ORIGINAL ARTICLE

Deep convolutional neural network designed for age assessment based on orthopantomography data

Seyed M. M. Kahaki[1] · Md. Jan Nordin[2] · Nazatul S. Ahmad[3] · Mahir Arzoky[4] · Waidah Ismail[5]

Abstract
In this paper, we proposed an age assessment method evaluated on Malaysian children aged between 1 and 17. The approach is based on global fuzzy segmentation, local feature extraction using a projection-based feature transform and a designed deep convolutional neural networks (DCNNs) model. In the first step, a global labelling process was achieved based on fuzzy segmentation, and then, the first-to-third molar teeth were segmented. The deformation invariant features were next extracted based on an intensity projection technique. This technique provided high-order features which were invariant to rotation and partial deformation changes. Finally, the designed DCNN model extracts a large set of features in the hierarchical layers which provided scale, rotation and deformation invariance. The method using this approach was evaluated using a comprehensive and labelled orthopantomographs of 456 patients, which were captured in the Department of Dentistry and Research at Universiti Sains Islam Malaysia. The results from the analysis have suggested that the method can classify the images with high performance, which enabled automated age estimation with high accuracy.

Keywords Age assessment · Orthopantomography data · Image processing · Deep learning

✉ Seyed M. M. Kahaki
  kahaki@neu.edu

  Md. Jan Nordin
  jan@ukm.edu.my

  Nazatul S. Ahmad
  drnazatul@gmail.com

  Mahir Arzoky
  mahir.arzoky@brunel.ac.uk

  Waidah Ismail
  waidah@usim.edu.my

[1] College of Science, Northeastern University, Boston, MA 02115, USA

[2] Center for Artificial Intelligence Technology, UKM, 43600 Bangi, Malaysia

[3] Faculty of Dentistry, Universiti Sains Islam Malaysia, 71800 Nilai, Malaysia

[4] Department of Computer Science, Brunel University, Uxbridge, UK

[5] Center of Holistic Intelligent, Universiti Sains Islam Malaysia, 71800 Nilai, Malaysia

1 Introduction

Automatic age assessment using dental imagery is an essential area of dentistry, orthodontics and forensics. Besides that, age estimation may also be useful for civil, criminal, law enforcement, border control, homeland security, airport security, forensic and anthropologic purposes [1]. Deep convolutional neural networks show promising robustness in different image processing tasks such as classification and segmentation [2, 3]. For example, in forensic studies, estimating dental age plays an important role in estimating the chronological age of children. Chronological age (CA) is typically defined as an age calculated based on an individual's birth date. A study by Smith and Brownlees [4] reported that only half of the children in the developing world had registered their date of birth. This important issue offers us a significant sign that automated age estimation is vital especially when involving forensic investigation and judicial proceedings. Also, age estimation is important for clinical dentistry to diagnose and plan for treatment as well as in other areas of dentistry. Likewise, estimating the age of children is crucial when the CA is missing or unrecorded. There are

several common methods for age assessment including skeletal age assessment [5, 6], X-ray of the sternal end of the clavicle [7, 8], age assessment based on psychosocial development [4, 9] and dental-based age assessment [10]. The most common method for skeletal-based age assessment is to analyse the X-ray image of the hand and wrist based on Greulich–Pyle [11] atlas. This method yields a standard deviation on 0.6–1.1 years and not designed to determine the chronological age. Methods based on X-ray of the sternal end of the clavicle are widely used in Belgium and the Netherland, but the variation between individuals of this method is also high [10]. The age assessment methods based on psychosocial development suggest that getting to know a child over a period of time and observing how they react to a situation can provide an age assessment. However, this method is not cost-efficient and is not always applicable since it needs observation over time. Age assessment based on dental images is usually based on extracted features from dental X-ray images. Dental features are known as the best form of evidence for post-mortem biometric identification as the tooth structure survives in most disasters such as accidents and violent crimes. In such events, where family members cannot identify the bodies, teeth information can be reliable identification proof. Likewise, the tooth development stage can also be used to identify the age of children which is not the case in other age assessment methods since they are usually determining whether a young person is a child or an adult [10]. Orthopantomographs are an excellent source of information in such cases to estimate the age of children. However, several factors, such as image quality, broken or missing teeth or the early stage of teeth, can be considered as obstacles in this field. Moreover, manual age assessment requires the effort and skills of an expert given the wide variation of skills between assessors. Therefore, an automatic assessment system is crucial in many cases to ease this process, which may help to reduce both the time and errors compared to observation using the naked eyes. Indeed, an automated system is vital in aiding the experts in identifying the CA of a patient based on their dental features.

The essential properties for any digital imaging system should include the following aspects: (a) the image produced is of diagnostic quality; (b) the radiation dose is equal or reduced compared to the film; (c) that digital radiology techniques are compatible with conventional X-ray generators; (d) that lossless archiving is allowed in an image file format that promotes interoperability within the Digital Imaging and Communications in Medicine (DICOM) standard; and (e) the time required for the total procedure should be equal to or less than the acquisition. The X-ray is a cheap and handy tool for dental diagnosis as it has a low level of radiation, is comfortable and is very quick for the patient to have taken. However, X-ray images suffer degradation from the low resolution that contributes to the existence of noise in the images. Therefore, the first step to process dental X-ray images is to first distinguish between the region of interest (ROI) and background. This task can be carried out through image segmentation, although this process has faced significant challenges due to poor quality. The object to be extracted from the image represents the ROI that contains valuable data used in later steps. The ROI can be defined as the part of the image that focuses on one object or part of an image. The global or local features should be extracted from the ROI to create the input of a classifier to obtain the identification result. The age assessment problem can then be solved using a deep convolutional neural network (DCNN) which is trained based on artificial intelligence (AI) techniques. This technique is compared in this study with other methods including support vector machine (SVM) and nearest neighbour (NN)-based approaches. This type of assessment is considered as a classification problem by segmenting the X-ray image globally and locally based on a fuzzy approach to extract the molar teeth sections in the first stage. In this case, the molar features are extracted using a proposed projection-based technique that is considered to be robust under illumination, in order to view the changes. These features are considered as input information to train the convolutional network. The convolutional network was considered in this study as it has been shown in the literature that other techniques such as SVM and NN-based methods were insufficient for the classification of these complex data.

2 Related works

Several biological ages have been developed during the last few decades such as the skeletal age, morphological age, secondary sex character age and dental age. These criteria can be applied separately or together to assess the degree of physiological maturity of a growing child [12]. Dental age based on tooth development stages was first introduced by Demirjian et al. [13] which is considered the most reliable and straightforward method as it has the highest values for both intra- and inter- observer agreements [14]. Demirjian and his colleagues introduced a system of age estimation based on tooth development stages where they classified teeth development into eight stages for a different sex, marked from stages A to H and zero for no appearance. Using this system, each stage was given a rating, which was then summed to achieve a final score called the maturity score. A table was also provided which was used to translate the maturity score into age.

To analyse X-ray images, three different stages can be considered including segmentation, feature extraction and classification. Segmentation of medical imagery is one of the most challenging and fundamental processes in almost all image analyses applications. The primary task in image segmentation is to divide an image into its constituent regions or objects, which helps the feature extraction stage to extract more accurate and distinctive features. In other words, segmentation eliminates redundant image sections. Segmentation also recognises and labels individual teeth in the X-ray image or parts of the tooth. For segmentation purposes, Ngan et al. [15] divided a dental X-Ray image into several segments and identified similar diseases using a classification method called affinity propagation clustering (APC+). However, their method was sensitive to noise and local variation.

In most segmentation algorithms, segmentation is performed either by extracting region-based features, which can identify different objects and regions or by applying a model and attempting to adjust its parameters to fit the processed objects or regions. Segmentation of dental X-ray images can also be divided into three basic techniques such as pixel based, region based and boundary based which were introduced by Rad et al. [16]. Other literature provides further categories of segmentation generalising on X-ray images such as pattern recognition based, deformable models, wavelet based and atlas-based techniques [17]. However, the most straightforward or simplest method that is typically used in image segmentation is based on the thresholding method and another popular threshold method called Otsu [18].

Applying the threshold method, pixels in a greyscale image are first separated into two classes where the pixels are either below the threshold value or above the threshold value. This technique can also be extended to include multiple threshold values to suit more types of images such as multi-colour images. However, selecting a threshold value is challenging, and it will determine whether an image is over a segment or otherwise. In dental X-ray image segmentation, Rad et al. [16] applied the level-set method for performing boundary-based segmentation which detected the edges and angles and other image characteristics on a surface covered by a curve. In their research, they found that choosing an appropriate function that represented the curve was a significant issue, and therefore, further research should be undertaken to address this issue. Tuan [19] later proposed a cooperative scheme that applied semi-supervised fuzzy clustering algorithms to the dental X-ray image called the Otsu method, which was used to remove the background area from the dental X-ray image. The fuzzy clustering algorithm (FCM) was chosen to remove the dental structure area from the results of the previous step. Finally, the semi-

supervised entropy regularised fuzzy clustering algorithm (eSFCM) was applied to clarify and improve the results based on the optimal result from the previous clustering method. Notably, the authors remarked that their framework did not use any dental features in their clustering algorithm.

Similarly, a study by Li and Wang [20] proposed that segmentation can be solved by a linear system defined by a discrete Laplace–Beltrami operator with Dirichlet boundary conditions. In this approach, a set of contour lines were sampled from the smooth scalar field, and candidate cutting boundaries were detected from concave regions with a large variety of field data. In this case, the algorithms were focusing on crowding problems. With the latest advancement of computer vision and machine learning (ML) algorithms like convolutional neural networks (CNN), it is now possible to implement in real-time systems including age estimation using dental X-ray images. CNNs are a class of hierarchical neural networks with multiple convolutions and pooling layers which exchange messages between each other [21].

In contrast to feature-based learning techniques that rely on extracting robust discriminative features from images with the aim of training a classifier to be used to discriminate between different categories, CNNs are based on feature representation learning. In fact, CNNs can learn useful representations of features directly from an image rather than using handcrafted discriminative features from pixel values in an image. Extracting meaningful features can also help the network to produce an output of better results. Also, in the CNNs, convolutional layers can extract the local features, while pooling layers reduce the resolution of features, making them more robust to noise [22, 23].

To use CNN for age estimation from X-ray images, some filters smaller than the original image are involved in the convolutional layers, and whose weights are learned in the training stage. These filters aim to extract the fine details of the image such as edges, corners and blobs. Pooling layers have also been applied to reduce the number of parameters to be computed and thus resulted in reducing computations at subsequent layers. Accordingly, CNN is trained by a collection of dental features and its corresponding to labels from the ROI extracted in the segmented sections. In particular, the training features are passed through the CNN and their specific features are then extracted using activation functions at particular layers of CNN. Although each layer of CNN responds to a training image, only a few layers are suitable for image feature extraction, where there is no specific condition for identifying these layers. Hence, a different number of layers are used to observe how well they perform. Štern et al. [24] proposed a framework of age assessment using dental MRI volumes and skeletal age by classifying the extracted

features. Their method focused on adult patients which were a more straightforward task compared to the age assessment of children. The ROI was also used in the Spampinato et al. [25] method to estimate the age of children based on left-hand MRI images. Their method used a pre-trained convolutional neural network with six layers, which showed promising results. However, these kinds of clear images without noise are not always available especially in disasters such as an accident or violent crimes.

3 Materials and methods

As mentioned earlier, automatic age assessment is considered to be a classification problem. In this study, a convolutional neural network was designed and trained using extracted features from the segmented image. In the first step, the molar teeth were extracted from orthopantomographs based on binarisation and localising of the teeth in the segmented image. For this goal, the orthopantomographs were analysed using a fuzzy technique and segmented globally, and the labelling process was carried out based on the segmented image using connected component analysis (CCA) followed by identifying the ROI based on local segmentation. The result of this process was then labelled using CCA, and the first-to-third molar images were then separated. Next, a combination of the different molar's features was extracted based on the projection-based transform, then, features were selected using principal component analysis (PCA), and phase, and magnitude spectrums were accumulated in a 2D matrix. The target information was defined using the ground truth information provided by metadata for each patient in the dataset. Finally, a deep CNN was designed to solve the classification problem using the achieved features in the previous step. The 2D matrix of features was used as the input matrix of the convolutional network to train the system. The visual structure of the deep convolutional neural network for the automatic age assessment technique is presented in Fig. 1. Each of the layers of CNN is described in Table 1.
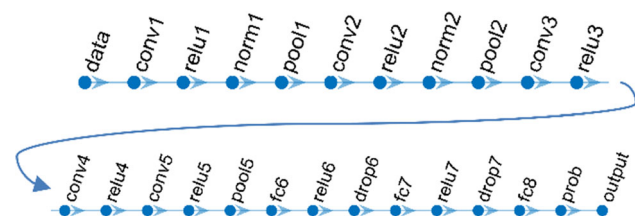

Fig. 1 Deep convolutional neural networks (DCNN) layers for automatic children age assessment

In this study, a dataset called Malaysian children dental development (MCDD) [26] was created containing X-ray images of 356 patients where each image and corresponding metadata file were stored in the dataset. The image capturing date and the birthdate of the patient were used to create the ground truth information for the classification stage of the age assessment. Following the data capturing stage and analysing the captured information, the images then needed to be analysed to achieve the results. To label the learning data, the orthopantomograph radiology images were captured, and the patient information was recorded in a metadata file along with the X-ray images. This process was undertaken in a lengthy process by medical experts at the Universiti Sains Islam Malaysia. This process was accomplished in several steps, and different experts monitored and evaluated the information to create a reliable and accurate dataset for further research studies. The dental image was next labelled based on a fuzzy approach to segment the molars, and the segmented data and the patient age were used to create the CNN input data.

3.1 Molars labelling

The segmentation problem was solved using fuzzy c-means (FCM) [27] which is a method allowing each data point to belong to multiple segments with varying degrees of membership. FCM is based on the minimisation of the following objective function:

$$j_k = \sum_{i=1}^{m} \sum_{j=1}^{n} \mu_{ij}^k \left| x_i - c_j \right|^2, \tag{1}$$

where m is the number of data points, n is the number of clusters, and k is the fuzzy partition matrix exponent for controlling the degree of fuzzy overlap. Fuzzy overlap refers to how fuzzy the boundaries are between the clusters, that is, the number of data points with significant membership in more than one cluster. x_i is the ith data point. c_j is the centre of the jth cluster. μ_{ij} is the degree of membership of x_i in the jth cluster. For a given data point, x_i, the sum of the membership values for all clusters is one. FCM performs the following steps during clustering: I. Randomly initialise the cluster membership values, μ_{ij}. II. Calculate the cluster centre:

$$c_j = \sum_{i=1}^{m} \mu_{ij}^k x_i \Bigg/ \sum_{i=1}^{m} \mu_{ij}^k \tag{2}$$

III. Update μ_{ij} according to the following:

$$\mu_{ij} = 1 \Bigg/ \sum_{k=1}^{n} \left(\left\| x_i - c_j \right\| / \left\| x_i - c_k \right\| \right)^{\frac{2}{m-1}} \tag{3}$$

Table 1 Detailed information of the DCNN

Layer	Name	Layer role	Description
1	data	Input	Input layer
2	conv1	Convolution	96 11 × 11 × 3 convolutions
3	relu1	Rectifier	Rectified-linear unit
4	norm1	Cross-channel normalisation	Cross-channel normalisation
5	pool1	Max pooling	3 × 3 max pooling
6	conv2	Convolution	256 5 × 5 × 48 convolutions
7	relu2	Rectifier	Rectified-linear unit
8	norm2	Cross-channel normalisation	Cross-channel normalisation
9	pool2	Max pooling	3 × 3 max pooling
10	conv3	Convolution	384 3 × 3 × 256 convolutions
11	relu3	Rectifier	Rectified-linear unit
12	conv4	Convolution	384 3 × 3 × 192 convolutions
13	relu4	Rectifier	Rectified-linear unit
14	conv5	Convolution	256 3 × 3 × 192 convolutions
15	relu5	Rectifier	Rectified-linear unit
16	pool5	Max pooling	3 × 3 max pooling
17	fc6	Fully connected	Fully connected
18	relu6	Rectifier	Rectified-linear unit
19	drop6	Dropout	50% dropout
20	fc7	Fully connected	4096 fully connected
21	relu7	Rectifier	Rectified-linear unit
22	drop7	Dropout	50% dropout
23	fc8	Fully connected	1000 fully connected
24	prob	Softmax	Softmax
25	output	Classification output	Cross-entropy-ex with four classes

IV. Calculate the objective function, j_m. V. Repeat steps two to four until j_m improves by less than a specified minimum threshold or until after a specified maximum number of iterations. After the global segmentation approach, the ROI, which is at the bottom right of the image, is extracted. This process is implemented to find the horizontal line, which separates the upper and lower teeth and identifying a vertical line that identifies the left and right sides of the mouth.

In the next step, the image is binarised, and the connected components are labelled based on the following procedures. First, scan all image pixels, assigning preliminary labels to nonzero pixels and recording label equivalences in a union–find table. Then, resolve the equivalence classes using the union-find algorithm [28]. Finally, relabel the pixels based on the resolved equivalence classes. The original input image, segmentation and molar labelling results are presented in Fig. 2.

3.2 Extract features

CNNs can extract features from raw images using convolutional filtering. However, the performance of the network can be improved by proposing a projection-based information extraction known as global feature extraction, which is based on the projection of locally segmented images. These projection features are combined with the co-occurrence matrix phase and the magnitude spectrum accumulated in a 2D feature matrix as the network inputs. This strategy also assists the network to have valuable information rather than extra redundant information such as the background. To extract the projection-based features, the input matrix which is an $N \times N$ of the first-to-third molar teeth images is rotated in $[0°, 90°]$, and the integral information is extracted. This projection-based method also helps to extract rotation invariant information from the raw image. Moreover, this image transform is general and is widely applicable to capturing different input signals to conduct sufficient and invariant features. Furthermore, this transform is an extension of the previously proposed transform [29] which was optimised and designed explicitly to generate the appropriate features from the molar input images. Indeed, this image transform guarantees that the detector extracts highly informative global features from the input shape to accurately and invariantly classify the data. The proposed transform G_f is

Fig. 2 a Original image, **b** a part of the segmented image, **c** labelled image and **d** a sample of the molar-segmented image
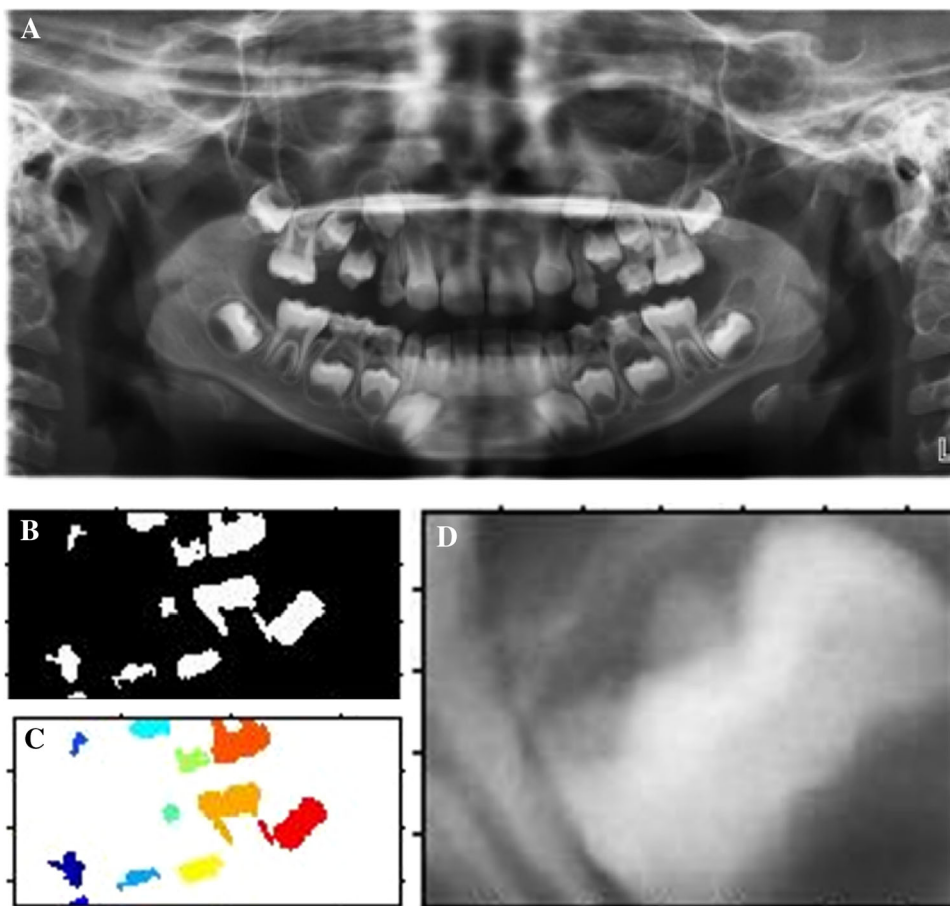
based on integrals of the geometric harmonic mean over straight lines in a digital image. Considering $f(x, y)$ as the function of the input image signal $S_{x|y}$ in $\Re$, then G_f is an image transform of $S_{x|y}$, where the harmonic mean of the horizontal and vertical integral is calculated as:

$$G_e = \sum_{Hr} \left(\int_{Sx,y} F(x, y) |dx, y| \right) \tag{4}$$

To parametrise any signal, S, with respect to the arc length and the Euclidean distance from the origin to S, G_e can be written as:

$$(x(t), y(t)) = \sum_{Hr} ([(t \sin(\alpha) + s \cos(\alpha)), \\ (-t \cos(\alpha) + s \sin(\alpha))]), \tag{5}$$

where α is the angle of the vector, and are the transform parameters for all signals, and G_e can be represented in the aforementioned coordinates according to Eq. (6):

$$G_e = \sum_{Hr} \left(\int_{-\infty}^{+\infty} F(x(t), y(t)) |dx, dy| \right), \tag{6}$$

which can also be written as:

$$G_e = \sum_{Hr} \left(\int_{-\infty}^{+\infty} F(t \sin(\alpha) + s \cos(\alpha)) |dy|, \\ \times \int_{-\infty}^{+\infty} F(-t \cos(\alpha) + s \sin(\alpha)) |dy| \right) \tag{7}$$

The G_e calculates the geometric mean of the integrals of an input image function in vertical and horizontal directions to calculate the feature vector matrix. These feature vectors are vectors of characteristic information extracted from the segmented image. In the next step, the magnitude and phase spectrums are achieved based on the shift zero-frequency component of the Fourier shift. Equation (8) defines the discrete Fourier transform y of an m-by-n matrix x:

$$y_{p+1,q+1} = \sum_{i=0}^{i=m-1} \sum_{j=0}^{j=n-1} \omega_m^{ip} \omega_n^{jq} x_{i+1,j+1}, \tag{8}$$

where ω_m and ω_n are the complex roots of unity: $\omega_m = e^{-2\pi/m}$ and $\omega_n = e^{-2\pi/n}$, i is the imaginary unit, p and j are indices that run from 0 to $m - 1$, and i and j are indices

which run from 0 to $n - 1$. In this case, the spectrum is calculated based on Eq. (9):

$$S = 100 \log \lVert 1 + y_{p+1,q+1} \rVert \tag{9}$$

This spectrum information of the local segmented parts in different images has been used as the input of the training set in the CNN. The target set was defined based on the given information for each patient in the metadata file as mentioned earlier. Therefore, the network was trained, and the classification result recorded as the result of the system.

3.3 Deep convolutional neural network

Deep convolutional neural networks (DCNNs) model (Table 1) has been proven effective and successful for complex image processing problems. To minimise the classification error rate in the automatic age assessment problem, a DCNN was trained using the information extracted from the first-to-third molar teeth images captured from the 356 patients aged between 1 and 17 years of age. These data were employed to solve the automatic age assessment issue as a classification problem.

The input features extracted from the segmented data are used in the network to eliminate the redundant information such as background and consider the valuable information from the image rather than using the raw data images, which usually contain many noises in the X-ray imagery. The DCNN was designed with seven totally hidden layers, five convolutional layers and three fully connected layers. Table 1 displays the detailed information of the network layers in the model including input, convolution, fully connected and output layers.

The convolution layer convolves the input data with a set of learnable filters, each producing a one-feature map in the output. The pooling layer, which is Max pooling, partitions the input image into a set of non-overlapping rectangles and, for each such sub-region, outputs the maximum. In the inner product or fully connected, the input is treated as a single vector with each point contributing to each point of the output vector. To train the network, a learning strategy is therefore needed. In the rectified-linear (ReLU) learning, given an input value x, the ReLU layer computes the output as x if $x > 0$ and negative slope x if $x \leq 0$. The rectified-linear unit can be written as:

$$\mathrm{ReL}(x) = \sum_{k=1}^{\infty} \sigma(x - k + 0.5) \approx \log(1 + \mathrm{e}^x) \tag{10}$$

where $\mathrm{ReL}(x)$ is the stepped sigmoid, and $\log(1 + \mathrm{e}^x)$ is a softplus function where it can be approximated by $(0, x + N(0, 1))$. The $\mathrm{sig}(x)$ has a range of $[0, 1]$ and can be used to model the probability, while $\mathrm{ReL}(x)$ has the range

in $[0, \infty]$ which can model a positive real number. If a hard max is used as the activation function, $\mathrm{ReL}(x)$ induces the sparsity of hidden units. Moreover, $\mathrm{ReL}(x)$ does not face the gradient vanishing problem which is the case in $\mathrm{sig}(x)$ and $\tanh(x)$ functions. Figure 3 demonstrates the weights of the convolutional layers following the learning process. These filters can be used in the test model to predict the probabilities of class belongings of new data.

However, some researchers believe that random convolutional connections can achieve adequate results. For example, Wan et al. [30] proved that the asymmetry in the layers' connection results in generalisation. Therefore, this structure has been chosen as the connection map, and the hyperbolic tangent sigmoid was used as the output layer function. Also, $\mu = 0.001$ is the coefficient for stochastic study and the number of epochs is 3. The training coefficient is $\theta = [50\ 50\ 20\ 20\ 20\ 10\ 10\ 10\ 5\ 5\ 5\ 5\ 1]/100{,}000$. The Hessian approximation is recalculated every iteration. Therefore, the number of iterations to pass for Hessian recalculation is 300; the number of samples for Hessian recalculation is 50, and α for the decrease coefficient is 0.4.

4 Results

To evaluate the propose method, a dataset called MCDD [26] was developed as a part of this research containing X-ray images and the image information of 456 patients was collected from Malaysian children aged between 1 and 17 years of age from the collection stored at the Department of Dentistry Research at Universiti Sains Islam Malaysia. These images were taken from the same source, which the pixel size and image sizes were the same. The information related to the age and sex distribution of the

Fig. 3 Convolutional layer weights

collected data is shown in Table 2. Based on Demirjian's [13] research as presented earlier, the development of teeth for different genders is noticeably different. This is the reason why the data were separated based on gender, and the classification and identification results were shown separately and combined.

Data were divided into five different age categories. Most of the age assessment techniques aim to determine whether a young person is a child or adult [10], which is a binary classification. Here, our method designed to investigate the age of children further in more details and make a multi-classification which is a more complicated task. This process can give the potential applications to determine the age categories more precisely for further analysis and legal decisions.

The molars were also selected for classification because they are considered more reliable in the population case. More straightforward stages of segmentation and classification in collected data lead us to conclude that the molars should have a high potential for age estimation purposes. In this case, the images are segmented, and the extracted features are accumulated in the input sets of the convolutional network as presented in the previous section. Cross-validation was used to evaluate the predictive model by partitioning the original sample into a training set to train the model, along with a test set to evaluate it. The maximum number of iterations chosen was 200, although the results were achieved in less than 100 iterations. Figure 4 presents the result of the identification error rate, which is presented as the accuracy and loss plots of the DCNN. There is a noticeable fluctuation at the beginning of the learning in both plots, which are smoothed during the learning process.

Different types of input data for training by convolutional networks are possible. In some applications, the original input images without pre-processing or features extraction are inputted to the convolutional network. In this type of application, usually, the number of input images is enormous. While, in some techniques, the original images are analysed by pre-processing methods, useful information extracted using segmentation to produce the network input. In the second type, the number of input images does

not need to be large because an enormous number of useful features can be extracted and different possibilities for the network are available based on a combination of feature vectors. In this study, the second approach was considered, and the results were presented. To evaluate the method, CNN was compared with the SVM and NN-based methods using 11 different distance metrics. Sequential minimal optimisation (SMO) was used as the SVM training solver, and the Bayesian optimiser was employed in this network. SMO uses heuristics to partition the training problem into smaller problems that can be solved analytically. Further, it works well but depends mainly on the assumptions behind the heuristics (working set selection) which speeds up the training.

The SMV classification searches merely for a hyperplane that separates the data into different classes based on their features. For these classes, the optimal plane maximises the margin, which is the space that does not contain any data to maximise the boundaries between data from different classes. For some data, which are not technically separable, the algorithm will consider a penalty on the length of the margin for each false observation. The linear SVM score function can be written as:

$$f(x) = x'\beta + b \tag{11}$$

where x is the row of X which is an observation. β is the coefficient vector that describes the hyperplane. Figure 5 presents the hyperplane results of SVM for different classes.

To minimise the error of the hyperplane, one solution that can be used is the NN approach using different distance functions. In the classification phase using KNN, k is a constant value that indicates the neighbour size, and the data can be classified by assigning a label that is most frequent among the neighbour of the training samples. Accordingly, the KNN is developed using a different number of neighbours and different distance functions including, Spearman, Seuclidean, Minkowski, Mahalanobis, Jaccard, Hamming, Euclidean, Cosine, Correlation, Chebyshev, and Cityblock. The best result was achieved using Minkowski distance with 12 neighbours, and the Bayesian function was used as the optimisation method in this approach. Figure 5 presents the classification objective function using different neighbour size and different distance functions.

To minimise the error of the hyperplane, one solution that can be used is the NN approach using different distance functions. In the classification phase using KNN, k is a constant value that indicates the neighbour size, and the data can be classified by assigning a label that is most frequent among the neighbour of the training samples. Accordingly, the KNN is developed using a different number of neighbours and different distance functions

Table 2 Data population information

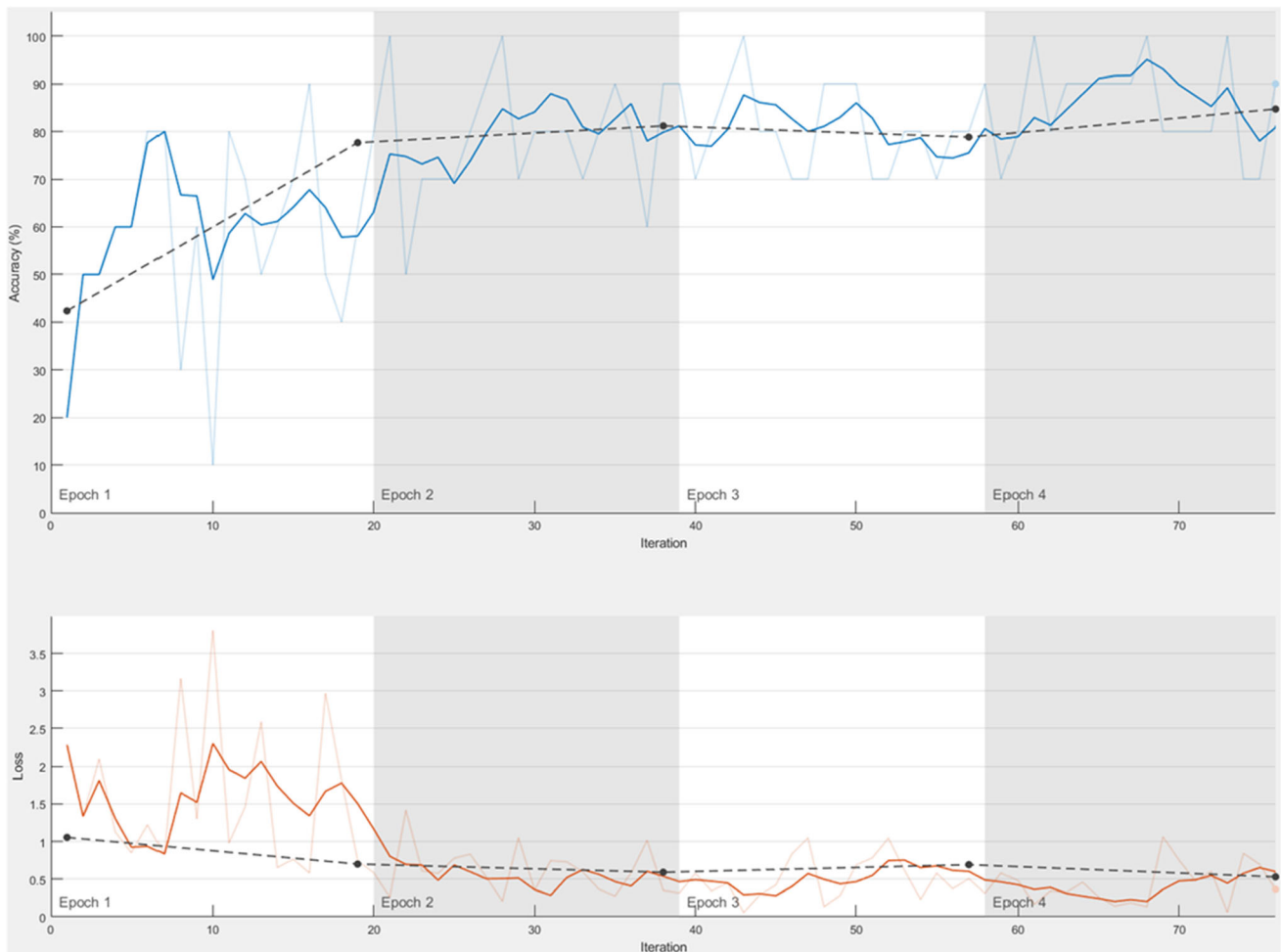| Age	First–third molars	
	Male	Female
1–4	18	19
5–7	46	58
8–10	58	73
11–13	44	52
14–17	41	47
Total	207	249

Fig. 4 Accuracy and loss plot of the deep convolutional neural network (DCNN)

including Spearman, Seuclidean, Minkowski, Maha-lanobis, Jaccard, Hamming, Euclidean, Cosine, Correlation, Chebyshev, and Cityblock. The best result was achieved using Minkowski distance with 12 neighbours, and the Bayesian function was used as the optimisation method in this approach.

Table 3 presents the classification results of the automatic age assessment technique using the convolutional network, SVM and proposed approaches. Based on the patient information extracted from the available metadata in the dataset, the patients were categorised based on the patient ages into four age categories as a class of patient (CoP). The classification rate result was also separated based on the male and female population given the teeth evaluation is different. However, most of the age assessment techniques aim to determine whether an individual is a child or an adult [10] as a binary classification; Table 3 demonstrates that the proposed method in this study successfully investigates the age of a child as a multi-class classification.

Based on the results presented in Table 3, it is observed that the classifier performance tends to increase with the increase in the age range. Based on the data population (see Table 2), since the number of samples for the first category of patients is limited compared to other categories, it achieved slightly lower performance. Since the adolescents have complete teeth structure and the size of the teeth was different from children and toddlers, the highest accuracy achieved by the last category (age 14–17).

The classification rate results in the table above indicate that the automatic age assessment can be achieved without human operation and with reasonable accuracy. This enables the experts to analyse the result and collect meaningful information to meet their goals in dentistry, orthodontics and forensics. The results indicate that in most cases, the system can identify the male patient age more precisely than the age of female patients. Therefore, one can conclude that the development of male patients is slightly higher compared to females in similar age groups.
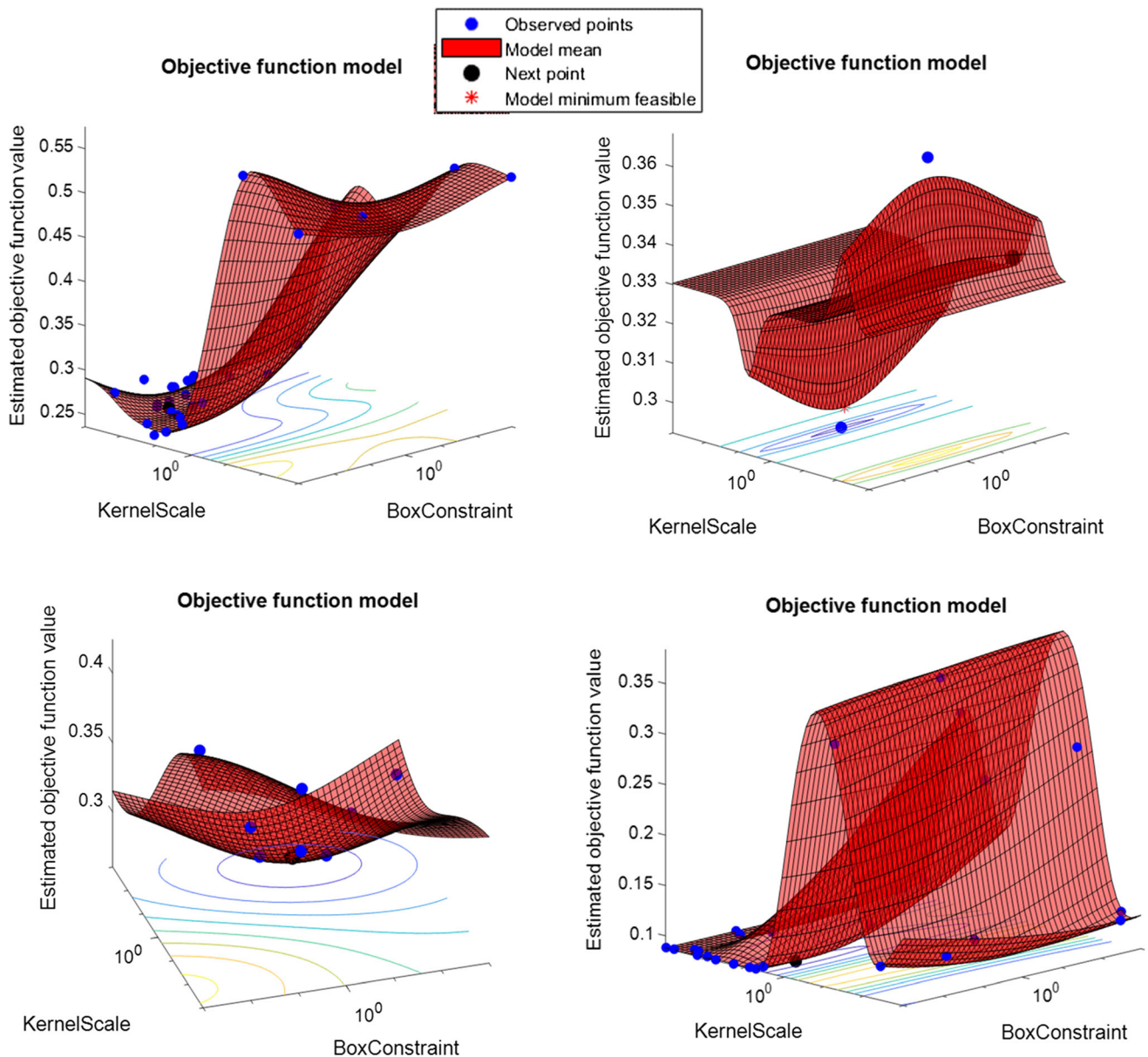
Fig. 5 Object function model means of SVM classifier solution for the age assessment problem

Table 3 Classification rate for automatic age assessment method

Category	Age	Male	Female	SVM	Proposed
Cop1	1–4	80.02	80.64	80.33	80.05
Cop2	5–7	80.11	80.93	80.52	77.15
Cop3	8–10	81.35	82.48	81.915	80.15
Cop4	11–13	81.24	82.91	82.075	85.37
Cop5	14–17	90.36	91.05	90.705	86.45
Cop1-5	Average	82.62	83.60	83.11	81.83

5 Conclusions

An approach was proposed in this study for the automatic age assessment technique based on a pre-trained DCNN using orthopantomographs. This automated estimation potentially helps experts in different fields to estimate the age of children based on teeth information. To achieve this goal, in the first stage, a fuzzy-based segmentation was first developed for global segmentation purpose to enable a more accurate local segmentation. Then, the first-to-third molar segments were localised based on a shape analysis method, which was followed in the next step, where the features were extracted based on the projection of locally

segmented images to speed up and enhance the learning process. Finally, a DCNN was designed and trained to identify the pre-labelled teeth classes, which lead to estimating the children's age. The proposed approach was tested on a captured dataset from the laboratory at the Faculty of Dentistry, Universiti Sains Islam Malaysia. These data were manually gathered from 365 patients, and the patient's age was used as the target set for the network. The results conclude that the method can efficiently classify the images with high performance that enables automated age estimation with high accuracy and precision. In the future study, the intention is to extend the data population to include elderly patients to expand the generality of the method. It is also proposed that anisotropic filtering in the early stages of CNN is considered for future research.

Compliance with ethical standards

References

1. Alsaffar H, Elshehawi W, Roberts G, Lucas V, McDonald F, Camilleri S (2017) Dental age estimation of children and adolescents: validation of the Maltese Reference Data Set. J Forensic Leg Med 45:29–31
2. Kang G, Liu K, Hou B, Zhang N (2017) 3D multi-view convolutional neural networks for lung nodule classification. PLoS ONE 12(11):e0188290
3. Qiao K, Chen J, Wang L, Zeng L, Yan B (2017) A top-down manner-based DCNN architecture for semantic image segmentation. PLoS ONE 12(3):e0174508
4. Smith T, Brownlees L (2011) Age assessment practices: a literature review and annotated bibliography. United Nations Children's Fund (UNICEF), New York
5. Terada Y, Kono S, Tamada D, Uchiumi T, Kose K, Miyagi R, Yamabe E, Yoshioka H (2013) Skeletal age assessment in children using an open compact MRI system. Magn Reson Med 69(6):1697–1702
6. Haiter-Neto F, Kurita LM, Menezes AV, Casanova MS (2006) Skeletal age assessment: a comparison of 3 methods. Am J Orthodont Dentofac Orthoped 130(4):435.e415–435.e420
7. Ji L, Terazawa K, Tsukamoto T, Haga K (1994) Estimation of age from epiphyseal union degrees of the sternal end of the clavicle. Hokkaido J Med Sci 69(1):104–111
8. Kreitner K-F, Schweden F, Riepert T, Nafe B, Thelen M (1998) Bone age determination based on the study of the medial extremity of the clavicle. Eur Radiol 8(7):1116–1122
9. Crawley H (2007) When is a child not a child? Asylum, age disputes and the process of age assessment. Immigration Law Practitioners' Association (ILPA), London
10. Hjern A, Brendler-Lindqvist M, Norredam M (2012) Age assessment of young asylum seekers. Acta Paediatr 101(1):4–7
11. Greulich W, Pyle S (1959) Radiographic atlas of skeletal development of the hand and wrist. Am J Med Sci 238(3):393
12. Chaillet N, Nyström M, Kataja M, Demirjian A (2004) Dental maturity curves in Finnish children: Demirjian's method revisited and polynomial functions for age estimation. J Forensic Sci 49(6):JFS2004211–JFS2004218
13. Demirjian A, Goldstein H, Tanner J (1973) A new system of dental age assessment. Human Biol 45:211–227
14. Olze A, Reisinger W, Geserick G, Schmeling A (2006) Age estimation of unaccompanied minors: Part II. Dental aspects. Forensic Sci Int 159:S65–S67
15. Ngan TT, Tuan TM, Minh NH, Dey N (2016) Decision making based on Fuzzy aggregation operators for medical diagnosis from dental X-ray images. J Med Syst 40(12):280
16. Rad AE, Mohd Rahim MS, Rehman A, Altameem A, Saba T (2013) Evaluation of current dental radiographs segmentation approaches in computer-aided applications. IETE Tech Rev 30(3):210–222
17. Stolojescu-Crişan C, Holban Ş (2013) A comparison of X-Ray image segmentation techniques. Adv Electr Comput Eng 13(3):85–92
18. Zhu N, Wang G, Yang G, Dai W (2009) A fast 2D Otsu thresholding algorithm based on improved histogram. In: Chinese conference on pattern recognition, 2009. CCPR 2009. IEEE, pp 1–5
19. Tuan TM (2016) A cooperative semi-supervised fuzzy clustering framework for dental X-ray image segmentation. Expert Syst Appl 46:380–393
20. Li Z, Wang H (2016) Interactive tooth separation from dental model using segmentation field. PLoS ONE 11(8):e0161159
21. Ciresan DC, Meier U, Masci J, Maria Gambardella L, Schmidhuber J (2011) Flexible, high performance convolutional neural networks for image classification. In: IJCAI Proceedings-international joint conference on artificial intelligence, 2011. Barcelona, Spain, vol 1, p 1237
22. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
23. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. pp 3431–3440
24. Štern D, Kainz P, Payer C, Urschler M (2017) Multi-factorial age estimation from skeletal and dental MRI volumes. In: International workshop on machine learning in medical imaging, 2017. Springer, Berlin, pp 61–69
25. Spampinato C, Palazzo S, Giordano D, Aldinucci M, Leonardi R (2017) Deep learning for automated skeletal bone age assessment in X-ray images. Med Image Anal 36:41–51

26. Kahaki SMM, Nordin MJ, Ahmand NS (2017) Malaysian children dental development (MCDD). USIM, Malaysia. http://www.kahaki.ir/source/MCDD.zip

27. Bezdek JC (2013) Pattern recognition with fuzzy objective function algorithms. Springer, Berlin

28. Cormen TH (2009) Introduction to algorithms. MIT Press, Cambridge

29. Kahaki SMM, Nordin MJ, Ashtari AH (2014) Contour-based corner detection and classification by using mean projection transform. Sensors 14(3):4126–4143. https://doi.org/10.3390/s140304126

30. Wan L, Zeiler M, Zhang S, Cun YL, Fergus R (2013) Regularization of neural networks using dropconnect. In: Proceedings of the 30th international conference on machine learning (ICML-13), 2013, pp 1058–1066