FROM INDIVIDUAL HAND BONE AGE ESTIMATES TO FULLY AUTOMATED AGE ESTIMATION VIA LEARNING-BASED INFORMATION FUSION

Darko Stern[1,2], *Martin Urschler*[1,2]

[1]Institute for Computer Graphics and Vision, Graz University of Technology, BioTechMed Graz, Austria
[2]Ludwig Boltzmann Institute for Clinical Forensic Imaging, Graz, Austria

ABSTRACT

Increasingly important for both clinical and forensic medicine, radiological age estimation is performed by fusing independent bone age estimates from hand images. In this work, we show that the artificial separation into bone independent age estimates as used in established fusion techniques can be overcome. Thus, we treat aging as a global developmental process, by implicitly fusing developmental information from different bones in a dedicated regression algorithm. With 0.82 ± 0.56 years absolute deviation from chronological age on a database of 132 3D MR hand images, the results of this novel automatic algorithm are inline with radiologists performing visual examinations.

Index Terms— age estimation, hand bones, magnetic resonance (MR), regression random forest, information fusion

1. INTRODUCTION

Estimation of the unknown age of individuals is an important topic for both clinical and forensic practice. Clinical medicine is focused on estimating biological age to determine endocrinological diseases in children [1] or in paediatric orthopaedic surgery [2]. Approximated by the biological age, forensic medicine is focused on estimating unknown chronological age, e.g. the age of victims or minor asylum seekers lacking valid identification documents [3].

Age estimation (AE) from left hand images is a widespread radiological method for assessing unknown age, where ossification of individual hand bones, i.e. epiphyseal plate fusion, is predominantly examined visually. When examining such images, radiologists have to take into account the highly nonlinear relationship between the development of individual bones, that finish ossification with different timings. Thus, it is an important and non-trivial task how to combine the information about the ossification progress of distinct hand bones to derive an age estimate of a subject. The most commonly used Greulich-Pyle [4] (GP) method is easy to apply and treats aging as global developmental process, since all hand bones are simultaneously compared to the best matching reference image from the X-ray image atlas. However, it also requires the rater to visually examine and mentally perform the fusion of information from different hand bones, which makes the GP method prone to intra- and inter-reader variability. The more accurate and complex Tanner-Whitehouse [5] (TW2) staging scheme reduces observer-related subjectivity by independently estimating individual bone age according to textual and visual descriptions. The individual bone age estimates are then fused using a pre-defined nonlinear fusion function derived from characteristics of a sample population. BoneXpert [6], the most prominent automatic AE method, imitates the atlas matching procedure of TW2 with an automatic hand bone segmentation and an extraction of image features describing ossification. By calibration to TW2, BoneXpert performs fusion of individual age estimates based on the same pre-defined nonlinear function. A current trend in AE is to use non-ionizing Magnetic Resonance Imaging (MRI) data [7] instead of radiographs. In the automatic MRI-based AE method [8], a heuristic fusion strategy was presented, based on a decision tree excluding metacarpal and phalanx information from AE when averaging individual hand bones for older subjects. Empirical determination of this exclusion age is considered ad-hoc and depends on parameter tuning.

In this work, we investigate strategies for the complex task of fusing information describing ossification in 3D MRI hand bones for automated AE. As a baseline, we show the performance of fixed rule fusion like summing or multiplying [9] individual hand age estimates, exemplified by a novel regression-based bone AE algorithm. To better model the nonlinear relationship between the development of individual bones, we further propose to learn how to combine bone age distributions by training a meta-estimator of unknown age [10]. To overcome the need for splitting training data as needed for training the meta-estimator, we develop as our main contribution a fusion strategy integrated into the regression algorithm. In the proposed approach, the algorithm internally decides from which bones to learn the subjects age. Thus, aging is treated as a global developmental process. Our experiments confirm that this strategy is both accurate and efficiently using available data without the need for pre-defined nonlinear functions or heuristic fusion schemes.

978-1-4799-2349-6/16/$31.00

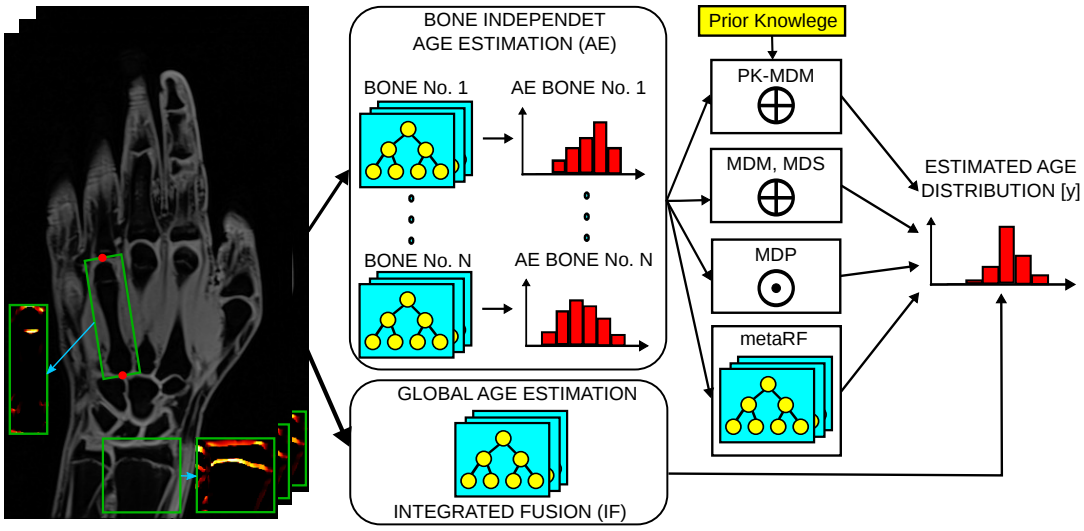

Fig. 1. Features generated from cropped MRI hand data are used in either bone independent or combined information fusion approaches to estimate the subject age.

2. METHOD

Fusion of age estimates from the subject hand bones is built upon a machine learning algorithm that automatically estimates age of individual bones. Inspired by [8], the algorithm uses Random Regression Forests (RRF) [11] to nonlinearly model age estimates of individual hand bones from training data. The method assumes hand bones have been automatically localized, aligned and cropped. To estimate a subjects age from individual predictions, different fusion strategies based on fixed rules or learning-based schemes are developed, as shown in Fig. 1.

2.1. Feature Generation

Image features are generated in a pre-processing step enhancing the appearance of the ossifying epiphyseal plate in the 3D MR images compared to surrounding anatomical structures (see Fig. 1). Due to the planar structure of the epiphyseal plate, an intermediate image representation I_i^b is generated for each cropped bone volume $b = \{1, .., N_B\}$ of hand $i = \{1, .., N\}$, based on an eigenanalysis of the Hessian matrix [12]. Thus, to enhance plate structures we compute

$$I_i^b = \frac{1}{1 + exp\left(-\frac{|\lambda_1| - \zeta_1}{\zeta_2}\right)} \cdot exp\left(-\frac{|\langle \vec{v}_1, \vec{n}_z \rangle - 1|}{\zeta_3}\right), \quad (1)$$

where the left term exploits the ratio between Hessian eigenvalues in planar structures, i.e. $|\lambda_1| \gg 0$, $|\lambda_{2,3}| \approx 0$ with $\zeta_1 = 40$ and $\zeta_2 = 5$ chosen to enhance $|\lambda_1|$ inside the epiphyseal plate. The right term penalizes the deviation of eigenvector $\vec{v}_1$, i.e. the plane normal, from the longitudinal axis of the aligned bone $\vec{n}_z$, via their dot product, scaled by $\zeta_3 = 0.25$. To provide the input for RRF, for each bone separately feature responses $f_i^b(\vec{x}; d)$ are calculated as average intensity values along the line from random coordinate $\vec{x} = (x, y, z)$ in direction of $\vec{n}_z$ with random length d.

2.2. Bone AE with Random Forest Regression

The nonlinear RRF can cope with arbitrarily large pools of randomly generated image features $f_i^b(\vec{x}; d)$ for computing trees of a forest during training the model. Thus, shrinkage and relevant feature selection for mapping ossification progress of individual bones b to chronological age are performed implicitly in each tree node by maximizing information gain IG over a random pool of features:

$$IG = |Var(S)| - \sum_{i \in \{L,R\}} \frac{|S_i|}{|S|} |Var(S_i)|. \quad (2)$$

Here $Var(\cdot)$ is the variation of age in a set of bone images, and S, S_L and S_R are the set of the bone images reaching the node and its left and right split subsets, respectively. The node splits are defined according to the binary test $\vec{f}^b(\vec{x}_j; d_j) > \tau_k$, $j \in \{1..N_F\}$ and $k \in \{1..N_T\}$. At each node, feature parameters $(\vec{x}_j; d_j)$ and a threshold τ_k that best discriminates over the ages are stored. When the maximum tree depth N_D is reached or there is no improvement in IG, the recursive splitting procedure is finished and histograms of the age distribution for the bone images that reach the leaf node are stored in the tree.

During testing, feature responses are computed for a test input I_i^b based on the stored parameters $(\vec{x}_j; d_j)$, and the image is passed to the left or right child node according to the result of testing with τ_k. The estimated bone age distribution h^b is obtained by averaging the histograms from the reached leaf nodes of all trees in the RRF.

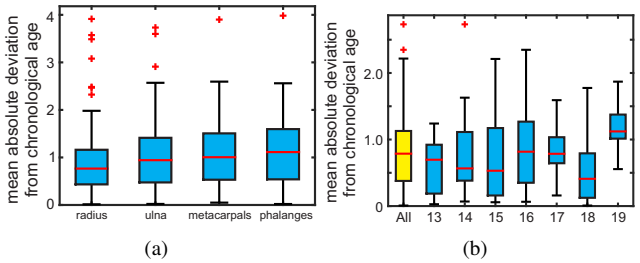

Fig. 2. (a) The results in independent bone age estimation of RRF for groups of bones and (b) separately for each age group of the best performing IF strategy.

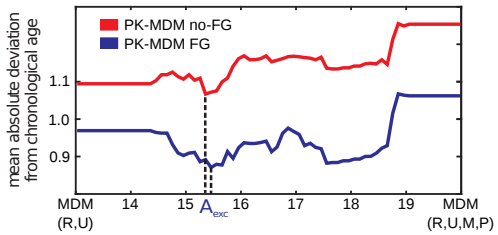

Fig. 3. Variations of mean of distribution median (MDM) fusion results according to the exclusion age of metacarpals (M) and phalanges (P) with (PK-MDM FG) and without (PK-MDM no-FG) feature generation step. Radius (R) and ulna (U) are always used in estimation.

2.3. Fusion of the Age Distribution

To estimate the subjects age from the individual bone age distributions h^b (Fig. 1), the following fixed rule fusion methods [9] are applied: mean of the distributions median (MDM), and maximum of the distribution sum (MDS) and product (MDP). Enabling comparison to recent work from [8], MDM is combined with incorporating prior knowledge by removing phalanges and metacarpals from mean calculation for subjects older than a pre-specified exclusion age (PK-MDM).

To investigate learning the nonlinear relationship between the development of individual bones [10], we also create a meta Random Forest (metaRF) based on the bone age distributions h^b, splitting the available training data into two parts according to a ratio $r = 0.5$, i.e. half of the data used to train the RRF method and the other half to train the metaRF with the same parameter settings as RRF. A feature response $f_i(\xi; b)$ of the metaRF is a value at a randomly selected bin of the histogram ξ, representing an estimated bone age distribution of a randomly selected bone b. Based on these feature responses, metaRF is trained for each node by maximizing the information gain according to Eq. 2 on the part of the training data reserved for learning the fusion of age distributions. When testing a novel input image, we first apply the RRF method to the individual hand bones to obtain age distributions separately for each bone, and use these distributions as input for the metaRF, giving final estimated age distributions h_{mRF}, see Fig. 1.

As our main contribution, we propose to integrate the fusion step into regression (integrated fusion, IF), such that the algorithm internally selects relevant bones according to the subject age, see Fig. 1. Such an approach is very well supported by the RF architecture, since it can easily cope with a large number of features needed for simultaneously learning ossification in individual bones as well as how to fuse that information to nonlinearly relate their development. Thus, a simple feature parameter extension allows that a bone, in which the bone age regression is the most prominent, is selected in each node. By making the bone number $b \in \{1, .., N_B\}$ a feature parameter $(\vec{x}; d; b)$, the features are generated from all hand bones at each node simultaneously and the parameters whose feature response produce the maximum IG (see Eq. 2) are stored in the tree. At testing, the estimated age of a subject is calculated as median of the estimated age distribution h_{IF} obtained by using all bone volumes of a subject as input for a single RRF.

3. EXPERIMENTAL SETUP AND RESULTS

Experimental setup: Our dataset of left hand T1-weighted 3D gradient echo MR images consists of scans from $N = 132$ male volunteers of known chronological age between 13 and 20 years, with in-plane resolution and slice thickness of 0.5 and 1mm, respectively. For evaluation, $N_B = 11$ bones (radius, ulna, five metacarpal and four proximal phalanges) are automatically cropped using [13] as well as empirically determined widths ensuring that bones are contained.

In RRF, $N_T = 1000$ trees of depth $N_D = 5$ were used and $N_F = 20$ features were generated per node. The number of features N_F generated in each node of the IF approach was increased by a factor of N_B compared to the individual bone RRFs, to guarantee the same probability for generating bone image features. The results of all experiments were computed in a leave-one-out cross-validation with the training time around 2 hours and testing time below one minute.

Results: Age estimation results for each bone separately are shown as box-whisker plots in Fig. 2a. Fig. 3b reveals how the pre-selected exclusion age of metacarpal and phalanges information from the MDM calculation affects the estimated age of a subject in PK-MDM. The mean absolute deviations from chronological age $\pm$ its standard deviation in cross-validation for the MDM fusion method is 0.98$\pm$0.62, for MDS is 0.97$\pm$0.61 and for MDP is 1.12$\pm$0.98. The best results 0.88$\pm$0.64 for the PK-MDM is obtained for the exclusion age of 15.45 years. When a metaRF is trained by splitting the available data into two equal parts, the result is 0.85$\pm$0.64, and for the IF strategy it is **0.82$\pm$0.56**. Detailed results of IF are shown in Fig. 2b.

4. DISCUSSION AND CONCLUSION

In this work, we investigated fusion strategies to automatically combine growth information from individual hand bones into an AE of a subject. The RRF is chosen for bone AE (Fig. 2a) since, compared to most other machine learning methods, it has the ability to cope with a large pool of image features that may be selected during training, without significantly increasing training time. Furthermore, RRF is a nonlinear approach with good generalization due to randomized feature generation when constructing trees.

When the ossification of the bone is about to be finished, epiphyseal plate location is hard to be recognized even for the experienced radiologist, and impossible when ossification is finished. As an inaccurate cropping of epiphyseal plate regions leads to unpredictable BAE results, we experimented with omitting this step and taking the whole bone images between the localized anatomical landmarks for training the BAE methods. However, when using the same fusion strategy PK-MDM on our larger dataset of 132 subjects (compared to 56 subjects in [8]), which better represents the inherent biological variation present in growth development, this resulted in a larger estimation error, i.e. 0.98 ± 0.70 (see Fig. 3, red) compared to 0.85 ± 0.58 reported in [8]. We found the main reason for this decrease in accuracy in image artifacts and intensity inhomogeneities. To remove unreliable epiphyseal plate detection step, and to make our method more robust against image artifacts and intensity inhomogeneities, we allowed the use of the whole bone for learning the age regression function and introduced a preprocessing step enhancing the epiphyseal plate. To conclude, compared to the work presented in [8], our method is not depending on accurate detection of the bounding box around the epiphyseal plate.

Our main goal of this work was to find the best fusion strategy capturing the highly nonlinear relationship between the development of individual bones. Since individual hand bones follow the stages of ossification similarly, but with different timings, fixed rule fusion (MDM, MDS, MDP) offers the lowest potential to perform information fusion from different bones for estimating unknown chronological age. By integrating prior knowledge on when metacarpals and phalanges finish ossification into the MDM fixed rule fusion strategy as proposed in [8], an increase in estimation accuracy ($p < 0.1$) can also be noted in our implementation. Thus, our results of PK-MDM, achieved on a much larger dataset, are similar to [8], where additionally more complex epiphyseal gap region extraction is required. However, such an approach is highly dependable on tuning the exclusion age parameter that is additionally dataset specific, e.g. in [8] the exclusion age is 17 years, while for our dataset the best exclusion age is 15.5 years, see Fig. 3. A significant improvement ($p < 0.04$) for RRF fusion compared to fixed rule fusion is achieved if the nonlinear relationship between the development of individual bones is learned (metaRF), thus interpreting the different bones as an ensemble of experts for AE [10]. Unfortunately, by splitting the training data into two parts for individual BAE and training the fusion model, this approach is inefficient regarding the use of available data. To overcome this problem and to also reflect that aging of subjects is a global development process not requiring an artificial separation into individual bones, we finally proposed the IF approach, i.e. a novel regression method that implicitly combines age information from different bones. Although, there is no significant difference in results between metaRF (0.85 ± 0.64) and IF (0.82 ± 0.56) approaches, the IF approach makes best use of the given data and we expect that it will scale better to a larger data set, since it simultaneously combines most discriminative features from all bones.

Age estimation results on our database of 132 MRI subjects obtained with the novel IF approach (0.82 ± 0.56 years) are comparable with the clinically established X-ray based methods, where deviations from 0.5 up to 2.0 years are reported [14]. To interpret these results, the reader has to bear in mind that the estimated age is the biological age of a subject, which due to different development speed of individuals might vary from the chronological age used in the comparison. To overcome this ambiguity, the automatic BoneXpert [6] method is calibrated by the same nonlinear function as used in the TW2 staging scheme and the reported results of 0.72 years on 1700 X-ray images (from 2 to 17 years) are obtained as deviation from the visual GP [4] atlas matching result. Nevertheless, when using the GP matching system to estimate the ground truth biological age, its high intra- and inter-observer variability has to be considered for interpreting results. The evaluation results of the BoneXpert [6] method are also depending on the quality of the nonlinear fusion function, which was originally introduced by TW2 in an "ensemble scheme" intended to reduce observer-related subjectivity from information fusion. As the main goal of this work is to find the best fusion strategy, the nonlinear fusion function introduced by the TW2 method therefore cannot be used as the established ground truth bone age during evaluation. For example, our metaRF scheme calculates the nonlinear fusion of age estimates from individual bones and additionally the rater subjectivity from individual BAE is eliminated.

To sum up, in this work we have shown that the best fusion strategy is achieved in our automatic IF approach, if aging is treated as a global developmental process, by removing the artificial separation into individual bones as it is done in TW2 or BoneXpert.

5. ACKNOWLEDGEMENTS

This work was supported by the Austrian Science Fund (FWF): P 28078-N33 (FAME).

6. REFERENCES

[1] D D Martin, J M Wit, Z Hochberg, L Saevendahl, R R van Rijn, O Fricke, N Cameron, J Caliebe, T Hertel, D Kiepe, K Albertsson-Wikland, H H Thodberg, G Binder, and M B Ranke, "The Use of Bone Age in Clinical Practice - Part 1," *Horm. Res. Paediatr.*, vol. 76, pp. 1–9, 2011.

[2] S C Lee, J S Shim, S W Seo, K S Lim, and K R Ko, "The accuracy of current methods in determining the timing of epiphysiodesis," *Bone Jt. J*, vol. 95-B, no. 7, pp. 993–1000, 2013.

[3] A Schmeling, P M Garamendi, J L Prieto, and M I Landa, "Forensic Age Estimation in Unaccompanied Minors and Young Living Adults," in *Forensic Med. - From Old Probl. to New Challenges*, chapter 5, pp. 77–120. InTech, 2011.

[4] W W Greulich and S I Pyle, *Radiographic atlas of skeletal development of the hand and wrist*, Stanford University Press, Stanford, CA, 2nd edition, 1959.

[5] J M Tanner, R H Whitehouse, Cameron N, W A Marshall, M J R Healy, and H Goldstein, *Assessment of skeletal maturity and predicion of adult height (TW2 method)*, Academic Press, 2nd edition, 1983.

[6] H H Thodberg, S Kreiborg, A Juul, and K D Pedersen, "The BoneXpert Method for Automated Determination of Skeletal Maturity," *IEEE Trans. Med. Imaging*, vol. 28, no. 1, pp. 52–66, 2009.

[7] Y Terada, S Kono, D Tamada, T Uchiumi, K Kose, R Miyagi, E Yamabe, and H Yoshioka, "Skeletal age assessment in children using an open compact MRI system," *Magn. Reson. Med.*, vol. 69, no. 6, pp. 1697–1702, 2013.

[8] D Stern, T Ebner, H Bischof, S Grassegger, T Ehammer, and M Urschler, "Fully Automatic Bone Age Estimation from Left Hand MR Images," in *MICCAI*, 2014, vol. 8674 of *LNCS*, pp. 220–227.

[9] J Kittler, M Hatef, R P W Duin, and J Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, 1998.

[10] R Polikar, "Bootstrap-inspired techniques in computation intelligence," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 59–72, 2007.

[11] A Criminisi and J Shotton, Eds., *Decision Forests for Computer Vision and Medical Image Analysis*, Springer, 2013.

[12] A F Frangi, W J Niessen, K L Vincken, and M A Viergever, "Multiscale Vessel Enhancement Filtering," in *MICCAI*. 1998, vol. 1496 of *LNCS*, pp. 130–137, Springer.

[13] T Ebner, D Stern, R Donner, H Bischof, and M Urschler, "Towards Automatic Bone Age Estimation from MRI: Localization of 3D Anatomical Landmarks," in *MICCAI*, 2014, vol. 8674 of *LNCS*, pp. 421–428.

[14] S Ritz-Timme, C Cattaneo, M J Collins, E R Waite, H W Schuetz, H J Kaatsch, and H I Borrman, "Age estimation: The state of the art in relation to the specific demands of forensic practise," *Int. J. Legal Med.*, vol. 113, no. 3, pp. 129–136, 2000.