

# Fully Automatic Localisation of Vertebrae in CT images using Random Forest Regression Voting

Paul A. Bromiley<sup>1</sup>, Eleni P. Kariki<sup>2</sup>, Judith E. Adams<sup>2</sup>, and Timothy F. Cootes<sup>1</sup>

<sup>1</sup> Imaging Sciences Research Group, University of Manchester, UK.

{paul.bromiley, timothy.f.cootes}@manchester.ac.uk

<sup>2</sup> Radiology & Manchester Academic Health Science Centre, Central Manchester University Hospitals NHS Foundation Trust, UK. eleni.kariki@cmft.nhs.uk, judith.adams@manchester.ac.uk

**Abstract.** We describe a system for fully automatic vertebra localisation and segmentation in 3D CT volumes containing arbitrary regions of the spine, with the aim of detecting osteoporotic fractures. To avoid the difficulties of high-resolution manual annotation on overlapping structures in 3D, the system consists of several 2D operations. First, a Random Forest regressor is used to localise the spinal midplane in a coronal maximum intensity projection. A 2D sagittal image showing the midplane is then produced. A second set of regressors are used to localise each vertebral body in this image. Finally, a Random Forest Regression Voting Constrained Local Model is used to segment each detected vertebra.

The system was evaluated on 402 CT volumes. 83% of vertebrae between T4 and L4 were detected and, of these, 97% were segmented with a mean error of  $\leq 1mm$ . A simple classifier was applied to perform a fracture/non-fracture classification for each image, achieving 69% recall at 70% precision.

## 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) [13] is used as a criterion suggesting osteoporosis. It significantly increases the risk of fractures, most commonly occurring in the hip, wrist or vertebra. Approximately 40% of postmenopausal Caucasian women are affected, increasing their lifetime risk of fragility fractures to as much as 40% [13]. Osteoporosis therefore presents a significant public health problem for an ageing population. However, between 30%-60% of vertebral fractures may be asymptomatic and only about one third of those present on images come to clinical attention; they are frequently not reported by radiologists, not entered into medical records, and do not lead to preventative treatments [8]. Many of these cases involve images acquired for purposes other than assessment for the presence of vertebral fractures, so identification may be opportunistic. However, a recent multi-centre, multinational prospective study [9] found a false negative rate of 34% for reporting vertebral fractures from lateral radiographs of the thoracolumbar spine. The potential utility of computer-aided vertebral fracture identification systems is therefore considerable. Modern clinical imaging is primarily digital,

with images acquired in Digital Imaging and Communications in Medicine (DICOM) format and stored on a Picture Archiving and Communication System (PACS). A system that could query a PACS to extract images that include the spine, automatically segment vertebrae, detect any abnormal shape, and report suspect images for further investigation by a radiologist, would therefore be particularly valuable.

CT is arguably the ideal modality for opportunistic osteoporotic vertebral fracture identification, due to the large number of procedures (4.3 million per year within the UK National Health Service [12]) and the high image quality. However, a recent audit at the Manchester Royal Infirmary revealed that only 13% of such fractures visible on CT images were identified [15], similar to identification rates reported in the literature [1]. Proposed reasons for such low rates [1] include the difficulty of identifying vertebral height reduction on axial images. Routine production of coronal and/or sagittal reformatted images has been proposed, and is being adopted, but reporting rates on such images remain low [1].

We describe a system for fully automatic localisation and segmentation of vertebrae in sagittal reformatted CT image volumes covering arbitrary regions of the spine, based on landmark point annotation. Manual annotation on 3D spinal images for model training would be a challenging task, due to the large number of points required to quantify vertebral shape accurately. Therefore, several 2D operations are used. A coronal maximum intensity projection (MIP) of the volume is produced, highlighting the bony structures. Random Forest (RF) Regression Voting (RFRV) is used to localise points on the spine. This takes advantage of the fact that the patient is supine in the CT scanner, and so is not subject to arbitrary rotation in the axial image plane. A single, thick-slice, 2D sagittal image is then produced, showing the midplanes of all vertebrae present. A second set of RF regressors is used to localise the posterior-inferior vertebral corners in this image. Both of these initialisation stages are based on the algorithm described in [4]. Finally, the vertebral corner points are used to initialise a Random Forest Regression Voting Constrained Local Model (RFRV-CLM), based on [3], which provides a high-resolution segmentation of the vertebrae allowing subsequent shape measurement. These algorithms are described briefly in Section 2. The reader is referred to [3] and [4] for a more complete description, and discussion of related literature.

## 2 Method

### 2.1 Random Forest Regression Voting

Random Forest Regression Voting (RFRV) uses a RF [2] regressor to localise a landmark, trained to predict the offset to that point based on local patches of image features. The training data consists of a set of images  $\mathbf{I}$  with manual annotations  $\mathbf{x}_l$  of the point on each. Random displacements  $\mathbf{d}_j$  are generated by sampling from a uniform distribution with apothem  $d_{max}$  and the same dimensionality as the images. Image patches of area  $w_{patch}^2$  are extracted at these displacements from  $\mathbf{x}_l$  in each training image, and features  $\mathbf{f}_j$  are derived from them. Haar-like features [14] are used, as they have proven effective for a range of applications and can be calculated efficiently from integral images. To allow for inaccurate initial estimates of pose during model fitting, and to make the detector locally pose-invariant, the process is repeated with random perturbations in scale

and orientation. A RF is then constructed; each tree is trained on a bootstrap sample of pairs  $\{(\mathbf{f}_j, \mathbf{d}_j)\}$  from the training data using a standard, greedy approach. At each node, a random set of  $n_{feat}$  features is chosen, and a feature  $f_i$  and threshold  $t$  that best split the data into two compact groups are selected by minimising an entropy measure [11]. The process is terminated at a maximum depth  $D_{max}$  or minimum number of samples  $N_{min}$ , and repeated to generate a forest of  $n_{trees}$ .

## 2.2 RFRV Initialiser Fitting

The coronal and sagittal initialisation algorithms used here are based on [4], and use RF regressors trained as described in Section 2.1. An exhaustive search is performed over a query image, by defining a grid of positions with a spacing of 3 pixels. The RFs are applied at each position, and give predictions of the displacement to the landmarks. The search is repeated at a range of angle and scale variation:  $-0.8$  to  $0.8$  radians in steps of  $\theta_r = 0.1$ , and scales from 0.1 to 4 in rational/integer steps. The predicted landmark locations from each tree are collected in a Hough-style voting array. An RF trained to localise a point on a specific vertebral level will respond strongly to the equivalent points on neighbouring vertebrae, due to their similar shapes, predicting the closest to each search position. Full coverage of the spine can be achieved by training a single RF on concatenated data  $\{(\mathbf{f}_j, \mathbf{d}_j)\}$  from multiple levels. Alternatively, RFs can be trained on each level and applied in parallel, voting into a single array. The array is then smoothed using a Gaussian kernel of standard deviation twice the resolution of the search grid, allowing detection of modes using nine-way maximum. Modes with weights lower than 20% of the strongest response are discarded. (In contrast to [4], no additional weighting of the modes was used here). A graphical method is then used to extract an ordered, linked set of modes, representing landmarks on all visible vertebra, and to discard false detections. Starting from the strongest mode, an iterative search is performed in the local inferior and superior directions, determined from the average pose of the RFRV detections for that mode. At each iteration, the closest mode within an angle constraint of  $\theta_l = 2\theta_r$  is added to the set, terminating when no further modes meet the constraint.

## 2.3 Constrained Local Models (CLMs)

The CLM [7] uses a statistical shape model (SSM) to constrain the fitting of multiple, independent RFRVs for a set of landmarks. The training data consists of a set of images  $\mathbf{I}$  with manual annotations  $\mathbf{x}_l$  of a set of  $N$  points  $l = 1 \dots N$  on each. The images are first aligned into a standardised reference frame using a similarity registration, giving a transformation  $T$  with parameters  $\theta$ , and then resampled into this 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, in pixels, is controlled by a parameter  $w_{frame}$ , allowing variation of the resolution of the resampled images. 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 set of training shapes [5]. This yields a linear model of shape variation, giving the position of point  $l$

$$\mathbf{x}_l = T_\theta(\bar{\mathbf{x}}_l + \mathbf{P}_l \mathbf{b} + \mathbf{r}_l) \quad (1)$$

where  $\bar{\mathbf{x}}_l$  is the mean point position in the reference frame,  $\mathbf{P}_l$  is a set of modes of variation,  $\mathbf{b}$  encodes the shape model parameters, and  $\mathbf{r}_l$  allows small deviations from the model. For each point  $l = 1 \dots N$ , an RF  $R_l$  is trained as described in Section 2.1, using data from the resampled images.

#### 2.4 RFRV-CLM Fitting

The fitting of a RFRV-CLM to a query image  $\mathbf{I}_q$  is initialised via an estimate of pose ( $\mathbf{b}$  and  $\theta$ ) from a previous model or a manual initialisation. The image is resampled in the reference frame using the current pose  $\mathbf{I}_{qr}(m, n) = \mathbf{I}_q(T_\theta^{-1}(m, n))$ . For each point  $l$ , a grid of locations  $\mathbf{z}_l$  is defined covering a search range of apothem  $d_{search}$  around the initial estimate of its position. Regressor  $R_l$  is applied to the image features extracted from the local patch around each grid location. Each tree in  $R_l$  predicts the offset to the true point position, and casts a vote into an accumulator array  $C_l$  at the predicted position. This is performed independently for each point. The shape model places a constraint on the results from all regressors. The quality of fit  $Q$  is given by

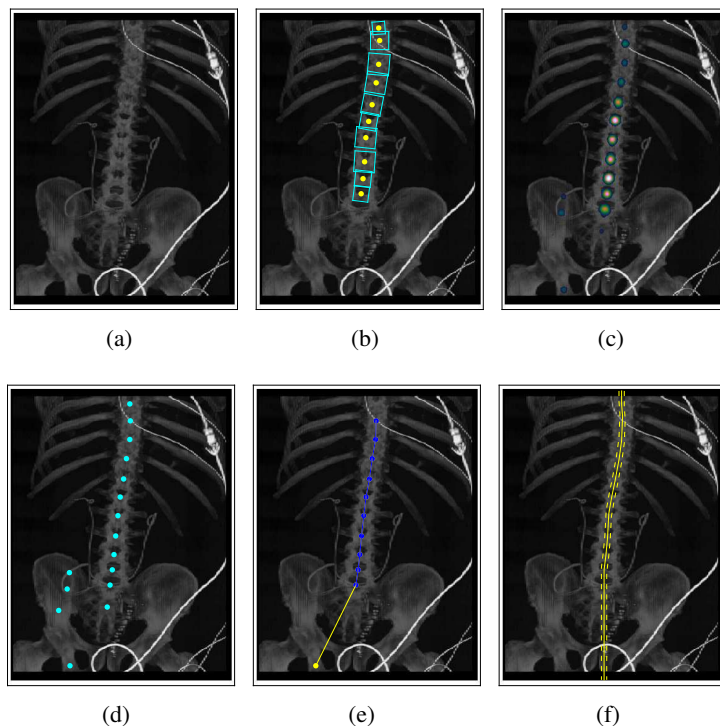
$$Q(\mathbf{p}) = \sum_{l=1}^N C_l(T_\theta(\bar{\mathbf{x}}_l + \mathbf{P}_l \mathbf{b} + \mathbf{r}_l)) \quad \text{s.t.} \quad \mathbf{b}^T \mathbf{S}_b^{-1} \mathbf{b} \leq M_t \quad \text{and} \quad |\mathbf{r}_l| < r_t \quad (2)$$

where  $\mathbf{S}_b$  is the covariance matrix of shape model parameters  $\mathbf{b}$ ,  $M_t$  is a threshold on the Mahalanobis distance, and  $r_t$  is a threshold on the residuals.  $M_t$  is chosen using the cumulative distribution function (CDF) of the  $\chi^2$  distribution so that 98% of samples from a multivariate Gaussian of the appropriate dimension would fall within it. This ensures a plausible shape by assuming a flat distribution for model parameters  $\mathbf{b}$  constrained within hyper-ellipsoidal bounds [6].  $Q$  is iteratively optimised, over parameters  $\mathbf{p} = \{\mathbf{b}, \theta, \mathbf{r}_l\}$ , as described in [11].

#### 2.5 Data Collection and Manual Annotation

The PACS (Centricity Universal Viewer, GE Healthcare, Little Chalford, Buckinghamshire, UK) at Central Manchester University Hospital NHS Trust (CMFT) was queried to produce a list of CT scans acquired during May and June 2014 and January to September 2015. The scans that a) were from non-trauma patients, b) included any part of the thoracic or lumbar spine and c) were of patients over 18 years of age, were selected. This gave a list of 868 patients' scans. The CMFT 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 vertebral fractures, 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 for the models, including the 132 fracture-rich images to ensure high fracture prevalence. The remaining images were reserved for validation purposes. The 402 image list was divided into quarters for leave-1/4-out training and testing, with the fracture-rich images distributed evenly. Each volume was up-sampled to give isotropic voxel dimensions, equal to the smallest voxel dimension from the original volume, using tri-cubic interpolation.

A coronal MIP was generated from each image volume, and manual annotation of a landmark on the neural arch of each visible vertebra was performed. 2D sagittal



**Fig. 1.** (a) An example coronal MIP of a CT volume; note the presence of confounding structures both outside (cardiac monitoring equipment) and inside (from previous abdominal surgery) the subject. (b) Manual annotations of the neural arch, with the undisplaced sample regions used in RF training. (c) Density plot of the Hough voting array from the RF search. (d) Modes of the array. (e) Result of linking and filtering; yellow links are those rejected by the filter. (f) Extrapolated piecewise-linear curve through the filtered modes (solid line) and the  $\pm 5mm$  range (dashed line) over which sagittal rasters were summed to produce the sagittal projection.

images were generated from each volume, as described Section 2.6, by summing all sagittal slice rasters within  $\pm 5mm$  of the plane defined by the coronal annotations. This thickness was chosen by manual inspection of the results, to minimise blurring of the endplates whilst ensuring that the middle of each endplate was visible. High-resolution manual annotation of 33 points on each vertebral body between T4 and L4 inclusive was then performed on the sagittal images by trained radiographers. Finally, each annotated vertebra was classified by an expert radiologist as normal, deformed but not fractured, or grade 1, 2 or 3 osteoporotic fracture, according to the Genant definitions [10].

## 2.6 Midplane Image Extraction

Osteoporotic vertebral fractures typically develop as a depression of the middle of the vertebral endplates (biconcave fracture), followed by anterior collapse of the vertebral

body (wedge fracture) and posterior collapse (crush fracture). Therefore, height reductions must be measured at the endplate midplanes to avoid underestimation of the fracture severity. If the superior-inferior axis of the subject is not aligned exactly with the CT scanner, or if any degree of scoliosis is present, then no single slice of the sagittal reformatted volume will pass through all midplanes. Therefore, an algorithm was developed to extract a 2D image along the spine midplane. First, a coronal maximum intensity projection (MIP) was produced from the volume, to show the bony structures. In particular, the point at which the laminae join to form the spinous process of the neural arch is a distinctive, U-shaped structure on each vertebra in such images (Fig. 1a). These points were manually annotated on each image (see Section 2.5).

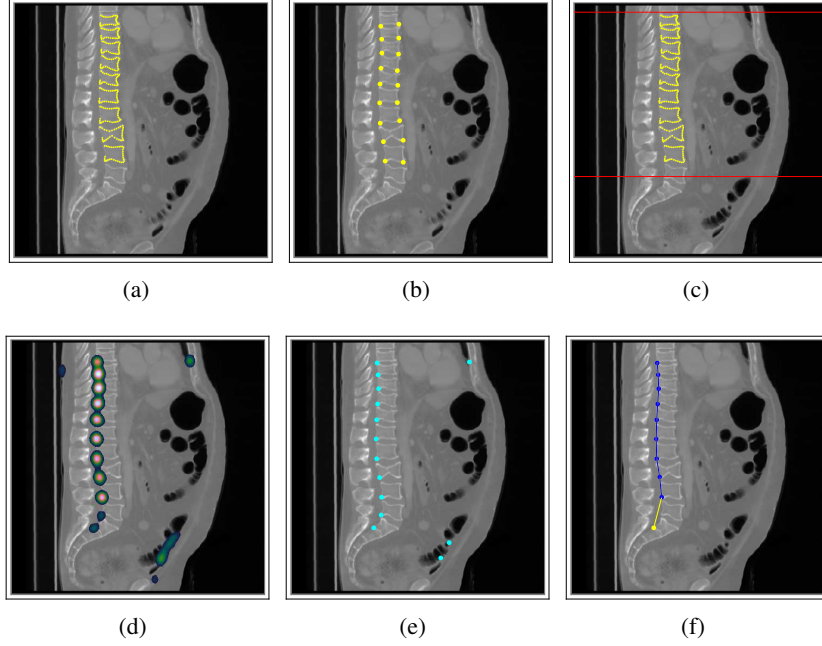
A RF regressor was then trained to localise the neural arch points as described in Section 2.1. Undisplaced sample patches were defined by using half of the average vector to the neighbouring points as the apothem of a square region of interest (ROI) (Fig. 1b). Free parameters were set to the values given in [4], and a single RFRV was trained using data from all points. Example images of each stage of the algorithm are shown in Fig. 1. Several confounding structures are visible inside and outside the body. The algorithm was robust to such features, but did produce false detections on some non-spine bony structures, such as the pelvis and mandible. Therefore, a filtering stage was implemented. Any image with fewer than four detections was removed from the analysis. The median  $L_m$  of the distances between neighbouring modes was then calculated. If the first or last mode in the list was further than  $3L_m$  from its neighbour, it was removed.

The final set of ordered, filtered modes defined a midplane through the volume, and was used to extract a 2D sagittal image. A piecewise-linear curve was defined through the modes; at the extremities, it was extrapolated vertically to the boundary of the volume (Fig. 1f). For each axial slice from the original volume, all anteroposterior raster lines (i.e. rasters of sagittal slices) that passed within  $D_r$  of this curve were averaged to give a single raster line of a sagittal image. Repeating this for all axial images gave a single, thick-slice, 2D sagittal image that showed the midplane of each vertebra, but remained in the coordinate system of the original volume, so points annotated onto it could be directly translated to projections of a different  $D_r$ . In the remainder of this paper, “manual” and “automatic” projection refer to images produced from the manual and automatic annotations on the coronal MIP images, respectively.

## 2.7 Vertebra Localisation

Next, a set of RF regressors was trained to detect the inferior corners of each vertebral body present in the sagittal projection images. Manual annotation of the vertebrae from T4 to L4 was performed as described in Section 2.5 (Fig. 2a). As in [4], undisplaced sample patches were defined as square ROIs with the two lower endplate corner points at proportional positions of (0.25, 0.75) and (0.75, 0.75) (Fig. 2b). One RF regressor was trained for each vertebral level from T5 to L3, using only images where both neighbors were present, to prevent strong responses to the boundaries of the image volume.

Fitting and extraction of a linked set of modes  $\mathbf{x}_l$ ,  $l = 1 \dots n_m$  proceeded as described in Section 2.6 (Fig. 2). The aim was to use the detected vertebral corners to initialise an RFRV-CLM that modeled a triplet of neighbouring vertebrae, so it was essential to deal with any missing detections. Therefore, several filters were applied. First, all images



**Fig. 2.** (a) An example  $\pm 5mm$  manual sagittal image projection from a CT volume, with high-resolution manual annotation. (b) Inferior corner points, used to define sampling ROIs for RF training. (c)  $\pm 5mm$  automatic sagittal image projection, with the manual annotations superimposed; the red lines show the extent of the RFRV annotations on the coronal maximum intensity projection. (d) Smoothed Hough voting array of the posterior-inferior corner point regressors. (e) Modes of the Hough voting array, detected using a nine-way maximum. (f) Result of linking and filtering; yellow points are those rejected by the filters.

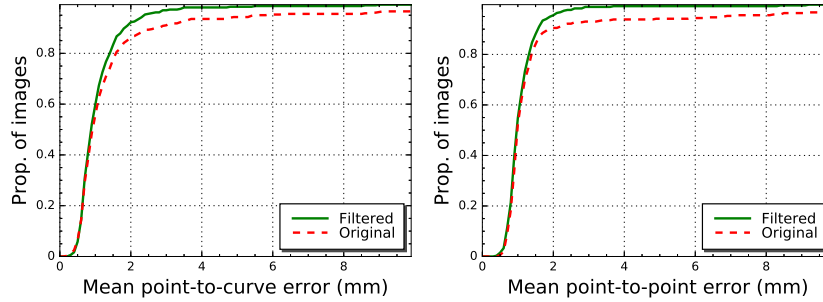
with fewer than three detections were discarded, as they could not provide a reliable initialisation. In each image, the distance between neighbouring modes was compared to the median distance between all pairs of neighbours. Where the ratio was greater than 1.5, the most probable number of missed detections  $n_l$  was

$$n_l = \left\lfloor \frac{L_l}{\mu_{1/2}(L)} + 0.5 \right\rfloor - 1 \quad \text{where} \quad L = \{L_l \mid L_l = \|\mathbf{x}_{l+1} - \mathbf{x}_l\| \forall l \in \{1, n_m - 1\}\} \quad (3)$$

where  $\mu_{1/2}(\cdot)$  represents the median, and  $n_l$  points  $(0.0, 0.0)$  were entered into the list to represent missing detections. Where this left a singlet or doublet of modes at the end of the list, these were removed. Finally, all modes outside the range of the detections from the coronal initialisation (Fig. 2c) were removed.

## 2.8 High-resolution Vertebral Segmentation

Finally, a high-resolution segmentation of the vertebrae detected by the sagittal initialisation algorithm was performed using an RFRV-CLM (see Section 2.3-2.4). The model



**Fig. 3.** Cumulative distribution functions showing the accuracy of the coronal and sagittal initialisation algorithms. (Left) The mean P2C error of points on the spine midline in each image produced by the coronal regressors, before and after filtering. (Right) The mean P2P error of posterior-inferior vertebral corner points in each image produced by the sagittal regressors, before and after filtering.

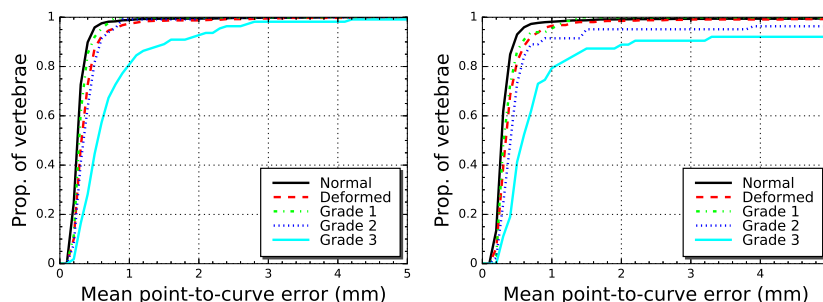
used a 2-stage, coarse-to-fine RFRV-CLM covering a triplet of vertebrae with 33 points on each. It was trained on all triplets of vertebrae from the training images. All free parameters were set to the values given in [3]. Fitting was initialised using the filtered list of posterior-inferior corner points from the sagittal regressor described in Section 2.7. All points represented as  $(0.0, 0.0)$  were considered to be undefined. The model was fitted to all triplets of neighbouring vertebrae with at least two defined points. Points from the fitted models were then concatenated to give the final segmentation (Fig. 5a). Averaging was not applied; where two models covered a single vertebra, points from the central vertebra in a triplet were used in preference to those from an extremal vertebra, and only points on vertebrae with a defined initialisation point were used.

### 3 Evaluation

Training and testing of the system on the 402 images was performed in a leave-1/4-out fashion. Errors for the coronal initialisation were measured as the mean of the minimum Euclidean distances, over each image, between the detected points and a piecewise linear curve through the manual annotations (P2C error). For the sagittal initialisation, they were measured as the mean of the Euclidean distance, over each image, between the detected points and the closest manually annotated posterior-inferior vertebral corner point (P2P error). In both cases, detections outside the axial range of the manual annotations,  $\pm$ half of the median vertebra height, were removed from the analysis to avoid penalising accurate detections of vertebra that had not been annotated manually.

Figure 3 shows CDFs of the coronal initialisation errors. Prior to filtering, 94.3% of the midplanes had a mean error of  $\leq 5\text{mm}$ , and this rose to 98.3% after filtering. The difference at  $\leq 10\text{mm}$  was small (98.3% and 99.2%). Therefore, as with the manual





**Fig. 4.** Cumulative distribution functions of P2C error for high-resolution annotations on vertebrae using manual (left) and automatic (right) coronal and sagittal initialisation, divided by vertebral classification.

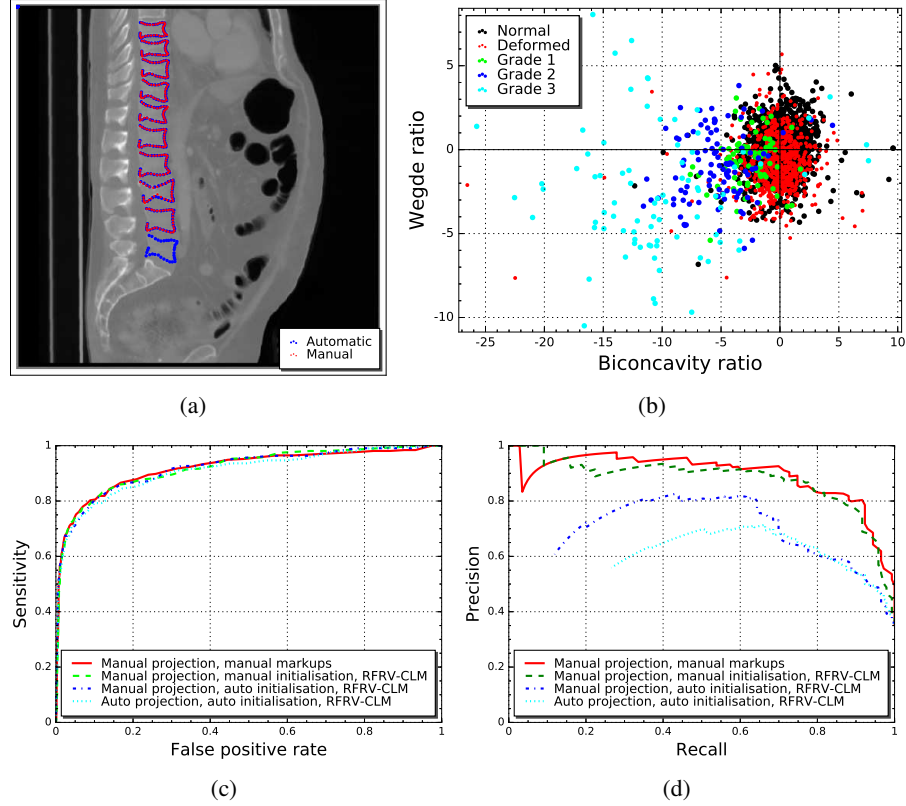
**Table 1.** Statistics of the mean point-to-curve errors on each vertebra after RFRV-CLM fitting, using manual and automatic initialisation.

Diagnosis	% of Sample	Manual Initialisation		Automatic Initialisation		
		Median (mm)	% > 2mm	% Detected	Median (mm)	% > 2mm
Normal	64.7%	0.24	0.55%	84.7%	0.27	0.83%
Deformed	25.2%	0.30	1.29%	84.9%	0.32	1.51%
Grade 1	2.84%	0.27	0.00%	75.0%	0.30	0.00%
Grade 2	3.71%	0.34	0.68%	71.3%	0.41	4.87%
Grade 3	3.59%	0.56	7.27%	56.8%	0.57	11.11%

projections, a thickness of  $D_t = \pm 5mm$  was used for automatic sagittal projection<sup>1</sup>. The filtering removed 41 images (10.2%). Figure 3 also shows CDFs of the sagittal initialisation errors. The mean errors across all points in all images were 2.14mm prior to filtering, and 1.34mm after; the medians were 0.98mm and 0.96mm, respectively. At the higher end of the CDF, 97.5% of all points in all images achieved  $\leq 5mm$  prior to filtering, rising to 99.4% after filtering. The filtering removed 27 images from the analysis i.e. 6.7%, for a total of 16.9% removed during both initialisation stages.

An example of RFRV-CLM annotation on an automatically projected image with automatic sagittal initialisation is shown in Fig. 5a. Again, any vertebrae where the centroid lay outside the axial range of the manually annotated vertebrae,  $\pm$ half of the median vertebral height, were eliminated from the analysis. The error for each vertebra was then calculated as the mean of the minimum Euclidean distances between each automatic annotation and a piecewise-linear curve through the manual annotations (P2C error). Correspondence between automatically and manually annotated vertebrae was established by calculating this error for all manual vertebrae, and taking the smallest

<sup>1</sup>The remainder of the evaluation was repeated with  $D_t = \pm 10mm$ , but this produced no improvements in the accuracy of subsequent stages, and the results are not reported here.



**Fig. 5.** (a) Example RFRV-CLM fit based on automatic coronal and sagittal initialisation. (b) Whitened biconcavity and wedge ratios for all detected vertebrae. (c) ROC curves for classification of vertebrae, based on 6-point morphometry, for manual annotations and RFRV-CLM fits with stages of automatic initialisation. (d) Precision-recall curves for classification of images.

response. Figure 4 shows CDFs of these errors for both the fully automatic system, and for RFRV-CLM fits to manual sagittal projections, initialised using manual annotations on the vertebral corners. Numerical data derived from these curves, together with the percentages of all vertebrae between T4 and L4 detected (including those in images discarded during the initialisation stages) are given in Table 1, using a mean error of  $\geq 2mm$  to indicate fit failure. The results show that automatic coronal and sagittal initialisation had little effect on the accuracy of successful RFRV-CLM fits. However, they did lead to a 4 percentage point rise in fit failures on moderate and severe fractures. Overall, 67.2% of the fractured vertebrae were detected by the fully automatic system, of which 89.1% were successfully fitted according to the  $\geq 2mm$  threshold.

The significance of the segmentation accuracy was evaluated by applying a simple classifier, based on six-point morphometry, as described in [3]. The anterior  $h_a$ , middle  $h_m$  and posterior  $h_p$  heights of each detected vertebra were calculated from the relevant

points, 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, and so this process whitened to the distribution of the normal class. A scatter plot of the whitened  $r_b$  and  $r_w$  for all detected vertebrae between T4 and L4 is shown in Fig. 5b. A simple fracture/non-fracture classification was performed by applying a threshold to  $r_c^2 + r_b^2 + r_w^2$ ; deformed vertebrae were counted correct when classified into either class. This was applied to the manual annotations, the RFRV-CLM fits on manually projected images initialised from both manual and automatic corner points, and to the fully automatic system. Receiver-operator characteristic (ROC) curves produced by varying the threshold are shown in Fig. 5d. The classifier achieved 80% sensitivity at a 10% false positive rate. More importantly, however, the fully automatic system achieved sensitivities no worse than 2 percentage points lower than classification from manual projection and annotation, at any threshold.

The classifier was also applied on a per-image basis. This simulated the use of the system in clinical practice, as described in Section 1, to generate a list of potentially fracture-containing images. A threshold on  $r_c^2 + r_b^2 + r_w^2$  was used to classify each automatically detected vertebra, and classify the images into two groups: all vertebrae normal; some vertebrae fractured. Images filtered out during initialisation were classified as fractured. Manual diagnoses were used to classify the images into normal and fractured groups, counting non-fracture deformities as normal. Figure 5d shows precision-recall curves produced by varying the threshold. Note that curves for automatic initialisation do not reach  $(0, 1)$ , due to the filtered images being classified as fractured. The fully automatic system achieved 69% recall (higher than current clinical practice; see Section 1) at 70% precision (i.e. 2/3 of reported images contained fractures).

## 4 Conclusion

The strikingly low detection rates for osteoporotic vertebral fractures on CT image volumes in clinical practice create an opportunity for an automatic system that can draw attention to images containing fractured vertebrae. The high image quality and 3D nature of CT volumes allow the automatic extraction of a single, thick, 2D sagittal slice that shows the vertebral midplanes, and does not suffer the problems of overlapping bony structures (ribs, scapulae and iliac crests) that make accurate vertebral segmentation difficult in alternative modalities such as DXA. Robust and accurate segmentation can then be achieved using a RFRV-CLM, allowing quantification of vertebral shape. This paper has shown that, even using a simple classifier, detection rates can be achieved that exceed those found in clinical practice. In future work, we intend to investigate the use of more accurate classifiers. The shape parameters of the SSM that forms part of the RFRV-CLM would provide a more complete quantification of vertebral shape than the six-point morphometry approach described above. However, osteoporosis also changes the texture of bone, since it affects horizontal trabeculae more than vertical ones. Therefore, classifiers based on both shape and texture will also be investigated.

## Acknowledgment

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. The views expressed in this publication are those of the authors and not necessarily those of 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. Adams, J.E.: Opportunistic Identification of Vertebral Fractures. *Journal of Clinical Densitometry* 19(1), 54–62 (2016)
2. Breiman, L.: Random Forests. *Machine Learning* 45, 5–32 (2001)
3. Bromiley, P.A., Adams, J.E., Cootes, T.F.: Localisation of Vertebrae on DXA Images using Constrained Local Models with Random Forest Regression Voting. In: *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*. Lecture Notes in Computational Vision and Biomechanics, vol. 20, pp. 159–172. Springer (2015)
4. Bromiley, P.A., Adams, J.E., Cootes, T.F.: Automatic Localisation of Vertebrae in DXA Images using Random Forest Regression Voting. In: T Vrtovec et al (ed.) *Computational Methods and Clinical Applications for Spine Imaging*. LNCS Image Processing, Computer Vision, Pattern Recognition, and Graphics, vol. 9402, pp. 38–51. Springer (2016)
5. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. *IEEE TPAMI* 23, 681–685 (2001)
6. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active Shape Models - Their training and application. *Comput Vis Image Und* 61(1), 38–59 (1995)
7. Cristinacce, D., Cootes, T.: Automatic Feature Localisation with Constrained Local Models. *Pattern Recogn* 41(10), 3054–3067 (2008)
8. Cummings, S.R., Melton, J.: Epidemiology and outcomes of osteoporotic fractures. *The Lancet* 359(9319), 1761–1767 (2002)
9. Delmas, P.D., van de Langerijt, L., Watts, N.B., Eastell, R., Genant, H.K., Grauer, A., Cahall, D.L.: Underdiagnosis of vertebral fractures is a worldwide problem: the IMPACT study. *J Bone Miner Res* 20(4), 557–563 (2005)
10. Genant, H.K., Wu, C.Y., Kujik, C.V., Nevitt, M.C.: Vertebral fracture assessment using a semi-quantitative technique. *J Bone Miner Res* 8(9), 1137–1148 (1993)
11. Lindner, C., Bromiley, P.A., Ionita, M., Cootes, T.F.: Robust and Accurate Shape Model Matching using Random Forest Regression-Voting. *IEEE TPAMI* 37(9), 1862–1874 (2015)
12. Operational Information for Commissioning: Diagnostic imaging dataset statistical release. Tech. rep., NHS, UK (May 2016), [www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2015/08/Provisional-Monthly-Diagnostic-Imaging-Dataset-Statistics-2016-05-19.pdf](http://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2015/08/Provisional-Monthly-Diagnostic-Imaging-Dataset-Statistics-2016-05-19.pdf)
13. Rachner, T.D., Khosla, S., Hofbauer, L.C.: Osteoporosis: now and the future. *The Lancet* 377(9773), 1276–1287 (2011)
14. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: *Proc. CVPR*, pp. 511–518. IEEE Computer Society (2001)
15. Williams, A.L., Al-Busaidi, A., Sparrow, P.J., Adams, J.E., Whitehouse, R.W.: Under-reporting of osteoporotic vertebral fractures on computed tomography. *European Journal of Radiology* 69(1), 179–183 (2009)