


RESEARCH

Open Access



Accuracy of advanced deep learning with tensorflow and keras for classifying teeth developmental stages in digital panoramic imaging

Norhasmira Mohammad¹, Anuar Mikdad Muad², Rohana Ahmad³ and Mohd Yusmiaidil Putera Mohd Yusof^{1,4*} 

Abstract

Background: This study aims to propose the combinations of image processing and machine learning model to segment the maturity development of the mandibular premolars using a Keras-based deep learning convolutional neural networks (DCNN) model.

Methods: A dataset consisting of 240 images (20 images per stage per sex) of retrospect digital dental panoramic imaging of patients between 5 and 14 years of age was retrieved. In image preprocessing, abounding box with a dimension of 250 × 250 pixels was assigned to the left mandibular first (P1) and second (P2) permanent premolars. The implementation of dynamic programming of active contour (DP-AC) and convolutions neural network on images that require the procedure of image filtration using Python TensorFlow and Keras libraries were performed in image segmentation and classification, respectively.

Results: Image segmentation using the DP-AC algorithm enhanced the visibility of the image features in the region of interest while suppressing the image's background noise. The proposed model has an accuracy of 97.74%, 96.63% and 78.13% on the training, validation, and testing set, respectively. In addition, moderate agreement (Kappa value = 0.58) between human observer and computer were identified. Nonetheless, a robust DCNN model was achieved as there is no sign of the model's over- or under-fitting upon the learning process.

Conclusions: The application of digital imaging and deep learning techniques used by the DP-AC and convolutions neural network algorithms to segment and identify premolars provides promising results for semi-automated forensic dental staging in the future.

Keywords: Deep learning, Tooth segmentation, Premolar, Digital dental panoramic radiographs

Background

Age estimation is essential in forensic medicine for identifying deceased victims and assessing the age of those involved in crimes and accidents. Forensic odontology is one of the forensic branches used for human identification through teeth. It is a term that refers to a variety of clinical, analytical, radiographic, and other methods for estimating dental age, which is then converted to

*Correspondence: yusmiaidil@uitm.edu.my

¹ Center for Oral and Maxillofacial Diagnostics and Medicine Studies, Faculty of Dentistry, Universiti Teknologi MARA, Sungai Buloh Campus, 47000 Sungai Buloh, Selangor, Malaysia

Full list of author information is available at the end of the article



chronological age. Since 1982, dental radiographic imaging, as obtained by the dental x-ray equipment, has been used in age estimation approaches as a non-destructive and straightforward technology used in dental practice [1]. In addition, the age estimation technique based on dental development has been widely used in other clinical set-up like the orthodontics and paediatric dentistry. This method entails identifying the mineralization stage on radiographic images and comparing it to a standard stage to determine the approximate age range [2, 3].

Demirjian's method is the most commonly used method for age estimation and has gained worldwide acceptability due to its objective criteria for describing the stages of tooth development [4–6]. This method involves human expertise to assess eight stages of dental development of the seven left permanent mandibular teeth. Then, each tooth stage is assigned biologic weights, which are numerical and computed following the method described in study on skeletal maturity [7]. Next, the dental maturity score is calculated by combining the weights. Finally, to convert maturity scores to dental age, separate tables of dental maturity for males and girls [4] are employed. The conventional way of estimating an individual's age can be an intricate procedure involving a significant number of forensic identifications, especially in the pandemic situation. Hence, automation of the conventional method may serve as a starting point for improving efficiency and reproducibility during identification process.

Digital image processing and deep learning techniques are applied to resolve the matters concerning the conventional method of forensic dental age estimation especially during large scale calamities for disaster victim identification. Issues such as the time taken to estimate age using manual atlas, subjective scoring due to operator's bias, and large number of cases to be completed within short period are among the reasons to expedite the use of semi-automated or fully automated dental scoring. Recent publications related to the segmentation of x-ray images for multipurpose medical application have been reported. Segmentation categories can be divided into four which are region-based, cluster-based, threshold-based and watershed-based segmentation. For region-based segmentation approach, dynamic programming of gradient inverse coefficient of variation (DP-GICOV) and Chan-Vese (CV) was employed in Muad et al. to perform a segmentation of mandibular third molar tooth for forensic human identification [8]. Results demonstrated that DP-GICOV performs a better segmentation with an accuracy of 95.3%. A method based on clustering segmentation has been presented where a novel semi-supervised fuzzy clustering algorithm with spatial constraints for dental segmentation is proposed [9]. The result shows

that the proposed method performs a better segmentation result compared to semi-supervised fuzzy clustering and other relevant methods. Hence, a new parameter has been suggested by the author for better performance.

For the threshold-based method, a novel semi-supervised Hyperbolic Tangent Gaussian kernel fuzzy C-Means clustering (HTGkFCM)-Otsu method is proposed in Kumar et al. [10]. The hyperbolic tangent Gaussian kernel method associates the input data space nonlinearly into attribute space having high dimensional and less sensitive to noise with robustness while adding kernel information to the proposed algorithm enhanced the original FCM algorithm. Watershed-based segmentation method has been widely tested on countless medical images and produced promising results. Recent studies on this method have been presented in various publications [11–13]. This approach's fundamental is based on the morphological mathematics operation where the visualization of a grey level in the images is transformed into a topographic representation that includes three basic concepts: local minima, catchment basins and watershed lines. The computation of marker image plays a significant role in determining a good segmentation output and minimalizing the over-segmentation result.

The deep learning convolutional neural networks (DCNN) have recently been introduced to perform automated tooth segmentation where these artificial intelligence methods have shown better performance than other mathematical approaches. A pilot study to stage lower third molar development on panoramic radiographs for age estimation was proposed in De Tobel et al. where the pre-trained AlexNet network model was adapted [14]. However, improvements have been made using DenseNet201 network for automated stage allocation [15]. Based on the improvement, the authors made a new hypothesis that only the third molar could improve the automated stage allocation performance. As of now, there are few studies utilizing DCNN in their methods [16, 17]. In the dental age estimation, literature stated that the third molar variation might affect age estimation accuracy on different populations [18–20]. The automated approach, the classification accuracy may also be affected due to the tooth's morphological and its surrounding. The unwanted object, such as periodontal ligament, bony structures, and mandibular nerve canal, influences automated stage allocation performance, as stated by Boedi et al. [15]. Following several systematic reviews on dental age estimation methods among children and adolescents, the use of premolar development stages as one of the variables accurately estimate the age with less than one year based on the margin of error [21, 22]. These studies, however, were based on conventional radiographic method of dental age estimation.

Therefore, the present pilot study aimed to observe the feasibility of both combinations of image processing and machine learning approach to segment and classify the maturity development of the mandibular premolars using a DCNN model. As these monoradicular teeth have less variation than the third molar, the pilot segmentation of premolars tooth may improve the proposed method's performance. The current study is an extension from the previous study by Mohammad et al. to enhance the existing segmentation and classification method [23]. In addition, Demirjian's staging system is adopted in which premolar teeth will be segmented and classified accordingly following the atlas, as shown in Fig. 1.

Proposed methodology

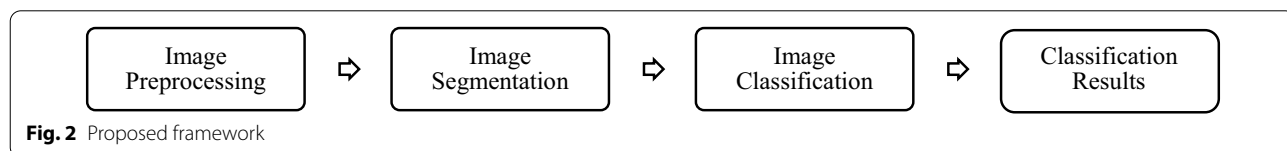
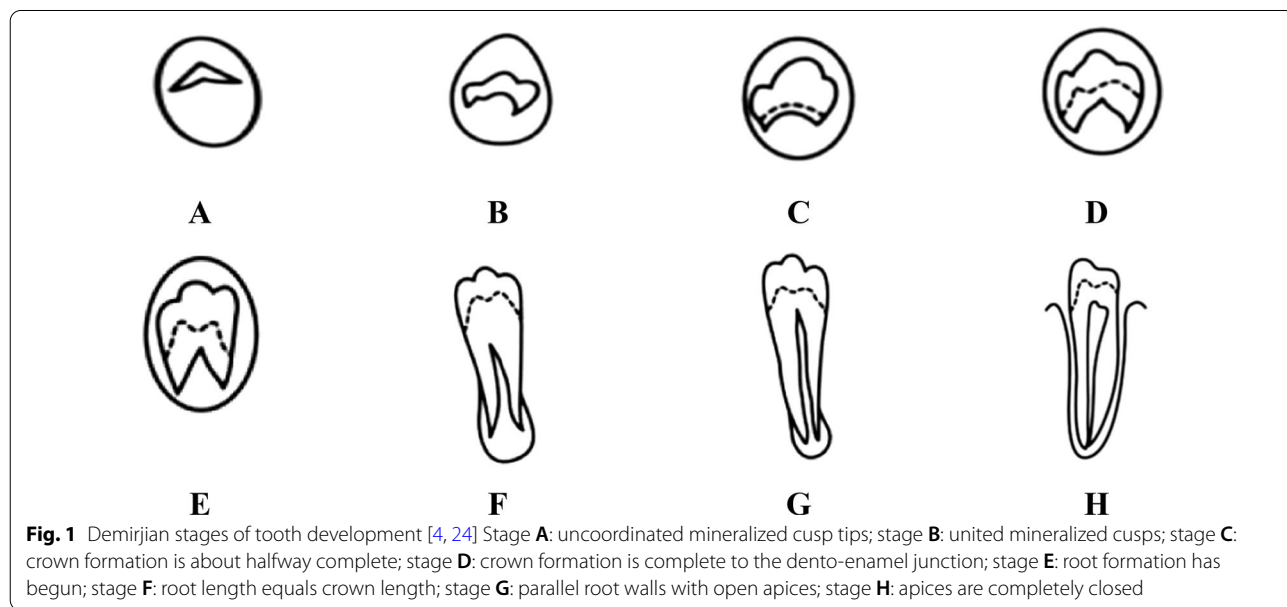
In this study, a new approach based on digital image processing and DCNN technique is applied to solve the original work issues. A dataset which consists of twenty radiographs per stage per sex were retrospectively selected. The methodologies include three phases which are image preprocessing, image segmentation and image classification. Figure 2 shows the sequence operation of the methodology. This study protocols have been reviewed and approved by the local institutional review board. The patients' informed consent was waived.

Image preprocessing

Image preprocessing includes the extraction of the region of interest (ROI) and image enhancement technique. Abounding box with a dimension of 250 × 250 pixels is assigned to all the mandibular first (P1) and second (P2) permanent premolars in panoramic dental radiographs where the object is approximately located at the centre of the image. Then, the intensity of the image is adjusted, and the median filter is applied. This operation works by highlighting the foreground and suppressing the image's background, which will help later processing. A median filter is used to remove an impulsive noise while preserving the object edges where a kernel filter with a size of 7 × 7 is applied.

Image segmentation

Active contour (AC) is one of the active models in segmentation techniques, which uses the image's energy constraints and forces to separate the region of interest. AC outlines a distinct boundary or curvature in the target region for segmentation. The basic computation of AC is based on optimizing a cost function where two main approaches were developed from the calculus of variations and dynamic programming (DP) [25]. The approach has the capability of dealing with a variety of



cost functions. However, it depends on gradient descent methods for local optimization and requires a good initialization for the contour whilst DP-AC computation does not require contour initialization, but it needs a confined form of the cost function [26].

For object delineation and segmentation, the DP-AC implementation is based on the formula proposed in the study by Ray et al. [27]. The original work was done to perform the particle object segmentation, a leukocyte in a microscopic image. To begin the segmentation process, as the ROI is located central to the segmentation setting, concerning Fig. 3e, an assumption is made by assuming a point inside the object and emerging radial lines come out from that point. Each radial line will intersect the object boundary only once. Based on Fig. 3a, assume there are M radial lines and on each M , let there be N discrete points. Therefore, a constructed MN different closed contours are obtained, in which programming uses only one point in each of M lines which denotes as v_i , shown in Fig. 3b. By v_i we denote a variable, which can take any of the N discrete graduations on the i th radial line. The following equality is a typical overlapping and additive form of a cost function that DP can optimize in $O(NM^2)$ computations:

$$E(v_1, v_2, \dots, v_N) = E_1(v_1, v_2) + E_2(v_2, v_3) + \dots + E_{N-1}(v_{N-1}, v_N) + E_N(v_N, v_1) \tag{1}$$

If $g(v_i)$ denotes the directional image gradient computed at location v_i , each additive cost component E_i can be defined as:

$$E_i(v_i, v_{i+1}) = \begin{cases} -g(v_i) & \text{if } D(v_i, v_{i+1}) \leq \delta \\ \infty & \text{Otherwise} \end{cases} \tag{2}$$

The cost component (2) implies that if the distance D between two consecutive points on the AC are within a certain user-defined distance δ , then the cost is the negative of the directional gradient; otherwise, the cost is assigned a large value. With (2) as the individual cost component, the net effect of minimizing (1) would be to obtain a contour with the maximum directional gradient strength and with any two consecutive contour points within δ distance. The end effect is a smooth contour as shown in Fig. 3c. Next, the ROI is extracted by converting polygon to a region mask in which the output shown in Fig. 3d. The mask image is then superimposed with the original image to obtain only the ROI without background objects. The image was saved in the JPEG file for later processing. Figure 4 depicts some random samples of results obtained after the superimposition procedure. These images will undergo the learning process using the classification algorithm, which is discussed in the next section.

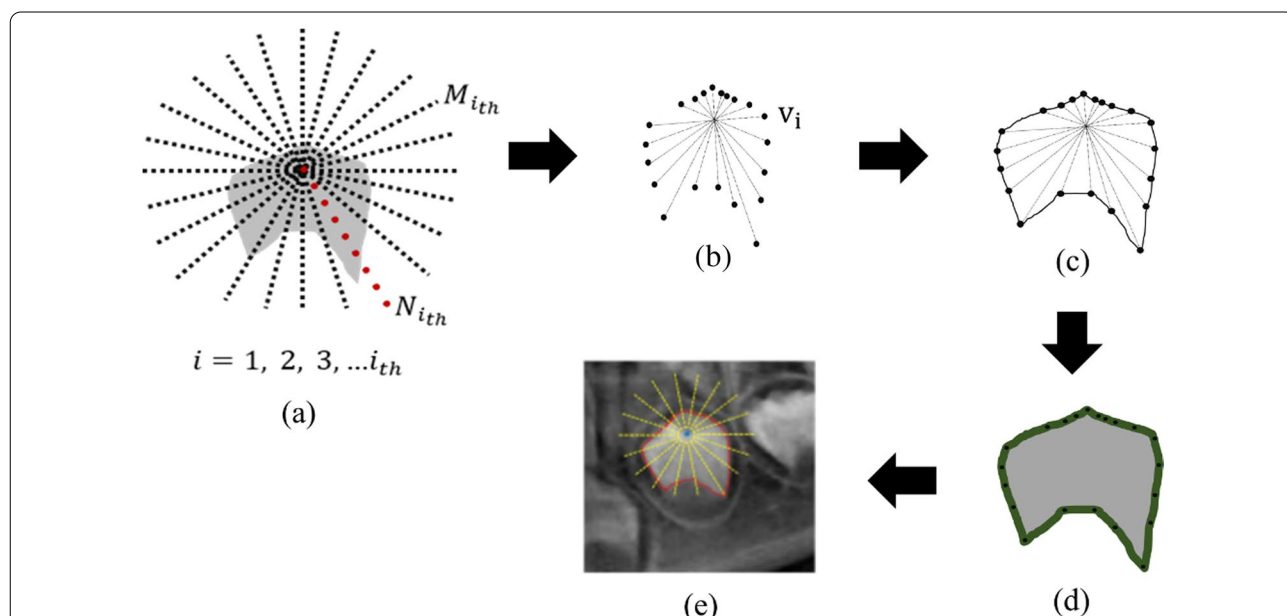


Fig. 3 The example of constructed radial lines in the schematic diagram forms a closed contour on ROI. Image **a** depicts the assumption of radial lines emerging from the centroid, **b** the detected $N_{i_{th}}$ point on $M_{i_{th}}$, **c** the contour with the greatest directional gradient strength, **d** the region mask obtained after performing polygon to binary conversion, and **e** delineated ROI based on a region mask on the original input image

Image classification

Multi-class classification with the Python Deep Learning Library is implemented in order to perform image classification. Figure 5 demonstrates the steps taken to implement DCNN using the Python TensorFlow and Keras libraries.

Figure 6 shows the graphical representation of the DCNN structure, where it consists of convolution (conv), pooling (pool) and fully connected (FC) layers. The input images consist of the segmented image of P1 and P2 are fed into the DCNN model where the classification takes place at the FC layer with an arbitrary output of 0 to 5, which involves six stages of dental growth (stage C to stage H). Eighty per cent of the dataset was allocated for training and validation, while the remaining 20 per cent was allocated for testing. As the available dataset consists of only 240 images, the data augmentation techniques are applied before the image classification is carried out. A new dataset consisting of 2400 images (200 images per stage per sex) is therefore obtained. The DCNN model with 3-convolutional layers-64 nodes-2 dense layers were

implemented in this report. Table 1 indicates the parameter assigned to DCNN architecture. Other parameters set for the experiment; optimizer = "Adam," epoch number = 10 and batch size = 8.

Results

Premolars segmentation

The proposed DP-AC method required the user to manually assigned the initial point of close contour. In this step, the placement of the initial point is crucial as it will determine the success rate of the image segmentation. Figure 7 shows the sequential steps of the image segmentation process, which involves the image filtration operation, placement of the initial point of closed contour, image conversion from polygon to binary and superimposition of the ROI with the original input image to retain the image pixels.

Based on the proposed segmentation algorithm, two important parameters are assigned before the segmentation process: the number of radial lines and the radius range of the radial line. These two may affect the overall

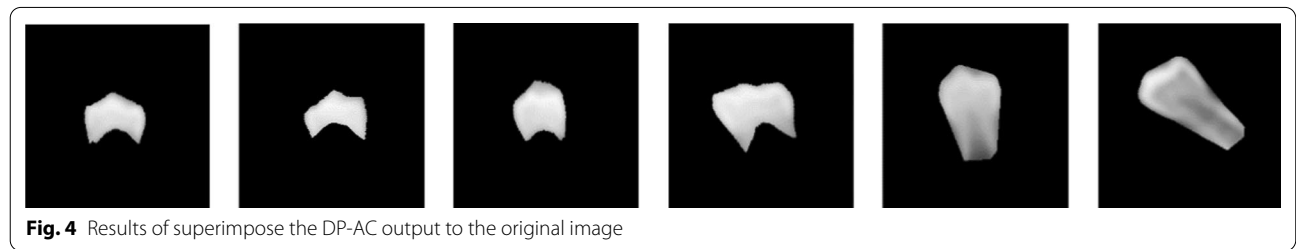


Fig. 4 Results of superimpose the DP-AC output to the original image

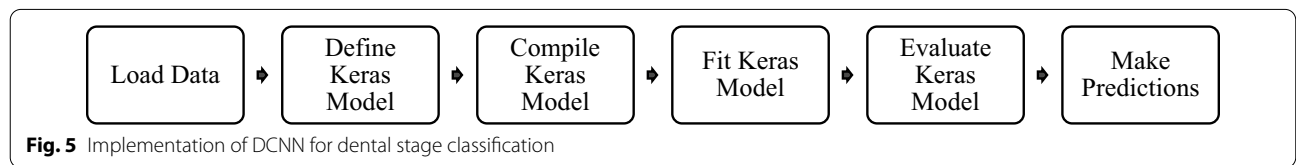


Fig. 5 Implementation of DCNN for dental stage classification

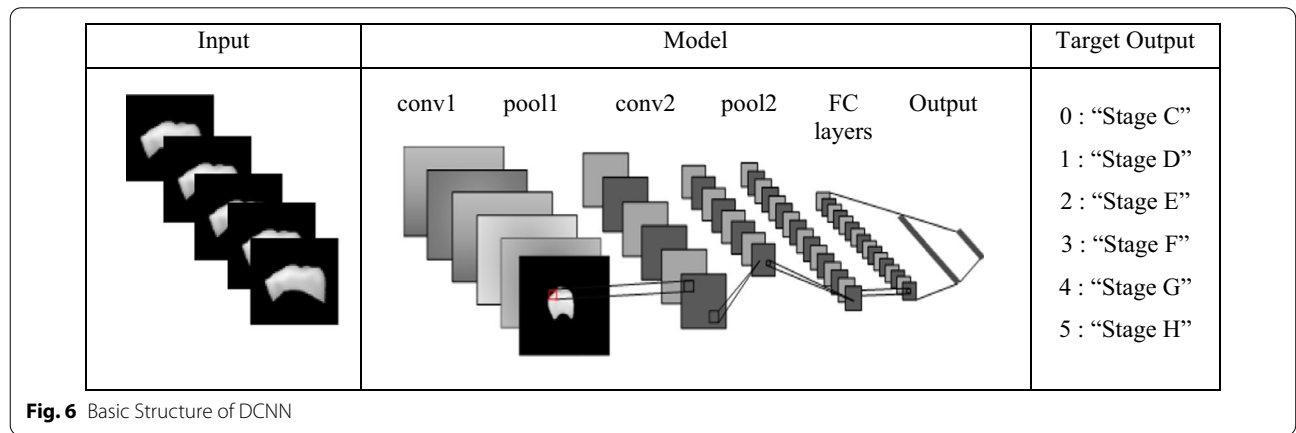


Fig. 6 Basic Structure of DCNN

Table 1 Defining DCNN model parameter

Layer (type)	Output shape	Parameter
conv2d (Conv2D)	(None, 90, 90, 64)	640
activation (Activation)	(None, 90, 90, 64)	0
conv2d_1 (Conv2D)	(None, 30, 30, 64)	36,928
activation_1 (Activation)	(None, 30, 30, 64)	0
max_pooling2d (MaxPooling2D)	(None, 15, 15, 64)	0
conv2d_2 (Conv2D)	(None, 15, 15, 64)	36,928
activation_2 (Activation)	(None, 15, 15, 64)	0
conv2d_3 (Conv2D)	(None, 5, 5, 64)	36,928
activation_3 (Activation)	(None, 5, 5, 64)	0
max_pooling2d_1 (MaxPooling2D)	(None, 2, 2, 64)	0
flatten (Flatten)	(None, 256)	0
dense (Dense)	(None, 30)	7710
activation_4 (Activation)	(None, 30)	0
dense_1 (Dense)	(None, 6)	186
activation_5 (Activation)	(None, 6)	0
Total parameters: 119,320		
Trainable parameters: 119,320		
Non-trainable parameters: 0		

segmentation accuracy. Hence, thirty images were randomly selected to undergo the segmentation process to measure the performance of the proposed algorithm. The average accuracy of the analysis was tabulated in Table 2. The results show that the low number of radius lengths may distort the object’s outline, resulting in under-segmentation. In contrast, a higher number may result in over-segmentation as well as longer computation time.

A significant correlation is seen between these two variables: the radius length and the number of radial lines assigned. The higher the number of radial lines in any radius length set, the longer the computation time. Besides, less computational time yields most of the output to be under-segmented, while the longer time taken results in over-segmented output. Therefore, the optimal parameters must be chosen based on the ability of the algorithm to segment the ROI in a decent length of time, with the majority of segmentation outputs being well-segmented and having a high segmentation accuracy. Based on the analysis, it can be seen that 100 pixels of radial length associated with 1200 radial lines is superior to other variables assigned. They produce a

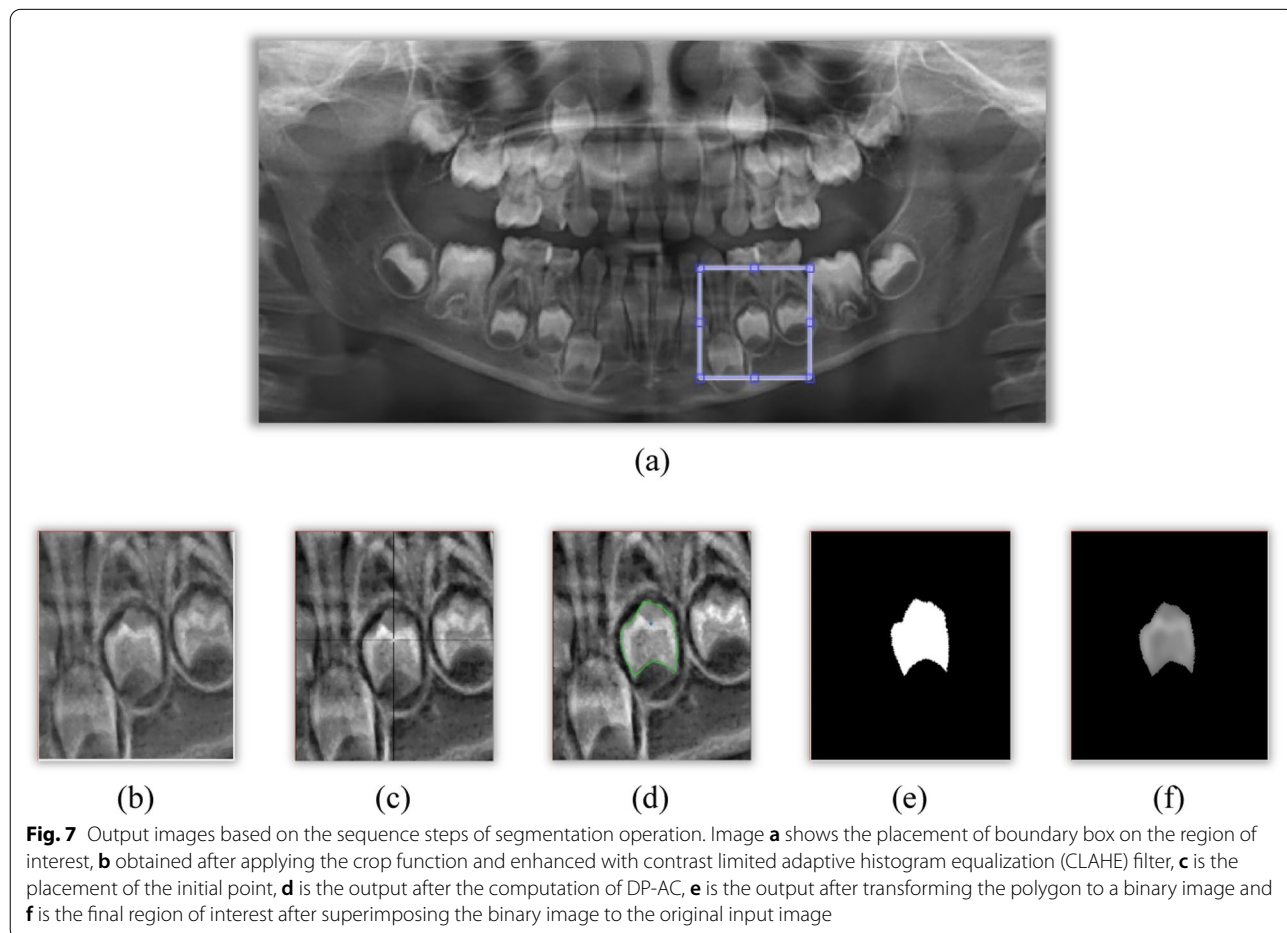


Fig. 7 Output images based on the sequence steps of segmentation operation. Image **a** shows the placement of boundary box on the region of interest, **b** obtained after applying the crop function and enhanced with contrast limited adaptive histogram equalization (CLAHE) filter, **c** is the placement of the initial point, **d** is the output after the computation of DP-AC, **e** is the output after transforming the polygon to a binary image and **f** is the final region of interest after superimposing the binary image to the original input image

Table 2 Segmentation accuracy according to parameters assigned based on P1 and P2

Radius length, in pixels	Number of radial lines	P1				P2			
		Output state	Time (s)	Dice Coefficient (F1 Score)	Jaccard similarity	Output state	Time (s)	Dice Coefficient (F1 score)	Jaccard similarity
80	800	*	1.64	0.85	0.76	*	1.74	0.81	0.73
	1000	*	1.75	0.87	0.78	*	1.83	0.83	0.74
	1200	***	2.36	0.88	0.75	*	1.97	0.85	0.78
100	800	*	2.01	0.85	0.74	*	2.13	0.86	0.75
	1000	**	2.22	0.86	0.77	**	2.29	0.91	0.86
	1200	**	2.42	0.96	0.90	**	2.31	0.94	0.88
120	800	**	2.54	0.94	0.85	*	2.68	0.93	0.84
	1000	*	2.92	0.89	0.73	**	2.77	0.80	0.71
	1200	**	3.26	0.92	0.83	*	3.64	0.92	0.83
140	800	**	2.97	0.92	0.81	*	3.21	0.93	0.84
	1000	**	3.03	0.90	0.79	***	3.98	0.89	0.81
	1200	***	3.93	0.85	0.72	***	4.13	0.84	0.76

*Under-segmented, **well-segmented, ***over-segmented

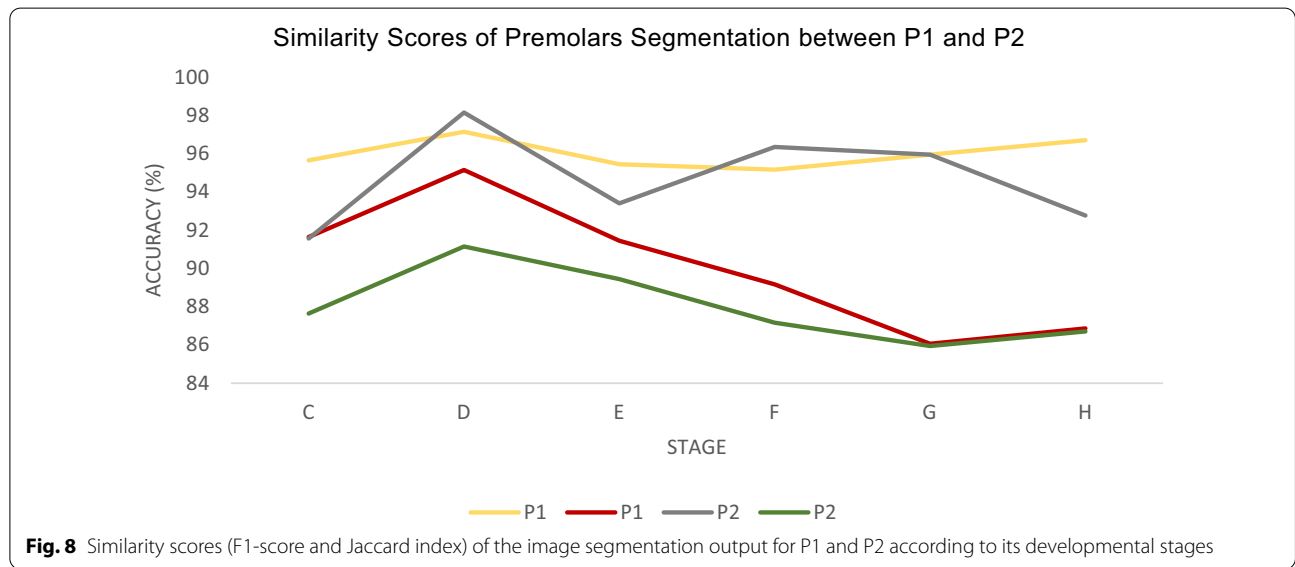


Fig. 8 Similarity scores (F1-score and Jaccard index) of the image segmentation output for P1 and P2 according to its developmental stages

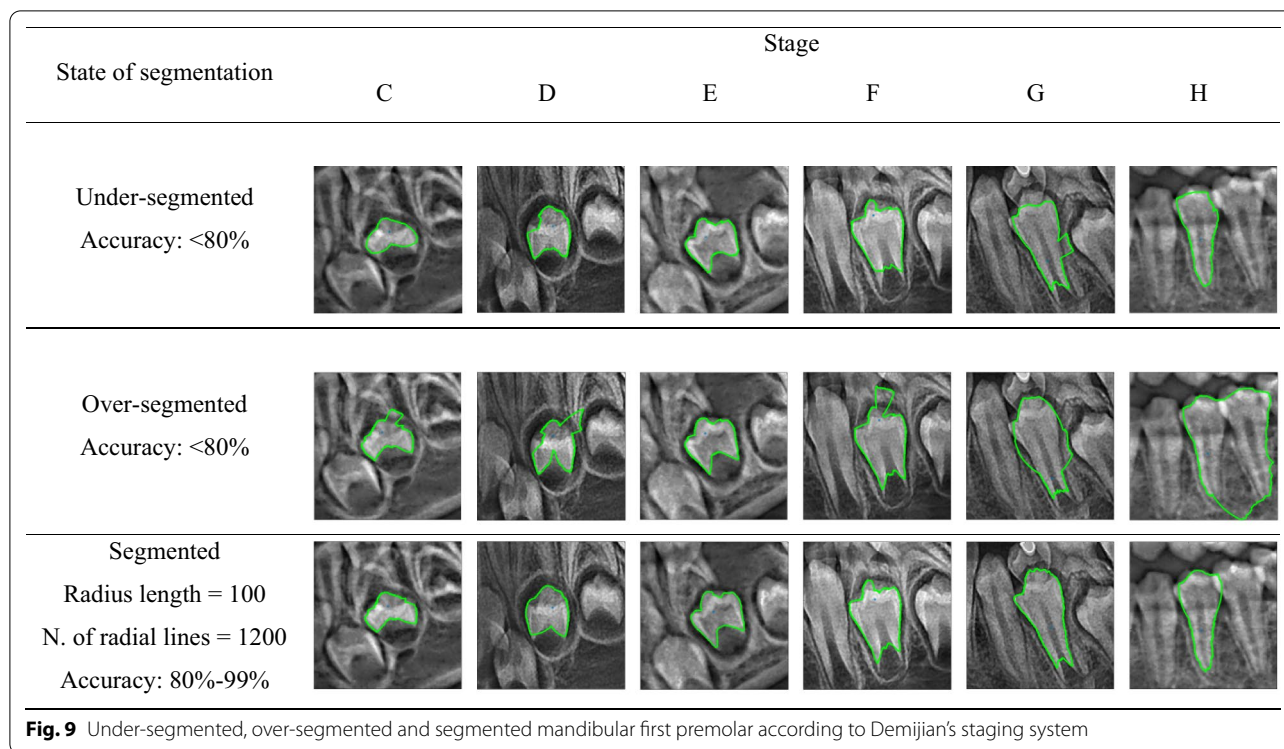
well-segmented object for almost all of the tested datasets and good segmentation accuracy in a reasonable time. The similarity scores between F1-score and Jaccard index according to the developmental stages was plotted in Fig. 8. On the same figure, it shows that the average segmentation accuracies are above 80%, indicating promising output.

Meanwhile, Fig. 9 shows an example of segmentation results from stages C to H, divided into three groups: under-segmented, over-segmented, and segmented. The following section will discuss the significance of

performing the image segmentation before the classification operation and the effect of performing the image segmentation on the classification accuracy.

Convolutions neural network model selection

The convolutions neural network model was selected based on the lowest validation loss obtained after the model optimization process. The algorithm was run using the Python TensorFlow and Keras library while the output was visualized in the Python Tensorboard.



Several parameters need to be assigned in the algorithm before performing the model selection which is as follows; Denselayers = [0, 1, 2], layer_sizes = [16, 32, 64] and conv_layers = [1, 2, 3].

Figure 10 shows the output for the convolutions neural network model optimization where the best three models were presented in the bounding box with a black colour where the model with 3-convolutional layers-64 nodes-2 dense layers indicates the best convolutions neural network model for our datasets by presenting the lowest validation loss among all tested model followed by the 3-convolutional layers-64 nodes-1 dense layers and 3-convolutional layers-64 nodes-0 dense layers.

Optimizers selection

In this analysis, "Adam" optimizer was selected with the default hyperparameter values which are $\beta_1 = 0.9$, $\beta_2 = 0.999$ and learning rate, $\epsilon = 10^{-3}$ to reduce the overall loss function proposed by the CNN model. The "Adam" optimizer was chosen by conducting cross-validation on the datasets. Four forms of optimizers were tested, including "Adam," "SGD," "RMSProp" and "AdaGrad," where the CNN model ran across 50 epochs. Figure 11 shows the efficiency of the network, depending on the optimizing assignment. Among all, "Adam" reveals the most stable optimizer as it converged in 30 epochs.

Accuracies and losses

A learning curve is a plot of model learning performance over time. Figure 12 shows the behaviour of the proposed DCNN model. A plot of learning curves shows slightly overfitting as the plot of validation loss decreases to the point that happened at the 4th epoch and begins increasing again. The proposed model has an accuracy of 97.74% on the training set and 96.63% on the accuracy plot's validation set. In other words, the model is expected to perform classification with 96.63% accuracy on new data. An extensive neural network trained on relatively small datasets will overfit training data. Adding a dropout layer is an easy way to avoid overfitting.

Table 3 demonstrates the learning accuracy of the dropout layer in the proposed DCNN model. Without dropout, the accuracy of the training is slightly higher than the accuracy of the validation while adding 10% (0.1) dropout, both the accuracy of the training and the accuracy of the validation is synchronized. However, the increase in dropout values can reduce the accuracy of learning. Furthermore, image segmentation using the DP-AC algorithm helps enhance the visibility of the image features in the ROI while suppressing the image's background noise. Hence, this may impact the classification accuracy of the test dataset.

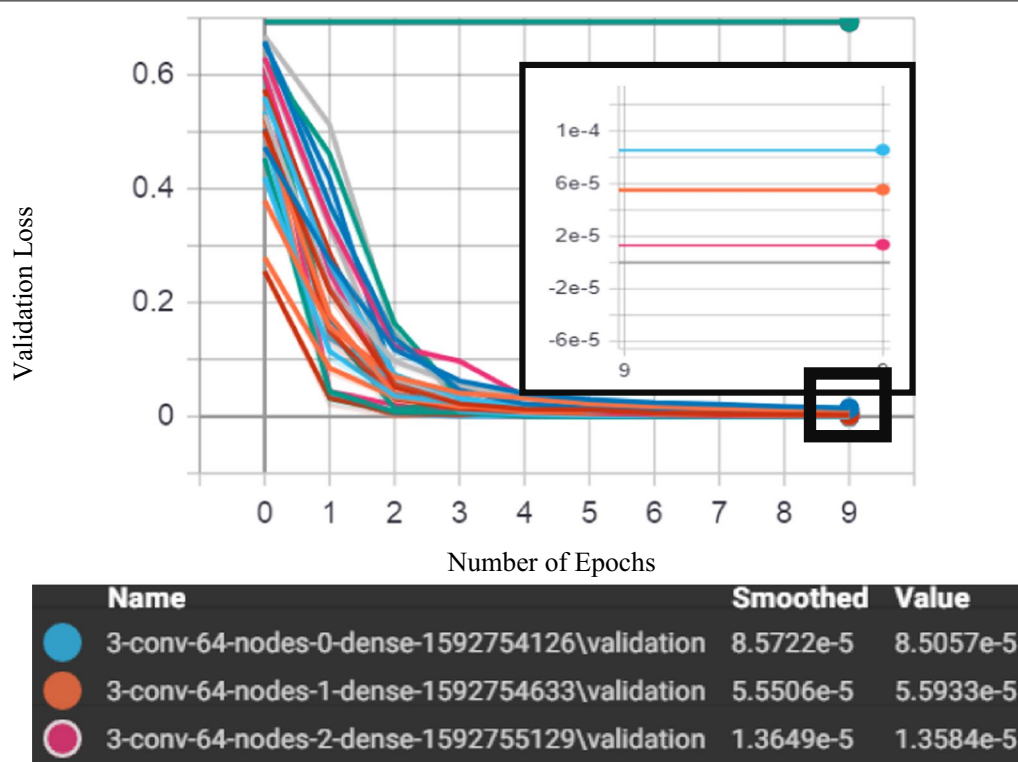


Fig. 10 Selection model of convolutional neural network

In machine learning, a confusion matrix depicts the summary of prediction results on a classification problem. It is also often used to describe the performance of the classifier, in this case, the proposed DCNN model presented in this study. This study categorized as a multiclass classification problem as the desired output would be to classify images into six categories that involve six stages of dental development. Thus, the confusion matrix would be a 6×6 matrix, as shown in Fig. 13. The classification accuracy on a set of test data obtained is 0.781. The confusion matrix reveals no misclassification in stage C, while small errors were detected in stages G and H. However, the accuracy appears to be quite promising. Meanwhile, most misclassified stages minimally happened in the neighbouring stages only for stages D, E, and F.

Cohen’s Kappa was also employed in the context of a classification model to compare the machine learning model predictions with the manually established scores. It is also used to evaluate the performance of a classification model. Table 4 shows the allocated stages by the machine (rows) and by the human observers (columns). The value for Kappa is 0.58, indicating a moderate level of agreement.

Discussion

Digital image processing (DIP) is the process of transforming a digital image using a set of algorithms. It includes simple tasks like picture filtration, as well as more complicated tasks like image segmentation, classifications, emotion identification, anomaly detection, and more. Image segmentation is the process of dividing a digital image into many subgroups based on the pixels known as image objects, which can minimize the image’s complexity and thus make image analysis easier. It has been utilized in the medical profession for effective and faster diagnosis, as well as the detection of illnesses, tumors, and cell and tissue patterns obtained from various medical imaging techniques such as radiography, MRI, CT scan, ultrasound, and so on.

The proposed method, which requires segmentation of mandibular premolar teeth before image classification, has some flaws that need to be worked out in future studies. The original images for digital panoramic dental imaging were of various sizes and resolutions. The input images should be normalized as a result. When the chosen parameters affect image classification accuracy, using the proper image normalization techniques is crucial.

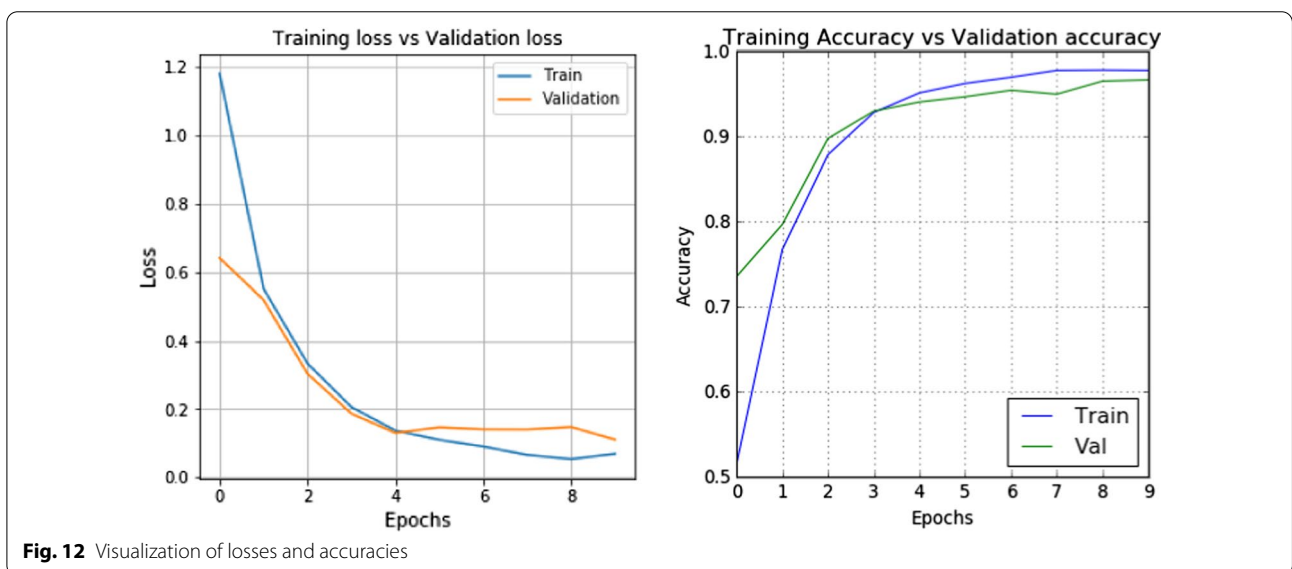
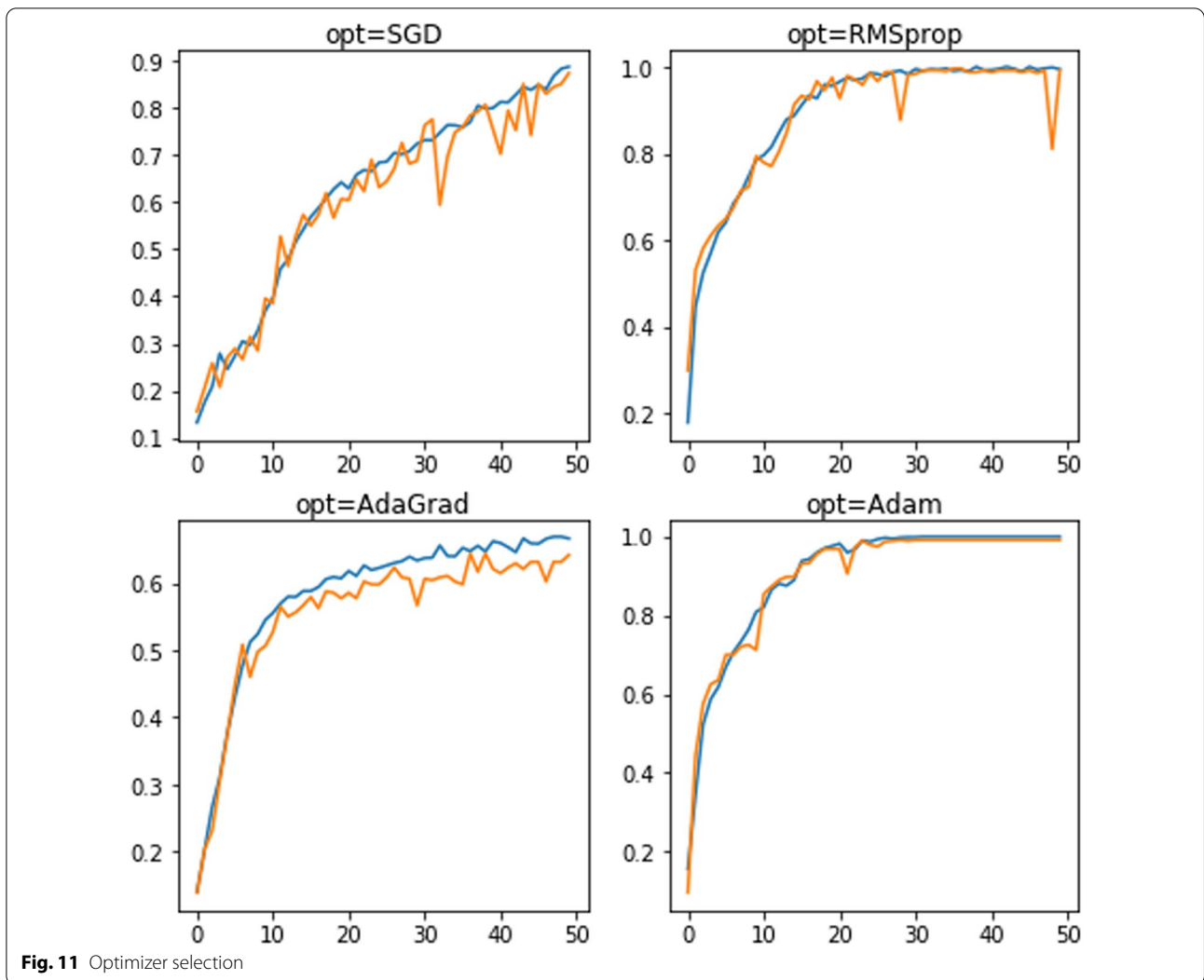
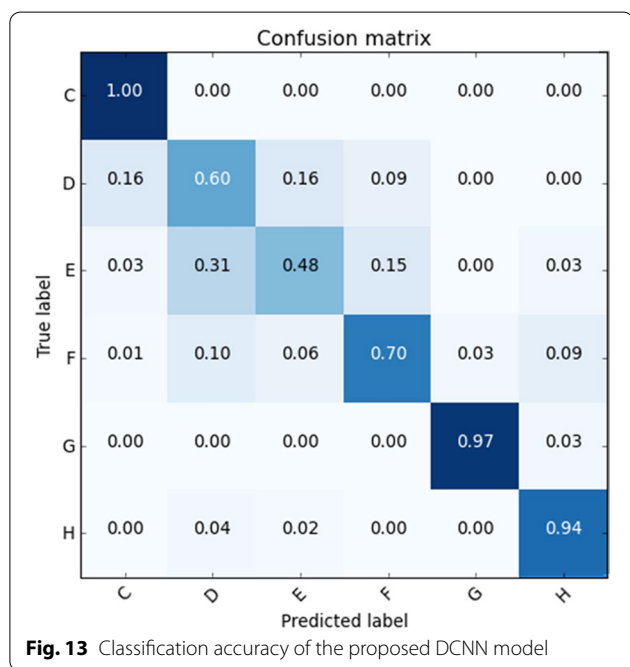


Table 3 Training and validation accuracy with and without dropout layer

	Training accuracy	Validation accuracy
Without dropout	97.74	96.63
With dropout ($p=0.1$)	96.17	96.17
With dropout ($p=0.5$)	86.45	91.42
With dropout ($p=0.7$)	78.60	90.66
With dropout ($p=0.9$)	52.45	73.35



Resizing an image, for example, may increase its general size but reduce its resolution and distort the edges of the ROI, lowering classification accuracy. Furthermore, the dental developmental stages A and B were omitted and will not be tested due to the limited number of datasets. It is because the dental development has proceeded to the stage of the lower-bound chronological age (stages A

and B). Hence, the input dataset for image classification only involved the image of mandibular premolars' developmental stage from stage C onwards until stage H.

The quality of digital panoramic dental imaging varies greatly from one patient to another, depending on the patient's position during the treatment as well as the expertise of the human operators. As a result, rather of going through the automated procedure, a semi-automated technique has been implemented. For example, getting a decent quality image requires proper placement of a bite-blocker and the patient's head. Furthermore, the panoramic x-ray can give a somewhat fuzzy image from time to time, making precise measurements of your teeth and jaw problematic. As a result, due to the wide range of data sources, developing a fully automated system for age estimation can be difficult.

Image segmentation is employed in this study to segment the first and second mandibular of permanent teeth before performing image classification, based on the adaptability of digital image processing technique. In this entire semi-automated approach, the use of the DP-AC method has proven to be successful. The easy steps to implement the DP-AC method are presented in Fig. 5, making this approach handleable by the end-user. Rather than constructing a completely automated system, which would require a considerable financial investment, a semi-automated system has produced promising results and satisfactory performance in the assessment of dental age.

The dental age for permanent teeth can be estimated by monitoring dental calcification development using radiographic images. Demirjian et al. suggested the staging of teeth based on the development of the teeth' outline instead of its proportions using the lower-left seven permanent teeth, except the third molar. The implementation of DCNN for the classification of dental stages yields promising results as the accuracy of the classification obtained is very high in some of the predicted stages.

Misclassification has occurred due to the variety of factors that may be linked to the deep neural network's effectiveness and the significance of assigned

Table 4 Cross-tabulation of the stages assigned using the ground truth data (rows) and the DCNN model (columns)

Stages	C	D	E	F	G	H
C	0.96	0.04	0	0	0	0
D	0.11	0.63	0.13	0.14	0	0
E	0	0.16	0.56	0.28	0	0
F	0	0	0.15	0.66	0.19	0
G	0	0	0	0.03	0.94	0
H	0	0	0	0	0.09	0.91

Kappa index: 0.58, Overall accuracy: 0.77

parameters or other factors related to dental morphology that affect the neural network's ability to achieve a useful classification. The proposed DCNN model was explicitly built based on our datasets. The experiment was performed by assigning parameters to a network model involving layer sizes, several dense layers and convolution layers. As a result, the stage classification accuracy obtained was 0.78. A pre-trained model of CNN which are DenseNet201 and AlexNet was adapted by Merdietio et al. [15], Banar et al. [28] and De Tobel et al. [14], respectively. Based on the performance, the proposed method performed superior compared to the other three methods, which are 0.61, 0.54 and 0.51, respectively.

Moreover, the misclassification of the test datasets is likely due to the behavior of the dataset itself. For example, the new test data implemented did not adequately reflect the broader domain cases. Therefore, this would potentially impact the accuracy of the test. However, based on Table 4, the proposed DCNN model looks promising as more than 90% of the test data in stages C, G and H were correctly allocated. Lower accuracies in stages D, E and F may be due to the significant variation of the morphological structure of dentition between stages. As the human interpretations of allocated stages are highly dependent on skills and experiences, a mutual agreement could not be achieved in some observation samples. Hence, the kappa value obtained is 0.58, indicating moderate agreement. However, most misclassified stages were seen only in the neighboring stages. Although it was not a perfect agreement, the proposed DCNN model showed a robust network. There is no sign of whether the model is over- or underfitting detected during the learning process, whereby the training accuracy was noted higher than the validation and testing accuracy.

Common challenges in deep learning models include the lack of data available for training, model overfitting, model underfitting and high training time. In this research, data augmentation techniques involving image resizing, rescaling, spinning, flipping, cropping, filtering, and brightness modification have been used to increase the number of training datasets. This technique is achieved using the open-source Python pre-processing package known as Scikit-image. Therefore, the model was able to perform well and escape underfitting in the validation collection.

Model overfitting is the most common problem that data scientist has faced in the field of machine learning [29]. Introduction of the dropout feature to the design of the model is one of the methods used to resolve the overfitting problem. Some of the neurons in the neural network were switched off using the dropout. For

example, in an experiment, a drop of 0.1 to a layer initially had 30 neurons that removed three neurons out of its total number of neurons. As a result, a less complicated architecture was obtained, and the model will not learn the intricate pattern. Overall, it can be argued that the DCNN structure plays a critical role in the classification process as it determines the overall performance of the automated stage allocation.

Conclusion

The application of digital imaging and Keras-based deep learning techniques used by the DP-AC and convolutions neural network algorithms to segment and identify premolars provides promising results for semi-automated forensic dental staging in the future. The techniques used to optimize the DCNN model can be expanded by adding some other hyperparameters to pick a better model with better performance. In addition, wide ranges of datasets will be tested on the proposed model to reduce the inter-rater discrepancy and enhance reproducibility.

Abbreviations

DCNN: Deep learning convolutional neural networks; DP-AC: Dynamic programming of active contour; P1: Left mandibular first permanent premolars; P2: Left mandibular second permanent premolars; DP-GICOV: Dynamic programming of gradient inverse coefficient of variation; CV: Chan-Vese; HTGkFCM: Hyperbolic tangent gaussian kernel fuzzy C-means clustering.

Acknowledgements

The authors would like to acknowledge Universiti Teknologi MARA, Cawangan Selangor (UCS) for the financial support via Dana UCS and all radiographers at Diagnostic Imaging Unit, Faculty of Dentistry, Universiti Teknologi MARA, Malaysia.

Author contributions

Author NM carried out the data collection, experiments, analysis and wrote the manuscript draft. Authors AMM and RA were involved in the formulation of study design, preparation of inclusion and exclusion criteria and revised the manuscript. Author MYPMY assisted in data analysis, interpretation and statistical analysis, and revised the manuscript. All authors read and approved the final manuscript.

Funding

GPK Research Grant, Universiti Teknologi MARA Malaysia [600-RMC/GPK 5/3 (188/2020)].

Availability of data and materials

The datasets generated and analyzed during the current study are not publicly available due to the security of data but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved by the institutional review board of the Universiti Teknologi MARA Research Ethics Committee (REC/380/19). Informed consent was waived due to the nature of the retrospective study.

Consent for publication

Not applicable.

Competing interests

The authors had no conflicts of interest to declare in relation to this article.

Author details

¹Center for Oral and Maxillofacial Diagnostics and Medicine Studies, Faculty of Dentistry, Universiti Teknologi MARA, Sungai Buloh Campus, 47000 Sungai Buloh, Selangor, Malaysia. ²Center for Integrated Systems Engineering and Advanced Technologies (INTEGRA), Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor, Malaysia. ³Center for Restorative Dentistry Studies, Faculty of Dentistry, Universiti Teknologi MARA, Sungai Buloh Campus, 47000 Sungai Buloh, Selangor, Malaysia. ⁴Institute of Pathology, Laboratory and Forensic Medicine (I-PPerForM), Universiti Teknologi MARA, Sungai Buloh Campus, 47000 Sungai Buloh, Selangor, Malaysia.

Received: 18 February 2022 Accepted: 4 April 2022

Published online: 08 April 2022

References

1. Panchbhai A. Dental radiographic indicators, a key to age estimation. *Dentomaxillofacial Radiology*. 2011;40(4):199–212.
2. Ciapparelli L. The chronology of dental development and age assessment. *Pract Forensic Odontol*. 1992;22–42.
3. George GJ, Chatra L, Shenoy P, Veena K, Prabhu RV, Kumar LV. Age determination by schour and massler method: a forensic study. *Int J Forensic Odontol*. 2018;3(1):36.
4. Demirjian A, Goldstein H, Tanner J. A new system of dental age assessment. *Hum Biol*. 1973;211–227.
5. Demirjian A, Goldstein H. New systems for dental maturity based on seven and four teeth. *Ann Hum Biol*. 1976;3(5):411–21.
6. Ismail AF, Othman A, Mustafa NS, Kashmoola MA, Mustafa BE, Mohd Yusof MYP. Accuracy of different dental age assessment methods to determine chronological age among malay children. *J Phys Conf Ser*. 2018;1028:012102.
7. Tanner JM. A new system for estimating skeletal maturity from the hand and wrist, with standards derived from a study of 2600 healthy British children. Part II: the scoring system; 1959.
8. Muad AM, Bahaman NSM, Hussain A, Yusof MYPM. Tooth segmentation using dynamic programming-gradient inverse coefficient of variation. *Bull Electric Eng Inf*. 2019;8(1):253–60.
9. Tuan TM. Dental segmentation from X-ray images using semi-supervised fuzzy clustering with spatial constraints. *Eng Appl Artif Intell*. 2017;59:186–95.
10. Kumar A, Bhadauria H, Singh A. Semi-supervised OTSU based hyperbolic tangent Gaussian kernel fuzzy C-mean clustering for dental radiographs segmentation. *Multimed Tools Appl*. 2020;79(3):2745–68.
11. Fan Y, Beare R, Matthews H, Schneider P, Kilpatrick N, Clement J, Claes P, Penington A, Adamson C. Marker-based watershed transform method for fully automatic mandibular segmentation from CBCT images. *Dentomaxillofacial Radiol*. 2019;48(2):20180261.
12. Ihya R. Segmentation of tooth using watershed transform and region merging. *J Theor Appl Inf Technol*. 2019;97(24).
13. Mohammad N, Yusof M, Ahmad R, Muad A. Region-based segmentation and classification of mandibular first molar tooth based on Demirjian's method. *J Phys Conf Ser*. 2020;2020:012046.
14. De Tobel J, Radesh P, Vandermeulen D, Thevissen PW. An automated technique to stage lower third molar development on panoramic radiographs for age estimation: a pilot study. *J Forensic Odontostomatol*. 2017;35(2):42.
15. Merdieto Boedi R, Banar N, De Tobel J, Bertels J, Vandermeulen D, Thevissen PW. Effect of lower third molar segmentations on automated tooth development staging using a convolutional neural network. *J Forensic Sci*. 2020;65(2):481–6.
16. Lee J-H, Han S-S, Kim YH, Lee C, Kim I. Application of a fully deep convolutional neural network to the automation of tooth segmentation

- on panoramic radiographs. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2019;129(6):635–42.
17. Kahaki SM, Nordin MJ, Ahmad NS, Arzoky M, Ismail W. Deep convolutional neural network designed for age assessment based on orthopantomography data. *Neural Comput Appl*. 2019;32:1–12.
18. Tafrount C, Galić I, Franchi A, Fanton L, Cameriere R. Third molar maturity index for indicating the legal adult age in southeastern France. *Forensic Sci Int*. 2019;294:218–e1.
19. Mohd Yusof MY, Cauwels R, Martens L. Stages in third molar development and eruption to estimate the 18-year threshold Malay juvenile. *Arch Oral Biol*. 2015;60(10):1571–6.
20. Franco A, Vetter F, Coimbra EF, Fernandes A, Thevissen P. Comparing third molar root development staging in panoramic radiography, extracted teeth, and cone beam computed tomography. *Int J Legal Med*. 2020;134(1):347–53.
21. Mohd Yusof MYP, Wan Mokhtar I, Rajasekharan S, Overholser R, Martens L. Performance of Willem's dental age estimation method in children: a systematic review and meta-analysis. *Forensic Sci Int*. 2017;280:245–e1.
22. Jayaraman J, Wong HM, King NM, Roberts GJ. The French-Canadian data set of Demirjian for dental age estimation: a systematic review and meta-analysis. *J Forensic Leg Med*. 2013;20(5):373–81.
23. Mohammad N, Muad AM, Ahmad R, Mohd Yusof MYP. Reclassification of Demirjian's mandibular premolars staging for age estimation based on semi-automated segmentation of deep convolutional neural network. *Forensic Imaging*. 2021;24:200440.
24. Trakinienė G, Andriuškevičiūtė I, Šalomskienė L, Vasiliauskas A, Trakinis T, Šidlauskas A. Genetic and environmental influences on third molar root mineralization. *Arch Oral Biol*. 2019;98:220–5.
25. Acton ST, Ray N. Biomedical image analysis: segmentation. *Synth Lect Image Video Multimed Process*. 2009;4(1):1–108.
26. Amini AA, Weymouth TE, Jain RC. Using dynamic programming for solving variational problems in vision. *IEEE Trans Pattern Anal Mach Intell*. 1990;12(9):855–67.
27. Ray N, Acton ST, Zhang H. Seeing through clutter: snake computation with dynamic programming for particle segmentation. In: *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*. IEEE; 2012, pp. 801–804.
28. Banar N, Bertels J, Laurent F, Boedi RM, De Tobel J, Thevissen P, Vandermeulen D. Towards fully automated third molar development staging in panoramic radiographs. *Int J Legal Med*. 2020;134:1–11.
29. Payal P, Goyani MM. A comprehensive study on face recognition: methods and challenges. *Imaging Sci J*. 2020;68(2):114–27.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

