



Towards an Automatic Estimation of Skeletal Age Using $k - NN$ Regression with a Reduced Set of Tinny Aligned Regions of Interest

José Luis Tonatiuh Banda-Escobar¹, Salvador E. Ayala-Raggi^{1(✉)},
Aldrin Barreto-Flores¹, Susana Sánchez-Urrieta¹,
José Francisco Portillo-Robledo¹, Alinne Michelle Sánchez-Tomay¹,
and Verónica Edith Bautista-López²

¹ Facultad de Ciencias de la Electrónica,
Benemérita Universidad Autónoma de Puebla, Av. San Claudio and 18 sur,
Col. Jardines de San Manuel, C.P. 72570 Puebla, Puebla, Mexico
tonatiuhbanda@gmail.com,
{saraggi, abarreto, surrieta, portillo}@ece.buap.mx,
michelle.stomay@correo.buap.mx
² Facultad de Ciencias de la Computación,
Benemérita Universidad Autónoma de Puebla, Av. San Claudio and 18 sur,
Col. Jardines de San Manuel, C.P. 72570 Puebla, Puebla, Mexico

Abstract. Human skeletal maturity has been typically estimated from radiographic images of the non-dominant hand through a subjective analysis performed by expert radiologists. In this paper we present a semi-automatic learning approach for estimating bone age. We consider five regions of interest, shortly *ROIs*, located between metacarpal and phalanges, which are obtained by placing strategic landmarks. *ROI* images are reshaped in the form of vectors which are merged in order to generate aligned feature vectors of each hand. The method consists of two stages, training and testing, for which radiographic images of female gender were used in a range of 1 to 18 years old. The training stage focuses on structuring the feature vectors of 300 bone-age-labeled images to generate a set of prototypes for a regression classifier. The second step is to approximate the bone age of a novel testing image, by computing its respective feature vector and comparing it with the set of prototypes. The age was determined using regression through a weighted $k - NN$ classifier. By using a set of 100 testing images, we demonstrate that it is possible to obtain an error comparable with state of the art algorithms by using only five small *ROIs* within the hand image.

Keywords: Skeletal maturity recognition · Bone age estimation
 $k - NN$ regression · *ROI* alignment

1 Introduction

Bone age assessment, also known as skeletal maturity test, is a medical practice, commonly performed by radiologists, which provides important information for physicians from other areas who are looking for possible growth disorders. Typically, a radiographic image from the non-dominant hand (usually the left hand) is analyzed by the radiologist to accomplish the test. The useful range for bone age assessment is typically between 1 to 18 years because this is the most important period related to growth in children. Subsequently, after 18 years old the medical interest for estimating bone age decreases while changes in bone structure are small and less noticeable than at younger age.

The most common clinical methods for performing the bone age assessment are usually subjective because they are based on a visual comparison of the test radiographic image with a set of labeled standard images contained in a handbook [1]. In an attempt to reduce subjectivity, other methods like [2,3] are based on individually scoring different regions of different bones and then calculate a weighted sum in order to obtain the bone age. Although less subjective than the former method [1], the later [2] is time-consuming and impractical to perform on a day-to-day basis. Finally, the subjectivity inherent in the above traditional assessment methods causes the result to be different depending on the particular physician who performs it.

Inherent subjectivity present in traditional bone age assessment can be avoided by using computerized recognition approaches. Many of those approaches have been proposed [4–9]. Some of them work as expert systems and usually are based on extracting specific high level features from bones and comparing them with pre-established values defined by human experts [8]. Other approaches use again human defined high level features, but classification is carried out by machine learning methods that usually require a training stage based on a large set of examples [7,9,10].

In [9], Hsieh et al. calculate geometric features from *ROIs* defined over the Carpal bones for ages between 1 and 8 years, and propose an artificial neural network for estimating bone age. In [10], Giordano et al. automate the known clinical method from Tanner and Whitehouse [2] by applying image processing techniques to segment metaphysis, epiphysis, and diaphysis of bones and then calculate a feature vector composed by a reduced number of lengths and areas computed from those regions. Then, a classification algorithm based in hidden Markov models is used to estimate bone age.

In an attempt to develop pure machine-learning approaches, other authors like Spanpinato et al. [11] proposed not only to classify with known methods like neural networks, but to allow the machine to infer the classification features which better differentiate bone ages by using training examples. In a deep learning approach, they use a convolutional neural network to automatically learn features. Whole hand images are used and no special regions of interest are needed. Even though the accuracy of the above method is high (a MAE or Mean Absolute Error of 0.8 years), it must be mentioned that it requires a large amount of training images (1400 images taken from a data set described in [12]).

There is little work involving low-level features such as pixels. This happens because pixels in an image do not always represent the same place in the object to recognize. The same object in a second shot may have been displaced, rotated, scaled or even adopted another perspective. However, pixels can be used as classification features as long as images are properly aligned before carrying out the comparison. In [13] Ayala-Raggi et al. use the aligned appearance of the whole hand as a feature vector to be classified by a $k - NN$ regression classifier which computes bone age. An specially designed Active Appearance Model [14] for radiographic hand images, is computed to segment the test hand and align it to a *standard shape*. Then, it is compared with a data set of prototype aligned hands. Despite the method works (MAE of 1.8 years), we think the reduced data set they used is not enough to cope with the large number of features involved, the whole hand image is used to classify!

In this paper, we show that by selecting a few, and very small, regions of interest, it is possible to reach a high accuracy in bone age estimation as long as those regions are properly aligned in scale and rotation.

According to [1, 2, 15] there are specific regions in a radiographic hand image that change markedly as the age changes. These regions are: 1. the carpal bones region, 2. the regions between metacarpal and proximal phalanges, and 3. the regions between proximal, middle and distal phalanges. Different methods for automatic bone age estimation use different regions. For instance, in [16] a total of 18 *ROIs* are used, and 5 of these are the ones used by us. However, in [12], other 7 different *ROIs* are utilized.

In this paper, we wanted to answer the question of whether it was possible to calculate bone age using only the five regions between metacarpal bones and proximal phalanges, which to our subjective opinion present a more noticeable appearance change, observed between 0 and 18 years, than the other regions.

In our work, pixels are used as low level features after a proper alignment of our small *ROI's*. We propose a simple but original method to compute the size (scale), of each *ROI* based on the size of the hand in the image. Similarly, we also calculate a rotation angle in order to normalize *ROIs* both in size and angle. Normalized *ROIs* are merged to generate a feature vector.

2 System Overview

The proposed method for bone age estimation consists of two main stages: training and testing, as shown in Figs. 1 and 2. A pre-processing step is carried out in both training and testing stages as a first step before feature extraction. This step segments the hand in the picture, eliminates possible radiological markers and undesirable objects in the background, and finally adjusts the contrast of the images in order to homogenize them before entering the system.

A second step in both training and testing stages is a manually placement of the landmarks (points of interest) over strategic locations within radiographic image. The third step, also present in both stages, corresponds to segmentation and normalization in scale and angle of five *ROIs* used to generate a feature vector.

Finally, the fourth step is different for training and testing. In training, we store the feature vector as an age-labeled prototype within a *prototypes* database. In testing, we use the feature vector as a test unlabeled prototype to be classified by a $k - NN$ regression classifier based on radial-basis functions. This regression classifier estimates bone age by regression from the age-labeled training prototypes stored during the training stage.

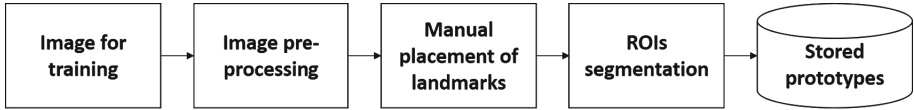


Fig. 1. Training stage.

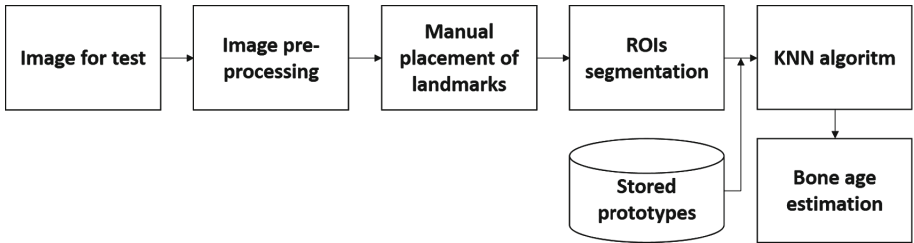


Fig. 2. Testing stage.

3 Image Pre-processing

Original radiographical images could be different each other, either by a different contrast or by intrusive objects or radiological markers present in the background surrounding the hand. In this section, we describe the two phases used for pre-processing radiological images.

3.1 Hand Segmentation

The contrast or intensity distribution in the *ROIs* used in this paper must be adjusted in such a way that gray intensity of bone regions and gray intensity of background should be both the same two intensities in all images in our system so we can make comparisons between them. Since the *ROIs* used in this paper are small regions located between metacarpal and phalangeal bones, then the amount of visible bone and background depend greatly on bone age. If the amount of visible bone is different in two images, we will obtain different gray

intensities for bone and background when we apply the same contrast adjustment criterion to both images, for example an histogram equalization. In such a condition, it is not possible to compare the images satisfactorily.

In the whole hand image, even though the amount of bone is different for each bone age, this difference is much smaller and less noticeable than that present in our small selected *ROIs*. Therefore, instead of carrying out the contrast adjustment to each *ROI* separately, we decided to adjust the contrast to the whole hand images. However, the background surrounding the hand is not part of it, so we needed to segment the hand region in order to adjust the contrast only to this hand region.

Thus, a hand segmentation step is needed before carrying out the contrast adjustment of the hand region.

We use a variation of the *floodingfill* algorithm described in [17] to segment the hand's region. Once the hand is segmented, we use a binary mask such as that illustrated in Fig. 3 in order to make the contrast adjustment to that region.



Fig. 3. Example of a binary mask used for local contrast adjustment.

3.2 Contrast Adjustment

The binary image of the hand obtained in last section is used for adjusting the contrast only within the hand's region. We propose to perform this contrast adjustment by using a simple linear mapping based on a *mean maximum* and a *mean minimum* values of the gray level intensities in the image. In order to calculate the *mean maximum* and a *mean minimum* values use compute first the mean μ and the standard deviation σ of gray levels intensities. Then, the *mean maximum* can be calculated as $MeanMax = \mu + 1.5\sigma$ and the *mean minimum* as $MeanMin = \mu - 1.5\sigma$. From these two values it is possible to do a linear mapping of all the gray values to a new range between 0 and 255.

Figure 4 illustrates this process of contrast adjustment.

4 Manual Placement of Strategic Landmarks

In order to obtain five strategic *ROIs*, we propose a manual placement of 10 points of interest that we call *landmarks*, five of them located between proximal and intermediate phalanges, and the other five between metacarpal and proximal

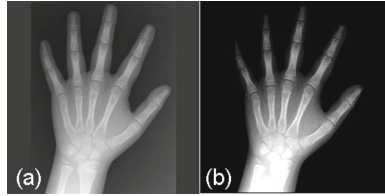


Fig. 4. Contrast adjustment of the hand's region. (a) Radiographical image with original contrast. (b) Radiographical image with corrected contrast

phalanges. The layout of the 10 landmarks is depicted in Fig. 5. In addition, we propose to locate the landmark exactly in the intermediate position between the bones where there is not some type of ossification, as is shown in Fig. 6.



Fig. 5. Location of the 10 landmarks aimed to segment the proposed *ROIs*.

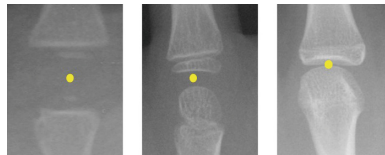


Fig. 6. Landmarks are located in the intermediate region where no ossification exists.

5 Segmenting *ROIs*

Once the process of placement of the landmarks is completed, the next step is segmenting the *ROIs*. In this paper we propose to use only five *ROIs* to determine bone age. The five landmarks located between proximal and intermediate phalanges are used just as a geometric reference aimed to be used for computing an inclination angle θ of the *ROI* with respect to the vertical, as shown in Fig. 7.

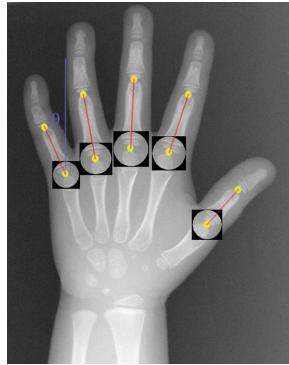


Fig. 7. Five *ROIs* used. Angle θ is computed by using the landmark location between proximal and intermediate phalanges in the same finger. The size or scale of the *ROI* is calculated using the distance between the two landmarks in the same finger multiplied by a constant factor.

The size of the *ROI* to segment is calculated based on the distance between the two landmarks in the same finger multiplied by a constant factor. We summarize the process for creating *ROIs* aligned in size and orientation in the following algorithm:

- Compute the distance between landmarks belonging to each finger.
- Multiply the distance by a parameter D . Thus, we obtain the size of the *ROI*.
- Segment the square *ROI* for each finger.
- Compute the angle θ between the vertical to the imaginary line between the two landmarks for each finger.
- Rotate each *ROI* so that the new angle θ is equal to zero.
- Resize each *ROI* to have a new size of 32×32
- Apply a circular binary mask to each *ROI* image ($diameter = 32$) in order to preserve only the same image pixels before the rotation.

Figure 8 shows the process already described. Once the five *ROIs* for a hand image are computed, the next step is to create a features vector or prototype which will be stored in a database or used as a test prototype for bone age estimation.

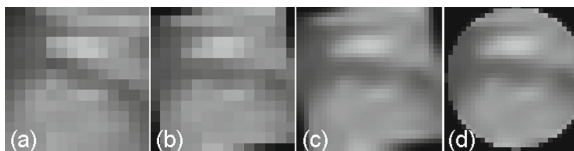


Fig. 8. The process of aligning and normalizing each *ROI*. (a) Inicial *ROI*. (b) Rotated *ROI*. (c) Resized *ROI*. (d) Masked *ROI*

6 Creating a Features Vector or Prototype

The prototype is created by reshaping or vectorizing each one of the five *ROIs* in such a way that its new size is 1×1024 (lines by columns) instead of 32×32 . The five *line* vectors are then concatenated to form only one *line* vector with size of 1×5120 . During the training stage, prototypes are stored, and each one is labeled with its corresponding actual bone age from the database. During testing, the created prototype will be analyzed by a $k - NN$ regression classifier to estimate its bone age.

7 $k - NN$ Regression Classifier

Bone age is finally estimated by a simple $k - NN$ regression classifier similar to the classifier used in [13], where ages of the nearest k neighbors are weighted by a factor which depends on the Euclidean distance d between the test prototype and each neighbor, and it is calculated as:

$$W_i = \exp\left(\frac{-d_i^2}{2\alpha^2}\right)$$

where α is the smallest distance (d_i) divided by 2. Finally, the estimated bone age is

$$age = \frac{\sum_{i=1}^K W_i BA_i}{\sum_{i=1}^K W_i}$$

where (BA_i) are the respective bone ages of the k prototypes.

8 Setup and Results

We used the public data set described in [12], which contains 1391 X-ray left-hand images of children of age up to 18 years old. These images have been evaluated for bone age by two different experts. Images in the data set are divided by gender (males and females) and by race (asian, afro-american, hispanic, and caucasian). Regarding race, in our approach, images were randomly mixed. In order to generate balanced training and testing sets, from each gender in the original dataset, we taken 300 images balanced in age and race for training, and other 100 different images balanced in age and race for testing. Therefore, a total of 800 images were used in our work.

8.1 Resizing the Original Images

Because the original images in the data set are different in size. Usually the vertical dimension (*lines*) is 256 and the horizontal dimension (*columns*) is less than 256 but not always the same. Then, we cropped the central part of images (where the hand is located) and merged two lateral bands which color was calculated from the pixels in each lateral edge of the cropped image. The final was a 256×256 image.

8.2 Estimating Bone Age

We tested our system for males and females separately using 100 test images with ages and races randomly mixed. Figure 9 shows two histograms of bone age for both test sets (males and females), showing a balance in age suitable for demonstrating the capability of our algorithm for estimating bone age independently of age and ethnicity.

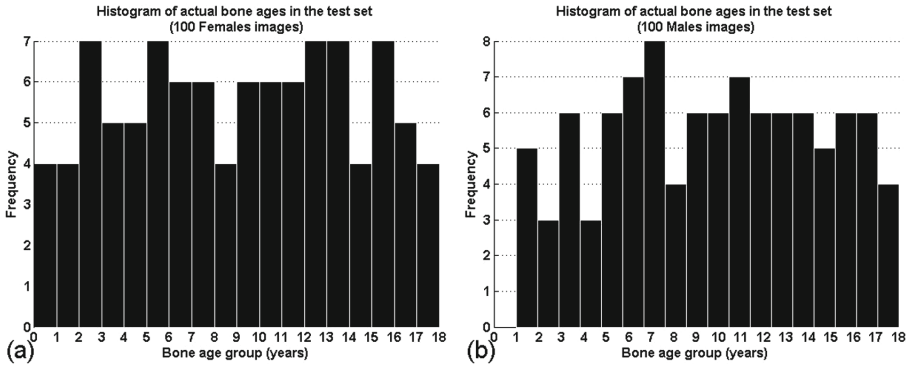


Fig. 9. Histograms of actual bone ages of the images used for testing (100 for each group). (a) Histogram for the female set. (b) Histogram for the male set.

300 images, different to those used for testing, of all ethnicities and ages were used for training. Each image was manually labeled with the 10 landmarks, and a prototype vector was created for each one. We test our system by computing the mean absolute error MAE between the a vector formed with the 100 actual bone ages and a vector formed with 100 estimated bone ages returned by the system. Similarly, we computed the square root of the mean square error calculated between the above vectors.

The test was performed varying k from $k = 2$ to $k = 26$, and we observed the best results in $k = 7$ for female images and $k = 10$ for male images as is shown in Fig. 10.

Finally, Fig. 11 illustrates graphically a comparison between actual bone ages and the estimated ones, sorted from lowest to highest. We observe in both plots a larger separation of actual and estimated age values just in the boundaries of the used age range, 0 and 18 years. The explanation could be the nature of $k - NN$ approach for *interpolating* but not for *extrapolating* ages.

Table 1 shows reported errors for different methods found in literature. In our case, by averaging MAE for females and MAE for males, we obtained a $MAE = 0.95$ years.

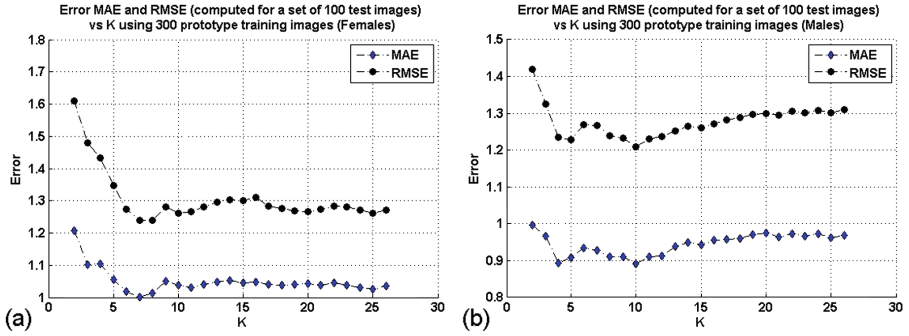


Fig. 10. Age error with 300 training images for each set of 100 test images, and varying k parameter in $k - NN$ algorithm. (a) Age error (years) as a function of k (females). (b) Age error (years) as a function of k (males)

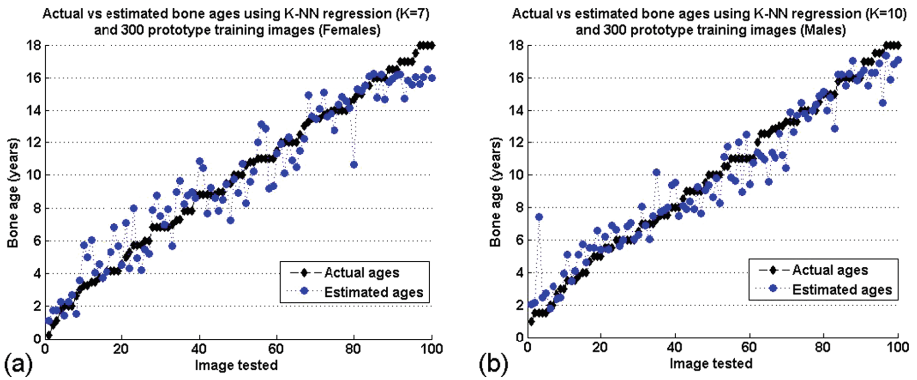


Fig. 11. Actual bone age vs estimated bone age, using 300 training prototypes and 100 test images where actual bone age was sorted from lowest to highest. k is the optimum value for each group. (a) Actual bone age vs estimated bone age (females). (b) Actual bone age vs estimated bone age (males).

Table 1. Methods found in literature

| Method | Dataset | MAE (years) |
|-------------------------|---|-------------|
| Our proposed | Digital hand atlas database system [12] | * 0.95 |
| Ayala-Raggi et al. [13] | Imagen exakta laboratories (165 images) | 1.8 |
| Spampinato et al. [11] | Digital hand atlas database system [12] | 0.8 |
| Giordano et al. [10] | Private dataset (360 images) | 0.41 |

9 Conclusions and Future Work

In this paper we proposed a simple algorithm for estimating bone age from five small *ROIs* centered around five landmarks strategically located over a radiographically image of a hand. Our experimental results demonstrate that our estimation errors are very close to those reported in state of the art approaches $MAE = 1.0$ and $RMSE = 1.24$ years for females and $MAE = 0.89$ and $RMSE = 1.21$ years for males. In contrast to other machine learning techniques, our approach needs relatively few training images to reach practically the same age error that the other methods report. We consider that our contributions are the following: 1. An original algorithm for aligning regions of interest inside radiographical images. Our method calculates the size of the *ROIs* to be segmented based on the relative positions of the placed landmarks. Then, normalizes (in angle and scale) the *ROIs* in order to be used in the creation of feature vectors. 2. An original way to create aligned vectors of features useful for successful classification. 3. A way for obtaining consistent and discriminant classification features based on applying an adequate correction of contrast to the images involved. Finally, as a future work, we are developing a completely automatic algorithm for detecting the landmarks used in this work.

References

1. Greulich, W., Pyle, S.: Radiographic Atlas of Skeletal Development of Hand and Wrist, 2nd edn. Stanford University Press, Palo Alto (1971)
2. Tanner, J., Whitehouse, R., Cameron, N., Marshall, W., Healy, M., Goldstein, H.: Maturity and Prediction of Adult Height (TW2 Method), 2nd edn. Academic Press, London (1975)
3. Molinari, L., Gasser, T., Largo, R.: TW3 bone age: RUS/CB and gender differences of percentiles for score and score increments. *Ann. Hum. Biol.* **31**(4), 421–435 (2004)
4. Adeshina, S.A., Cootes, T.F., Adams, J.E.: Evaluating different structures for predicting skeletal maturity using statistical appearance models. In: Proceedings of the MIUA (2009)
5. Aja-Fernández, S., de Luis-Garcia, R., Martin-Fernandez, M.A., Alberola-López, C.: A computational TW3 classifier for skeletal maturity assessment. A computing with words approach. *J. Biomed. Inf.* **37**, 99–107 (2004)
6. Cunha, P., Moura, D.C., López, M.A.G., Guerra, C., Pinto, D., Ramos, I.: Impact of ensemble learning in the assessment of skeletal maturity. *J. Med. Syst.* **38**, 87 (2014)
7. Liu, H., et al.: Bone age pre-estimation using partial least squares regression analysis with a priori knowledge. In: 2014 IEEE International Symposium on Medical Measurements and Applications, MeMeA 2014, Lisboa, Portugal, 11–12 June 2014, pp. 164–167 (2014)
8. Niemeijer, M., van Ginneken, B., Maas, C., Beek, F., Viergever, M.: Assessing the skeletal age from a hand radiograph: automating the tanner-whitehouse method. In: Sonka, M., Fitzpatrick, J. (eds.) *SPIE Medical Imaging*, vol. 5032, pp. 1197–1205. SPIE, Bellingham (2003)

9. Hsieh, C.W., Jong, T.L., Chou, Y.H., Tiu, C.M.: Computerized geometric features of carpal bone for bone age estimation. *Chin. Med. J.* **120**(9), 767–770 (2007)
10. Giordano, D., Kavasidis, I., Spampinato, C.: Modeling skeletal bone development with hidden markov models. *Comput. Methods Programs Biomed.* **124**, 138–147 (2016)
11. Spampinato, C., Palazzo, S., Giordano, D., Aldinucci, M., Leonardi, R.: Deep learning for automated skeletal bone age assessment in x-ray images. *Med. Image Anal.* **36**, 41–51 (2017)
12. Gertych, A., Zhang, A., Sayre, J., Pospiech-Kurkowska, S., Huang, H.: Bone age assessment of children using a digital hand atlas. *Comput. Med. Imaging Graph.* **31**, 322–331 (2007). Computer-aided Diagnosis (CAD) and Image-guided Decision Support
13. Ayala-Raggi, S., Montoya, F., Barreto-Flores, A., Sánchez-Urrieta, S., Portillo-Robledo, J., Bautista-López, V.: A supervised incremental learning technique for automatic recognition of the skeletal maturity, or can a machine learn to assess bone age without radiological training from experts? *Int. J. Pattern Recogn. Artif. Intell.* (2017)
14. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 681–685 (2001)
15. Gilsanz, V., Ratib, O.: *Hand Bone Age: A Digital Atlas Of Skeletal Maturity*. Springer, Heidelberg (2005)
16. Kashif, M., Deserno, T.M., Haak, D., Jonas, S.: Feature description with SIFT, SURF, BRIEF, BRISK, or FREAK? A general question answered for bone age assessment. *Comput. Biol. Med.* **68**, 67–75 (2016)
17. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 3rd edn. Prentice-Hall Inc., Upper Saddle River (2006)