Forensic Science International 258 (2016) 81–87

Contents lists available at ScienceDirect

Forensic Science International

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

Probabilistic age classification with Bayesian networks: A study on the ossification status of the medial clavicular epiphysis

Emanuele Sironi [a,*], Vilma Pinchi [b], Franco Taroni [a]

[a] School of Criminal Justice, University of Lausanne, Building Batochime, 1015 Lausanne-Dorigny, Switzerland
[b] Department of Health Sciences, Section of Forensic Medical Sciences, University of Florence, Largo Brambilla 3, 50134 Florence, Italy

ARTICLE INFO

Article history:
Received 31 March 2015
Received in revised form 1 September 2015
Accepted 17 November 2015
Available online 27 November 2015

Keywords:
Forensic medicine
Forensic age estimation
Skeletal age assessment
Medial clavicular epiphysis development
Probabilistic approach
Bayesian networks

ABSTRACT

In the past few decades, the rise of criminal, civil and asylum cases involving young people lacking valid identification documents has generated an increase in the demand of age estimation. The chronological age or the probability that an individual is older or younger than a given age threshold are generally estimated by means of some statistical methods based on observations performed on specific physical attributes. Among these statistical methods, those developed in the Bayesian framework allow users to provide coherent and transparent assignments which fulfill forensic and medico-legal purposes. The application of the Bayesian approach is facilitated by using probabilistic graphical tools, such as Bayesian networks. The aim of this work is to test the performances of the Bayesian network for age estimation recently presented in scientific literature in classifying individuals as older or younger than 18 years of age. For these exploratory analyses, a sample related to the ossification status of the medial clavicular epiphysis available in scientific literature was used. Results obtained in the classification are promising: in the criminal context, the Bayesian network achieved, on the average, a rate of correct classifications of approximatively 97%, whilst in the civil context, the rate is, on the average, close to the 88%. These results encourage the continuation of the development and the testing of the method in order to support its practical application in casework.

1. Introduction

Since knowing the age of an individual is necessary for a large number of social and juridical situations (both civil and criminal), the practice of age estimation of living persons has extremely gained in importance in the past few decades [1–4]. This is mostly due to the increase in cross-border migration movements or human trafficking involving people unable to provide valid documentary evidence for attesting their age [5]. Generally, the range of ages having a particular legal interest is between 14 and 22 years of age, thus it is essential to be able to estimate the chronological age across this interval and/or to decide whether an individual belongs to a given age class included in this range [3,4]. According to some approaches, such as those proposed by the Study Group on Forensic Age Diagnostic (AGFAD) and the Forensic Anthropology Society of Europe (FASE) [6], age estimation should be appropriately based on the examination of some physical attributes, generally related to bones and dental maturation [4,7,8]. The age estimation is usually performed in two steps: first, the degree of maturity reached by a physical indicator is assessed in ordered developmental stages or through the assignment of score values [9–11]. Then, the observed degree of maturity is converted into the estimated chronological age or in the probability that the individual is included in a meaningful range of ages by means of different statistical methods, usually related to a specific physical attribute. Results should be provided in an appropriate form in order to fulfill the specific forensic and legal needs. From this point of view, the Bayesian approach is particularly suited because it allows the user to coherently deal with the uncertainty related to age estimation and to provide results in a transparent and logical form [12]. Some Bayesian methods allow one to update an initial (or *prior*) belief about the chronological age of the examined individual in the light of the observed degree of maturity of a given physical attribute [12–14]. Results are then provided in the form of a *posterior probability distribution*, which encapsulates all uncertainty associated with the estimated quantity, i.e., the chronological age. This distribution is then used to compute the probability that the examined individual is younger or older than an age threshold of interest, such as the

* Corresponding author. Tel.: +41 21 692 46 21.
E-mail addresses: emanuele.sironi@unil.ch (E. Sironi), vilma.pinchi@unifi.it (V. Pinchi), franco.taroni@unil.ch (F. Taroni).

http://dx.doi.org/10.1016/j.forsciint.2015.11.010

age of majority [12–14]. The practical application of Bayesian methods can be facilitated to a great extent by using specific probabilistic graphical tools, such as Bayesian networks [15]. Bayesian networks combine elements of both graph and probability theories. The first one is applied to qualitatively define the structure of the model, taking into account the variables of interest (represented by nodes) and their respective probabilistic relationships (represented by directed arcs). The probability theory is then applied to define the nature of these relationships and to quantify their strength [15,16]. Recently, a Bayesian network for age estimation was presented in scientific literature [14]. The network allows users to take into account evidence related to the degree of maturity observed in an examined individual and to provide both posterior probability distribution on the chronological age, as well as the posterior probabilities that the individual is older or younger than a given age threshold.

The aim of this paper is to study the performances of the Bayesian network in classifying individuals as younger or older than 18 years of age (i.e., the age of majority in numerous countries worldwide) on the basis of data related to the medial clavicular epiphysis development. Therefore, misclassifications (i.e., the rate of false minors and false majors) were investigated. Furthermore, the influence of prior probabilities assignment was also inspected. Analyses were performed using a sample available in scientific literature [17]. Note that the sample cannot serve as reference database and it was used only for exploratory analysis.

2. Material

For the analysis exposed in this paper, the data sample presented by Kreitner et al. was used [17]. The sample contains 380 European white subjects (males and females) under the age of 30 years lacking of development disorders. Each subject is described by his or her age and the classification of the degree of maturity reached by the medial clavicular epiphysis. The degree of maturity was assessed by an experienced radiologist according to a traditional four-stage classification [17,18] by means of a Computed Tomography (CT) scan examination. For the analyses carried out in this work, subjects counted in groups of heterogeneous ages (i.e. 0–4 and 5–9) were considered as being the maximal age of the range (i.e. 4 and 9 years old). Table 1 shows the sample distribution per age. Further details on the data sample are available in the paper of Kreitner et al. [17].

The data sample was randomly subdivided into a training dataset composed of 250 individuals and a testing dataset with the remaining 130 individuals. The testing dataset was then treated in order to remove individuals younger than 14 and older than 24 years of age, so that it would be considered for the analyses only individuals for whom an appraisal for age classification according to a threshold of 18 years of age appears to be reasonable. The procedure was randomly repeated 1000 times, producing 1000 independent training data sets and 1000 related testing data sets. The subsamples were used for the analyses on the classification performances, whilst the whole data sample was used for investigating the influence of the prior beliefs on results.

3. Methods

Analyses were carried out with the Bayesian network for age estimation [14]. The structure of the network considers three variables: the chronological age (node **Age** in Fig. 1), the degree of maturity reached by the medial clavicular epiphysis, expressed in form of the assigned developmental stage (node **Stage** in Fig. 1) and the event that the examined individual is younger or older than 18 years of age. This latter is expressed in the form of two mutually exclusive propositions, P_1: the examined individual is older than 18 years of age and P_2: the examined individual is younger than 18 years of age (node **Class** in Fig. 1) [14]. Each node in the network is defined in a finite set of mutually exclusive states which cover the domain of existence of the variable. Nodes' states are described in the Table 2 [14].

Table 1
Subjects by age (years).

Age	Subjects
4	18
9	22
10	7
11	8
12	5
13	11
14	13
15	18
16	19
17	13
18	17
19	20
20	22
21	15
22	16
23	22
24	20
25	34
26	27
27	17
28	23
29	13
Total	380

The network structure was built in order to consider the dependences between the variables of interest (Fig. 1). The arc from node **Age** to node **Stage** represents the biological consideration that the developmental stage observed on an individual depends on its chronological age, whilst link between **Age** and **Class** reflects the fact that the two considered age classes are a merely subdivision of the chronological age space.

As prior belief on the chronological age, a Skew–Normal distribution with a location parameter of 15, a scale parameter of 8 and a slant parameter of 6 ($\mathcal{SN}$ ($\xi = 15$, $\omega = 8$, $\alpha = 6$)) was assigned. This probability distribution was filled in into the node **Age** in the network structure presented in Fig. 1 [19]. Probabilities in the node **Class** are then automatically computed by the Bayesian network. Probabilities on the node **Stage** are assigned by means of *transition analysis* [20] employing a *continuation ratio logit* as

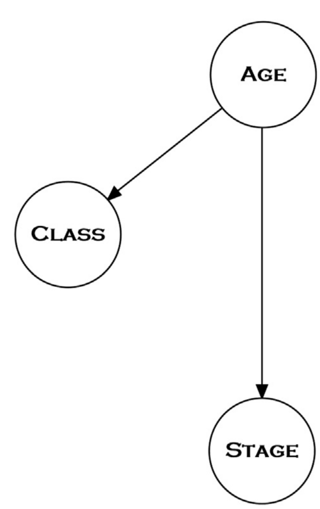

Fig. 1. Structure of the Bayesian network for age estimation.