

Diagnostic of Pathology on the Vertebral Column with Embedded Reject Option

Ajalmar R. da Rocha Neto¹, Ricardo Sousa², Guilherme de A. Barreto¹, and Jaime S. Cardoso²

¹ Depto. Engenharia de Teleinformática, Universidade Federal do Ceará (UFC),
{ajalmar,guilherme}@deti.ufc.br,

WWW home page: <http://www.deti.ufc.br/~{ajalmar,guilherme}>

² INESC Porto, Faculdade de Engenharia da Universidade do Porto, Portugal,
{rsousa,jaime.cardoso}@inescporto.pt,

WWW home page: <http://www.inescporto.pt/~{rsousa,jsc}>

Abstract. Computer aided diagnosis systems with the capability of automatically decide if a patient has or not a pathology and to hold the decision on the difficult cases, are becoming more frequent. The latter are afterwards reviewed by an expert reducing therefore time consumption on behalf of the expert. The number of cases to review depends on the cost of erring the diagnosis. In this work we analyse the incorporation of the option to hold a decision on the diagnostic of pathologies on the vertebral column. A comparison with several state of the art techniques is performed. We conclude by showing that the use of the reject option techniques is an asset in line with the current view of the research community.

Keywords: Computer Aided Diagnosis System, Pattern Recognition, Support Vector Machines, Reject Option

1 Introduction

Over the last decade, we have been assisting to an increasing number of Machine Learning (ML) techniques, such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN), applied to several medical fields. One reason for this behaviour relies on the capacity of human diagnostic, which is significantly worse than the neural system's diagnostic under adverse conditions, as stress, fatigue and little technical knowledge [2].

In the literature there are some reviews on the application of machine learning techniques in medicine. Some studies regarding to this problem describe several applications over different fields, namely: cardiology, ECG analysis, gastroenterology, pulmonology, oncology, neurology, EEG analysis, Otolaryngology, gynecology and obstetrics, ophthalmology, radiology, pathology, cytology, genetics, biochemistry among others. In another study, related to the use of ANN in the medical and biological areas, it is shown the distribution of a significant number of articles (more than 800) produced in the years 2000 and 2001 in more than 40 countries [7].

Moreover, recent studies show the use of ANNs in the recovery of images of the vertebral column based in contents [8] , in the modality classification of medical images [5] and in the classification of individuals with normal or abnormal osteophyte pathology [7].

Although the use of ML techniques is already widespread in Medicine Diagnosis, in general the application of these techniques in Traumatic Orthopedics is rather sparse in the literature. This fact is due to the absence of numerical attributes that quantitatively describe the pathologies of interest to the field of orthopedics, to generate a suitable database for the design of classifiers [7].

As already stated, the incorporation of ML techniques on medical decision aiding procedures is becoming more frequent. Due to this common acceptance, Dreiseitl et al. [4] studied the credibility that physicians give to decision making support systems. By analysing the physicians' reaction when the system contradicts to their diagnostic, they noticed that specialists are very susceptible to accept the recommendations from these kinds of systems.

However, on complex cases with very similar attributes, these kind of systems can become unreliable. As consequence, the automation of decisions in these situations lead invariably to many wrong predictions. On the other hand, and although items in the historical data are labelled *only* as 'good' or 'bad', 'normal' or 'abnormal' pathology, the deployment of a decision support system in many environments has the opportunity to label critical items for manual revision, instead of trying to automatically classify every and each item. The system automates only those decisions which can be reliably predicted, labelling the critical ones for a human expert to analyse. Therefore, the development of classifiers with a third output class, the reject class, in-between the good and bad classes, is attractive.

This work intends to present an auxiliary system to medical decision aiding. The framework of study to the incorporation of the reject option will be the Intelligent System for Diagnosis of Pathologies of Vertebral Column (SINPATCO) [8]. This framework is composed by three subsystems. Namely, graphical interface, pathologies classification and knowledge extraction.

1.1 Pathologies of the Vertebral Column

The vertebral column is a system composed by a group of vertebrae, intervertebral discs, nerves, muscles, medulla and joints. The main functions of the vertebral column are as follow: (i) human body support axle; (ii) osseous protector of the spine medulla and nervous roots; and (iii) body's movement axles, making movement possible in three levels: frontal, sagittal and transversal.

This complex system can suffer dysfunctions that cause backaches with very different intensities. Disc hernia and spondylolisthesis are examples of pathologies of the vertebral column that cause intense pain. They result of small or several traumas in the column that gradually injures the structure of the intervertebral disc.

Disc hernia appears when the core of the inter-vertebral disc migrates from its place (from the center to the periphery of the disc). Once heading towards

the medullary channel or to the spaces where the nervous roots lie, this leads inevitably to their compression. Spondylolisthesis occurs when one of the 33 vertebrae of the vertebral column slips in relation to the others. This slipping occurs generally towards the base of the spine in the lumbar region, causing pain or symptomatology irritation of the nervous roots. In the following section we will briefly describe characteristics (attributes) that are used to quantitatively describe each patient.

Biomechanical Attributes The database applied in this work was kindly supplied by Dr. Henrique da Mota, who collected it during a medical residence in spine surgery at the *Centre Médico-Chirurgical de Réadaptation des Massues*, placed in Lyon, France. This database contains data about 310 patients obtained from sagittal panoramic radiographies of the spine. From this, 100 patients are volunteers that do not have any pathology in their spines (normal patients). The remaining data are from the patients operated due to disc hernia (60 patients) or spondylolisthesis (150 patients). Therefore, the database is composed of 210 abnormal patients.

Each patient in this database is represented as a vector (or pattern) with six biomechanical attributes, which correspond to the following parameters of the spino-pelvic system: angle of pelvic incidence, angle of pelvic tilt, lordosis angle, sacral slope, pelvic radius and grade of slipping. The correlation between the vertebral column pathologies and this attributes was originally proposed in reference [1].

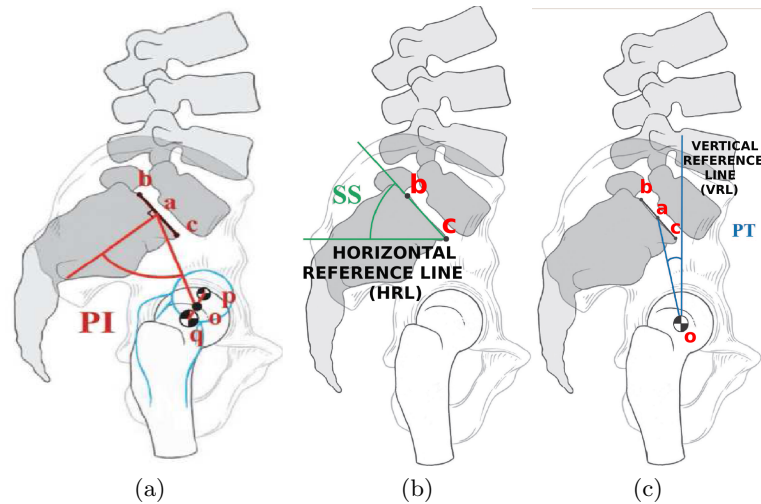


Fig. 1: Spino-pelvic system.

Pelvic incidence (PI) is defined as an angle subtended by line \overline{oa} , which is drawn from the center of the femoral head to the midpoint of the sacral endplate and a line perpendicular to the center of the sacral endplate in Fig. 1a. The sacral

endplate is defined by the line segment \overline{bc} constructed between the posterior superior corner of the sacrum and the anterior tip of the $S1$ endplate at the sacral promontory. For the case when the femoral heads are not superimposed, the center of each femoral head is marked, and a connecting line segment will connect the centers of the femoral heads. Pelvic radius (RP) \overline{ao} will be drawn from the center of this line to the center of the sacral endplate (Fig. 1a).

Lordosis angle is the bigger sagittal angle between the sacrum superior plate and the lumbar vertebra superior plate or thoracic limit. Sacral Slope (SS) is defined as the angle between the sacral endplate (\overline{bc}) and the horizontal reference line (HRL), in Fig. 1b, while Pelvic Tilt (PT) is defined as the angle between the vertical reference line (VRL) and the line joining the middle of the sacral endplate and the axis of the femoral heads in Fig. 1c. It is positive when the hip axis lies in front of the middle of the sacral endplate. Finally, the level of slipping is the percentage level of slipping between the inferior plate of the fifth lumbar vertebra and the sacrum.

The occurrence of pathologies in the vertebral column is conditioned to the morphological types of the pelvis-spine system. The pelvic incidence, being in an elevated level, is conditioned to a higher sacral slope, that generates increasing shear by the increase of the support plan inclination for lumbar lordosis, besides facilitating the conflict of posterior structures, leading to the appearing of a fracture of fatigue in the arc that supports the vertebra and generating a slope called Spondylitics. The low pelvic incidences lead to the contrary effect, with the occurrence of an increasing pressure in the intervertebral disc and facilitate the occurrence of degeneration and disc hernias. The incidence angle determines a normal condition.

The design of automatic classifiers based in biomechanical attributes of real clinical cases allows that linear and/or non-linear relations, as well as their influences in the diagnostic, are captured in a transparent way for the orthopedist, in a way to help him in the decision making. Following, we briefly describe the machine learning models evaluated in this work.

2 Problem Statement and Standard Solutions

Predictive modelling tries to find good rules (models) for guessing (predicting) the values of one or more variables (target) in a dataset from the values of other variables. Our target can assume only two values, represented by ‘normal’ and ‘pathological’ classes. When in possession of a “complex” dataset, a simple separator is bound to misclassify some points. The design of classifiers with reject option can be systematised in three different approaches:

- the design of two, *independent*, classifiers. A first classifier is trained to output \mathcal{C}_{-1} only when the probability of \mathcal{C}_{-1} is high and a second classifier trained to output \mathcal{C}_{+1} only when the probability of \mathcal{C}_{+1} is high.
- the design of a single, standard binary classifier (SBC). If the classifier provides some approximation to the a posteriori class probabilities, then a pattern is rejected if the maximum of the two posterior probabilities is lower

- than a given threshold. If the classifier does not provide probabilistic outputs, then a rejection threshold targeted to the particular classifier is used.
- the design of a single classifier with embedded reject option. This approach has consisted in the design of algorithms specifically adapted for this kind of problems [3, 6, 9].

3 An Ordinal Data Approach for Detecting Reject Regions

The rejection method to be presented—rejoSVM—is an extension of a method already proposed in the literature but for the classification of ordinal data. For more detail about this method, the reader should consult [3, 9]. For completeness, we summarise here the rejoSVM model.

3.1 The Data Replication Method for Detecting Reject Regions

The scenario of designing a classifier with reject option shares many characteristics with the classification of ordinal data. It is also reasonable to assume for the reject option scenario that the three output classes are naturally ordered as $\mathcal{C}_1, \mathcal{C}_{reject}, \mathcal{C}_2$. In the scenario of designing a classifier with reject option, we are interested on finding two boundaries: a boundary discriminating \mathcal{C}_1 from $\{\mathcal{C}_{reject}, \mathcal{C}_2\}$ and a boundary discriminating $\{\mathcal{C}_1, \mathcal{C}_{reject}\}$ from \mathcal{C}_2 .

We proceed exactly as in the data replication method for ordinal data [9]. We start by transforming the data from the initial space to an extended space, replicating the data, according to the rule:

$$\mathbf{x} \in \mathbb{R}^d \begin{cases} \nearrow [\mathbf{x}] \in \mathbb{R}^{d+1} \\ \searrow [\mathbf{x}] \in \mathbb{R}^{d+1} \end{cases}, \text{ where } h = \text{const} \in \mathbb{R}^+$$

If we design a binary classifier on the extended training data, without further considerations, one would obtain the same classification boundary in both data replicas. Therefore, we modify the misclassification cost of the observations according to the data replica they belong to. In the first replica (the extension feature assumes the value zero), we will discriminate \mathcal{C}_1 from $\{\mathcal{C}_{reject}, \mathcal{C}_2\}$; therefore we give higher costs to observations belonging to class \mathcal{C}_2 than to observations belonging to class \mathcal{C}_1 . This will bias the boundary towards the minimisation of errors in \mathcal{C}_2 . Similar approach is conducted for the second replica.

3.2 Selecting the Misclassification Costs

The typical adoption of the same cost for erring and rejecting on the two classes is as follows: assign C_{low} cost when classifying a class as reject and assign C_{high} cost when misclassifying. Therefore, $C_{reject} = \frac{C_{low}}{C_{high}} = w_r$ is the cost of rejecting (normalised by the cost of erring). The data replication method with reject option tries to minimise the empirical risk $w_r R + E$, where R accounts for the rejection rate and E for the misclassification rate.

4 Experiments

The aim of our experimental study is to evaluate the interest of embedded reject options methodologies for aiding the diagnostic of pathology on the Vertebral Column. This dataset is thoroughly described in Section 1.1. For this study, we transformed our three class problem into a binary one. We aggregated the disc hernya and spondylolisthesis pathologies classes into a just one pathology class. The normal class remained unchanged.

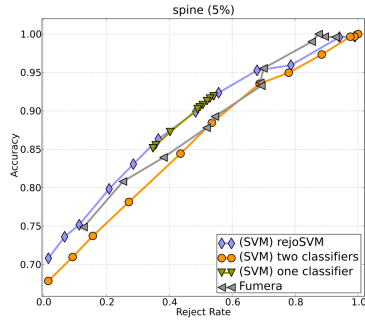
The training dataset was composed in different experiments with 5%, 40% and 80% of the data. The splitting of the data into training and test sets was repeated 100 times in order to obtain more stable results for accuracy by averaging and also to assess the variability of this measure. The best parametrisation of each model was found by ‘grid-search’, based on a 5-fold cross validation scheme conducted on the training set. In our experiments we used a linear kernel on all methods performing the ‘grid-search’ over the C parameter spanning values between $2^{-5}, \dots, 2^5$. The C value is a penalty factor for each point misclassified. Finally, the error of the model was estimated on the test set.

The performance of a classifier with reject option can be represented by the classification accuracy achieved for any value of the reject rate (the so-called Accuracy-Reject curve). The trade-off between errors and rejections depends on the cost of a rejection w_r . This implies that different points of the A-R curve correspond to different values of w_r . We considered values of w_r less than 0.5, as above this value it is preferable to just try to guess randomly.

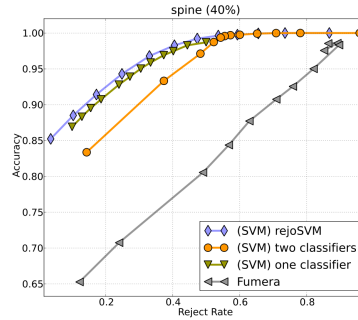
In Fig. 2 we present the results obtained for all used methods. We can first identify that as training size increases, both methods attain similar results with exception with Fumera technique (Fig. 2a-2c). This behaviour can be justified by the fact that the optimisation function is not convex and therefore not attaining the global optimum.

In order to emphasise the benefits of the incorporation of the reject option, we trained five standard learning algorithms. The training and testing evaluation was performed exactly as before. We have selected a SVM (with a linear kernel) with the C parameter encompassing the same range values (baseline learning method for the one and two classifiers). We have also used a SVM with the KMOD kernel [7]. Finally, we also conducted our experiments with a GRNN (General Regression Neural Network) and a MLP [7].

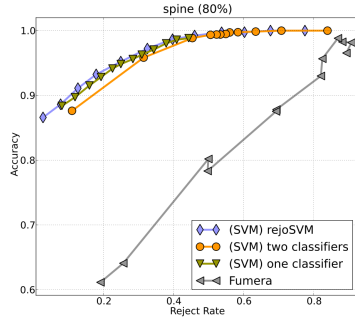
Considering further the performance evaluation function of the embedded reject option learning method, we can fairly compare all techniques. Table 1 shows that standard learning methods provide the baseline results for some of the reject option schemes. Within the SINPATCO context, the incorporation of a reject option can be an asset. Moreover, tools like SINPATCO are designed as decision aiding system which could be used on healthcare offices located on remote areas with limited access to modern resources and funding. In this way, systems with high rates of True Positive (sensitivity) and True Negatives (specificity) are required. Such techniques besides imposing high accuracies rates and a higher confidence on the diagnosis, they also avoid misclassifications. In doing so, there will not be any influence by SINPATCO on the expert to take wrong decisions



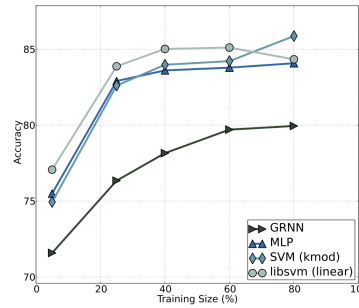
(a) A-R performance curves (Models trained with 5% of the full dataset.)



(b) A-R performance curves (Models trained with 40% of the full dataset.)



(c) A-R performance curves (Models trained with 80% of the full dataset.)



(d) Several learning algorithms for the Spine dataset with the reject option (oSVM technique is detailed in [3]).

Fig. 2: The A-R curves for the Spine dataset with different training sizes.

Training Size	Method \ wr	0.04	0.24	0.48	Method	Accuracy
40%	reoSVM	96.5	87.9	83.5	SVM (linear)	85.0
	one classifier	96.7	87.7	82.1	SVM (KMOD)	83.9
	two classifier	96.2	86.0	76.5		
80%	reoSVM	96.9	89.1	85.2	SVM (linear)	84.3
	one classifier	97.1	88.8	84.4	SVM (KMOD)	85.9
	two classifier	96.6	86.3	82.3		

Table 1: Performance rules according to measure $P = w_r R + E$ in p.p.

which could lead to some interventions (being invasive or not). For instance, on the Fig. 2c, only 50% of the cases would have to be reviewed by the expert while the remaining would be correctly classified. As a final remark, we could verify that reoSVM and two classifiers do not outperform the other. However, and as a feature of this work, reoSVM benefits of simplicity and interpretability which could aid the medical expert in future evaluations.

5 Conclusion

Here we incorporate the reject technique proposed in [9] to the diagnosis of pathologies on the Vertebral Column. This incorporation of the reject option provided to be an asset obtained better results than traditional learning techniques, even when rejecting few instances. Finally, further studies can be developed on the analysis with the incorporation of the reject option on the original problem. Furthermore, a comparison with the ensemble learning technique as proposed in [7] should also be assessed.

Acknowledgments This work has been partially supported by Fundação para a Ciência e a Tecnologia (FCT) - Portugal through project PTDC/EIA/64914/2006 and by CNPq - Brazil through Programa CNPq/Universidade do Porto/590008/2009-9.

References

1. Berthonnaud, E., Dimnet, J., Roussouly, P., Labelle, H.: Analysis of the sagittal balance of the spine and pelvis using shape and orientation parameters. *Journal of Spinal Disorders & Techniques* 18(1), 40–47 (2005)
2. Brause, R.: Revolutionieren neuronale netze unsere vorhersagefahigkeiten. *Zentralblatt für Chirurgie* pp. 692–698 (1999)
3. Cardoso, J.S., da Costa, J.F.P.: Learning to classify ordinal data: the data replication method. *Journal of Machine Learning Research* 8, 1393–1429 (2007)
4. Dreiseitl, S., Binder, M.: Do physicians value decision support? A look at the effect of decision support systems on physician opinion. *Artificial Intelligence in Medicine* 33(1), 25–30 (2005)
5. Florea, F., Rogozan, A., Bensrhair, A., Darmoni, S.J.: Comparison of feature-selection and classification techniques for medical images modality categorization. *Rapport Interne PSI No. 01/FIF* (2004)
6. Fumera, G., Roli, F.: Support Vector Machines with Embedded Reject Option. In: *SVM 2002: Proceedings of the First International Workshop on Pattern Recognition with Support Vector Machines*. pp. 68–82. Springer-Verlag, London, UK (2002)
7. Neto, A.R.R., Barreto, G.A.: On the application of ensembles of classifiers to the diagnosis of pathologies of the vertebral column: A comparative analysis. *IEEE Transactions on Latin America* 7(4), 487–496 (Aug 2009)
8. Neto, A.R.R., Barreto, G.A., Cortez, P.C., da Mota, H.: SINPATCO: Sistema inteligente para diagnóstico de patologias da coluna vertebral. In: *Congresso Brasileiro de Automática*. pp. 929–934 (2006)
9. Sousa, R., Mora, B., Cardoso, J.S.: An ordinal data method for the classification with reject option. In: *Proceedings of The Eighth International Conference on Machine Learning and Applications (ICMLA 2009)* (2009)