Bone age assessment using support vector regression with smart class mapping

Daniel Haak[a1], Jing Yu[a], Hendrik Simon[a], Hauke Schramm[b], Thomas Seidl[c], Thomas M. Deserno[a]

[a]Department of Medical Informatics, RWTH Aachen University, Germany
[b]University of Applied Sciences, Kiel, Germany
[c]Data Management and Data Exploration Group, RWTH Aachen University, Germany

ABSTRACT

Bone age assessment on hand radiographs is a frequently and time consuming task to determine growth disturbances in human body. Recently, an automatic processing pipeline, combining content-based image retrieval and support vector regression (SVR), has been developed. This approach was evaluated based on 1,097 radiographs from the University of Southern California. Discretization of SVR continuous prediction to age classes has been done by (i) truncation. In this paper, we apply novel approaches in mapping of SVR continuous output values: (ii) rounding, where 0.5 is added to the values before truncation; (iii) curve, where a linear mapping curve is applied between the age classes, and (iv) age, where artificial age classes are not used at all. We evaluate these methods on the age range of 0-18 years, and 2-17 years for comparison with the commercial product BoneXpert that is using an active shape approach. Our methods reach root-mean-square (RMS) errors of 0.80, 0.76 and 0.73 years, respectively, which is slightly below the performance of the BoneXpert.

Keywords: Bone Age Assessment, Support Vector Regression, Classification, Cross Correlation, Prototypes

1. INTRODUCTION

Bone age assessment (BAA) on hand radiographs is a frequently and costly task in radiological diagnostics. Discrepancy between chronological age and skeletal maturity, reflected by the bone age, is indicating growth disturbances, other diseases, and forms an important parameter in forensic medicine, too. Usually, BAA is performed manually by trained radiologists, requiring domain knowledge and experiences [1,2]. Conventional approaches suffer from being very subjective or sophisticated in application. Hence, various approaches in automation of this process have been published [3-15].

Already in 1996, an automatic BAA method was presented [3], classifying distal and middle phalanx bones of hands by shape deformation features with classification rates of 70.5% and 73.7%, respectively. A private data set with 120 images has been used. In [4], feature extraction on the gap between metaphysis and diaphysis of fingers was analyzed and epiphyseal region of interest (eROI) introduced. Methods for handling various issues in eROI extraction have been published in a process pipeline and evaluated on 540 radiographs. The idea of ROI location was readopted using landmarks for modeling finger positions in a wire model [5]. Registration has been done by affine transformations, comparing with a prototype image of a template hand. However, the work of [4,5] was focused on eROI extraction and registration, BAA actually was not performed. In [6], decision trees have been trained with six computational features, classifying ulna and proximal phalange I bones with an accuracy of 97.6% and 95.3%, respectively. The approach of [7] applied artificial neural networks to BAA by training of several models on various bone complexes. In preprocessing, Gabor transform and Gaussian filtering have been used to reduce noise and enhance structural features. Here, training and testing has been done on 120 and 40 images, respectively. Both publications tried to adopt the manual method of [2] with automatic approaches. The work of [8] was focused on carpal shape and area descriptors. For optimal selection of resulting features, principal component analysis and discriminate analysis have been used. Various classifiers have been trained on a private dataset, including 909 radiographs, and results compared to feature statistics of the hand atlas from [1]. A fully automatic process pipeline has been presented for phalange bones [9]. Physiological features of the medial finger and morphological features of the joint between distal and middle phalanx are extracted and classified by neural

---

networks on a dataset composed of 917 images. In [10], the discrete cosine transformation (DCT) and a linear discriminant analysis (LDA) were applied to BAA. DCT and LDA coefficients are used as features and compared to prototypes from a database. On 396 radiographs, provided by Hanyang University Medical Center, an average error of 0.6 years has been achieved. For improvement of comparability of automatic BAA approaches, a new hand atlas dataset has been published by the University of Southern California (USC) [11]. The image set reflects a standard reference database for evaluation of automated BAA methods. In addition, first experiments on the data set with fuzzy classifiers on carpal bone and phalangeal ROIs have been done.

Summarizing the state of the art, it is obvious that various promising approaches have been presented for automatic BAA. However, two significant weak points in previous work raised: (i) comparability of results is restricted by evaluation on various private data sets and (ii) in usual, heuristic approaches for feature extraction have been applied, which suffer from high variability in bone development resulting from different age, gender and ethnic origin.

Improving these drawbacks, the USC hand atlas has been used in our previous work that is based on content-based image retrieval (CBIR) [12]. Here, eROI patches are automatically extracted from hand radiographs, similar and labeled regions retrieved from the Image Retrieval in Medical Applications (IRMA) framework [16,17], and k nearest neighbor (kNN) is used for classification of algebraic combinations of reference ages of the retrieved images. A mean error of 0.97 years was reported on the age range of 0-18 years. This method was improved applying a support vector machine (SVM) for classification [13,14]. Similarity between extracted eROIs and prototypes was expressed by the cross-correlation function (CCF), resulting in feature vectors with additional gender information. A mean error of 0.83 was achieved on the age range of 0-18 years.

Although this is a promising method, the natural age range is not applicable for binary classifiers like SVM. Hence, a classification scheme with 30 classes (Tab. 1) was introduced according to [18] and several SVMs were combined to predict the bone age by a one-against-one voting. Disregarding the combination scheme of binary SVM classifiers, i.e. one-against-one or one-against-all, one has to cope with problems in training since the classifiers usually require same class sizes, and with a high number of classes (here 30), the entire scheme may become instable. These weak points were improved in [15] using support vector regression (SVR) [19,20,21] instead as SVM for BAA. The support of continuous values by SVR avoided the application of voting methods gaining performance.

Class	Age range in years	Class	Age range in years
00	0.00-0.66	15	5.00-5.50
01	0.66-0.83	16	5.50-6.00
02	0.83-1.00	17	6.00-7.00
03	1.00-1.16	18	7.00-8.00
04	1.16-1.33	19	8.00-9.00
05	1.33-1.50	20	9.00-10.00
06	1.50-1.66	21	10.00-11.00
07	1.66-2.00	22	11.00-12.00
08	2.00-2.33	23	12.00-13.00
09	2.33-2.50	24	13.00-14.00
10	2.50-3.00	25	14.00-15.00
11	3.00-3.50	26	15.00-16.00
12	3.50-4.00	27	16.00-17.00
13	4.00-4.50	28	17.00-18.00
14	4.50-5.00	29	18.00-99.00

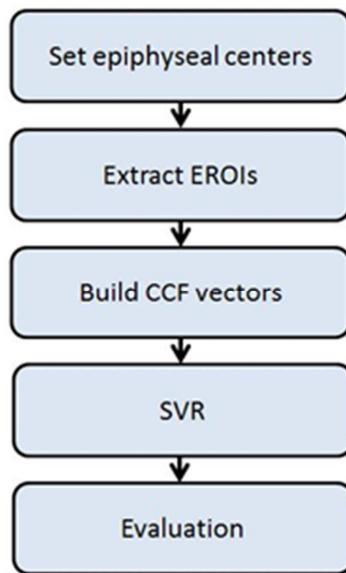**Table 1:** Age classes and corresponding age range [18].

Figure 1: Processing pipeline with SVR integration and continuous value preprocessing.

However, the continuous output of SVR was mapped to the classification scheme by simple truncation of prediction values. In this work, we aim at optimizing the prediction of bone age classes by applying novel approaches in mapping the continuous SVR output to a resulting bone age measurement.

2. MATERIAL AND METHODS

Fig. 1 illustrates the processing pipeline of our method. At first, epiphyseal centers on hand radiographs are located. The selected regions are processed, eROIs extracted and rotated into reference position. CCF values are computed and the feature vectors are built. SVR is trained and the resulting model is used for evaluation.

2.1 Evaluation data

The USC hand atlas represents the standard data available for research in automatic BAA approaches. Today, the image set contains 1,103 hand radiographs from various age classes, gender and ethnics [11]. The data is processed by semi-automatic eROI location and extraction. In Fig. 2, the epiphyseal hand regions with their corresponding numbers are shown.

In total, 14 available eROIs are used disregarding the epiphyseal centers close to the wrist. The eROIs are extracted, normalized, and rotated into vertical alignment. Such an eROI is represented by a 60 x 50 pixels. After processing of all images, 29,050 eROIs are available. To provide comparability to other work, 15 radiographs yielding highest mean error have been excluded, and the resulting data set is labeled USC-15.

2.2 Feature extraction

In the first step of feature extraction, prototype images from each class of USC data are randomly chosen and fixed as reference images. The eROI number 15 (cf. Fig. 2) of those prototype images are shown in Fig. 3 for all 30 classes. The feature vector is based on the cross correlation function (CCF) to express the similarity between eROIs extracted from unclassified hand radiographs and the prototypes. Additionally, gender information (male vs. female) is used to acknowledge their different characteristics in growth spurts [13,14,15].
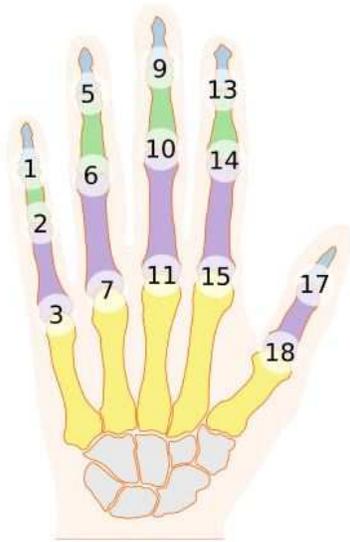
Figure 2: Epiphyseal ROIs and the corresponding region numbers.

2.2 SVM and SVR

Due the high amount of classes, the SVM needs to be extended by the one-against-one approach, resulting in several SVMs [13,14]. Each model divides two classes and the final prediction is determined by voting. In contrast, the SVR is a regression method naturally handling continuous labels and prediction values. Given a training data set $\{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \ldots, (\vec{x}_n, y_n)\}$ of size n with feature vectors $\vec{x}_i$ and label y_i, SVR tries to find a target function $f(\vec{x})$. One goal in approximation of this function is a maximal deviation of ϵ for every target value y_i.

For this, an ϵ-intensive loss function is introduced that penalizes data points with higher deviation than ϵ. The choice of ϵ is critical: if the value of ϵ is set too low, $f(\vec{x})$ approximates optimally the target values in the training data but can lead to conceivable bad results on the test data (over fitting). On the other hand, if ϵ is too high, the approximation of the target values is imprecise and again results in a poor prediction. In such cases, the v-SVR method is used that implicitly optimize ϵ by means of the parameter v [20]. Now, ϵ is traded off against complexity of the model and certain error tolerances of data points in the training set, so called slack variables.

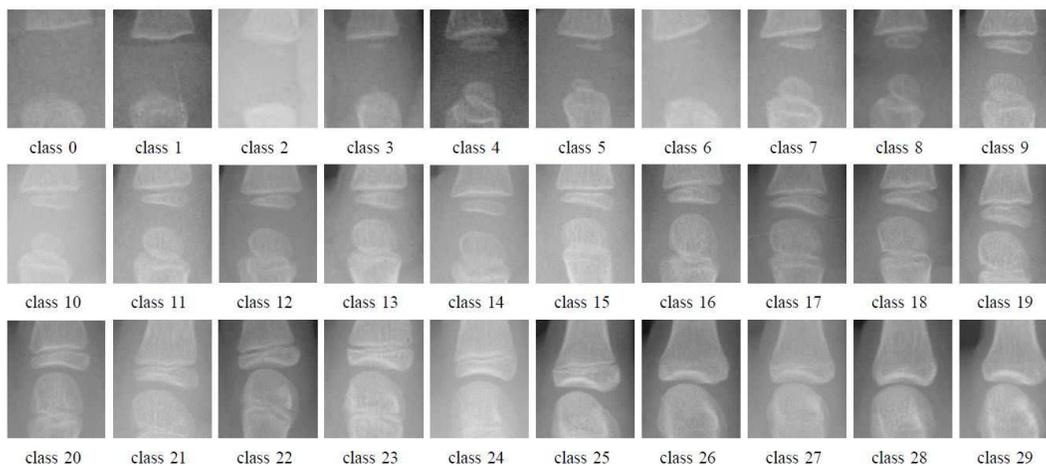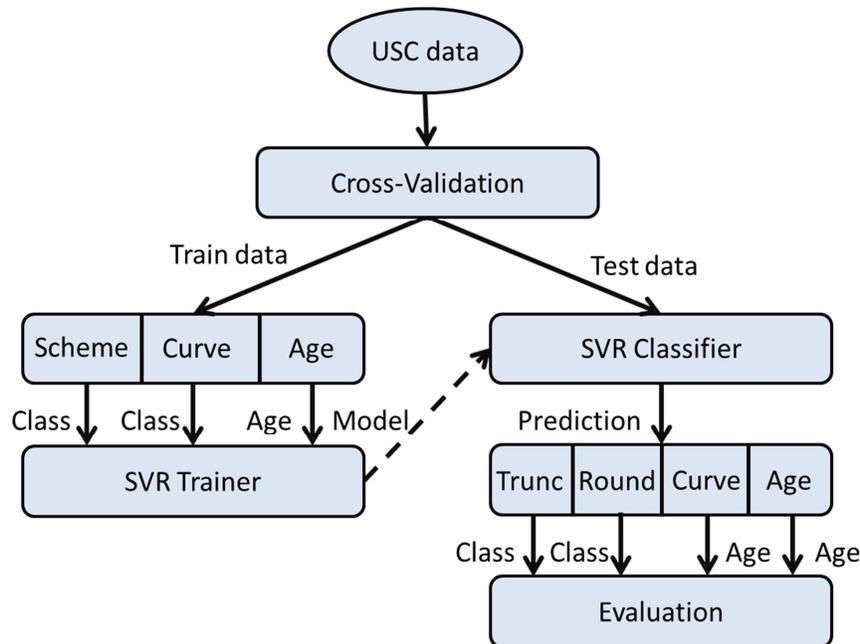**Figure 3:** Prototype images of each class [14].

Figure 4: Various mapping approaches in train and test phase.

2.3 Mapping

Due to the fact that SVR handles continuous values, various forms of input and output data transformation in train and test phase are possible. In general, training can be done on age readings or age classes. In case of training on classes the continuous class prediction of SVR has to be mapped back to an age value in test phase. Fig. 4 illustrates evaluated approaches, and their integration in train (left) and test phase (right). Those methods are denoted as SVR-Trunc, SVR-Round, SVR-Curve and SVR-Age:

1. *SVR-Trunc* was presented in [15]. Here, continuous SVR output is simply truncated in test phase to fit in the class scheme. The predicted age is determined as the center value of the corresponding class bounds.
2. *SVR-Round* is a modified approach of SVR-Trunc. Instead of truncation of the output values, the prediction is rounded in two steps. First, all values outside the classification scheme bounds are mapped to the closest class. In the second step, floating point numbers are rounded as usual with a threshold of 0.5. Age is accordingly calculated to [15].
3. *SVR-Curve* referees to the most investigative but also promising transformation of the class-trained SVR output values. This approach uses a map reflecting relationship between age values and classes in train and test phase. Fig. 5 depicts the mapping curve for the SVR output range of classes labeled 0-29 (Tab. 1). As illustrated, this method supposes linear slope between the class bounds and accordingly performs assignment.
4. *SVR-Age* interprets BAA as natural regression problem without detouring (artificial) age classes. In contrast, all mapping approaches used so far are based on the age classes. Since SVR is designed to cope with continuous values, age readings are directly used as input values in training phase. The prediction value reflects an age and can be immediately compared to radiologist readings for evaluation in test phase.

2.3 Evaluation

To produce comparable results, all experiments are done accordingly to settings in [13,14,15]. The random chosen, but for all experiments fixed, class prototypes have been carried over from [13,14]. SVR-Round, SVR-Curve and SVR-Age are evaluated using a five-fold cross-validation scheme. Optimal parameters have been determined in [15] by grid search and grid regression. In addition, experiments on each possible subset provided an optimal set of best regions. These parameters have been reclaimed in our experiments.
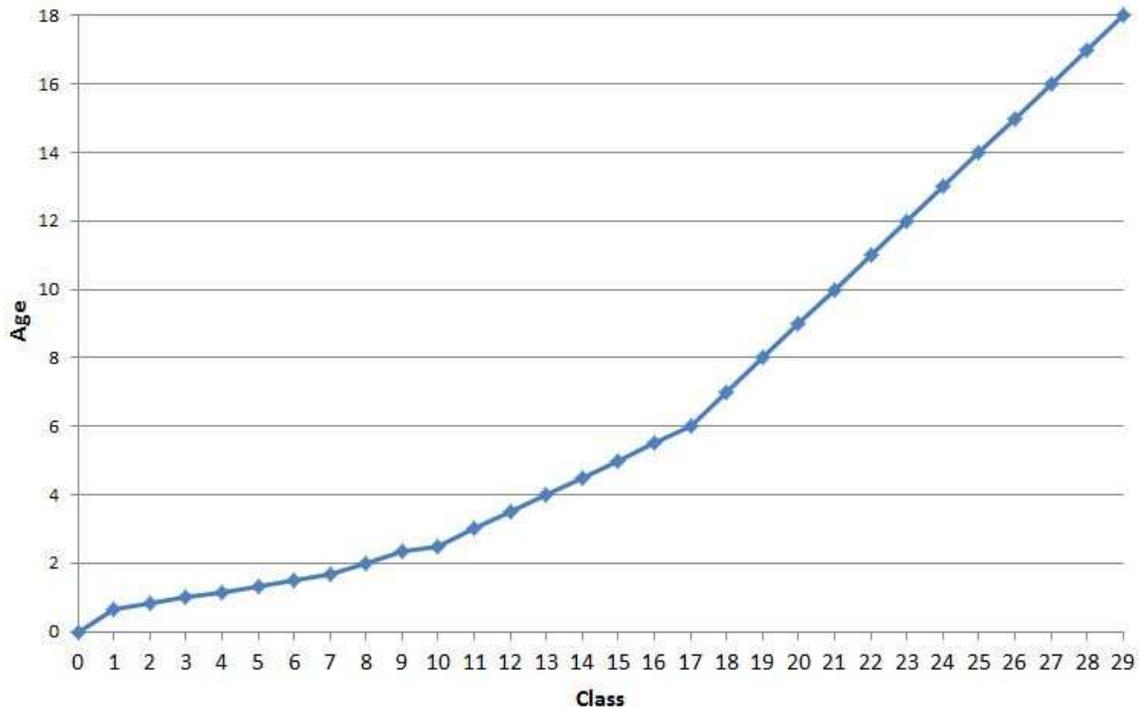
Figure 5: Mapping between age and class output in SVR-Curve.

For all experiments, the mean error was computed

$$\text{Mean} = \frac{1}{n} \sum_{i=1}^{n} |r_{\text{est}}{}^i - r_{\text{rad}}{}^i|$$

with n denoting the size of the data set, which here is the number of hand radiographs. The variables r_{est} and r_{rad} denote the estimated age and the radiologists reference reading, respectively.

For comparison, the root mean squared (RMS) error is additionally calculated in some experiments. It is defined as

$$\text{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (r_{\text{est}}{}^i - r_{\text{rad}}{}^i)^2}$$

and gives more weight to high deviations from estimated age than the mean error.

3. RESULTS

The results of our experiments and the experiments done by [15,22] can be found in Tab. 2. Comparing SVR–Trunc with SVR-Round, the mean error decreases remarkably from 0.768 to 0.720 and rather slightly from 0.692 to 0.679 on the age ranges of 0-18 years and 2-17 years, respectively. SVR-Curve further decreases the mean error to 0.695 years and 0.627 years for hands of 0-18 years and 2-17 years, respectively. Using SVR-Age without class transformation, the mean error increases again on the age range of 0-18 years to 0.729 and remains nearly static for the age range of 2-17 years.

Dataset	Age	SVR-Trunc	SVR-Round	SVR-Curve	SVR-Age	BoneXpert[22]
USC	0-18	0.768 ± 0.657	0.720 ± 0.631	0.695 ± 0.612	0.729 ± 0.595	
	2-17	0.692 ± 0.572	0.679 ± 0.560	0.627 ± 0.542	0.639 ± 0.540	
USC-15	2-17	0.799 (RMS)	0.801 (RMS)	0.737 (RMS)	0.756 (RMS)	0.61 (RMS)

For comparison with BoneXpert, the evaluation of all approaches on USC-15 shows only slightly different results between SVR-Trunc and SVR-Round. SVR-Curve reduces the RMS value drastically to 0.737, but does not reach the performance of BoneXpert. SVR-Age results in a RMS value of 0.756, which is below that of SVR-Round but still higher than SVR-Curve.

For those approaches with training on age classes and therefore comparable class prediction values, the evaluated class hits are illustrated in Table 3. The percentage of correctly classified hands rises, comparing SVR-Trunc and SVR-Round, from 35.52% to 38.87%. The class hit ratios with SVR-Curve of 36.24% shows only slightly differences to SVR-Trunc.

4. DISCUSSION

In this work, SVR has been readopted for BAA and improved by smarter mapping methods. With SVR-Round, SVR-Age and SVR-Map three new approaches have been presented, all improving performance on USC data. Evaluation shows that, regarding mean and RMS error, SVR-Curve produces the best results on both age ranges. Focusing on percentage of correctly classified hands SVR-Round performance better, resulting in higher accuracy.

Comparing with the commercial product BoneXpert, the process has been improved by application of SVR-Age and SVR-Curve, recognizable by significantly reduced RMS value, but still does not reach the performance of BoneXpert. Noticeable here is that with the application of SVR-Round, the mean error decreases but the RMS value remains nearly constant. A reason for this behavior could be the ambiguous experiment setting with removal of worst 15 hands. If there are many hands with high mean error values in the set, the RMS value decreases proportionately more significant from removing of these outliers. The high percentage of exact class hits for SVR-Round and therefore low amount of majorly misclassified hand radiographs supports this assumption. Anyway, the setting of the experiments for comparison with BoneXpert is not optimal and might be improved using new reference values that are produced with the commercial product.

Furthermore, it is conspicuous that SVR-Map performs better than SVR-Age. Differing from expectation, the application of SVR-Age without detour over age classes, and therefore without loss of information in trainings phase, is not mirrored in the results. Training on age classes seems to perform better than directly on age, even this discretization step is not necessary anymore with SVR.

In addition, the USC images reflect a standard data set composed of carefully selected images. This is very useful for the definition of prototypes but does not represent typical hand radiographs from daily routine. To evaluate the performance of the presented methods in practical work, hand radiographs with typical artifacts, e.g. noise, misplacement, and superimposition of irrelevant structures such as fracture fixations, should be processed.

Table 3: Class distance hits of all approaches in the age range of 0-18 years

Class distance	0	1	2	3	4	5	6	7	8	9	10	≥11
SVR-Trunc (%)	32.52	46.23	15.53	3.36	1.09	0.73	0.09	0.27	0.09	0.00	0.00	0.09
		94.28						5.72				
SVR-Round (%)	38.87	43.78	12.62	2.72	1.18	0.27	0.27	0.09	0.09	0.00	0.09	0.02
		95.27						4.73				
SVR-Curve (%)	36.24	44.23	13.71	3.72	1.00	0.45	0.18	0.18	0.18	0.00	0.09	0.02
		94.18						5.82				

Also, the current method of prototype selection is improvable. Alternatively, prototypes might be generated by mean values or selection of optimal prototypes. In further experiments best regions selection should be optimized by parameter optimization methods. Evaluation of all possible subsets ensures to find the optimal solution, but is heavily time-consuming.

5. CONCLUSION

In this work, we comprehensively examined the integration of SVR in the processing pipeline of BAA automation. On a USC image set eROIs are located and extracted. Then, feature vectors are built up from CCF similarity measurements, compared to prototype images. For classification, the regression model SVR is used in combination with various smart age mapping methods for pre- and post-processing of age readings. The introduction of mappings approaches for SVR continuous output improved the performance of the process pipeline. With SVR-Round, SVR-Curve and SVR-Age, three approaches have been presented decreasing mean and RMS error, and increasing class hits, respectively. Despite these improvements, these methods do not yet reach the performance of the commercial product BoneXpert that is based on the active shape model. Data-driven SVM methods, like SVR, however, are more general and applicable other image-based classification problems, for instance, the continuous tumor staging in screening mammography [23].

Anyway, investigative evaluation of hand radiographs data promises good performance for automatic BAA in computer-aided diagnostics (CAD) that in special should be analyzed with images from daily routine to evaluate suitability in practical application.

REFERENCES

[1] Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist. Am J Med Sci. 1959;238(3):393.

[2] Tanner JM, Healy MJR, Goldstein H, Cameron N. Assessment of skeletal maturity and prediction of adult height (TW3). London: WB Saunders;2001.y

[3] Al-Taani AT, Ricketts IW, Cairns AY. Classification of hand bones for bone age assessment. Proc ICECS 1996;2:1088–91.

[4] Pietka E, Gertych A, Pospiech S, Cao F, Huang HK. Computer-assisted bone age assessment: image preprocessing and epiphyseal/metaphyseal ROI extraction. IEEE Trans Med Imaging. 2001;20(8):715–29.

[5] Martin-Fernandez MA, Martin-Fernandez M, Alberola-Lopez C. Automatic bone age assessment: a registration approach. Proc SPIE. 2003;5032:1765–76.

[6] Aja-Fernandez S, Luis-Garcia RD, Martin-Fernandez MA, Alberola-Lopez C. A computational TW3 classifier for skeletal maturity assessment. a computing with words approach. J Biomed Inform. 2004;37(2):99–107.

[7] Bocchi L, Ferrara F, Nicoletti I, Valli G. An artificial neural network architecture for skeletal age assessment. Proc ICIP. 2003;1:1077–80.

[8] Hsieh CW, Jong TL, Chou YH, Tiu CM. Computerized geometric features of carpal bone for bone age estimation. Chin Med J (Engl).2007;120(9):767-70.

[9] Chang CH, Hsieh CW, Jong TL, Tiu CM. A fully automatic computerized bone age assessment procedure based on phalange ossification analysis. Proc IPPR.2003;16:463–8.

[10] Kim HJ, Kim WY. Computerized bone age assessment using dct and lda. Lect Notes Computer Sci. 2007;4418:440-8.

[11] 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. 2007;31(4-5):322.

[12] Fischer B, Welter P, Grouls C, Günther RW, Deserno TM. Bone age assessment by content-based image retrieval and case-based reasoning. Proc SPIE. 2011;7963.

[13] Harmsen M, Fischer B, Schramm H, Deserno TM. Support vector machine classification using correlation prototypes for bone age assessment. Proc BVM. 2012; 434-9.

[14] Harmsen M, Fischer B, Schramm H, Seidl T, Deserno TM. Support vector machine classification based on correlation prototypes applied to bone age assessment. IEEE Trans Inf Technol Biomed. 2012;PP(99):1.

[15] Haak D, Simon H, Yu J, Harmsen M, Deserno TM. Bone age assessment using support vector regression. Proc BVM. 2013.

[16] Lehmann TM, Gold MO, Thies C, Fischer B, Spitzer K, Keysers D, Ney H, Kohnen M, Schubert H, Wein BB. Content-based image retrieval in medical applications. Methods Inf Med.2004;43(4):354-61.

[17] Gueld M, Thies C, Fischer B, Lehmann TM. A generic concept for the implementation of medical image retrieval systems. Methods Inf Med. 2007;76(2-3):252-9.

[18] Gilsanz V, Ratib O. Hand Bone Age: A Digital Atlas of Skeletal Maturity. 2nd ed. Springer; 2011.

[19] Basak D, Pal S, Patranabis DC. Support vector regression. Neural Information Processing - Letters and Reviews. 2007;11(10):203-24.

[20] Schölkopf B, Smola AJ, Williamson RC, Bartlett PL. New support vector algorithms. Neural Comput. 2000;12(5):1207-45.

[21] Schölkopf B, Burges CJC, Smola AJ. Advances in kernel methods: support vector learning. Cambridge: MIT press; 1999.

[22] Thodberg HH, Sävendahl L. Validation and reference values of automated bone age determination for four ethnicities. Acad Radiol. 2010;17(11):1425-32.

[23] Deserno TM, Soiron M, de Oliveira JEE, de Albuquerque Araújo A. Computer-aided diagnostics of screening mammography using content-based image retrieval. Proc SPIE. 2012;8315:271-9.