boxes. Notice that both box distances and normalized distances always lie between 0 and 1.

The last feature is particularly useful not only in the occupational context, but also in view of possible later surgical operations, whose invasivity could be substantially reduced.

It is worth mentioning that our results seem to show that the combined use of EMG and US tests is not recommendable, since the accuracy of prediction of EMG alone appears to be superior. This might be due to a "noise" increase caused by the interaction of too many variables.

Some final remarks about the automated diagnosis techniques that have been used in this work must be clarified.

All of them show their potential as a reliable aid in the clinician's practice and as a valuable support for diagnosis. In particular, on the one hand, BC is less flexible than classical cluster analysis, in the sense that the geometrical shape of clusters of observations is rigidly constrained; on the other hand, the intervals of values produced 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 and explanation of the pathology profiles.

References

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

P. L. Hammer, Y. Liu, B. Simeone, S. Szedmák (2004), "Saturated systems of homogeneous boxes and the logical analysis of numerical data", 24th Annual Conference of the German Classification Society, Passau 2000, *Discrete Applied Mathematics*, 144, 103-109.

J. Han, M. Kamber, "Data Mining: Concepts and Techniques", The Morgan Kaufmann Series in Data Management Systems, 2001.

L. Lebart, A. Morineau, K. Warwick, "Multivariate Statistical Analysis", John Wiley & Sons, 1984.

T.M. Mitchell, (1999), "Machine Learning and Data Mining", Communications of the ACM, 42, 11, 30-36.

SPAD (2000), Système Portable pour l'Analyse des Données. CISIA.

W. N.Venables, B. D. Ripley, "Modern Applied Statistics", Springer, 2002.