Assessing the Bone Age of Children in an Automatic Manner Newborn to 18 Years Range

Farzaneh Dehghani[1] · Alireza Karimian[1] · Mehri Sirous[2]

Abstract

Bone age assessment (BAA) is a radiological process to identify the growth disorders in children. Although this is a frequent task for radiologists, it is cumbersome. The objective of this study is to assess the bone age of children from newborn to 18 years old in an automatic manner through computer vision methods including histogram of oriented gradients (HOG), local binary pattern (LBP), and scale invariant feature transform (SIFT). Here, 442 left-hand radiographs are applied from the University of Southern California (USC) hand atlas. In this experiment, for the first time, HOG–LBP–dense SIFT features with background subtraction are applied to assess the bone age of the subject group. For this purpose, features are extracted from the carpal and epiphyseal regions of interest (ROIs). The SVM and 5-fold cross-validation are used for classification. The accuracy of female radiographs is 73.88% and of the male is 68.63%. The mean absolute error is 0.5 years for both genders' radiographs. The accuracy a within 1-year range is 95.32% for female and 96.51% for male radiographs. The accuracy within a 2-year range is 100% and 99.41% for female and male radiographs, respectively. The Cohen's kappa statistical test reveals that this proposed approach, *Cohen's kappa coefficients are 0.71 for female and 0.66 for male radiographs*, p value < 0.05, is in substantial agreement with the bone age assessed by experienced radiologists within the USC dataset. This approach is robust and easy to implement, thus, qualified for computer-aided diagnosis (CAD). The reduced processing time and number of ROIs facilitate BAA.

Keywords Bone age assessment (BAA) · Left-hand radiographic image · Computer vision operators · Computer-aided diagnosis (CAD) · Carpal and epiphyseal regions of interest (ROIs) · Support vector machine

Introduction

The process of bone growth indicates bone age. Therefore, in general, bone age assessment (BAA) is of great importance and

✉ Alireza Karimian
  karimian@eng.ui.ac.ir

  Farzaneh Dehghani
  f.dehghani@eng.ui.ac.ir

  Mehri Sirous
  Sirous@med.MUI.ac.ir

[1] Department of Biomedical Engineering, Faculty of Engineering, University of Isfahan, Isfahan, Iran

[2] Department of Radiology, AL-Zahra Hospital, Isfahan University of Medical Science, Isfahan, Iran

can precisely diagnose growth disorders, hormonal, and genetic problems in children. Bone age is typically assessed by measuring the maturity of bones in children through left-hand and wrist radiographs. A pediatrician will make the final decision based on the disparity between the estimated bone age and chronological age. The BAA is applied to other fields like judgment in some criminal investigations and forensic medicine [1].

In general, there exist two basic methods for clinical assessment of bone age which are the Greulich and Pyle (GP) method [2] and Tanner and Whitehouse (TW) method [3]. The GP method is an atlas-based method against which the patient's radiograph is compared. The pediatrician then identifies the most corresponding radiograph to the patient's radiograph and reports an estimate of bone age based on the evidence. The TW method is a scoring method, where a score is given to the maturation of the patient's bones. Eventually, the pediatrician translates the calculated score to the bone age by applying the standard table. According to the available reports, the TW method is time-consuming, unbearable, and complex; therefore, 76% of pediatricians prefer to estimate bone age by the

GP method [4]. Based on previous studies, the average time needed to assess bone age by GP and TW methods is 1.4 and 7.9 min, respectively [5]. The manual BAA depends on the pediatricians' skill and is not accurate for either inter-observation or intra-observation.

In recent years, researchers focusing on BAA have revealed an increasing interest in applying computer-based methods. In 2001, Pietka et al. presented a method to extract features from epiphyseal/metaphyseal regions, with the objective of measuring the gap between metaphysis and diaphysis. In their study, 540 radiographic images were assessed from the University of Southern California (USC) database within a 0–18years range. The extracted features of these regions clearly exhibit the bone maturity stages [6].

In 2012, Fischer et al. assessed the bone age of children within the 0 to 18years range through content-based image retrieval. In their study, where USC radiographs are applied to extract features from 14 epiphyseal regions of interest (eROIs), a mean absolute error (MAE) of 0.99 years and a standard deviation of 0.76 years are obtained in comparison with the mean USC–BAA [7].

In 2016, Kashif et al. run a similar study on BAA, where five computer vision methods including SIFT, SURF, BRIEF, BRISK, and FREAK are applied to extract features from 14 eROIs. Their suggested methods are evaluated by SVM and 5-fold cross-validation. The best results are obtained by dense SIFT feature extraction with 45.69% accuracy, 0.605 years MAE; 89.38% accuracy within the range of 1 year; and 98.36% within the range of 2 years [8].

In 2017, Lee et al. applied a fully automatic deep-learning system for BAA within a 5 to 18years range. Their models used an ImageNet pretrained, fine-tuned convolutional neural network (CNN) to achieve 57.32% and 61.40% accuracies for female and male test radiographs, respectively. In their study, female test radiographs were assigned a BAA within the range of 1 year 90.39% and within the range of 2 years 98.11% of the time. Male test radiographs are assigned 94.18% within 1 year and 99.00% within 2 years in their study [9].

In 2017, Larson et al. compared the performance of a deep-learning neural network model for BAA with that of expert radiologists and the existing automated models. The data set, accumulated from Lucile Packard Children's Hospital at Stanford University and Children's Hospital in Colorado, applied there is composed of left-hand radiographs. These images are interpreted by pediatric radiologists according to the GP atlas. Their data set is divided into two separate subsets of training and validation. Ninety percent of the total data is applied for training and 10% is applied for validation which is to tune the hyper-parameters of the model. In their study, 200 hand radiographs are applied to assess the performance of the model in relation to that of radiologists. They applied USC radiographs to assess the performance of their model in relation to that of existing automated software. The mean difference between the performance of their model and the mean of the radiologists' estimates in bone age is 0 years. There the root mean square (RMS) is 0.63 years and the mean absolute difference (MAD) 0.50 years. When applied to the USC data set, their model is assigned an RMS of 0.73 years [10].

In 2018, Halabi et al. briefed the results of the Radiological Society of North America (RSNA) Pediatric Bone Age Machine Learning Challenge, where 90 % of the total data set is applied in training and 10% is applied in validation. Their data is accumulated from Lucile Packard Children's Hospital at Stanford University and Children's Hospital in Colorado. A separate test set containing 200 radiographs from Lucile Packard Children's Hospital is applied to assess the performance of the given algorithms. In general, the best result of this challenge based on MAD is 4.2 months [11].

In this study, the histogram of oriented gradients (HOG) [12], local binary pattern (LBP) [13], and scale invariant feature transform (SIFT) [14] are applied for BAA. These methods are applied in feature extraction of carpal ROI and eROI (which belongs to the epiphyseal center of proximal phalanx) (Fig. 1). The HOG, LBP, and dense SIFT features are concatenated in order to improve the BAA. Before selecting ROIs, several hierarchical stages are observed on radiographic images. For evaluating the results, SVM and 5-fold cross-validation are applied as the classification process. According to GP atlas, bone evolution rate is significantly higher in females than in males; consequently, this study is run on both genders' radiographs in a separate manner.
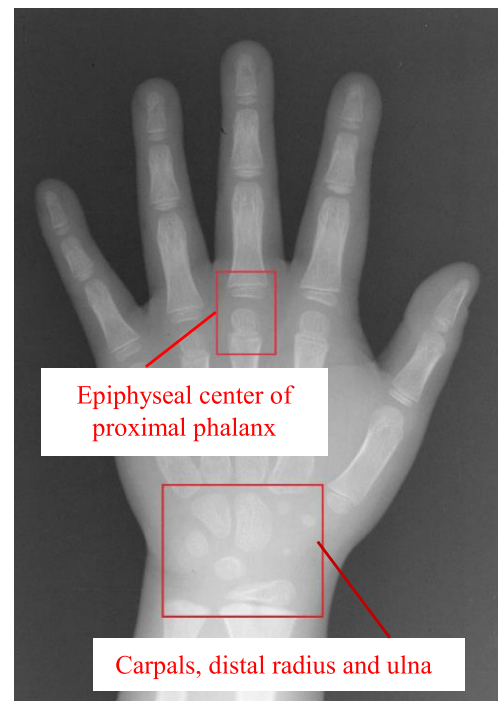

Fig. 1 Carpal and epiphyseal regions of interest

Materials and Methods

The method adopted in this study is analytic.

The computer-aided diagnosis (CAD) bone age chain here consists of : pre-processing, ROI extraction, background subtraction, feature extraction, feature selection, and classification (Fig. 2).

In order to implement the above stages, the ROIs are extracted from the carpal area and epiphyseal center of proximal phalanx after pre-processing, followed by applying background subtraction on ROIs where features are extracted through HOG, LBP, and dense SIFT. The HOG, LBP, and dense SIFT features are concatenated to yield better features as to BAA. First, the dimensions of features are reduced through statistical feature selection methods, and next, data are classified.

USC hand atlas radiographs, proposed by Gertych et al. [15], are applied here. All images of this database consist of two reports from two radiologists. In this study, 442 radiographs (220 radiographs for females and 222 radiographs for males) consisting of four ethnic groups of Asian, African American, Caucasian, and Hispanic origins within the 0 to 18years range are applied. All of the selected radiographs are authorized by an experienced radiologist, with 30 years of experience, in Isfahan University of Medical Sciences.

The implemented stages are the following:

Pre-processing

This stage consists of noise reduction and contrast enhancement. For the purpose of noise reduction of x-ray images, anisotropic diffusion filtration is applied, which is capable of reducing the noise of homogeneous areas while maintaining the contrast and the edges of the image in a simultaneous manner [16, 17]. Here, homomorphic filtrating is applied to normalize the brightness of the images and increase their contrast [17, 18].

ROI Extraction

Two ROIs including carpal ROI and eROI are extracted here. The carpal area which consists of carpal bones, distal radius, and ulna is contributive in BAA because these bones contain discriminative features for assessing the bone age of young children. However, at older ages, they do not have desirable performance in a single manner. According to Kashif et al., epiphyseal centers which belong to the proximal phalanges are more reliable than the other epiphyseal regions [8]. Therefore, here, one eROI which belongs to proximal phalanx is selected out of five in order to reduce the processing time of the proposed approach. These two ROIs are applied in the background subtraction stage.

Background Subtraction

In order to increase the performance of this approach, the background subtraction is made through top-hat transform operation. One of the main applications here is to remove objects from the image by applying a structural element. This approach is applied to correct the non-uniform brightness of images. Here, a disk-shaped structuring element with a special radius is applied in order to subtract background from the images (extracted ROIs), where a uniform background and only the bones are exposed. The output of this stage is applied for the purpose of feature extraction.

Feature Extraction

In this article, the three computer vision methods HOG, LBP, and dense SIFT are applied in order to extract features from the obtained ROIs. HOG and LBP are known as global feature extraction techniques, with the objective of object detection. Object detection deals with detecting instances of semantic objects of a certain class in digital images. Descriptions like HOG are applied for this purpose. According to Wang, the
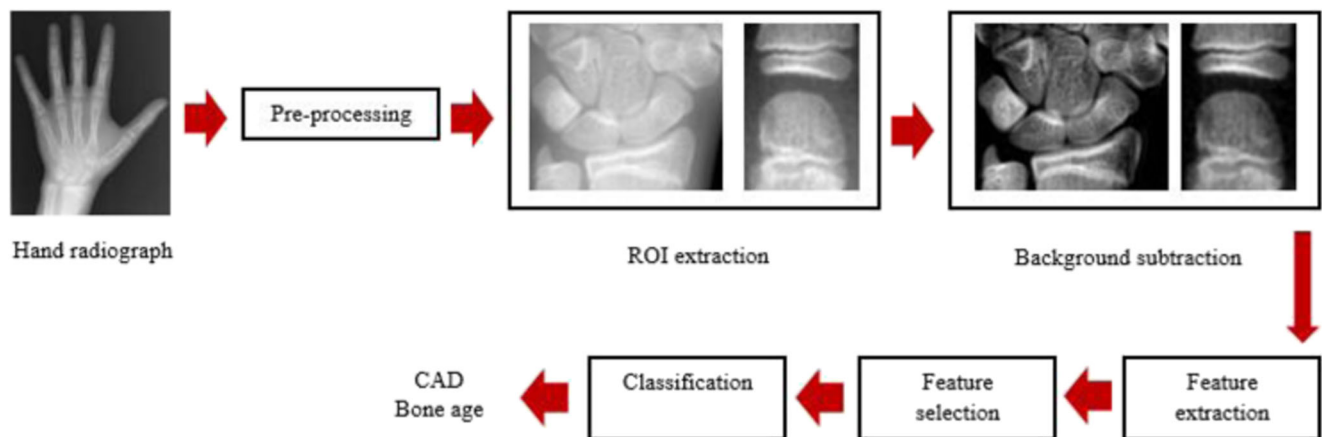

Fig. 2 Flowchart of the proposed approach

concatenation of HOG and LBP is of a good performance in some studies (i.e., human detection and face detection) [19]; therefore, this concatenation is applied for BAA in this study. SIFT is known as a local feature extraction technique, with the objective of object recognition. Object recognition deals with identifying objects in an image through their identities like size or scale.

In general, there exist no bones in the carpal area at birth while, upon growth, they appear. The number of carpal bones in children older than 5 to 7 years will be completed; therefore, only the size of the bones is considered as bone maturation index in these age ranges. Hence, the HOG–LBP feature extraction method is applied as an object detection in young children, and SIFT feature description method is added to extract more accurate features for BAA. In this context, the radiograph of all age ranges are examined in a simultaneous manner; therefore, the concatenation of HOG, LBP, and SIFT is applied here. The construction of each method is investigated in more details in the following sections.

Scale Invariant Feature Transform

SIFT extracts distinctive local features, which are used to match objects in different images. The feature extraction procedure of SIFT consists of two major steps which are feature detection and feature description. In feature detection, an algorithm determines the *keypoints* which represent the most informative parts of the image. In feature description, a local descriptor is computed for each *keypoint*.

Keypoints (feature points) are image points which are extremum in the scale space of an image. The scale space of an image is yielded from the convolution of a Gaussian function with an input image. This process is performed several times by changing the scale of the Gaussian function. The scale space is calculated through Eq. (1):

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y, \sigma) \tag{1}$$

where L, G, I, σ, and $*$ are the scale space of an image, variable-scale Gaussian function, input image, width of the Gaussian, and convolution operation in x and y, respectively. $G(x, y, \sigma)$ is calculated through Eq. (2):

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2 + y^2)/2\sigma^2} \tag{2}$$

The difference between each two nearby scales is computed in order to acquire the difference-of-Gaussian function. The extremum of this function is achieved by comparing a pixel with its 8 neighbors at the current scale and 18 neighbors at the adjacent scales in 3×3 regions.

In this method, a feature vector, which consists of 128 components, is computed for each *keypoint*. The window around each *keypoint* is divided into 4×4 sub-windows. The histogram of gradients, which contains eight values, is calculated for each sub-window. The feature vector size of the *keypoint* is $4 \times 4 \times 8 = 128$. Consequently, the output dimension of SIFT with n keypoints is $n \times 128$. *Keypoints* can be selected in a sparse or dense manner [14].

Histogram of Oriented Gradients

HOG is a feature descriptor which applies the distribution of local gradients or edge directions to describe an image. This method is reliable in explaining an image even if there is no accurate information about the exact position of the gradients or the orientation of edges. In this method, first, the image window is divided into small spatial regions names *Cell*, and next, the direction of image gradients is specified for each *Cell*. Ultimately, the histogram of gradients directions is calculated for pixels of the *Cell*. The calculated histograms are combined together in order to represent the feature vector of HOG [12].

Local Binary Pattern

LBP is one of the best descriptors for texture description [13]. This operator is capable to determine microstructures like edges, lines, spots, and flat areas in the image [20, 21]. In this descriptor, an input image is divided into smaller regions of *Cell*. Each pixel is compared with its 8 neighbors in this *Cell*. To be more specific, the value of the central pixel is considered as the threshold (Fig. 3) where, first, it is revealed that whether the intensity value of each binary pixel is 1, higher than the central pixel and 0, or lower than central pixel, and
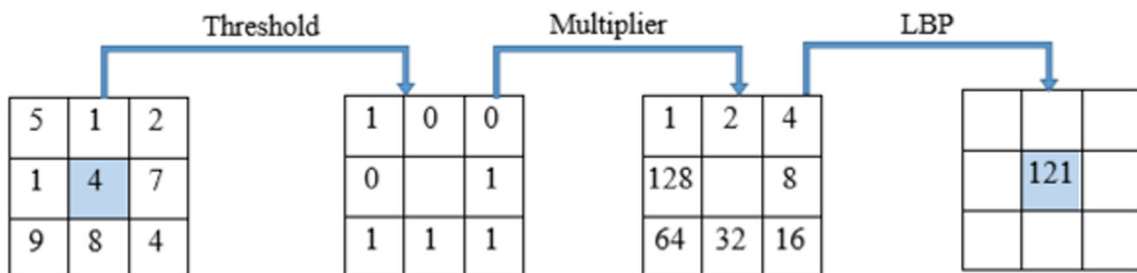

Fig. 3 An example to find the binary labels of eight neighbors of a pixel, where LBP is $1 + 8 + 16 + 32 + 64 = 121$

next, the histogram is calculated for each *Cell*. The calculated histograms are accumulated and the feature vector of LBP is yield [13].

Feature Selection

This process is of two steps: a statistical feature selection (SFS) and correlation-based feature selection (CFS). In order to select appropriate features, the assumption of normality or non-normality of features must be assessed. The one sample Kolmogorov–Smirnov test (K-S test) is run for this purpose. According to the results obtained through the K-S test, because the extracted features from both female and male radiographs are of abnormal distribution, the Kruskal–Wallis SFS is run to select appropriate features. A p value is assigned for each feature and those with p value ≤ 0.05 are selected. The correlation of selected features is examined through CFS because Kruskal-Wallis is a univariate SFS (the features of which are examined separately). At this step, the correlation of features is calculated pairwise among all, and when they are highly correlated, one of the features is removed.

Classification

The support vector machines are well known as powerful classifiers, and are binary and applicable for the purpose of two-class classification; consequently, the existing methods suggested for extending SVMs of multi-classification become necessary [22]. According to Hsu et al., the one-against-one and one-against-all methods are reliable for the purpose of SVM multi-classification [23].

Validation Experiments

The feature vectors' size of three computer vision methods applied here is related to the image size. These vectors should be of the same size for the purpose of classification; therefore, all the ROIs extracted from radiographs are rescaled to 48×48 pixels. This size is recognized as the best as to time of feature extraction and accuracy of BAA aspects. For this classification, SVM is executed with a radial basis kernel function (RBF) and one-against-all approach. Twenty percent of the total data is selected on a random basis as a validation set in order to tune the hyper-parameters of RBF and select appropriate features as well. The objective of this measure is to avoid over-fitting. This number is associated with the size of data set, which is normally within a 10–30% range of the total data. It was decided to choose 20% of the total data to have a smaller validation set error. If a smaller validation set was chosen, it would give a relatively noisy estimate of predictive performance [24]. The results obtained in this step are applied on the remaining 80% of the total data for classification in order to evaluate the performance of this newly proposed

method on BAA. To increase the reliability of this classification, a 5-fold cross-validation is run. Because bone growth is of different rates in females and males, an attempt is made to run this approach on both their radiographs, separately.

In this study, the accuracy of correctly labeled classes, the accuracy within the range of 1 year and 2 years, mean absolute error (MAE), standard deviation (SD), precision, recall, F-score, and Cohen's kappa are determined for the both genders' radiographs. The F-score is calculated through Eq. (3) [25]:

$$F-\text{score} = \frac{2 \times \text{Precision}_M \times \text{Recall}_M}{\text{Precision}_M + \text{Recall}_M} \tag{3}$$

where Precision_M and Recall_M are defined as:

$$\text{Precision}_M = \frac{\sum_{i=1}^{l} \frac{\text{tp}_i}{\text{tp}_i + \text{fp}_i}}{l} \tag{4}$$

$$\text{Recall}_M = \frac{\sum_{i=1}^{l} \frac{\text{tp}_i}{\text{tp}_i + \text{fn}_i}}{l} \tag{5}$$

where tp is the true positive, fp is the false positive, fn is the false negative, and l is the number of classes.

The mean absolute error and its SD are calculated through Eqs. (6) and (7) [7]:

$$\mu_{\text{err}} = \left(\frac{1}{H} \sum_{h=1}^{H} \left| \frac{a^{\text{read1}}(h) + a^{\text{read2}}(h)}{2} - a^{\text{predict}}(h) \right| \right) \tag{6}$$

$$\sigma_{\text{error}} = \sqrt{\frac{1}{H} \sum_{h=1}^{H} \left(\left| \frac{a^{\text{read1}}(h) + a^{\text{read2}}(h)}{2} - a^{\text{predict}}(h) \right| - \mu_{\text{err}} \right)^2} \tag{7}$$

where μ and σ are the mean absolute error and the SD, respectively, and a^{read1} and a^{read2} are the predicted age reported by the first and the second experienced radiologists, respectively. These two reports are available in the USC database for each hand h. Definition a^{predict} is the predicted age of this proposed method, and H is the total number of hand radiographs. The predicted age is calculated according to the equation presented by Kashif et al. as follows:

$$\text{Age} = 1/2(U_B(c) + L_B(c)) \tag{8}$$

where c is the predicted age class, and $U_B(c)$ and $L_B(c)$ are the upper and lower bands of this class [8].

Implementation

The accomplishment procedures are run completely in MATLAB. The computer vision toolbox of MATLAB provided built-in support for HOG and LBP. *Keypoint* detection and description procedures of SIFT are applied through the VLFEAT library version 0.9.20.

Results

The HOG, LBP, and dense SIFT are applied on ROIs of uniformed backgrounds in which the size of their feature vectors are 900, 360, and 1024, respectively. The HOG, LBP, and dense SIFT features are concatenated to improve the performance of this newly proposed method for BAA. A number of features are removed through the feature selection methods. The results of accuracy and accuracy in the range of 1 year and 2 years for HOG–LBP–dense SIFT feature extraction method and SD together with the minimum and maximum of accuracy in 5-folds are tabulated in Table 1. The results of MAE and its SD are also tabulated in this table.

The precision and the recall of different age ranges are calculated and illustrated for female and male radiographs in Figs. 4 and 5, respectively.

The obtained precision means are 0.7 ± 0.16 for female and 0.69 ± 0.16 for male radiographs. The recalls' means are 0.68 ± 0.22 and 0.65 ± 0.22 for female and male radiographs, respectively.

The results of F-score and Cohen's kappa with 95% CI are calculated and tabulated in Table 2.

The p value is reported in Cohen's kappa method. Here, the p value ≤ 0.05 is obtained for both the genders' radiographs indicating that the observed agreement is not accidental. The Cohen's kappa mean with the SD, the minimum and maximum in 5 folds are 0.71 ± 0.06 [0.63, 0.77] and 0.66 ± 0.03 [0.6, 0.68] for female and male radiographs, respectively.

Discussion

The objective of this study is to find a manner in increasing the reliability of automatic BAA. In this context, for the first time, HOG, LBP, dense SIFT computer vision methods and background subtraction are applied for assessing the bone age of 0- to 18-year-old children. The obtained results indicate that HOG–LBP–dense SIFT features and background subtraction, which constitute the new features of this article, are promising measures approach in assessing bone age.

The results of precision and recall reveal that this proposed algorithm is significantly reliable for estimating the

bone age in the subject group. The results indicate the acceptable performance of this proposed method in all age ranges except 11–12-years age range for females and 7–8 and 15–16-years age ranges for males, where precision fail slightly; the same indication is for ages 7–8-, 8–9-, 11–12-, and 12–13-years ranges for females, and 6–7, 10–11, and 14–15 for males, as to the recall, which fail slightly. One of the main reasons for decreasing the results in some age ranges is that the radiographs of four ethnic groups are examined in a simultaneous manner. In general, in these age ranges, the bone evaluation rate decreases. Furthermore, bone growth is at different rates in different racial origins; consequently, BAA error is inevitable in these age ranges. More investigations through USC dataset revealed that the bone ages reported by two experienced radiologists differ for more than 1 year in a large number of radiographs. This indicates that BAA is even challenging for experienced radiologists. The other reason is that the system is not well trained in some of the age ranges because of a limited number of data within the USC database.

To provide more accurate BAA, Cohen's kappa statistical test is calculated. The results obtained from Cohen's kappa indicate the substantial agreement between the results here and the gold standard (the mean of bone age reported by two experienced radiologists in the USC database). The p value ≤ 0.05 indicates that this agreement is not accidental. Here, better results are obtained in case of female radiographs as compared with that of male radiographs, thus, assuring the fact that female bones develop faster than that of the male [8].

There exist many studies in this context where the extraction methods of morphological features like the area, perimeter, and number of wrist bones face serious problems due to their being semi-automatic, time-consuming, and depending heavily on the proficiency of the user. Another restriction of these methods is in their non-applicability in images where bones are overlapped. For instance, due to the morphological feature extraction methods proposed by Zhang et al. [26], Somkantha et al. [27], and Güraksin et al. [28], the areas of the hamate and capitate should be calculated in a separate manner, and this may cause a problem if they overlap. Thus, their

Table 1 Validation results

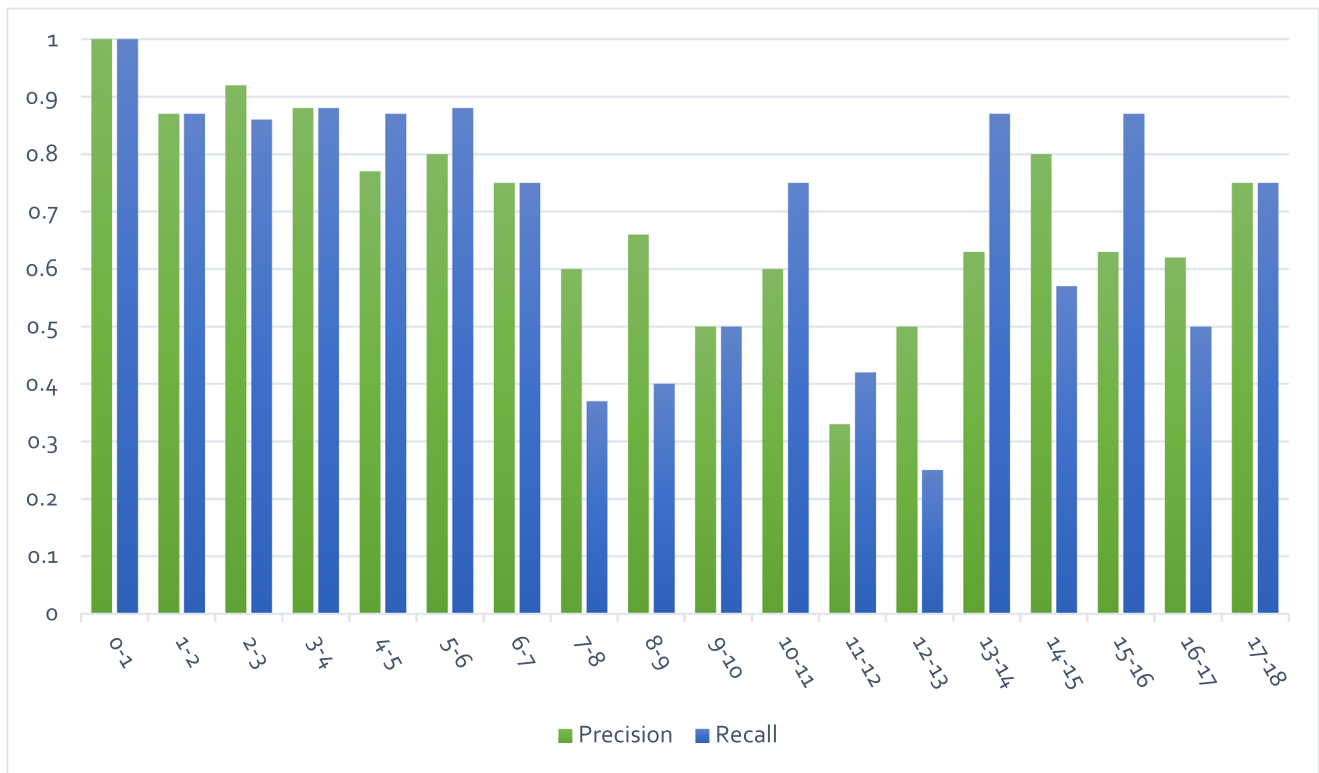| Outcome	Female	Male
Accuracy (%) = mean ± SD [min, max], 5 folds	73.88 ± 5.9 [65.78, 79.41]	68.63 ± 3.34 [62.85, 70.58]
Accuracy in 1 year (%)	95.32	96.51
Accuracy in 2 year (%)	100	99.41
MAE (year) = mean ± SD [min, max], 5 folds	0.55 ± 0.1 [0.42, 0.69]	0.56 ± 0.06 [0.47, 0.65]
SD (year)	0.49	0.49

Fig. 4 Precision and recall of different age ranges for female radiographs

studies are not applicable for children older than 5 to 7-years old. An automatic feature extraction method

proposed here overcomes these drawbacks. The background subtraction is not only more executable than

Fig. 5 Precision and recall of different age ranges for male radiographs

Table 2 The results of F-score and Cohen's kappa for female and male radiographs

Statistical test	Female	Male
F-score	68.98%	66.94%
Cohen's kappa	0.71 [CI 95% 0.64–0.78]	0.66 [CI 95% 0.59–0.73]

segmentation proposed in their publications but also has a better performance. The average time to extract HOG–LBP–dense SIFT feature from one radiograph is 1.04 s. This method is completely independent of the user's skill and is applicable to the entire radiographs.

The comparison between this proposed approach and the other prior attempts on BAA is tabulated in Table 3.

In studies proposed by Fischer et al. and Kashif et al., 14 eROIs are applied which causes the higher processing time for BAA. Their study is run on both genders' radiographs in a simultaneous manner which causes a serious error in assessing bone age. The Bonexpert is a commercial product which is considered as the most successful attempt in automatic BAA. In Bonexpert, bone ages are estimated within 2–15 and 2.5–17 years for females and males, respectively. This approach obtained a root mean square error of 0.61 years [29].

Here, in spite of using 2 ROIs, the higher accuracy and lower error are obtained. The carpal area has not been used in BAA within the 0 to 18 years range, while the results obtained in this study revealed that these bones accompanied by the epiphyseal center of proximal phalanx provide important indices for BAA. The presented approach is robust, easy to implement, and applicable to the entire age ranges, although it cannot be compared with that of Larson and the winning entry in the RSNA bone age challenge because of the difference in their datasets. According to the results tabulated in Table 3, this proposed approach outperforms the prior published methods on BAA. Various statistical tests are applied in BAA in order to determine the reliability of this approach more accurately. The main achievement of this study is its good performance on relatively small data sets for BAA. Because in many hospitals the data on bone age is limited, this newly proposed approach is advantageous.

Conclusions

In this article, by combining HOG, LBP, and dense SIFT features extracted from the carpal area and epiphyseal center of the proximal phalanx, an attempt is made to assess the bone age of 0- to 18-year-old children more accurately. The radiographic images applied here are available in the USC hand atlas database. The background of ROIs is uniformed through background subtraction in order to increase the efficiency of applied computer vision methods. SVM and 5-fold cross-validation are applied for the purpose of classification. According to the obtained results, this approach is appropriate for assessing the bone age of the subject group. The accuracy and MAE of female radiographs are 73.88% and 0.55 years, respectively. The accuracy and MAE of male radiographs are 68.63% and 0.56 years, respectively. The female and male radiographs assigned accuracy within the 1-year range are 95.32% and 96.51%. The accuracy within the 2-year range is 100% for female and 99.41% for male radiographs. To evaluate the performance of this proposed approach more accurately, precision, recall, F-score, and Cohen's kappa statistical test are run. The obtained precisions are 0.7 ± 0.16 for female and 0.69 ± 0.16 for male radiographs. The recalls are 0.68 ± 0.22 and 0.65 ± 0.22 for female and male radiographs, respectively. The obtained F-score is 68.98% for female radiographs and 66.94% for male radiographs. A detailed examination of Cohen's kappa reveals that these results are in substantial agreement with the gold standard. The obtained p value ≤ 0.05 for Cohen's kappa indicates that the observed agreement is not accidental for both genders' radiographs. This issue proves the reliability of this approach for assessing the bone age of the subject group. This proposed approach is qualified for computer-aided diagnosis (CAD) because it is low time-consuming, robust, and easy to implement. Here, it is revealed that this approach outperforms its prior counterparts. This proposed approach reveals that it is possible to develop a promising CAD system through a data set of approximately 440 radiographs. This approach may prove to be very beneficial for researchers with limited data sets.

Table 3 Comparison to published studies at BAA

Studies on BAA	Age range	Number of ROIs	MAE ± SD (years)	Accuracy (%)	Accuracy in 1 year (%)	Accuracy in 2 years (%)
Fischer et al. [7]	0–18	14	0.99 ± 0.76	–	–	–
Kashif et al. [8]	0–18	14	0.605	45.69	89.38	98.36
Lee et al. [9]	5–18	Different ROIs	–	57.32 for female 61.40 for male	90.39 for female 94.18 for male	98.11 for female 99 for male
Our work	0–18	2	0.55 ± 0.49 for female 0.56 ± 0.49 for male	73.88 for female 68.63 for male	95.32 for female 96.51 for male	100 for female 99.41 for male

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Moradi M, Sirous M, and Morovatti P The reliability of skeletal age determination in an Iranian sample using Greulich and Pyle method. Forensic science international, vol. 223, pp. 372. e1–372. e4, 2012

2. Greulich WW, Pyle SI: Radiographic atlas of skeletal development of hand wrist. California: Stanford University Press, 1971

3. Tanner J, Healy M, Goldstein H, Cameron N: Assessment of Skeletal Maturity and Predication of Adult height (TW3 Method). WB Saunders, 2001

4. De Sanctis V, Soliman AT, Di Maio S, Bedair S: Are the new automated methods for bone age estimation advantageous over the manual approaches? Pediatric endocrinology reviews: PER 12:200–205, 2014

5. Satoh M: Bone age: assessment methods and clinical applications. Clinical Pediatric Endocrinology 24:143–152, 2015

6. Pietka E, Gertych A, Pospiech S, Cao F, Huang H, Gilsanz V: Computer-assisted bone age assessment: Image preprocessing and epiphyseal/metaphyseal ROI extraction. IEEE transactions on medical imaging 20:715–729, 2001

7. Fischer B, Welter P, Günther RW, Deserno TM: Web-based bone age assessment by content-based image retrieval for case-based reasoning. International journal of computer assisted radiology and surgery 7:389–399, 2012

8. Kashif M, Deserno TM, Haak D, Jonas S: Feature description with SIFT, SURF, BRIEF, BRISK, or FREAK? A general question answered for bone age assessment. Computers in biology and medicine 68:67–75, 2016

9. Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, Choy G, Do S: Fully automated deep learning system for bone age assessment. Journal of digital imaging 30:427–441, 2017

10. Larson DB, Chen MC, Lungren MP, Halabi SS, Stence NV, Langlotz CP: Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. Radiology 287:313–322, 2017

11. Halabi SS, Prevedello LM, Kalpathy-Cramer J, Mamonov AB, Bilbily A, Cicero M et al.: The RSNA Pediatric Bone Age Machine Learning Challenge. Radiology:180736, 2018

12. Dalal N, Triggs B: Histograms of oriented gradients for human detection In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 2005, pp. 886–893.

13. Houam L, Hafiane A, Boukrouche A, Lespessailles E, Jennane R: One dimensional local binary pattern for bone texture characterization. Pattern Analysis and Applications 17:179–193, 2014

14. Lowe DG: Distinctive image features from scale-invariant keypoints. International journal of computer vision 60:91–110, 2004

15. Gertych A, Zhang A, Sayre J, Pospiech-Kurkowska S, Huang H: Bone age assessment of children using a digital hand atlas. Computerized Medical Imaging and Graphics 31:322–331, 2007

16. Perona P, Malik J: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on pattern analysis and machine intelligence 12:629–639, 1990

17. Güraksin GE, Uğuz H, Baykan ÖK: Bone age determination in young children (newborn to 6 years old) using support vector machines. Turkish Journal of Electrical Engineering & Computer Sciences 24:1693–1708, 2016

18. Wang Q, Yang J: Eye location and eye state detection in facial images with unconstrained background. Journal of Information and Computing Science 1:284–289, 2006

19. Wang X, Han TX, Yan S: An HOG-LBP human detector with partial occlusion handling. In Computer Vision, 2009 IEEE 12th International conference on, 2009, pp. 32–39.

20. Ojala T, Pietikäinen M, Mäenpää T: Gray scale and rotation invariant texture classification with local binary patterns. European Conference on Computer Vision:404–420, 2000

21. Ojala T, Pietikainen M, Maenpaa T: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence 24: 971–987, 2002

22. Lingras P, Butz C: Rough set based 1-v-1 and 1-vr approaches to support vector machine multi-classification. Information Sciences 177:3782–3798, 2007

23. Hsu C-W, Lin C-J: A comparison of methods for multiclass support vector machines. IEEE transactions on Neural Networks 13:415–425, 2002

24. Bishop CM: Pattern recognition and machine learning (information science and statistics). New York: Springer, 2006

25. Sokolova M, Lapalme G: A systematic analysis of performance measures for classification tasks. Information Processing & Management 45:427–437, 2009

26. Zhang A, Gertych A, Liu BJ: Automatic bone age assessment for young children from newborn to 7-year-old using carpal bones. Computerized Medical Imaging and Graphics 31:299–310, 2007

27. Somkantha K, Theera-Umpon N, Auephanwiriyakul S: Bone age assessment in young children using automatic carpal bone feature extraction and support vector regression. Journal of digital imaging 24:1044–1058, 2011

28. Güraksın GE, Haklı H, Uğuz H: Support vector machines classification based on particle swarm optimization for bone age determination. Applied Soft Computing 24:597–602, 2014

29. Thodberg HH, Kreiborg S, Juul A, Pedersen KD: The BoneXpert method for automated determination of skeletal maturity. IEEE transactions on medical imaging 28:52–66, 2009