COMPUTERS AND BIOMEDICAL RESEARCH **28,** 239–256 (1995)

# Automatic Analysis of Hand Radiographs for the Assessment of Skeletal Age: A Subsymbolic Approach

M. RUCCI,* G. COPPINI,† I. NICOLETTI,‡ D. CHELI,‡ AND G. VALLI§

**Scuola Superiore S. Anna, Pisa, Italy*; *†CNR Institute of Clinical Physiology, Pisa, Italy*; *‡Centro Studi Auxologici, Firenze, Italy*; *and §Department of Electronic Engineering, University of Florence, Florence, Italy*

Received December 11, 1994

The assessment of skeletal maturity is crucial for the analysis of growth disorders and plays an important role in paediatrics. For this reason, several methods have been developed for estimating skeletal maturity. Among them, the Tanner and Whitehouse method (TW2), which is based on the analysis of hand radiographs, is usually considered the most accurate and reliable. Nevertheless, TW2 is applied only in a small fraction of cases, due to its complexity and long examination times. Thus, the development of automated systems which reliably implement this method is highly desirable. However, major difficulties have been found in the development of computer-based systems for the assessment of skeletal maturity. In particular the extraction of the bones of interest has proved to be extremely challenging. In this paper, we propose a system architecture for the implementation of the TW2 method, which is based on artificial neural networks. For each bone considered, the maturation stage is determined by means of a two-step process which first locates the position of the bone in the radiograph and then analyzes the bone shape. Experimental results obtained with our implementation of the carpal version of TW2 are in good agreement with those provided by trained observers. 

## 1. INTRODUCTION

The assessment of skeletal maturity is crucial for a proper diagnosis and treatment of growth disorders and is an important clinical tool in pediatrics. As already pointed out by several researchers at the beginning of this century (*1*, *2*), the skeleton shows maturational changes which occur continuously from intrauterine life to full maturity. On the grounds of these findings, several methods have been developed for evaluating skeletal maturity, which are based on the measurement of the development stage of sets of bones.

The most commonly used methods are based on the analysis of radiographs of the left hand. The most accurate of these is commonly agreed to be the method developed in the early '70s by Tanner and Whitehouse (*3*) (TW2 method), which operates on a selected set of bones of the hand.

However, as shown by a statistical analysis of examinations performed in the



0010-4809/95 $6.00

United States (*4*), the Tanner and Whitehouse method is usually applied in no more than 20% of the cases, whereas in approximately 70% of examinations more simple methods, such as the Greulich and Pyle atlas (*5*), are preferred. This limited use of TW2, in spite of its superior accuracy, can be attributed to the longer time required for examination and the greater complexity of the method. In order to produce reliable results, a well trained observer is usually required, and the maturity stages of several bones have to be carefully evaluated.

Automated computer-based systems which reliably implement the TW2 method could find a wide application in hospital and clinical centers. In fact, they could make it possible to achieve accurate estimates with considerable savings of human time and work. Also, they could contribute to solving the critical problem of performing comparisons among data from different institutes by providing a standard measurement. Furthermore, they could be valuable in the training of specialized personnel.

Recently, attempts at the development of such systems have been made by many researchers. Different approaches have been proposed, involving two separate phases of *bone extraction* and *feature analysis*. In the first phase, the radiographic image is segmented and the bones of interest for the TW2 method are extracted. The shapes of the bones and the features which significantly characterize the maturation stage are then analyzed in the second phase.

Due to the difficulties of accurately extracting the bones in X-ray images, most of the works described in literature have focused on the problem of radiograph segmentation, and many efforts have been made toward the development of reliable techniques.

Both edge-based (*6*, *7*) and region-based (*8*) approaches have been attempted, often including *a priori* knowledge on the possible shape of the bones (*9*, *10*), but results have not always been satisfactory. In particular, as noted by Manos *et al.* the use of edge-based methods is limited by the fact that, in medical radiographs, boundaries are often weak and diffused, obscured by other tissues, and subject to spatial and biological variability.

In addition, the phase of feature analysis presents serious difficulties as well, because the shape of the bones has to be represented in a suitable way for evaluating the presence of salient characteristics. An approach proposed by some authors directly relates features which can be extracted more easily from a segmented image (for example, the perimeter and the area of a given bone) to the chronological age (*11–13*). However, the effectiveness of such features for the assessment of maturity has to be proved.

Unfortunately, even if many studies have focused on specific aspects of the problem, a complete and accurate system for the evaluation of skeletal maturity, reliable enough to be used in clinical practice, is not currently available.

In this paper we propose a radically different approach to the problem of automating the Tanner and Whitehouse method, and present a neural network-based system for the assessment of bone maturity. Following connectionist methodologies, knowledge about the typical shapes of bones during the various maturation stages is stored in a set of neural networks, so that it is not necessary to

explicitly extract and describe the salient features of the bones. In this way, the phase of bone segmentation can be avoided. As a result, our approach overcomes one of the major problems of other computational methods.

This paper is organized as follows: in section 2 we briefly summarize the most common methods for evaluating skeletal maturity, with particular reference to the Tanner and Whitehouse approach. The proposed approach is described in section 3, and the neural network implementation, along with the experimental results, are shown in section 4. Finally, conclusions are drawn in section 5.

## 2. The Problem of Skeletal Maturity Assessment and the TW2 Method

The evaluation of skeletal maturity is based on the presence and shape of some calcified regions which significantly characterize the development of the skeleton. In nearly a century of research, a large number of methods have been developed which operate on the radiographic analysis of several parts of the body.

In the past, two lines of research have been pursued based on either

- the number of ossification centers present in the skeleton (*chronological methods*) or
- the maturation stages of a number of ossification centers in terms of shapes of the corresponding bones (*qualitative methods*).

Only methods belonging to the second group are commonly in use, as they require the acquisition of a small number of radiograms. However, some methods of the first class are more reliable in special cases, such as for children younger than 2 years old.

Qualitative methods are based on the analysis of the evolution of the shape of short bones and the epiphyses of long bones. For each bone, a set of *maturity indicators*, i.e., features and characteristics of the shape of the bone, which "tend to occur regularly and in a definite and irriversible order" (*5*), have been determined. The evaluation of maturity requires the identification of specific discrete stages in the continuous process of growth. The most widely applied methods operate on radiograms of the left hand and wrist, such as the one illustrated in Fig. 1.

The atlas of Greulich and Pyle (*5*) is probably the most common method and is based on previous studies by Todd (*1*). It is a collection of radiograms chosen on a statistical basis as prototypical radiographs, where all the bones show a skeletal maturation stage which is typical for a given age. For each bone, a specific maturity indicator is associated to each chronological age.

With such an atlas the examination implies the comparision of the radiogram whose maturity has to be assessed, with the images of the atlas, in order to find the one which best matches the input image. The age associated to the prototypical radiogram selected is the estimated *skeletal age*, that is the chronological age at which subjects usually present that specific ossification pattern.

The major problem with the Greulich and Pyle method stems from the fact that the bones of the hand, even if they can often be considered as a single entity, can have different maturity stages, which makes the comparison with a

FIG. 1. Typical radiological image of the left hand and wrist used for evaluating skeletal maturity.

single pattern extremely difficult and arbitrary. That is, a degree of dismaturity within the hand and wrist can easily occur, and some bones may well be ahead of others even in normal healthy children. A second problem results from the intrinsic accuracy of the atlas which is not high.

The technique proposed by Tanner and Whitehouse (*3*), overcomes all the major problems connected with the atlas methods. The single match operation is replaced by a multiple matches procedure, which determines the maturation stage for each bone of a specific set. The average level of maturation is then converted into a corresponding skeletal age. In this way, the maturation stage of each bone contributes separately to the final estimate.

According to Tanner and Whitehouse a set of twenty bones of the left hand and wrist can be used (TW2 method). The maturation of each bone has to be evaluated by classifying the bone as belonging to one of eight (nine for the radius) classes, usually labeled with letters A to I. Figure 2 shows the maturity indicators for the hamate.

In a second phase, the maturity indicator of each bone is converted into a corresponding numerical factor by means of suitable correspondence (look-up)

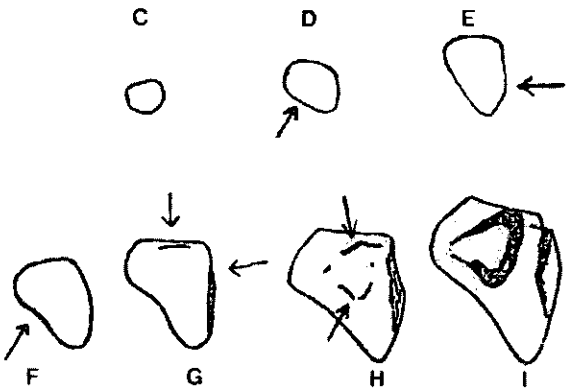


FIG. 2. Sequence of maturity indicators for the hamate.

tables. These scores have been evaluated through optimality criteria based on the statistical analysis of the maturation process.

The sum of all the bone scores is used as a pointer in another look-up table which gives the skeletal age. This final estimate depends to a different degree on all the selected bones of the hand.

In addition to the classical TW2 method, which makes use of all 20 selected bones, other reduced versions are available. By using different conversion tables, estimates of the skeletal age can be carried out only on the basis of the seven bones of the carpus (Carpal bone maturity) as well as on the radius, ulna, and short bones (RUS maturity). For each bone, the procedure for evaluating the maturity indicators is unchanged, only different look-up tables are used. All of these versions are expected to produce equivalent estimates of skeletal age, though the complete TW2 is usually considered more accurate due to its redundance.

Since TW2 is a method for evaluating the maturation stage and not directly the skeletal age, it can be easily adapted to different growth rates. For example, it can be applied both to males and females simply making use of different look-up tables. In general, even if the age associated with a particular maturation stage can change for different populations, the sequence of maturity indicators is always strictly followed. Thus, the extension of the method to other populations requires only the production of suitable conversion tables. For istance, the adaptation of TW2 to the Italian population is provided by Nicoletti *et al.* (*14*).

The evaluation of skeletal maturity by means of a single numerical factor provides a measurement which is extremely convenient for statistical analysis. It is worth noting that differences between the chronological and skeletal ages up to two years could still be considered normal. For example, 68% of subjects of 8 years of age have a skeletal age which ranges from 7 to 9, and they are considered healthy even if the skeletal age is in the range of 6–10. However, many different situations are possible depending on the pattern of maturity

indicators, and the actual interpretation should be made by physicians on the basis of their experience.

## 3. Automated Estimate of Skeletal Maturity: A Subsymbolic Approach

The main difficulties in developing an automated system which implements the TW2 method are related to the analysis of the maturity indicators.

To illustrate the problems which need to be solved, in Fig. 3 we show some images of different maturation stages of the hamate, as well as some images of the same bone at a given stage in different patients. As it can be noted, the main differences among classes are due to modifications of shape and cannot be attributed simply to changes of scale. Also, due to the intrinsic nature of the growth process, the organization into a set of discrete classes is strongly artificial. A continuum of intermediate configurations exists which makes the classes fuzzy.

In addition, radiograms are affected by several sources of noise: statistical noise is generated for example by the imaging chain. Troublesome structural noise is present, due in particular to the partial overlapping which is usually present among the various structures under examination, as well as to soft tissue shadows, which are characterized by a nonuniform gray-level distribution.

All these considerations suggest that very robust processing methods are needed for a reliable computer implementation of the TW2 method.

### *3.1. Computational Architecture*

As pointed out in the previous sections, the computational implementation of the TW2 method imposes severe design requirements. In particular, we have tried to meet the following global constraints:

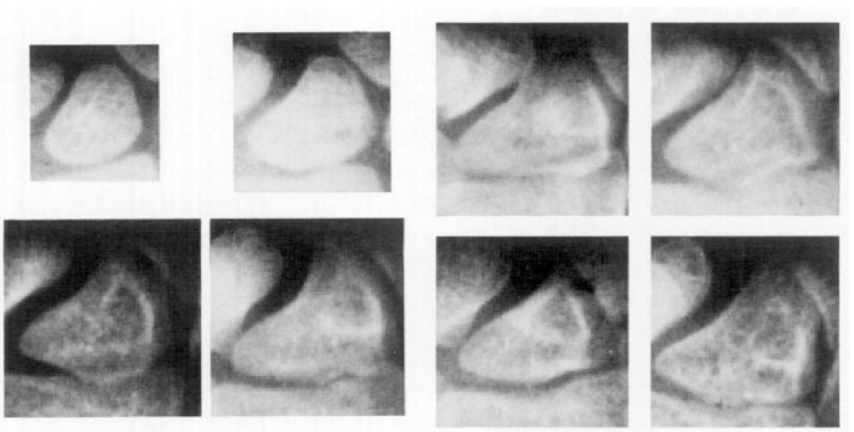

Fig. 3. Different maturation stages of the hamate. (Left) classes E, F, G, and H. (Right) different patients at level H.

1. The system architecture should be flexible in order to allow the implementation of the different versions of TW2 method.
2. The acquisition of expert knowledge should take limited time and effort.
3. The system should be very robust producing classification results comparable to those of trained operators.

The use of a modular approach based on separate processes for each bone under consideration, integrated with an adequate control strategy, is probably the simplest way to meet item 1.

Items 2 and 3 constrain the basic structure of the various subsystems as well as the overall design scheme. Actually, knowledge acquisition is usually very cumbersome and is one of the major drawbacks of symbolic knowledge-based approaches. Furthermore, the knowledge involved in hand-bone classification is fuzzy and depends on visual perception. For these reasons building a symbolic knowledge-base about the shape of hand bones during the growing process is hard, time consuming, and is expected to exhibit a large interobserver variability.

On the basis of previous experience about the use of artificial neural networks (ANN) for visual tasks (*15*), we have decided to adopt a subsymbolic approach for the assessment of skeletal maturity. The typical properties of ANN-based systems have been widely documented by many authors (see, for example, (*16*)). In particular the properties of learning from examples and producing fuzzy representations, which are typical of ANN, are well suited for complex tasks like those concerning visual perception (*17*). Trainable systems are adequate for the acquisition of the *a priori* knowledge involved in medical-image understanding. The expert has only to indicate and/or properly label the relevant features in the pictures composing the training set. Since no explicit formulation of expert knowledge is needed, this task can usually be achieved with a reasonable effort using a suitable graphic interface.

It must be also pointed out that the learning procedure results in the adaption of the system which, after convergence, stores, in a distributed fashion, the prototypical knowledge embedded in the training set in its connection weights. As pointed out by Hinton *et al.* (*18*) distributed representations are often more flexible, precise, and reliable as compared to local ones which are characterized by a one-to-one correspondence between concepts and representation elements.

Due to their intrinsic operation mechanisms, neural networks are expected to provide an optimal use of acquired knowledge: for a given input stimulus, the output is usually obtained by an efficient integration of the data stored in the connection weights (which can be seen as a sort of knowledge base) with the actual input. It is also worth noting that a strong noise immunity is typical of neural systems. This is crucial to ensure an adequate system reliability.

On the other hand, neural networks usually impose a huge computational load, and training algorithms often require long convergence times. However, the use of a proper architecture can dramatically reduce the network sizes (*19*, *15*). This usually produces also a considerable reduction of learning time.

From a mathematical point of view, a supervised neural network can be seen as implementing a mapping $\mathbf{c} = \mathbf{f_W}(\mathbf{d})$ between two different vector spaces $D$ and $C$. This mapping depends on a set of parameters $\mathbf{W} = (w_1, \ldots, w_K)$ which are the weights of the network connections. That is,

$$\mathbf{f_W} : D \subset R^n \text{ R } C \subset R^m \qquad [1]$$

where $R^n$ and $R^m$ are the $n$-dimensional and $m$-dimensional Euclidean spaces, respectively. Given a suitable training set $T$ composed by input/output pairs (**d**, **c**) where $\mathbf{d} \in D$ and $\mathbf{c} \in C$, learning can occur by modifying the weights **W** on the basis of $T$ according to specific optimality criteria.

In our case, **d** is a radiograph and **c** is the corresponding vector of bone maturity indicators $(c_1, \ldots, c_M)$ where $M$ is the number of bones considered (20 in the full version of the TW2). It should be noted that according to the TW2 method, each $c_i$ can only have a discrete value (the 9 classes of the method), whereas in the proposed neural network-based implementation they range in $[0, 1]$. Thus, the problem of estimating skeletal maturity can be formulated as the determination of vector **c** in

$$\mathbf{f(d)} = \begin{pmatrix} f_1(\mathbf{d}) \\ f_2(\mathbf{d}) \\ \vdots \\ f_M(\mathbf{d}) \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_M \end{pmatrix},$$

where each function $f_i$ estimates the maturation stage of the $i$-th bone (the dependence on the weight vector **W** has been dropped for notational convenience). Actually, $f_i$ needs to process only a small subset of the data of the whole image. Thus it can be assumed that it operates on a subset $\mathbf{d}_i$ of the image **d**

$$\begin{array}{l} f_1(\mathbf{d}_1) \text{ R } c_1, \\ f_2(\mathbf{d}_2) \text{ R } c_2, \\ \quad \vdots \quad \vdots \quad \vdots \\ f_M(\mathbf{d}_M) \text{ R } c_M. \end{array} \qquad [2]$$

This has led us to adopt the system architecture which is schematized in Fig. 4. It includes two main subsystems: the attention focuser (AF), which, by localizing the bones, creates the subsets $\mathbf{d}_i$, and the bone classifier (BC) which implements the mappings $f_i$. Basically, the BC subsystem includes as many ANN-based modules (called classifier modules (CMs)) as the bones considered: each CM classifies a given bone by estimating the corresponding maturity indicator. The CMs operate on the portions of the input radiograph generated by the AF, which locates the various bones and automatically extracts $M$ windows, each one centered on one of the bones considered (regions of interest (ROI)).

It must be pointed out that attention focusing is performed by processing the entire image at a reduced spatial resolution. In fact, only large features of the

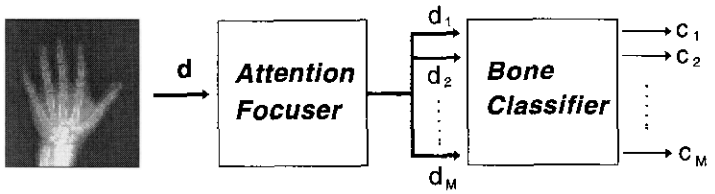


FIG. 4. The proposed system architecture includes an attention focuser which selects the region of interest of the input radiograph and a bone classifier which produces the estimates of the maturity indicators.

bone (those in which low spatial frequency components are dominant) must be considered. Small details are easily corrupted by noise, are heavily affected by biological variability, and can be misleading. This allows a rather simple topology of the AF as detailed in the next subsection. On the contrary, classification is carried out at a higher spatial resolution with respect to attention focusing, by processing a limited area of the radiograph. Obviously, small details can considerably affect the classification task and cannot be neglected in this phase. It should be clear that this strategy allows the optimization of the overall system performances, since each CM has only to process those patterns for which it has been implemented. Furthermore, it is possible to reduce the computational burden, because there is no need for highly complex networks.

A further consideration is needed about bone classification. As discussed in the introductory sections, classification must be primarily based on bone shape, rather than on the gray-level distribution of the radiographs. The use of a band-pass filtering, such as smoothed derivative operators, which removes the continuous component, as well as the low-frequency background variability, is useful for enhancing bone boundaries. Therefore, for each ROI, we compute the magnitude of the gradient of the original picture.

### *3.2. The Attention Focuser Subsystem*

The AF (see Fig. 5) includes an input section we have called retina and a set of focuser networks, one for each bone to be analyzed. The retina operates as a two layer net: the first layer has as many linear units as the image pixels ($X \times Y$) connected to the units of the second layer which includes $M \times N$ units, with $M \ll X$ and $N \ll Y$. The weights of the connections between the two layers are spatially arranged into overlapping, isotropic receptive fields, the profile of each field being a bidimensional Gaussian distribution $G(x, y)$ centered on each output unit, with

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \qquad [3]$$

The output of the retina is a decimated version, with a Gaussian smoothing, of the input radiograph with $\sigma = k \max(X/M, Y/N)$, where $k$ is a constant which

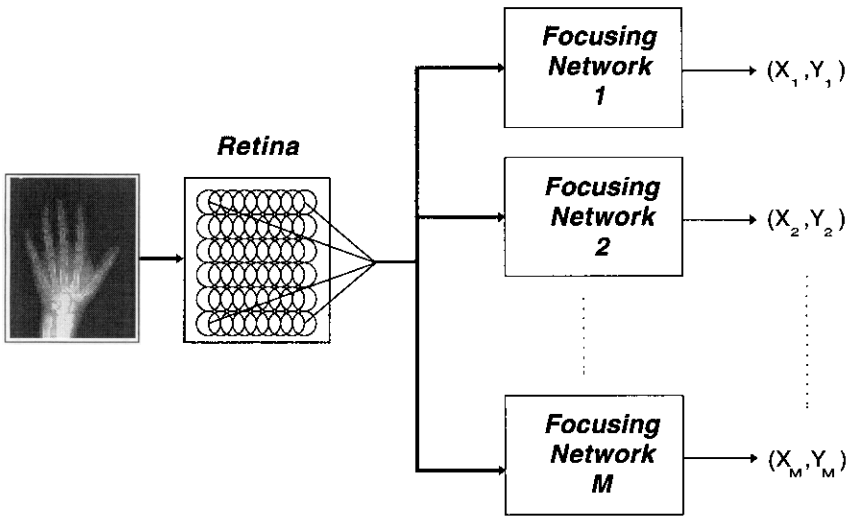


FIG. 5. The attention focuser locates the position of each bone in the input radiograph by estimating the coordinates of the bone centroid.

determines the degree of overlapping of adjacent receptive fields. This image is the input for $i$-th focusing network which provides the coordinates $(X_i, Y_i)$ of the centroid of the $i$-th bone. Each of these nets (see Fig. 6) is feed-forward, block-connected, and has four layers, with $M \times N$ input units, two hidden layers, and $M + N + 2$ output units. The number of units of the first hidden layer is determined experimentally, while the last two layers are split into two paths, one for each coordinate, with $M + 1$ units for $X_i$, and $N + 1$ units for $Y_i$, respectively. The values of $X_i$ and $Y_i$ are represented by means of the following sparse code:

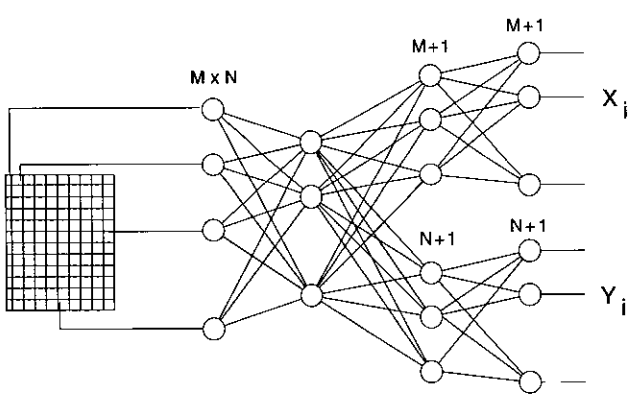


FIG. 6. The topology of the networks included in the attention focuser.

$$x(k) = \begin{cases} \dfrac{(k+1)X/M - X_i}{X/M} & \text{if } k = \left[ X_i \Big/ \dfrac{X}{M} \right] \\ \dfrac{X_i - (k-1)X/M}{X/M} & \text{if } k = \left[ X_i \Big/ \dfrac{X}{M} \right] + 1 \\ 0 & \text{elsewhere} \end{cases} \quad [4]$$

$$y(k) = \begin{cases} \dfrac{(k+1)Y/N - Y_i}{Y/N} & \text{if } k = \left[ Y_i \Big/ \dfrac{Y}{N} \right] \\ \dfrac{Y_i - (k-1)Y/N}{Y/N} & \text{if } k = \left[ Y_i \Big/ \dfrac{Y}{N} \right] + 1, \\ 0 & \text{elsewhere} \end{cases} \quad [5]$$

where $x(i)$ is the output value of the $i$-th unit in the $X_i$ branch and $y(j)$ is the output value of the $j$-th unit in the $Y_i$ branch (the operator $[z]$ gives the maximum integer $p$ such that $p \leq z$). In this way, $x(i)$ where $(x = 0, \ldots, M)$ is maximum when $X_i$ is equal to $i(X/M)$, while two adjacent units code intermediate positions. The same is true for $y(j)$ where $(y = 0, \ldots, N)$.

The output of the network can be decoded as follows. With reference to the $x$ coordinate, let $\hat{\imath}$ be the index of the output unit with maximum activation value and $\hat{\jmath}$ be the index of the adjacent unit with higher activation ($\hat{\jmath} \in \{\hat{\imath} - 1, \hat{\imath} + 1\}$). The decoded value $\tilde{X}_i$ is obtained as

$$\tilde{X}_i = \frac{x'_r + x''_r}{2}, \quad [6]$$

where

$$x'_r = \frac{X}{M}\{\hat{\imath} + (\hat{\jmath} - \hat{\imath})(1 - x(\hat{\imath}))\}$$
$$x''_r = \frac{X}{M}\{\hat{\jmath} - (\hat{\jmath} - \hat{\imath})(1 - x(\hat{\jmath}))\}. \quad [7]$$

Similar equations hold for the coordinate $y$. Once the centroid coordinates $(X_i, Y_i)$ have been computed, for each bone a predefined $a \times b$ window (the ROI) is extracted from the image. The dimensions of the ROIs are *a priori* set on the basis of the average size of the bone considered.

### *3.3 Bone Classifyer Subsystem*

The ROIs produced by the AF are processed by a dedicated CM. All CMs include an input section and a classifying network, as illustrated in Fig. 7 The input section is a retina with $a \times b$ input units and $m \times n$ output units and includes two different types of oriented receptive fields which compute the two components of a gradient of Gaussian; i.e., their profiles are

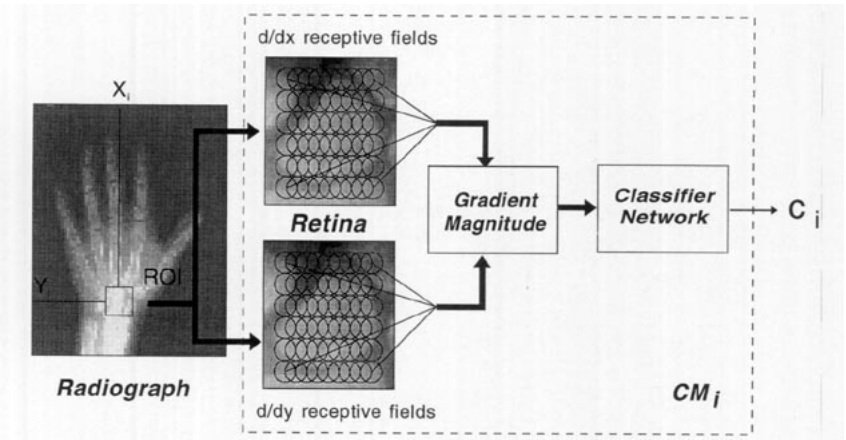


FIG. 7. Each classifier module processes the region of interest centered on the corresponding bone and evaluates the maturation stage $c_i$ of the bone by means of a neural network.

$$-x \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right), -y \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right), \quad [8]$$

where $\sigma = k \max(a/m, b/n)$. Afterwards, the magnitude of the gradient is computed by summing up the squared outputs of the two groups of receptive fields. Following this process, the gray levels are expanded to match the available dynamic range. In this way, intense edges and/or areas with a high edge density are strongly enhanced. A fixed number $m$ of output units is adopted. Since the size of the input ROI is variable, it results that large bones are processed with a lower spatial resolution as compared to small bones.

The classifying network is a four layer fully connected feed-forward net with $m \times n$ input units, two hidden layers whose number must be determined in the implementation stage, and a number of output units equal to the number of possible classes for the considered bone.

## 4. Implementation of the TW2 Carpal Bone Method

In order to test the performances which can be achieved with the proposed approach, we have implemented the TW2 carpal bone method which is based on the classification of seven bones of the carpus, namely the hamate, capitate, triquetral, lunate, scaphoid, trapezium, and the trapezoid. According to the described system architecture, we have designed and trained seven focusing networks and seven classifying networks, one for each considered bone. To this end, we have collected 72 hand radiographs; 56 of them were used to train the system and the remaining 16 composed the test set. Original radiographs were digitized by means of a laser scanner into 850 × 640 × 12 bit pictures, using a pixel size of 280 $\mu$m. All the radiographs were examined by an expert physician

who evaluated the maturity indicator of each bone and computed the skeletal age according to TW2 tables. The estimates performed by the specialist were used as reference values in training the system, and the performances were evaluated in comparison with them.

As concerns the AF subsystem, special effort was dedicated to minimizing the number of output units of the retina. We have found that with $M = 18$, $N = 14$, and $k = 0.33$ in the $\sigma$ value, the overall system performances are still satisfactory. All the focusing networks were implemented with an input layer composed by 252 ($18 \times 14$) units and 34 ($19 + 15$) output units. The first hidden layer was implemented with 18 units; this number was determined by experimenting with different network topologies and selecting the simplest one (that is the one with fewer units) (*19*). The ROIs generated by the AF were squares with a different number of pixels for each considered bone, and they ranged from $61 \times 61$ for the trapezoid (the smallest bone) to $95 \times 95$ for the capitate (the largest one), as illustrated in Table 1.

For each CM, the size of the receptive fields of the retinas changed according to the ROI dimension, by having a different value $\sigma$. These values were obtained so as to produce an output image with 400 pixel ($m = n = 20$). In this way, all the classifier networks included $20 \times 20$ units in the first layer. The unit number in the hidden layers was determined with similar considerations of the focusing networks, by selecting the minimal net-topology, which produces satisfactory results. The number of output units of each classifier network was equal to the number of maturity indicators available for the bone considered in the training set. The topology of the seven classifying networks is summarized in Table 2.

All the networks consisted of linear input units and hidden and output sigmoidal units. The system was software implemented with a proper simulation program written in C language. Supervised training was carried out by means of the back-propagation algorithm (*20*). Since our final goal is to develop a system

TABLE 1

STRUCTURE OF THE ATTENTION FOCUSER MODULES

| Bone | ROI | Network topology |
|---|---|---|
| Capitate | $95 \times 95$ | $18 \times 14 - 18 - 19 + 15 - 19 + 15$ |
| Hamate | $77 \times 77$ | $18 \times 14 - 18 - 19 + 15 - 19 + 15$ |
| Triquetral | $67 \times 67$ | $18 \times 14 - 18 - 19 + 15 - 19 + 15$ |
| Lunate | $65 \times 65$ | $18 \times 14 - 18 - 19 + 15 - 19 + 15$ |
| Scaphoid | $77 \times 77$ | $18 \times 14 - 18 - 19 + 15 - 19 + 15$ |
| Trapezium | $69 \times 69$ | $18 \times 14 - 18 - 19 + 15 - 19 + 15$ |
| Trapezoid | $61 \times 61$ | $18 \times 14 - 18 - 19 + 15 - 19 + 15$ |

*Note.* For each bone, the size (pixel) of the ROI used for processing and the topology of the related net are given. Each network includes $18 \times 24$ input units, a hidden layer with 18 units, and two output paths (one for each coordinate) having two layers with 19 and 15 units, respectively (see also Fig. 6).

TABLE 2

STRUCTURE OF THE BONE CLASSIFIER MODULES

| Bone | Network topology |
|---|---|
| Capitate | 20 × 20 − 10 − 5 − 9 |
| Hamate | 20 × 20 − 10 − 7 − 9 |
| Triquetral | 20 × 20 − 10 − 5 − 9 |
| Lunate | 20 × 20 − 10 − 7 − 9 |
| Scaphoid | 20 × 20 − 10 − 9 − 9 |
| Trapezium | 20 × 20 − 10 − 10 − 9 |
| Trapezoid | 20 × 20 − 10 − 9 − 9 |

*Note.* Each network includes 20 × 20 input units, a first hidden layer with 10 units, a second hidden layer with a variable number of units, and 9 output units.

which can be efficiently used in a medical environment, without the need for sophisticated and expensive hardware tools, we have run the programs on a low cost CPU 80486 clocked at 33 Mhz. Training times have been fairly long, with several hours of training for each network. The actual time required varied from bone to bone, depending also, for the classifier networks, on the number of maturity indicators present in the training set for each bone. However, once the system is ready and all the networks are trained, the assessment of skeletal maturity is very fast compared to the human operator; the evaluation of the maturation stage of each bone required less than 10 sec, most of which were spent for simulating the receptive fields. The global estimate of skeletal maturity with the carpal TW2 method can thus occur in less than a minute.

To verify the convergence of the training procedure, we preliminarily tested the overall system using the radiographs of the training set. The system was able to exactly replicate the reference values indicated by the physician. Then, the real behavior of the system was analyzed by means of the test images, not previously examined by the system. Table 3 shows the performance of the AF subsystem, measured by the MSQ error in locating the centroid of the related bone. It is worth noting that, even in the worst case, the ROI includes the analyzed bone completely, the average error being 7.5 pixel for the $X$ coordinate and 4.8 for the $Y$ coordinate.

Table 4 shows the results obtained by the BC subsystem for each classifier module on the images of the test set. The rate of correct classification is about 65% (classification (a)). However, if errors between contiguous classes are ignored, system performances reach 97% (classification (b)). The final estimate of skeletal maturity provided by the system for the patients of the training set is illustrated in Table 5, along with the estimates carried out by the physician. The average difference between the two values is 0.62 years, with a standard deviation equal to 0.58.

TABLE 3

PERFORMANCES OF THE ATTENTION FOCUSER MODULES (PIXEL UNITS)

| Bone | $X$ | $Y$ |
|---|---|---|
| Capitate | 8.8 ± 5.8 | 5.6 ± 4.9 |
| Hamate | 8.2 ± 5.9 | 4.5 ± 4.1 |
| Triquetral | 7.0 ± 6.1 | 5.5 ± 4.1 |
| Lunate | 8.0 ± 5.3 | 3.8 ± 3.5 |
| Scaphoid | 4.7 ± 4.9 | 4.2 ± 3.1 |
| Trapezium | 9.8 ± 7.5 | 7.5 ± 6.4 |
| Trapezoid | 6.3 ± 4.2 | 2.9 ± 2.2 |

*Note.* For each of the two coordinates ($X$, $Y$) of bone centroids, the MSQ positioning error and its standard deviation are given.

System misclassification seemed to be related to the intrinsic fuzziness of TW2 categories. Indeed, estimation disagreements are commonly encountered among expert physicians. To clarify this point, we show in Table 6 the results of some comparative experiments carried out by different experts. In two tests, the skeletal maturity has been estimated by different specialists on 10 and 8 radiographs, respectively. These classifications have been compared with those of a senior expert. As illustrated in Table 6, the two groups indicated the same classification of the supervisor (classification (a)) only in 67 and 56% of the cases, respectively. When classification errors between contiguous classes are disregarded (classification (b)), these values rose to 100 and 94%. These results suggest that the behavior of the proposed system is similar to that of trained operators.

TABLE 4

ESTIMATES OF THE MATURITY INDICATORS

| Bone | Classification (a) | Classification (b) |
|---|---|---|
| Capitate | 60 | 100 |
| Hamate | 75 | 100 |
| Triquetral | 80 | 100 |
| Lunate | 70 | 100 |
| Scaphoid | 60 | 95 |
| Trapezium | 40 | 95 |
| Trapezoid | 70 | 90 |

*Note.* For each bone, the rate of correct classification (a) and the rate of correct classification ignoring errors between contiguous classes (b) are shown (percentage values).

TABLE 5

ESTIMATES OF THE SKELETAL MATURITY (YEARS)

| Case No. | Reference age | Estimated age |
|---|---|---|
| 1 | 13.50 | 13.20 |
| 2 | 16.00 | 15.70 |
| 3 | 16.00 | 14.80 |
| 4 | 16.00 | 16.00 |
| 5 | 16.00 | 15.60 |
| 6 | 14.20 | 15.30 |
| 7 | 7.20 | 8.45 |
| 8 | 15.30 | 14.70 |
| 9 | 9.40 | 9.50 |
| 10 | 15.70 | 15.70 |
| 11 | 16.00 | 14.50 |
| 12 | 16.00 | 15.60 |
| 13 | 6.80 | 6.45 |
| 14 | 16.00 | 15.40 |
| 15 | 16.00 | 14.30 |
| 16 | 13.70 | 13.85 |

## 5. CONCLUSIONS

It is commonly agreed that the TW2 method produces estimates of skeletal maturity which are usually more accurate than other, widely used, methods. As explained in section 2, this is mainly due to the fact that the TW2 estimate of skeletal maturity is based on the shape features of a set of different bones rather than on a single, global evaluation of the radiogram, as in the case of atlas-based methods. The set of maturity indicators so obtained provides an accurate representation of the growth process, which permits one to deal with inhomogeneities in the maturation stages of the various bones. Unfortunately, this is also the source of the major limitations of this method. In fact, for each examination,

TABLE 6

RESULTS OF COMPARISONS BETWEEN HUMAN EXPERTS

| Test | Classification (a) | Classification (b) |
|---|---|---|
| A (10 subjects) | 67% | 100% |
| B (8 subjects) | 56% | 94% |

*Note.* For two groups of experts, we show in (a) the rate of correct classification (with respect to a senior expert) and in (b) the same rates obtained disregarding errors between contiguous classes.

the operator has to carefully classify a fairly large number of bones with long execution times. In addition, the assessment of the bone maturity indicators is a nontrivial task, and a specialist with considerable experience is needed to produce accurate results.

Thus, the automation of the TW2 method is highly desirable, since it would allow a much more extended use of the method and more accurate clinical examinations. The system could be useful also in the training of specialized personnel. Moreover it could provide a reference standard for statistical analysis.

However, the computational analysis of hand radiographs has proved to be very hard, due to the complexity and variability which is typical of the growth process. The various sources of noise that are present in the imaging chain, along with the interferences among the shadows of the projected bones contribute to making the task harder.

In recent years, artificial neural network techniques have been successfully applied to many computer vision problems. Their properties of robustness, noise-tolerance, and learning are very important for dealing with perceptual problems. In particular, learning paradigms are well suited for the approximation of complex multivariate functions. Clearly, this is relevant in those cases where samples of the function are available for discrete sets of the independent variables, but the function itself is difficult to formalize explicitly. This often happens in the analysis of medical images, where the physician is able to provide quantitative results, but the underlying perceptual and reasoning processes are in most cases unknown.

In this work, we have faced the problems involved in the computational analysis of hand radiographs and developed a neural network architecture for implementing the TW2 method. We have proposed a simple modular system which, in principle, can be applied to any version of the TW2 method. Preliminary results obtained with the carpal version of TW2 are extremely encouraging.

We are presently working on the implementation of the complete TW2 method based on twenty bones. We believe that, thanks to the higher redundancy of the data set, even more accurate results will be achieved, without a significant increase of the computational load.

## ACKNOWLEDGEMENTS

This work has been supported by the Italian Ministry of University and Scientific and Technological Research (MURST) and Italian Research Council (CNR).

## REFERENCES

1. TODD, T. "Atlas of Skeletal Maturation (Hand)." Mosby, St. Louis, 1937.
2. HELLMAN, M. Ossification of epiphysial cartilages in the hand. *Am. J. Phys. Anthropol.* **11,** 223–257 (1928).
3. TANNER, J., AND WHITEHOUSE, R. "Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method)." Academic Press, New York, 1975.
4. MILER, G., LEVICK, R., AND KAY, R. Assessment of bone age: A comparison of the Greulich and Pyle and the Tanner and Whitehouse methods. *Clin. Radiol.* **37,** 119–121 (1986).
5. GREULICH, W., AND PYLE, S. "Radiographic Atlas of Skeletal Development of the Hand and Wrist." Stanford Univ. Press, Stanford, 1950.

6. PAL S., AND KING, R. On edge detection of X-ray images using fuzzy sets. *IEEE Trans. Pattern Anal. Machine Intelligence.* **5**(1), 69–77 (1983).
7. PATHAK A., AND PAL, S. Fuzzy grammars in syntactic recognition of skeletal maturity from X-ray. *IEEE Trans. Syst. Man Cybernetics.* **16**(5), 657–667, 1986.
8. MANOS, G., CAIRNS, A., RICKETTS, I., AND SINCLAIR, D. Automatic segmentation of hand-wrist radiographs. *Image Vision Comput.* **11**(2), 110–111 (1993).
9. MICHAEL, D., AND NELSON, A. HANDX: A model-based system for automatic segmentation of bones from digital hand radiographs. *IEEE Trans. Med. Imaging.* **8,** 64–69 (1989).
10. EFFORD, N. Knowledge-based segmentation and feature analysis of hand and wrist radiographs. *Proc. SPIE-Int. Soc. Opt. Eng.* **1905,** 512–525 (1993).
11. PIETKA, E., MCNITT-GRAY, M., KUO, M., AND HUANG, H. Computer assisted phalangeal analysis in skeletal age assessment. *IEEE Trans. Med. Imaging* **10**(4), 616–620 (1991).
12. PIETKA, E., KAABI, L., KUO, M., AND HUANG, H. Feature extraction in carpal-bone analysis. *IEEE Trans. Med. Imaging.* **12**(1), 44–49 (1993).
13. SUN, Y., KO, C., MAO, C., AND LIN, C. A computer system for skeletal growth measurement. *Comp. Biomed. Res.* **27,** 2–12 (1994).
14. NICOLETTI, I. "Crescita e Maturazione Scheletrica." Ed. Centro Studi Auxologici, Firenze, Italy, 1990.
15. COPPINI, G., POLI, R., RUCCI, M., AND VALLI, G. A neural network architecture for understanding discrete 3d scenes in medical imaging. *Comp. Biomed. Res.* **25,** 569–585 (1992).
16. RUMELHART D., AND MCCLELLAND, J. "Parallel Distributed Processing." MIT Press, Cambridge, 1986.
17. CARPENTER G., AND GROSSBERG, S. "Neural Networks for Vision and Image Processing." MIT Press, Cambridge, 1992.
18. HINTON, G., MCCLELLAND, J., AND RUMELHART, D. Distributed representations. *In* "Parallel Distributed Processing. "MIT Press, Cambridge, 1986.
19. MAREN, A., HARSTON, C., AND PAP, R. "Handbook of Neural Computing Applications." Academic Press, San Diego, 1990.
20. RUMELHART, D., HINTON, G., AND WILLIAMS, R. Learning internal representations by error propagation. *In* "Parallel Distributed Processing." MIT Press, Cambridge, 1986.