Computers in Biology and Medicine 62 (2015) 196–205

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Semiautomatic computer-aided classification of degenerative lumbar spine disease in magnetic resonance imaging

Silvia Ruiz-España [a], Estanislao Arana [b], David Moratal [a,*]

[a] Center for Biomaterials and Tissue Engineering, Universitat Politècnica de València, 46022 Valencia, Spain
[b] Radiology Department, Fundación Instituto Valenciano de Oncología, Valencia, Spain

ARTICLE INFO

Article history:
Received 16 December 2014
Accepted 16 April 2015

Keywords:
Lumbar intervertebral discs
Disc degeneration
Herniation
Lumbar spinal stenosis
Segmentation
Reproducibility

ABSTRACT

Background: Computer-aided diagnosis (CAD) methods for detecting and classifying lumbar spine disease in Magnetic Resonance imaging (MRI) can assist radiologists to perform their decision-making tasks. In this paper, a CAD software has been developed able to classify and quantify spine disease (disc degeneration, herniation and spinal stenosis) in two-dimensional MRI.
Methods: A set of 52 lumbar discs from 14 patients was used for training and 243 lumbar discs from 53 patients for testing in conventional two-dimensional MRI of the lumbar spine. To classify disc degeneration according to the gold standard, Pfirrmann classification, a method based on the measurement of disc signal intensity and structure was developed. A gradient Vector Flow algorithm was used to extract disc shape features and for detecting contour abnormalities. Also, a signal intensity method was used for segmenting and detecting spinal stenosis. Novel algorithms have also been developed to quantify the severity of these pathologies. Variability was evaluated by kappa (k) and intra-class correlation (ICC) statistics.
Results: Segmentation inaccuracy was below 1%. Almost perfect agreement, as measured by the k and ICC statistics, was obtained for all the analyzed pathologies: disc degeneration (k=0.81 with 95% CI=[0.75..0.88]) with a sensitivity of 95.8% and a specificity of 92.6%, disc herniation (k=0.94 with 95% CI=[0.87..1]) with a sensitivity of 60% and a specificity of 87.1%, categorical stenosis (k=0.94 with 95% CI=[0.90..0.98]) and quantitative stenosis (ICC=0.98 with 95% CI=[0.97..0.98]) with a sensitivity of 70% and a specificity of 81.7%.
Discussion: The proposed methods are reproducible and should be considered as a possible alternative when compared to reference standards.

1. Introduction

Nowadays, lumbar disc degeneration, herniation and spinal stenosis are very common entities that affect millions of people, causing lower back pain (LBP) which can restrict mobility and interfere with daily routine of posture [1]. Degenerative changes as loss of disc height or osteophyte formation and degenerative disc herniation cause most cases of lumbar spinal central and lateral stenosis (>90%) [2,3]. So, approximately one in every 1000 individuals over the age of 65 undergoes laminectomy surgery annually for spinal stenosis [2,4]. One-third of adults over the age of 20 show evidence of herniated discs [5] and 90% of herniations occur in the lumbar and lumbosacral regions of the spine [6]. Although imaging techniques have limitations, magnetic resonance imaging (MRI) is the preferred modality for the accurate diagnosis of intervertebral disc pathology and spinal stenosis [7]. As inter-rater agreement among radiologists is often moderate, reliable methods to quantify and classify these entities are needed [5,7]. Also, a demand of computer-aided diagnosis (CAD) methods has increased in the past decade as a way to reduce radiologist workload in the imaging diagnosis of lower back pain [8] and improve repeatability.

Several algorithms have been developed, with variable success. For disc degeneration, some studies measured the T1, T1ρ and T2 relaxation times and the apparent diffusion coefficient (ADC) [9–11]. However, their main drawback is long acquisition times and the necessity of specific image acquisition protocols, which makes its routine clinical use difficult, showing also controversial outcomes relationship [7,12]. In addition, other studies based on shape, context, intensity, and texture information differentiated only between normal and degenerated discs, without specifying degeneration grade [13–15]. By contrast, several approaches used classification systems [12,16,17], such as Pfirrmann classification.

* Correspondence to: Center for Biomaterials and Tissue Engineering, Universitat Politècnica de València, Camí de Vera, s/n, 46022 Valencia, Spain. Tel.: +34 96 387 70 07x88939; fax: +34 96 387 72 76.
E-mail address: dmoratal@eln.upv.es (D. Moratal).

http://dx.doi.org/10.1016/j.compbiomed.2015.04.028
0010-4825/

Pfirrmann classification is a 5-level grading system for classifying the severity of disc degeneration. Disc degeneration is graded using MRI images to evaluate the homogeneity of disc structure, signal intensity, differentiation between nucleus and annulus and disc height. This information is converted into five grades, being considered as grade I when signal intensity is homogeneous and bright, and there is a clear distinction between nucleus and annulus and height is normal. As grade V shows low signal intensity, it is not possible to differentiate between nucleus and annulus and disc space is collapsed [6]. However, most of these approaches, based on Pfirrmann classification, were not capable to distinguish between grades IV and V [12,16]. An early degenerative change normally seen in MRI is a decrease of the mean intensity on T2 sequences, so this is a widely accepted and the most common method for disc degeneration classification [7], and it has been taken into account to develop the method presented in this work.

There are not many clinically useful CAD systems for the detection of disc herniations in the lumbar area. Some of these studies proposed methods based mainly on geometrical features (shape size and location) [18–21]. However, the majority of them only distinguished discs as normal or abnormal [18,21], or between bulging and herniation [19]. Tsai et al. [20] developed a method able to distinguish among bulging, protrusion, extrusion and separation but it was only patented for educational purposes. In this work, we present a method, also based on shape features, capable of classifying among normal, focal-based protrusion, broad-based protrusion and extrusion, regardless of their location.

Methods developed to diagnose lumbar spinal stenosis with standard definitions are limited [21–24] and even quantitative techniques are scarcer [2]. In the method presented in this work, it is possible not only to detect but also to quantify spinal stenosis.

In addition, existence of one abnormality provokes the development of other abnormalities [1,2]. So, CAD systems available to detect several related pathologies would be very useful for clinical routine. However, combined methods to detect several pathologies are rarely reported [15,21], and they are only capable of classifying the discs as normal or abnormal. Another drawback is the absence of the gold standard, only obtained in cadaver specimens [25,26]. Currently, the only accepted source for the definition of a ground truth (GT) is based on signal intensities and boundary markings performed by expert radiologists [27], absent in several spine imaging studies [2,5,7]. Therefore, grading classifications, as Pfirrmann, for disc degeneration [6] and disc contour [28] according to standardized nomenclature are encouraged [29]. The purpose of supervised learning on a CAD system is to deduce a functional relationship from validating data that generalizes well to unknown data. This requisite is not always accomplished in diagnostic imaging studies [5,7,30,31].

In this work, we restrict ourselves to two-dimensional MRI as it is the gold standard in clinical practice [2], due to its relative simplicity and low computational requirements. The purpose of this paper is to present a three-way CAD methodology for detecting and quantifying degenerative disc disease, according to Pfirrmann classification, disc contour abnormalities and spinal stenosis with minimal user input. To the best of our knowledge, no full CAD system is available to detect and also to quantify spinal stenosis.

2. Materials and methods

2.1. Subjects

For validating and testing, 14 (9 male, 5 female) and 53 (25 male, 28 female) subjects, respectively, were randomly selected among patients referred to lumbar MRI in our Radiology Department for LBP and/or sciatica in 2013, and also randomly assigned to each group. All were assessed by visual analog pain scale (VAS, range 0–10). There were no statistical differences regarding age, gender, or VAS pain scale between both groups (t-test, $p=0.12$). Their characteristics are shown in Table 1.

All lumbar intervertebral levels were selected for Pfirrmann's grade analysis. Disc contour was studied where axial images were obtained; as in clinical practice discs observed by MRI technologist as normal in sagittal images are not explored in axial sequences. Prevalence of disc degeneration was available for the 70 discs used for validation and for the 265 discs used for testing. Its global value was 31.5% (32.3% for the validation cases and 31.1% for the test cases) (Table 1). No statistical differences were found between both datasets (ANOVA F:1.04, $p=0.08$). Disc contour abnormalities were validated in 52 discs of the former group and tested in 180 discs of the testing dataset (Table 1).

2.2. Magnetic resonance imaging

All examinations were performed on a 1.5-T MRI (Siemens Symphony, Erlangen, Germany) with a 6-channel phased-array spine coil. Same image acquisition protocols were used for validating and testing. Common sequences usually used for detecting spinal pathology, axial and sagittal T2-weighted were used in this study without fat suppression [32]:

- Sagittal T2-weighted turbo spin echo 2896–3300 ms/102–120 ms (TR/effective TE), 416–576 × 448–1024 matrix, 270 mm field of view, 11 slices of 4 mm thickness and a pixel spacing of [0.4492–0.8203] × [0.4492–0.8203], 2 acquisitions, 12 echo train length.
- Axial T2-weighted turbo spin echo 2896–3040 ms/103–120 ms (TR/effective TE), 256–512 × 256–512 matrix, 180 mm field of view, 15 slices of 4 mm thickness and a pixel spacing of a [0.3906–0.8594] × [0.3906–0.8594], 3 acquisitions, 5 echo train length. Slices were placed in the plane of the five lower discs.

2.3. Disc and spinal stenosis qualitative classification

Qualitative classification of disc degeneration based on Pfirrmann grading system was made by an experienced radiologist (15 years' experience in spine imaging). Discs were classified into 5 grades (from grade I: normal disc to grade V: collapsed disc space).

Table 1
Characteristics of patients included in the study. M, male; F, female; LBP, low back pain; VAS, visual analog scale; y, years.

	Validation group (14 patients)	Testing group (53 patients)
Gender	9 M/5 F	25 M/28 F
Age (y)[a]	46.1 ± 13.7	47.3 ± 12.7
LBP intensity (VAS)	6.1 ± 1.2	6.3 ± 1.7
Disc degeneration (Pfirrmann's grade) (%)	70 discs	265 discs
I	–	–
II	7.14	7.2
III	21.3	21.6
IV	42.8	43
V	28.8	28.2
Disc contour (%)	52 discs	180 discs
Normal	28.8	29.4
Bulging	48.1	50.5
Herniation	23.1	20.1
Spinal stenosis (yes)	50	48.8

[a] Mean and standard deviation.