

## What automated age estimation of hand and wrist MRI data tells us about skeletal maturation in male adolescents

Martin Urschler, Sabine Grassegger & Darko Štern

To cite this article: Martin Urschler, Sabine Grassegger & Darko Štern (2015): What automated age estimation of hand and wrist MRI data tells us about skeletal maturation in male adolescents, *Annals of Human Biology*

To link to this article: <http://dx.doi.org/10.3109/03014460.2015.1043945>



Published online: 27 Aug 2015.



Submit your article to this journal [↗](#)



Article views: 9



View related articles [↗](#)



View Crossmark data [↗](#)

RESEARCH PAPER

## What automated age estimation of hand and wrist MRI data tells us about skeletal maturation in male adolescents

Martin Urschler<sup>1,2</sup>, Sabine Grassegger<sup>1,3</sup>, and Darko Štern<sup>2</sup>

<sup>1</sup>Ludwig Boltzmann Institute for Clinical Forensic Imaging, Graz, Austria, <sup>2</sup>Institute for Computer Graphics and Vision, Graz University of Technology, BioTechMed, Graz, Austria, and <sup>3</sup>Institute of Forensic Medicine, Medical University of Graz, Graz, Austria

### Abstract

**Background:** Age estimation of individuals is important in human biology and has various medical and forensic applications. Recent interest in MR-based methods aims to investigate alternatives for established methods involving ionising radiation. Automatic, software-based methods additionally promise improved estimation objectivity.

**Aim:** To investigate how informative automatically selected image features are regarding their ability to discriminate age, by exploring a recently proposed software-based age estimation method for MR images of the left hand and wrist.

**Subjects and methods:** One hundred and two MR datasets of left hand images are used to evaluate age estimation performance, consisting of bone and epiphyseal gap volume localisation, computation of one age regression model per bone mapping image features to age and fusion of individual bone age predictions to a final age estimate.

**Results:** Quantitative results of the software-based method show an age estimation performance with a mean absolute difference of 0.85 years (SD = 0.58 years) to chronological age, as determined by a cross-validation experiment. Qualitatively, it is demonstrated how feature selection works and which image features of skeletal maturation are automatically chosen to model the non-linear regression function.

**Conclusion:** Feasibility of automatic age estimation based on MRI data is shown and selected image features are found to be informative for describing anatomical changes during physical maturation in male adolescents.

### Keywords

Automatic software, forensic age estimation, hand and wrist, MRI

### History

Received 1 April 2015

Accepted 20 April 2015

Published online 26 August 2015

### Introduction

An important but challenging branch of human biology dedicates its research efforts towards the estimation of age in living individuals or human remains. Due to gradual anatomical changes during physical maturation and growth, age estimation is of special interest when determination of the development of children and young adults is needed (Tanner, 1978). These studies are important for clinical and legal medical applications, but other domains like forensic anthropology and sports also show considerable interest in age estimation methods. There are two categories of applications that involve the need for age estimation. The estimation of biological (developmental) age assesses an individual's state of physical maturation at the time of examination, while estimation of unknown chronological age is of interest for living or deceased individuals in case no identification documents are available. Due to considerable biological variation in the speed of human development between

individuals, the assessment of unknown chronological age, which is performed by approximating it with a biological age estimate, is prone to a significant amount of uncertainty. Thus, when estimating chronological age, it has to be considered that the biological clock of an individual ticks differently compared with the chronological clock.

Estimation of biological age is relevant for applications in paediatrics. In endocrinology, children showing short or tall stature, early or late puberty or genetically induced changes in sexual development (e.g. congenital adrenal hyperplasia) are examined regarding their biological age to evaluate the need for growth hormone treatment, to plan the dosage of hormone therapy (Martin et al., 2011a,b) and to control therapy success. Another example is orthopaedic surgery, where biological age estimation is the foundation to predicting the remaining growth of bones, thus determining the optimal time-point for corrective surgery in cases of leg-length discrepancy (Lee et al., 2013) or scoliosis (Wang et al., 2009).

Recently, a rising interest in age estimation for legal and forensic medicine applications, targeting children and young adolescents lacking valid identification documents, can be observed (Ritz-Timme et al., 2000). In criminal and civil law,

Correspondence: Martin Urschler, Ludwig Boltzmann Institute for Clinical Forensic Imaging, Universitätsplatz 4, 8010 Graz, Austria. Tel: 0043 316 380 4067. Fax: 0043 316 380 9696. E-mail: martin.urschler@cfi.lbg.ac.at

determination of age thresholds such as 14 or 21 years is relevant to establish the proper legal context in which to treat a subject under investigation. With the growing number of asylum seekers entering the developed countries of the so-called Western world, the age threshold of 18 years (i.e. legal age) has also received a lot of interest, since unaccompanied minors have the right of protection and education in the country where they apply for asylum according to the United Nations Conventions on the Rights of the Child. Therefore, a reliable, objective and accurate chronological age estimation technique is required to guarantee fair treatment of children in this psychologically stressful situation (Schmeling et al., 2011). Forensic need for chronological age estimation is also required in sports applications, where the International Olympic Committee and FIFA (Fédération Internationale de Football Association) have a particular interest in preventing age fraud in junior-level competitions. Since competitions like under-17 soccer tournaments, in which there is much commercial interest, may be significantly manipulated by older, physically more developed players, forged identification documents need to be prevented by chronological age estimation techniques (Dvorak et al., 2007). On the other hand, manipulation of age in the other direction has to be detected as well, for example in the Olympic gymnastics competition, where a lower age limit of 14 years exists for participation, since younger children benefit due to their stature and mobility. Finally, a branch of forensic anthropology (Dirkmaat, 2012) is dedicated to age estimation of the deceased with the goal of dating human remains, either for age-at-death estimation in archaeology (Dean et al., 2014) or for identification of disaster victims (Bassed, 2012). In contrast to age estimation of the living, here dental, osteological and histological methods that involve destruction of teeth or bones may be used (Dutour, 2012).

These various application areas indicate the need for a reliable, objective and accurate age estimation method for practical use (Black et al., 2010). In the following we focus on non-destructive biological age estimation, which may be used to assess physical maturation and growth development or to approximate chronological age in forensic applications. The established biological age estimation procedure involves radiological examinations of skeletal (Freitas et al., 2004; Wenzel et al., 1984) and/or dental development (Liversidge, 2012), often together with an external body examination of sexual maturation to reach a more robust estimate. External body examinations are commonly assessed using the Tanner stages (Marshall & Tanner, 1969, 1970). The combination of several types of examinations is referred to as multi-factorial age estimation (Schmeling et al., 2011), with each factor giving complementary information about growth. In the context of age assessment, the anatomical region of the hand and wrist is very important for radiological and forensic examination, as the development of numerous ossification centres and the progressive fusion of a large number of growth plates can be observed and assessed from infancy to attaining adulthood. The depicted osseous structures and growth plates of a subject of unknown age are either compared to an atlas of reference images of known age, the Greulich-Pyle (GP) standard (Greulich & Pyle, 1959), or to a staging scheme for individual bones, the

Tanner-Whitehouse (TW2) system (Tanner et al., 1983). Due to biological variation, both reference images and stages of the developmental staging scheme represent a range of possible chronological ages in subjects, as derived from a reference population sample used for constructing these systems. Building upon the visual comparison of the whole hand and wrist with a reference atlas, the GP method is fast and easy to use compared to TW2, but shows lower accuracy and larger inter-observer variability (Ritz-Timme et al., 2000). Exploiting the fact that the ageing progress is not the same for all bones of the hand, bone age estimation is improved by the TW2 method by visually comparing each individual hand bone to the X-ray image-based developmental staging scheme, and combining scores describing individual stages to a final biological age estimate. Recently, automated image analysis methods for age estimation have appeared (Mansourvar et al., 2013), most prominently the X-ray image based BoneXpert method (Thodberg et al., 2009, 2012), which mimics the atlas matching procedures of both GP and TW2. The performance of BoneXpert for automated hand bone age estimation has been demonstrated in several validation studies (Martin et al., 2009; Thodberg & Saevendahl, 2010).

A severe drawback of radiographic bone age estimation techniques is exposure to ionising radiation, which cannot be ethically justified for screening healthy children and adolescents. Thus, non-invasive magnetic resonance imaging (MRI) has gained in importance for bone age estimation (Dvorak et al., 2007; Terada et al., 2013), since the laws of many countries prohibit the application of ionising radiation without medical indication. Another benefit of MRI compared to projective 2D X-ray examinations is its volumetric nature, which may provide a foundation for more accurate and reliable bone age estimation. Most current methods proposed for bone age estimation in MR images are restricted to best-view cross-sections to imitate the estimation methods developed for X-ray images (Terada et al., 2013; Tomei et al., 2014). Two recent exceptions are the works of Hillewig et al. (2013), who perform staging-based age estimation from 3D MRI of the clavicle, taking left hand radiographs into account to improve their estimates in a Bayesian analysis step, and the work of Saint-Martin et al. (2014), who study the epiphyseal gap in the distal tibial physis for age estimation. The idea of developing automated image analysis software for age estimation is even more beneficial when using MR images due to the need for interpreting additional complex 3D information that is not available in projective radiographic images. With modern medical computer vision (Sonka & Fitzpatrick, 2000) and machine learning techniques (Bishop, 2007), analysis of this complex information becomes possible, allowing automatic learning of image features that discriminate over age from 3D MR training data, thus promising a more objective and reliable estimation. While the only reported work on automatic age estimation from MRI data (Saint-Martin et al., 2014) solely proposes a static analysis of grey level variations in the epiphyseal-metaphyseal junctions of the femur bone, we are very confident that learning the image features discriminating age from training data is a key benefit to improve the objectivity in age assessment by an automatic method.

In this work, we examine the 3D MR image information extracted by a software-based age estimation technique and verify its sufficiency to define gradual anatomical changes during physical maturation on the example of skeletal ossification. Therefore, we use our previously presented fully automatic bone age estimation method (Stern et al., 2014), based on volumetric MR images showing the hand and wrist simultaneously, which is part of our ongoing effort towards establishing a novel automated multi-factorial MR imaging-based method for biological age estimation. In this method, we locate the region of the epiphyseal gap in individual long bones automatically and extract nearby image features that discriminate the fusion stages of the epiphyseal gap in a given dataset based on known chronological age. Both localisation of the epiphyseal gap region and mapping of the age relevant image features to chronological age can be seen as a regression task, where machine learning is employed to learn the non-linear regression function from annotated training data. For age regression this has the additional benefit of performing image feature selection as part of the training procedure, thus allowing one to visualise extracted age-relevant image features and to interpret them regarding their ability to discriminate different ages. This allows us to report answers to the title of this study: What automated age estimation of hand and wrist MRI data can tell us about skeletal maturation in male adolescents.

## Materials and methods

### Subject data from MRI

For development of an automatic machine learning-based age estimation software, a dataset of left hand and wrist MR images of volunteering healthy male European-derived children from Austria was used. This data was collected at the Ludwig Boltzmann Institute for Clinical Forensic Imaging in Graz as part of a study investigating the role of MRI in forensic age estimation, which was performed in accordance with the Declaration of Helsinki and approved by the ethical committee of the Medical University of Graz. All eligible participants provided written informed consent and for under-age participants written consent of the legal guardian was obtained additionally. Exclusion criteria were history of endocrinal, metabolic, genetic or developmental disease.

MRI data of 102 subjects with known chronological age were taken from the study collective, selecting them with a nearly uniformly distributed age range between 13–20 years (mean = 16.9 years, SD = 2.02 years). All MRI examinations were performed on a 3.0T scanner (Tim Trio, Siemens AG, Germany), in prone position with outstretched fixed arm using both a head and a neck coil to cover the hand and the wrist simultaneously. The MRI protocol used for acquiring the 3D volumetric images was an isotropic T1 weighted 3D FLASH VIBE sequence with a water selective pre-pulse, a resolution of 0.9 mm and a matrix size of 256. The voxel size of the resulting MRI scans was  $0.45 \times 0.45 \times 0.9 \text{ mm}^3$  and the image size  $288 \times 512 \times 72$  voxels. Voxels were interpolated from the isotropic scanning set-up and acquisition time was slightly below 4 min.

### Automated MRI bone age estimation method

As illustrated in Figure 1, our proposed machine learning-based age estimation approach (Ebner et al., 2014; Stern et al., 2014) first locates individual long bones from MR images of the left hand and wrist, followed by a localisation of the expected epiphyseal gap position in every bone. A bounding box capturing the region of the epiphyseal gap is extracted at this location, which allows cropping the 3D region of individual bones, where relevant image features for bone age estimation can be expected that describe the appearance of osseous structures in terms of its MRI signal intensity. After image feature extraction, a machine learning-based regression step automatically maps these image features to a chronological age. The regression function is derived from training data with known chronological age for every bone individually. The last step is the fusion of individual bone age estimates to determine the final bone age prediction. In the following, more details on the automatic age estimation method will be given to develop an intuition on how the algorithm derives age relevant image features which may be visualised and inspected.

#### Extraction of epiphyseal gap images

In the first two steps of our algorithm illustrated in Figure 1, the aim is to automatically extract the epiphyseal gaps of hand and wrist bones from the 3D MRI volumes. To achieve this, we have developed a fully automatic localisation algorithm

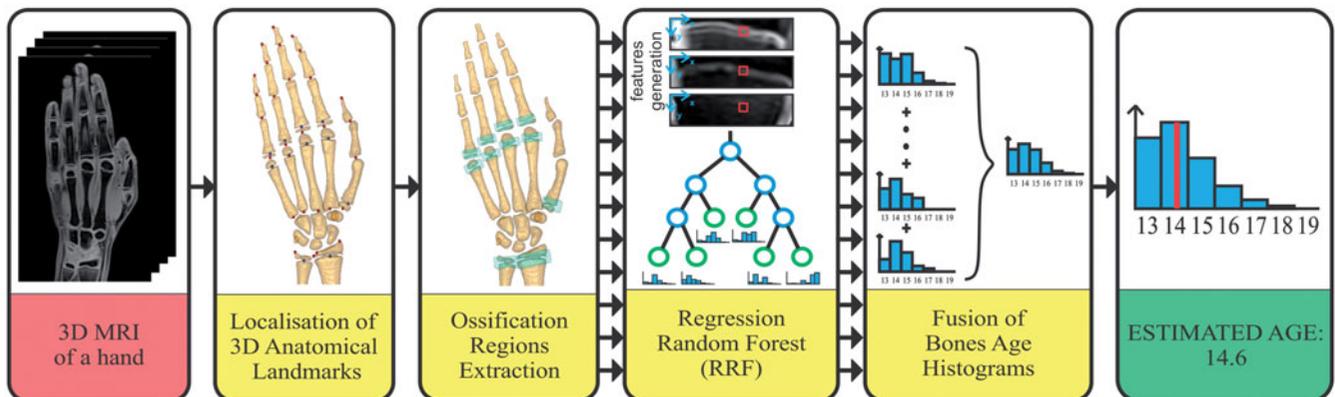


Figure 1. Flowchart diagram for the proposed automatic age estimation method. The major components are bone and epiphyseal gap region localisation, individual bone age estimation and fusion of the age estimates for a final bone age prediction.

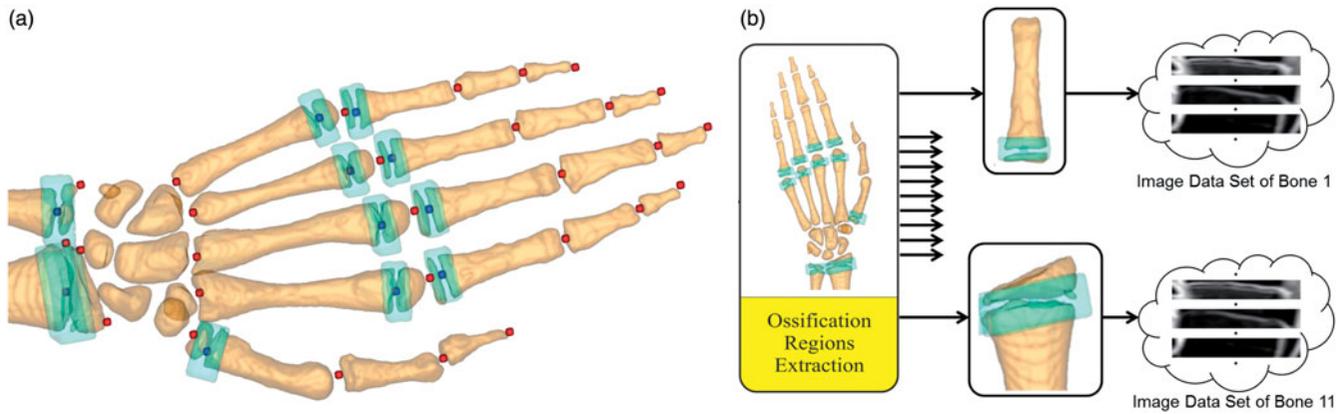


Figure 2. (a) The 28 anatomical hand and wrist bone landmarks (dark-grey, red in color version) as well as 11 landmarks (light-grey, blue in color version) and bounding boxes (transparent green) of the epiphyseal gaps. (b) Extraction of 11 individual bones and corresponding epiphyseal gap regions (transparent green) creates a normalised and aligned set of training data providing the 3D volumetric input data for automatically learning an age estimation function. (Please refer to the online version of this paper for the color figures.)

(Ebner et al., 2014), which is able to predict the location of 28 anatomical landmarks defined between the hand and wrist bones. These anatomical landmarks are shown in red in Figure 2(a). The algorithm derives this landmark prediction by using a machine learning technique (Criminisi & Shotton, 2013), which learns the landmark's position from the surrounding anatomical structures based on a training data set with manually annotated landmarks. The obtained anatomical landmarks are used to crop each individual bone based on its maximally expected 3D extent, as estimated from the training dataset and to rotate the cropped image to a standard orientation defined by the bone axis, thus aligning each individual bone for further processing (see Figure 2(b)).

Since they are most informative for the following automatic bone age estimation step in our investigated age range, we now restrict the localisation of the epiphyseal gap volumes (the green regions in Figure 2(a)) to 11 bones: radius, ulna, the five metacarpals and four proximal phalanges. To automatically predict the epiphyseal gap position, the same algorithm that was used for bone localisation is applied, with a prediction model derived from training data involving known manually annotated epiphyseal gap centre locations. In those bones, where ossification has already finished (corresponding to older subjects), a prediction of where the epiphyseal gap was located during maturation is obtained. This leads to a localisation of the blue points in Figure 2(a). The bounding box of the epiphyseal gap is defined with this centre location, with the estimated diameter of the bone in the axial slice where the centre is located and with a statistical estimate for the height of the gap which is derived from training images. The gap volume is generated by cropping the individual bone image with the obtained bounding box and then normalising the size of the image. Thus, all individual gap images from the 11 investigated bones are aligned and normalised to allow image features discriminating age to be generated in the same co-ordinate system of the respective epiphyseal gap region. Finally, for each of the 11 individual bones a dataset of aligned, normalised epiphyseal gap images is collected from the training set, that enables age estimation in the following step (see Figure 2(b)).

### Age estimation for individual bones

To estimate the age of individual bones, a machine learning approach is used, which determines the mapping of image appearance features to chronological age automatically from a set of training images with known age (see Figure 2(b)). For that purpose, a Regression Random Forest (RRF) is applied, which is based on the idea that combining an ensemble of weak (i.e. only slightly better than random) predictors forms a strong predictor of age in the form of a non-linear regression function (Bishop, 2007; Criminisi & Shotton, 2013). An important advantage of the RRF framework is its ability to perform image feature selection, which means that the algorithm internally selects image features that best discriminate age automatically, thus allowing their visualisation and inspection. Using normalised and aligned epiphyseal gap volumes of individual bones together with their known chronological age as input, an RRF is constructed as an ensemble of binary decision trees. Each decision tree consists of two distinct node types, internal nodes and leaf nodes (see Figure 3(a)). Including the root node where a decision tree starts, internal nodes store a simple, weak decision function that aims to separate the input epiphyseal gap volumes reaching that node according to their age distribution. This is achieved by computing a discriminative image feature and thresholding its computed value for all volumes reaching that node (see Figure 4 for a comparison of a discriminative and a non-discriminative image feature). At each internal node, an image feature is generated by randomly selecting one of the following three types of features: the intensity value at a randomly generated point, the intensity difference between two randomly generated points or the average intensity value along the line that connects two randomly generated points. Motivated by the appearance of the epiphyseal gap in the MR images, we designed these feature generation mechanisms to focus on the age relevant changes in image appearance (see Figure 5). With the intensity value at a specific location, the algorithm can determine whether the epiphyseal gap is present in an image, while the difference of intensity values at two locations also covers the intensity variation between epiphyseal gap and bone. The third generation mechanism is designed to describe intensity changes more smoothly along lines across

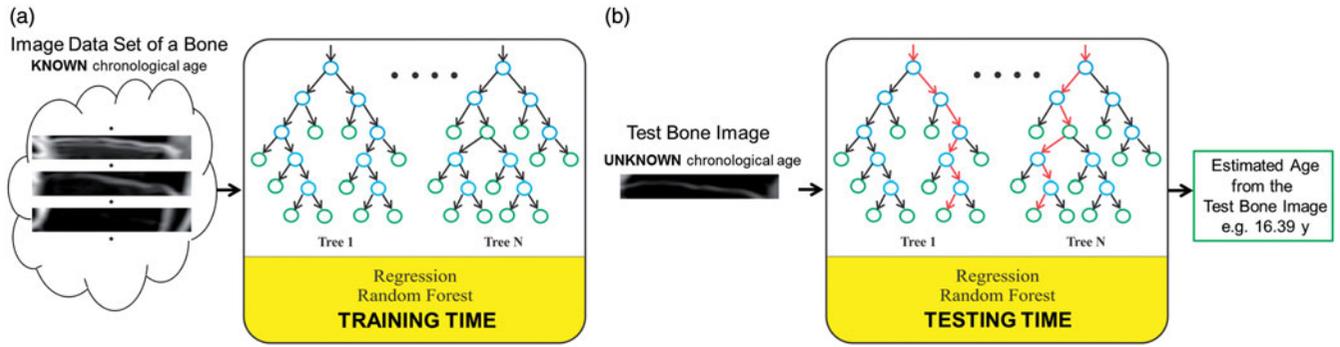


Figure 3. Set-up of a Regression Random Forest consisting of decision trees with nodes and their connections (black arrows). (a) Epiphyseal gap images with known chronological age form a training dataset that allows computation of a forest of decision trees with internal (blue) and leaf (green) nodes, modelling a non-linear age regression function. (b) During testing, epiphyseal gap images are sent through each decision tree (paths with light-grey arrows, red in color version) to achieve age distributions from the reached leaves of the decision tree. Age distributions are combined to form a final age estimate. (Please refer to the online version of this paper for the color figures.)

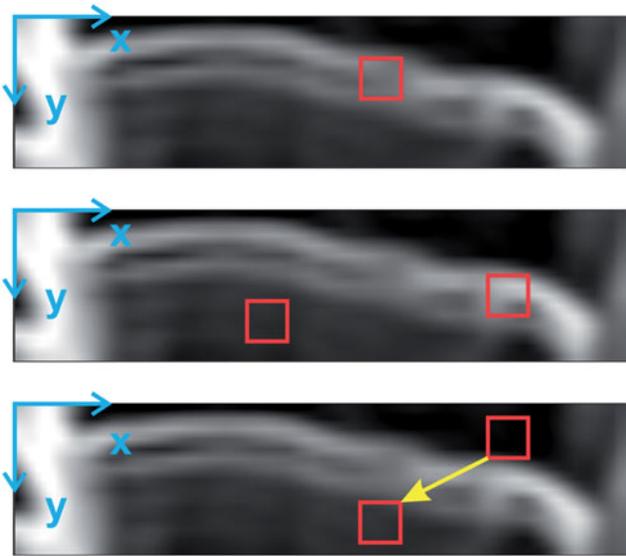


Figure 4. Image features used for training internal nodes of an RRF. A discriminative image feature like the image intensity at a specific, randomly selected location allows distinction between younger and older subjects, while a non-discriminative feature shows no clear relationship with chronological age (CA).

the epiphyseal gap. The second node type is leaf nodes, storing an age distribution of those input epiphyseal gap volumes that have reached it during training. This age distribution is represented as a seven bin histogram (see Figure 6 corresponding to the age groups between 13–20 years, e.g. all input volumes between 14.0–14.99 are grouped into one histogram bin).

Training of the RRF involves an optimisation procedure such that for each node the best discriminating feature and threshold combination is selected from a set of randomly generated image features belonging to a randomly generated feature type and from the set of randomly generated thresholds (see Figure 4 for two random feature examples). The best combination is selected according to an information-theoretic measure, obtained after splitting the input volumes based on the comparison of the computed feature value with the selected threshold  $T_S$  and stored within the internal node. The determined split of input volumes provides the input for the connected left and right nodes of the binary decision tree and training continues in a recursive manner. At the root node

of a decision tree, all uniformly distributed input volumes initialise the recursive procedure. Due to termination criteria like maximum tree depth or minimum number of input volumes reaching a node, the splitting procedure ends within a leaf node of the decision tree. In the leaf node, the age distribution among the input volumes reaching that node is stored as a histogram representing age distribution. Due to the random choices involved in this tree construction and training procedure, it has been shown that the RRF framework excellently generalises to new input data during testing (Criminisi & Shotton, 2013). In the testing stage, where data of unknown age is processed (see Figure 3(b)), epiphyseal gap volumes have to be localised in the same manner as described above and are pushed through all the trees of the forest. Until the gap volume reaches the leaf node, a feature response of the gap volume is generated based on the stored image feature and the volume is passed to the left or right child node depending on the result of the comparison with the stored threshold value. This mechanism is illustrated in Figure 6 for two distinct decision trees. The estimated age distribution of a bone is obtained as a sum of the histograms in the reached leaf nodes of all trees in the forest.

#### Final age estimation

The ageing progress of epiphyseal gaps is not the same for all bones of the hand and wrist (Tanner et al., 1983), since the gradual ossification changes occur at different times for distinct bones. Thus, the ossification of the epiphyseal gap in the metacarpal and proximal phalanges bones is faster than for the ulna and radius bone, except for the thumb metacarpal bone which lies in-between. Depending on this prior knowledge about the differing maturation progress of the individual bones, a bone is included or excluded when age distributions are combined to a final age distribution. This fusion stage is the last step of the pipeline shown in Figure 1, producing a final age distribution for the evaluated subject based on an MR image. The estimated age is obtained as the median value of the subject's final age distribution.

#### Experimental set-up for evaluation

After localising bones and epiphyseal gap volumes per bone, we construct separate RRFs for the radius, ulna, metacarpal

Figure 5. The three types of feature generation mechanisms. Single intensity values (top row), differences between two intensity values (middle row) and mean intensity along a path (bottom row). All features are generated at random locations.

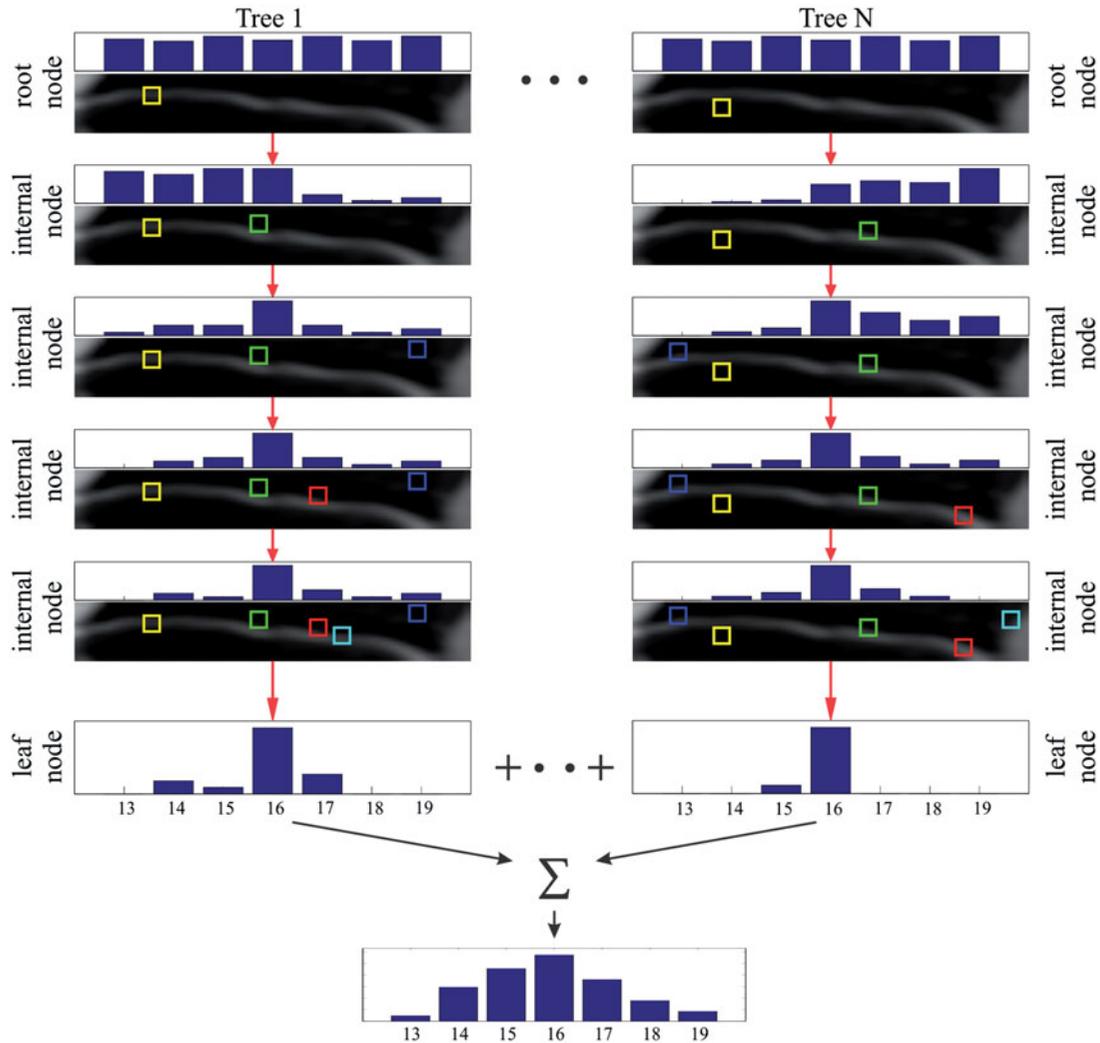
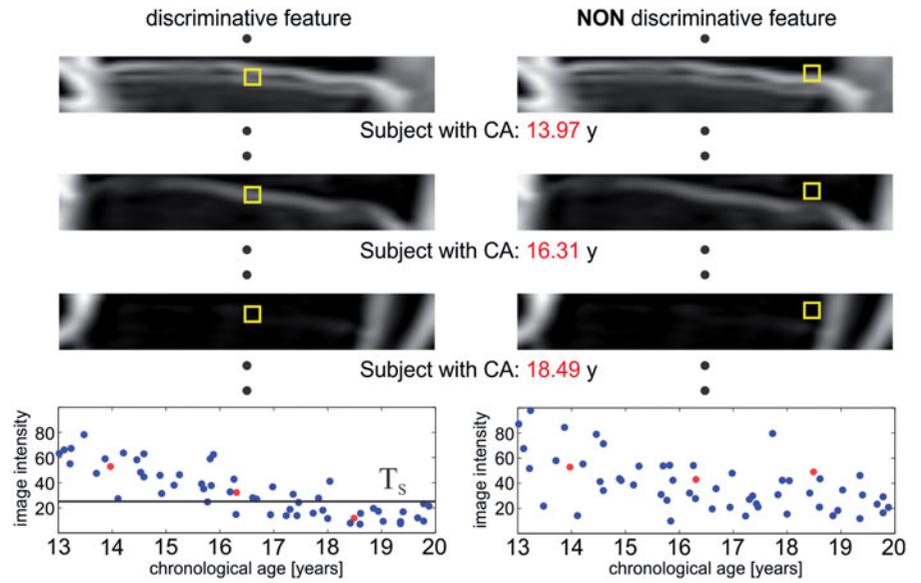


Figure 6. During testing an MRI input volume with unknown chronological age, the image features stored in the RRF are evaluated one after the other for each decision tree. Age distributions are refined while following the path through decision trees, finally reaching a predicted age distribution. These are summed up to finally derive the likely age estimate.

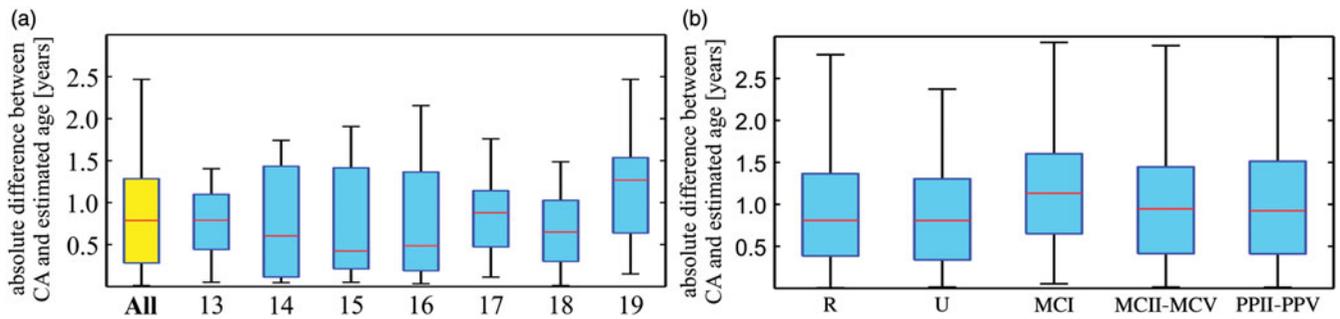


Figure 7. Quantitative box-whisker plot results of our automatic age estimation method showing (a) overall absolute deviations of estimated age from chronological age (CA) as well as separate results for all age groups and (b) absolute deviations for individual bones (R = radius, U = ulna, MCI = thumb metacarpal, MCII-MCV = other metacarpals, PPII-PPV = proximal phalanges).

bone (MCI, i.e. thumb metacarpal bone) and a single forest for the other metacarpal bones (MCII-MCV), as well as a single forest for the proximal phalanx bones (PPII-PPV). We group metacarpals and proximal phalanges, since they show a similar appearance and ageing progress of the epiphyseal gap, while the thumb is treated independently due to its different time of ossification compared to the other metacarpals. Each leaf node of the RRF stores a histogram over the age range with a 1-year resolution between 13–20 years. Decision trees are built with a maximum depth of 5 and each RRF contains 1000 trees, i.e. for every bone a forest with 1000 root nodes was started. Age estimation results are computed in a leave-one-out cross-validation using the 102 available datasets, where we compute the mean absolute difference between estimated and chronological age and its standard deviation over the 102 cross-validation experiments. This involves using 101 images in each round for training a model and testing the model on the remaining image excluded from training. Presentation of statistical results is performed using MATLAB<sup>®</sup> software (The MathWorks Inc., Natick, MA). Box-whisker plots were created using the MATLAB<sup>®</sup> “boxplot” function, while regression of estimated age vs chronological age was created using the “fit” function with a second order polynomial model and the “predint” command for the 95% confidence interval.

## Results

The mean absolute difference between ground truth chronological and estimated age in our dataset of 102 MR volumes was 0.85 years with a standard deviation of 0.58 years. Figure 7(a) shows box-whisker plots that incorporate the results for all images and for each year separately. The error in age estimation for each bone separately is shown in Figure 7(b) as box-whisker plots for radius (R), ulna (U), thumb metacarpal (MCI), long bone metacarpals combined (MCII-MCV) and proximal phalanges combined (PPII-PPV). Figure 8 shows a plot of estimated ages vs ground truth chronological ages along with its 95% confidence interval, obtained by fitting a second order polynomial to the data. Qualitative results of bone age estimation for a radius bone can be seen in Figure 6, where the subject’s radius bone is estimated as being 16.5 years by our algorithm. The corresponding chronological age of the subject is 16.2 years.

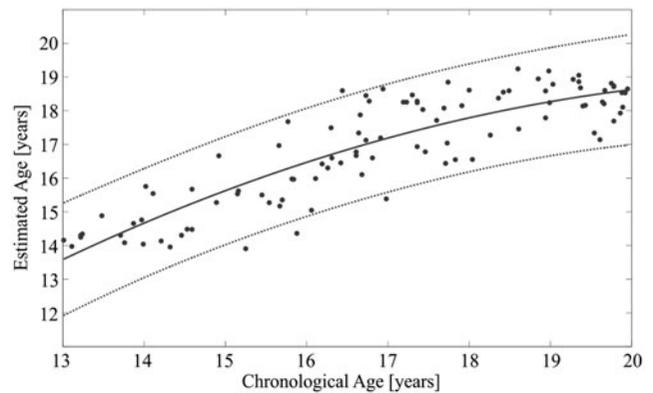


Figure 8. Regression of estimated age vs ground truth chronological age as obtained by fitting a second order polynomial to the data; 95% confidence interval shows uncertainty in prediction due to biological variation.

## Discussion

Our results demonstrate the feasibility of automatic bone age estimation from left hand and wrist MRI volumes based on medical computer vision and machine learning. Such an automated non-linear regression technique introduces benefits compared to established methods, which are lack of exposure to ionising radiation, improved objectivity due to automation and the potential for higher accuracy due to exploiting 3D volumetric data of bone as opposed to 2D projective radiographic images.

Gradual changes during physical maturation are captured by our RRF algorithm by assigning a chronological age distribution learned from a training data set to a selected image feature. Since each image feature corresponds to a maturation characteristic, the sum of all chronological age distributions obtained for all selected features in the image of a subject of unknown age corresponds to its biological age. With the use of biological age to approximate unknown chronological age, uncertainty in the prediction of chronological age has to be taken into account. In Figure 8 this uncertainty can be seen as a variation from the regression curve between estimated and chronological age for all subjects, as determined by our cross-validation experiment. The deviation of the obtained curve from a linear regression for subjects older than 18 years can be explained by the lack

of maturation characteristics available in hand and wrist images. This uncertainty can also be seen quantitatively from Figure 7(a), where the box-whisker plots show the absolute deviation between estimated and chronological age, both for the overall results among all age groups and the single age groups. The latter box-whisker plots show that a larger difference between the estimated and chronological age is present for the age groups between 14–16 years, which may be due to the fact that male children enter puberty during that time period, thus having larger biological variation. When the ossification of the hand and wrist is finished, there are a lack of maturation characteristics available in images for a regression algorithm to learn discrimination among ages, which might be the reason for the larger deviation of the age group older than 19 years. From Figure 7(b), it can be seen that the performances of radius and ulna bones are better than metacarpals and proximal phalanges, due to physical maturation finishing earlier in more distal bones. The highest uncertainty is obtained for the thumb metacarpal bone, since it was trained separately from the other metacarpal bones, i.e. on a smaller data base compared to the pooled metacarpals. The overall results for age estimation of adolescents with a mean absolute difference of 0.85 years are comparable with clinically established methods like manual Greulich-Pyle (Greulich & Pyle, 1959) or Tanner-Whitehouse (Tanner et al., 1983) X-ray atlas comparison, with reported differences from 0.5 up to 2.0 years depending on the age, sex and origin of the examined population (Ritz-Timme et al., 2000). BoneXpert (Thodberg et al., 2009), the most prominent automatic method for bone age estimation from X-ray images, reports a deviation of 0.72 years; however, this result is hard to compare to ours, since it resembles deviation from the manual Greulich-Pyle (Greulich & Pyle, 1959) atlas matching result. The database on which the BoneXpert system is trained, i.e. more than 1700 X-ray images of boys and girls in the age range from 2–17 years, significantly exceeds our dataset size. A current limitation of our work is the restriction to MRI data from males; however, given a database of female volunteers, which we currently do not have available, we expect our method to automatically derive a non-linear regression function from the training data in the same manner. Due to male and female children developing at different speeds, we hypothesise that both regression functions will look similar, but will be shifted by at least a year.

Our age estimation algorithm internally selects image features related to changes in visual appearance due to the gradual progress of growth in subjects. This selection process is performed in a large number of internal nodes of the RRFs decision trees, with each node selecting image features from the epiphyseal gap volumes of the subjects, which are discriminative regarding chronological age. This feature extraction only works if the epiphyseal gap volumes of individual bones are first aligned and normalised in a standardised co-ordinate system, such that features are evaluated at anatomically corresponding locations. Therefore, our automatic anatomical landmark localisation algorithm has to be used (Ebner et al., 2014), which is robust to variations in size and pose of hand and wrist among different subjects. This landmark localisation shows an accuracy of 1.4 mm ( $SD = 1.5$  mm), as has been shown in

Ebner et al. (2014), aligning the bones of all subjects with each other for further processing. Epiphyseal gap volume extraction, which is based on the same algorithm used for anatomical landmark localisation, shows a localisation accuracy of 1.1 mm ( $SD = 0.8$  mm). These two localisation steps have been shown to be sufficient for the following age estimation in Stern et al. (2014), since the goal is to roughly, but robustly, locate the gap region, where the RRF then extracts the image features for learning the non-linear age regression model.

Building upon aligned bones and extracted normalised epiphyseal gap regions, each internal node of the decision trees finds image features, separating input volumes and their visual appearance according to a threshold on the computed feature value. This idea is illustrated in Figure 4, where an example for a discriminative image feature during training is shown compared with a non-discriminative one. This comparison is shown for an intensity value image feature at different locations in the 3D input volumes, corresponding to three subjects of increasing age. The distribution of intensities at this location among all subjects in the training dataset is shown in the bottom row. It can clearly be seen that the discriminative feature shows a relationship with increasing chronological age and that the selected threshold  $T_S$  among the image intensity will split the subjects epiphyseal gap volumes into younger and older ones. Therefore, this image feature, together with the threshold value, resembles a weak decision function as determined by the node optimisation procedure and stored in an internal node of the RRF. On the other hand, the non-discriminative image feature does not show a clear relationship of its feature value with age; therefore, it will not be selected during node optimisation, since no threshold can be found that splits the subject's epiphyseal gap volumes into two sub-sets, which are more informative than the whole set regarding age. The example intuitively demonstrates the ability of the algorithm to pick up image features discriminating age. This mechanism is shown for single intensity features, since they are easier to visualise, but the same arguments hold for the other types of image features, as illustrated in Figure 5.

The evaluation of an individual bone's epiphyseal gap volume of unknown age during testing also allows a deeper understanding of the workings of the RRF algorithm regarding image feature selection. When evaluating an epiphyseal gap volume, it is pushed through all decision trees of the forest (see Figure 3(b), following the path of red arrows), thus visiting a number of internal nodes where image feature tests are performed. Figure 6 shows how the selected path through a tree gradually improves certainty in age prediction. At the root node of a decision tree, before evaluating the image feature test, the initial age distribution corresponds to the age distribution in the whole training set, e.g. in our case it is nearly uniformly distributed, suggesting that each age group is equally likely to be predicted. After evaluating the feature test in the root node (the yellow feature in Figure 6), tree 1 decides for this epiphyseal gap volume that an age below 17 is likely, while tree N votes for an age above 15. These predictions can be seen from the age distribution histograms below the shown slice of the epiphyseal gap volume, indicated by the red arrow. The comparison of the computed feature value with the

threshold value stored in the internal node determines the next node, which is visited in each decision tree. With the second feature in both trees, the age distribution histograms get sharper, thus allowing a step-by-step elimination of uncertainty in the prediction while the input volume traverses towards a leaf node for all trees. After reaching a leaf node, the stored age distribution histogram of each decision tree is reported and a sum over all histograms of the forest is computed, thus generating a strong age prediction from many weak predictors. From this process of decision tree traversal, it can be seen that it is not single decisions of internal nodes or individual decision trees that are relevant for the success of the RRF, but the power of the method comes from the combination of image features and the ensemble of its weak predictions. In the example of Figure 6, tree 1 starts with a prediction of being in the younger half of the age range, while tree N predicts the opposite situation. However, the combination of image features lets both trees end up with a more accurate age distribution, with tree N finally showing a very different prediction in the leaf node (very likely in age group 16 or 15) compared to the root node (likely older than 15, with a mean estimate of above 17). Thus, it can be seen that the RRF framework automatically selects exactly those anatomical structures from the training input volumes that are most informative for describing age in a given dataset.

## Conclusion

From the presented investigation of the automatically selected image features and their combination, it can be concluded that the RRF framework is well suited for non-linear regression of MR image features to chronological age. In addition it allows the visualisation and interpretation of selected image features in a clear and intuitive way, thus finding those image structures in MR images that are most informative for describing age in a given dataset. Extending this idea to analyse novel, yet unexplored soft tissue features visible in MRI like cartilage or bone marrow (Tomei et al., 2014) is an interesting step for further research.

The image feature selection is embedded in an automatic age estimation pipeline based on medical computer vision and machine learning, consisting of automated bone and epiphyseal gap localisation, age regression of individual bones from aligned epiphyseal gap volume data and fusion of bone age distributions to a final age prediction. On a database of 102 male European-derived subjects between 13–20 years, we were able to estimate the subjects age with a mean absolute difference of 0.85 years ( $SD = 0.58$  years) compared to the chronological age, which is in line with results using established methods based on X-ray projections that involve radiation exposure. In future work we will investigate the use of this approach on a larger database, if possible including female subjects as well, and investigate how to extend the method to a larger age range involving younger children as well and we will explore its use for specific clinical and legal medicine applications.

## Acknowledgements

This work was partly supported by a European Community FP7 Marie Curie Intra European Fellowship (331239).

## Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

## References

- Bassed RB. 2012. Advances in forensic age estimation. *Forensic Sci Med Pathol* 8:194–196.
- Bishop CM. 2007. *Pattern recognition and machine learning*. Secaucus, NJ: Springer-Verlag New York, Inc.
- Black S, Aggrawal A, Payne-James J, editors. 2010. *Age estimation in the living – The Practitioner's Guide*. West Sussex, UK: Wiley-Blackwell, John Wiley and Sons.
- Criminisi A, Shotton J, editors. 2013. *Decision forests for computer vision and medical image analysis*. London, UK: Springer Publishing Company, Inc.
- Dean C, Liversidge HM, Elamin F. 2014. Combining radiographic and histological data for dental development to compare growth in the past and present. *Ann Hum Biol* 41:336–347.
- Dirkmaat D, editor. 2012. *A companion to forensic anthropology*. Chichester, UK: Wiley-Blackwell.
- Dutour O. 2012. Age estimation of the human skeleton (review). *Hum Biol* 84:445–447.
- Dvorak J, George J, Junge A, Hodler J. 2007. Age determination by magnetic resonance imaging of the wrist in adolescent male football players. *Br J Sports Med* 41:45–52.
- Ebner T, Stern D, Donner R, Bischof H, Urschler M. 2014. Towards automatic bone age estimation from MRI: localization of 3D anatomical landmarks. In: Golland P, Hata N, Barillot C, Hornegger J, Howe R, editors. *Medical image computing and computer-assisted intervention – MICCAI 2014: 17th International Conference, Boston, MA, Vol. 8674 of LNCS*. Heidelberg, Germany: Springer. p 421–428.
- Freitas D, Maia J, Beunen G, Lefevre J, Claessens A, Marques A, Rodrigues A, et al. 2004. Skeletal maturity and socio-economic status in portuguese children and youths: the madeira growth study. *Ann Hum Biol* 31:408–420.
- Greulich WW, Pyle SI. 1959. *Radiographic atlas of skeletal development of the hand and wrist*. 2nd ed. Stanford, CA: Stanford University Press.
- Hillewig E, Degroote J, van der Paelt T, Visscher A, Vandermaele P, Lutin B, D'Hooge L, et al. 2013. Magnetic resonance imaging of the sternal extremity of the clavicle in forensic age estimation: towards more sound age estimates. *Int J Legal Med* 127:677–689.
- Lee SC, Shim JS, Seo SW, Lim KS, Ko KR. 2013. The accuracy of current methods in determining the timing of epiphysodesis. *Bone Joint J* 95-B:993–1000.
- Liversidge HM. 2012. The assessment and interpretation of Demirjian, Goldstein and Tanner's dental maturity. *Ann Hum Biol* 39:412–431.
- Mansourvar M, Ismail MA, Herawan T, Raj RG, Kareem SA, Nasaruddin FH. 2013. Automated bone age assessment: motivation, taxonomies and challenges. *Comput Math Method M* 2013: 391626.
- Marshall WA, Tanner JM. 1969. Variations in pattern of pubertal changes in girls. *Arch Dis Child* 44:291–303.
- Marshall WA, Tanner JM. 1970. Variations in the pattern of pubertal changes in boys. *Arch Dis Child* 45:13–23.
- Martin D, Deusch R, Schweizer R, Binder G, Thodberg HH, Ranke MB. 2009. Clinical application of automated Greulich-Pyle bone age determination in children with short stature. *Pediatr Radiol* 39: 598–607.
- Martin DD, Wit JM, Hochberg Z, Saevendahl L, van Rijn RR, Fricke O, Cameron N, et al. 2011a. The use of bone age in clinical practice – Part 1. *Horm Res Paediatr* 76:1–9.
- Martin DD, Wit JM, Hochberg Z, van Rijn RR, Fricke O, Werther G, Cameron N, et al. 2011b. The use of bone age in clinical practice – Part 2. *Horm Res Paediatr* 76:10–16.
- Ritz-Timme S, Cattaneo C, Collins MJ, Waite ER, Schuetz HW, Kaatsch HJ, Borrman HI. 2000. Age estimation: the state of the art in relation to the specific demands of forensic practise. *Int J Legal Med* 113: 129–136.
- Saint-Martin P, Rerolle C, Dedouit F, Rousseau H, Rouge D, Telmon N. 2014. Evaluation of an automatic method for forensic age estimation by magnetic resonance imaging of the distal tibial epiphysis – a preliminary study focusing on the 18-year threshold. *Int J Legal Med* 128:675–683.
- Schmeling A, Garamendi PM, Prieto JL, Landa MI. 2011. Forensic age estimation in unaccompanied minors and young living adults. In:

- Vieira DN, editor. Forensic medicine – From old problems to new challenges, chapter 5. Rijeka, Croatia: InTech. p 77–120.
- Sonka M, Fitzpatrick JM, editors. 2000. Handbook of medical imaging: volume 2 – Medical image processing and analysis. Bellingham, WA: SPIE Press.
- Stern D, Ebner T, Bischof H, Grassegger S, Ehammer T, Urschler M. 2014. Fully automatic bone age estimation from left hand MR images. In: Golland P, Hata N, Barillot C, Hornegger J, Howe R, editors. Medical image computing and computer-assisted intervention – MICCAI 2014: 17th International Conference, Boston, MA, USA, Vol. 8674 of LNCS. Heidelberg, Germany: Springer. p 220–227.
- Tanner JM. 1978. Foetus into man: physical growth from conception to maturation. London: Open Books.
- Tanner JM, Whitehouse RH, Marshall WA, Healy MJR, Goldstein H. 1983. Assessment of skeletal maturity and prediction of adult height (TW2 method). 2nd ed. London, New York, San Francisco: Academic Press.
- Terada Y, Kono S, Tamada D, Uchiumi T, Kose K, Miyagi R, Yamabe E, Yoshioka H. 2013. Skeletal age assessment in children using an open compact MRI system. *Magn Reson Med* 69:1697–1702.
- Thodberg HH, Jenni OG, Ranke MB, Martin DD. 2012. Standardization of the Tanner-Whitehouse bone age method in the context of automated image analysis. *Ann Hum Biol* 39:68–75.
- Thodberg HH, Kreiborg S, Juul A, Pedersen KD. 2009. The BoneXpert method for automated determination of skeletal maturity. *IEEE Trans Med Imaging* 28:52–66.
- Thodberg HH, Saevendahl L. 2010. Validation and reference values of automated bone age determination for four ethnicities. *Acad Radiol* 17:1425–1432.
- Tomei E, Sartori A, Nissman D, Al Ansari N, Battisti S, Rubini A, Stagnitti A, et al. 2014. Value of MRI of the hand and the wrist in evaluation of bone age: preliminary results. *J Magn Reson Imaging* 39:1198–1205.
- Wang WW, Xia CW, Zhu F, Zhu ZZ, Wang B, Wang SF, Yeung BH, et al. 2009. Correlation of Risser sign, radiographs of hand and wrist with the histological grade of iliac crest apophysis in girls with adolescent idiopathic scoliosis. *Spine* 34:1849–1854.
- Wenzel A, Droschl H, Melsen B. 1984. Skeletal maturity in Austrian children assessed by the GP and the TW-2 methods. *Ann Hum Biol* 11:173–177.