Computers in Biology and Medicine 46 (2014) 61–70

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Pairwise FCM based feature weighting for improved classification of vertebral column disorders

Yavuz Unal [a], Kemal Polat [b,*], H. Erdinc Kocer [c]

[a] Education Faculty, Amasya University, Amasya, Turkey
[b] Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Abant Izzet Baysal University, 14280 Bolu, Turkey
[c] Technology Faculty, Department of Electrical and Electronics, Selcuk University, Konya, Turkey

ARTICLE INFO

Article history:
Received 7 October 2013
Accepted 6 December 2013

Keywords:
Vertebral column
Pairwise Fuzzy C-means clustering based feature weighting
Classification
Data pre-processing

ABSTRACT

In this paper, an innovative data pre-processing method to improve the classification performance and to determine automatically the vertebral column disorders including disk hernia (DH), spondylolisthesis (SL) and normal (NO) groups has been proposed. In the classification of vertebral column disorders' dataset with three classes, a pairwise fuzzy C-means (FCM) based feature weighting method has been proposed. In this method, first of all, the vertebral column dataset has been grouped as pairwise (DH-SL, DH-NO, and SL-NO) and then these pairwise groups have been weighted using a FCM based feature set. These weighted groups have been classified using classifier algorithms including multilayer perceptron (MLP), *k*-nearest neighbor (*k*-NN), Naive Bayes, and support vector machine (SVM). The general classification performance has been obtained by averaging of classification accuracies obtained from pairwise classifier algorithms. To evaluate the performance of the proposed method, the classification accuracy, sensitivity, specificity, ROC curves, and *f*-measure have been used. Without the proposed feature weighting, the obtained *f*-measure values were 0.7738 for MLP classifier, 0.7021 for *k*-NN, 0.7263 for Naive Bayes, and 0.7298 for SVM classifier algorithms in the classification of vertebral column disorders' dataset with three classes. With the pairwise fuzzy C-means based feature weighting method, the obtained *f*-measure values were 0.9509 for MLP, 0.9313 for *k*-NN, 0.9603 for Naive Bayes, and 0.9468 for SVM classifier algorithms. The experimental results demonstrated that the proposed pairwise fuzzy C-means based feature weighting method is robust and effective in the classification of vertebral column disorders' dataset. In the future, this method could be used confidently for medical datasets with more classes.

1. Introduction

Lower back pain is among the most frequent ailments in the world. It is the second commonest illness after headaches. Besides lowering quality of life, it also causes labor loss and early retirement [1]. This ailment brings huge burdens to the relatives of ill people and to the economies of countries. The fact that the imaging techniques that are used in the diagnosis of this ailment are high-cost and the radiologists who can diagnose this ailment are few in number make it more important for this ailment to be treated in a computer-assisted manner [2].

Lower back pain is a very common ailment in adults. About 90% of all adults have at least one lower back pain attack in their whole lives. If all adults were questioned at the same time, it would become obvious that 15% of them have lower back pain [3]. Lower back pain may be due to trauma, lifting heavy objects or a reverse action; however, it may not be due to a certain cause. Moreover, some cancer types may also cause lower back pain; however, this is rare. Again, brucella and some similar infections may also appear together with lower back pain. For all these reasons, lower back pain should be taken serious and the diagnosis should be fast and accurate [4].

Today, the plain radiography in the research of lower back pain has become less important. The patient does not receive any X-rays and the superior imaging ability in various plans; and its being able to image the spinal cord and other soft tissues clearly makes magnetic resonance become more important day by day.

The magnetic resonance imaging method makes regular diagnosis and discriminatory diagnosis possible. Although magnetic resonance is a very useful method, the assessment of the images requires great experience. Wrong comments on the images lead to wrong treatments. If the light disk chamber, which is very common, is assessed as a herniated disk, the way of treatment will go far away from the required treatment [5,6].

In this paper, we have used the vertebral column dataset taken from UCI (University of California, Irvine) machine learning [7]. In this dataset, there are six biomechanical features and 310 data

* Corresponding author. Tel.: +90 530 561 9226; fax: +90 374 253 45 58.
E-mail address: kemal_polat2003@yahoo.com (K. Polat).

0010-4825/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2013.12.004

including three classes including disk hernia (DH), spondylolisthesis (SL), and normal (NO). These biomechanical features are derived from the shape and orientation of the pelvis and lumbar spine. The names of these features are pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius and grade of spondylolisthesis [8].

There have been several studies on classification and determination of vertebral column diseases in the literature. These are as follows: Neto and Barreto applied standalone support vector machine (SVM), multiple layer perception (MLP) and generalized regression neural network (GRNN) which are among machine learning methods to UCI pathologies of the vertebral column dataset (VCP dataset). They then applied a combination of these machine learning (ML) methods and released the results in a comparative method after classifying their work [8]. Mattos and Barreto introduced two new ensemble methods in their work. One of these methods is fuzzy adaptive resonance theory (FA), and the other one is self-organizing map (SOM) neural networks based classifiers. Using many datasets including VCP dataset, they released comparative results [9]. Neto et al. called their study as "Reject Option" and applied VCP dataset in this study where they introduced a new technique. They used several different ML classifiers in their study [10]. Abdrabou designed a hybrid system in his study and joined the two machine learning (ML) techniques: case-based reasoning (CBR) and artificial neural network. This system was applied to the UCI vertebral column dataset. The hybridization of CBR and ANN showed that the classification accuracy increased [11]. Ansari et al. diagnosed disease through machine learning classifiers to vertebral column dataset taken from UCI. These classifiers are: feed forward back propagation neural network, generalized regression neural network and support vector machine. They achieved 93.87% classification accuracy with feed forward back propagation neural network [12].

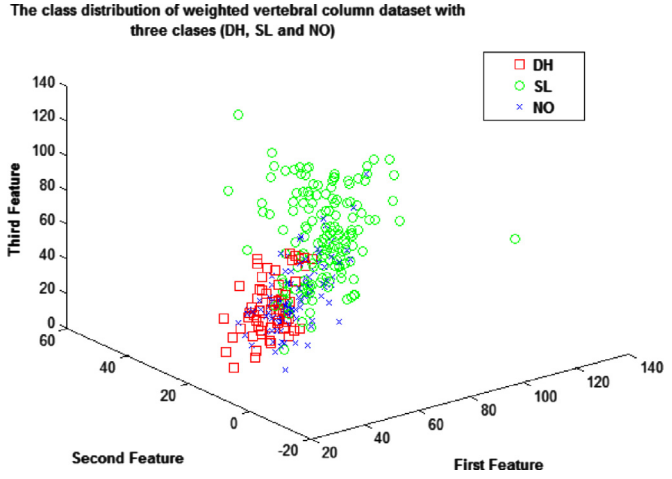

Fig. 1. The class distribution of raw vertebral column disorders' dataset with three classes (DH, disk hernia; SL, spondylolisthesis; NO, normal).

In this study, a novel data pre-processing method called the pairwise fuzzy C-means based feature weighting method has been proposed both to determine of type of vertebral column disorders including disk hernia group, spondylolisthesis group, and normal group to improve the classification accuracy of used classifier algorithms in the classification of vertebral column disorders' dataset. As seen from the literature review, the proposed feature weighting method is firstly proposed and also applied to this dataset by us. The proposed hybrid system consists of two stages: in the first stage (feature weighting procedure), each feature in the vertebral column dataset has been weighted to transform from a non-linearly separable case to a linearly separable case by pairwise fuzzy C-means based feature weighting. In the second stage, the weighted vertebral column dataset that has a linearly separable distribution is classified by classifier algorithm including multilayer perceptron (MLP), *k*-nearest neighbor (*k*-NN), Naive Bayes, and support vector machine (SVM). Before applied the pairwise fuzzy C-means based feature weighting (PFCMBFW) to vertebral column dataset, the obtained classification accuracies were 78.3871, 85.4839, 83.2258, and 81.6129 by *k*-NN, MLP, Navie Bayes, and SVM classifiers, respectively. After applied the PFCMBFW to this dataset, the obtained classification accuracies were 95.4839, 96.7742, 97.4194, and 96.4516 by *k*-NN, MLP, Navie Bayes, and SVM classifiers, respectively. The obtained results have shown that the proposed feature weighting method could be confidently used to classify the vertebral column disorders.

The organization of the paper is as follows: the dataset used in this paper is explained in Section 2. A detailed description of the methodology adopted in this paper is described in Section 3 which includes feature extraction, feature weighting methods and classifiers. Section 4 explains the interpretation of the results and discussion. Section 5 concludes the paper.

2. Material: vertebral column dataset

In this study, the used vertebral column with three classes (pathologies of the vertebral column dataset) was formed by Dr. Henrique Mota et al. [7]. In this dataset, there are 310 samples consisting of 100 normal group, 60 disk hernia group, 150 spondylolisthesis group and six biomechanical features called pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius and grade of spondylolisthesis. These used features were derived from the shape and orientation of the pelvis and

Table 1
The statistical evolution of raw and weighted vertebral column datasets.

The case of dataset	The name of feature in dataset	Min. value	Max. value	Mean value	Std. dev. value	The correlation coefficient between feature and class label
Raw	Pelvic_incidence	26.148	129.834	60.497	17.237	−0.029
Weighted		18.58	195.32	69.27	40.75	−0.167
Raw	Pelvic_tilt	−6.555	49.432	17.543	10.008	−0.211
Weighted		−9.84	74.26	20.55	16.86	−0.372
Raw	Lumbar_lordosis_angle	14	125.742	51.931	18.554	0.036
Weighted		8.45	212.31	64.69	46.69	−0.131
Raw	Sacral_slope	13.367	121.43	42.954	13.423	0.12
Weighted		7.94	182.79	49.81	29.43	−0.026
Raw	Pelvic _radius	70.083	163.071	117.92	13.317	0.234
Weighted		65.86	153.27	116.63	15.08	0.202
Raw	Degree_spondylolisthesis	−11.058	418.543	26.297	37.559	−0.119
Weighted		−0.84	5859	351.58	533.107	−0.120