Expert Systems With Applications 50 (2016) 75–88

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Expert system for automated bone age determination

Jinwoo Seok[a,*], Joséphine Kasa-Vubu[b], Michael DiPietro[c], Anouck Girard[a]

[a] Department of Aerospace Engineering, University of Michigan, Ann Arbor, Michigan 48109, United States
[b] Department of Pediatrics-Division of Pediatric Endocrinology, University of Michigan Health System, Ann Arbor, Michigan 48109, United States
[c] Department of Radiology, Section of Pediatric Radiology University of Michigan Health System, Ann Arbor, Michigan 48109, United States

ARTICLE INFO

Keywords:
Bone age
Expert system
Automated system
Fusion rule
Classification

ABSTRACT

A novel automated bone age determination algorithm using left hand X-ray images, which provides consistent overall bone age as well as five part bone ages, is presented in this paper. Based on the descriptive narrative from the Greulich and Pyle atlas as well as those from other more recent studies, 17 region of interests are selected and based on anatomical similarity, five clinically relevant groupings (or "parts") are defined on left hand X-ray images. When disharmonious maturations for different regions of interest are large, providing part bone ages with overall bone age is helpful to pediatricians. Based on interviews with two experts to get input on their bone age determination strategy, overall bone age determination can be viewed as the weighted sum of "part bone ages" of the five parts. Using the method of least squares and inputs from five (human) readers, we extract weights for bone age determination using all five parts (as well as reduced algorithms using only four, three or two parts). The weights indicate that part 1 (distal joints) has the highest priority. Overall bone age is then estimated based on the weights and bone ages of available parts. In our work, a computer vision algorithm provides bone ages of individual regions of interest. To combine the region of interest computer classifiers and generate each of the five part bone ages, we develop and analyze fusion rules of multiple classifiers with more than three classes each. The fusion rules take into account performance of each region of interest classifier. Once the part bone ages are obtained based on the fusion rule and region of interest classifiers, the overall bone age is determined in a fully automated way. Finally, we give a use case for the whole automated bone age determination system and validation of the algorithm based on given performance of each region of interest computer classifier. Results indicate that the algorithm works well with reasonably good classifiers.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Motivation

Early detection of possible growth disorders or abnormal pubertal maturation is an important aspect for wellbeing. The assessment of growth and pubertal maturation is central to the practice of pediatric endocrinology. Bone age (BA) is a measure of the degree of maturation of a child's skeleton. Its reliable assessment is a key reference in growth and maturation evaluation.

The Greulich and Pyle (GP) (Greulich & Pyle, 1959) atlas is the predominant clinical reference to determine BA in pediatric endocrinology in the United States. To determine BA, a radiologist compares a patient's X-ray to those contained in the reference atlas. The patient's X-ray is then assigned the BA corresponding to the closest matching atlas image.

When comparing the X-ray and the atlas, clinicians focus on a number of areas of special interest, subsequently called Regions of Interest (ROI). Disharmonious maturations may be found in BA assessment for different ROIs. However, there is no standard rule to determine the overall BA when the maturation of each ROI is different. In this situation, the clinicians rely on their experience and personal opinions. From an engineering perspective, the method is human-centric and subject to bias in interpretation. In other words, different radiologists may use different criteria when matching the patient's X-ray to the atlas, which may cause different overall BA readings. Furthermore, if disharmonious maturations are significant, overall BA may not be representative, and clinicians may benefit from additional information on maturation of different parts in the hand. However, no formal method to provide this information exists. Generating and providing that additional information would be helpful for the clinicians.

* Corresponding author. Tel.: +17183168589.
E-mail addresses: sjinu@umich.edu (J. Seok), jzkv@med.umich.edu (J. Kasa-Vubu), dipietro@med.umich.edu (M. DiPietro), anouck@umich.edu (A. Girard).

http://dx.doi.org/10.1016/j.eswa.2015.12.011
0957-4174/© 2015 Elsevier Ltd. All rights reserved.

There are other reasons why the current methodology is problematic. The GP standard is outdated: it was created in the 1950s, from a small sample of Caucasian children. The images in the standard are irregularly spaced, with large gaps in clinically significant ranges. The overall goal of our research is the establishment of more modern standards that are also more representative of the current population. However, over the years, several different new standards have been proposed (Eklőf & Ringertz, 1967; Gilsanz & Ratib, 2008; Tanner, Healy, Goldstein, & Cameron, 2001), and none has taken root in the United States. By developing an automated assistant that gives BA readings in both the GP standard and the new, superior standard, acceptance should be broader.

Thus, we aim to build an automated bone age determination system that is accurate, consistent and provides additional subpart information. From an engineering perspective, our system can be divided into two main modules; (1) computer vision and ROI classification, and (2) expert system for integration of results into overall BA. The computer vision module first detects each ROI and then uses a classifier specific to each ROI to determine its BA. The expert system module integrates the results from the computer vision and classification modules to generate a final BA reading result. In this paper, we focus on the expert system and integration of individual classification results into an overall BA.

Here, we structure our primary system based on in-depth interviews with two clinicians, one a pediatric radiologist, one a pediatric endocrinologist. We then calibrate the system using five expert BA readers. Readers were selected for experience. Three are pediatric radiologists; two are pediatric endocrinologists. Readers assessed BA on 30 images, all in the female 6 years 10 months (female standard 15) to 13 years 6 months (female standard 22) range, which is of greatest clinical interest to pediatric endocrinologists for girls, using the GP standard.

1.2. Literature review

The most common methods to determine BA (Mughal, Hassan, & Ahmed, 2014) are the use of the GP atlas and the Tanner Whitehouse (TW) method (Tanner et al., 2001). To determine a BA using the GP atlas, radiologists compare a patient's left hand X-ray images with the GP atlas to find the closest matching image in the atlas. The GP atlas contains 31 standard X-ray images for males from newborn to 19 years old and 27 standard X-ray images for females from newborn to 18 years old. Two more X-ray images for females are included after female standard 27, that are 28 and 50 years old. As stated in the GP atlas, all the children were white, all had been born in the United States, and almost all were of North European ancestry. The children were examined at three-month intervals during the first postnatal year, at six-month intervals from 12 months to five years of age, and annually thereafter; however, only a selection of the X-ray images is provided in the atlas. Consequently, the gaps between the standards are not consistent. For males, the gaps between the standards are 3 months to 12 months and for females, the gaps between the standards are 3 months to 14 months. These inconsistencies make rating images by using the GP BA difficult.

The atlas contains text explanations next to each standard to help the radiologist find the right match. The length of explanations varies from two to 23 lines of text, and some standards have BA of specific parts (e.g. Distal end of radius, Metacarpal 1, etc...) above the text explanation. The detailed explanation of BA rating of specific parts with pictures is in the appendix at the end of the atlas. Eventually, radiologists rely on their experience and personal opinions when they determine GP BA.

There is a more recent book to help to decide GP BA (Gaskin, Kahn, Bertozzi, & Bunch, 2011) with clearer images and a graphical explanation of each standard on each image, so that the radiologist can easily see important features of each standard. However, it can still be difficult to determine BA by comparing patients X-ray images with the atlas, and the overall BA assessment when ROIs present different degrees of maturation is left to the physician.

There are three versions of the TW method (TW1, TW2 and TW3) with the most recent one being the TW3 method. To determine TW BA (TW3), radiologists evaluate each ROI's maturation by a "grade" of A through I and assign a weighted score to each ROI according to the maturation using the table given in (Tanner et al., 2001). Then, they combine all scores to get a final score and convert this to TW BA using the given table. TW3 method defines 20 regions of interests (ROIs) to determine a TW BA (Tanner et al., 2001). It is time consuming until one gets used to the method because all 20 ROIs have to be evaluated, then combined into the final answer. In the TW2 method, three BA rating methods are provided: (1) radius, ulna and short bones (RUS) only; (2) the carpal bones (CARPAL) only; (3) full 20-bone score (TW2 (20)). In the TW3 method, the authors use only the first two ratings (RUS/CARPAL) and discontinue the full 20-bone score because the maturities of RUS and carpal bones are independent. The TW method is generally more accurate than GP BA.

However, one problem is that neither method is really accurate. Inconsistent BA ratings have been described in the literature. In Bull, Edwards, Kemp, Fry, and Hughes (1999), BA of 362 radiographs were evaluated using the GP method by a succession of 12 radiology trainees over a four year period, and using the TW method by one of two nurse auxologists who had received specific training in the use of the TW2 method. Then, for both methods, 39 radiographs were reevaluated by the same reader for intra-observer variation. Intra-observer variation for the GP method was 0.14 ± 1.16 years (-2.46 to 2.18 in 95% confidence limit) and for the TW method was 0.01 ± 0.71 (-1.41 to 1.43 in 95% confidence limit). In another example, four radiologists evaluated GP BA of 107 radiographs and reevaluated 25 randomly selected radiographs (Berst et al., 2001). The average difference in bone age assessment among observers was 0.69 ± 0.48 years and the average difference in bone age assessment within individual observers ranged from 0.29 to 0.49 years.

To resolve the inconsistency, attempts have been made to obtain computer-based BA determination for both GP and TW BA. A first step is computer-aided BA determination systems such as computer-assisted skeletal age scores (CASAS) (Pietka et al., 2001; Tanner, Gibbons, & Bock, 1992). By image preprocessing and extraction of the epiphyseal/metaphyseal region of interest (EMROI), four features (epiphyseal diameter, metaphyseal diameter, distance between lower diaphysis and epiphyseal diameter) are provided to help radiologists determine BA (Pietka et al., 2001). This work has been extended to an automated BA determination system. Reference Gertych, Zhang, Sayre, Pospiech-Kurkowska, and Huang (2007) provides a database of a collection of 1400 digital left hand X-ray images of Caucasian, Asian, African-American and Hispanic patients for male and female standards, ranged from 1 to 18 years old, and a computer-assisted diagnosis (CAD) module to determine BA automatically. The CAD module detects 7 ROIs (carpal and six phalangeals), extracts features and trains fuzzy classifiers, then obtains BA by combining bone age value in fuzzy classifiers. However, they use CAD BA which is their own BA rating system.

An image registration approach can be found in Martín-Fernández, Martín-Fernández, and Alberola-López (2003). Image registration is the determination of a geometrical transformation that aligns points in one image of an object with corresponding points in another image of the same or another object (Martín-Fernández et al., 2003). The authors claim that registration of testing images to the standard images, which have been previously computed and taken to be the ground truth, allows for automatic computation of GP BA. They also claim that a

Download English Version:

https://daneshyari.com/en/article/382010

Download Persian Version:

https://daneshyari.com/article/382010

Daneshyari.com