Scholars Journal of Engineering and Technology (SJET) **ISSN 2321-435X (Online)**
Sch. J. Eng. Tech., 2015; 3(5A):567-571 **ISSN 2347-9523 (Print)**
©Scholars Academic and Scientific Publisher
(An International Publisher for Academic and Scientific Resources)
www.saspublisher.com

Research Article

Support Vector Machine Classification Using Psychological and Medical-Social Features in Patients with Fibromialgya and Arthritis

Yolanda Garcia-Chimeno[1*], Begoña Garcia-Zapirain[2], Heather Rogers[3]
[1,2,3] University of Deusto, Avda. Universidades, 24, 48007, Bilbao, Spain

***Corresponding author**
Yolanda Garcia-Chimeno
Email: yolanda.garcia@deusto.es

Abstract: The SVM classifier is a very powerful tool for helping to diagnose illnesses. Subjects can be classified according to certain characteristics related to pathology. In this paper, the aim is to undertake a classification of arthritis and fibromyalgia pathologies using medico-social and psychopathological characteristics obtained from questionnaires, with a very high classification percentage having been obtained. A 96.4035% success rate was obtained using the SVM classifier only by introducing the psychopathological characteristics. Only specific questionnaires could be put together and the subject diagnosed if they have either fibromyalgia or arthritis, whereby the cost of tests that these types of pathology entail might be considerably reduced.

Keywords: Arthritis, fibromyalgia, classification, SVM, sensitivity, specificity, pathological, medico-social

INTRODUCTION

The classification involves the fact of performing multiple tests to get a reliable diagnosis in some cases. Existing classification techniques help that taking certain characteristics of pathology, to obtain a subject's classification, giving the doctor a basis to offer the final diagnosis.

This classification is done through techniques of 'Machine Learning' [1,2], in which some algorithms can perform a categorization based on the defined features.

In this manuscript, it classified subjects which suffer arthritis and fibromyalgia, through medical, social and psychopathology parameters, is intended to understand the importance of psychopathological assessment in the diagnosis of two similar chronic pain disorders.Fibromyalgia (FM) is a disorder of unknown etiology characterized by widespread pain, abnormal pain processing, sleep disturbance, fatigue and it is often accompanied by psychological distress [3]. It affects 2-3% of the general population and 90% of patients are women [4]. Rheumatoid arthritis (RA) is another type of painful musculoskeletal disease. It is an autoimmune condition with chronic inflammation that affects various joints of the body. RA has a worldwide prevalence of 0.5–1% and tends to affect three times as many women than men [5]. Physicians diagnose FM based on the level of tenderness on some spots of the body when pressure is applied, the duration of the presence of the symptoms, level of fatigue, and cognitive difficulties [6]. In contrast, Rheumatoid arthritis (RA) is diagnosed through the presence of symptoms and results of a physical exam revealing swollen and painful joints, and sometimes laboratory exams detecting the presence of Rheumatoid factor in the blood [6]. Using a classification algorithm the physician has the ability to understand the probably of a specific disease given certain parameters. The physician can use results of classification to guide his clinical decision making regarding a differential diagnosis.

The supervised classification [7], requires a previous training phase, this means that it has to build a group of samples, which have been classified as true, to be able to be identified by the classifier. The features introduced, described these samples and must be discriminatory for an efficient classification [8]. In conclusion, a supervised classifier produces a mathematical function which, from training samples previously tagged, deduced what kind or group belong the set of input samples. Finally, after training, it found the validation group of the classifier, introducing another sample of subjects, which have not participated in previous training. These samples will get the success ratio of the classifier.

To make optimal classification, it exist 'cross-validation' technique [9]; guarantee that the results of the classification are fully independent samples of training and the validation partition.

MATERIAL AND METHODS

Participants

74 women with RA and 53 women with FM were recruited from ambulatory centres in Neiva, Colombia between January 2013 and January 2015. All individuals were diagnosed according to the American College of Rheumatology/European League against Rheumatism (ACR/EULAR) criteria, were aged 18 to 79, and cognitively able to participate. Exclusion criteria were: currently hospitalized, comorbid neurological or psychiatric disorders interfering with independent decision making, terminal illness, or history of alcohol or other drug abuse.

Patients were assessed by a rheumatologist or internal medicine specialist to determine eligibility. After signing an informed consent, a trained research assistant met to obtain demographic and medical information and complete the self-report scales. This study received ethics committee approval.

MEASURES

Visual Analog Scale (VAS) pain rating on a scale from 0 to 10.

Psychopathology

The Symptom Checklist-90-R [13] consists of 90 symptoms of psychological problems that are assessed on a five-point Likert scale. The scale has nine sub-scales: somatization, obsessive–compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. Total scales include the Global Severity Index (GSI) as a measure of total distress, the Positive Symptom Distress Index (PSDI) as an indicator of average symptom intensity, and the Positive Symptom Total (PST) as a count of symptoms rated more than 0. Higher scores indicate more symptoms and/or more distress.

The medico-social (med-soc) and psychopathological (psy) characteristics that were used for classification purposes were as follows:

- Meds3 (med-soc)
- Meds4 (med-soc)
- Comorbi10 (med-soc)
- Yrs_with_disease (med-soc)
- Somavg (psy)
- MRoleEmotional (psy)
- DAS28VAS (med-soc)
- Meds7 (med-soc)
- MRolePhysical (psy)
- Depavg (psy)
- MVitality (psy)
- Anxavg (psy)
- Totalphq (psy)
- Comorbi4 (med-soc)
- Comorbi3 (med-soc)
- Psdi (psy)
- Gsi (psy)
- Living_comparison (med-soc)
- Meds5 (med-soc)
- Totalzung (psy)
- Total60stait (psy)
- Comorbi5 (med-soc)
- Comorbi8 (med-soc)
- MMentalHealth (psy)
- Pst (psy)
- Marital_ra (med-soc)
- Isavg (psy)
- Psyacg (psy)
- Paravg (psy)
- Comorbi5 (med-soc)
- Yrs_schooling (med-soc)
- Income (med-soc)
- MPhysicalFunct (psy)
- Occupation (med-soc)
- Comorbi1 (med-soc)
- Meds9 (med-soc)
- Total60stais (psy)
- Social_stratum (med-soc)
- MGenHealth (psy)
- Ocavg (psy)
- Phobavg (psy)
- MSocialFunct (psy)
- MPain (psy)
- Hosavg (psy)
- Behdiseng (psy)
- Comorbi2 (med-soc)
- Age (med-soc)

The order in which the characteristics appear is from greatest to least importance according to subjects, i.e. a ranking was established taking into account the scores obtained by the subjects in each characteristic. Thus, the importance can be ascertained of those characteristics that account for the greatest load for subsequent classification purposes, to ensure the classification is correct.

SVM Classifier

This classifier is specifically to the classification and regression problems. In this way, it trains the set of training samples, creating a model that classifies each new sample. These, depending on their proximity are classified in one or the other class.

This type of classifier was introduced by Vapnik [10, 11], which it separates a set of binary labelled training data with a hyper-plane (maximum distant).

Given a set of training samples, it can be labelled samples and train a SVM machine to build a model to classify the class of a new sample. Samples are represented as points in space by separating the groups as possible. Therefore, when new samples are introduced into the model, depending on its proximity they can be classified into one class or another. That is, a good separation between classes allows a correct classification.

In this way, given a set of points that each belongs to one of the possible categories, based on SVM algorithm constructs a model to predict the group to which concerned, whose classification is unknown, belongs to one category or another that exist in the model.

Support Vector Machine looking for an optimal hyperplane that separates the points of a class from each other [12]. This algorithm is a supervised classification, so the input data are represented as a dimensional vector of length. The optimal separation is the key feature of this algorithm, to be known as maximum margin classifier as it is intended that the hyperplane at the maximum distance with the points that are closer to him. Thus, the vector points that are labelled with a category are on one side of hyperplane and which are another category, on the other side.

The advantages of this classifier are that SVM has good properties and are robust in terms of overtraining and in terms of dimensionality.

For two-class classification, it has a input samples vectors (1) with corresponding labels (2), that indicates the two classes (-1 and +1).

$$\vec{x_i} \in \mathbb{R}^d (i = 1,2,\dots,N) \qquad (1)$$

$$y_i \in \{+1,-1\} (i = 1,2,\dots,N) \qquad (2)$$

The decision function implemented can be written as:

$$f(\vec{x}) = sgn\left(\sum_{i=1}^{N} y_i \alpha_i * K(\vec{x},\vec{x}_i) + b\right) \qquad (3)$$

The coefficients are obtained the maximize problem:

$$\sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\,\alpha_j * y_i y_j * K(\vec{x},\vec{x}_i)\ \ subject\ to\ 0 \le \alpha_i \le C \qquad (4)$$

$$\sum_{i=1}^{N} y_i \alpha_i = 0 \quad i = 1,2,\dots,N \qquad (5)$$

C is a regularization parameter that controls the trade-off between margin and classification error.

RESULTS

Psychopathology Features

The success rate introducing psychopathology features in SVM classifier is 96.4035%

In addition to the capacity the former has for detecting positive and negative cases, the contingency table was obtained in order to be able to measure the efficiency of the classifier itself, showing whether the subjects were classified correctly or not.

Table-1: SVM contingency table (psychopathology features).

		Reality (Unknown)	
		+ (YES)	- (NO)
Test	+ (YES)	23	1
	- (NO)	1	32

From table 1 can be obtained the sensitivity and specificity rates. The sensitivity rate is 95.83%, meaning it has a very good capacity for detecting positive cases.

96.97% was obtained for the specificity rate; it's also having a very good capacity for detecting negative cases.

Table-2: Sensitivity and specificity results (psychopathology features). TP: True Positive; FN: False negative; TN: True Negative; FP: False Positive.

SENSITIVITY	TP / (TP + FN)	0.9583
SPECIFICITY	TN / (TN + FP)	0.9697

Medico-Social Features

In this new section, we once again calculate all the previous parameters, but only modifying the characteristic matrix used for classification purposes, including only medico-social characteristics.

From the SVM classifier we obtain a good success rate of 85.5263%.

In the contingency table we can see the comparative scores obtained from the classification results, and what this means in reality based on the subjects who underwent the study (cross-validation). On this occasion the scores remain very good, albeit somewhat less so in the case of the psychopathological characteristics.

Table-3:.SVM contingency table (medico-social features).

		Reality (Unknown)	
		+ (YES)	- (NO)
Test	+ (YES)	20	5
	-(NO)	3	29

Certain rates below 90% are obtained when calculating the sensitivity and specificity rates: 86.96% in the case of sensitivity and 85.29 in the case of specificity.

Table-4: Sensitivity and specificity results (medico-social features).TP: True Positive; FN: False negative; TN: True Negative; FP: False Positive.

SENSITIVITY	TP / (TP + FN)	0.8696
SPECIFICITY	TN / (TN + FP)	0.8529

Psychopathological + Medico-Social Features

The psychopathological and medico-social characteristics were combined by moving onto the last block before undertaking classification according to committees, and the three classifiers were once again used. 22 medico-social and 25 psychopathological characteristics were selected.

The classifier subject to study is obtaining a 95% success rate.

When observing the contingency table for this classifier, we note that there were only 3 subjects who were not classified as either true positives" or "true negatives".

Table-5:SVM contingency table (psychopathology and medico-social features).

		Reality (Unknown)	
		+ (YES)	- (NO)
Test	+ (YES)	23	2
	- (NO)	1	31

Therefore, with these scores we can see that the sensitivity and specificity rates are very good – over 93% in both cases, which suggests that the classifier is a very efficient one.

Table-6: Sensitivity and specificity results (psychopathology and medico-social features). TP: True Positive; FN: False negative; TN: True Negative; FP: False Positive.

SENSITIVITY	TP / (TP + FN)	0.9583
SPECIFICITY	TN / (TN + FP)	0.9394

DISCUSSION & CONCLUSSIONS

In all cases percentages over 85.5263% were obtained for each of the cases using the SVM classifier: psychopathological characteristics (96.4035%), medico-social characteristics (85.5263%) and a combination of the two (95.8246%).

The SVM classifier provides a powerful method for classification of samples. This is because it has a solid grounding in statistical learning, enabling the optimum decision-making function to be found for the set of training data [14].

In view of the previous results, it can be ascertained that the SVM algorithm performs well in classifying subjects with arthritis and fibromyalgia, obtaining a 96.4035% success rate with regard to psychopathological characteristics. However, it is also true to say that this rate does not drop excessively in the case of the medico-social characteristics, and the combination of the two types is the same as the first-mentioned of them. These percentages constitute an improvement on those obtained in the Benedikt Sundermann study[15], which obtains a 78.8% success rate for the classification of fibromyalgia in terms of arthritis, albeit with characteristics obtained using fMRI images.

Therefore, we can consider both the psychopathological and medico-social characteristics selected to be clear and concise for the classifier, although it is also true to say that the medico-social ones are not so accurate, with the classifier ultimately being confused by a greater percentage.

These preliminary results using a single classifier need to be explored more specifically, in addition to including more types of classifier in the study and enabling a comparison to be made between them.

The weights of the characteristics should also be studied and an understanding gained as to why there are differences between the psychopathological and medico-social characteristics in the case of the classifier.

REFERENCES

1. Murphy KP; Machine learning: a probabilistic perspective. MIT press, 2012.
2. Yoshida K, Sakurai A; Machine Learning. Bidgoli H (Ed.), Encyclopedia of Information Systems, Elsevier, New York, 2003; 103–114.
3. Centers for Disease Control and Prevention, 2010 Available from: http://www.cdc.gov/arthritis/basics/fibromyalgia.htm, http://www.cdc.gov/arthritis/basics/rheumatoid.htm
4. Salaffi F, Sarzi P, Puttini P, Girolimetti R, Atzeni F, Gasparini S, et al.; Health-related quality of life in fibromyalgia patients: a comparison with rheumatoid arthritis patients and the general population using the SF-36 health survey. Clinical and Experimental Rheumatology, 2009; 27: 67-74.
5. Scott DL, Wolfe F, Huizinga TW; Rheumatoid arthritis. Lancet, 2010; 25(376): 9746.
6. American College of Rheumatology, 2010.

7. Kotsiantis SB, Zaharakis I, Pintelas P; Supervised machine learning: A review of classification techniques. 2007.
8. Swiniarski RW; Rough sets methods in feature reduction and classification. International Journal of Applied Mathematicsand Computer Science, 2001; 11(3): 565–582.
9. Kohavi R; A study of cross-validation and bootstrap for accuracy estimation and model selection.Ijcai, 1995; 2: 1137–1145.
10. Boser BE, Guyon IM,Vapnik VN; A training algorithm for optimal margin classifiers. In Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory ACM Press, Pittsburgh, PA, 1992; pp. 144-152.
11. Vapnik V; Statistical Learning Theory.Wiley, New York, 1998.
12. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D; Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics, 2000; 16(10): 906-914.
13. Derogatis L; SCL-90. Administration, scoring, and proceduresmanual-I for the R (revised) version and other instruments ofthe Psychopathology Rating Scales Series. Chicago: JohnsHopkins University School of Medicine, 1977; 37.
14. Garrett D, Peterson D, Anderson CW, Thaut MH; Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. Neural Systems and Rehabilitation Engineering, IEEE Transactions, 2003 11(2): 141-144.
15. Sundermann B, Burgmer M, Pogatzki-Zahn E, Gaubitz M, Stüber C, Wessolleck E, et al.; Diagnostic classification based on functional connectivity in chronic pain: model optimization in fibromyalgia and rheumatoid arthritis. Academic radiology, 2014; 21(3): 369-377.