Age Estimation based on Dental Radiographs using Hybrid CNN

M.Sc. Muaamar Khamees Mijwil [1]
Muamaraltaee78@gmail.com

Assist. Prof. Zainab Mohammad Hussain [2]
zainabmohammadhussain@aliraqia.edu.iq

Abstract: Nowadays, computer vision had been inserted on a large scale in multiple applications in several fields, such as age estimation using dental images. Age estimation based on Convolutional-Neural-Networks (CNN) is a crucial aspect of forensic dentistry, by analyzing the growth and development of teeth. Recently, different techniques based on CNN had been integrated with multiple approaches to extracting a various feature of human and used for age estimation. In this paper, a Convolutional Neural Network was developed according to two scenarios to estimate an age based on dental radiographs. In the first scenario, a CNN classifier is proposed for age estimation from dental images, in the second scenario, the CNN classifier is modified by replacing the last dense layer with a logistic regression classifier to enhance performance. The results showed that the effectiveness of the proposed method in improving the accuracy of age estimation, such that its improved to 95.15 % and the performance measures improved such that: the Precision by 1.83%, the Recall by 2.82% and the F1-score by 2.16%.

Keywords: Age Estimation, Convolutional-Neural-Networks, Dental Radiographs, dental images, logistic regression classifier.

1. Introduction

There are many factors that contribute to a person's identity, the most prominent being age. Many methods are relied upon to estimate the age of a person, such as skeletal analysis or dental analysis [1], where many doctors specializing in forensic dentistry can be found who are able to estimate the age of a person at death by studying the age of teeth. Teeth are one of the most flexible parts of the skeleton,

[1] M.Sc. Iraqi Ministry of Education-, Al-Mu'iniyah Intermediate School, Baghdad, Iraq
[2] Assist. Prof. Department of Political Science, College of Law and Political Science, Al- Iraqia University, Baghdad, Iraq

distinguished by their ability to loosen slowly and resist many stresses and various factors. Therefore, the teeth are of great importance in estimating the age of the humen at death. Human tooth development continues until the age of twenty, so it is a complex and time-consuming process, and teeth can be exposed to many developmental changes. During the growth and development of teeth, a person's age (up to puberty) can easily be estimated by observing the growth process, multiple dental radiography, as well as the radiographic technique of the third molar tooth. After the end of the tooth growth process and the emergence of the third molar, it becomes difficult to accurately determine the age of a person, especially in the elderly. The methods used in studying teeth can be divided into four different methods that can be used to estimate a person's age. The methods used in clinical situations to estimate the age of living subjects (adolescents and adults) are the morphological method, radiography, and the method used in euthanasia to estimate the age of deceased 3 persons is the histological method and the biochemical method. Dental radiographs are one of the surgical techniques that help in showing several morphological measurements of the tooth (P: pulp, R: root, and T: tooth) as shown in figure (1) [2].

Estimation of age from teeth age assessment methods are relatively simple and involve determination of mineralization stages on radiographs followed by comparison with the standard stage to estimate the approximate age range. The different radiographs that can be used in determining age are intraoral radiographs, lateral oblique radiographs, vertical radiographs, panoramic radiographs, digital imaging, and advanced imaging techniques [3].

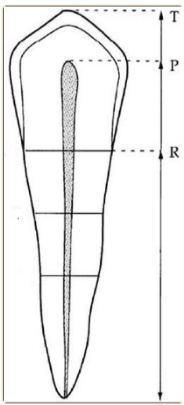

Figure (1) Tooth Morphological Measurement

In this paper, an age estimation based on Hybrid CNN is proposed. A CNN classifier is proposed for age estimation from dental radiographs images, and it's modified by replacing the last dense layer with a logistic regression classifier to enhance performance.

Dental radiographs are x-rays obtained from laboratories using specific equipment. These radiographs help obtain details about the shape of the teeth and bones, as well as identify the tissues that surround the teeth and jaws. Knowledge radiography can provide different measurements of teeth and molars, as shown in figure (2). On the other hand, radiography can help in identifying many diseases (caries, gum disease, bone disease, tumors), and it can also explain how teeth develop and identify any injuries and disorders in the process of development and growth of teeth [7]. Dental radiography gives the dentist a view of the hard tissues "teeth and bone" and also, soft tissues that surround the teeth and jawbones. Dental radiographs can alarm the dentist to changes in the hard and soft tissues. In children, radiographs allow the dentist to know how their teeth and jawbones are developing. Therefore, dental radiographs can be used to distinguishing diseases and growth problems before they become serious health problems. Further damage to other areas of the mouth can be limit or prevent with early detection of infection or injury.

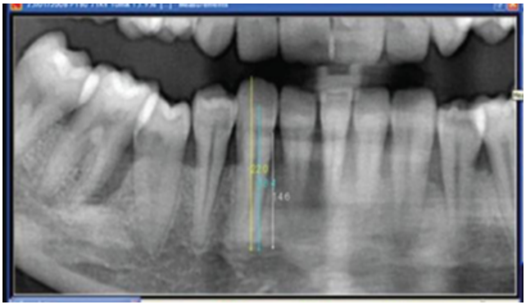

Figure (2) Dental radiographs image

Deep learning is data science that helps analyze large amounts of data quickly and easily in order to build predictive models or analytics. One of the deep learning algorithms is Convolutional Neural Networks (CNN) which is "a type of feed forward neural network, operating similarly to biological processes in the brain". It contains several layers including "convolutional layers" that are applied to the input or previous layer neurons and is the process of using the mathematical convolution function. Subsampling layers, which are "additional layers that come after convolutional layers, and are not necessary in the network design". These layers reduce the number of neurons, shorten them and remove unnecessary information. Finally, "fully connected" to the neurons of the previous layer, these layers are added at the end of the network, usually two successive layers[8]. When a dental radiographs image (input) is entered into a deep learning model, the features are extracted from the image (the important pixels in the image) using convolution and pooling layers. The pixel values are stored in an array of numbers "a two-dimensional (2D) grid", and an optimizable feature extractor, a small grid of parameters called "kernel", is applied at each image position, which makes CNNs highly efficient for image processing, since a feature may occur anywhere in the image. Then these pixels are entered into a classifier called full connected neural network in hidden layers, after that the image is classified into age value and showing the result using output layer as shown in figure (3) below [9].

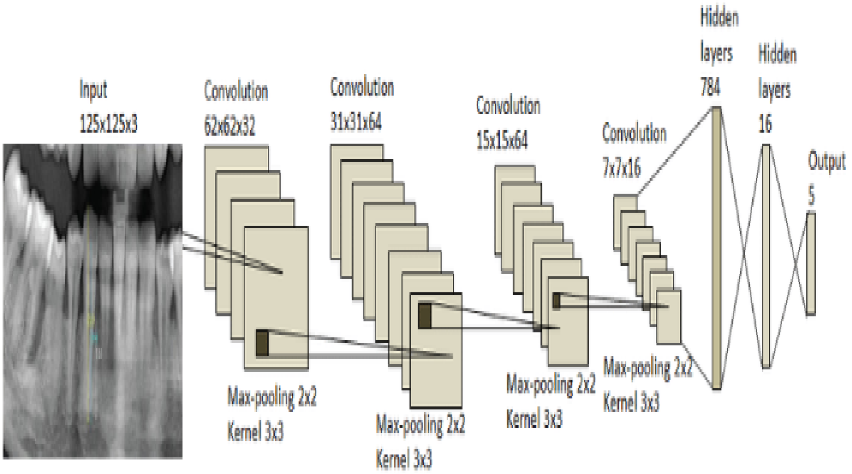

Figure (3) Working of CNN model.

A neural network is a collection of algorithms designed to detect trends and relationships in a set of training data. These algorithms are based on how neurons in the human brain process information. The neural network modifies the hyperparameters, weights (w), weightings (x), biases ( ), and (b) to satisfy the equation (1) during training (the process through which the model identifies the relationship between the training data and the output). In a process known as forward propagation, each training input is incorporated into the neural network. When the model generates an output, it is compared to the given target output in a process known as back propagation [10].

$$\hat{y} = \sigma(w^T x + b) \quad \ldots (1)$$

The loss function (along with the optimization function) is one of the most important aspects of neural networks, as it is directly involved in fitting the model to specific training data. A loss function compares the target value to the expected output value and assesses how effectively the neural network predicts the training data. During training, we try to minimize the difference between the expected and actual output. The weights, weightings, and biases, b that minimize the value of (j -average loss) and change the hyperparameters to minimize it. This can be thought of as a residual value in a statistic that measures the distance of the actual (y-value) from the regression line (the expected value) as shown in equation (2):

$$J(w^T, b) = \frac{1}{m} \sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)}) \quad \ldots (2)$$

One of the most common loss functions is the Mean Square Error (MSE), which finds "the average squared difference between the target and the expected output" as defined in [11] using the equation:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y^{(i)} - \hat{y}^{(i)})^2 \quad \ldots (3)$$

Since the difference is a square, it does not matter whether the expected value is higher or lower than the target value, but the values with large errors are penalized. A Hopper loss can be used using:

$$\text{Huber Loss} = \begin{cases} \frac{1}{n} \sum_{i=1}^{n} (y^{(i)} - \hat{y}^{(i)})^2 & |y^{(i)} - \hat{y}^{(i)}| \leq \delta \\ \frac{1}{n} \sum_{i=1}^{n} \delta(|y^{(i)} - \hat{y}^{(i)}| - \frac{1}{2}\delta) & |y^{(i)} - \hat{y}^{(i)}| > \delta \end{cases} \quad \ldots (4)$$

If the absolute difference between the actual value and the expected value is less than or equal to the threshold, the MSE applies. This is the "loss function" which is used in the "binary classification model" – a model that takes inputs and must classify them into one of 2 predefined categories. There are only two possible actual values for p-0 or 1 in binary classification. To accurately calculate the difference between the actual and expected values, it must compare the actual value (0 or 1) with the probability of the input alignment for that class (y(i) = p (i)).

$$CE\ Loss = \frac{1}{n}\sum_{i=1}^{N} - (y_i \cdot log(p_i) + (1 - y_i) \cdot log(1 - p_i) \quad ... (5)$$

If the number of classes is greater than 2, use class entropy - this follows a process very similar to binary entropy. Binary entropy is a special case of class entropy, where m = 2—the number of classes is 2 [11]:

$$CE\ Loss = -\frac{1}{M}\sum_{i=1}^{N}\sum_{i=1}^{M} y_i \,.\, \log(p_i) \quad ... (6)$$

The most related works to this work are:

Seyed Mostafa Mousavi Kahak, et al. in 2020 [4], 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 orthopantomography of 456 patients, which were captured in the Department of Dentistry and Research at University 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".

Sharifonnasabi F, et al. in 2022 [5], proposed "an ensemble method of image classifiers to enhance the accuracy of age estimation using dental radiographs Orthopantomography OPGs from 1 year to a couple of months (1-3-6). This hybrid model is based on convolutional neural networks (CNN) and K nearest neighbors (KNN). The hybrid (HCNN-KNN) model was used to investigate 1,922 panoramic dental radiographs of patients aged 15 to 23. These OPGs were obtained from the various teaching institutes and private dental clinics in Malaysia. To minimize the chance of overfitting in the model and eliminated the features with high correlation, the principal component analysis (PCA) algorithm was used. The innovative model, successfully estimated the age in classified studies of 1 year old, 6 months, 3 months and 1-month-old cases with accuracies of 99.98, 99.96, 99.87, and 98.78 respectively".

Arofi Kurniawan, et al. in 2025 [6], proposed "a convolutional neural network (CNN) algorithm trained on panoramic radiographs to develop an automated dental age estimation method based on the London Atlas of Tooth Development and Eruption. Material and Methods A dataset of 801 panoramic radiographs from outpatients aged 5 to 15 years was used. A CNN model for dental age estimation was developed using a 16-layer CNN architecture implemented in Python with TensorFlow and Scikit-learn, guided by the London Atlas of Tooth Development. The model included 6 convolutional layers for feature extraction, each followed by a pooling layer to reduce the spatial dimensions of the feature maps. A confusion matrix was used to evaluate key performance metrics, including accuracy, precision, recall, and F1 score. Results The proposed model achieved an overall accuracy, precision, recall, and F1 score of 74% on the validation set. The highest F1 scores were observed in the 10-year and 12-year age groups, indicating superior performance in these categories. In contrast, the 6-year age group demonstrated the highest misclassification rate, highlighting potential challenges in accurately estimating age in younger individuals".

The rest of this paper is organized into: Second part explains the research methodology. The results of this work illustrated in part three. Finally, the conclusion explained at the last part.

2. Research Methodology

This work is divided into two scenarios, in the first scenario, a CNN Classifier was proposed for age estimating from dental images, in the second scenario the proposed CNN classifier modified, with replacing the last dense layer by logistic regression classifier to improve the performance.

This research runs on the free Google Colab platform [12], with the following free resources: RAM = 12.7 GB, GPU = 15 GB, and Hard Disk = 78.2 GB, the research framework consists of four stages as shown in figure (4).

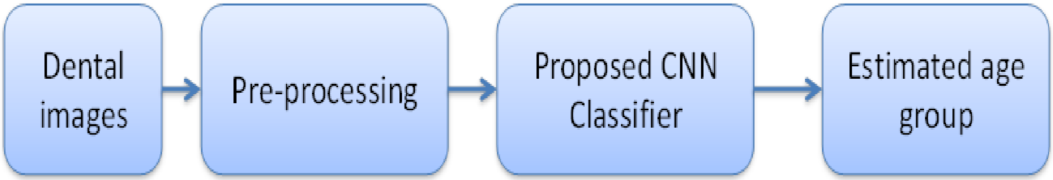

Figure (4) Research Framework

2.1. Dental Images Dataset

The data set from Google Colab platform [12] that used in this research contains 400 jpg images, and divided into 240 images as a "training set", 16 images as a "validation set", and 144 images as a "test set". The figure (5) shows some of the radiographs representing images of teeth at different ages.

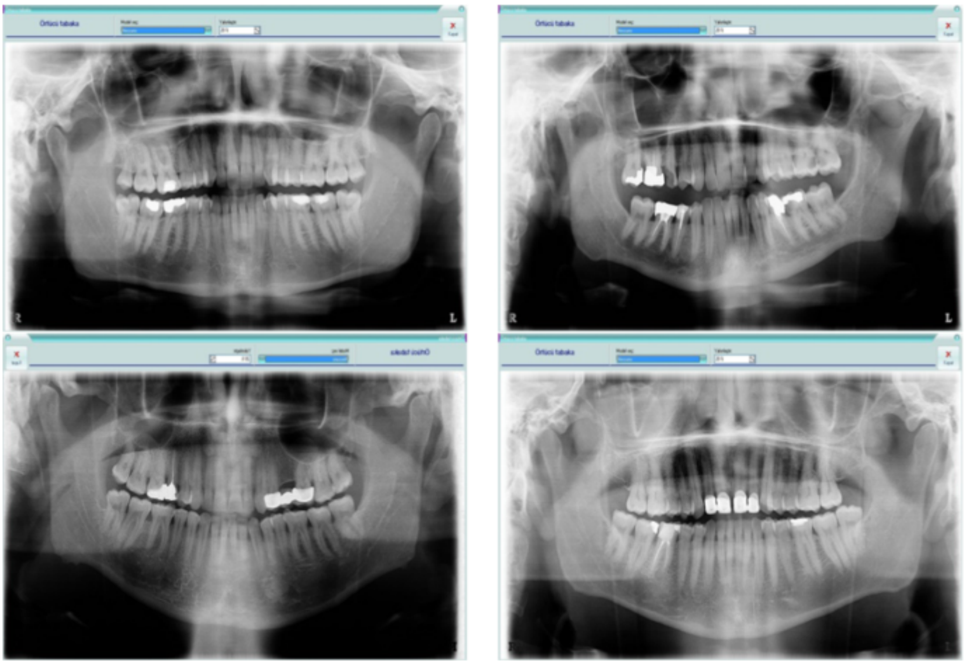

Figure (5) images from dataset

2.2. Pre-processing

Pre-processing operations done to prepare the images before training, first the images converted to gray-scale images as shown in figure (6), then re-sized to the shape of (112*92) to simplify and fasting the process. Before starting the training process, the training data augmented by performing random shuffling on the training set as show in figure (7), to ensures that the model does not learn any artificial patterns or biases from a specific order, which leads to a more generalized and robust model and enhancing robustness and feature extraction.

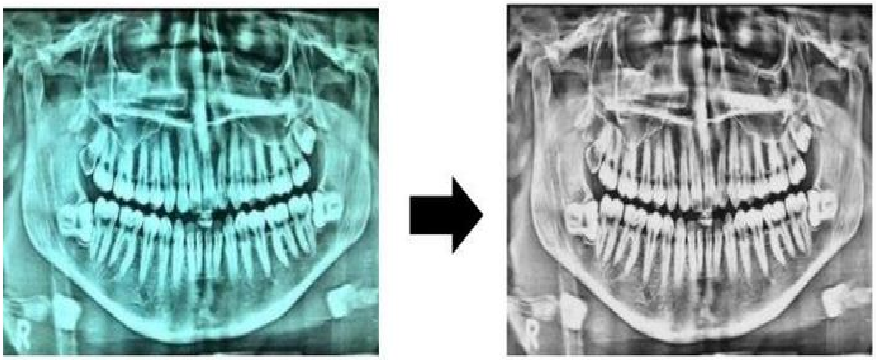

Figure (6) Convert RGB images to gray-scale

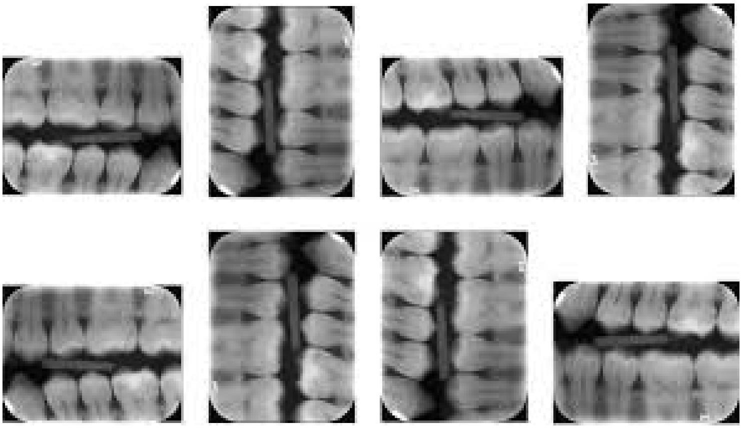

Figure (7) Image cropping, resizing and shuffling

2.3. Proposed Classifiers

Two models were proposed in this research: The first model is a CNN classifier built from scratch as shown in figure (8), and in the second model, is hybrid such that the last layer from the CNN classifier was replaced with a logistic regression classifier.

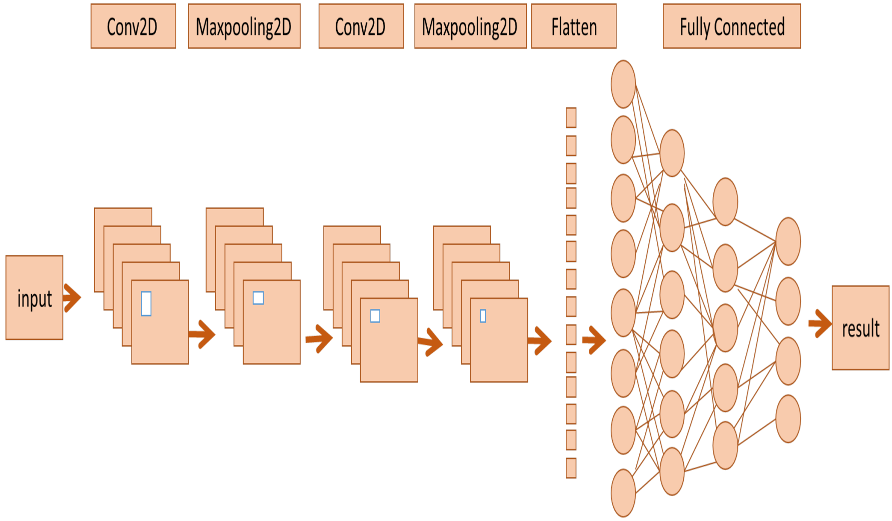

Figure (8) The proposed CNN Classifier

2.3.1. Build CNN from Scratch

The proposed classifier consists of two convolution layers for features extracting, two maxpooling layers to decrease network size, a flattening layer to convert features to vector, and five dense layers for recognition task, table (1) shows the parameter and the specification of this CNN. The proposed CNN is 32 Mb size, and all parameters trained successfully, the training time is 7.378s, and the testing time for all the testing set is 506ms, that means the proposed CNN is fast and accurate. The accuracy and loss curves for 20 epochs of the training stage shown in Figure (9) and Figure (10) respectively shows the good performance of this model.

Table (1): Proposed CNN Classifier

Model = Sequential			
Layer (Type)	Output Shape	Kernel	Parameters
Conv2d (Conv2D)	(None,112,92,32)	(5,5)	832
Maxpooling2D (Maxpooling2D)	(None,37,30,32)	(3,3)	0
Conv2d_1 (Conv2D)	(None,35,28,32)	(3,3)	9248
Maxpooling2D_1 (Maxpooling2D)	(None,17,14,32)	(2,2)	0

Flatten (Flatten)	-	-	
Dense (Dense, 1024)	-	-	7799808
Dense_1 (Dense, 512)	-	-	524800
Dense_2 (Dense, 256)	-	-	131328
Dense_3 (Dense, 128)	-	-	32896
Dense_4 (Dense, 20)	-	-	2580
Total Parameters	8501492 (32.43 MB)		
Trainable Parameters	8501492 (32.43 MB)		
Non-trainable Parameters	0 (0 MB)		
Training Time	7.378 s		
Testing Time	0.506 s		

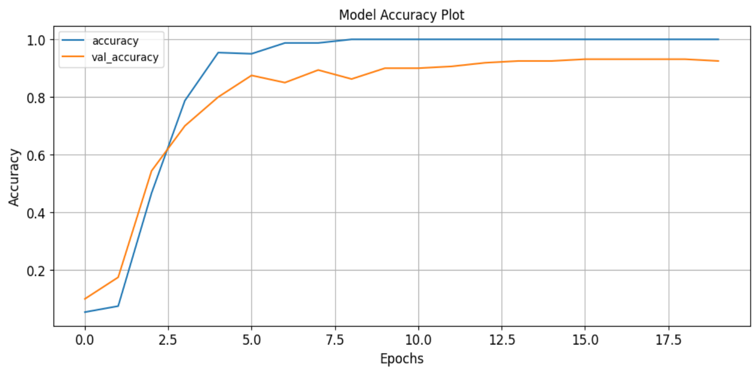

Figure (9) Proposed Model Training Accuracy

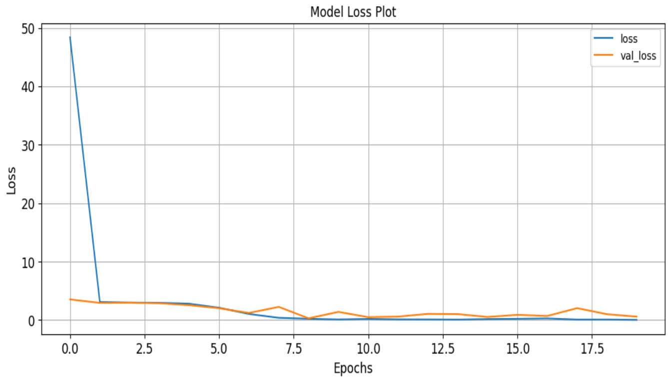

Figure (10) Proposed Model Training Loss

2.3.2 Build Hybrid Model

In this scenario, the logistic regression classifier used as the last layer in the proposed CNN network, so that its input is the output of the (dense_3) layer of the proposed CNN classifier and its output is the final estimation result, as shown in Figure (11).

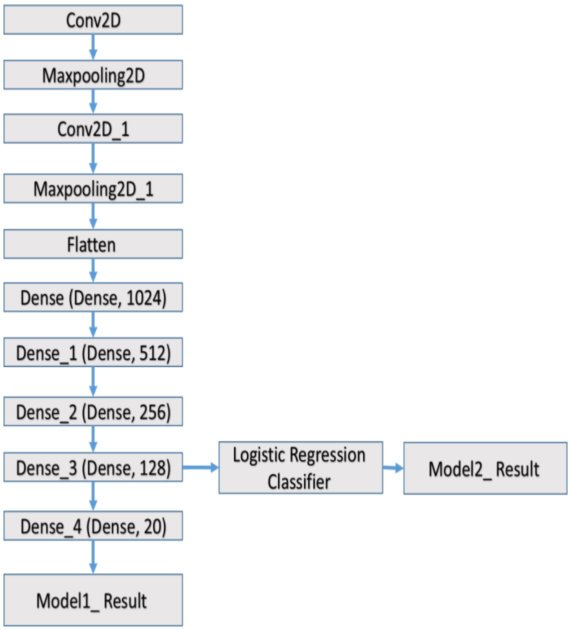

Figure (11) The steps of Second Proposed Classifier

3. Results

The performance of the two proposed CNN models applied for dental age estimation was evaluated using many key measures such as: "accuracy, precision, recall, and F1-score" derived from the "confusion matrix", which provides a detailed analysis of the proposed model's predictions by categorizing the outcomes. Accuracy indicates "the overall correctness of the model's predictions". The proposed system was tested on the testing set and reached a high value of accuracy (93.5%). Figure (12) shows the confusion matrix of the model, such that the true ages (True labels) on the y-axis and the predicted ages (Predicted labels) on the x-axis classified into four age group ((20-24), (25-29), (30-34), and (35-39)). From the confusion matrix it can be shown that the proposed model misclassified 10 times out of 144 times, this means the misclassified for the proposed model is (6.9%), which is a good value. Precision, Recall, and F1-score measures for the proposed model1 calculated for each class from the "confusion matrix", to obtain the good results shown in figure (13). The results were calculated based on a confusion matrix and it can be shown that the highest "precision" value is (95.5%) and highest "recall" value is (94.3%), were observed in the (3034) year age group and highest "F1-Score" value is (94.9%) observed in the (25-29) year age group. While the lowest "precision" value is (91.9%) observed in the (35-39) year age group and lowest "recall" value and "F1-Score" is (91.1%), were observed in the (2024) year age group.

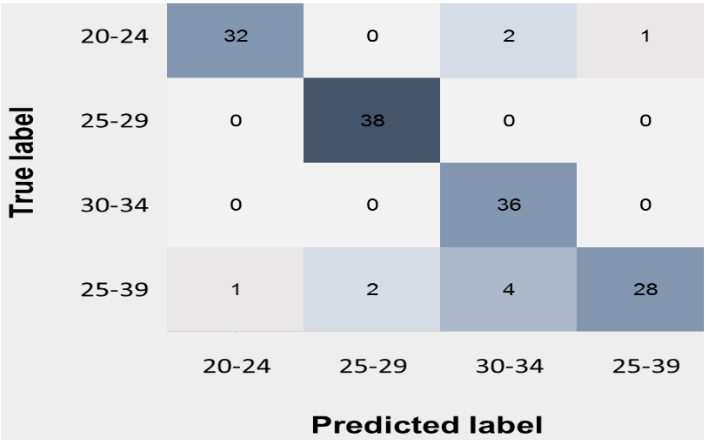

Figure (12) The Proposed Model Confusion Matrix

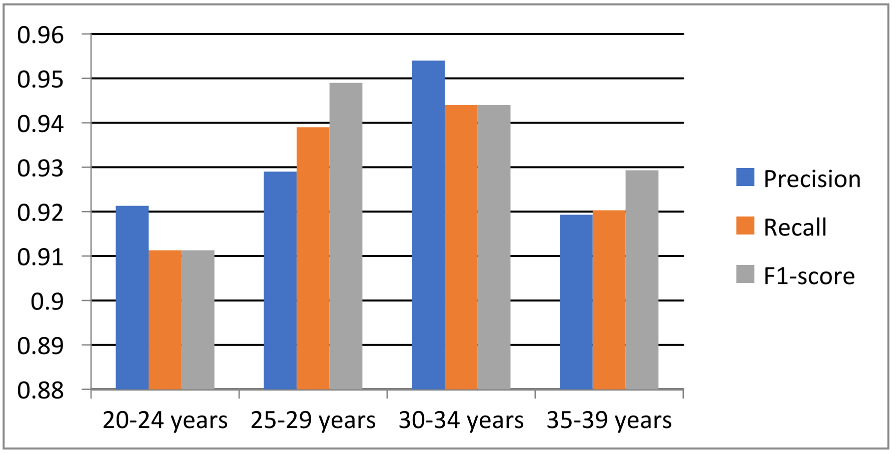

Figure (13) Precision, Recall, and F1-score for proposed model1

After training and testing the model2 proposed system, the confusion matrix that obtained shown in the figure (14). From this figure, it can be note that the second proposed model misclassified 7 times out of 144 times, this means the misclassified for the proposed model is (4.86%), which is means a less misclassified than model1 by (2.04%) such that the model2 is better performance. The measures, "Precision, Recall, and F1-score" were calculated for the proposed model2 for each class from the "confusion matrix" to obtain the higher results than model1 as illustrated in figure (15). The results were calculated based on a confusion matrix and it can be shown that the highest "precision" value is (96.1%) and highest "F1-Score" value is (96.3%), were observed in the (3539) year age group and highest "recall" value is (96.4%), observed in the (30-34) year age group. While the lowest "precision" value is (93.39%) observed in the (25-29) year age group and lowest "recall" value is (95%), observed in the (35-39) year age group and lowest "F1-Score" value is (94.2%), were observed in the (2024) year age group.

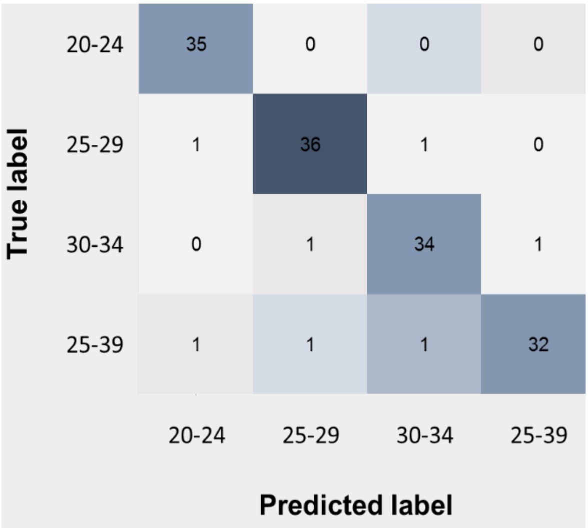

Figure (14) Confusion Matrix for a Proposed Model 2

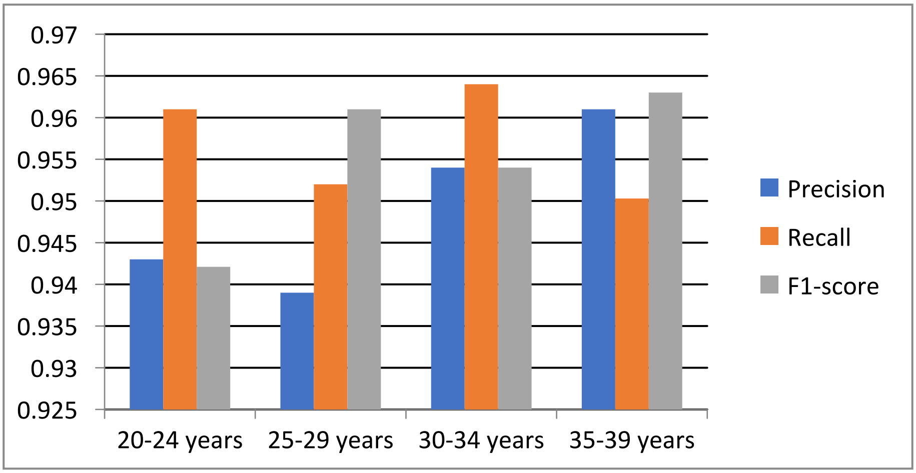

Figure (15) Precision, Recall, and F1-score for proposed model2

By comparing the two proposed models in the first and second scenarios applied on the same dataset using average of accuracy, Precision, Recall, and F1-score the results illustrated in figure (16). Figure (16) shows that, adding the logistical regression classifier instead of the last layer of the CNN network led to an improvement of:

- Accuracy by 2.1%.
- Precision by 1.83%.
- Recall by 2.82%.
- F1-score by 2.16%.

In this research, the proposed system achieved high accuracy (95.15%), while classifying age using dental images for 4 -age groups which is better than traditional CNN by 2.1%, and better than almost of other systems. These results can be summarized using table (2) to show the better measurement results of the proposed model2 over model1.

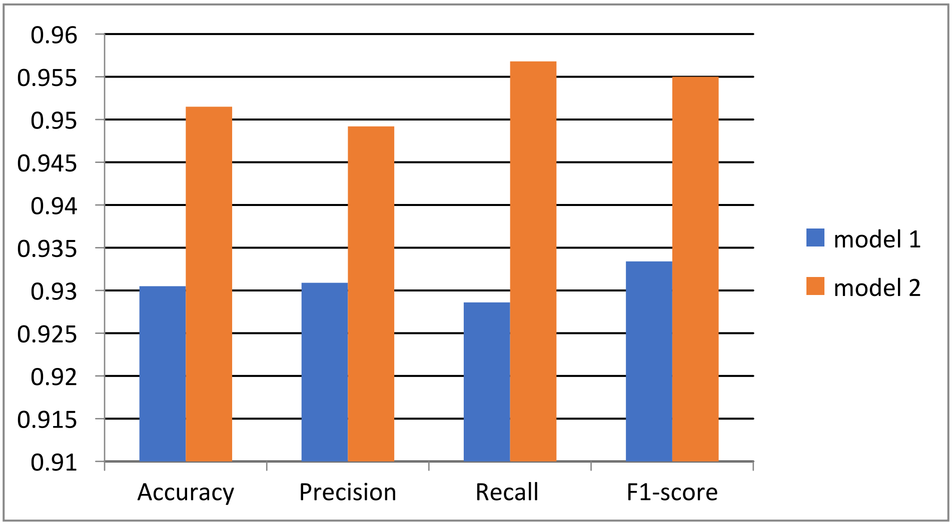

Figure (16) Performance Comparison between two Proposed Models

Table (2) Comparison results between Model1 and Model2

	Model1			Model2		
Age Group	precision	recall	F1-score	precision	recall	F1-score
20-24	92.1%	91.1%	91.1%	94.2%	96.1%	94.2%
25-29	92.9%	93.9%	94.9%	93.39%	95.1%	96.1%
30-34	95.5%	94.3%	94.9%	95.4%	96.4%	95.4%
35-39	91.9%	92%	92.9%	96.1%	95%	96.3%
Accuracy	93.5%			95.15%		
misclassified	6.9%			4.86%		

Also, for more precise, table (3) shows a comparison results with the three related studies mentioned previously in this paper. From table (3) it can be show that the second study it reached higher accuracy, but for the comparison to be fair, it must be done on the same data set and the same division ratio. Also, the proposed work gives better accuracy than the closest work the third study.

Table (3): Comparison Accuracy Results

Study	Task	Classifier	Accuracy Result
First [4]	Age Estimation/17 classes	DCNN	High Accuracy
Second [5]	Age Estimation/4 classes	HCNN-KNN	99%
Third [6]	Age Estimation/11 classes	CNN	74%
Proposed	Age Estimation/4 classes	CNN-LR	95.15%

4. Conclusions

In this research, two models were developed to estimate the age of a person from his teeth images classified into four groups.

The first model is a deep learning network CNN which reached an accuracy of 93.5%.

The second model is a hybrid deep-machine learning model, which uses a logistic regression classifier as the last layer of the proposed CNN model, and the accuracy was improved by 2.1% compared to the first model. Also, the misclassified rate is less by (2.04%), and the performance measures is improved such that the precision by 1.83%, the recall by 2.82%, and the F1-score by 2.16%. The proposed method gives a higher accuracy result as compared with the most relate work.

5. References:

[1] Mohammad N, Yusof MYPM, Ahmad R, and Muad AM., "Region-based segmentation and classification of mandibular first molar tooth based on Demirjian's method", J Phys, 2020.

[2] Kim S, Lee YH, Noh YK, Park FC, and Auh Q, "Age-group determination of living individuals using first molar images based on artificial intelligence", Sci Rep. 11:1–11, 2021.

[3] McGill D., "Bone Age. Endocrine Conditions in Pediatrics", Cham: Springer, 2021.

[4] Seyed Mostafa Mousavi Kahak, Md Jan Nordin, Nazatul Sabariah Ahmad, and Waidah Ismail, "Deep convolutional neural network designed for age assessment based on orthopantomography data, Neural Computing and Applications", July 2020.

[5] Fatemeh Sharifonnasabi, Noor Zaman Jhanjhi, Jacob John, Peyman Obeidy, Shahab S Band, Hamid Alinejad-Rokny, and Mohammed Baz, "Hybrid HCNN-KNN Model Enhances Age Estimation Accuracy in Orthopantomography", Frontiers in Public Health, Volume 10, 879418, May 2022.

[6] Arofi Kurniawan, Michael Saelung, Beta Novia Rizky, and Anand Marya, "Dental age estimation using a convolutional neural network algorithm on panoramic radiographs: A pilot study in Indonesia, Imaging Science in Dentistry", March 2025.

[7] Dahlberg PS, Mosdøl A, Ding Y, Bleka Ø, Rolseth V, Straumann GH, et al, "Systematic review of the agreement between chronological age and skeletal age based on the greulich and pyle atlas", Eur Radiol, 2019.

[8]Alkaabi S, Yussof S, and Al-Mulla S, "Evaluation of convolutional neural network based on dental images for age estimation", ICECTA. 2019.

[9] Ezhil I, Jagannathan N, and Kumar MP, "Estimation of age from physiological changes of teeth", Drug Invent Today,2018.

[10] B. Martinez, M. F. Valstar, B. Jiang, and M. Pantic, ''Automatic analysis of facial actions: A survey'', IEEE Trans. Affect. Compute., to be published, 2017.
[11] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, ''Local binary patterns and its application to facial image analysis: A survey'', IEEE Trans. Syst., Man, C, Appl. Rev., vol. 41, no. 6, pp. 765–781, Nov. 2011.
12. https://colab.research.google.com/

تقدير العمر بناءً على صور الأشعة السينية للأسنان باستخدام شبكة عصبية تلافيفيه هجينة

م . م . معمر خميس مجول[1]
Muamaraltaee78@gmail.com

أ. م . زينب محمد حسين[2]
zainabmohammadhussain@aliraqia.edu.iq

المستخلص: في الوقت الحاضر، باتت تقنيات رؤية الحاسوب تُستخدم على نطاق واسع في تطبيقات متعددة في مجالات متنوعة، مثل تقدير العمر باستخدام صور الأسنان. ويُعدّ تقدير العمر باستخدام الشبكات العصبية الالتفافية(CNN) جانبًا بالغ الأهمية في طب الأسنان الشرعي، وذلك من خلال تحليل نمو الأسنان وتطورها. وقد دُمجت مؤخرًا تقنيات مختلفة تعتمد على الشبكات العصبية الالتفافية مع مناهج متعددة لاستخلاص سمات بشرية متنوعة واستخدامها في تقدير العمر. في هذه الورقة، طُوّرت شبكة عصبية التفافية وفقًا لاثنين من السيناريوهات لتقدير العمر بناءً على صور الأشعة السينية للأسنان. في السيناريو الأول، يُقترح استخدام مُصنِّف شبكة عصبية التفافية لتقدير العمر من صور الأسنان، بينما في السيناريو الثاني، عُدِّل مُصنِّف الشبكة العصبية الالتفافية باستبدال الطبقة الكثيفة الأخيرة بمُصنِّف الانحدار اللوجستي لتحسين الأداء. أظهرت النتائج فعالية الطريقة المقترحة في تحسين دقة تقدير العمر، حيث بلغت 95.15%، كما تحسنت مؤشرات الأداء بنسبة 1.83% في الدقة، و2.82% في الاستدعاء، و2.16% في مقياس F1.

الكلمات المفتاحية: تقدير العمر، الشبكات العصبية الالتفافية، صور الأشعة السينية للأسنان ، صور الأسنان، مصنف الانحدار اللوجستي.

[1] ماجستير العلوم، وزارة التربية والتعليم العراقية - مدرسة المعينية المتوسطة، بغداد، العراق
[1] أستاذ مساعد، قسم العلوم السياسية، كلية القانون والعلوم السياسية، الجامعة العراقية، بغداد، العراق