Early Diagnosis of Osteoporosis using Active Appearance Model and Metacarpal Radiogrammetry

Mathew Sam*, Anu Shaju Areeckal†, Sumam David S.‡
*BTech student, †PhD student, ‡Professor
Department of Electronics and Communication Engineering
National Institute of Technology Karnataka, Surathkal
Mangalore, Karnataka, India
Email: *pmathewsam@gmail.com, †anu_shaju_ec13f06@nitk.edu.in, ‡sumam@ieee.org

***Abstract*—Osteoporosis is a condition of fragile bone with an increased susceptibility to fracture. Since the gold standard method used for the diagnosis of osteoporosis, Dual X-ray Absorptiometry (DXA), is expensive and not widely available in low economies, there is a need for low cost approaches to detect bone loss in people. A new automated radiogrammetric method for early diagnosis of osteoporosis from a single hand radiograph is proposed. In this technique, the third metacarpal bone is segmented from hand X-ray images using Active Appearance Models (AAM). Points of interest acquired from the segmented bone are used to take radiogrammetric measurements, from which bone indices are calculated. Data used in this work was acquired from 138 subjects in two hospitals in India. Significant radiogrammetric features were selected using statistical analysis. The bone indices are observed to be significantly correlated with Bone Mineral Density (BMD) of the lumbar spine measured using DXA. Different classification models were trained using the significant features. The results obtained are promising and can be used as a cost effective diagnostic tool for early detection of osteoporosis.**

***Keywords*-Osteoporosis; Active Appearance Model; Third metacarpal; Radiogrammetry**

I. Introduction

Osteoporosis is a disease characterized by low bone mass and structural deterioration of bone tissue, leading to bone fragility and an increased susceptibility to fractures of hip, spine and wrist. It is a silent disease. The condition often remains painless and undiagnosed until a fragility fracture occurs. Osteoporosis can be prevented and treated if diagnosed early.

The most commonly used method for diagnosis of osteoporosis is the measurement of Bone Mineral Density (BMD) using Dual X-ray Absorptiometry (DXA). World Health Organisation (WHO) has defined BMD criteria for the diagnosis of osteoporosis, taking as reference the BMD of a young Caucasian woman [1]. The reference mean value is assigned a T-score of 0. For each increase in standard deviation above or below the mean, T-score increases or decreases by 1, respectively. According to the T-scores, bone loss in people are classified as normal, osteopenia and osteoporosis.

The gold standard technique, DXA, can take BMD measurement of forearm, hip, spine and even the whole body. It is accurate and highly precise and has low radiation dose. However, cost of DXA scans and availability of DXA system are the major limiting factors in low economies. Hence, there is a need for developing a low cost technique to diagnose osteoporosis.

A simple and inexpensive method for measuring bone loss is computerized radiogrammetry. It is a technique by which measurements of the bone such as cortical width, cortical thickness, cortical area etc. are taken from a radiograph and indices derived from these measurements are used to diagnose osteoporosis. Radiogrammetry was originally proposed by Barnett and Nordin in 1960 [2]. Scores were developed for the hand, femur and spine radiographs and thresholds were estimated to discriminate between osteoporotic and non-osteoporotic groups. Various other indices based on radiogrammetry were later developed such as Combined Cortical Thickness (CCT), Pediatric Bone Index (PBI), Exton Smith Index (ESI), etc [3].

Digital Xray Radiogrammetry (DXR) is a computerized radiogrammetric technique by which the cortical measurements of the second, third and fourth metacarpals are used to measure BMD [4]. Even though DXR uses X-ray images of metacarpal bones to measure BMD, it is found to be well correlated with DXA-BMD of the forearm, hip and spine. Osteoporosis is a systemic disease in which bone loss affects all sites of the skeleton. Hence, metacarpal radiogrammetry from hand radiograph is sufficient to measure the bone loss of appendicular and axial skeletal sites.

Our work explores the ability of cortical bone indices of the third metacarpal bone alone to classify healthy and low bone mass groups. The objective of this paper is to develop an automated tool for segmentation and radiogrammetric measurements of third metacarpal and to develop a classifier model for diagnosis of people with low bone mass.

This paper is organized as follows: Section 2 describes the working of active appearance models and further explains the proposed segmentation approach, feature extraction and classification algorithms used. Section 3 discusses the segmentation and statistical results along with performance evaluation of the classifiers. Section 4 concludes the work.

978-1-5386-4283-2/17 $31.00 © 2017 IEEE
DOI 10.1109/SITIS.2017.38

II. Proposed Methodology

A. Active Appearance Models

X-ray images pose various challenges against proper segmentation due to image acquisition conditions. The challenges faced during image segmentation arise from non-uniform illumination, noise, edge blurring etc. Accuracy of segmentation of X-ray images rely heavily on contrast of the images, which depend on the image acquisition conditions. Hence, intensity-dependent segmentation algorithms fail for images with poor contrast.

Deformable models such as Active Appearance Model (AAM), Active Shape Models (ASM), etc. make use of prior information about the object shape and/or appearance and hence performs better for low contrast images [5]. They are quite robust under different acquisition conditions. ASM is a parametric deformable model where a mean shape model of the global shape variation of the metacarpal bone is built from manually annotated training dataset, by first aligning the shapes by translation, scaling and rotation and then using Principal Component Analysis (PCA) to describe the variation of the aligned shapes among individuals. Finally, the mean shape is used to fit the model to a new test image. Hence ASM makes use of the prior shape information to improve the segmentation accuracy.

Active Appearance Model (AAM) combines a statistical shape model and gray level appearance of the object and hence is more robust than ASM. AAM is a deformation model built by analyzing the appearance of manually labeled examples. A set of training images are taken to generate the appearance model. Each image is landmarked along the edges of the object to be segmented at the same positions with respect to the object in each image. The model is built from this set of training images and landmark files. While structures vary in shape or texture, it is possible to learn what are the plausible variations and what are not. A new image can be interpreted by finding the best plausible match of the model to the image data.

In this work, an appearance model is developed using a training set of hand X-ray images with landmark points marked manually along the outer and inner boundaries of the third metacarpal bone. The AAM model is generated by carrying out PCA to acquire the most likely locations of these points for different orientation of the hand and contrast levels in the image. Information about the relative position of these points are also acquired in the form of line profile, where the normal distance between a line joining adjacent landmarks and a contour edge is also recorded. When adjusting the landmarks on to a test image, this distance must be minimum (or ideally tending to 0).

The fitting of the model on a test image is done in an iterative process. At first, the center of gravity of the model is aligned with the center of gravity of the test image. The location of the landmarks are iteratively changed to minimize the distance between the acquired landmarks from fitting the model and the edge of the metacarpal being segmented. Thus the AAM model is fitted onto the test image.

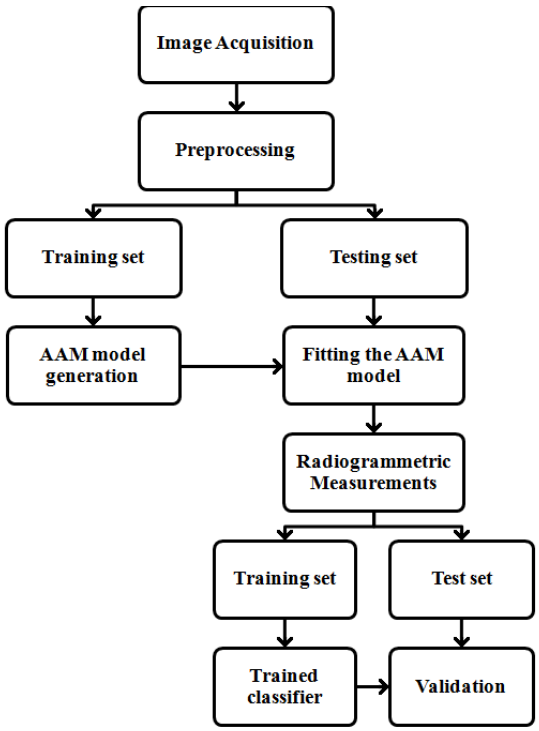

Figure 1. Flowchart of the proposed methodology

B. Segmentation Approach

The proposed methodology comprises of the following stages: Preprocessing, segmentation using AAM, radiogrammetric feature extraction and classification. Figure 1 shows the flowchart of the proposed approach.

1) Preprocessing: Orientation of the hand is an important factor that decides the accuracy of detection of third metacarpal. If the hand is highly tilted, the model may detect the wrong bone, as the metacarpal bones lie close together and are very similar in shape. The images are first rotated so that the metacarpals are aligned vertically. This is done by cropping the centre portion of the image, converting it to binary image by adapative thresholding, and estimating the orientation using image moments.

2) Segmentation of third metacarpal: In this work, AAM is implemented using Menpo package in Python [6]. To create the appearance model for the third metacarpal bone, a training set of hand X-ray images is randomly selected from the dataset. The training images are manually annotated using Landmark.io tool in Menpo. For the segmentation of third metacarpal bone, 46 landmark points were used. The annotated training images are cropped to the bounding box of the landmark points with 20% padding around the boundary. The cropped images are rescaled if the image diagonal has more than 400 pixels. The bounding box of these training images are trained using the dlib library. The AAM model is built from these training images using a holistic approach with Fast Dense Scale Invariant Feature Transform (DSIFT) features, 20

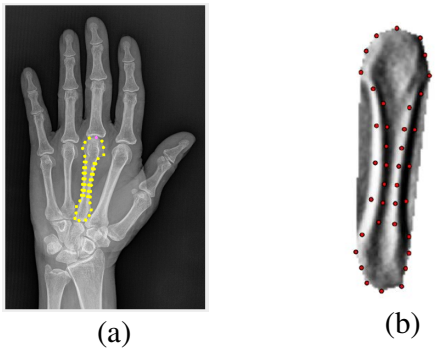

Figure 2. AAM model generation: (a) Manual annotations of the third metacarpal bone for a training image, and (b) AAM model built from a set of training images

shape components and 150 appearance components. AAM model of the third metacarpal bone is built from the training images, as shown in figure 2.

For a new test data, bounding box of the third metacarpal bone is detected using the dlib detector. The trained AAM model is then iteratively fitted onto the third metacarpal bone. This process is automated. The AAM model is fitted to the images using Lucas Kanade Optimisation. The fitting method finds the local optimum of the cost function in the bounding box. Figure 3 shows the bounding box detection and fitting of AAM model to a test image. Pixel positions of the landmark points are retrieved and used to find the cortical and medullary widths. Radiogrammetric measurements are determined at the centre of the third metacarpal bone shaft.

C. Feature Extraction

The radiogrammetric measurements acquired from the segmentation include cortical width (CW), medullary width (MW) and length (L) of the third metacarpal bone. Cortical width is measured as the diameter of outer bone at the midpoint of the metacarpal shaft. The diameter in pixels is converted to mm using pixel dimension given in the DICOM header file. Similarly, diameter of inner bone at the midpoint of metacarpal bone shaft is measured as medullary width.

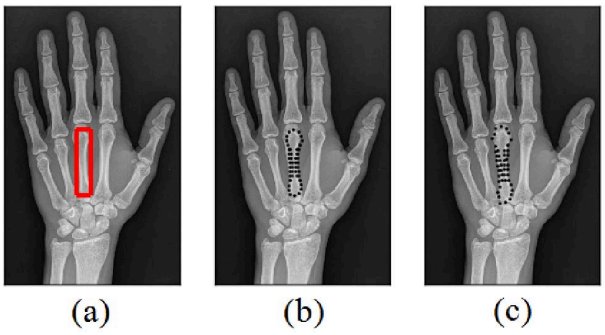

Figure 3. Testing the trained AAM model: (a) Initial bounding box detection, (b) Initial position of the AAM model and (c) Final segmented third metacarpal bone

The bone indices calculated from these radiogrammetric measurements are given in the equations below. Combined Cortical Thickness (CCT) denotes the combined value of the medial and lateral cortical bone and is calculated as the difference of the cortical and medullary widths at the midpoint of metacarpal shaft. Barnett Nordin Index (BNI) is the ratio of the combined cortical thickness to the cortical width at the midpoint. Pediatric Bone Index (PBI) is used to measure the bone loss in children.

$$CCT = CW - MW \tag{1}$$

$$BNI = \frac{CCT}{CW} \tag{2}$$

$$PBI = \frac{\pi \times CCT}{CW \times L^{0.33}} \tag{3}$$

D. Statistical Analysis

The ability of the extracted metacarpal radiogrammetric features to differentiate between healthy and low bone mass groups is analyzed using independent sample t-test. t-test measures the variation between two groups by analyzing the difference between the two group means. Correlation of the radiogrammetric features with DXA-BMD of lumbar spine are calculated using Pearson's correlation analysis. The significant features are determined and used to train different classification models.

E. Classification

Classification techniques are used to group images into healthy and low bone mass groups. Four classification models are developed in this work, namely K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Decision tree and Artificial Neural Network (ANN).

1) K-Nearest Neighbour: K-Nearest Neighbor (KNN) is a simple non-parametric supervised learning method by which the object class is determined by a majority vote of its neighbors. The nearest neighbours are measured using a distance metric. The object is assigned to the class which is the most common among its k nearest neighbors. The optimal value of k can be determined by cross validation of the dataset.

2) Support Vector Machines: Support Vector Machine (SVM) is a non-probabilistic binary linear classifier [7]. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. SVM builds a hyper plane in a high dimensional space for classification. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. Performance and accuracy depends on the hyper plane selection and kernel parameters. It can handle a large input data very efficiently.

3) Decision Trees: Decision tree is a non-parametric supervised learning method. It is used to create a classification model that can predict the target class by learning decision rules inferred from the input features. Decision tree learning uses a decision tree as a predictive model

which maps the features of the object (represented in the branches) to conclusions about the object's class (represented in the leaves) [8]. In decision tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Once the decision tree has been trained, a test image is classified by applying test condition starting from the root node and following the appropriate branch based on the decision rule. Thus it traces a path in the decision tree till the path terminates at a leaf node, whose label is assigned to the test image. The main advantage of decision trees is that it performs well even if some of the underlying assumptions of the true model is violated.

4) Artificial Neural Network: Artificial Neural Network (ANN) is a type of artificial intelligence that uses a data-driven self-adaptive technique to recognize patterns. The performance and accuracy depends on the network structure and number of inputs. ANN consists of three layers: an input layer, one or more hidden layers and an output layer. Extracted features are fed to the classifier through the input layer, which is communicated to the hidden layer, where the actual processing is done. Each node in the hidden layer is a weighted sum of input variables such that regression of output of nodes has better accuracy than regression on inputs. The processed signals are then sent to an output layer where the data is classified to one of the classes.

F. Performance metrics

The classification results may be grouped as four outcomes: True Positive, TP, False Positive, FP, True Negative, TN and False Negative, FN. True Positive denotes the case when a diseased subject is correctly classified as diseased. False Positive is the case when a healthy subject is misclassified as diseased. True Negative is the case when healthy people are correctly classified as healthy and False Negative denotes the diseased people wrongly classified as healthy. Various performance metrics such as sensitivity, specificity, accuracy, etc. may be derived from the four outcomes, as given in the equations below.

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{FP + TN} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{6}$$

III. Results and Discussion

A. Materials Used

The dataset used for this work is obtained from two hospitals in India. Postero-anterior view hand X-ray images of 138 men and women above the age of 30 years were acquired along with DXA scans of lumbar spine. The demographic and anthropometric details were also recorded. The T-scores obtained from DXA-BMD of lumbar spine were used to categorize the subjects into two groups: healthy and low bone mass. A T-score of greater than or equal to -1 is classified as healthy and less than -1 is classified as low bone mass.

B. Segmentation results

In this work, a training set of 39 hand X-ray images are used to create the AAM model. A total of 46 landmark points are used to delineate the inner and outer walls of the third metacarpal bone in the training images. The manually annotated landmark points are saved in pickle file in Python and the AAM model for the third metacarpal is generated.

For the remaining test data, the generated AAM model is placed onto the image and iteratively fitted into the third metacarpal bone. Menpo library in Python is used for the model generation and fitting. Out of total 138 hand X-ray images, the proposed segmentation approach was able to accurately segment the third metacarpal bone in 112 images. Hence, an accuracy of 81.2% has been obtained in correctly segmenting the third metacarpal bone.

Radiogrammetric measurements are then taken from the midpoint of the third metacarpal bone shaft and bone indices are calculated. The radiogrammetric features measured from the fitted AAM model were compared with ground truth measurements of 14 images obtained from experts. The absolute error and percentage error for each case is shown in table I. For cortical width measurement, mean absolute error of 0.43 mm (3.08 pixels) is obtained, resulting in 4.95% percent error. A mean absolute error of 0.26 mm (1.87 pixels) is obtained with medullary width, showing a percent error of 7.88%.

C. Statistical results

Images with good segmentation results were used for the statistical analysis and classification. Images of 52 healthy and 59 osteoporotic people were obtained after segmentation. Statistical tests and classification were carried out using Matlab R2015a. Table II shows the results of the statistical analysis. Independent sample t-test is used to determine the ability of a feature to discriminate between classes. The bone indices CCT, BNI and PBI shows a high discrimination capability with a significance value $p<0.001$.

The correlation between the extracted radiogrammetric features and DXA-BMD of lumbar spine is analyzed using Pearson correlation coefficient. This gives two values, namely the correlation coefficient r and the p-value to indicate the significance of the correlation. It is observed that the bone indices are significantly correlated with DXA-BMD of femoral neck with $p<0.0001$. This shows that metacarpal cortical radiogrammetry helps to detect bone loss in other weight-bearing skeletal sites.

The ability of features to distinguish between classes using two features at a time was further explored using scatter plots, as shown in figure 4. From the statistical analysis, it is observed that CCT, BNI and PBI can efficiently discriminate between the healthy people and low bone mass groups. Hence, these bone indices have been selected as the input features for classification.

Table I
COMPARISON OF RADIOGRAMMETRIC MEASUREMENTS WITH GROUND TRUTH

Images	Measured CW (mm)	Actual CW (mm)	Absolute error (mm)	Percent error (%)	Measured MW (mm)	Actual MW (mm)	Absolute error (mm)	Percent error (%)
1	6.86	6.61	0.25	3.79	3.14	2.89	0.24	8.42
2	8.94	10.27	1.33	12.94	4.97	4.92	0.05	1.00
3	8.33	8.75	0.42	4.76	4.18	4.19	0.00	0.08
4	8.00	7.94	0.06	0.77	3.09	2.63	0.46	17.55
5	8.69	9.14	0.45	4.96	4.58	4.42	0.16	3.58
6	7.97	7.89	0.08	1.03	4.21	3.18	1.04	32.59
7	9.74	10.45	0.71	6.78	5.68	5.85	0.17	2.87
8	7.99	8.48	0.50	5.86	4.68	4.20	0.48	11.41
9	8.16	8.44	0.28	3.32	2.96	2.92	0.04	1.22
10	8.56	8.89	0.34	3.77	3.92	3.98	0.06	1.45
11	7.34	6.73	0.61	9.02	3.51	3.17	0.34	10.71
12	7.10	7.55	0.45	5.94	2.79	2.62	0.17	6.47
13	8.82	8.48	0.33	3.94	5.16	4.88	0.27	5.63
14	7.53	7.72	0.18	2.36	2.38	2.22	0.16	7.30

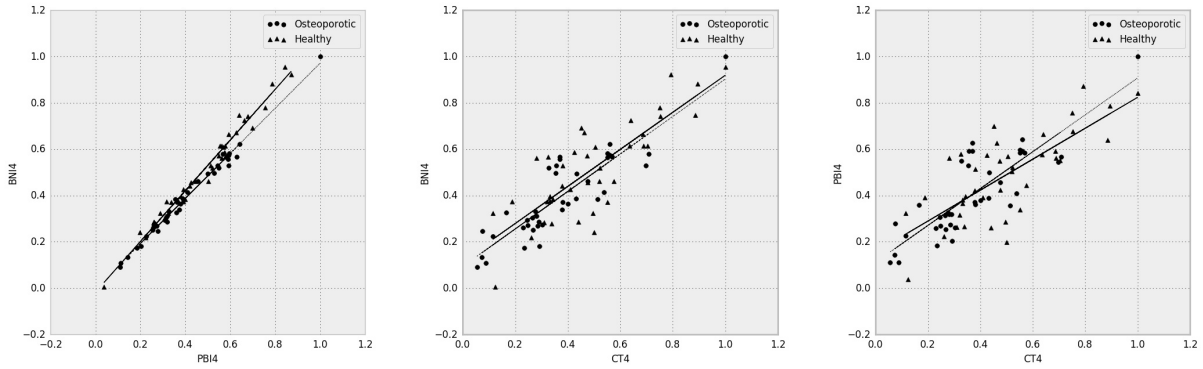

Figure 4. Scatter plots of bone indices

Table II
RESULTS OF STATISTICAL ANALYSIS

Radiogrammetric measures	Healthy Mean (SD)	Osteoporotic Mean (SD)	Correlation with BMD
CCT (mm)	4.41 (0.79)	3.79 (0.61)†	0.47†
BNI	0.56 (0.11)	0.48 (0.08)*	0.40†
PBI	0.44 (0.09)	0.39 (0.07)*	0.37†

*p<0.001, † p<0.0001, SD- standard deviation

Table III
PERFORMANCE OF TRAINED CLASSIFIERS

Classifier	Sensitivity (%)	Specificity (%)	Accuracy (%)
KNN	76.9	70	73.9
SVM	53.9	70	60.9
Decision tree	61.5	70	65.2
ANN	68.8	81.8	74.1

D. Performance of classification algorithms

The hand X-ray images are split into training set and test set. The training set consisted of features acquired from 86 images and the test set had features acquired from 27 images. Healthy case is taken as the negative class and low bone mass case is taken as the positive class. The selected significant features are used to train four classifier models: KNN, SVM, decision tree and ANN. For KNN classifier, the number of neighbours considered is three and Euclidean distance function is used. SVM is modeled with a Gaussian kernel of scale 0.43. A decision tree of 20 number of splits with Ginis Diversity Index as the split criteria is used. A ten-fold cross validation was employed for training these classifier models. ANN is modeled with one hidden layer with a single hidden node. The performance of the trained classifiers on the test set are listed in table III.

KNN and ANN classifiers gave good accuracy of above 70% for the test set. ANN gives the best accuracy of 74.1% with a sensitivity of 68.8% and specificity of 81.8%. This shows that a simple radiogrammetric measurement of the third metacarpal bone of hand is able to discriminate between healthy and low bone mass groups.

In this work, cortical bone features alone were considered for the detection of bone loss due to osteoporosis. However, osteoporosis is also caused by other determinants such as deterioration of the micro-architecture of the trabecular bone. Hence, a combination of the features determined from cortical radiogrammetry and texture analysis of trabecular bone would help to largely improve the accuracy of the diagnostic tool. This is a part of our ongoing research.

IV. Conclusion

An automated classification model to diagnose osteoporosis was proposed using Active Appearance Model (AAM) and radiogrammetry of the third metacarpal bone. AAM model of the third metacarpal bone was generated using landmark information from hand X-ray images. The AAM model was able to determine the third metacarpal bone in 81.2% of the images. Comparison of radiogrammetric measurements with ground truth showed a mean absolute error of 0.43 mm (4.95%) and 0.26 mm (7.88%) for cortical width and medullary width, respectively. Radiogrammetric measurements were taken and bone indices were calculated. CCT, BNI and PBI were found to be highly significant in differentiating healthy people and people with low bone mass. The bone indices were significantly correlated with DXA-BMD of lumbar spine, when analyzed using Pearson correlation. These indices were used as features for training different classifier models. ANN gave the best results as compared to KNN, SVM and decision trees, with a test accuracy of 74.1%. This shows that metacarpal radiogrammetry proves to be a useful diagnostic tool for low bone mass detection and cortical measurements of the third metacarpal bone alone is capable of detecting bone loss in people.

Acknowledgment

We would like to thank TEQIP-II Office, National Institute of Technology Karnataka, Surathkal, India, for the funding provided for collection of data. We would like to thank the Institutional Ethics Committee, Kasturba Medical College Hospital, Mangalore, Manipal University, India, for approving the study protocol. We would like to thank and acknowledge the Department of Orthopedics, District Wenlock Hospital, Mangalore, India and Tejaswini Hospital, Mangalore, India, for their help and support extended towards the data collection.

References

[1] J.A. Kanis, J.D. Adachi, C. Cooper, P. Clark, S.R. Cummings, M. Diaz-Curiel, N. Harvey, M. Hiligsmann, A. Papaioannou, D.D. Pierroz and S.L. Silverman, "Standardising the descriptive epidemiology of osteoporosis: recommendations from the Epidemiology and Quality of Life Working Group of IOF", *Osteoporosis International*, vol. 24, no. 11, pp. 2763-2764, 2013.

[2] E. Barnett and B.E.C. Nordin, "The Radiological Diagnosis of Osteoporosis: A New Approach", *Clinical Radiology*, vol. 11, no. 3, pp. 166-174, 1960.

[3] H.H. Thodberg, R.R. Van Rijn, T. Tanaka, D.D. Martin and S. Kreiborg, "A Pediatric Bone Index Derived by Automated Radiogrammetry", *Osteoporosis International*, vol. 21, no. 8, pp. 1391-1400, 2010.

[4] A. Rosholm, L. Hyldstrup, L. Baesgaard, M. Grunkin and H.H. Thodberg, "Estimation of Bone Mineral Density by Digital X-ray Radiogrammetry: Theoretical Background and Clinical Testing", *Osteoporosis International*, vol. 12, no. 11, pp. 961-969, 2001.

[5] T.F. Cootes, G.J. Edwards and C.J. Taylor, "Active appearance models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681-685, 2001.

[6] The Menpo Project: http://www.menpo.org/.

[7] C. Cortes and V. Vapnik, "Support-vector networks", *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.

[8] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, "Classification and regression trees" in Wadsworth and Brooks/Cole Advanced Books and Software, 1984.