journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Computer-aided diagnosis system: A Bayesian hybrid classification method

*F. Calle-Alonso**, *C.J. Pérez, J.P. Arias-Nicolás, J. Martín*

Department of Mathematics, Faculty of Veterinary Medicine, University of Extremadura, Avda. de la Universidad s/n, 10003 Cáceres, Spain

ARTICLE INFO

Article history:
Received 30 November 2012
Received in revised form
29 April 2013
Accepted 26 May 2013

Keywords:
Bayesian methodology
Classification
Computer-aided diagnosis
Relevance feedback

ABSTRACT

A novel method to classify multi-class biomedical objects is presented. The method is based on a hybrid approach which combines pairwise comparison, Bayesian regression and the *k*-nearest neighbor technique. It can be applied in a fully automatic way or in a relevance feedback framework. In the latter case, the information obtained from both an expert and the automatic classification is iteratively used to improve the results until a certain accuracy level is achieved, then, the learning process is finished and new classifications can be automatically performed. The method has been applied in two biomedical contexts by following the same cross-validation schemes as in the original studies. The first one refers to cancer diagnosis, leading to an accuracy of 77.35% versus 66.37%, originally obtained. The second one considers the diagnosis of pathologies of the vertebral column. The original method achieves accuracies ranging from 76.5% to 96.7%, and from 82.3% to 97.1% in two different cross-validation schemes. Even with no supervision, the proposed method reaches 96.71% and 97.32% in these two cases. By using a supervised framework the achieved accuracy is 97.74%. Furthermore, all abnormal cases were correctly classified.

1. Introduction

Pattern recognition solves the problem of assigning each input value or object to one of a given set of classes [1–3]. This research area has become very active in the last several years. Its importance has increased since the development of new Content-Based Image Retrieval (CBIR) methods and the improvement of classification techniques. The growth of information technologies has produced a huge increase of available information, especially images and videos.

Image classification methods are being used in many disciplines, but especially relevant are those applications related to health sciences. In fact, one of the most important pattern recognition application areas is diagnostic imaging in medicine [4]. Many modalities of diagnostic imaging (conventional X-ray, computed tomography, nuclear medicine, magnetic resonance imaging, ultrasound scans, etc.) are especially useful because they help to provide effective diagnoses in a noninvasive way. Besides, during the past several years there has been an important increase in the use of diagnostic medical imaging [5].

Computer-aided diagnosis (CAD) is a broad concept that integrates signal processing, artificial intelligence and statistics into computerized techniques that assist health professionals in their decision-making processes. These techniques seek to maximize the information that may be automatically extracted from medical images via objective and quantitative computations. Due to the high volume of images and the amount of information currently provided,

* *Corresponding author*. Tel.: +34 610764510.
E-mail address: fcalonso@unex.es (F. Calle-Alonso).
0169-2607/$ – see front matter
http://dx.doi.org/10.1016/j.cmpb.2013.05.029

CAD systems have become powerful tools to assist physicians in achieving high efficiency [6]. They have greatly increased the knowledge of normal and diseased anatomy, and can be determinant in diagnosis and treatment planning.

CAD systems play an important role in the early detection of diseases. For example, accurate and early diagnoses of Alzheimer's disease is key for the development of effective treatments that can palliate the effects of this neurodegenerative disease. Ref. [7] presented a CAD system for the early detection of Alzheimer's disease by means of single-photon emission computed tomography. Refs. [8–10] and references therein also consider CAD systems that use different types of images and classification algorithms for early detection of Alzheimer's disease. Diagnosing osteoporosis can also be benefited from these computerized techniques. Thoracic and lumbar vertebrae are the most common sites of osteoporosis-related fractures. It is very important to detect vertebral fractures as early as possible because timely pharmacologic intervention can reduce the risk of subsequent additional fractures. Ref. [11] developed a CAD system to detect vertebral fractures by using lateral chest radiographs. Ref. [12] retrospectively evaluated the usefulness of CAD systems to radiologist performance in the detection of vertebral fractures and lung nodules on chest radiographs. There are many other diseases that CAD systems may help to detect. In fact, they can help to diagnose any disease that can be detected from images or other biological signals by automatically extracting features [13–15].

A broad vision of CAD is not limited to image analysis. For example, certain specific characteristics obtained from voice recordings can be useful in diagnosing speech disorders. Numerous techniques for automatic evaluation of speech disorders have been proposed in the last several years [16]. Voice recordings have also been used to diagnose Parkinson's disease. Ref. [17] developed new speech signal processing algorithms to classify Parkinson's disease patients. The Parkinson's Voice Initiative[1] has opened the possibility to advance in the diagnosis of the disease by using a large database obtained by automatized telephone calls.

Both the feature extraction processes (pre-processing) and the classification techniques are currently challenges for the development of efficient CAD systems [15]. The pre-processing step provides the necessary ingredients to apply the classification algorithms. Sometimes dimensional reductions are performed [18,19]. There is a wide variety of available classification methods that have been used for CAD systems, for example, linear discriminant [20], logistic regression [21], Support Vector Machine (SVM) [22], segmentation techniques [23], k-nearest neighbor [24], and artificial neural networks [25], among others. Sometimes conceptually simple algorithms are considered, whereas other times more sophisticated methods are developed. However, there is no global best classifier, but some work better than others in specific contexts. The important issue is that the classifier can be properly integrated in a CAD system and that high levels of accuracy are obtained.

In spite of the advances in this field, there are still few powerful methods that consider learning processes and even fewer that apply the Bayesian methodology in a relevance feedback framework. The power of this methodology remains to be exploited because learning processes can be determinant for the classification results. Most of the developed Bayesian classification methods have been focused on binary classification, although multi-class problems have also been tackled. Ref. [26] classified patients in two classes (with or without Alzheimer's disease), using a Bayesian classifier without learning. The number of available objects was much lower than the dimension of the feature space, so first of all they applied a principal component analysis. Ref. [27] classified renal cell carcinoma (four classes) with a Bayesian approach without learning. Ref. [28] use several classification and regression methods, including Bayesian networks, and meta-learning algorithms such as bagging [29] and AdaBoostM1 [30]. These methods have been applied for pancreatic cancer detection in a multi-class setting, showing that Bayesian techniques provided the best overall performance. Ref. [31] test the Bayesian classifier versus classical logistic regression, finding out that the former performs better than the latter for automatic classification of polycystic ovary syndrome (two classes). Ref. [32] classified patients with and without depressive disorder by using neuroimaging scans with two machine learning techniques: relevance vector machine (Bayesian) and SVM (classical).

In this work, a novel Bayesian hybrid classification method that can be used with and without relevance feedback is presented. This method is specially indicated to classify biomedical objects containing a high number of characteristics and where there are not many objects in each class. Cancer classification problems are one of the interesting applications. By combining the power of three different techniques (pairwise comparison, Bayesian regression, and k-nearest neighbors (KNN)), this method provides a high accuracy in classification. The method can run in a fully automatic way or it can incorporate relevance feedback. In the latter case, the information obtained from both an expert and the automatic classification is used to improve the results through a relevance feedback framework. Although the applications presented in this paper focus on two concrete medical diagnostic problems, the approach is useful for many other purposes. The only limitation is that the features of the biomedical objects must have enough information to allow a discrimination.

The outline of the rest of the paper is as follows. Section 2 presents the approach by including a general description, the information about the features, the algorithm involved and the relevance feedback. In order to illustrate the applicability of the method, two different problems in the area of medical diagnosis are presented in Section 3. Finally, some conclusions are summarized in Section 4.

2. The approach

A novel hybrid classification approach is proposed. The interest is focused on classifying objects in several classes according to a similarity criterion.

The method can be summarized as follows. Firstly, a set of correctly classified objects are considered as a training dataset. The objects have features, represented by numerical vectors, that have been previously extracted. Then, a pairwise

[1] http://www.parkinsonsvoice.org/.