

Age-Net: An Advanced Hybrid Deep Learning Model for Age Estimation Using Orthopantomograph Images



Merve Parlak Baydogan^{1*}, Sumeyye Cosgun Baybars², Seda Arslan Tuncer³

¹ Department of Computer Technologies, Technical Sciences Vocational School, Firat University, Elazig 23119, Turkey

² Faculty of Dentistry, Department of Oral and Maxillofacial Radiology, Firat University, Elazig 23119, Turkey

³ Faculty of Engineering, Department of Software Engineering, Firat University, Elazig 23119, Turkey

Corresponding Author Email: mpbaydogan@firat.edu.tr

<https://doi.org/10.18280/ts.400423>

ABSTRACT

Received: 31 January 2023

Revised: 21 May 2023

Accepted: 25 July 2023

Available online: 31 August 2023

Keywords:

age estimation, deep learning, dental orthopantomographic image, machine learning

Forensic odontology, recognized as a fundamental and reliable technique in human identification, frequently employs orthopantomograph images in dental biometry. Despite the introduction of various techniques for age and identity estimation, the accurate and rapid interpretation of these images remains challenging. Manual methods, currently employed by forensic experts, present numerous limitations including time consumption, human error, and challenges in handling large data sets. Addressing these limitations, this study proposes a computer-aided hybrid age detection system, Age-Net, leveraging artificial intelligence. A total of 933 orthopantomograph images, categorized into three classes, were collected from Firat University Hospital for this study. These images were subsequently resized to be compatible with pre-trained Convolutional Neural Networks (CNNs) models, such as AlexNet, ResNet50, VGG16, SqueezeNet, EfficientNetB0, DenseNet201, and ResNet18. Following the extraction of feature vectors from these images, algorithms including Naive Bayes (NB), K-Nearest Neighbor (KNN), Multilayer Perceptron (MLP), XGBoost Algorithm, Support Vector Machine (SVM), Decision Tree (DT), and Linear Discriminant (LD) were implemented for detection. The use of an array of feature extraction models and algorithms aimed to best represent the features of the dataset, thereby enhancing classification performance. The proposed system's efficacy was assessed and validated using the 5-fold cross-validation test technique and statistical Friedman test. Of all models, the EfficientNetB0-SVM hybrid model demonstrated superior performance, achieving highest accuracy, precision, sensitivity, F-score, and AUC ratios of 0.846, 0.850, 0.846, 0.846, and 0.970, respectively. This hybrid detection system, Age-Net, is projected to provide time and cost benefits to forensic experts in their clinical studies. However, the data utilized in this study are limited to specific age groups. Future research could expand the number of age groups and data to observe potential enhancements in the system's success rate.

1. INTRODUCTION

Throughout the determination of human identity has a significant place in collective disaster and criminal investigations. According to the 2014 Interpol Disaster Victim Identification Guide, friction ridge analysis, DNA analysis, and forensic odontology are the fundamental and most confident techniques of recognition [1]. One of the important issues in forensic medicine applications is the demand for age determination for cases whose identity is unknown and whose age is suspected [2]. Additionally, age determination is also a necessary examination in anthropology, forensic medicine, pediatrics, and orthopedics. The person's age; is related to the physical characteristics of the individual's medical identity such as gender, height, body weight, hair, skin feature, eye color, fingerprints, bone tissue, and teeth. In cases where the soft tissues of the human body are damaged, it has been seen that the use of dental features in the identification phase is the most reliable method since the teeth are the hardest tissue of the human body [3].

Methods applied in age estimation are grouped into three groups: radiological, morphological, and histological.

Radiological and morphological methods are employed most frequently in forensic medicine [4]. Medical imaging techniques, which develop in parallel with technology, can also indicate bone and tooth structures. In this way, it contributes to forensic science both as evidence and as an important and valid criterion in forensic cases. Orthopantomography imaging technique is a type of radiological imaging that allows imaging of jaw bones, tooth roots, and teeth with a single film. Orthopantomography images ensure diverse and dissimilar biometric features for recognition, such as the contours of the teeth, and the relationship between dental restorations and adjacent teeth [5]. Although there are some age estimation studies utilizing orthopantomography images, accurate and rapid identification from these images remains a challenging issue. Current methods used in age determination require expert opinion. In addition, different expert opinions affect the consistency of the results of the study. These manual methods are also prone to human error. These traditional methods have the disadvantage that they are tiring and time-consuming. This problem increases the need for cost and manpower. Finally, it does not have a fast and efficient decision-making process in big data.

However, artificial intelligence-based methods offer more objective approaches. In addition, producing fast and effective results necessitated the orientation to these systems.

Recently, medical images such as x-ray images [6, 7], computed tomography [8, 9] and magnetic resonance imaging [10, 11] have been trained with CNNs. CNNs offer promising results in biomedical classification and segmentation studies [12, 13]. Previous research has shown that the success results of CNNs can compete with those of trained radiologists or clinicians. CNNs have high accuracy, efficiency, and processing speed especially for these tasks [14]. Therefore, it has also been stated that it can be utilized as a successful technique in clinical medicine [15-17] and forensic age estimation [18, 19].

An orthopantomographic images present developmental and physiological changes such as the appearance of tooth germs, the appearance of mineralization, the condition of the erupted teeth, the degree of enamel formation, the degree of root development of erupted teeth, the degree of root resorption, crown attrition, physiological secondary dentin formation, cementum formation, and root dentin transparency. These features can be listed as the main criteria that can be utilized to determine age [20].

Tooth eruption is a developmental and active process in the period from the first formation of primary and permanent teeth in the alveolar bone to the building occlusal contact with the opposing tooth [21]. From the moment the tooth begins to bud, it moves within the bone. Then it is formed occlusal contact. These changes are controlled by cellular, molecular and genetic mechanisms that control temporal and spatial differences [22]. In this process, teeth go through various mineralization and calcification stages, and these stages can also be detected in orthopantomography images. In the 6th month after birth, primary incisors are the first group of teeth to appear in the mouth. This is followed by other m primary ilk teeth at six-month intervals, and all primary teeth are visible in the mouth around 36 months [23]. Permanent dentition (approximately six years of age) begins after the eruption of the primary teeth at 3-3.5 years. The permanent teeth continue to complete their formation (around the age of 20) until the third permanent molars are terminated. The end and beginning of the dentition periods provide considerable information on age estimation. Especially six years of age, which is the beginning of the permanent dentition period, and 14 years, which is considered to be the end of the complex dentition period in infancy and childhood, are reliable periods used in age estimation [24]. Age estimation may be performed to 20, which is the end of the completion period of the formation of the third molar teeth. However, in some cases, third molars may erupt late or not erupt and remain impacted. For this reason, reliable results are obtained in age estimation until about 20 years of age, which is the end of the completion period of tooth formation. Estimates of adult age applied after this age period are less reliable [25].

Numerous studies including histological, morphological, and radiographic methods have been performed for estimating dental age in adult and non-adult individuals [26]. Orthopantomography films are one of the most frequently used methods in routine clinical practice of dentistry, diagnosis, and treatment planning. Although orthopantomography images allow only two-dimensional images of 3-dimensional structures to be examined, low radiation dose and rapid image acquisition compared to

advanced imaging methods are the most main reasons for their widespread use. This imaging technique provides a lot of information about the teeth in the maxilla and mandible with a single panoramic film. In addition, panoramic images have been frequently utilized in determining bone and tooth age in recent years, since they are easy to obtain and have common use [27]. The first group of the data set collected in this study consists of images of the age of six when permanent dentition begins. The second group is the period up to the age of 14, which is the time of eruption of the permanent number 7 teeth, which is also considered to be the end of the childhood period and the mixed dentition period at the same time. The third group is the 20-year-old period when permanent tooth formation is largely completed. In the proposed system, orthopantomograph images belonging to these 3 groups were examined. In the study, seven different feature extraction architectures based on deep learning were employed. Then, the results of seven different classifier algorithms are presented for each of the feature extraction architectures. In addition, various performance evaluation criteria were calculated to observe the performance of this proposed hybrid system. Finally, the results are compared with tables and graphs.

The x-ray dental image data set used in our study was collected for the first time within the scope of this study. With this dataset, a deep learning-based hybrid autonomous system for age determination was applied. This autonomous system provided many advantages over manual methods in terms of time and cost, error rate and reliability, processing speed, and efficiency, control, and management. Considering these differences, autonomous systems can generally offer higher processing speed, efficiency, and reliability. In this study, a hybrid system that provides automatic age estimation from dental images used in many fields such as forensics, anthropology, and international adoption was proposed. Finally, this hybrid system is based on state-of-the-art deep learning architectures and artificial intelligence techniques. The most prominent innovations of the study can be highlighted as follows:

- Since the teeth undergo some changes depending on age, it was once again accepted as a shred of reliable evidence that can be applied for age estimation.
- In this study, instead of performing the age determination process manually, an automatic, reliable, and objective hybrid system was designed using artificial intelligence based on state-of-the-art methods.
- The proposed hybrid system provided advantages in terms of time, cost, and accuracy, unlike manual methods.
- With this study, a new data set with a different grouping structure for age determination was applied for the first time in the literature. It is estimated that this hybrid system will shed light and guide the future of many different medical and forensic disciplines, thanks to the multiple feature extraction models and classification algorithms applied in the study.

The other parts of the study were organized as follows. In the second part of the study, a literature review was conducted for studies that perform age estimations with dental data. In the third part, the features of the data set employed in the study, feature extraction methods, and features of deep learning algorithms were listed. In the fourth chapter, the performances of algorithms were assessed together with performance evaluation. In the last section, information about the contribution of the study to the literature and future studies was presented.

2. RELATED WORKS

In this section, the studies related to age estimation in the literature were scrutinized in detail. In addition, the data sets used, the methods applied, and the performance criteria obtained in these studies were evaluated. These studies are listed as follows.

In the literature review, it was seen that more than one method and different evaluation criteria were used to determine the age from dental data. In the literature, some studies have been carried out by looking at the eruption/development of the tooth and the pulp-tooth volume ratio while determining the age. With these evaluation criteria, artificial intelligence-based autonomous systems have also been designed.

Some studies related to tooth eruption/development are as follows:

Age determination from teeth was examined by Gustafson as two separate main periods. These were primary teeth and permanent teeth. For temporary teeth, it was necessary to rely on the evaluation by microscopy. In the period over the age of 14, a continuous development continues in the jawbone and dentition with the loss of primary teeth, eruption, mineralization, and formation. However, permanent teeth also receive shape, mineralize, erupt, and then change. According to Gustafson, age determination in this period is generally performed by comparing the radiographs with the dentition development schema or table. In addition to, permanent teeth, only the development of the 3rd molar tooth can offer information about age in the period between the ages of 14-20, when all teeth are fully erupted or developed. While teeth are considered to be the most reliable age parameter until adulthood, it has been suggested that it is not reliable in the assessment of age after tooth eruption is complete. The use of structural changes in the tooth for age determination was suggested by Gustafson [28].

In another study, Hägg and Taranger argued that when measuring dental age, using a short-period parameter like tooth eruption is better than using a longer-period parameter like tooth formation, which is finished in 6-25 months [29]. Bernhard and Glocker found that there was no change in the development of permanent dentition for several centuries in their study on a population of girls aged 5-13 years [30]. This result revealed that tooth eruption in humans is primarily genetically determined [31]. Townsend and Hammel represented that eruption time in children develops independently of environment-related factors and provides a more accurate and objective estimation of age as well as the conclusions of Demirjian's technique, today's reference work [32, 33].

Some studies based on pulp-tooth volume ratio are as follows:

Yang et al. [34] aimed to appraise the ratio of pulp-to-tooth volume utilizing Cone Beam Computed Tomography (CBCT) images, and they employed CBCT to perform this using less radiation compared to microCT. In their study, CBCT images of 15 incisors, 12 canines, and 1 premolar tooth were used. The volumes of the pulps and teeth that they segmented from the images were compared with tooth volumes and the pulp that they measured in the laboratory with the Archimedes principle of the two extracted teeth. With the threshold value obtained, they demonstrated the accuracy and reliability of the segmentation. In addition, it was determined that there was a linear regression between the pulp-tooth volume ratios

obtained and the chronological age, with $R^2=0.29$, and the deviation between the chronological age and the predicted age was ± 8.3 years.

Aboshi et al. [35] conducted a study estimating the age of persons from a micro-CT image of the lower premolar teeth of a Japanese population. In this study, pulp-tooth volume ratios were measured in micro-CT images in 4 different sections: the crown of the tooth, the coronal third of the root, the middle third of the root, and the apical third of the root. According to the results, the highest correlation with age was obtained in the pulp-tooth volume ratios in the coronal third of the root ($R=0.81$ for the lower first premolars).

Some studies based on artificial intelligence are:

Fan et al. [36] proposed a system for human identification (DENT-net) in their study. In this study, panoramic dental radiographs (PDR) were trained by utilizing CNN. A total of 15,868 PDRs were collected from 6473 patients to train and test of CNN. During the training of the system, 15,369 PDRs from 6300 people were used. In the test phase, the remaining 499 PDRs from 173 people were employed. At least two PDRs were received from each person at various periods. In the test set of the study were obtained Rank-1 and Rank-5 accuracy of 85.16% and 97.74%, respectively.

Because it is difficult to segment dental x-ray images Tuan et al. [37] proposed SSFC-FS, a new semi-supervised fuzzy clustering algorithm based on Interactive Fuzzy Satisfaction. This new algorithm was applied to a dataset of 56 dental x-rays obtained from the Medical University of Hanoi in Vietnam during the 2014-2015 period. This new method was then validated experimentally and compared with the relevant ones in terms of clustering quality. The findings set out that the new approach outperforms existing clustering algorithms including Fuzzy C-Means, Otsu, eSFCM, SSCMOO, FMMBIS, and another variant of SSFC-FS called SSFC-SC that uses the native Lagrange technique.

Another study on gender determination from dental x-ray images was conducted by Ilić et al. [38]. A data set containing 4155 images was collected in the study. While 60% of the dataset was composed of female individuals, 40% was male individuals. After feature extraction with VGG16, a 10-fold cross-validation technique was applied. Mean accuracy of 94.3% was yielded across age groups in this study [38]. Oktay [39] conducted a study for the detection of 3 teeth (anterior, premolar, and molar) from tooth x-ray images. In the study, 100 x-ray images were classified with CNN. An accuracy value of over 90% was obtained.

Another study by Liu [40], a study was carried out to state chronological age from Lateral Cephalometric Radiographs. 20,174 Lateral Cephalometric Radiograph was collected in the study. The classification was performed by dividing the data set into different age groups. EfficientNet-B0 was applied as the feature extraction architecture. At all ages, EfficientNetB0 calculated the mean absolute error (MAE) and standard deviation of the absolute error (SD) to be 1.3 and 2.24 respectively [40].

A study on age determination was conducted by Sathyavathi and Baskaran [41]. In the study, 3508 Orthopantomogram (OPG) images collected from individuals aged 10-30 were used. 80% of the total dataset was set as training data. The remaining 20% was selected as test data. With the ResNet and Sequential CNN classification models, 91% and 93% accuracy were achieved, respectively.

In this study, age estimation was performed by taking into account the deficiencies in other studies in the literature. In the

proposed study, the methods used for age determination were processed in a more comprehensive and detailed way than other studies mentioned in the literature. Unlike other studies, our study was classified by considering the changes in eruption times of teeth. In addition, our proposed hybrid system, age detection from dental images was achieved with state-of-the-art seven different deep learning architectures and seven different algorithms for three different age groups. Moreover, promising results were yielded with this advanced hybrid system. In the next section, detailed information about the pipeline of this system is presented.

3. MATERIALS AND METHODS

It is known that the proposed hybrid autonomous system is applied effectively in solving many classification problems in many aspects such as time, cost, high performance, consistency, and minimum error margin. In addition, the performance of the proposed system is also highlighted in Section 4.

In this part of the study, the pipeline of the proposed system is presented, including the data characteristic, feature extraction, model development, classification, and assessment process.

3.1 Dataset

The data set operating in the proposed hybrid system was allocated from Elazig Firat University, Faculty of Dentistry, Department of Oral, Dental, and Maxillofacial Radiology. In addition, the protocol for the data to be used in the study was approved by the Non-Interventional Research Ethics Committee of Firat University (Date: April 2022). Each participant signed their written informed consent in order for their information to be utilized in the study. Images were acquired from Planmeca Promax (00880, Helsinki, FINLAND) at standard settings. The collected data set was labeled by the dentist.

Data or panoramic radiographs; It was classified by considering the primary dentition period (2-6 age) in which only primary teeth are present, the mixed dentition period (6-13 age), which begins with the eruption of permanent teeth, and the age of 13, which is the end of the mixed dentition and the beginning of the permanent dentition.

The data set was separated into three different age groups. The data set consists of a total of 933 dental x-ray images. Information about the data set was offered in Table 1. Examples of data groups are demonstrated in Figure 1. In addition, it is seen in Figure 2. that the classes have a balanced distribution.

Since data belonging to only one gender group would cause bias in the generalization of the results of the study, data from both male and female individuals were used in the data collection process. Also, labeling errors or missing labeling in the dataset can distort results. Therefore, the data labeling process was fine-tuned to minimize potential physician errors.

Table 1. Numerical distribution of the classes in the data set used in age estimation

Data Groups	Count
2-6 Age	315
6-13 Age	316
13-21 Age	302
Total	933

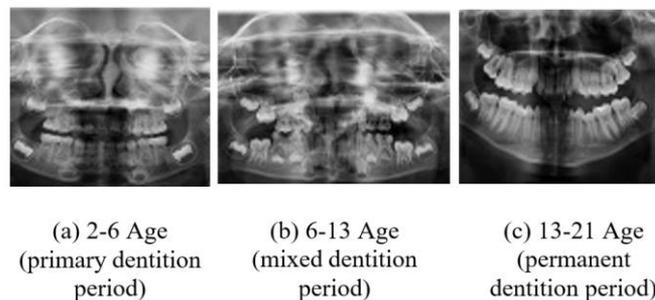


Figure 1. Example images of classes in the data set collected in age estimation

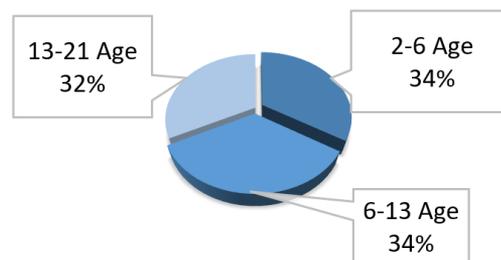


Figure 2. Circular chart showing the distribution of classes

3.2 Structure of proposed system

CNN is a sort of artificial neural network built for many tasks such as image processing and recognition. CNNs produce more accurate and promising results, particularly with multidimensional image data. Thanks to CNNs, the distinctive feature map of the images in the data set is directly extracted. In addition, through the CNN architecture built, both the new network can be trained and the previously trained network can be operated. With this feature of CNNs, operations are executed easier and faster. CNN architectures generate feature maps using a certain number of filters. CNN architectures generally consist of a pooling layer, convolutional layer, and linked layer. The fully connected layer is employed to classify or score the input data depending on extracted features from the previous layers. Compared to other models, EfficientNetB0 is designed as an efficient model and makes better use of computational resources than others. In addition, it attracts attention with its smaller size compared to other models and has the ability to exhibit high performance. It is also quite successful in obtaining better performance results with better generalization and less training data.

NB, KNN, SVM, DT, MLP, XGBoost, and LD algorithms applied in the developed system are the most preferred algorithms based on machine learning for many different tasks. They are the most reliable techniques, especially in classification tasks and detection systems. They are also known as methods used in many studies in the classification of medical images. Briefly, NB is a classifier based on Bayes' theorem. KNN, on the other hand, is an algorithm that decides according to the k nearest neighbor values. SVM has a structure that allows separating classes with the line it creates. DT and XGBoost are an algorithm that provides tree-based diagnostics. MLP is a basic model of neural network. Finally, LD is a classifier that provides linear distinction. The parameters of these algorithms are fixed with Table 2 values.

The proposed hybrid system created with CNN architectures consists of 6 stages. First of all, the data were collected according to certain standards and the labelling

process was carried out by the dentist. In addition, the process of labelling the dataset is also part of the feature extraction process, classifying it according to certain parameters. This classification process is done depending on the purpose of the study. In this study, classification was made by looking at the tooth eruption parameters of individuals aged 2-21. In the second stage of the study, each image taken from the data set was pre-processed and given to the system as input for feature extraction models. Feature extraction can improve the intelligibility, computation time, and performance of the designed model. Therefore, in the third stage of the study, seven different CNN architectures were applied to create the feature map. In the fourth stage, the feature maps obtained in

the third stage were used as the input for the classification algorithms. Comparing different classification methods is important to determine the best-performing method on the data set. Therefore, after the feature maps were extracted, seven different classification algorithms were applied for age determination in the fourth stage of the study. In addition, a 5-fold cross-validation technique was applied for the performance evaluation of CNN architectures and classification algorithms in the study. In the fifth step, the classification results of each algorithm were obtained. In the last stage, the highest-performance hybrid model was emphasized by calculating performance metrics. The general structure of the proposed system is shown in Figure 3.

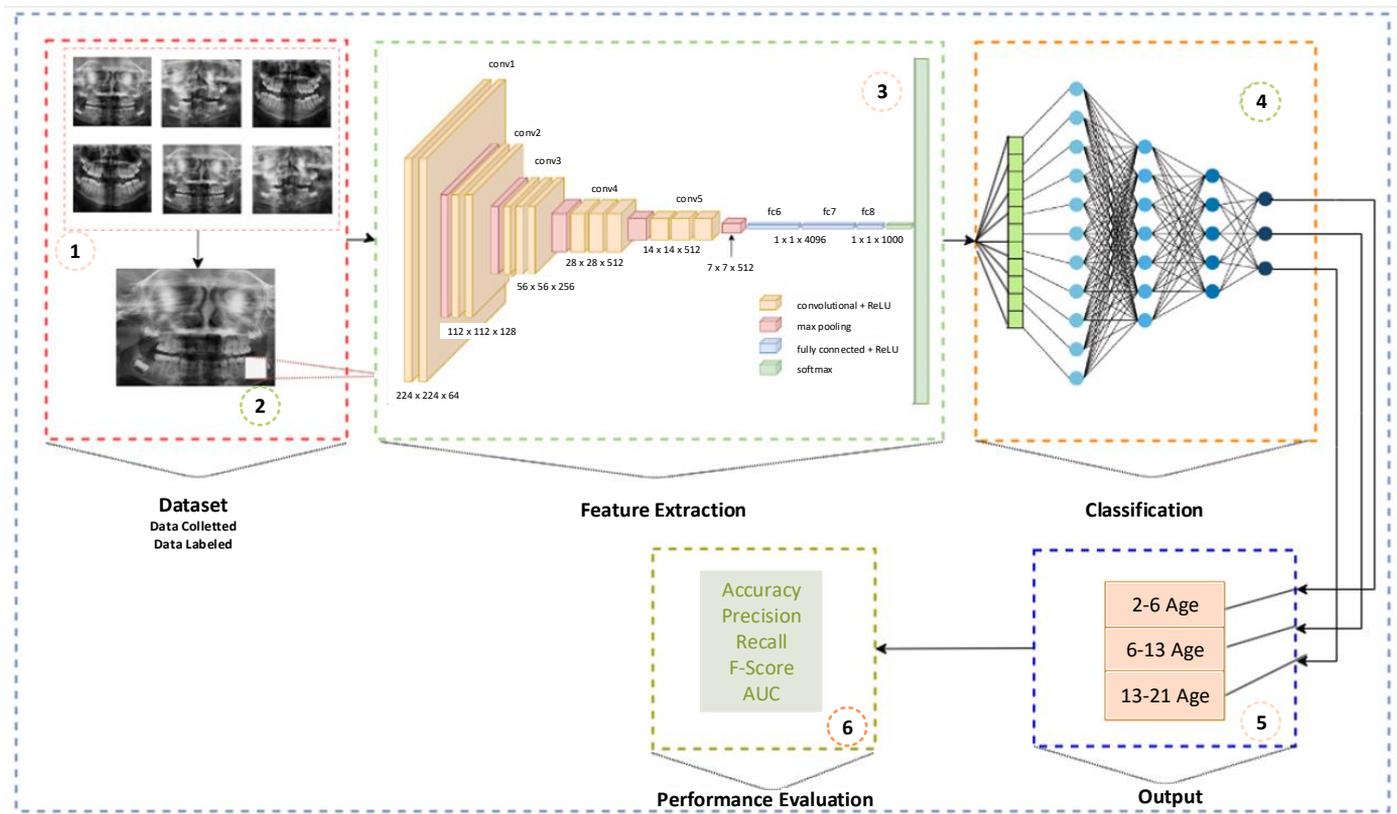


Figure 3. Structure of the proposed decision support system

Table 2. Parameter values of classifiers

Applied Classifiers	Parameters
KNN	k value: 5, distance metric: Minkowski, distance weight: equal, other parameters: default
LD	threshold: 1.0e-4, solver: svd, other parameters: default
SVM	kernel: linear, kernel scale: automatic, gamma: scale, degree: 3, C parameter: 1, other parameters: default
DT	maximum number of splits: 100, split criterion: gini, splitter: best, other parameters: default
MLP	hidden_layer_sizes: 100, fully connected layers: 1, solver: adam, activation: relu, iteration: 200, lambda: 0, epsilon: 1e-8, other parameters: default
XGBoost	type: decision tree, maximum number of splits: 20, number of learners: 30, learning rate: 0.1, other parameters: default
NB	predictors: gaussian, variance: 1e-9, other parameters: default

Different classification algorithms may perform differently for different datasets and problems. It is possible for one algorithm to perform better or worse than others. Therefore, it is important to try different algorithms and compare their performance to obtain the best classification results. For this reason, seven different classifiers were used to measure the performance of the proposed system in this study. The parameters of the applied classifiers are listed in Table 2.

The performance of a neural network is affected by the size and depth of the network and the preference for different parameters. The combination of different neural networks and fine-tuning of parameters in the network affect the success of the model positively or negatively. In this study, seven different architectures, namely Alexnet, EfficientNetB0, VGG16, ResNet18, ResNet50, SqueezeNet and Densenet201, known as pre-trained networks and whose performances have been proven on different problems, were applied.

In order to ensure the reliability of the proposed system, the 5-fold cross-validation test technique was operated on the system. Thus, the deviations and errors caused by the distribution and fragmentation of the data were minimized.

The features and parameters of the suggested and best-performing EfficientNetB0 model for the system are listed in Table 3 [40].

Table 3. The structure of EfficientNetB0

Stage	Operator	Resolution	#Channels	#Layers
i	\hat{F}_i	$\hat{H}_i \times \hat{W}_i$	\hat{C}_i	\hat{L}_i
1	Conv3x3	224x224	32	1
2	MBCConv1,k3x3	112x112	16	1
3	MBCConv6,k3x3	112x112	24	2
4	MBCConv6,k5x3	56x56	24	2
5	MBCConv6,k3x3	28x28	80	3
6	MBCConv6,k5x5	14x14	112	3
7	MBCConv6,k5x5	14x14	192	4
8	MBCConv6,k3x3	7x7	320	1
9	Conv1x1 & Pooling & FC	7x7	1280	1

3.3 Assessment criteria

The confusion matrix is a schema that shows the count of correctly and incorrectly classified data groups in a data set [42]. In this study, accuracy, precision, recall, f-score, and AUC values were calculated for the assessment of classification performance. The basic definition of these parameters is listed below [43].

- True Positive (TP): Orthopantomograph images that are classified as true when actually true.
- True Negative (TN): Orthopantomograph images that are classified as false when actually false.
- False Positive (FP): Orthopantomograph images that are classified as true when actually false.
- False Negative (FN): Orthopantomograph images that are classified as false when actually true.

Mathematically, the performance evaluation metrics applied are calculated by Eqs. (1)-(4), respectively. Also, performance metrics of confusion matrix are illustrated in Figure 4.

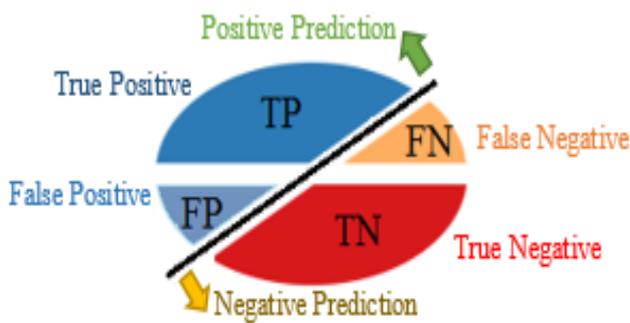


Figure 4. Confusion matrix

$$Accuracy (Acc) = \left(\frac{TP + TN}{(TP + FP + FN + TN)} \right) \quad (1)$$

$$Precision (P) = \left(\frac{TP}{(TP + FP)} \right) \quad (2)$$

$$Recall(R) = \frac{TP}{(TP + FN)} \quad (3)$$

$$F - Score (F1) = \left(\frac{2DP}{(2DP + YP + YN)} \right) \quad (4)$$

Another performance evaluation criterion Receiver Operating Characteristic (ROC) curves are applied to test classifier performance. Methods of classification aim to balance sensitivity and precision. Briefly, the ratio of sensitivity to precision generates the ROC curve. In other words, the curve is utilized to assess the balance between precision and sensitivity. Also ROC curve shows both discrimination and the AUC in the performance comparison of distinct tests. The value of the field under the ROC curve generates the AUC ratio. When this value approaches 1, it means that the positives are efficiently divided from the negatives.

4. EXPERIMENTAL STUDY

The results of the system developed for determining age from orthopantomograph images are presented in Table 4. In order to create Table 4, a feature map was first created with seven different models. Then SVM, DT, KNN, NB, MLP, XGBoost, and LD algorithms were run separately for each model. A different complexity matrix was generated for each run experiment. Using the 5-fold cross-validation test technique, approximately 186 tooth images for each fold on the data set were used for testing and the remaining images were used for training. Accuracy, precision, sensitivity, f-score, and AUC values were calculated for each experiment. The highest outcomes of each architecture are indicated and highlighted in bold on the table. All experiments were performed under equal conditions and with Table 2 parameters.

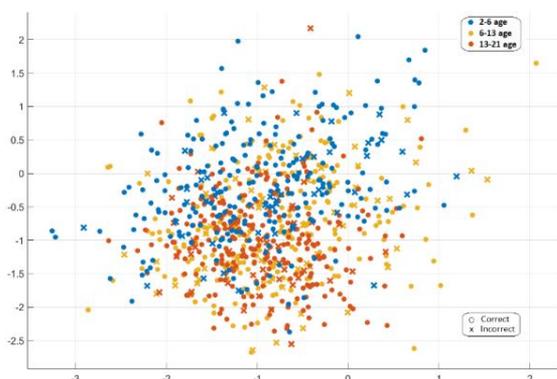
The SVM algorithm outperformed other classifiers with the best results for all pre-trained models based on CNN. It is clearly seen that the highest accuracy, sensitivity, recall, f-score, and AUC ratios are achieved by the hybridization of EfficientNetB0 architecture and SVM algorithm. It was observed that the accuracy performance of this hybrid system, which performs age estimation from orthopantomograph images, was approximately 85%. The highest precision, recall and f-score values were also obtained from the EfficientNetB0 and SVM duo. The highest AUC value of 97% was achieved by ResNet18, ResNet50 and EfficientNetB0. The ResNet50 model and the SVM duo finished second in this performance race. Then, ResNet18, Densenet201, VGG16, SqueezeNet and AlexNet models took place in the performance, respectively.

In evaluating the performance of the classification algorithms, the SVM algorithm prevailed in all experiments. Except for the EfficientNetB0 and KNN duo, the LD algorithm always finished second in the performance assessment. LD, NB and KNN algorithms have obtained similar results in other experiments. This ranking was similar for other performance evaluation criteria. DT algorithm produced the worst performance in all experiments. Parameter optimization can be applied to improve the performance of this algorithm.

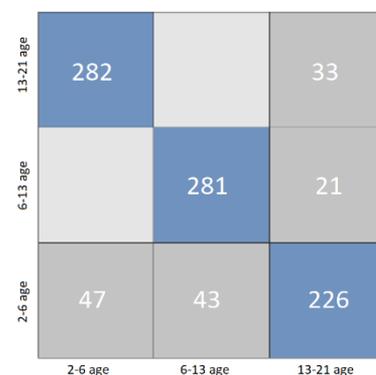
Figure 5 is presented the data distribution, complexity matrix, and ROC curve for the best-performing architecture and algorithm. The closer the ROC curve is to the upper left corner, the higher the positive ratio of the right and the larger the area under the curve. From this, it can be seen whether the positives are successfully divided from the negatives. When Figure 5 (c) is examined, it is seen that the SVM algorithm performs a more successful distinction with a value of 0.97 AUC than other models. The performance assessment results of all architectures and classifiers operated are summarized in Figure 6.

Table 4. Accuracy, Precision, Recall, F-Score, AUC, and standard deviation (σ) performance values of algorithms

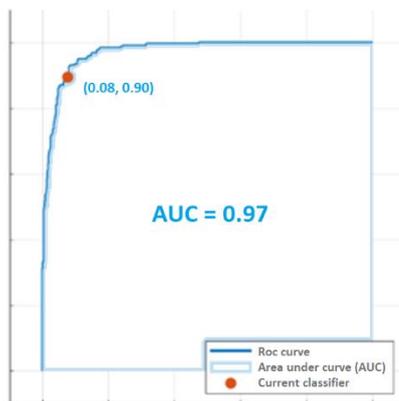
Models	Algs.	Accuracy Values for Each K-Fold					Performance Assessment Criteria					σ (%)
		k1	k2	k3	k4	k5	Acc	P	R	F1	AUC	
AlexNet	KNN	0.770	0.738	0.781	0.753	0.726	0.750	0.753	0.746	0.746	0.910	± 2.013
	LD	0.717	0.561	0.706	0.710	0.500	0.796	0.800	0.796	0.800	0.940	± 9.035
	SVM	0.765	0.765	0.813	0.817	0.801	0.811	0.810	0.806	0.810	0.960	± 2.299
	DT	0.695	0.684	0.695	0.688	0.597	0.634	0.683	0.686	0.686	0.860	± 3.782
	MLP	0.813	0.733	0.824	0.806	0.812	0.769	0.766	0.763	0.763	0.900	± 3.289
	XGBoost	0.845	0.743	0.791	0.780	0.780	0.723	0.723	0.616	0.723	0.910	± 3.283
	NB	0.770	0.743	0.738	0.688	0.737	0.733	0.736	0.723	0.726	0.890	± 2.646
ResNet50	KNN	0.770	0.786	0.754	0.769	0.747	0.778	0.780	0.776	0.773	0.940	± 1.355
	LD	0.738	0.583	0.706	0.645	0.570	0.788	0.790	0.790	0.790	0.940	± 6.602
	SVM	0.797	0.759	0.786	0.780	0.796	0.839	0.840	0.836	0.836	0.970	± 1.364
	DT	0.652	0.679	0.717	0.694	0.656	0.713	0.716	0.716	0.716	0.870	± 2.393
	MLP	0.813	0.802	0.850	0.817	0.812	0.802	0.803	0.800	0.800	0.890	± 1.646
	XGBoost	0.834	0.807	0.807	0.790	0.812	0.736	0.737	0.730	0.737	0.920	± 1.407
	NB	0.775	0.754	0.775	0.677	0.731	0.751	0.753	0.743	0.740	0.890	± 3.650
VGG16	KNN	0.749	0.695	0.674	0.737	0.731	0.784	0.786	0.786	0.786	0.940	± 2.803
	LD	0.690	0.519	0.668	0.618	0.516	0.806	0.810	0.810	0.810	0.910	± 7.308
	SVM	0.797	0.765	0.717	0.769	0.780	0.822	0.823	0.820	0.820	0.950	± 2.677
	DT	0.658	0.663	0.674	0.699	0.715	0.752	0.753	0.726	0.750	0.890	± 2.187
	MLP	0.797	0.759	0.786	0.780	0.785	0.750	0.750	0.747	0.750	0.870	± 1.233
	XGBoost	0.840	0.738	0.797	0.790	0.796	0.711	0.713	0.700	0.703	0.910	± 3.232
	NB	0.759	0.674	0.668	0.710	0.667	0.763	0.766	0.760	0.760	0.890	± 3.553
SqueezeNet	KNN	0.727	0.759	0.742	0.758	0.689	0.753	0.757	0.750	0.750	0.920	± 2.558
	LD	0.647	0.551	0.652	0.651	0.554	0.495	0.497	0.487	0.483	0.700	± 4.791
	SVM	0.802	0.743	0.824	0.817	0.758	0.814	0.817	0.810	0.813	0.960	± 3.226
	DT	0.701	0.642	0.701	0.656	0.613	0.659	0.660	0.657	0.660	0.800	± 3.414
	MLP	0.781	0.749	0.818	0.769	0.677	0.771	0.773	0.767	0.770	0.880	± 4.655
	XGBoost	0.813	0.759	0.802	0.790	0.767	0.713	0.713	0.713	0.713	0.920	± 2.001
	NB	0.733	0.695	0.770	0.742	0.731	0.729	0.730	0.720	0.723	0.860	± 2.400
EfficientNetB0	KNN	0.770	0.727	0.802	0.812	0.790	0.800	0.803	0.800	0.793	0.960	± 2.996
	LD	0.668	0.556	0.668	0.699	0.457	0.771	0.770	0.773	0.773	0.920	± 9.061
	SVM	0.845	0.791	0.845	0.796	0.763	0.846	0.850	0.846	0.846	0.970	± 3.205
	DT	0.690	0.652	0.733	0.683	0.704	0.695	0.696	0.690	0.696	0.870	± 2.630
	MLP	0.807	0.807	0.845	0.812	0.780	0.793	0.793	0.790	0.790	0.900	± 2.079
	XGBoost	0.802	0.775	0.850	0.844	0.780	0.744	0.743	0.743	0.740	0.920	± 3.152
	NB	0.765	0.711	0.749	0.677	0.710	0.723	0.726	0.710	0.713	0.880	± 3.095
DenseNet 201	KNN	0.759	0.738	0.738	0.731	0.704	0.762	0.763	0.756	0.760	0.940	± 1.769
	LD	0.610	0.567	0.652	0.575	0.543	0.804	0.806	0.803	0.803	0.950	± 3.805
	SVM	0.802	0.770	0.791	0.780	0.796	0.824	0.826	0.820	0.823	0.960	± 1.152
	DT	0.674	0.626	0.663	0.565	0.624	0.677	0.680	0.670	0.670	0.860	± 3.837
	MLP	0.834	0.781	0.802	0.817	0.774	0.758	0.757	0.757	0.757	0.870	± 2.233
	XGBoost	0.840	0.791	0.797	0.796	0.780	0.721	0.720	0.713	0.717	0.910	± 2.041
	NB	0.733	0.717	0.706	0.677	0.704	0.794	0.793	0.793	0.793	0.920	± 1.806
ResNet18	KNN	0.743	0.695	0.786	0.688	0.726	0.758	0.760	0.753	0.746	0.940	± 3.543
	LD	0.733	0.674	0.668	0.656	0.672	0.798	0.796	0.800	0.800	0.950	± 2.677
	SVM	0.781	0.759	0.791	0.753	0.758	0.831	0.833	0.830	0.830	0.970	± 1.496
	DT	0.658	0.610	0.701	0.624	0.640	0.689	0.693	0.686	0.692	0.850	± 3.154
	MLP	0.765	0.807	0.829	0.780	0.817	0.778	0.776	0.780	0.767	0.880	± 2.387
	XGBoost	0.781	0.775	0.813	0.742	0.801	0.736	0.740	0.737	0.737	0.920	± 2.435
	NB	0.754	0.770	0.765	0.640	0.769	0.739	0.743	0.736	0.733	0.870	± 5.016



(a) EfficientNetB0-SVM scatter plot



(b) Confusion matrix of EfficientNetB0-SVM



(c) ROC curve of EfficientNetB0-SVM

Figure 5. Highest results achieved by the EfficientNetB0-SVM

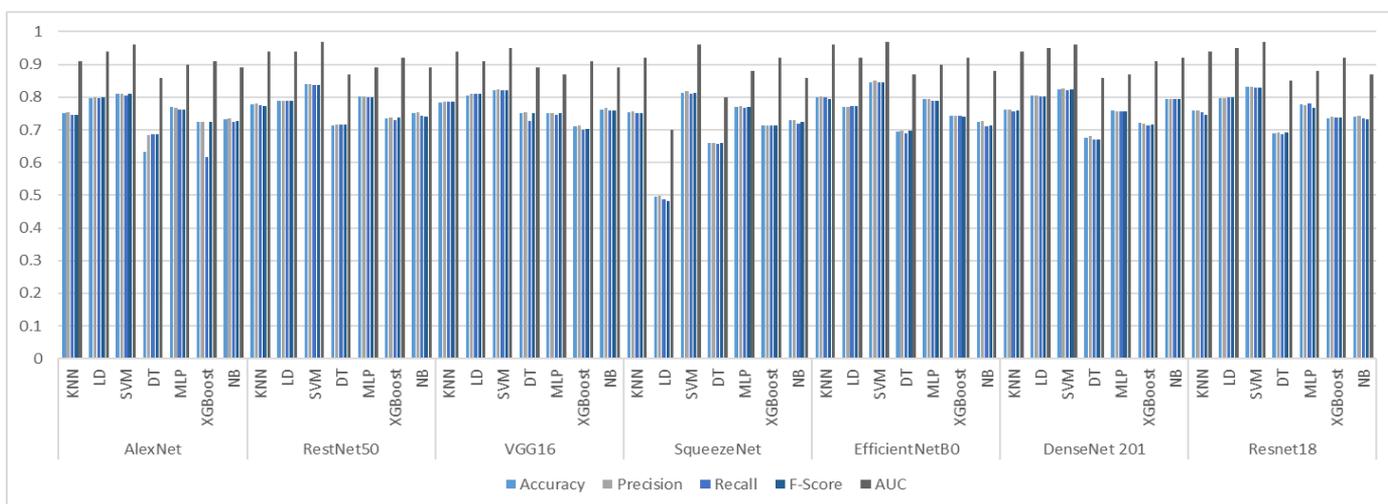


Figure 6. General performance comparison of proposed system

Friedman test, which is one of the non-parametric test techniques, was used to verify the reliability of 7 different CNN-based models used in this study. In the Friedman test, the values of two or more interrelated variables are compared and it is questioned whether there is a significant difference between them. In other words, it is a statistical test technique used to test whether there is a difference between repeated values on distributions. H_0 and H_1 hypotheses must be determined for the execution of the Friedman test. The hypotheses defined for seven different models used in the age determination system are defined as follows:

- H_0 : There is no statistically significant difference in the results obtained from 7 different models used for age determination.
- H_1 : There is a statistically significant difference on the results obtained from 7 different models used for age determination.

$p=0.115$ value was determined in the Friedman test. The n value represents the accuracy value in each fold obtained as a result of 5-fold cross validation. The degrees of freedom (fd) were calculated as $fd=k-1$. In this study, $fd=6$ was accepted as $k=7$ (7 different CNN models). With reference to p and fd values, the limit value was taken as 7.841 in the chi-square distribution table. If the calculated value is greater than the limit value (7.841) in the light of the statistical results obtained, it is accepted as H_1 . Otherwise, if it is less than the limit value,

the H_0 hypothesis is accepted. It is aimed to show statistically the values of the models used in the study with the box plot technique in Figure 7. Statistical results obtained at the end of this test are listed in Table 5 and Table 6.

Whether the difference was significant on 7 different models used was examined with the Friedman test. The average ranking values obtained as a result of the Friedman test of the model values are listed in Table 5.

The chi-square value was calculated for these models using the mean ranking values in this table. The Friedman test results of the models are listed in Table 6.

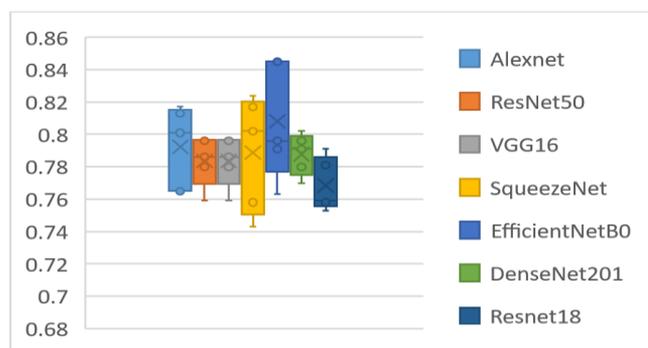


Figure 7. Statistical box plot representation of the applied models

Table 5. Mean ranking values of models

Models	Mean Rank
Alexnet	4.90
ResNet50	3.20
VGG16	3.20
SqueezeNet	4.10
EfficientNetB0	5.80
DenseNet 201	4.60
ResNet18	2.20

Table 6. Friedman test result of models

Parameters	Values
<i>n</i>	5
Chi-Square	10.237
<i>fd</i>	6
<i>p</i> -value	0.115

Considering these results, since the *p* value is less than 0.25 and the chi-square value is greater than the chi-square limit value calculated as 10.237, the H0 hypothesis will be rejected and the H1 hypothesis will be accepted. In other words, at least one of the 7 different models used is different from the others. As a result, it is understood that there is a statistically significant difference on the results of the applied models.

5. DISCUSSIONS AND CONCLUSIONS

Changes in the tooth are a significant diagnosis factor in the assessment of the developmental age of the individual. Generally, judicial authorities request forensic medicine experts to perform age determinations for the solution of many legal and social problems. This study offers a deep learning-based solution to the estimation and classification problems of age determination with orthopantomograph images. Orthopantomograph images were committed as input to CNN models to categorize age groups. The feature vectors acquired from the pooling layer of the CNN models were classified utilizing SVM, KNN, LD, NB, MLP, XGBoost, and DT. The conclusions showed that the suggested method accomplished a high level of classification accuracy. State-of-the-art CNN models and algorithms are applied and the results were presented as comparisons. It was determined that the designed deep learning models have an accuracy rate of 81%-84% in general.

The most distinctive aspects of the study are listed as follows:

- It has been seen that teeth can be used for diagnostic purposes because they show some changes depending on age.
- Age determination in forensic medicine is a very significant issue in terms of punishment and law. Therefore, identifications performed based on anatomical features and life-long changes in the organism should be based on objective evidence with the least margin of error. The purpose of the study is to perform the age determination operation automatically, not manually, but through computer-based systems. Thus, a reliable and objective system was designed for age estimation.
- The suggested system provides advantages in time, accuracy, and cost, unlike manual methods.
- A data set that has not been used before in the literature is used. Data duplication was not performed in this study; only raw images from patients were used.

- In order to increase the reliability and accuracy of the study, more than one solution was sought by using multiple feature extraction architectures and classification algorithms. It is foreseen that it will contribute to the literature by comparing these results.

- Since the proposed method is developed with orthopantomograph images without human error, it can be a standard solution as it is an automatic evaluation tool.

6. RESULTS AND RECOMMENDATIONS

The proposed study is to present an objective and alternative method that can help forensic experts to determine age. The study consists of four different stages. In the first stage, the effect of the change in the eruption time of the teeth in determining the age was observed and the data set was created with orthopantomograph images. The data set consists of three different age groups. In the second stage, feature maps of orthopantomograph images were obtained with seven different CNN architectures. In the third stage, classification was performed with seven different classification algorithms according to the values in the feature map. 5-fold cross-validation was used for the validation of the classifier. In the last stage, the performances of the classification algorithms were evaluated. The performances of the architectures used in the study were compared with accuracy, precision, recall, *f*-score, and AUC values. The highest performance was achieved with the EfficientNetB0 architecture and the SVM algorithm. The proposed system presents new findings and results with the new data set. It also contributes to the literature by referencing previous studies. Although there are studies on age determination in the literature, no studies have been conducted with these age groups before. The proposed hybrid autonomous system provides many benefits for forensics, law enforcement or other related fields. Some of those; autonomous systems are capable of quickly performing repetitive or time-consuming tasks without human intervention. In addition, these systems can produce more consistent and accurate results by reducing human errors. However, there are some limitations in the study.

- The data used in the study are limited to certain age groups. The success rate of the system can be observed by increasing the number of age groups.

- The study consists of three classes. The performance of the results can be increased by further increasing the number of classes.

- In the study, seven feature extraction models and seven different classification algorithms were used. By increasing the number of these models and methods, the performance evaluation of the study can be improved.

- The number of data in the study is limited due to the lack of equipment and personnel. Though the findings of the study were satisfactory, better accuracy values could have been obtained with more data.

- Setting the correct parameter is a time-consuming and experience-requiring process. Therefore, in the study, default parameters were used in the performance evaluation phase. In future studies, the results can be compared by setting the default parameter.

- In future studies, the results can be evaluated by estimating the age of the individuals whose dental images are taken according to their gender.

The results obtained from the study are promising and it has

been determined that the designed deep learning model plays an effective role in determining age. Moreover, better results can be obtained in future studies by removing the limitations mentioned above.

REFERENCES

- [1] INTERPOL. (2018). Disaster victim identification guide. DVI Guide: INTERPOL 2018: 1-31. file:///C:/Users/Administrator/Downloads/18Y1344%20E%20DVI_Guide.pdf, accessed on 20 May 2023.
- [2] Çöloğlu, A.S. (1999). Adli olaylarda kimlik belirlenmesi. In: Soysal Z, Çakalır C; eds. Adli Tıp Cilt 1. 1. baskı, İstanbul: İ.Ü., Tıp Fakültesi, 73-92.
- [3] Ajaz, A., Kathirvelu, D. (2013). Dental biometrics: Computer aided human identification system using the dental panoramic radiographs. In 2013 International Conference on Communication and Signal Processing. IEEE, pp. 717-721. <https://doi.org/10.1109/iccsp.2013.6577149>
- [4] Banerjee, K.K., Agarwal, B.B.L. (1998). Estimation of age from epiphyseal union at the wrist and ankle joints in the capital city of India. *Forensic Science International*, 98(1-2): 31-39. [https://doi.org/10.1016/S0379-0738\(98\)00134-0](https://doi.org/10.1016/S0379-0738(98)00134-0)
- [5] Karunya, R., Askarunisa, A., Athiraja, A. (2014). Human identification using dental biometrics. *International Journal of Applied Engineering Research*, 9(20).
- [6] Olze, A., Reisinger, W., Geserick, G., Schmeling, A. (2006). Age estimation of unaccompanied minors: Part II. Dental aspects. *Forensic Science International*, 159: S65-S67. <https://doi.org/10.1016/j.forsciint.2006.02.018>
- [7] Ngan, T.T., Tuan, T.M., Son, L.H., Minh, N.H., Dey, N. (2016). Decision making based on fuzzy aggregation operators for medical diagnosis from dental X-ray images. *Journal of Medical Systems*, 40(12): 1-7. <https://doi.org/10.1007/s10916-016-0634-y>
- [8] Rad, A.E., Mohd Rahim, M.S., Rehman, A., Altameem, A., Saba, T. (2013). Evaluation of current dental radiographs segmentation approaches in computer-aided applications. *IETE Technical Review*, 30(3): 210-222.
- [9] Stolojescu-Crişan, C., Holban, Ş. (2013). A comparison of X-ray image segmentation techniques. *Advances in Electrical and Computer Engineering*, 13(3): 85-92. <https://doi.org/10.4316/AECE.2013.03014>
- [10] Zhu, N., Wang, G., Yang, G., Dai, W. (2009). A fast 2d otsu thresholding algorithm based on improved histogram. In 2009 Chinese Conference on Pattern Recognition. IEEE, pp. 1-5. <https://doi.org/10.1109/CCPR.2009.5344078>
- [11] Tuan, T.M. (2016). A cooperative semi-supervised fuzzy clustering framework for dental X-ray image segmentation. *Expert Systems with Applications*, 46: 380-393. <https://doi.org/10.1016/j.eswa.2015.11.001>
- [12] Kahaki, S.M., Nordin, M., Ahmad, N.S., Arzoky, M., Ismail, W. (2020). Deep convolutional neural network designed for age assessment based on orthopantomography data. *Neural Computing and Applications*, 32(13): 9357-9368. <https://doi.org/10.1007/s00521-019-04449-6>
- [13] Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639): 115-118. <https://doi.org/10.1038/nature21056>
- [14] Jayaraman, J., Wong, H.M., King, N.M., Roberts, G.J. (2013). The French-Canadian data set of Demirjian for dental age estimation: A systematic review and meta-analysis. *Journal of Forensic and Legal Medicine*, 20(5): 373-381. <https://doi.org/10.1016/j.jflm.2013.03.015>
- [15] 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. <https://doi.org/10.1371/journal.pone.0174508>
- [16] Smith, T., Brownlees, L. (2011). Age assessment practices: A literature review & annotated bibliography. United Nations Children's Fund (UNICEF), New York.
- [17] Hjern, A., Brendler-Lindqvist, M., Nørredam, M.L. (2012). Age assessment of young asylum seekers. *Acta Paediatrica (Oslo, Norway: 1992)*, 101(1): 4-7. <https://doi.org/10.1111/j.1651-2227.2011.02476.x>
- [18] Yarımoğlu, H.B. (2005). Yaş tayini uygulamalarında epifiz plağı kapanma derecelerinin incelenmesi. Çukurova Üniversitesi, Tıp Fakültesi, Adli Tıp Anabilim Dalı, Uzmanlık Tezi.
- [19] Arslan, M.M., Çekin, N., Akçan, R., Saylak, E. (2008). Hatay ağır ceza ve asliye hukuk mahkemelerine 2007 yılında yansıyan yaş tespiti davalarının incelenmesi. *Adli Tıp Dergisi*, 22(2): 8-13.
- [20] Shamim, T., İpe, V.V., Shameena, P.M., Sudha, S. (2006). Age estimation: A dental approach. *Journal of Punjab Academy of Forensic Medicine and Toxicology*, 6(1): 14-16.
- [21] Suri, L., Gagari, E., Vastardis, H. (2004). Delayed tooth eruption: pathogenesis, diagnosis, and treatment. A literature review. *American Journal of Orthodontics and Dentofacial Orthopedics*, 126(4): 432-445. <https://doi.org/10.1016/j.ajodo.2003.10.031>
- [22] Wang, X.P. (2013). Tooth eruption without roots. *Journal of Dental Research*, 92(3): 212-214. <https://doi.org/10.1177/0022034512474469>
- [23] Demirel T, Bodrumlu E.H. (2018). Sürme anomalileri. *Uluslararası Diş Hekimliği Bilimleri Dergisi*, (3): 141-146. <https://doi.org/10.21306/jids.2018.191>
- [24] Yaşar, Z.F., Büken, E., Tekindal, M.A. (2016). Demirjian metodu farklı ülkelerde yaş tayininde kullanılabilir mi? *Adli Tıp Bülteni*, 21(3): 144-152. <https://doi.org/10.17986/blm.2016323747>
- [25] Harris, M.J., Nortje, C.J. (1984). The mesial root of the third mandibular molar. A possible indicator of age. *The Journal of Forensic Odonto-Stomatology*, 2(2): 39-43.
- [26] Akay, G., Atak, N., Güngör, K. (2018). Adli dişhekimliğinde dişler kullanılarak yapılan yaş tayini yöntemleri. *Ege Üniversitesi Diş Hekimliği Fakültesi Dergisi*, 39(2): 73-82.
- [27] Constantine, S., Clark, B., Kiermeier, A., Anderson, P. (2019). Panoramic radiography is of limited value in the evaluation of maxillary sinus disease. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 127(3): 237-246. <https://doi.org/10.1016/j.oooo.2018.10.005>
- [28] Gustafson, G. (1950). Age determination on teeth. *Journal of the American Dental Association (JADA)*, 41(1): 45-54. <https://doi.org/10.14219/jada.archive.1950.0132>
- [29] Hägg, U., Taranger, J. (1985). Dental development, dental age and tooth counts. *The Angle Orthodontist*, 55(2): 93-107. [1562](https://doi.org/10.1043/0003-</div><div data-bbox=)

- 3219(1985)055<0093:dddaat>2.0.co;2
- [30] Bernhard, W., Glöckler, C. (1995). New investigations on the question of secular acceleration of permanent dentition. *Zeitschrift Fur Morphologie und Anthropologie*, 81(1): 111-123.
- [31] Liliequist, B., Lundberg, M. (1971). Skeletal and tooth development: A methodologic investigation. *Acta Radiologica. Diagnosis*, 11(2): 97-112. <https://doi.org/10.1177/028418517101100201>
- [32] Townsend, N., Hammel, E.A. (1990). Age estimation from the number of teeth erupted in young children: An aid to demographic surveys. *Demography*, 27(1): 165-174. <https://doi.org/10.2307/2061560>
- [33] Demirjian, A., Goldstein, H. (1976). New systems for dental maturity based on seven and four teeth. *Annals of Human Biology*, 3(5): 411-421. <https://doi.org/10.1080/03014467600001671>
- [34] Yang, F., Jacobs, R., Willems, G. (2006). Dental age estimation through volume matching of teeth imaged by cone-beam CT. *Forensic Science International*, 159: S78-S83. <https://doi.org/10.1016/j.forsciint.2006.02.031>
- [35] Aboshi, H., Takahashi, T., Komuro, T. (2010). Age estimation using microfocus X-ray computed tomography of lower premolars. *Forensic Science International*, 200(1-3): 35-40. <https://doi.org/10.1016/j.forsciint.2010.03.024>
- [36] Fan, F., Ke, W., Wu, W., Tian, X., Lyu, T., Liu, Y., Liao, P., Dai, X., Chen, H., Deng, Z. (2020). Automatic human identification from panoramic dental radiographs using the convolutional neural network. *Forensic Science International*, 314: 110416. <https://doi.org/10.1016/j.forsciint.2020.110416>
- [37] Tuan, T.M., Ngan, T.T., Son, L.H. (2016). A novel semi-supervised fuzzy clustering method based on interactive fuzzy satisficing for dental X-ray image segmentation. *Applied Intelligence*, 45(2): 402-428. <https://doi.org/10.1007/s10489-016-0763-5>
- [38] Ilić, I., Vodanović, M., Subašić, M. (2019). Gender estimation from panoramic dental X-ray images using deep convolutional networks. In *IEEE Eurocon 2019-18th International Conference on Smart Technologies*, pp. 1-5. <https://doi.org/10.1109/EUROCON.2019.8861726>
- [39] Oktay, A.B. (2017). Tooth detection with convolutional neural networks. In *2017 Medical Technologies National Congress (TIPTEKNO)*. IEEE, pp. 1-4. <https://doi.org/10.1109/TIPTEKNO.2017.8238075>
- [40] Liu, N. (2021). Chronological age estimation of lateral cephalometric radiographs with deep learning. *arXiv Preprint arXiv: 2101.11805*. <https://doi.org/10.48550/arXiv.2101.11805>
- [41] Sathyavathi, S., Baskaran, K.R. (2023). Human age estimation using deep convolutional neural network based on dental images (Orthopantomogram). *IETE Journal of Research*, 1-8. <https://doi.org/10.1080/03772063.2023.2165177>
- [42] Taşçi, M.E., Şamlı, R. (2020). Veri madenciliği ile kalp hastalığı teşhisi. *Avrupa Bilim ve Teknoloji Dergisi*, 88-95. <https://doi.org/10.31590/ejosat.araconf12>
- [43] Saito, T., Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS One*, 10(3): e0118432. <https://doi.org/10.1371/journal.pone.0118432>