

SUPERVISED SHAPE ANALYSIS FOR RISK ASSESSMENT IN OSTEOPOROSIS

Marleen de Bruijne

University of Copenhagen, Denmark
Nordic Bioscience A/S, Herlev, Denmark
BIGR, Erasmus MC Rotterdam, The Netherlands

Paola Pettersen

Center for Clinical and Basic Research
Ballerup, Denmark

ABSTRACT

Early diagnosis and treatment of patients at high risk of developing fragility fractures is crucial in the management of osteoporosis. In this paper we propose to estimate the risk of future vertebral fractures using a training set of longitudinal data to learn the shape characteristics of vertebrae and spines that will sustain a fracture in the near future. A discriminant classifier is trained to discriminate between subjects developing one or more vertebral fractures in the course of 5 years and subjects maintaining a healthy spine. This approach is compared to a one-class system where the classifier is trained only on the subjects staying healthy. In a case-control study with 218 subjects, all unfractured at baseline and matched for main vertebral fracture risk factors such as spine BMD and age, we were able to predict future fractures with a sensitivity of 76% and a specificity of 72%.

Index Terms— shape analysis, discriminant analysis, disease prognosis, vertebral fracture, osteoporosis

1. INTRODUCTION

Much of the burden of osteoporosis can potentially be avoided if individuals at risk of developing fragility fractures are identified and appropriate interventions (both preventive and therapeutic) are made in a timely manner [1]. Assessment of fracture risk is therefore receiving increasing attention [1, 2, 3, 4].

Previous studies indicated that irregularity in vertebral alignment or spine curvature as can be observed on lateral X-rays is associated with known risk factors for future fractures such as prevalent fractures and low bone mineral density (BMD) [5, 6], as well as with future fractures independent of the main risk factors [7]. These studies based their analysis on simple measures of vertebral heights or circles drawn through vertebral body center points and aim at capturing the known characteristics of deformed spines in the hope that these measures may also be suited for early diagnosis and prediction of future deformity.

We propose to learn explicitly the potentially complex characteristics of shapes of spines that will and will not fracture from a training set of X-ray images for which the fracture status of the subject after a number of years is known.

2. SUPERVISED SHAPE ANALYSIS

A training set of longitudinal data of which it is known whether a deformity developed in the next few years is separated in classes of spines staying normal and spines developing a fracture. The spine shapes are described by feature vectors \mathbf{x} containing the coordinates of a set of landmark points that correspond between different shape instances and that are ‘interesting’ points, such as corner and boundary points of the vertebral bodies. The collection of training shapes is aligned using Procrustes analysis [8] to remove position, orientation, and possibly also size variations.

We investigate two different approaches to determine for a previously unseen spine shape a measure of deviation from the ‘normal’, not fracturing, spines:

1. A linear discriminant approach, in which both classes of spine shapes are modeled and the posterior probability of the spine belonging to the class that will sustain a fracture in the near future is used.
2. A one-class approach, in which only the spines staying unfractured are modeled and the reconstruction error if the spine is approximated by this model is used as the measure of abnormality.

The two approaches are described in more detail in the following sections.

2.1. Linear Discriminant Analysis

A classifier can be written in terms of discriminant functions d_i [9]. An object with feature vector \mathbf{x} is then assigned to the class that corresponds to the largest $d_i(\mathbf{x})$. If we assume a so-called 0-1 loss function — false positives and false negatives are equally bad — the optimal discriminant function that minimizes the risk associated with misclassifications is

$$d_i(\mathbf{x}) = p(\mathbf{x}|w_i)P(w_i).$$

That is, each object should be assigned the class that maximizes the posterior probability.

We will assume multivariate Gaussian distributions $p(\mathbf{x}|w_i)$ for each of the classes w_i and, to reduce the complexity of

the model and therewith the risk of overfitting to the data in the case of a relatively small dataset, we assume that both distributions have equal covariance. The resulting optimal classifier is the linear discriminant classifier:

$$d_i(\mathbf{x}) = \ln P(w_i) - \frac{1}{2}[(\mathbf{x} - \mu_i)^T \Sigma^{-1}(\mathbf{x} - \mu_i)]$$

where Σ is the pooled covariance matrix, i.e. the average covariance matrix weighted by the class prior probabilities, and μ_i are the class means.

In the experiments in this paper we use a regularized linear discriminant classifier [10] and replace Σ by Σ' :

$$\Sigma' = (1 - \alpha)\Sigma + \alpha\sigma^2 I, \quad (1)$$

where I is the identity matrix, σ^2 is the mean variance, and α a regularization parameter with $0 \leq \alpha \leq 1$. Varying α makes the classifier vary between the non-regularized linear discriminant classifier and the nearest mean classifier weighted by the class prior probabilities. What is a suitable value for α depends on the number of features and training samples and on the distribution of the data; a good value can be selected for instance using cross-fold validation on the training set.

2.2. One-class model

Similarly to the two-class case, the shapes for the one-class model are assumed to be normally distributed. If the model is trained on ‘normal’ shapes only, the Mahalanobis distance to the mean shape can be seen as a measure of abnormality.

To reduce the effect of noise in landmark placement, we follow the procedure of linear point distribution models (PDM) [11] and apply a principal component analysis (PCA) to the aligned shape vectors to model the shape probability distribution in a subspace of reduced dimensionality. To this end, the mean shape $\bar{\mathbf{x}}$, the covariance matrix Σ , and the eigensystem of Σ are computed. The eigenvectors ϕ_i of Σ provide the so-called *modes of shape variation* that describe a joint displacement of all landmarks. The eigenvectors corresponding to the largest eigenvalues λ_i account for the largest variation; a small number of modes usually captures most of the variation. Each shape \mathbf{x} in the set can then be approximated by a linear combination of the mean shape and these modes of variation:

$$\mathbf{x} = \bar{\mathbf{x}} + \Phi_t \mathbf{b} + \mathbf{r}$$

where Φ_t consists of the eigenvectors ϕ corresponding to the t largest eigenvalues, $\Phi_t = (\phi_1 | \phi_2 | \dots | \phi_t)$, \mathbf{b} is a vector of model parameters that specify the contribution of each of the modes, and \mathbf{r} is a vector of residual shape variation outside of the model subspace.

As a measure of abnormality we now use the approximation error $|\mathbf{r}|$ when the observed shape is approximated by its projection on the PCA subspace derived from the training set

of normal shapes. To ensure that the model describes only plausible, normal shapes, the shape parameters of the projected shape are constrained according to

$$b_i \leq k\sqrt{\lambda_i}. \quad (2)$$

3. EXPERIMENTS

3.1. Data

A case-control study was performed using X-rays from 218 post-menopausal women selected from a cohort of Danish women that was followed for assessment of osteoporosis and atherosclerosis in the Prospective Epidemiological Risk Factors (PERF) study [12]. Out of these 218 women, 109 maintained skeletal integrity whereas the other 109 developed one or more vertebral fractures in the course of approximately five years. None of the women received treatment for osteoporosis and none experienced an osteoporotic fracture before the baseline visit, including non-vertebral fractures. Cases and controls were matched for age, follow-up time, BMD of the lumbar spine, and weight.

Lateral X-rays of the lumbar and thoracic spine were obtained at baseline and at follow-up. The X-rays were digitized and analyzed by experienced radiologists. All vertebrae that were visible in the images, i.e. L5 to L1, T12, or T11 in the X-rays of the lumbar region, and from L1 or T12 to T4 in the thoracic region, were annotated with at least one vertebra overlapping in each pair of lumbar and thoracic images. The annotations consisted of six points placed on the corners and in the middle of the vertebra end plates, defining the anterior, middle and posterior heights. The lumbar and thoracic parts of the spine were combined by rigidly matching the landmarks of the overlapping vertebra(e) and averaging the doubly annotated landmarks. All landmark coordinates from the six points on L5 to T4 were used as features in the classification. A few examples of observed spine shapes are given in Fig. 1.a.

Fractures in the follow-up images were identified and graded according to the Genant et al. method of semi-quantitative visual assessment [13] in severity mild, moderate, or severe. Fractures are indicated as mild if one of the three heights is between 20% and 25% smaller than the maximum of the heights, moderate if the difference is between 25% and 40% and severe if the difference is larger than 40%. A total of 156 fractures was found in 109 spines at follow up. Of these, 106 were mild fractures, 39 moderate, and 11 severe. If these semi-quantitative scores are added up into an overall spinal deformity index (SDI [4], sum of individual vertebra scores with normal=0, mild fracture=1, moderate fracture=2, severe fracture=3), 76 spines had SDI=1 (one mild fracture), 25 had an SDI of 2 (one moderate or two mild fractures), 6 had SDI \geq 3.

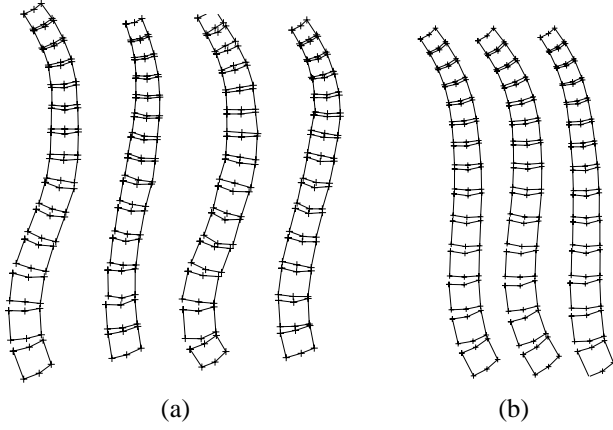


Fig. 1. (a) Two randomly selected spine shapes from the group maintaining skeletal integrity (left) and two from the group developing a fracture within the next five years (right) (b) Shape variation across the classification boundary: mean spine shape (left), spine likely to fracture (middle) and spine likely to stay intact (right).

3.2. Experimental set-up

A set of leave-one-out experiments was performed in which the classifiers for each shape were trained on the remaining 217 shapes. For the one-class classification, only the group maintaining skeletal health was used for training. Since the size of the spine could carry important information on fracture risk, the shapes are aligned using Procrustes alignment with translation and rotation, without scaling. Discriminant classifications are performed using the Matlab pattern recognition toolbox *prtools* [14]. The regularized linear discriminant classifier of Eq. (1) is used with α set to 0.1. For the one-class model, the number of modes selected was $t = 10$ and shape parameters were constrained according to Eq. (2) with $k = 3$. The selected values for parameters α , k , and t maximize the area under the ROC curve (AROC).

3.3. Results

Fig. 1.b. illustrates the discriminating direction found by the classifier. A spine shape (in this case the mean shape) was deformed in the most discriminating direction, across the decision boundary. The discriminating shape variation is a complex combination of various changes; prominent features seem to be an overall accentuation of the spinal curve and slight enlargement of the intervertebral spaces in the lumbar area.

The difference between the two groups at baseline is significant with $p = 6.0 \times 10^{-5}$ in a Wilcoxon rank sum test. Future fracturing of the spine is predicted with an accuracy (percentage correct classification) of 0.67 and area under the ROC curve (AROC) of 0.66. At a sensitivity of 76% frac-

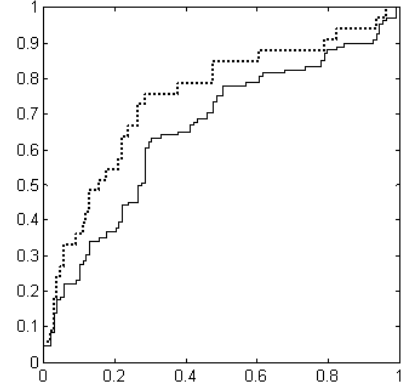


Fig. 2. ROC curves for fracture prediction using discriminant analysis, for the complete dataset (solid line) and for SDI ≥ 2 versus SDI=0 (dotted).

tures were predicted with a specificity of 72%. If the same classifier, trained on the same data, is applied to predicting more severe deformity (SDI ≥ 2) based on the baseline shape, accuracy is 0.71, AROC 0.71, $p = 2.0 \times 10^{-6}$.

The one-class model, trained on the normal subjects only, performs slightly worse with accuracy=0.56, AROC=0.63, $p = 5.9 \times 10^{-4}$.

4. DISCUSSION

As expected, the one-class model, which measures only deviation from normal spines without incorporating the examples of spines at risk is less accurate in its prediction of future fractures than the two-class, linear discriminant approach. However, it may in some cases be difficult to construct a reasonably sized training set of spines that are known to fracture, and if this is the case the one-class model could be a viable alternative.

The proposed methods relies on six landmark points placed on each of the vertebrae, for which we in this paper used manual annotations by radiologists. Although this is common practice in current quantitative morphometry studies, manual annotation is of course time consuming and might hamper large-scale use of these methods. Several authors have previously proposed methods for automatic and semi-automatic spine segmentation from X-ray or dual X-ray absorptiometry (DXA) images which might be used instead [15, 16, 17].

One shortcoming of the current six-point representation may be that osteophytes and more subtle vertebral shape variations can not be captured. In vertebral fracture classification, use of a full contour representation seems to give a slight improvement with respect to conventional height measurements [15, 18], and we would expect a similar improvement if a full contour annotation could be used in the supervised shape

analysis.

Finally, since vertebral fractures often develop without clear symptoms, the time point at which the subjects in this study experienced the first fracture is unknown. More experiments on data with varying follow-up time would therefore be required to find out the time scale at which prediction is possible.

5. CONCLUSIONS

We propose to use supervised learning techniques in a longitudinal setting to aid in disease prognosis, and have applied this to estimating spine fracture risk from the shape observed in lateral X-rays. The results in this paper suggest that both a discriminant analysis and a one-class approach can detect variations in spine shape that contribute to the risk of future vertebral fractures, independent of important risk factors such as age, spine BMD, and prevalent fractures. Supervised spine shape analysis may form a valuable addition to current vertebral risk assessment strategies.

6. REFERENCES

- [1] Department of Health and Human Services, Rockville: US Department of Health and Human Services, Office of the Surgeon General, 2004, *Bone health and osteoporosis: a report of the Surgeon-General*, 2004.
- [2] J.A. Kanis, F. Borgstrom, C. De Laet, H. Johansson, O. Johnell, B. Jonsson, A. Oden, N. Zethraeus, B. Pfleger, and N. Khaltaev, "Assessment of fracture risk.," *Osteoporos Int*, vol. 16, no. 6, pp. 581–589, 2005.
- [3] M.H. Huang, E. Barrett-Connor, G.A. Greendale, and D.M. Kado, "Hyperkyphotic posture and risk of future osteoporotic fractures: The Rancho Bernardo study," *J Bone Miner Res*, vol. 3, pp. 419–423, 2006.
- [4] G.G. Crans, H.K. Genant, and J.H. Krege, "Prognostic utility of a semiquantitative spinal deformity index," *Bone*, vol. 37, no. 2, pp. 175–179, 2005.
- [5] R.M.D. Zebaze, R.M.Z. Djoumessi, G. Maalouf, N. Maalouf, and E. Seeman, "Loss of regularity in the curvature of the thoracolumbar spine: a measure of structural failure," *J Bone Miner Res*, vol. 19, no. 7, pp. 1099–1104, 2004.
- [6] B. Cortet, E. Roches, R. Logier, E. Houvenagel, G. Gaydier-Souquires, F. Puisieux, and B. Delcambre, "Evaluation of spinal curvatures after a recent osteoporotic vertebral fracture," *Joint Bone Spine*, vol. 69, no. 2, pp. 201–208, 2002.
- [7] P.C. Pettersen, M. de Bruijne, J. Chen, Q. He, C. Christiansen, and L.B. Tankó, "A computer-based measure of irregularity in vertebral alignment is a BMD-independent predictor of fracture risk in postmenopausal women," *Osteoporos Int*, vol. 18, no. 11, 2007.
- [8] C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society B*, vol. 53, no. 2, pp. 285–339, 1991.
- [9] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley and Sons, New York, 2nd edition, 2001.
- [10] J.H. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 165–175, 1989.
- [11] T.F. Cootes and C.J. Taylor, "Combining point distribution models with shape models based on finite element analysis," *IVC*, vol. 13, no. 5, pp. 403–409, 1995.
- [12] Y.Z. Bagger, L.B. Tankó, P. Alexandersen, H.B. Hansen, G. Qin, and C. Christiansen, "The long-term predictive value of bone mineral density measurements for fracture risk is independent of the site of measurement and the age at diagnosis: results from the prospective epidemiological risk factors study," *Osteoporos Int*, vol. 17, no. 3, pp. 471–477, 2006.
- [13] H.K. Genant, C.Y. Wu, C. van Kuijk, and M.C. Nevitt, "Vertebral fracture assessment using a semiquantitative technique.," *J Bone Miner Res*, vol. 8, no. 9, pp. 1137–1148, 1993.
- [14] F. van der Heijden, R.P.W. Duin, D. de Ridder, and D.M.J. Tax, *Classification, Parameter Estimation and State Estimation: An Engineering Approach Using MATLAB*, Wiley, 2004.
- [15] P.P. Smyth, C.J. Taylor, and J.E. Adams, "Vertebral shape: Automatic measurement with active shape models," *Radiology*, vol. 211, no. 2, pp. 571–578, 1999.
- [16] G. Zamora, H. Sari-Sarrafa, and R. Long, "Hierarchical segmentation of vertebrae from X-ray images," in *Med Imaging: Image Process*, M. Sonka and M.J. Fitzpatrick, Eds. 2003, vol. 5032 of *Proc of SPIE*, pp. 631–642, SPIE Press.
- [17] M. de Bruijne and M. Nielsen, "Image segmentation by shape particle filtering," in *Proc ICPR*, J. Kittler, M. Petrou, and M. Nixon, Eds. 2004, pp. III:722–725, IEEE Computer Society Press.
- [18] M. de Bruijne, P.C. Pettersen, L.B. Tankó, and M. Nielsen, "Vertebral fracture classification," in *Med Imaging: Image Process*, J.P. Pluim and J.M. Reinhardt, Eds. 2007, vol. 6512 of *Proc of SPIE*, SPIE Press.