Classification of Osteoporotic Vertebral Fractures Using Shape and Appearance Modelling

Paul A. Bromiley[1(✉)], Eleni P. Kariki[2], Judith E. Adams[2], and Timothy F. Cootes[1]

[1] Centre for Imaging Sciences, School of Health Sciences,
University of Manchester, Manchester, UK
paul.bromiley@manchester.ac.uk

[2] Radiology and Manchester Academic Health Science Centre,
Central Manchester University Hospitals NHS Foundation Trust,
Manchester, UK

Abstract. Osteoporotic vertebral fractures (VFs) are under-diagnosed, creating an opportunity for computer-aided, opportunistic fracture identification in clinical images. VF diagnosis and grading in clinical practice involves comparisons of vertebral body heights. However, machine vision systems can provide a high-resolution segmentation of the vertebrae and fully characterise their shape and appearance, potentially allowing improved diagnostic accuracy. We compare approaches based on vertebral heights to shape/appearance modelling combined with *k*-nearest neighbours and random forest (RF) classifiers, on both dual-energy X-ray absorptiometry images and computed tomography image volumes. We demonstrate that the combination of RF classifiers and appearance modelling, which is novel in this application, results in a significant (up to 60% reduction in false positive rate at 80% sensitivity) improvement in diagnostic accuracy.

Keywords: Osteoporosis · Vertebral fracture · Shape modelling

1 Introduction

Osteoporosis is a common skeletal disorder characterised by a reduction in bone mineral density (BMD). This is commonly assessed using dual-energy X-ray absorptiometry (DXA); a T-score of < -2.5 (i.e. more than 2.5 standard deviations below the mean in young adults) [1] is used as a criterion suggesting osteoporosis. It significantly increases the risk of fractures, most commonly occurring in the hips, wrists or vertebrae. Approximately 40% of postmenopausal Caucasian women are affected, increasing their lifetime risk of fragility fractures to as much as 40% [1]. Osteoporosis therefore presents a significant public health problem for an ageing population. However, between 30–60% of vertebral fractures (VFs) may be asymptomatic and only about one third of those present on

B. Glocker et al. (Eds.): MSKI 2017, LNCS 10734, pp. 133–147, 2018.
https://doi.org/10.1007/978-3-319-74113-0_12

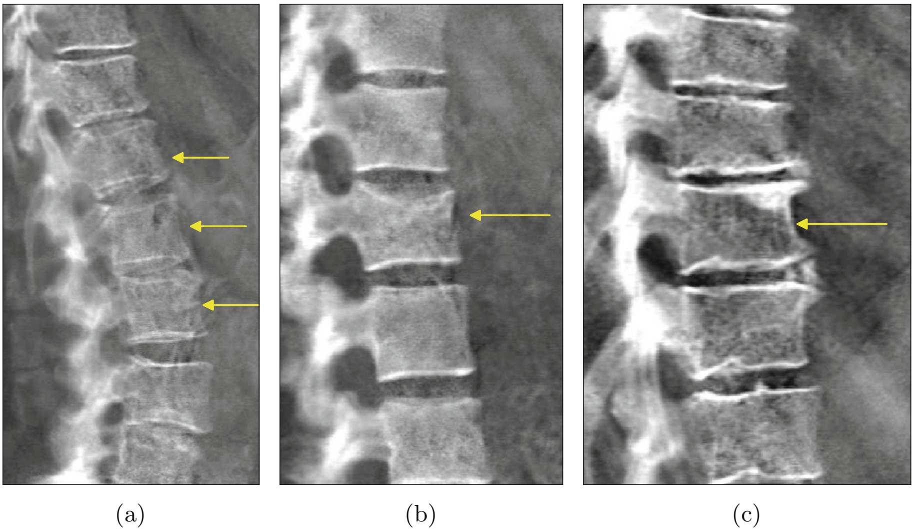


Fig. 1. Example vertebrae visualized on GE Luna iDXA images. (a) Mild scoliosis or incorrect positioning of the patient may cause tilting of the vertebrae relative to the beam direction, leading to the appearance of end-plate curvature although no fractures are present. (b) Osteoporotic vertebral fractures lead to loss of vertebral height and changes in texture due to the presence of micro-fractures; the upper end plate is not symmetrical with the lower and appears blurred. (c) Spondylosis also results in changes in vertebral shape and texture, due to the growth of osteophytes on the anterior portion of the end-plates and sclerosis, i.e. the high-intensity region within the vertebral body, below the anterior portion of the upper end-plate.

images come to clinical attention; they are frequently not reported by radiologists [2]. Many of these cases involve images acquired for other clinical indicators, so identification may be opportunistic. For example, computed tomography (CT) is arguably the ideal modality for opportunistic identification, due to the large number of procedures (4.3 million per year within the UK National Health Service [3]) and high image quality. However, a recent audit at the Manchester Royal Infirmary (MRI) revealed that only 13% of VFs visible on CT images were identified [4], similar to rates reported in the literature [2]. Proposed reasons for such low rates [2] include the difficulty of identifying vertebral height reduction on axial images. Routine coronal and/or sagittal reformatting has been proposed, and is being adopted, but reporting rates remain low [2,5]. The potential utility of computer-aided VF assessment (VFA) systems is therefore considerable.

Several authors have investigated the use of methods based on statistical shape models (SSMs) [6] to segment vertebrae in both radiographs, e.g. [7], and DXA images, e.g. [8,9]. In particular, the random-forest regression voting constrained local model (RFRV-CLM)[10] has been used for both semi- and fully automatic vertebral body segmentation in both DXA [11–13] and CT [14] images, providing superior segmentation accuracy on more severely fractured vertebrae compared to previous work using active appearance models (AAMs) [15]. However, all of these approaches share the common aim of providing a high-resolution segmentation of the vertebrae, typically as landmarks annotated on the vertebral body outline.

Several procedures for manual VFA have been described in the literature; see [16] for a recent review. Most attempt to remove the subjectivity of qualitative assessment [17] by defining fractures in terms of height reduction at the posterior, middle and anterior parts of the vertebral bodies, e.g. [18]. However, non-fracture deformities can affect vertebral shape, as shown in Fig. 1. In particular, tilting of the vertebrae can mimic the appearance of depressed end-plates. VFA therefore requires a subjective assessment of whether any apparent shape change is due to osteoporotic fracture or some other cause. The algorithm-based quantitative (ABQ) method for VFA [19] defined a heuristic for this process. The result is a complex procedure that involves consideration of multiple, interacting factors, including the apparent shape of each vertebra and the spine as a whole, which is difficult to translate into a machine vision algorithm.

Whilst machine vision based VFA methods suffer from difficulties in terms of replicating such complex, heuristic approaches, they have a potential advantage in that techniques based on SSMs can provide a precise mathematical description of the entire shape of a vertebra, and quantitatively compare this between vertebrae. We investigate the interaction of these two effects by constructing VF classifiers based on comparing the parameters of shape and appearance models of vertebrae using both k-nearest neighbours (kNNs) and random forests (RFs) [20]. These methods are compared to simple, height-based classifiers on two data sets; 320 DXA VFA images and spinal mid-line sagittal images projected from 402 CT volumes. When RFs were used to classify appearance model parameters, significant reductions in false positive rate (FPR) of ≈30% and ≈60% were achieved at 80% sensitivity for VF identification from automatic and manual landmark annotations, respectively, on both data sets.

2 Method

2.1 Data Collection and Manual Annotation

The picture archiving and communication server (PACS; Centricity Universal Viewer, GE Healthcare, Little Chalford, Buckinghamshire, UK) at the MRI was queried to produce a list of CT scans acquired during May and June 2014 and January to September 2015. Scans from non-trauma patients that included any part of the thoracic or lumbar spine and were of patients over 18 years of age, were selected. This gave a list of 868 patients' scans. The PACS was also queried for non-trauma CT scans during January to April and July to December 2014 in patients over 60 years of age that contained osteoporotic VFs, producing a second list of 132 patients. The sagittal reformatted volumes from both lists were downloaded in DICOM format. 402 volumes were selected to form a training set,

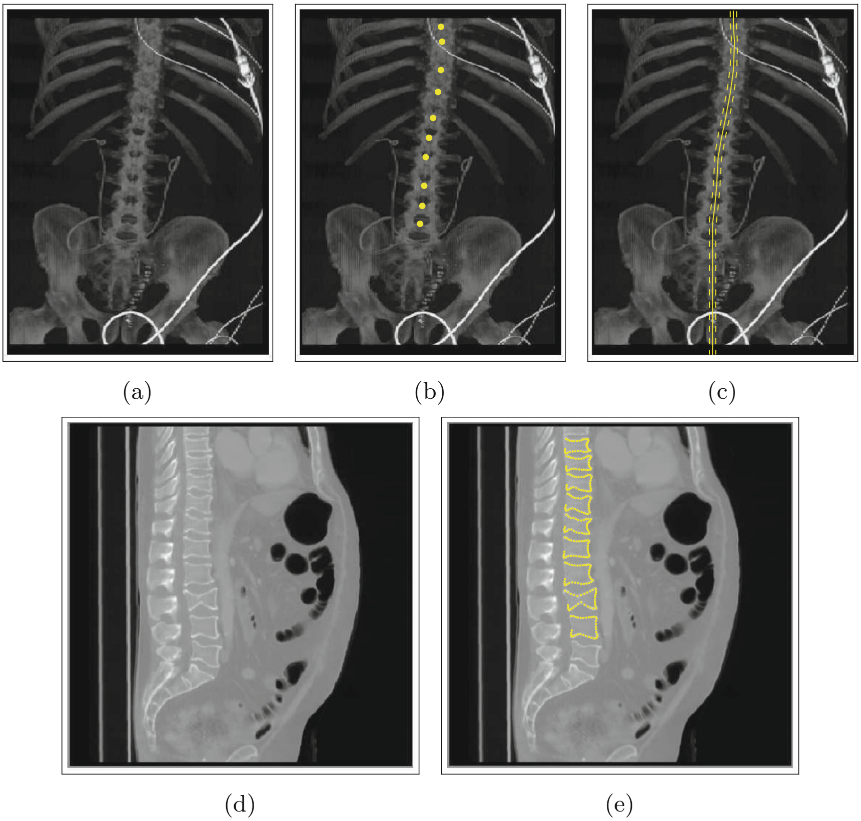


Fig. 2. (a) An example coronal maximum intensity projection of a computed tomography volume. (b) Manual annotations of the neural arch. (c) Extrapolated piecewise-linear curve and the ±5 mm range (dashed line) over which sagittal rasters were summed to produce the sagittal projection (d). (e) Manual landmark annotation.

including the 132 fracture-rich images to ensure high fracture prevalence. The remaining images were reserved for future validation purposes. The 402 image list was divided into quarters for four-fold cross validation, with the fracture-rich images distributed evenly. Each volume was up-sampled to give isotropic voxel dimensions using tri-cubic interpolation.

To avoid the difficulties of performing a high-resolution annotation of landmarks on vertebral bodies in 3D, analysis was limited to a single, two-dimensional (2D) image produced from each volume using the procedure described in [14]. The orientation of the subject within the CT scanner was highly constrained, allowing production of a maximum intensity projection showing an approximately anteroposterior view without registration. Landmarks were manually annotated on the MIP images at the distinctive, U-shaped structure on each vertebra where the laminae join to form the spinous process of the neural arch (Fig. 2(a), (b)).

A piecewise-linear curve was defined through the points and extrapolated vertically to the boundary of the volume (Fig. 2(c)). For each axial slice from the original volume, all anteroposterior raster lines (i.e. rasters of sagittal slices) that passed within D_t of this curve were averaged to give a single raster line of a sagittal image. Repeating for all axial images gave a single, thick-slice, 2D sagittal image that showed the midplane of each vertebra (Fig. 2(d)). The thickness $D_t = \pm 5$ mm was chosen by manual inspection to optimise endplate visualisation.

The images derived from CT volumes were projected onto the spinal midline, and so were unaffected by issues such as mild scoliosis. DXA images are projections through the full body, and so will show the tilting of the vertebrae encountered in this condition, making accurate diagnosis based on vertebral shape more difficult. Therefore, a second data set of 320 DXA VFA images scanned on various Hologic (Bedford, MA, USA) scanners was also used. This comprised: 44 patients from a previous study [21]; 80 female subjects in an epidemiological study of a UK cohort born in 1946; 196 females attending a local clinic for DXA BMD measurement, for whom the referring physician had requested VFA (approved by the local ethics committee).

Manual annotation of 33 landmarks on each visualized vertebra from T7 to L4, for the DXA images, and T4 to L4, for the CT midline images (Fig. 2(e)), was performed by a trained radiographer. The vertebrae were also classified by an expert radiologist into five groups: normal; deformed but not fractured; and mild (grade 1), moderate (grade 2) and severe (grade 3) fractures using the Genant definitions [18].

2.2 Height-Based Fracture Classification

A baseline for VF classification accuracy was derived by applying a simple classifier, based on six-point morphometry, as described in [13]. The anterior h_a, middle h_m and posterior h_p heights of each detected vertebra were calculated from the relevant landmarks, together with a predicted posterior height $h_{p'}$, calculated as the maximum of the posterior heights of the four closest vertebrae. The wedge $r_w = h_a/h_p$, biconcavity $r_b = h_m/h_p$, and crush $r_c = h_p/h_{p'}$ ratios were derived, and the data were whitened by subtracting the medians of each ratio and dividing by the square-root of the covariance matrix, calculated using the median standard deviation. The data contained far more normal than deformed or fractured vertebrae, so this process whitened to the distribution of the normal class. A simple fracture/non-fracture classification was performed by applying a threshold t_{class} to $r_c^2 + r_b^2 + r_w^2$; deformed vertebrae were counted correct when classified into either class.

2.3 Shape and Appearance Model Based Classifiers

SSMs provide a linear model of the distribution of a set of landmarks in an image. The training data consists of a set of images $\mathbf{I}$ with manual annotations $\mathbf{x}_l$ of a set of N points $l = 1, \ldots, N$ on each. The images are first aligned into a standardised

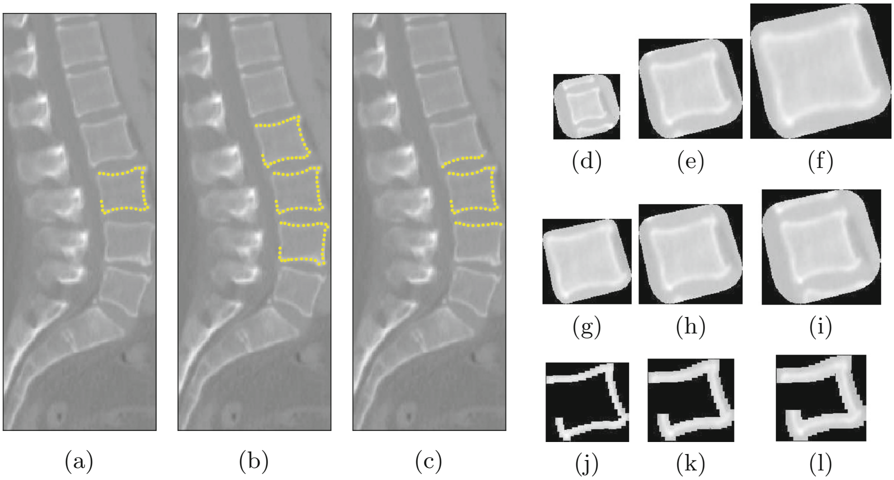


Fig. 3. Point sampling strategies, using (a) a single vertebra, (b) a triplet and (c) a vertebra plus its neighbouring end-plates (nEP). Texture sampling strategies in the single vertebra case. (d)–(f) Sampling using Delaunay triangulation with increasing w_{frame}. (g)–(i) Sampling using Delaunay triangulation with increasing w_{border}. (j)–(l) Patch-based sampling with increasing w_{patch}.

reference frame using a similarity registration, giving a transformation T with parameters θ. The concatenated, reference-frame coordinates of the points in each training image define its shape. The SSM is generated by applying principal component analysis (PCA) to the training shapes [15], generating a linear model where the position of point l is given by

$$\mathbf{x}_l = T_\theta(\bar{\mathbf{x}}_l + \mathbf{P}_{sl}\mathbf{b}_s), \tag{1}$$

where $\bar{\mathbf{x}}_l$ is the mean point position in the reference frame, $\mathbf{P}_s$ is a matrix of modes of variation, $\mathbf{P}_{sl}$ is the sub-matrix of $\mathbf{P}_s$ relevant to point l, and $\mathbf{b}_s$ encodes the shape model parameters. The matrix $\mathbf{P}_s$ is orthogonal and so

$$\mathbf{b}_s = \mathbf{P}_{sl}^T(T_\theta^{-1}(\mathbf{x}_l) - \bar{\mathbf{x}}_l). \tag{2}$$

A compact description of the shape in a query image can therefore be derived by annotating the landmarks, performing a similarity registration into the reference frame of the model, and applying (2) to generate the vector $\mathbf{b}_s$.

The SSM considers only the distribution of landmarks on a shape. However, with reference to osteoporosis and potentially confounding pathologies, information is also present in the pixel intensities. Osteoporotic VF proceeds as a cascade of micro-fractures in the vertebral end-plates [16] leading to a blurred appearance (Fig. 1). Appearance models (APMs) such as those used by AAMs [15] adopt the same PCA-based linear modelling approach as the SSM to characterize both shape and intensity information. Each training image is resampled into the reference frame by applying $\mathbf{I}_r(m,n) = \mathbf{I}(T_\theta^{-1}(m,n))$, where (m,n) specify pixel coordinates. The reference frame width w_{frame} acts as a free parameter controlling the resolution. The intensities of each pixel within an image patch covering the points are then concatenated into a vector $\mathbf{g}$, and PCA applied as before to generate a linear model based on modes of variation $\mathbf{P}_g$

$$\mathbf{g} = T_\phi(\bar{\mathbf{g}} + \mathbf{P}_g\mathbf{b}_g) \quad \text{and} \quad \mathbf{b}_g = \mathbf{P}_g^T(T_\phi^{-1}(\mathbf{g}) - \bar{\mathbf{g}}), \tag{3}$$

where T_ϕ represents an intensity normalisation. Correlations may exist between the shape $\mathbf{b}_s$ and intensity $\mathbf{b}_g$ parameters, and so the models are concatenated and a further PCA performed to extract the independent modes of variation of both shape and intensity $\mathbf{P}_c$, referred to as appearance modes

$$\mathbf{b} = \mathbf{P}_c\mathbf{c} \quad \text{where} \quad \mathbf{b} = \begin{pmatrix} \mathbf{W}_s\mathbf{b}_s \\ \mathbf{b}_g \end{pmatrix}. \tag{4}$$

The weights $\mathbf{W}_s$ scale the relative magnitude of the shape and intensity parameters, and are derived by sampling the change in $\mathbf{g}$ per unit change in $\mathbf{b}_s$ in the training images.

Two approaches for sampling the intensities contributing to $\mathbf{g}$ were tested (Fig. 3). The first used a Delaunay triangulation of the landmarks to define a region of interest, with an optional, additional border of width w_{border} to ensure the whole edge was included. However, most relevant intensity information was expected to be located close to the end-plates i.e. the site of the fractures. Therefore, an alternative strategy that involved sampling a square patch of width w_{patch} around each landmark was also implemented. Multiple approaches for defining the landmarks used were also tested (Fig. 3), including sampling from a single vertebra, a triplet of neighbouring vertebrae, and a vertebra plus the closest end-plates of its neighbours (nEP sampling). The latter were intended to aid in identification of tilted vertebra, since these are distinguished from fractures, in clinical practice, by the symmetry of adjacent end-plates (Fig. 1).

The SSM or APM extract all significant shape and intensity information from an image as a compact vector of features. A variety of classifiers could then be applied to compare the features of a query image to those of annotated and diagnosed training images. Two were studied here. First, kNN was applied, measuring Euclidean distance in the feature space of $\mathbf{b}$ or $\mathbf{b}_s$ to identify N_{kNN} neighbours. However, kNN has the drawback that all features are considered; some will not be relevant to fracture status and so will potentially act as confounding information. To determine whether this effect was significant, classification was also performed using RFs. Since each split node considers a single feature, an RF has the capability to identify only those features relevant to the target. In both cases, the problem was treated as a regression task. The gold-standard diagnosis was translated into a numerical score with $0 = \text{normal}$, $1 = \text{deformed but not fractured}$ and 2, 3, and $4 = \text{mild, moderate and severe fracture}$. The output of the kNN was the mean of this score across the identified neighbours, and RFs were trained as regressors to predict the score. This created a potential problem as it assumed that deformed vertebrae are intermediate in

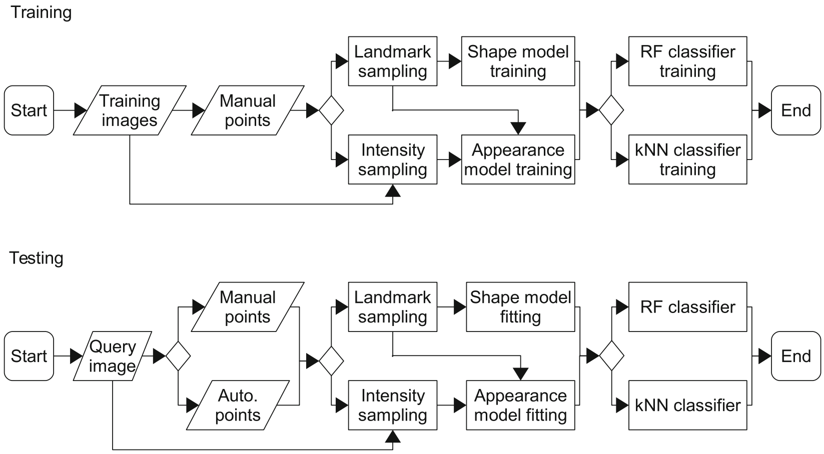


Fig. 4. Flow-chart summarising the various algorithmic choices. See main text for description.

shape between normal and fractured ones, which may not be the case. However, it more accurately represented fracture status as a position within a continuum, rather than discrete classes. The alternative, treating the problem as an explicit classification task, was investigated using a multi-class RF classifier. However, this led to significantly worse accuracy and full results are not reported here. As with the height-based classifier, a simple fracture/non-fracture classification was performed by thresholding the kNN or RF output.

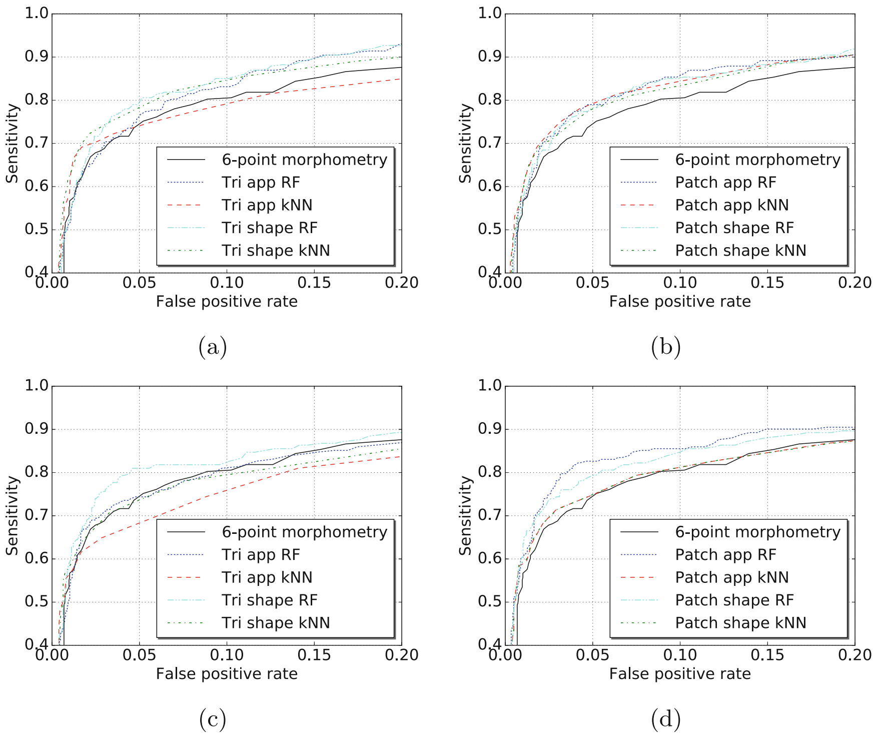
Figure 4 summarises the various algorithmic choices that were evaluated. During testing, a 2D query image was input. This was either a DXA image or a thick-slice sagittal midline projection from a CT volume, produced as described in Sect. 2.1 using manual annotations of neural arch landmarks. The authors have previously described an algorithm that can automatically produce these projections [14]. This was not considered here due to lack of space. Landmark points outlining each visualised vertebral body, annotated either manually or using an automatic approach as described in Sect. 3, were then input (choice 1). The feature space (shape or appearance) and sampling procedure (single vertebra, nEP or triplet) were then chosen as described in Sect. 2.3 (choice 2). Shape modelling required sampling only from the landmark points; appearance modelling required sampling from both the points and image intensities. For each visualised vertebra in the query image, the chosen model was fitted to the sampled data. The resultant shape or appearance features were then passed to a kNN or RF classifier (choice 3) to obtain the final classification for each vertebra. All combinations of feature space, sampling procedure and classifier were evaluated. Model and classifier training was performed using manual annotations but otherwise followed a similar work-flow.

3 Evaluation

Throughout the evaluation, classifiers were trained and tested in a leave-1/4-out procedure using the data from all vertebral levels. During SSM and APM training, the number of modes of variation was constrained to model 98% of the variation in the data. Free parameters were empirically optimised on the CT midline images and manual annotations, such that the DXA images served as an independent evaluation set. Initially, the shape and appearance model parameters were optimised in combination with a kNN classifier. The latter had only one free parameter, N_{kNN}, greatly reducing the dimensionality of the parameter space compared to using a RF. Receiver operator characteristic (ROC) curves showing sensitivity against false positive rate (FPR) were generated by varying t_{class}, and the parameters leading to the highest value at which sensitivity equalled FPR were selected, giving $w_{frame} = 80$ pixels, $w_{border} = 30$ pixels for triangulated intensity sampling, $w_{patch} = 24$ pixels for patch-based sampling, and $N_{kNN} = 10$. This was repeated for all sampling strategies described in Sect. 2.3, and the optimae were consistent. In general, dependence on the parameters was weak for all except w_{frame}. A second round of parameter optimisation focused on the RF, using the optimised parameters for SSM/APM described above. The same ROC-curve based pattern search procedure was used to optimise the number of trees n_{tree}, the maximum depth of each tree D_{max}, and the minimum number of training samples n_{min} allowed at a split node, leading to $n_{tree} = 200$, $D_{max} = 30$ and $n_{min} = 1$.

Finally, semi-automatic annotation of the vertebrae in both the DXA and CT midline images, the latter projected using the manual neural arch annotations, was performed using a RFRV-CLM, initialised using manual annotations of vertebral centre points, following the procedure described in [12,14]. The classifiers were then applied to the automatic annotations using the optimised parameters. Classification accuracies for the optimised procedure with all combinations of features, classifier, patch and intensity sampling procedure and manual or automatic annotation were then compared.

Figures 5 and 6 show the evaluation of various sampling, feature extraction and classification procedures for the CT and DXA images, respectively, compared to baselines established by the six-point morphometry approach. A universal trend was noted across all experiments; accuracy for triplet sampling was always significantly worse than the alternatives (full results are not shown for this reason). Triplet sampling results in more modes of variation and so more features in $\mathbf{b}$ and $\mathbf{b}_s$, and adding the neighbouring vertebrae spreads the training data across the feature space depending on the fracture status of the neighbours. Both reduce data density. In contrast, whilst 6-point morphometry benefits from comparison of posterior vertebral heights between neighbours to identify crush fractures, where the height is reduced throughout a vertebral body, shape models can extract equivalent information from a single vertebra through quantification of its aspect ratio. Triplet sampling therefore adds little information, but makes it more difficult for a classifier to extract the information present.



(a)

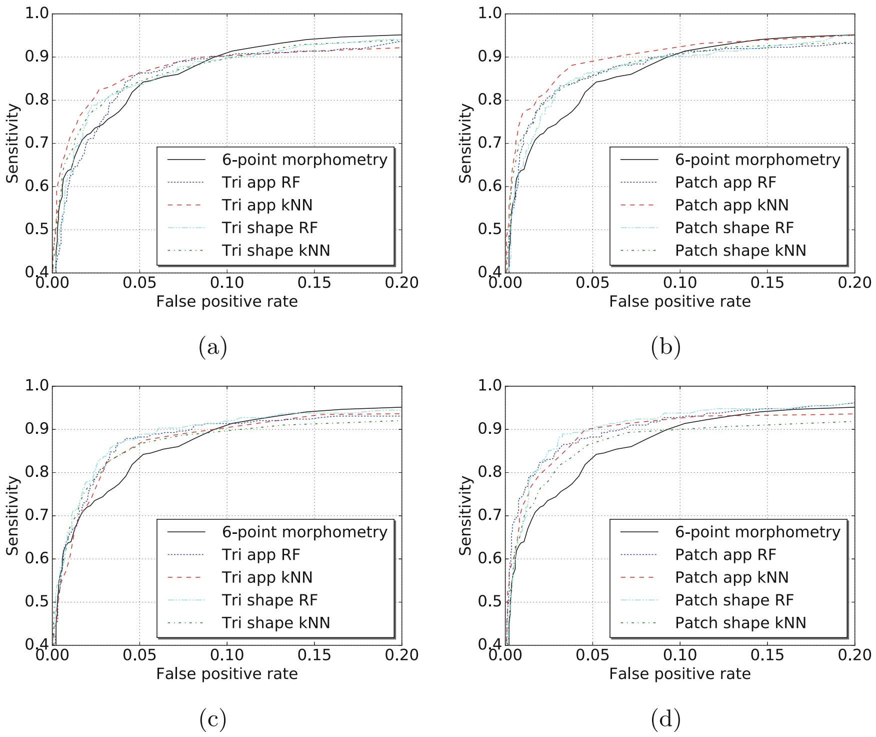
(b)

(c)

(d)

Fig. 5. Receiver operator characteristic curves for vertebral classification on computed tomography midline images. (a), (b) Point sampling from single vertebrae. (c), (d) Point sampling from vertebrae plus neighbouring end-plates. "Tri" refers to intensity sampling from Delaunay triangulated regions, "patch" to sampling from patches around each landmark point, "shape" to shape and "app" to appearance features.

Differences between the remaining sampling/modelling/classification procedures were usually small and frequently not significant, but the optimal procedure was always significantly better than 6-point morphometry. Several trends emerged from the results. On CT images (Fig. 5), there was little evidence of additional information in appearance compared to shape features. The kNN classifier showed a marked reduction in performance when triangulated intensity sampling was used, adding large numbers of uninformative appearance features (Fig. 5(a), (c)). The RF classifier also showed some evidence of this effect, with the combination of triangulated nEP sampling and appearance features resulting in performance no better than the baseline (Fig. 5c). However, the RF, in general, resulted in better accuracy than the kNN classifier and did not lose accuracy when nEP sampling was used, or when using appearance features if the number of features was controlled using patch-based sampling (Fig. 5(b), (d)).



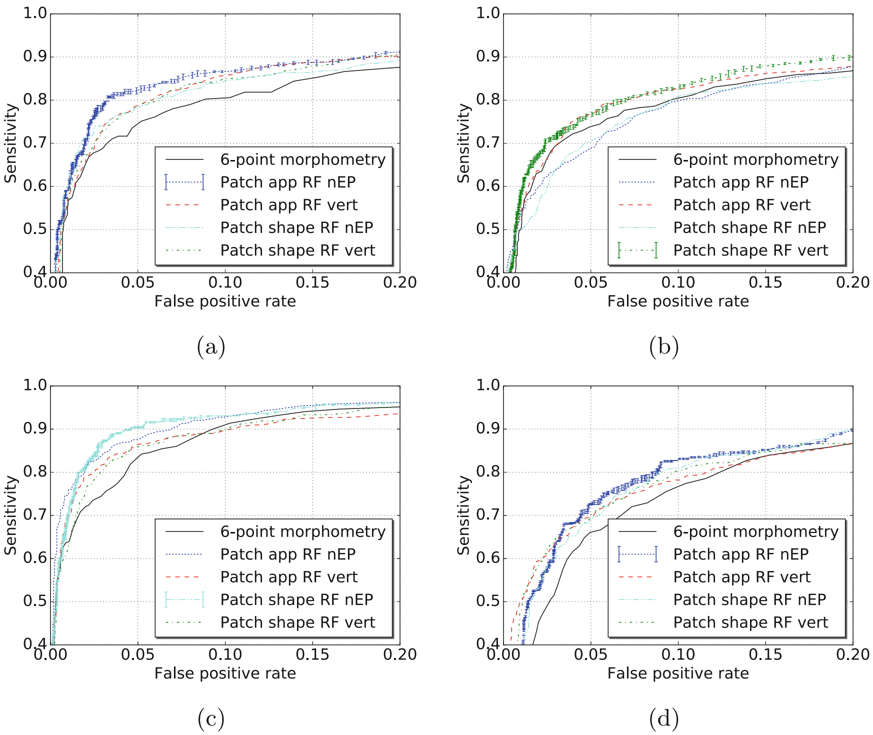
(a) (b)

(c) (d)

Fig. 6. Receiver operator characteristic curves for vertebral classification on dual-energy X-ray absorptiometry images. (a), (b) Point sampling from single vertebrae. (c), (d) Point sampling from vertebrae plus neighbouring end-plates. “Tri” refers to intensity sampling from Delaunay triangulated regions, “patch” to sampling from patches around each landmark point, “shape” to shape and “app” to appearance features.

DXA images (Fig. 6) represented a more challenging task since, being projections of the full vertebral body, the shape as visualised in the images was more complex, and the image quality was lower. Increased noise on individual features resulted in kNN outperforming RF when single-vertebra sampling was used (Fig. 6(a), (b)), with accuracy gains resulting from using appearance as long as the length of the feature vector was controlled by using patch-based sampling (Fig. 6(b)). However, using shape information from neighbouring vertebrae through nEP sampling, to deal with cases where the vertebrae were tilted relative to the beam direction, allowed the RF to achieve equal or better performance on shape alone (Fig. 6(d)).

Figure 7 shows results from the optimised procedure, using patch-based sampling and RF classification, for both manual and automatic annotations on both image sets. Each experiment was repeated five times, using the stochastic nature of RF training to support error estimation, and the figures show the mean and (where shown) standard deviation of the repeats. As described above, nEP



(a)

(b)

(c)

(d)

Fig. 7. Receiver operator characteristic curves for random forest classification with patch sampling in computed tomography midline (a), (b) and dual-energy X-ray absorptiometry (c), (d) images, from manual (a), (c) and RFRV-CLM (b), (d) annotations, compared to 6-point morphometry. "Shape" refers to shape and "app" to appearance features.

sampling resulted in increased accuracy for both image types when classifying from manual annotations (Fig. 7(a), (c)). However, this requires accurate annotations of all vertebrae in the triplet, increasing the risk of a fit failure being present in the automatic annotations. Single-vertebra sampling was therefore more accurate on RFRV-CLM annotations of CT images and appearance provided no additional information over shape (Fig. 7(b)). However, the difficulties of the classification task in DXA images, described above, resulted in higher accuracies when nEP sampling and appearance parameters were used, to disambiguate tilted vertebra and confounding pathologies (Fig. 7(d)). At 80% sensitivity, the optimal classifier reduced the FPR from 4.4% to 1.7% for manual and from 12.7% to 8.8% for automatic DXA annotations. For CT images, FPR was reduced from 8.9% to 3.3% for manual and 9.7% to 7.0% for automatic annotations. These equate to a significant ($p < 0.01$) reduction in FPR of ≈30% for automatic, and ≈60% for manual annotations.

4 Conclusion

We have evaluated several different methods for osteoporotic VF classification in DXA VFA and CT midline images. In this context, the 6-point morphometry approach can be viewed as the construction of a low-parameter shape model based on hand-crafted modes of variation. Since these reflect the clinical definitions used in fracture grading [18] they are guaranteed to be informative. A SSM or APM can extract all statistically significant shape and appearance information from the images as a compact feature vector, but not all of these features are guaranteed to be informative in terms of fracture classification. The way in which the subsequent classifier handles non-informative dimensions of the feature space therefore influences accuracy. Straightforward application of kNN considers all features equally, whilst a RF considers only the most informative at each split node, and so is more robust to non-informative features.

The results showed that statistically significant gains in classification accuracy can be achieved by applying kNN or RF classifiers to shape or appearance model features. The optimal procedure across both CT and DXA was to apply a RF classifier to features sampled from a vertebra and the closest end-plates of its neighbours with manual landmark annotations, but fitting errors in automatic annotations resulted in single vertebra sampling being more accurate for CT images. Evidence for improved performance when using appearance, as opposed to shape, features was generally weak and inconsistent. However, it did not result in significant reductions in classifier performance. In conclusion, the combination of appearance features and RF classification with patch-based nEP sampling, for manual annotations, and single vertebra sampling, for automatic annotations, provided optimal results. Significant accuracy gains compared to 6-point morphometry were achieved for both manual and automatic annotations on both DXA VFA and CT images using these approaches.

This work used a single model/classifier for all vertebral levels between T4 (for CT) or T7 (for DXA VFA) and L4, inclusive. Vertebral shape varies gradually across the spine, and so further improvements in accuracy might be gained through using multiple, level-specific classifiers. Roberts et al. [9] achieved higher classification accuracies on DXA VFA images using this approach, and so we intend to investigate this in future work. However, the use of level-specific classifiers requires a reasonably accurate method of level detection if combined with an automatic vertebral segmentation method, which may prove challenging given the similarity between vertebra.

Acknowledgements. This publication presents independent research supported by the Health Innovation Challenge Fund (grant no. HICF-R7-414/WT100936), a parallel funding partnership between the Department of Health and Wellcome Trust, and by the NIHR Invention for Innovation (i4i) programme (grant no. II-LB_0216-20009). The views expressed are those of the authors and not necessarily those of the NHS, NIHR, the Department of Health or Wellcome Trust. The authors acknowledge the invaluable assistance of Mrs Chrissie Alsop, Mr Stephen Capener, Mrs Imelda Hodgkinson, Mr Michael Machin, and Mrs Sue Roberts, who performed the manual annotations.

References

1. Rachner, T., Khosla, S., Hofbauer, L.: Osteoporosis: now and the future. Lancet **377**(9773), 1276–1287 (2011)
2. Adams, J.: Opportunistic identification of vertebral fractures. J. Clin. Densitom. **19**(1), 54–62 (2016)
3. Operational Information for Commissioning: Diagnostic imaging dataset statistical release. Technical report, NHS, UK (2016). http://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2015/08/Provisional-Monthly-Diagnostic-Imaging-Dataset-Statistics-2016-05-19.pdf
4. Williams, A.L., Al-Busaidi, A., Sparrow, P.J., Adams, J.E., Whitehouse, R.W.: Under-reporting of osteoporotic vertebral fractures on computed tomography. Eur. J. Radiol. **69**(1), 179–183 (2009)
5. Kariki, E., Bromiley, P., Cootes, T., Adams, J.: Opportunistic identification of vertebral fractures on computed radiography: need for improvement. Osteoporos. Int. **27**(S2), 621 (2016)
6. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models - their training and application. Comput. Vis. Image Understand. **61**(1), 38–59 (1995)
7. Brett, A., Miller, C., Hayes, C., Krasnow, J., Ozanian, T., Abrams, K., Block, J., van Kuijk, C.: Development of a clinical workflow tool to enhance the detection of vertebral fractures. Spine **34**(22), 2437–2443 (2009)
8. Roberts, M., Cootes, T., Adams, J.: Vertebral morphometry: semiautomatic determination of detailed shape from dual-energy X-ray absorptiometry images using active appearance models. Invest. Radiol. **41**(12), 849–859 (2006)
9. Roberts, M.G., Cootes, T.F., Adams, J.E.: Automatic location of vertebrae on DXA images using random forest regression. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7512, pp. 361–368. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33454-2_45
10. Lindner, C., Bromiley, P., Ionita, M., Cootes, T.: Robust and accurate shape model matching using random forest regression-voting. IEEE Trans. Pattern Anal. Mach. Intell. **37**(9), 1862–1874 (2015)
11. Bromiley, P., Adams, J., Cootes, T.: Localization of vertebrae on DXA VFA images using constrained local models with random forest regression voting. In: Proceedings of 20th International Bone Densitometry Workshop - IBDW 2014 (2014). J. Orthop. Translat., vol. 2, pp. 227–228
12. Bromiley, P., Adams, J., Cootes, T.: Localisation of vertebrae on DXA images using constrained local models with random forest regression voting. In: Yao, J., et al. (eds.) CSI 2014. LNCVB, vol. 20, pp. 159–171. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-14148-0_14
13. Bromiley, P.A., Adams, J.E., Cootes, T.F.: Automatic localisation of vertebrae in DXA images using random forest regression voting. In: Vrtovec, T., Yao, J., Glocker, B., Klinder, T., Frangi, A., Zheng, G., Li, S. (eds.) CSI 2015. LNCS, vol. 9402, pp. 38–51. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41827-8_4
14. Bromiley, P.A., Kariki, E.P., Adams, J.E., Cootes, T.F.: Fully automatic localisation of vertebrae in CT images using random forest regression voting. In: Yao, J., Vrtovec, T., Zheng, G., Frangi, A., Glocker, B., Li, S. (eds.) CSI 2016. LNCS, vol. 10182, pp. 51–63. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-55050-3_5

15. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. IEEE Trans. Pattern Anal. Mach. Intell. **23**(6), 681–685 (2001)
16. Griffith, J.: Identifying osteoporotic vertebral fracture. Quant. Imaging Med. Surg. **5**(4), 592–602 (2015)
17. Jensen, G., McNair, P., Boesen, J., Hegedus, V.: Validity in diagnosing osteoporosis. Observer variation in interpreting spinal radiographs. Eur. J. Radiol. **4**(1), 1–3 (1984)
18. Genant, H., Wu, C., Kuijk, C., Nevitt, M.: Vertebral fracture assessment using a semi-quantitative technique. J. Bone Miner. Res. **8**(9), 1137–1148 (1993)
19. Jiang, G.: Diagnosis of vertebral fracture using an ABQ method. Osteoporos. Rev. **18**(3), 14–18 (2010)
20. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
21. McCloskey, E., Selby, P., de Takats, D., Bernard, J., Davies, M., Robinson, J., Francis, R., Adams, J., Pande, K., Beneton, M., Jalava, T., Loyttyniemi, E., Kanis, J.: Effects of clodronate on vertebral fracture risk in osteoporosis: a 1-year interim analysis. Bone **28**(3), 310–315 (2001)