ORIGINAL ARTICLE

Compression fracture diagnosis in lumbar: a clinical CAD system

**Samah Al-Helo · Raja S. Alomari · Subarna Ghosh ·
Vipin Chaudhary · Gurmeet Dhillon ·
Moh'd B. Al-Zoubi · Hazem Hiary · Thair M. Hamtini**

Abstract

Purpose Lower back pain affects 80–90 % of all people at some point during their life time, and it is considered as the second most neurological ailment after headache. It is caused by defects in the discs, vertebrae, or the soft tissues. Radiologists perform diagnosis mainly from X-ray radiographs, MRI, or CT depending on the target organ. Vertebra fracture is usually diagnosed from X-ray radiographs or CT depending on the available technology. In this paper, we propose a fully automated *Computer-Aided Diagnosis System (CAD)* for the diagnosis of vertebra wedge compression fracture from CT images that integrates within the clinical routine.

Methods We perform vertebrae localization and labeling, segment the vertebrae, and then diagnose each vertebra. We perform labeling and segmentation via coordinated system that consists of an *Active Shape Model* and a *Gradient Vector Flow* Active Contours (GVF-Snake). We propose a set of clinically motivated features that distinguish the fractured vertebra. We provide two machine learning solutions that utilize our features including a supervised learner (Neural Networks (NN)) and an unsupervised learner (K-Means).

Results We validate our method on a set of fifty (thirty abnormal) Computed Tomography (CT) cases obtained from our collaborating radiology center. Our diagnosis detection accuracy using NN is 93.2 % on average while we obtained 98 % diagnosis accuracy using K-Means. Our K-Means resulted in a specificity of 87.5 % and sensitivity over 99 %.

Conclusions We presented a fully automated CAD system that seamlessly integrates within the clinical work flow of the radiologist. Our clinically motivated features resulted in a great performance of both the supervised and unsupervised learners that we utilize to validate our CAD system. Our CAD system results are promising to serve in clinical applications after extensive validation.

Keywords Vertebrae fracture · Computed Tomography (CT) · K-Means · Neural Network · Active Shape Model (ASM)

S. Al-Helo · R. S. Alomari (✉) · M. B. Al-Zoubi ·
H. Hiary · T. M. Hamtini
University of Jordan, Amman 11942, Jordan
e-mail: ralomari@buffalo.edu

R. S. Alomari · S. Ghosh · V. Chaudhary
University at Buffalo, SUNY, Buffalo, NY 14260, USA

G. Dhillon
Proscan Imaging Inc., Williamsville, NY, USA

Introduction

Computer-Aided Diagnosis (CAD) systems have been attracting wide range of researchers. They aim at helping radiologists before, during, and after the diagnosis of diseases from various medical imaging modalities including Computed Tomography (CT), X-ray radiographs, digital mammography, Magnetic Resonance Imaging (MRI), etc. Computer-Aided Detection and Computer-Aided Diagnosis are differentiated in the literature despite they are both referred by CAD. The former refers to the system that identifies suspicious features in an image and brings them to the attention of the radiologist, while the later refers to the system that estimates the likelihood that the image is abnormal.

Many CAD systems exist and are approved for clinical and (or) research use for various medical imaging modalities and various organs for the human body. We work on the lower back and, specifically, design a CAD system for lumbar vertebrae. Lower back pain is a common condition that affects 80–90 % of people at some point during their life. It is considered

as one of the most common reasons people visit the doctor or miss work. It comes in many forms, lower-, middle-, or upper-back pain. However, most cases of back pain are associated with pain and stiffness in the lower back "lumbar area" because lumbar area is responsible for the major body load.

Disease in the lumbar area might occur either in the disc, the vertebrae, or the soft tissues. Any of these diseases cause high volume of pain that irritates to the knees and might cause immobility of the patient. We work on lumbar vertebra fractures and, specifically, diagnosis of the most common traumatic fracture and wedge compression fracture. Clinical practice for lumbar vertebrae diagnosis utilizes X-ray radiographs and/or CT depending on the available modality, the severity of the disease, the target of the disease, and even the allowed insurance expenses. X-ray radiographs are usually the initial diagnostic modality because it is cheap, available, and give an indication of the target abnormal organ. Usually, the X-ray test recommends a CT when a bone fracture is suspected. However, in some cases, X-ray might be enough to diagnose simple and specific disease if neither of the other modalities is available. In this paper, we utilized CT images to diagnose wedge compression fracture because CT is the clinical standard for such fracture condition in the lumbar area.

There are two main steps that the radiologist follow during the work flow of diagnosis of lumbar area: (1) detection and labeling of the vertebrae and (2) diagnosis of each vertebra level by specifying all possible abnormalities in that vertebra. In diagnosis, there are many pieces of information that the radiologists utilize before making the decision such as age, height, history, and type of pain. They also utilize the relative context of the vertebra.

In this paper, we propose a fully automated CAD system for wedge compression fracture diagnosis from CT images. We initially detect, label, and segment the five lumbar vertebrae and then automatically detect the wedge fracture, if exists, based on a set of clinically motivated features that we extract based on our experience and consulting with our collaborating radiologist. We utilize our previous work in the first two steps, namely: detection and labeling [2], and vertebrae segmentation [1]. In this paper, we present the diagnosis of the vertebra wedge fracture.

The rest of this paper is organized as follows: section "Related work" reviews the literature; section "Available data" presents the available data. We then discuss the materials and methods in section "Materials and methods". We present our results and comparative study in section "Results" and conclude in section "Conclusion".

Related work

Vertebrae fracture is one of the main disorders that affects lumbar region of the spine. It is caused by many reasons such as violent trauma, car accidents, frequent flexion of the lower back, and jumps or falls from heights. However, other major causes are due to disc abnormalities such as disc *Herniation*, *Desiccation*, *Budgling*. We refer the reader for our major previous efforts in CAD systems for abnormalities in discs within the lumbar area [3].

Clinical classification for the fracture types is not yet of agreement among all radiologists. However, many classifications exist such as Eastell et al. [7] who classify the vertebral fractures by two standards: Type of deformity (including wedge, biconcavity, and compression) and degree of deformity (grades 1 and 2).

Wedge fractures are the most common type of lumbar fracture [8]. Figure 1 shows a model for the lumbar area with wedge fracture to show the deformity of such vertebra. Many other types of fracture exist and we refer the reader to [8] for an exhaustive survey of various clinical conditions.

Most of the literature working on the detection of vertebrae fracture works on X-rays due to many reasons including the availability of such clinical data. Furthermore, most existing literature detects and segments vertebrae for automating the diagnosis of *Osteoporosis* which is diagnosed from X-ray radiographs or dual X-ray radiographs as the clinical standard and for bone mass estimation.

Smyth et al. [13] used lateral dual X-ray absorptiometry scans to statistically model the shape and appearance of the spine with an Active Shape Model (ASM) to quantify the bone mass to diagnose *Osteoporosis*. The technique obtained entire shape information, and the segmentation they found was comparable to manual segmentation but the lower lumbar and upper thoracic has more error than the rest of vertebrae.
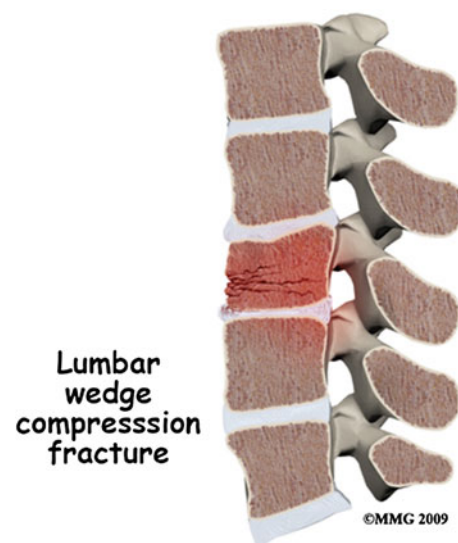

Fig. 1 Lumbar wedge compression fracture. Image used courtesy of Medical Multimedia Group, LLC. More information is available at eOrthopod.com [8]

Roberts et al. [12] presented a method for helping in early diagnosis of *Osteoporosis* and its clinical trials treatments. They used an ASM to detect and quantify vertebral fracture from X-ray radiographs for the lumbar and thoracic area (L4 up to T7) using extracted shape and appearance features for performing quantitative fracture classification. Because of differences in vertebrae, they trained a shape model for each of three classes: upper thoracic (T7–T9), lower thoracic (T10–T12), and lumbar (L1–L4). They presented a comparison study between appearance and shape effect on classification in each vertebral group.

Mastmeyer et al. [11] developed a new hierarchical 3D technique to segment the vertebral bodies in order to measure bone mineral density (BMD) with high trueness and precision in volumetric CT datasets. The tests were analyzed using phantom scans, and intra- and inter-operator precision errors of the segmentation procedure were analyzed using existing clinical patient datasets. Results for segmented volume, BMD, and coordinate system position were below 2.0 %, 0.6 %, and 0.7 %, respectively.

Tan et al. [15] developed an algorithm using high-resolution CT images that provide quantitative measures of the *Syndesmophytes* where this abnormal bone structures grow at intervertebral disc spaces. The algorithm first segments the whole vertebral body using a 3D multi-scale cascade of successive level sets, and then, it extracts the continuous ridge line of the vertebral body where *Syndesmophytes* are located. The third part of the algorithm segments the *Syndesmophytes* from the vertebral body using local cutting planes and quantifies them. They tested the algorithm with ten abnormal 3D CT scan images and compared the results with a medical expert. Correlation between the two evaluations was found to be 90 %.

Cherukuri et al. [5] presented an image processing technique using X-ray radiographs to study the bony growth on vertebrae: *Osteophytes*. For individual vertebra analysis, manual vertebral segmentation is performed. They used convex hull-based features to highlight anterior *Osteophytes*. They tested their work on 714 X-ray radiographs and achieved an average accuracy of 86.6 %.

Kasai et al. [10] developed a computerized method for detection of vertebral fractures on lateral chest X-ray radiographs in order to assist radiologists' image interpretation and thus allow the early diagnosis of *Osteoporosis*. They used 20 patients with severe vertebral fractures and 118 patients without fractures. The sensitivity of their computerized method for detection of fracture cases was 95 % (19/20), with 1.03 (139/135) false-positive fractures per image. The accuracy of identifying vertebral end plates, marked by radiologists in a *morphometric* study, was 76.6 % (400/522) and 70.9 % (420/592) for cases used for training and those for testing, respectively.

Most recently, our work in [9] presented a fully automated method for robustly localizing and segmenting the vertebrae for preparation of vertebral fracture diagnosis. However, the amount of data and technique is different from our proposed work in this paper. The main steps are as follows: (1) Localization of the intervertebral discs, (2) Localization of the vertebral skeleton, (3) Segmentation of the individual vertebra, (4) Detection of the vertebrae center line, and (5) Detection of the vertebrae major boundary points. We used five classifiers to detect the wedge fractures. Segmentation results achieved an average error of 1.5 mm on 50 clinical CT and their classification accuracy was 97.33 %. In this paper, we perform segmentation using ASM and GVF-snake, and then, we obtain new shape features incorporating inter-vertebrae shape, intra-vertebra shape, and inter-vertebrae contextual information.

Available data

Our data are obtained from our collaborating radiology center. All data are in DICOM format and are anonymized before we receive them. We also receive an anonymized clinical report along with each case showing all abnormalities at each vertebra level. Among fifty cases, there are thirty abnormal cases and twenty normal ones. Each abnormal case has at least one vertebra with an abnormality including various types of fracture and, specifically, compression fracture, wedge compression fracture, and *Spondolysis*.

Each CT volume contains a set of sagittal images with an average of 88 slices per case. However, the far lateral slices from right and left have minimal to no information about the vertebrae. Instead, they mainly consist of the fat and soft tissues before reaching the vertebral column. Thus, we pre-process each volume by including only 25 slices. Upon examination of all the fifty cases, we find that 25 slices are completely representative for the lumbar vertebrae in each case. We compute the index for the middle slice (with floor operator for fractions) and obtain the 14 slices below and 11 slices up the middle slices. This brings the total to 25 slices. In all these 25 slices, the vertebrae are visible, clear, and distinguishable from soft tissues.

We point out that due to the many known legal limitations, this dataset will not be available online. Moreover, a larger dataset is desired to ensure the robustness of the system.

Materials and methods

Our proposed CAD system consists of three major steps: vertebrae detection and labeling, vertebrae segmentation,

and vertebrae diagnosis (wedge fracture detection). Figure 2 shows the work flow of our proposed CAD system.

In our previous work [2], we proposed a two-level probabilistic model for automatically localizing and labeling the lumbar discs. We utilize our model for localization of the vertebra using the localized disc centers. Each vertebrae center is computed by the mean location between the two enclosing discs. Then, we perform automated vertebrae segmentation using an ASM and then refine the segmentation via the *Gradient Vector Flow Active Contour* (GVF-Snake). Our automated segmentation method and its experimental results were presented in [1]. Figure 3 shows two sample cases after vertebrae detection, labeling and segmentation resulting from our previous work [1,2].

Active shape model (ASM) [6,16] has proven its robustness to many segmentation problems in medical imaging. However, its most success depends on clear boundary of the target organ. In our clinical CT scans, vertebra shows a decent level of vertebra boundary that is extremely suitable for ASM. In our segmentation step, we have a separate model for each vertebra level. To prepare the training data, we ask the radiologist to manually mark 16 landmark points for each vertebra guided by the model shown in Fig. 4. We name these landmark points from P_1 to P_{16}. Upon the guidance of the ASM [6], we calculate the mean shape $\bar{P} = \frac{1}{N}\sum_{n=1}^{N} P_n$ where $N = 16$. Then, each vertebra shape P_n, where $n \in \{1, 2, \ldots, N\}$ is recursively aligned to the mean shape $\bar{P}$ using generalized Procrustes analysis to remove translational, rotational, and isotropic scaling from the shape. Then, we model the remaining variance around the mean shape for each vertebra with principal components
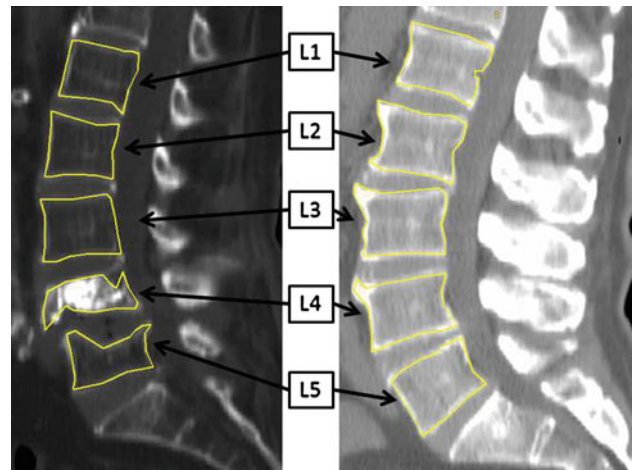

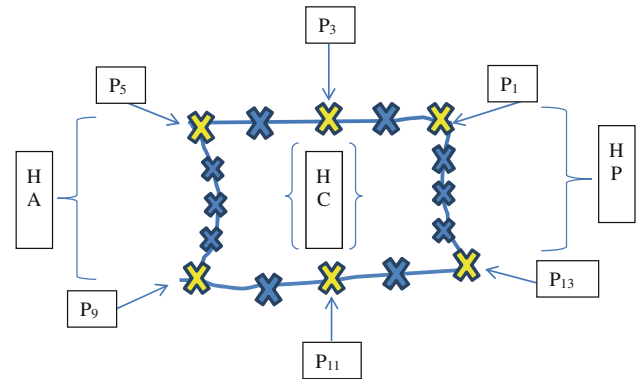**Fig. 3** Segmentation results from the GVF-snake and ASM

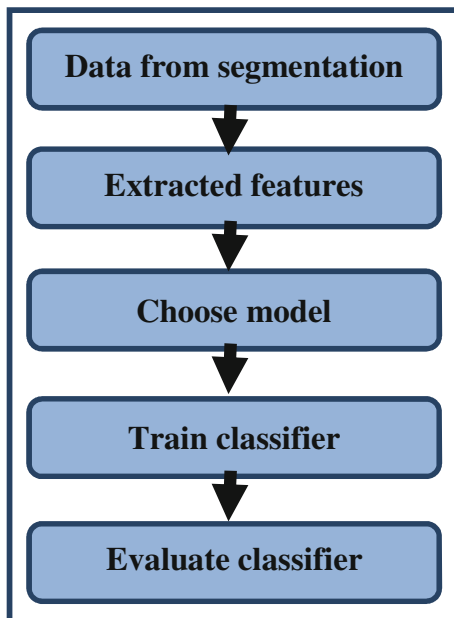**Fig. 4** Vertebra sample showing the training inputs and the selected features

analysis (PCA) to extract the eigenvectors of the covariance matrix associated with 98 % of the remaining point position variance according to the standard method for deriving the ASMs linear shape representation. For the testing step, we apply the mean shape $\bar{P}$ around the vertebra point produced by our localization step. Then, we allow the ASM to converge and obtain the boundary. We feed this boundary to the GVF-snake in the next step.

Gradient Vector Flow (GVF) snake [17] has been proven over the years to work robustly in refining an initial edge map into a smooth final shape. GVF-snake is the parametric curve that solves:

$$\mathbf{x(s, t)} = \alpha \mathbf{x''(s, t)} - \beta \mathbf{x''''(s, t)} + \mathbf{v} \qquad (1)$$

where α and β are weighting parameters that control the contour's tension and rigidity, respectively. $\mathbf{x''}$ and $\mathbf{x''''}$ are the second and fourth derivatives, respectively, of $\mathbf{x}$. $\mathbf{v(x, y)}$ is the gradient vector flow (GVF), $\mathbf{s} \in [0, 1]$, and $\mathbf{t}$ is the time component to make a dynamic snake curve from $\mathbf{x(s)}$ yielding $\mathbf{x(s, t)}$.

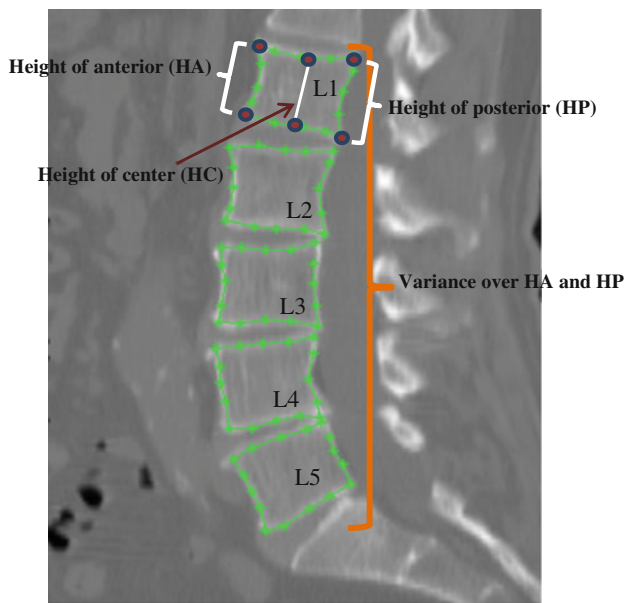**Fig. 2** Work flow of our CAD system

Fig. 5 Sample middle slice CT from our dataset with segmentation results from our previous work [1]

The smooth boundary outcome from the GVF-snake is utilized to obtain more refined locations for the 16 landmarks given by the ASM. We extract a set of features based on these 16 converged landmark points as shown in Fig. 4. In this section, we discuss our selected features and our proposed diagnosis scheme.

Features extraction

Features are characteristics of the objects of interest such as shape, texture, and color. They provide us with the relevant information we need if we select them carefully. Feature extraction methods analyze objects and images to extract the most prominent features that are representative of the various classes of objects [14]. This crucial step is mainly responsible for the discriminative power of the decision-maker (classifier).

We base our feature extraction on the set of points resulting from applying the ASM during the segmentation step [1]. Figure 4 shows a model for the vertebra with the set of points labeled. Our segmentation step results in two outcomes: a refined contour for each vertebra and a set of 16 points surrounding each vertebra as shown in the model in Fig. 4. We utilize these points to extract the set of relevant features that distinguish wedge-fractured vertebrae from normal ones. Figure 5 shows one sample CT middle slice with the converged contour for the five lumbar vertebrae. It also shows the set of the 16 points for each vertebra.

To clarify our proposed features, we define the three distances: HP_i, HA_i, and HC_i where HP_i is the posterior

distance at vertebra (i), HA_i is the anterior distance at vertebra (i), and HC_i is the center distance at vertebra (i) as shown in Fig. 4. We define:

$$HP_i = \sqrt{(x_{1i} - x_{13i})^2 + (y_{1i} - y_{13i})^2} \tag{2}$$

$$HA_i = \sqrt{(x_{5i} - x_{9i})^2 + (y_{5i} - y_{9i})^2} \tag{3}$$

$$HC_i = \sqrt{(x_{3i} - x_{11i})^2 + (y_{3i} - y_{11i})^2} \tag{4}$$

where (x_{ji}, y_{ji}) is the (x, y)-coordinates for point (j) at vertebra level (i). For each vertebra (i), we have the sixteen ASM points $P_1, \ldots, P_{16}$. There are five lumbar vertebra ($i = 1, \ldots, 5$). Using these distances, we extract the following four features for each vertebra:

(a) F_1: the ratio of posterior height to anterior height

$$F_1 = \frac{HP_i}{HA_i} \tag{5}$$

(b) F_2: The absolute difference between the anterior height and the posterior height normalized with respect to the central height

$$F_2 = \left| \frac{HA_i - HP_i}{HC_i} \right| \tag{6}$$

(c) F_3: Inter-vertebrae anterior heights variance

$$F_3 = \sigma^2 (HA_i) \tag{7}$$

(d) F_4: Inter-vertebrae posterior heights variance

$$F_4 = \sigma^2 (HP_i) \tag{8}$$

Below we address each feature, its motivation, and the individual effect.

F_1: the ratio of posterior height to anterior height

Upon the clinical examination of all available cases, we find that the anterior height (HA) and posterior height (HP) are quite similar for normal vertebrae and thus the ratio $F_1 \approx 1$. However, they tend to diverge for the wedge fracture vertebrae. Thus, the ratio value is either $F_1 >> 1$ or $F_1 << 1$ for wedge fracture vertebrae. Table 1 shows sample cases from our dataset with numeric values for each feature at each vertebra level.

Table 1 Three sample cases with feature numeric values and the ground truth decision (1 is normal and 2 is wedge fracture)

Case ID	Vertebra label	F_1	F_2	F_3	F_4	Truth decision
1	L1	1.055	0.061			1
	L2	0.940	0.075			1
	L3	0.859	0.168	0.46532	0.08892	1
	L4	1.399	0.437			2
	L5	1.424	0.496			2
2	L1	1.094	0.098			1
	L2	1.094	0.108			1
	L3	0.947	0.065	0.02287	0.08293	1
	L4	0.932	0.083			1
	L5	0.897	0.112			1
3	L1	1.162	0.154			1
	L2	1.060	0.055			1
	L3	0.952	0.048	0.0081	0.02693	1
	L4	0.996	0.004			1
	L5	1.126	0.113			1

F_2: The absolute difference between the anterior height and the posterior height normalized with respect to the central height

This feature represents the intra-shape within the vertebra taking the central height into consideration:

$$F_2 = \left| \frac{HA_i}{HC_i} - \frac{HP_i}{HC_i} \right| = \left| \frac{HA_i - HP_i}{HC_i} \right| \quad (9)$$

This feature's value approaches zero ($F_2 \approx 1$) for normal cases and goes higher ($F_1 >> 1$) because of wedge fracture existence. We consider the absolute value to have only positive resulting value because the negative value is irrelevant. The purpose here is to see whether there is significant height difference between anterior and posterior with respect to the central height. Table 1 shows sample numeric results for this feature.

F_3 and F_4: inter-vertebrae anterior and posterior heights variance, respectively:

These two features represent the inter-vertebra context information. It confirms a small value for highly similar vertebrae and a higher value for higher variation within the respected features (anterior and posterior heights). The five lumbar vertebrae tend to have similar (low variance) anterior heights and posterior heights in the normal case. They are not of the same heights but their collective variance is small as shown in Table 1 for the normal cases. However, when some vertebrae have wedge fracture, the variance of the anterior and posterior (or either of them depending on the location

of the wedge) heights across the five vertebrae level increases. Table 1 shows sample data supporting our selected features.

Decision-making

Collecting these clinically motivated features together requires automated machine intelligence for decision-making which is to give the outcome of being normal versus abnormal vertebra. To that end, we utilize two learners from the two broad families in machine learning: K-Means and Neural Networks (NN). K-Means is considered one of the major classical unsupervised learning methods while Neural Networks is among the robust and well-studied supervised learners. To achieve concrete and reliable decision-making learner, we utilize these two learners and validate them on the whole dataset as shown in the following sections. We present each learner with its results and then compare their performance.

Unsupervised learning: K-Means

K-Means is an unsupervised learner, *i.e.,* does not require domain knowledge guidance. It starts with a set of K cluster centers: $\mu_1^{(0)}, \mu_2^{(0)}, \ldots, \mu_K^{(0)}$. Each center μ_k is a *D-dimensional* vector. The overall goal of K-Means is to minimize J which is the sum of squares of distances for each point from its assigned center μ_k:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} ||\mathbf{x_n} - \mu_k||^2 \quad (10)$$

where r_{nk} is a binary indicator describing which of the k-clusters the data point $\mathbf{x_n}$ corresponds to:

$$r_{nk} = \begin{cases} 1 & \text{if } k = \arg\min_j ||\mathbf{x_n} - \mu_j||^2; \\ 0 & \text{Otherwise.} \end{cases} \quad (11)$$

The optimization step (minimization of J) is performed using the Expectation-Maximization (EM) algorithm by initially choosing a set of initial cluster centers μ_k, then: (i) minimize J with respect to the r_{nk} with μ_k fixed, (ii) minimize J with respect to μ_k keeping the r_{nk} fixed. The first step (E-step) is the assignment step of the data points for the current cluster centers μ_k, while the second step (M-step) is the re-computation of the cluster centers μ_k upon the new data point's assignment from the E-step. This proceeds iteratively until some stopping criteria is met [4].

In our K-Means step, preparation of the feature vector for each vertebra is automatically performed as illustrated in the previous section. We performed K-Means clustering based on a global consideration over all the vertebrae. Feature vectors are fed to the K-Means regardless of the vertebra level. The two later features F_3 and F_4 are repeated for each

vertebra to make the full vectors. The data point vector is four-dimensional and the initial means are randomly generated. Moreover, we performed local K-Means for each vertebra level separately but the results were not promising at all; thus, we only report the global K-Means across all vertebrae regardless of the vertebra level.

We run our global K-Means on the whole dataset that consists of fifty cases. Considering that we have five lumbar vertebrae in each case, we have $5 \times 50 = 250$ data points with four dimensions (features F_1, F_2, F_3, and F_4). Because K-Means is unsupervised, the truth value for each record is not used. However, we use it for measuring accuracy in terms of the number of misclassified vertebrae, false positives, and false negatives.

Supervised learning: neural network

Neural network emerged as a robust and powerful supervised learner. It has wide variability in its types and structure. A two-layer (one hidden layer) neural network overall function $y_k(\mathbf{x}, \mathbf{w})$ is defined by:

$$y_k(\mathbf{x}, \mathbf{w}) = \sigma \left(\sum_{j=0}^{M} w_{kj}^{(2)} h \left(\sum_{i=0}^{D} w_{ji}^{(1)} x_i \right) \right) \qquad (12)$$

where y_k is the output neuron k, $\mathbf{x}$ and $\mathbf{w}$ are the data vector (feature input) and the learned weights of the function, respectively. σ and h are the transition functions of the hidden layer and input layer, respectively. The superscript on the weight variable $w^{(\cdot)}$ corresponds to the layer order [4].

In our neural network, we use a two-layer neural network (one hidden layer). The input layer has four neurons $\mathbf{x}$ that correspond to the four dimensions of our feature vector for each data point, *i.e.*, $\mathbf{x} = < x_i >$ where $i = 1, 2, \ldots, 4$ and $D = 4$. The hidden layer contains 10 neurons, *i.e.*, $M = 10$. Our network is a feed-forward-back propagation that uses Levenberg-Marquardt optimization. We used sigmoid transition functions and set the rate as 0.05. We point out that the selection of the number of hidden neurons ($M = 10$) is empirical. We selected this amount after testing the results with various amount of hidden neurons.

Results

Diagnostic test performance is usually expressed by the terms: accuracy, sensitivity, and specificity. When a single test is performed, the person may have the disease (positive) or may not (negative). An ideal test should have high sensitivity, high specificity, and high accuracy. We define these accuracy measures as follows:

(a) Accuracy: we measure the accuracy as the ratio of the number of correctly classified vertebrae (N_c) to the total number of vertebrae (N):

$$Accuracy = \left(\frac{N_c}{N} \right) \times 100\% \qquad (13)$$

(b) Sensitivity: It is defined as the probability that the test says a person has the disease (positive) when in fact he has it (true positive) and is defined by:

$$Sensitivity = \left(\frac{TP}{TP + FN} \right) \times 100\% \qquad (14)$$

(c) Specificity: It is defined as the probability that the test says a person does not have the disease (negative) when in fact he does not (true negative) and is defined by:

$$Specificity = \left(\frac{TN}{TN + FP} \right) \times 100\% \qquad (15)$$

In our experiment diagnosis task, FP is the number of false positives (normal vertebrae diagnosed as wedge fractured), TP is the number of true positives (correctly diagnosed wedge-fractured vertebrae), FN is the number of false negatives (misclassified wedge-fractured vertebrae), and TN is the number of true negatives (correctly classified normal vertebrae). Below, we show the evaluation for both learners on our dataset.

Evaluation of K-Means

Upon the evaluation of the global K-Means described in section "Unsupervised learning: K-Means", we evaluate K-Means on the whole dataset at once and then compare the resulting automated outcome with our gold standard results and obtained the detailed Table 2.

As per the results in Table 2, we obtained 98% classification accuracy as the K-Means missed five vertebrae ($FP + FN$) out of the total of 250 vertebrae. Moreover, the sensitivity of our solution equals 87.5% while we obtained very high specificity of 99.1%.

Table 2 Global K-Means evaluation

		Gold standard	
		Fracture	Non-fracture
Automated	Fracture	21 (TP)	2 (FP)
Results	Non-fracture	3 (FN)	224 (TN)

Table 3 Results of the cross-validation experiment with an average detection accuracy of 93.2 %

Round	Testing cases	Training cases IDs	No. misclassified training (out of 225)	Training accuracy (%)	No. misclassified testing (out of 25)	Testing accuracy (%)
1	1–5	6–50	3	98.67	1	96.00
2	6–10	1–5, 11–50	4	98.22	3	88.00
3	11–15	1–10, 16–50	2	99.11	2	92.00
4	16–20	1–15, 21–50	5	97.78	2	92.00
5	21–25	1–20, 26–50	5	97.78	2	92.00
6	26–30	1–25, 31–50	5	97.78	1	96.00
7	31–35	1–30, 36–50	2	99.11	1	96.00
8	36–40	1–35, 41–50	3	98.67	3	88.00
9	41–45	1–40, 46–50	5	97.78	2	92.00
10	46–50	1–45	1	99.56	0	100.00
Average				98.44		93.20

Evaluation of neural network

Because Neural Network is a supervised learner, it requires training and thus the experiment is quite different from the K-Means. Thus, we experiment with cross-validation which is the standard way for such thorough experiments that require concrete and convincing accuracy measures. Hence, we perform five-cases-leave-out cross-validation experiment on the fifty cases in hand. In each round, we leave five cases out for testing and train the neural network on the remaining 45 cases. We repeat this ten times to round over all the cases in sequence as shown in Table 3. Our average training accuracy is 98.44 % while our testing accuracy average is 93.2 %.

Comparison between K-Means and neural network

The two learners we used show high and robust measures in terms of classification accuracy, sensitivity, and specificity. The average results are also highly comparable between the two learners. Our proposed features for inter-vertebrae shape (F_1), intra-vertebrae shape (F_2), and inter-vertebrae contextual information (F_3 and F_4) played the major role in the high accuracy for the utilized learners. Despite that K-Means showed higher classification accuracy, it, on the other hand, showed lower sensitivity. Measuring specificity and sensitivity for neural network was not appropriate due to the nature of the cross-validation experiment because of the repetition of the vertebrae instances within the ten rounds unlike the K-Means experiment.

Conclusion

In this paper, we proposed a set of features that showed great potential for automating wedge fracture compression diagnosis from clinical CT images. We presented our full system starting from our previous efforts in vertebrae localization and segmentation and our new work in the diagnosis of wedge compression fracture. Our system is fully automated and of great clinical use as per our experimental results. We proposed four features representing inter-vertebrae shape (F_1), intra-vertebrae shape (F_2), and inter-vertebrae contextual information (F_3 and F_4) that collectively allowed both K-Means and neural network to perform the decision-making with high accuracy, sensitivity, and specificity. Our overall accuracy was 98 % for K-Means and an average of 93.2 % for neural network testing set. K-Means showed high specificity of 99.1 % and acceptable sensitivity of 87.1 %.

Conflict of interest None.

References

1. Al-Helo S, Alomari R, Chaudhary V, Al-Zoubi M (2011) Segmentation of lumbar vertebrae from clinical CT using active shape models and GVF-snake. In: Proceedings international IEEE conference in engineering in medicine and biology society (EMBC'11), pp 8033–8036. IEEE, Boston, MA, USA
2. Alomari R, Corso J, Chaudhary V (2011) Labeling of lumbar discs using both pixel- and object-level features with a two level probabilistic model. IEEE Trans Med Imaging 30(1):1–10
3. Alomari R, Corso J, Chaudhary V, Dhillon G (2011) Toward a clinical lumbar CAD: Herniation diagnosis. Int J Comput Assist Radiol Surg (IJCARS) 6(1):119–126
4. Bishop C (2006) Pattern recognition and machine learning (Information Science and Statistics). Springer, New York
5. Cherukuri M, Stanley R, Long R, Antani S, Thoma G (2004) Anterior osteophyte discrimination in lumbar vertebrae using size invariant features. Comput Med Imaging Graph 28(1–2):99–108
6. Cootes TF, Taylor CJ (2001) Statistical models of appearance for medical image analysis and computer vision. In: Sonka M, Hanson KM (eds) Proceedings of SPIE medical imaging, p 236

7. Eastell R, Cedel S, Wahner H, Riggs B, Melton L (1991) Classification of vertebral fractures. J Bone Miner Res 6(3):207–215

8. eOrthopod: Image used courtesy of medical multimedia group, llc. more information is available at eorthopod.com. http://www.eorthopod.com/content/adult-lumbar-spine-fractures-types. Accessed September 2012

9. Ghosh S, Alomari R, Chaudhary V, Dhillon G (2011) Automatic lumbar vertebra segmentation from clinical CT for wedge compression fracture diagnosis. In: Proceedings SPIE medical imaging, vol 7963, p 796303. Orlando, Florida, USA

10. Kasai S, Li F, Shiraishi J, Li Q, Doi K (2006) Computerized detection of vertebral compression fractures on lateral chest radiographs: preliminary results with a tool for early detection of osteoporosis. Med Phys 33(12):4664–4674

11. Mastmeyer A, Engelke K, Fuchs C, Kalender W (2006) A hierarchical 3D segmentation method and the definition of vertebral body coordinate systems for QCT of the lumbar spine. Med Image Anal 10(4):560–577

12. Roberts M, Cootes T, Pacheco E, Adams J (2007) Quantitative vertebral fracture detection on DXA images using shape and appearance models. Acad Radiol 14(10):1166–1178

13. Smyth P, Taylor C, Adams J (1999) Vertebral shape: automatic measurement with active shape models. Radiology 211(2):571–578

14. Spine Universe: http://www.spineuniverse.com/. Accessed September 2012

15. Tan S, Yao J, Ward M, Yao L, Summers R (2008) Computer aided evaluation of ankylosing spondylitis using high-resolution CT. IEEE Trans Med Imaging 27(9):1252–1267

16. T.F Cootes D Copper CT, Graham J (1995) Active shape models—their training and application. Comput Vis Image Underst 61(1):38–59

17. Xu C, Prince JL (1998) Snakes, shapes, and gradient vector flow. IEEE Trans Image Process 7(3):359–369