



Assessment of osteoporotic vertebral fractures using specialized workflow software for 6-point morphometry

Giuseppe Guglielmi^{a,b,*}, Francesco Palmieri^b, Maria Grazia Placentino^b,
Francesco D’Errico^b, Luca Pio Stoppino^a

^a Department of Radiology, University of Foggia, Viale Luigi Pinto, Foggia 71100, Italy

^b Department of Radiology, Scientific Institute Hospital “CSS”, Viale Cappuccini, San Giovanni Rotondo 71013, Italy

Received 7 May 2007; received in revised form 6 August 2007; accepted 3 December 2007

Abstract

Purpose: To evaluate the time required, the accuracy and the precision of a model-based image analysis software tool for the diagnosis of osteoporotic fractures using a 6-point morphometry protocol.

Materials and methods: Lateral dorsal and lumbar radiographs were performed on 92 elderly women (mean age 69.2 ± 5.7 years). Institutional review board approval and patient informed consent were obtained for all subjects. The semi-automated and the manual correct annotations of 6-point placement were compared to calculate the time consumed and the accuracy of the software. Twenty test images were randomly selected and the data obtained by multiple perturbed initialisation points on the same image were compared to assess the precision of the system.

Results: The time requirement data of the semi-automated system (420 ± 67 s) were statistically different ($p < 0.05$) from that of manual placement (900 ± 77 s). In the accuracy test, the mean reproducibility error for semi-automatic 6-point placement was $2.50 \pm 0.72\%$ [95% CI] for the anterior–posterior reference and $2.16 \pm 0.5\%$ [95% CI] for the superior–inferior reference. In the precision test the mean error resulted averaged over all vertebrae was $2.6 \pm 1.3\%$ in terms of vertebral width.

Conclusions: The technique is time effective, accurate and precise and can, therefore, be recommended in large epidemiological studies and pharmaceutical trials for reporting of osteoporotic vertebral fractures.

© 2007 Elsevier Ireland Ltd. All rights reserved.

Keywords: Radiograph; Computer aided diagnosis; Spine fractures; Vertebral shape analysis; Statistical shape models (SSM); Osteoporosis

1. Introduction

Osteoporotic fractures are one of the most common causes of disability and a major contributor to medical care costs in many regions of the world. As the population ages, the social and financial burden of fractures will increase throughout the world. Compared with hip fractures, the prevalence of vertebral fractures is underestimated since a substantial proportion of fractures may escape clinical diagnosis and thus necessitate radiological assessment [1].

The conventional method for diagnosing an osteoporotic vertebral fracture relies on the subjective assessment of the vertebral deformity on lateral radiographs (crush, wedge or biconcave deformity). In order to minimise the subjective bias intrinsic to qualitative readings and to homogenise data analysis, a number of morphometric systems for the characterisation of osteoporosis have been developed [2–6]. There are currently two accepted methods for the objective assessment of vertebral fracture. The first method is the “semiquantitative” approach [3,7,8] in which a radiologist classifies each vertebra by means of visual examination of the radiograph according to strict definitions of degree and type of fracture. In the second method, six (or more) points are placed manually in the vertebral margins and are used to calculate anterior, middle and posterior heights, which are subsequently often used to categorize fracture type and referred to as morphometry [2,4,5]. However, the manual placement of points on the vertebrae is still somewhat subjective, which introduces variability into the process of detecting vertebral fractures,

* Corresponding author at: Viale Cappuccini, 1, San Giovanni Rotondo 71013, Italy. Tel.: +39 0882 410686; fax: +39 0882 453861.

E-mail addresses: g.guglielmi@unifg.it (G. Guglielmi), francescopalmieri@hotmail.com (F. Palmieri), placentino_mg@tiscali.it (M.G. Placentino), fra.derrico@tiscali.it (F. D’Errico), l.stoppino@hotmail.it (L.P. Stoppino).

and is time consuming, requiring more than 10 min to analyse a single radiograph. In order to address these problems, a number of semi-automated software solutions have been described, and indeed commercialised in Dual Energy X-ray Absorptiometry (DXA) systems, such as those produced by GE Lunar and Hologic [9–13]. In particular, Smyth et al. [14] have applied a statistical model-based approach, the Active Shape Model (ASM), to DXA scans. A number of attempts have also been made to analyse radiographs [15–17].

In the work presented here, we have used a commercial statistical model-based vision system (MorphoXpress, P&G Pharmaceuticals, Rusham Park, Egham, UK) to digitise, using a scanner, and analyse radiographs for semi-automated morphometric assessment. Analysis starts with manual indication of the centres of the upper and lower vertebrae and with automatical software positioning of landmarks. These points, which can be moved by the operator, are then used by the software to calculate anterior, middle and posterior vertebral heights.

The purpose of our study was to evaluate the time consumed by, the accuracy and the precision of the 6-point annotation aspect of the software tool, being an important component of the workflow improvement in the assessment of osteoporosis.

2. Materials and methods

2.1. Test images

Institutional review board approval and patient informed consent were obtained for all subjects. Experiments were carried out on a set of 92 digitised radiographs; these images had not been used for training the statistical model used in MorphoXpress. Although our study is retrospective, the original radiographs were selected randomly over a period of several months from patients examined at an osteoporosis centre in a large Italian hospital. No specific criteria or protocols were specified for image selection, except that the radiographs were centred on vertebra T7. Thus, the images arose from normal clinical activity and represent a randomly selected population of female patients of mean age 69.2 ± 5.7 years, mean height

151.8 ± 5.3 cm and mean weight 64.2 ± 11.7 kg. The diagnosis of osteoporosis was performed by visual inspection employing the semiquantitative method, previously described by Genant et al. [3]. In this technique, fractures assessed on lateral thoracolumbar spine radiographs were defined as reductions of more than approximately 20% in anterior, middle or posterior vertebral height. Fractures were defined mild, moderate and severe based on a height ratio decrease of 20–25%, 25–35% and more than 35%, respectively.

Women with known or suspected malignant disease or with secondary osteoporosis were excluded from the study. The population also contained a variety of normal, under- and over-exposed films, indicating both normal anatomy and vertebral fractures.

2.2. Software workflow

The software tool investigated in this paper was developed by Image Metrics for Procter & Gamble Pharmaceuticals (Image Metrics PLC, Regent House, Cheshire, UK). The MorphoXpress workflow is as follows: original lateral vertebral radiographs are digitised using a TWAIN scanner (UMAX Power Look 1000). Analysis is then initialised by the manual indication of the centres of the upper and lower vertebrae to be analysed. The software then automatically finds the positions of landmarks for a standard 6-point morphometry measurement. The software then allows these points to be moved by the operator, if deemed necessary, before the points are confirmed as being correct. The positions of the confirmed points are then used by the software to calculate anterior, middle and posterior vertebral heights, which may also be used for the determination of a deformity metric. Fig. 1 shows the MorphoXpress user interface for image analysis.

2.3. Modelling shape

The statistical modelling technique that has been applied to medical images is the Active Shape Model (ASM) [18–21]. The approach to model-based vision, including the one under investigation, uses the following framework.

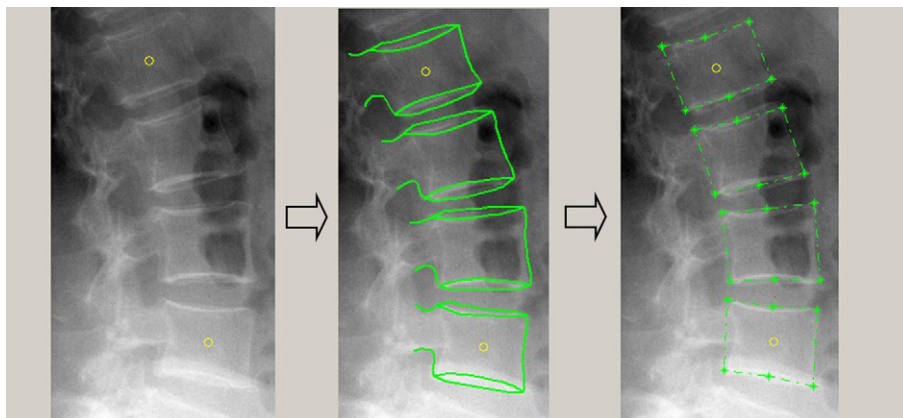


Fig. 1. Schematic operation of the software. A 2-point initialisation is used to define the mid-points of the upper (L1) and lower (L4) vertebrae to be measured. A full annotation by model-fitting is then performed which is used to determine the positions of 6-point morphometry landmarks.

There are two distinct stages: a one-time ‘model-building’ stage and a ‘runtime’ stage. To build the model, a set of representative images of the anatomy to be modelled are annotated manually by an expert, i.e., training samples of the anatomy to be modelled are defined by a set of landmark points placed along the anatomical boundary. Each example is then described by a vector $\mathbf{x} = [x_1, y_1, x_2, y_2, \dots, x_n, y_n]$, where (x_i, y_i) is the position of the i th landmark on an example. In cases where certain boundaries are occluded or missing, the expert must use his/her anatomical knowledge to interpolate an estimated boundary. This set of annotated training images efficiently captures the expert’s domain knowledge of how the anatomy of interest is presented in the chosen modality. The annotated objects are treated as a ‘training set’ and are aligned in scale and rotation, before a form of statistical decomposition is applied to describe the variation of the aligned shapes. This analysis results in the establishment of anatomical shape parameters, which can be understood as a set of independent weighted modes; adjusting the weights produces different shape examples. At runtime, the model is scaled, rotated, translated and further modified by adjusting the weights to create a statistically based best fit model describing the underlying patient image data. The shape captured by a statistical model may be described by a Point Distribution Model (PDM) that is generated by the statistical analysis of the variation in landmark point coordinates in a set of annotated images. The PDM represents shape in terms of a mean shape and a set of linearly independent modes of shape variation, requiring only the most significant variations of shape to be annotated. Any new shape, \mathbf{x} , of the same type as those observed in the training set can then be generated by summing combinations of these modes of variation, contained within a matrix \mathbf{P} , to the mean shape \mathbf{x}_m , with a vector of weight \mathbf{b} controlling the influence of each mode: $\mathbf{x} = \mathbf{x}_m + \mathbf{P}\mathbf{b}$.

A PDM of a single vertebra (with the points connected) can be seen in Fig. 2, showing the mean shape and 3 standard deviation (S.D.) limits for each of the most important 5 modes of shape variation. This diagram indicates the way in which, if the model is constrained to the 3 S.D. limits, only plausible vertebral shapes

may be generated while using the model to find and analyse vertebrae within a patient image.

The models at the core of the MorphoXpress technology developed using a set of 92 digitised lateral vertebral radiographs, obtained randomly from two large studies regarding the glucocorticoid induced osteoporosis and recognition of fractures in digital radiographs [22,23]. This set of radiographs was chosen to ensure that a sufficient number and variety of vertebral deformities would be included in the training set data. The database had an incidence of one or more vertebral fractures categorised as grades 1, 2 or 3 using the Genant semi-quantitative technique [3] in 31.1% of images. In particular, the image set consists of 62 analogue images and 30 digital images. Digital in this context means acquired directly by a Digital Radiography (Fuji Medical System FCR 5000) scanner at source and subsequently printed onto high quality photographic film. The 92 radiographic films were digitised using a commercial flat bed scanner with full size film back lighting adapter (UMAX Power Look 1000) connected via USB port to a desktop PC. Digitisation was at a resolution of 300 dpi resulting in an image of 2000 pixels height and variable width. The scanned image files were then saved to file without compression. In every case, the scanning protocol was performed by the same radiologist who subsequently annotated the images using MorphoXpress.

2.4. Overview of experimental analysis

The MorphoXpress software system is designed to reduce the time required to assess vertebral fractures in studies of osteoporosis by using an accurate and precise semi-automated system of examining a radiograph. However, it is not intended to replace the radiologist. Therefore, three analysis experiments were performed: (1) to compare the time take of the MorphoXpress with the manual placement of the 6 points per vertebra; (2) to investigate the accuracy of the system in placing landmark points and (3) to investigate precision of this placement.

The anterior–posterior (AP) and superior–inferior (SI) measurements were obtained automatically by the software and manually moved by the operator if necessary.

Errors in both accuracy and precision were calculated in both AP and SI directions, within a vertebral frame of reference. The AP axis was defined by the line passing through the user-corrected posterior and anterior points on the inferior vertebral edge. The SI axis was then defined by a line drawn perpendicular to the AP axis. Error distances were projected onto these axes. Furthermore, the radiographs contained no calibration fiducials or scales and so pixels could not be given a physical length. Therefore, in order to normalise measurements for comparison, both AP and SI projected errors were recalculated in terms of vertebral width. The halfway point between superior and inferior annotation points at both the posterior and anterior vertebral edges were calculated and the width was then defined as the distance between these two mean points. Fig. 3 shows schematically how the vertebral frame of reference and vertebral width were defined.

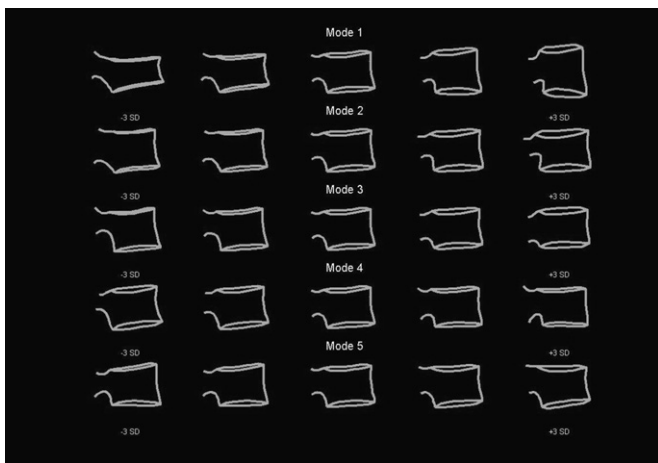


Fig. 2. A schematic of a Point Distribution Model (PDM) of a single vertebra showing the mean and 3 S.D. limits of the first five principal modes of shape variation ranked in order of the amount of shape variation capture by each.

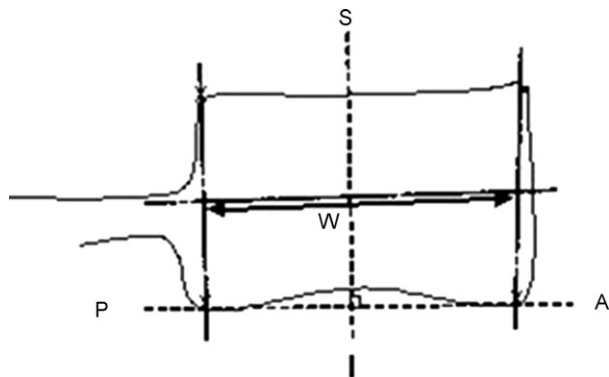


Fig. 3. Definition of vertebral width and frame-of-reference for error projection. Abbreviations are P(osterior), A(nterior), S(uperior), I(nferior) and W(idth).

2.5. Time consuming: comparison with the manual placement

The time required by the procedures in the assessment of the 92 radiographs were measured by an electronic chronometer. The measured time started with the digitisation of the film and stopped when the software calculated the anterior, middle and posterior vertebral heights. The mean with the standard deviation for both was calculated. A paired two-tail Student's *t*-test was calculated to assess a statistically significant difference between the two procedures. A difference with a *p* value less than 0.05 was considered statistically significant.

2.6. Accuracy: comparison with adjusted annotation

This study was carried out in order to test the accuracy of the semi-automated annotation (initialised by selecting two initialisation points, defined as the centres-of-gravity of the upper and lower vertebrae) compared to a 'true values' as defined by manually corrected automated annotations. It could be argued that this might introduce a 'bias'; however, this simulates the context in which the tool is used and secondly, a comparison to purely manual annotation will be the subject of a further study. The results of both semi-automated annotation, initialised by a well-trained musculoskeletal radiologist with a specific expertise in the radiology of osteoporosis, and the subsequently corrected annotations by the same radiologist were gathered for the 92 image sample pool described above. The semi-automated vertebrae annotation points were then compared with the finally corrected annotation data. None of the images were repeated in the trial. For reference, a schematic labelled vertebra is shown in Fig. 4. It should be noted that lumbar and thoracic vertebrae were assessed by the same study, thus, a test image could contain vertebrae T5–L4, in which case analysis would be initialised by indicating the centres of the most superior vertebra (T5) and the most inferior vertebra (L4).

To measure accuracy, the mean root-mean-square (RMS) point-to-point AP and SI errors between the semi-automated

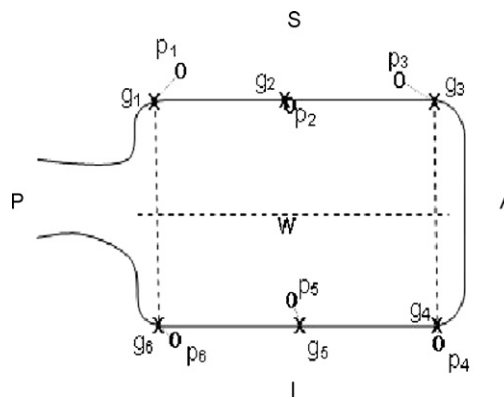


Fig. 4. Schematic width vertebra showing automatic (*p_j*) and corrected (*g_j*) annotations. *j*= means the six points used to define vertebral height. Point positions have been exaggerated for emphasis.

annotation points and corrected results were calculated for each vertebra types T5–L4 using Eq. (1) in Appendix A. Vertebrae that exhibited AP or SI errors greater than 3 S.D. away from the mean error were treated as failures of automatic annotation and excluded from calculation of the overall mean error. This prevented distortion of the accuracy results, although the number of excluded vertebra was recorded.

2.7. Precision

This study was performed in order to test the precision, or reproducibility, of the system using simulated multiple perturbed initialisation points on the same image. This aspect of the analysis reveals if the system is too sensitive to the position of the initialisation points on the upper and lower vertebrae such that performance is no longer operator-independent.

The automatic annotation algorithm was run on 20 randomly selected test images from the 92 image set. As described above, initialisation points were defined as the centres-of-gravity of the manually corrected 6-point morphometry landmarks on the upper and lower vertebrae. The simulated user input data were varied by adding a radial offset to these initialisation points, computed from a random Gaussian distribution (standard deviation = 2.5% vertebral width). A standard deviation of 2.5% was chosen as representative of the user-correction tolerance observed in the results of the accuracy experiment in order to simulate clinical usage. In a cumulative histogram analysis, no manual correction of automatic annotation was made below this threshold. Each test image was automatically annotated in this way 10 times.

The mean RMS point-to-point AP and SI errors between the automatically generated annotation points and the mean of those points across all 10 examples were measured for each vertebra type (labelled T5–L4) in the test set, using Eq. (2) in the appendix. As with the accuracy experiment, any vertebrae exhibiting AP or SI errors greater than 3 S.D. from the mean were treated as failures of automatic annotation and excluded from the calculation of the overall mean error.

Table 1

Overall mean accuracy errors as a fraction of vertebral width, projected onto AP and SI axes, the total numbers of vertebrae that were analysed and the number of failures (error > 3 S.D. from the mean error) are indicated

Level	AP error (%)	SI error (%)	Analysed	Failures
T5	1.3874	1.4994	12	5
T6	2.8387	2.3441	17	5
T7	3.902	3.2172	17	4
T8	2.6039	2.0704	19	5
T9	3.5491	2.5499	19	6
T10	2.5219	2.0141	21	8
T11	2.8519	2.2665	28	4
T12	2.1325	2.023	53	3
L1	2.4516	2.4982	76	3
L2	2.1517	2.3066	73	4
L3	2.0125	1.6678	70	2
L4	1.6249	1.522	61	4

3. Results

3.1. Time consuming: comparison with the manual placement

The time consume of the semi-automated system was 420 ± 120 s which was statistically different ($p < 0.05$) from the one of the manual placement (900 ± 77 s).

3.2. Accuracy: comparison with adjusted annotation

Accuracy results were calculated by vertebral label (T5–L4) and are presented for all 92 images in Table 1 (indicating the total numbers of vertebrae), in Table 2 (indicating the analogue images) and in Table 3 (indicating the digital images). Table 1 reveals a mean percentage error (for non-failures) ranging from 1.39% to 3.90% for the AP reference (mean of $2.50 \pm 0.72\%$ [95% CI]) and from 1.50% to 3.22% for the SI reference (mean of $2.16 \pm 0.5\%$ [95% CI]), with higher accuracy on the lumbar spine. The mean of the signed point errors is approximately 0.4% vertebral width indicating little or no systematic error in automatic annotation.

Table 2

Overall mean accuracy errors for analogue test images

Level	AP error (%)	SI error (%)	Analysed	Failures
T5	–	–	3	3
T6	2.6203	1.4608	5	2
T7	2.4658	2.0857	5	1
T8	1.7082	2.4289	7	1
T9	4.5738	2.8254	7	3
T10	1.8286	1.3113	9	4
T11	3.8745	3.1493	15	3
T12	2.2989	2.0767	32	1
L1	2.6601	2.5926	51	1
L2	2.4713	2.8852	53	2
L3	2.2379	1.9078	52	2
L4	2.2835	2.0106	44	3

Table 3

Overall mean accuracy errors for digital test images

Level	AP error (%)	SI error (%)	Analysed	Failures
T5	2.7451	2.0865	9	1
T6	2.3389	4.1503	12	2
T7	4.5403	3.7201	12	3
T8	4.6386	2.7808	12	2
T9	4.2467	3.1273	12	2
T10	5.6957	2.9062	12	1
T11	1.8292	1.3837	13	1
T12	1.8611	1.9353	21	2
L1	1.9984	2.293	25	2
L2	1.8947	1.7475	20	2
L3	1.0912	0.7326	18	2
L4	0.4506	0.6297	17	2

3.3. Precision

As mentioned previously, we desire to observe that system performance is operator independent. Precision results are presented in Table 4. In the precision test, no vertebrae were discounted as a failure, i.e., none had mean errors greater than 3 S.D. from the mean error of all tests. Values range from a minimum of 0.8% for lower lumbar vertebrae to a maximum of 3.1% for T5 in the AP reference. The mean result averaged over all vertebrae in the test set is $2.6 \pm 1.3\%$ in terms of vertebral width. Fig. 5 shows AP precision for each vertebral label.

3.4. Fracture's data

Forty of the 92 digitised radiographs were diagnosed with vertebral fractures according to our criteria, including mild (19.6%), moderate (14.2%) and severe (9.8%) fractures.

4. Discussion

The main purpose of this study was to determine the aspects of efficacy of a semi-automatic workflow improvement software tool to aid the clinician when using a 6-point morphometry protocol for analysis of vertebrae. The mean reproducibility averaged over all (non-failure) vertebrae in the test set was 2.6% when

Table 4

Precision of semi-automated 6-point assessment measured in the frame of reference of local vertebral level, errors are normalised by local vertebral widths

Level	AP precision (%)	SI precision (%)	Analysed
T5	3.1	5.0	6
T6	2.6	3.9	7
T7	2.1	4.0	7
T8	2.2	4.5	7
T9	2.5	3.9	7
T10	2.6	4.6	7
T11	2.1	3.7	6
T12	2.0	3.0	7
L1	1.1	1.6	13
L2	1.1	1.5	13
L3	0.8	1.4	13
L4	0.8	1.3	13

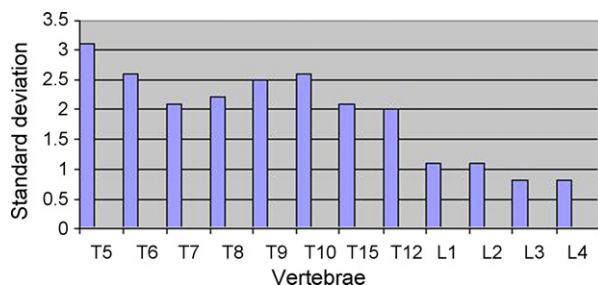


Fig. 5. AP precision for each vertebral label.

normalised in terms of vertebral width. We conclude, therefore, that for the vast majority of images, system precision is high enough to lead to operator independence, that is, the precision of the system is comparable to published figures for intra-operator reproducibility for manual 6-point vertebral morphometry [24]. This is an extremely important aspect of any semi-automatic system.

MorphoXpress is an automated 6-point morphometric system operating on radiographs, as opposed to DXA. Radiographs suffer from the disadvantage that firstly, more than one film is needed to image the whole spine and secondly, they exhibit projection artefacts. Nonetheless, unlike DXA, radiographs are still the only approved modality for diagnosis of vertebral fracture in prospective clinical trials [25].

Medical images may present noisy and, in many cases, incomplete visualisation of anatomical structures and typically represent complex anatomical structures that may vary significantly in both shape and textural appearance. Model-based methods offer solutions to these difficulties [19]. The major benefit of statistical shape models is that by applying simple constraints on the parameters, the generated shape used to describe the underlying image will always remain within physically reasonable bounds. Also, correlations between different parts of the shape are modelled; thus, even if portions of the anatomical object being described are either missing in the image or else obscured by noise, their approximate location and shape may be inferred.

However, for most of the upper thoracic vertebrae (T5–T10), the failure rate is greater than 20% but further investigation reveals that the majority of failures can be attributed to a small number of poor quality images for which multiple vertebrae were discounted. Also, the greater accuracy and precision in the L1–L4 region may be due to less visual clutter in the image from overlying structures such as ribs. In our study all vertebral deformities were due to fractures and not due to other pathologies, and the software failures were due to poor quality images as well as due to the presence of severe vertebral fractures. These observations require further investigation. Another interesting observation is that there is a minimum in the precision error results for the thoracic vertebrae at T7, which is the same vertebra used as the centre of the radiograph acquisition protocol. This suggests that precision is affected by parallax error; an acknowledged factor in radiographic analysis.

It is also interesting to separate the results by analogue and digital acquisition to examine if there is any difference between

analogue and digital mean RMS errors. The digital images appear to have slightly higher errors in T5–T10 region and lower errors in T11–L4 compared to analogue images but statistical analysis shows a lack of significance due to a paucity of successfully analysed vertebrae in this region to provide sufficient statistical power.

Better system performance is to be expected on the lumbar spine as these vertebrae often exhibit greater clarity in radiographic appearance. It is interesting to note that the number of failed searches for a given vertebra type does not appear to be related to the number of vertebrae analysed. This adds weight to the argument that some vertebrae present a bigger problem to the automated analysis, rather than failures being the result of some form of systematic global error. One may also speculate as to the cause of the different accuracy figures for AP and SI projections. For example, the superior and inferior vertebral edges are more salient than the anterior and, particularly, the posterior edges. As with human annotation, greater image evidence will lead to a more accurate result. Further analysis would be necessary, however, to statistically verify this conjecture. We compared the accuracy of the device by defining the ground-truth landmark data as those which were the result of manual correction to the semi-automatic annotation result. As previously discussed, it may be argued that it might introduce some bias into the measurement and it would be instructive to carry out a further study using previously manually annotated images to test this.

Given the workflow context of the tool's use, in our next study we plan to compare the time taken for typical analyses with the tool and compare it with other published approaches. Initial feasibility work indicates that it takes a similar time to perform a complete quantitative 6-point morphometric analysis on a "full" spine (T5–L4) using MorphoXpress as it does to perform the Genant semiquantitative analysis.

The introduction of software technologies such as the one tested here that allows digital clinical images to be analysed automatically or semi-automatically is increasing. This technology has the scope to help the clinician in diagnosis but is not designed to substitute the clinician's role. In the diagnosis of osteoporosis using radiographs, a vertebral deformity is not always a vertebral fracture, but a vertebral fracture is always a vertebral deformity. From the perspective of the radiologist, there is a list of potential differential diagnoses for vertebral deformities, such as osteoporosis, trauma, degenerative disease, Scheuermann's disease, congenital anomaly, neoplastic disease and haematopoietic disorders, infectious disease and Paget's disease, that should be taken into consideration and the correct classification of vertebral deformities can be achieved only by expert interpretation of the radiograph [26]. However, we believe that the technology tested here is highly reproducible, not time intensive or cumbersome in routine clinical use, and can, therefore, be recommended in large epidemiological studies and pharmaceutical trials for the reporting of osteoporotic vertebral fractures.

Conflict of interest

The authors declared no conflict of interests.

Appendix A

To measure accuracy, RMS point-to-point AP and SI errors between the semi-automated annotation points and corrected results were calculated for each vertebra type using Eq. (1).

$$e_{T(\text{AP or SI})} = \frac{\sum_{i=1 \text{ to } N} \sqrt{\sum_{j=1 \text{ to } 6} ([p_{ij} - g_{ij}]_{\text{AP or SI}} / w_i)^2}}{N} \quad (1)$$

where e_T is the defined as the mean error (AP or SI) for the vertebra type labelled T (T5–L4) in the test set; p_{ij} is the position (in pixels) of the j th semi-automatically annotated point for the i th vertebra of type T ; g_{ij} is the position (in pixels) of the j th manually corrected point for the i th vertebra of type T ; w_i is defined as the width (in pixels) of the vertebra, as described above; N is the number of vertebra of type T under investigation in the test set.

To measure precision, the mean RMS point-to-point AP and SI errors between the automatically generated annotation points and the mean of those points across all 10 examples were measured for each vertebra type using Eq. (2).

$$e_{T(\text{AP or SI})} = \frac{\sum_{i=1 \text{ to } N} \sqrt{\frac{(1/10) \sum_{j=1 \text{ to } 6} \sum_{k=1 \text{ to } 10} \times ([p_{ijk} - (1/10) \sum_{k=1 \text{ to } 10} p_{ijk}]_{\text{AP or SI}} / w_i)^2}}{N}} \quad (2)$$

where e_T is the defined as the mean error (AP or SI) for the vertebra labelled T in the test set; p_{ijk} is the position (in pixels) of the j th semi-automatically annotated point for the i th vertebra of type T in the k th experiment; w_i is defined as the width (in pixels) of the i th vertebra as described above; N is the number of vertebra of type T under investigation in the 20 image test set.

References

- [1] Cummings SR, Melton III LJ. Epidemiology and outcomes of osteoporotic fractures. *Lancet* 2002;359:1761–7.
- [2] Eastell R, Cedel SL, Wahner HW, Riggs BL, Melton III LJ. Classification of vertebral fractures. *J Bone Miner Res* 1991;6:207–15.
- [3] Genant H, Wu C, Kijik CC, Nevitt M. Vertebral fracture assessment using semiquantitative technique. *J Bone Miner Res* 1993;8:1137–48.
- [4] McCloskey E, Spector T, Eyres K, et al. The assessment of vertebral deformity: a method for use in population studies and clinical trials. *Osteoporos Int* 1993;3:138–47.
- [5] Melton III LJ, Kan S, Frye M, Wahner F, O'Fallon W, Riggs B. Epidemiology of vertebral fractures in women. *Am J Epidemiol* 1989;129:1000–11.
- [6] Jiang G, Eastell R, Barrington NA, Ferrar L. Comparison of methods for the visual identification of prevalent vertebral fracture in osteoporosis. *Osteoporos Int* 2004;15(11):887–96.
- [7] Ferrar L, Jiang G, Adams J, Eastell R. Identification of vertebral fractures: an update. *Osteoporos Int* 2005;16(7):717–28.
- [8] Ferrar L, Jiang G, Armbrecht G, et al. Is short vertebral height always an osteoporotic fracture? The Osteoporosis and Ultrasound Study (OPUS). *Bone* 2007;41(1):5–12.
- [9] Blake GM, Rea JA, Fogelman I. Vertebral morphometry studies using dual energy X-ray absorptiometry. *Semin Nucl Med* 1997;27:276–90.
- [10] Genant HK, Li J, Wu CY, Shepherd JA. Vertebral fractures in osteoporosis: a new method for clinical assessment. *J Clin Densitom* 2000;3:281–90.
- [11] Greenspan SL, von Stetten E, Emond SK, Jones L, Parker RA. Instant vertebral assessment. *J Clin Densitom* 2001;4:373–80.
- [12] Rea JA, Chen MB, Li J, et al. Morphometric X-ray absorptiometry and morphometric radiography of the spine: a comparison of prevalent vertebral deformity identification. *J Bone Miner Res* 2000;15:564–74.
- [13] Roberts M, Cootes T, Adams JE. Vertebral morphometry: semi-automated determination of detailed vertebral shape from dual-energy X-ray absorptiometry images using active appearance models. *Invest Radiol*; 41(12):849–59.
- [14] Smyth PP, Taylor CJ, Adams JE. Vertebral shape: automatic measurement with active shape models. *Radiology* 1999;211:571–8.
- [15] de Bruijne M, Nielsen M. Image segmentation by shape particle filtering. *IEEE ICPR Comp Soc Press*; 2004. pp. 722–25.
- [16] Howe B, Gururajan A, Sari-Sarraf H, Long R. Hierarchical segmentation of cervical and lumbar vertebrae using a customized generalized Hough transform and extensions to active appearance models. In: *Proceedings of IEEE 6th SSIAP*. 2004. p. 182–6.
- [17] Zamora G, Sari-Sarraf H, Long R. Hierarchical segmentation of vertebrae from X-ray images. *Proc SPIE Med Imag* 2003;5032:631–42.
- [18] Cootes TF, Taylor CJ. Statistical models of appearance for medical image analysis and computer vision. *Proc SPIE Med Imag* 2001;3:138–47.
- [19] Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models—their training and application. *Comput Vis Image Understand* 1995;61:38–59.
- [20] Cootes TF, Hill A, Taylor CJ, Haslam J. The use of active shape models for locating structures in medical images. *Image Vision Comput* 1994;6:276–85.
- [21] Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. In: Burkhardt H, Neumann B, editors. *5th European Conference on Computer Vision*, vol. 2. 1st edition Berlin, Germany: Springer (Berlin); 1998. p. 484–98.
- [22] Nuti R, Guglielmi G, Capelli G, et al. Vertebral deformities prevalence in postmenopausal women: a population-based national study. *J Bone Miner Res* 2003;18(S2):152.
- [23] Angeli A, Guglielmi G, Dovio A, et al. High prevalence of asymptomatic vertebral fractures in postmenopausal women receiving chronic glucocorticoid therapy: a cross-sectional outpatient study. *Bone* 2006;39(2):253–9.
- [24] Grados F, Roux C, de Vernejoul MC, Utard G, Sebert JL, Fardellone P. Comparison of four morphometric definitions and a semiquantitative consensus reading for assessing prevalent vertebral fractures. *Osteoporos Int* 2001;12:716–22.
- [25] Food and Drug Administration. Guidelines for preclinical and clinical evaluation of agents used in the prevention or treatment of postmenopausal osteoporosis; April 1994.
- [26] Link MT, Guglielmi G, van Kuijk C, Adams JE. Radiologic assessment of osteoporotic vertebral fractures: diagnostic and prognostic implications. *Eur Radiol* 2005;15:1521–32.