Deep Convolutional Neural Networks for Forensic Age Estimation: A Review

Sultan Alkaabi, Salman Yussof, Haider Al-Khateeb, Gabriela Ahmadi-Assalemi, and Gregory Epiphaniou

Abstract Forensic age estimation is usually requested by courts, but applications can go beyond the legal requirement to enforce policies or offer age-sensitive services. Various biological features such as the face, bones, skeletal and dental structures can be utilised to estimate age. This article will cover how modern technology has developed to provide new methods and algorithms to digitalise this process for the medical community and beyond. The scientific study of Machine Learning (ML) have introduced statistical models without relying on explicit instructions, instead, these models rely on patterns and inference. Furthermore, the large-scale availability of relevant data (medical images) and computational power facilitated by the availability of powerful Graphics Processing Units (GPUs) and Cloud Computing services have accelerated this transformation in age estimation. Magnetic Resonant Imaging (MRI) and X-ray are examples of imaging techniques used to document bones and dental structures with attention to detail making them suitable for age estimation. We discuss how Convolutional Neural Network (CNN) can be used for this purpose and the advantage of using deep CNNs over traditional methods. The article also aims to evaluate various databases and algorithms used for age estimation using facial images and dental images.

Keywords Deep learning · CNN · Forensic investigation · Information fusion · Magnetic resonant imaging (MRI) · Dental X-ray

S. Alkaabi · S. Yussof
Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, Kajang, Malaysia

H. Al-Khateeb (✉) · G. Ahmadi-Assalemi · G. Epiphaniou
Wolverhampton Cyber Research Institute (WCRI), University of Wolverhampton, Wolverhampton, UK
e-mail: H.Al-Khateeb@wlv.ac.uk

H. Jahankhani et al. (eds.), *Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity*, Advanced Sciences and Technologies for Security Applications, https://doi.org/10.1007/978-3-030-35746-7_17

1 Introduction

Forensic age estimation is one of the key research areas in the field of medical forensics. Although age estimation of unidentified cadavers or skeletal identification is a well- established forensic discipline, age estimation in living individuals is a relatively more recent area of applied research within forensic sciences that has attracted considerable attention [1]. The thriving integration of digital technologies into modern lives broadens the diversity and scope of forensic science and has created a need for new forensic science techniques including innovative computer vision and Machine Learning (ML) to support forensic investigations. Alongside the conventional forensic disciplines, Digital Forensics (DF) has developed as a branch of forensic science covering diverse digital technologies that can be exploited by criminals. Image-based evidence gained through sources like surveillance, monitoring or social media-driven intelligence that are commonly used by law enforcement in forensic investigations and by witnesses to describe suspects demonstrate the widening scope of forensic investigations. This creates specialised workload, generates backlog and requires highly specialised forensic practitioners [2, 3]. Therefore, more research is required to develop techniques and methods that are more efficient and automated thus reducing the backlog, workload and cost of the forensic investigation processes including the case studies when digital devices are involved as part of the crime scene or scope.

Soft biometric traits like age estimation, predicting a person's age using ancillary information from primary biometric traits like face, eye-iris, bones or dental structures, has attracted significant research in the past decade. Soft biometrics have a number of applications apart from medical forensics [1] including healthcare [4], age-related security control, human-computer interactions, law enforcement, surveillance and monitoring [5–7], socio-political related defence and security in border and immigration controls and to establish the age of illegal immigrants without valid proof-of-birth in adults or unaccompanied minors [8, 9], which is becoming an integral part of forensic practice [10]. Furthermore, without an accurate age estimation victims of child-trafficking, asylum seekers or illegal immigrants cannot receive the required instrumental support [11]. Due to the ease of online access, child sexual victimisation crimes are rising [12] with increased DF child exploitation investigations involving age estimation [13].

Apart from determining the age of cadavers or as part of the paleo-demographic analysis, the ability to estimate the age of living persons, which require accurate age estimation techniques, has become increasingly more important. In traditional approaches, most dental age estimation techniques like tooth emergence [14] or dental mineralisation [15] have limitations of age estimation beyond adolescence. Skeletal maturity with the development of X-ray was researched but due to the risks of exposure extensive X-ray based datasets were not produced. The development of highly detailed imaging techniques like ultrasound and Magnetic Resonant Imaging

(MRI), used to record dental and bone structures provide suitable opportunities for age determination of living persons [10].

Determining the age from image data is a highly complex task with numerous methods proposed by scientific research from measurement-driven analysis to the application of machine learning algorithms with constantly improving accuracy [16]. While a human face reflects significant amount of communicative information and facets about a person including gender, identity, ethnicity, expression and age, which humans have a capability to detect at a glance, there is a growing expectation that digital systems will have similar capabilities and recognition accuracy seamlessly [16–18]. Ancillary-related biological traits like the heterogeneity of the maturing process of human faces, bones, wrinkles, ethnicity or image-related traits including illumination, make-up or pose make age estimation challenging [19, 20]. Deep Learning (DL) methods result in higher accuracy compared to more traditional approaches like statistical [14], handcrafted methods that although require very small datasets, short training times and are computationally inexpensive their problem solving approach is modular relying on expert knowledge for complex feature extraction [21] or shallow learning which also requires feature extraction and classification [22]. Although DL methods require large-scale datasets, highly complex computational capability compared to the traditional approaches DL has automatic feature extraction with an end-to-end problem-solving approach that enables solving computer vision challenges [20, 23].

Furthermore, the large-scale availability of image dataset, the advantages of hardware, analysis techniques and parallel processing of High-Performance Computing (HPC) to deal with the computational requirement of image-based age estimation, although underexploited, are beneficial to the digital forensics' community and could reduce the computation time to expedite the processing and analysis of the DF investigation. Although traditionally GPU computing was considered difficult to utilise and targeted for very niche problem solving, the availability of multi-core CPU with GPU acceleration is increasingly more accessible and widely used in HPC enabling simpler programming models, better economies of scale and performance efficiency [2]. More precisely, recent research makes widespread use of deep Convolutional Neural Networks (CNN), automating and significantly increasing the age estimation accuracy. If applied, the use of CNN for automated age estimation could increase accuracy and reduce the human effort in forensic investigations.

This article addresses age estimation, introduces and discusses deep CNN in automated age estimation to support the medical community. The difference between the traditional approach and the deep learning approach for age estimation is discussed at length along with the reasons which made the deep learning approach more popular in recent years among researchers. A detailed comparison of deep CNN based methods for age estimation using different biological features is also covered including advantages and drawbacks of using dental MRI images for age estimation.

2 The Difference Between Traditional Approaches and Deep Learning for Age Estimation

We have found four distinctive approaches in the literature for estimating age from images. The first approach used statistical analysis of teeth and mandibular of child subjects [24]. proposed a method of age estimation based on the development of the seventh teeth from the left side of the mandible. And [25] proposes a method based on 14 stages of mineralization [25].

The second approach used handcrafted methods extracting features from the texture of the face, shape, the colour of the skin, appearance etc. [21] proposed a method for age estimation which can extract effective ageing pattern using a discriminant subspace learning algorithm. In [26], an automatic age estimation method based on ageing pattern representative subspace was proposed which mainly sorts face images by time order.

The third approach is related to shallow learning. It involves extracting features using local binary methods from the patches of the face and then classifying the extracted features using a classifier [27]. proposed Bio-inspired features which are widely used for age estimation, and [28] proposed the improvement base don using a scattering transform. This method added a filtering route to the biologically inspired future which improved the accuracy of age estimation [29]. proposed an orthogonal locality preserving projection technique (OLPP) which further increased the quality of features for age estimators. The second component in this method is a classifier or regressor. Classifiers can be a multi-layer Perceptron, k-nearest neighbours or Support Vector Machine (SVM). Polynomial regression [29] support vector regression and can be used as a regression method for age estimation. This method also requires some prior knowledge.

The fourth approach utilises deep learning algorithms to learn the hierarchical features automatically from images [30]. A detailed analysis of deep learning-based methods will be demonstrated in this article. These methods have the advantage of not requiring a feature selection process, instead, features are selected automatically according to the application.

The handcrafted and shallow learning approach requires a separate feature detection step, then these features are classified using a separate classifier. Whilst deep learning methods provides an end-to-end solution which removes the need of a separate classifier. However, the drawback of using deep learning can be manifested by the requirement of a big dataset and demand for a powerful processor. It has been observed that deep learning methods provide higher accuracy compared to other methods, but it is very difficult to interpret which features have been used to reach the conclusion with this higher level of accuracy.

Table 1 demonstrates differences between the traditional approaches namely shallow learning and hand-crafted feature learning methods, and deep learning methods.

Table 1 Comparison between deep learning and other traditional approaches

Comparison parameter	Deep learning	Shallow learning	Handcrafted methods
Data requirement	Large dataset	Small dataset	Very small dataset
Hardware requirement	CPU + GPU	CPU	Normal or embedded CPU
Feature extraction	Automatic	Handcrafted features + classification	Handcrafted features
Problem solving approach	End-to-end	Modular	Modular
Training time	Long	Short	Very short
Interpretability of features	Low	High	Very high

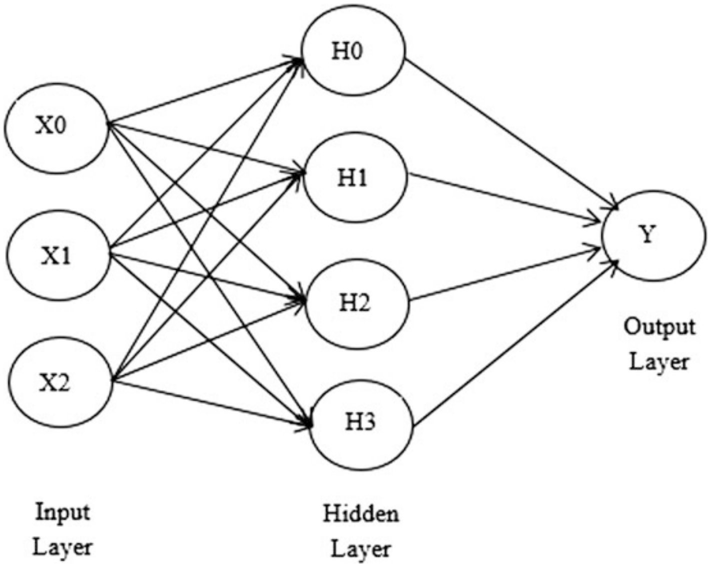

Fig. 1 Artificial neuron architecture

3 The Convolutional Neural Network (CNN)

The most widely used deep learning method for age estimation in literature is CNN. The basic type of neural network tries to mimic the behaviour of the human brain and is called Artificial Neural Network (ANN). The ANN architecture is a perceptron weighting a sum of inputs and applies a threshold activation function [31]. It contains multiple perceptrons connected with each other as shown in Fig. 1.

The ANN architecture in Fig. 1 contains an input layer with three neurons, an output layer with one neuron and a hidden layer with four neurons. The neurons in every layer are connected with each other so ANN is also known as a fully connected network. Each neuron performs the weighted sum of all the inputs and adds the bias term. This is a linear operation but most of the real word problems are non-linear.

Therefore, to make the network non-linear this sum is passed through an activation function. The output y for a neuron with k inputs can be represented as:

$$y = f\left(\sum_{i=0}^{k} X_i W_i\right) \tag{1}$$

The modern-day neural network contains many intermediate hidden layers so these networks are called Deep Neural Networks (DNN).

The number of weights between each layer can be calculated by multiplying neurons in a current layer by neurons in a previous layer. The number of weights will increase together with the number of neurons in the hidden layer. The number of hidden layers and number of neurons in each hidden layer is called hyperparameters which have to be chosen thoughtfully by the network designer according to the application.

The choice of activation function plays a very crucial role in determining the performance of the ANN. It will also determine how fast the network will converge while training and how much computational cost it requires. There are many activation functions used by network designers but Sigmoid, Tanh and ReLU are the most frequently used activation functions. The mathematical equations for these are given below.

$$\text{Sigmoid function}: f(y) = \frac{1}{1 + e^{-y}} \tag{2}$$

$$\text{Tanh function}: f(y) = \frac{e^{y} - e^{-y}}{e^{y} + e^{-y}} \tag{3}$$

$$\text{ReLU function}: f(y) = \max(0, y) \tag{4}$$

The Sigmoid function (3) is considered a smooth threshold function which is also differentiable. The output of a sigmoid function will be between 0 and 1. The issue with sigmoid function is that for a large value of activations it has a very small value of gradient so weights in initial layers will take a long time to update (also called the vanishing gradient problem). Tanh or hyperbolic tangent function as described in Eq. (4) is similar to sigmoid but it has an output in the range of -1 to 1. It will work better than sigmoid in most cases because it centres the data with zero Means. The vanishing gradient problem is also prevalent with the Tanh activation function. However, Rectified Linear Unit (ReLU) function as described in Eq. (5) can solve the problem of vanishing gradient. It is also easier to compute and the overall training of the network is relatively faster.

The final layer of ANN for a multiclass Image classification uses softmax activation function [32] described in Eq. (5) which is mainly an extension of the Sigmoid activation function. It gives the probability of each class by converting the vector to a range from 0 to 1.

$$\text{Softmax Activation function}: f(y) = \frac{e^{y_k}}{\sum_{k=1}^{k} e^{y_k}} \tag{5}$$

ANN uses weights and bias to store information related to the application. These weights and biases are updated during the training phase of the supervised learning approach by calculating the minima of a cost function. The cost function is an error function between the actual value and the predicted value and could be a Mean Square Error, Mean Absolute Error, Binary or sparse cross-entropy etc. The minima of the cost function can be found by using optimization algorithms like gradient descent, Adam, RMSProp etc.

There is a limit to using ANNs for computer vision tasks. The raw pixel values are used as input to the ANN. So for an image size of 1080 × 1080, there will be one million input neurons. Even if there is only one hidden layer with a small number of neurons, the network will have millions of trainable parameters which means a large dataset and a complex computational unit for training. The second drawback associated with using ANN for computer vision is that it does not take into account spatial neighbourhood information although it is essential for image processing.

These two drawbacks of ANN has led to the use of CNN in computer vision [33]. CNN uses convolution operation which takes into account the spatial neighbourhood information. It also uses the concept of parameter sharing which reduces the number of trainable parameters. It can do that because the same weights can be applied to find features from an entire image. A 3x3 Sobel filter can find edge features from an image of any size with only 9 weights.

The architecture of CNN for an age estimation problem is shown below in Fig. 2.

Figure 2 shows an input image (dental MRI scan) passing through a number of convolution and pooling layers. The convolutional layer tries to collect hierarchical features from the image. Then, the pooling layer is used to reduce the dimensions of the features map. The number of convolution operations in each layer along with the number of these layers should be chosen wisely by the network designer. The output is then converted to a single column vector by a Flattening layer. This single vector is given as an input feature vector to an ANN or a fully connected network for image classification.

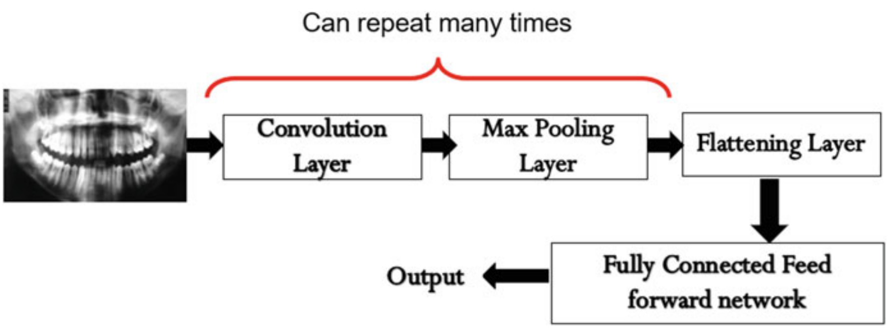

Fig. 2 CNN architecture

3.1 Techniques to Avoid Overfitting in CNN

When the network performs very well on the training data but poorly on the test data then it is called over-fitting. There are several techniques to avoid over-fitting. For instance, Regularization prevents the weights from getting too large. Batch normalization regularises the response after every convolution layer. Another technique is Dropout [34] where random neurons are dropped from the network during training, and the network will not be overly dependent on a single neuron.

3.2 Training of CNN

CNN stores information related to the application in the form of weights and bias and need to be trained for the given application. This can be done by showing the labelled training data to the CNN architecture, this approach is called Supervised Learning.

The weights and biases are initialized randomly with small values. Uniform Random distribution or Xavier initialization [35] is normally used to initialize the weights' value. When the labelled training image samples are given to the CNN architecture, it will calculate the prediction with a forward pass technique using the initialized weights. Then the error between the predicted output and the actual output will be calculated. Mean square error and Mean absolute error are two popular error function for regression problems. Binary cross-entropy is used for the binary classification problem, while the categorical or sparse cross-entropy is used as an error function for the multi-class classification problem.

The calculated error is backpropagated to update the weights using a gradient descent which is an optimization algorithm used to find the minimum of the error function. Other optimizers include Stochastic Gradient Descent, Adam, RMSProp and Adagrad.

There are different types of training methods depending on the number of times the weights are updated in a given timeframe. If weights are updates only once it is called full batch learning. The full batch learning method will take a long time to converge and it will require a large memory space to store images from the entire training set. The advantage of using full batch learning is that it will certainly converge to a global minimum. However, using a stochastic method as an alternative type of training updates the weights after every image, therefore, requires minimum memory and converges faster. It has the disadvantage of fluctuating around the minimum value. Moreover, an intermediate method is referred to as mini-batch learning where the training set is divided into several batches and the weights are updated after every batch of images.

4 Availability and Quality of Datasets for Age Estimation

The appropriateness and completeness of the training dataset can be the key factor to improve the accuracy of age estimation. CNN as a supervised algorithm requires a large number of labelled datasets for training. Datasets for age estimation should also contain a uniform distribution of images of all ages for accurate and inclusive detection. The widespread use of social networking sites has contributed to maintaining large scale facial datasets. Additionally, many open-source datasets designed specifically for age estimation have been created. Face and dental structure are the two most used biological features to estimate age in the literature. To investigate which of these have been successfully used in research studies, we have performed secondary data analysis of primary studies which we summaries in Tables 2 and 3.

Table 3 list all the datasets used in the literature to estimate age based on facial images.

The choice of datasets plays a very important role in getting an accurate result for a particular application. Therefore, a suitable database from the above tables can be chosen for estimating age using facial and dental images. In the next section, we compare between deep CNN methods trained using these databases for age estimation.

Table 2 Summary of dental datasets used for age estimation

Name	Number of subjects	Age range	Special note about the dataset
Southern Chinese Patient Dataset [36]	182	3–16 years	The dataset contained dental panoramic Tomograph (DPT) images from children and adults. The dataset contained the images in the range of 3 to 16 years. The selection of subjects was done from the archives of Prince Philip dental hospital, Hong Kong. The subjects were chosen randomly.
UK Caucasian Dataset [37]	5187	11–15 years	Aimed to develop a reference dataset for at the 13 year old threshold to support dental age assessment for Caucasian children.
French-Canadian Dataset [38]	274	2–21 years	This dataset is based on the dental maturity of French and Canadian population. This dataset overestimates the age by 6 months so you have to be very careful while choosing this dataset for a global population.
Darko Stern’s collected MRI Dataset [39]	103	13–25 years	This custom dataset contains 103 3D MRI images of the hand, thorax and dental structure out of that 44 subjects were of minors.

Table 3 Summary of facial age estimation datasets

Database	Images	Age Range (Years)	Special Notes about Dataset
FG-NET [40]	1002	0–69	This dataset is widely is used for estimating age. It is not available for download from its official site but can be downloaded from other sources.
MORPH [41]	1724	27–68	This dataset is provided for age estimation in adults for academic distribution.
Yamaha gender and age (YGA) [21]	8000	0–93	The dataset contains five labelled frontal face images of the same person. The images have different facial expression and illumination.
WIT-DB [42]	5500	3–85	The WIT-DB dataset contains images with large illumination variation and a large age group. The number of images in a particular illumination condition is also unbalanced.
AI & R Asian [43]	34	22–61	This dataset contains images taken in the diverse scenarios like different poses, illumination, ages etc.
Burt's Caucasian face database [44]	147	20–62	This dataset is used to estimate age by combining visual features of colour and shape of facial components.
Lotus Hill research institute (LHI) database [45]	8000	9–89	This dataset contains images of Asians adults with a wide age range. It is also very large dataset which can be used for deep CNN models.
Human and object interaction processing (HOIP) [46]	306,600	15–64	The dataset is divided into ten age groups with each group containing images of 30 subjects. Each age group contain an equal distribution of male and female.
Iranian face database [47]	3600	2–85	The images in the dataset contain large variation in pose and expressions. Every subject has at least one image with the glass. The dataset contains images in the age group of 2–85 years with the majority of them are of subjects before 40 years. This dataset is appropriate for formative and middle age estimation.
Gallagher's web-collected database [48]	28,231	0–66	This database is designed for studying group photos so most of the images in the database are front-facing images with artificial poses. It is a large database which can be used to estimate age in a wide range.
Ni's web collected database [49]	219,892	1–80	This dataset is collected from the web search engines like Google specifically for age estimation in the wide age range. The size of the dataset makes it suitable to use this dataset in estimating an age for children, middle age and old age persons.

(continued)

Table 3 (continued)

Database	Images	Age Range (Years)	Special Notes about Dataset
Kyaw's Web-Collected Database [50]	963	3–73	This dataset is manually created for age estimation by finding out images from Microsoft search engine Bing. The images are aligned manually. The images are cropped to the patches of size 65 by 75.
Combination of LFW, and images from the web [51]	13,466	–	This dataset is collected by the biometric engineering research center. There is uniform illumination in all the images of the dataset along with no variation in facial expression. The uniform distribution of subjects is there in terms of gender and age group.
FDDB Dataset [52]	5171	–	This database contains face images taken in a wide range of difficulties that include occlusion, different poses, and different illumination. The images are taken in either colour or grayscale scenario.
Adience Benchmark [53]	2284	0–60	This dataset is prepared for the study of age and gender estimation from facial images. The dataset contains images with a different appearance, different lighting, noise etc. it intends to take in to account all the challenges of real-world imaging conditions.
Apparent age dataset [54]	4691	–	The images are taken in a real-time environment and have variation in pose, occlusion, lighting, illumination, background, ethnicity etc.
IMDB-WIKI Dataset [55]	524,230	–	This dataset contains the web crawled images of celebrities taken from IMDB and Wikipedia. This is the largest public dataset available of facial images which are widely used particularly for deep CNN applications.

5 Deep CNN Based Methods for Age Estimation

5.1 Deep CNN Based Methods for Age Estimation from Facial Images

Most CNN-based methods seem to utilise well-known architectures (e.g. AlexNet, GoogleNet, ResNet and VGGNet) pre-trained usually on the ImageNet [56] dataset. Very few methods try to develop a new CNN architecture from scratch. This approach is simpler and faster because it does not require fine-tuning. The second approach, however, fine-tunes the weights of well-known pre-trained CNN architectures on a new facial dataset. This approach is an end-to-end method which requires additional training on new facial datasets.

In [57], the CNN architecture consisted of three convolution layers and two pooling layers. It used a combination of CNN and Gabor filter for achieving higher accuracy. The study also showed that going wider instead of deeper with increased filter size can achieve a good result for age and gender classification. The proposed method does not use a complete end-to-end approach as it uses a Gabor filter to find features [58]. proposed a large-scale 22-layers deep CNN framework (AgeNet) for age estimation which used a combination of real value-based regression and label-distribution based classification to estimate the final age. It also proposed a learning method which can be really helpful in avoiding overfitting on a small dataset. However, this method required separate training for regression and classification models. Another study [55] proposed a system called Deep Expectation (DEX) using CNN. It used VGG-16 as a base architecture which is pre-trained on the ImageNet dataset and then fine-tuned the model on face images with age labels. The VGG-16 network used 16 trainable layers with a smaller filter size of 3x3 compared to larger filter sizes in earlier networks. The results showed improvement over direct age regression using CNN. The authors in [30] utilised pre-trained CNN architectures as well but only to perform feature extraction. They used Principal Component Analysis (PCA), Mutual Information and Statistical dependency techniques for dimensionality reduction and ANN for classification.

Age estimation via fusion of depthwise separable CNN was proposed in [59], this has reduced the number of parameters for training without sacrificing accuracy. Three state-of-the-art deep learning models Xception, Inception V3 and ResNet were modified to use depth wise convolution for enhancing the performance and lowering the computational requirement of the system. Empirical results based on four publically available datasets showed superior performance compared to other methods on those datasets when it comes to age estimation [60]. proposed a cluster CNN architecture which significantly reduces the preprocessing steps. The facial image is normalized to a standard size according to the distance between two eyes, This normalized image is fed to the cluster CNN architecture for prediction. The cluster is integrated into the CNN architecture which is capable of multimodal transformation. It is also differentiable so the parameters of it can be learnt using backpropagation. A ranking CNN architecture was proposed in [61]. It is a series combination of normal CNN architecture trained on ordinal age labels. The outputs from the individual CNN architectures are combines to predict the final age. This approach of estimating error seem to obtain better results compared to multi-class classification approach. The performance of the method was evaluated with the MORPH dataset and compared with other state-of-the-art methods.

CNN2ELM was proposed in [62] as a more complex design that incorporates CNN and Extreme Learning Machine (ELM). It consists of three CNN architectures Age-Net, Gender-Net and Race-Net to extract features related to age, gender and race from the image of the same person. The architectures are pre-trained on the ImageNet database. Then it uses ELM classifier for age grouping and ELM regressor for age estimation. The network is fine-tuned on IMDB WIKI dataset and it outperforms other architectures on well-known datasets. It does that because it uses decision fusing to achieve a robust decision. This approach finds more discriminative features from the image then combines the prediction on them to estimate age. However, the performance of the system was poor for a dataset with varied poses or turned/tilted faces.

A consistent limitation affecting all the above methods was the amount of labelled facial data available for age estimation. In response, [62] proposed a data augmentation technique to increase the size of the training data for age estimation. This has produced new training samples from existing images and can be accomplished by applying small transformation like translation, rotation, flipping to the images in the existing dataset. The proposed method also take in to account the intrinsic information about the human face while creating the augmented dataset. The MORPH [41] dataset was used with the same CNN model trained using original and augmented dataset seen a rise of 10% in F-score after utilising the augmented dataset.

Furthermore, [63] proposed a transfer learning-based method. They used VGG19 and VGG Face architecture to explore the performance of transfer learning in age estimation. Techniques such as input standardization, data augmentation and label distribution age encoding were employed to enhance the quality of training while transfer learning. Although the performance of the proposed system was good, it was performing poorly on minorities in the dataset such as old age people, females, people of Asian or African origin. The gender prediction was only based on the length of the hair. These flaws can be overcome by establishing a balanced dataset and changing the architecture or training technique. In [64], the authors proposed an age estimation system by combining CNN with the other popular deep learning architecture called Long Short Term Memory (LSTM). They called the system recurrent age estimation. CNN was used to find discriminative features from the facial images and LSTM were used for learning ageing patterns from a sequence of personalized features.

Further comparison of deep CNN based methods for age estimation is demonstrated in Table 4.

5.2 Limitations When Using Facial Images for Age Estimation

Relying on features extracted from the face to estimate age has many limitations. The human face matures in different ways at different ages. Bone growth and wrinkles will be different from one person to another. It is also observed in the literature that women are more likely to develop wrinkles in the perioral region than men [65]. Other challenges include changes in illumination, application of makeup on the face, different face poses and different backgrounds. Hence, the face alone is not always reliable for accurate age prediction.

Table 4 Comparison of deep CNN facial age estimation methods

Deep CNN Architecture	Dataset Used	Performance	Note
Wide CNN [57]	Adience Benchmark Dataset [53]	Age accuracy: 61.3% Gender accuracy: 88.9%	The paper solved the problem of age estimation as a classification problem with eight classes of different age groups so the accuracy is in percentages.
AgeNet [58]	Apparent age dataset provided by the ICCV2015 looking at people challenge	Mean normalized error = 0.2872 Mean absolute error = 3.3345	The paper used 2476 images for training, 1136 for validation and 1087 for testing. The performance of the network is measured in terms of mean normalized and mean absolute error.
Deep Expectation [55]	IMDB-WIKI Dataset [55]	Mean absolute error = 3.221 ε error = 0.278	The paper used mean absolute error and ε error for evaluation on IMDB-WIKI and the ChaLearn LAP dataset.
DSC- Xception [59]	IMDB-WIKI dataset	Mean absolute error = 6.2898	This network used the Xception module and depth wise separable convolution.
	MORPH II [41]	Mean absolute error = 3.25	
DSC- inception v3 [59]	IMDB-WIKI dataset	Mean absolute error = 6.3571	This network used inception v3 module and depth wise separable convolution.
	MORPH II	Mean absolute error = 3.32	
DSC- ResNet [59]	IMDB-WIKI dataset	Mean absolute error = 6.5099	This network used the ResNet architecture and depth wise separable convolution.
	MORPH II	Mean absolute error = 3.52	
DSC-Xception + Inception v3 + ResNet [59]	IMDB-WIKI dataset	Mean absolute error = 5.8865	This network used the fusion of Xception module, inception v3 module and Resnet along with depthwise separable convolution.
	MORPH II	Mean absolute error = 3.08	
VGG Face CNN + Dimensionality Reduction + ANN [30]	IMDB-WIKI dataset	Mean absolute error = 5.4	This technique used VGG face technique for feature extraction which was applied to various dimensionality reduction techniques and ANN for classification.
AlexNet CNN Dimensionality Reduction + ANN [30]	IMDB-WIKI dataset	Mean absolute error = 5.86	This technique used VGG face technique for feature extraction which was applied to various dimensionality reduction techniques and ANN for classification

(continued)

Table 4 (continued)

Deep CNN Architecture	Dataset Used	Performance	Note
Cluster CNN [60]	MORPH II	Mean absolute error = 2.71	The GoogleNet architecture trained on ImageNet database is used as a base network.
Ranking CNN [61]	MORPH	Mean absolute error = 2.96	The age estimation problem in the range of 16 to 66 years was considered. 43,490 samples were used for training and 10,872 were used for testing results. The results were carried out with five-fold cross-validation.
CNN2ELM [62]	MORPH	Mean absolute error = 2.61	The architecture is pre-trained on ImageNet database and then fine-tuned on IMDB WIKI and MORPH II database.
VGG19 and VGG Face Transfer Learning [63]	MORPH	Mean absolute error = 4.10	The VGG 19 architecture is pre-trained on ImageNet database. The MORPH II database is used to fine-tune the weights with 80% of the images are used for training and 20% of the images are used for testing.
Recurrent age estimation (RAE) [64]	MORPH	Mean absolute error = 1.32	Two public dataset MORPH and FG-net are used to evaluate the performance of the system. VGG-16 was used as a base CNN architecture.
	FG-net	Mean absolute error = 2.19	

5.3 Deep CNN Based Methods for Age Estimation from Dental Images

Teeth are among the more reliable features for estimating age especially until the age of 20. The various stages of teeth development can be utilised as features to estimate the age of a person but results are more accurate during the dentition development stage because the changes are very prominent and easy to observe. Sometimes the third molar is used for age estimation between 16 and 23 years old though this method is not so accurate. The tooth formation process is over after this age so it becomes very hard to estimate age. Instead, the 'wear' and 'age' regressive changes of hard and soft tissues in the teeth are analyzed to estimate the age for adults.

Examples of imaging techniques include the two-dimension intraoral and panoramic radiographs, 3-dimensional cone-beam computed tomography (CBCT) and Magnetic resonant Imaging (MRI). Many researchers are working using CNN architecture for various applications in dentistry but until recently very little work was directed at deep CNN for age estimation using dentistry.

In [66], the authors produced a method based on a modified Demirjian staging Technique that includes ten development stages. It used transfer learning on a pre-trained AlexNet CNN architecture and the ImageNet dataset. The analysis included 400 panoramic radiographic images and the results showed 10% improvement in classification accuracy. In another recent study [39] the proposal combined features from Dental and Skeletal MRI images. Age estimation is performed by fusing features of three different CNN architectures. Three CNN architectures are used to extract features from cropped wisdom teeth, hand and clavicle bones. Each CNN Architecture consists of three stages of two Convolution and one Max Pulling Layer followed by a fully connected layer. The data augmentation technique was used for the training making the results of the system more accurate and robust. However, it only contains 103 studied subjects so generalization of these results has to be done carefully. This method can be used to estimate the age range up to 25 years. In [67] the method aimed for chronological age estimation using panoramic dental X-ray images. The dataset was divided into three age groups of 2–11 years, 12–18 years and 19 years onwards respectively. The DenseNet-121 [68] architecture with channel-wise attention module was used for age estimation. The curriculum learning strategy was employed in which the network was first trained on images of subjects up to 11 years old and then it slowly included other subjects from 12–18 years and 19 years onwards. The method yielded promising results including the 19 years onwards age group with a giving mean absolute error of 4.398 years only.

Table 5 shows a comparison of deep CNN based methods for age estimation based on facial images.

5.4 Limitations When Using Dental Images for Age Estimation

The empirical findings in current literature are based on in-house datasets which prevent objective cross-method comparisons. Results from small datasets cannot be generalised while deep CNN requires a large amount of data for training, a problem that can be partially solved using data augmentation techniques. There is a need to create a large public dataset for dentistry which can be used for age estimation from dental images. It is also observed that most datasets in dentistry contain more images of children and less number of images for adults. Therefore, a more uniform distribution of images at all ages will be good to support further research in this area.

The size of dental images is relatively large, a problem usually solved by reducing image size before applying detection methods to cut computational cost. However, we could argue against this practice since important information can be lost. Nonetheless, current methods require manual intervention e.g. to fix the “region of interest”. A process that can be automated as part of future work.

Table 5 Comparison of deep CNN dental age estimation methods

Deep CNN Architecture	Dataset	Accuracy	Note
AlexNet with Transfer Learning [64]	400 panoramic radiographic images	Mean accuracy = 0.51 Mean absolute difference = 0.6	The paper used the AlexNet CNN architecture trained on the ImageNet database as a base architecture which was fine-tuned on a custom dataset with 80% of images used for training and 20% of images used for testing. Five-fold cross-validation was used for training.
DCNN-MAJ-HAND [66]	103 3D MRI images of left hand, upper thorax and the jaw.	Classification accuracy = 90.3% Mean absolute error = 1.14 ± 0.96 years	The data consisted of images in the age ranges of 13–25 years. The data augmentation technique was used to increase the size of the dataset.
DenseNet-121 with channel-wise attention module [67]	Panoramic dental X-ray images of 9435 subjects.	Mean absolute error 2–11 years = 0.826 12–18 years = 1.229 19 years onwards = 4.398	The dataset contained an equal distribution of male and females with age range in 2–98 years. The size of the original image was 1024 × 2048 which was resized to 256 × 512 before giving it to CNN

Finally, developing age estimation methods is feasible up to 20 years of age, it is very challenging to develop a system to estimate age covering all age groups for several reasons. For example, there is a large variation in teeth conditions after puberty due to dietary habits and teeth management. Dental development is affected by various genetic, environmental, nutritional and endocrinal factors. Teeth eruption is also affected by a number of factors such as gender, ethnic origin, physical and sexual development. It is also observed that age estimation techniques developed for one population might not work for a population belonging to another ethnicity. As such, a typical error rate for adults using dental images is ±10 years which is a very large value. There is a need for researchers to minimize this error to as low as possible.

6 Conclusion

Age estimation plays a very important role in medical forensic as it provides confirmation of a most needed input based on biological features such as the face, bones, skeletal and dental structures. However, the application of age estimation goes beyond that to provide a form of authentication to computer systems. Think about a system's ability to offer personalised Human Computer Interaction (HCI) based on the user age group. Likewise, preventing unauthorized access to individuals as part of a proactive security and defence applications in connected cars [69], border control and more. Clearly, features from facial images or a live feed of the face will be more feasible to utilise for most of these applications.

The research in age estimation has seen a great amount of transformation in recent years following a surge in the use of deep learning algorithms for computer vision. This can be attributed to the availability of medical image datasets and an increase in computer processing power with GPUs or through Cloud Computing services. In this article, we have covered how deep CNN emerged and discussed several recent proposals to highlight the advantages and limitations associated with each approach. Performing secondary data analysis of deep CNN is inevitable to understand research gaps and opportunities.

References

1. Alkass K, Buchholz BA, Ohtani S, Yamamoto T, Druid H, Spalding KL (2010) Age estimation in forensic sciences, application of combined aspartic acid racemization and radiocarbon analysis. Mol Cell Probes 9(5):1022–1030. https://doi.org/10.1074/mcp.M900525-MCP200
2. Lillis D, Becker B, O'Sullivan T, Scanlon M (2016) Current challenges and future research areas for digital forensic investigation. arXiv preprint arXiv:1604.03850
3. Boddington R (2016) Practical digital forensics. Packt Publishing Ltd, Birmingham
4. Kim K, Choi Y, Hwang E (2009) Wrinkle feature-based skin age estimation scheme, pp 1222–1225. Published. https://doi.org/10.1109/ICME.2009.5202721
5. Guo G, Fu Y, Huang TS, Dyer CR (2008) Locally adjusted robust regression for human age estimation, pp 1–6. Published. https://doi.org/10.1109/WACV.2008.4544009
6. Han H, Otto C, Jain AK (2013) Age estimation from face images: human vs. machine performance, pp 1–8. Published. https://doi.org/10.1109/ICB.2013.6613022
7. Ahmadi-Assalemi G, Al-Khateeb HM, Epiphaniou G, Cosson J, Jahankhani H, Pillai P (2019) Federated blockchain-based tracking and liability attribution framework for employees and cyber-physical objects in a smart workplace, pp 1–9. Published. https://doi.org/10.1109/ICGS3.2019.8688297
8. Schmeling A, Garamendi PM, Prieto JL, Landa MI (2011) Forensic age estimation in unaccompanied minors and young living adults. In: Forensic medicine—from old problems to new challenges. InTech, Rijeka, pp 77–120. https://doi.org/10.5772/19261
9. Hjern A, Brendler-Lindqvist M, Norredam M (2012) Age assessment of young asylum seekers. Acta Paediatr 101(1):4–7. https://doi.org/10.1111/j.1651-2227.2011.02476.x
10. Schmeling A, Black S (2010) An introduction to the history of age estimation in the living. In: Age estimation in the living, pp 1–18. https://doi.org/10.1002/9780470669785.ch1
11. Sauer PJJ, Nicholson A, Neubauer D, Advocacy and Ethics Group of the European Academy of Paediatrics (2016) Age determination in asylum seekers: physicians should not be implicated. Eur J Pediatr 175(3):299–303. https://doi.org/10.1007/s00431-015-2628-z
12. Seigfried-Spellar KC (2012) Measuring the preference of image content for self-reported consumers of child pornography, pp 81–90. Published. https://doi.org/10.1007/978-3-642-39891-9_6
13. Gladyshev P, Marrington A, Baggili I (2015) Digital forensics and cyber crime. Springer, Berlin. https://doi.org/10.1007/978-3-642-35515-8
14. Demirjian A, Goldstein H, Tanner J (1973) A new system of dental age assessment. Hum Biol 45:211–227

15. Moorrees CF, Fanning EA, Hunt EE Jr (1963) Formation and resorption of three deciduous teeth in children. Am J Phys Anthropol 21(2):205–213. https://doi.org/10.1002/ajpa.1330210212
16. Anda F, Lillis D, Le-Khac N, Scanlon M (2018) Evaluating automated facial age estimation techniques for digital forensics, pp 129–139. Published. https://doi.org/10.1109/SPW.2018.00028
17. Sehrawat D, Gill NS (2018) Emerging trends and future computing technologies: a vision for smart environment. Int J Adv Res Comput Sci 9(2):839. https://doi.org/10.1109/TIFS.2014.2359646
18. Shejul AA, Kinage KS, Reddy BE (2017) Comprehensive review on facial based human age estimation, pp 3211–3216. Published. https://doi.org/10.1109/ICECDS.2017.8390049
19. C. f. D. C. a. P (2019) Chronic diseases: the leading causes of death and disability in the United States. 01/08/2019. https://www.cdc.gov/chronicdisease/resources/infographic/chronic-diseases.htm
20. Dantcheva A, Elia P, Ross A (2015) What else does your biometric data reveal? A survey on soft biometrics. IEEE Trans Inf Forensics Secur 11(3):441–467. https://doi.org/10.1109/TIFS.2015.2480381
21. Fu Y, Huang TS (2008) Human age estimation with regression on discriminative aging manifold. IEEE Trans Multimedia 10(4):578–584. https://doi.org/10.1109/TMM.2008.921847
22. Guo G, Mu G, Fu Y, Huang TS (2009) Human age estimation using bio-inspired features, pp 112–119. Published. https://doi.org/10.1109/CVPR.2009.5206681
23. Tian Q, Chen S (2015) Cumulative attribute relation regularization learning for human age estimation. Neurocomputing 165:456–467. https://doi.org/10.1016/j.neucom.2015.03.078
24. Demirjian A, Goldstein H, Tanner JM (1973) A new system of dental age assessment. Hum Biol 45(2):211–227
25. Moorrees CFA, Fanning EA, Hunt EE Jr (1963) Formation and resorption of three deciduous teeth in children. Am J Phys Anthropol 21(2):205–213. https://doi.org/10.1002/ajpa.1330210212
26. Wang J, Shang Y, Su G, Lin X (2006) Sim0075lation of aging effects in face images. In: Intelligent computing in signal processing and pattern recognition. Springer, Berlin, pp 517–527
27. Guo G, Guowang M, Fu Y, Huang TS (2009) Human age estimation using bio-inspired features, pp 112–119. Published. https://doi.org/10.1109/CVPR.2009.5206681
28. Chang K, Chen C (2015) A learning framework for age rank estimation based on face images with scattering transform. IEEE Trans Image Process 24(3):785–798. https://doi.org/10.1109/TIP.2014.2387379
29. Guo G, Fu Y, Dyer CR, Huang TS (2008) Image-based human age estimation by manifold learning and locally adjusted robust regression. IEEE Trans Image Process 17(7):1178–1188. https://doi.org/10.1109/TIP.2008.924280
30. Anand A, Labati RD, Genovese A, Muñoz E, Piuri V, Scotti F (2017) Age estimation based on face images and pre-trained convolutional neural networks, pp 1–7. Published. https://doi.org/10.1109/SSCI.2017.8285381
31. Rojas R (2013) Neural networks: a systematic introduction. Springer Science & Business Media, Berlin
32. Bishop CM (2006) Pattern recognition and machine learning. Springer, New York
33. Fukushima K (1988) Neocognitron: a hierarchical neural network capable of visual pattern recognition. Neural Netw 1(2):119–130. https://doi.org/10.1016/0893-6080(88)90014-7
34. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958
35. Kumar SK (2017) On weight initialization in deep neural networks. arXiv preprint arXiv:1704.08863
36. Jayaraman J, King N, Roberts G, Wong H (2011) Dental age assessment: are Demirjian's standards appropriate for southern Chinese children? J Forensic Odontostomatol 29(2):22

37. Chudasama PN, Roberts GJ, Lucas VS (2012) Dental age assessment (DAA): a study of a Caucasian population at the 13 year threshold. J Forensic Legal Med 19(1):22–28. https://doi.org/10.1016/j.jflm.2011.09.008
38. Jayaraman J, Wong HM, King NM, Roberts GJ (2013) The French–Canadian data set of Demirjian for dental age estimation: a systematic review and meta-analysis. J Forensic Legal Med 20(5):373–381. https://doi.org/10.1016/j.jflm.2013.03.015
39. Štern D, Kainz P, Payer C, Urschler M (2017) Multi-factorial age estimation from skeletal and dental MRI volumes, pp 61–69, Published
40. Panis G, Lanitis A, Tsapatsoulis N, Cootes TF (2016) Overview of research on facial ageing using the FG-NET ageing database. IET Biometrics 5., https://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2014.0053
41. Ricanek K, Tesafaye T (2006) MORPH: a longitudinal image database of normal adult age-progression, pp 341–345. Published. https://doi.org/10.1109/FGR.2006.78
42. Ueki K, Hayashida T, Kobayashi T (2006) Subspace-based age-group classification using facial images under various lighting conditions, pp 6–48. Published. https://doi.org/10.1109/FGR.2006.102
43. Fu Y, Zheng N (2006) M-face: an appearance-based photorealistic model for multiple facial attributes rendering. IEEE Trans Circuits Syst Video Technol 16(7):830–842. https://doi.org/10.1109/TCSVT.2006.877398
44. Burt DM, Perrett David I (1995) Perception of age in adult Caucasian male faces: computer graphic manipulation of shape and colour information. Proc R Soc Lond Ser B Biol Sci 259(1355):137–143. https://doi.org/10.1098/rspb.1995.0021
45. Suo J, Wu T, Zhu S, Shan S, Chen X, Gao W (2008) Design sparse features for age estimation using hierarchical face model, pp 1–6. Published. https://doi.org/10.1109/AFGR.2008.4813314
46. Fu Y, Guo G, Huang TS (2010) Age synthesis and estimation via faces: a survey. IEEE Trans Pattern Anal Mach Intell 32(11):1955–1976. https://doi.org/10.1109/TPAMI.2010.36
47. Azam B, Melika Abbasian N, Mohammad Mahdi D (2007) Iranian face database with age, pose and expression, pp 50–55. Published. https://doi.org/10.1109/ICMV.2007.4469272
48. Gallagher AC, Chen T (2009) Understanding images of groups of people, pp 256–263. Published. https://doi.org/10.1109/CVPR.2009.5206828
49. Ni B, Song Z, Yan S (2009) Web image mining towards universal age estimator. In: Proceedings of the 17th ACM international conference on multimedia, Beijing, China, pp 85–94. https://doi.org/10.1145/1631272.1631287
50. Sai Phyo K, Wang J, Eam Khwang T (2013) Web image mining for facial age estimation, pp 1–5. Published. https://doi.org/10.1109/ICICS.2013.6782962
51. Sun Y, Wang X, Tang X (2013) Deep convolutional network cascade for facial point detection, pp 3476–3483. Published. https://doi.org/10.1109/CVPR.2013.446
52. Jain V, Learned-Miller E (2010) Fddb: A benchmark for face detection in unconstrained settings, UMass Amherst Technical Report. http://works.bepress.com/erik_learned_miller/55/
53. Levi G, Hassncer T (2015) Age and gender classification using convolutional neural networks, pp 34–42. Published. https://doi.org/10.1109/CVPRW.2015.7301352
54. Escalera S, Fabian J, Pardo P, Baró X, Gonzàlez J, Escalante HJ, Misevic D, Steiner U, Guyon I (2015) ChaLearn looking at people 2015: apparent age and cultural event recognition datasets and results, pp 243–251. Published. https://doi.org/10.1109/ICCVW.2015.40
55. Rothe R, Timofte R, Gool LV (2015) DEX: deep expectation of apparent age from a single image, pp 252–257. Published. https://doi.org/10.1109/ICCVW.2015.41
56. Berg A, Deng J, Fei-Fei L (2010) Large scale visual recognition challenge (ILSVRC), 2010, vol 3. URL http://www.image-net.org/challenges/LSVRC
57. Hosseini S, Lee SH, Kwon HJ, Koo HI, Cho NI (2018) Age and gender classification using wide convolutional neural network and Gabor filter, pp 1–3. Published. https://doi.org/10.1109/IWAIT.2018.8369721

58. Liu X, Li S, Kan M, Zhang J, Wu S, Liu W, Han H, Shan S, Chen X (2015) AgeNet: deeply learned regressor and classifier for robust apparent age estimation, pp 258–266. Published. https://doi.org/10.1109/ICCVW.2015.42
59. Liu K, Liu H, Chan PK, Liu T, Pei S (2018) Age estimation via fusion of depthwise separable convolutional neural networks, pp 1–8. Published. https://doi.org/10.1109/WIFS.2018.8630776
60. Shang C, Ai H (2017) Cluster convolutional neural networks for facial age estimation, pp 1817–1821. Published. https://doi.org/10.1109/ICIP.2017.8296595
61. Chen S, Zhang C, Dong M, Le J, Rao M (2017) Using ranking-CNN for age estimation:742–751. Published. https://doi.org/10.1109/CVPR.2017.86
62. Duan M, Li K, Li K (2018) An ensemble CNN2ELM for age estimation. IEEE Trans Inf Forensics Secur 13(3):758–772. https://doi.org/10.1109/TIFS.2017.2766583
63. Smith P, Chen C (2018) Transfer learning with deep CNNs for gender recognition and age estimation, pp 2564–2571. Published. https://doi.org/10.1109/BigData.2018.8621891
64. Zhang H, Geng X, Zhang Y, Cheng F (2019) Recurrent age estimation. Pattern Recogn Lett 125:271–277. https://doi.org/10.1016/j.patrec.2019.05.002
65. Paes EC, Teepen HJLJM, Koop WA, Kon M (2009) Perioral wrinkles: histologic differences between men and women. Aesthet Surg J 29(6):467–472. https://doi.org/10.1016/j.asj.2009.08.018
66. De Tobel J, Radesh P, Vandermeulen D, Thevissen PW (2017) An automated technique to stage lower third molar development on panoramic radiographs for age estimation: a pilot study. J Forensic Odontostomatol 35(2):42–54
67. Kim J, Bae W, Jung KH, Song IS (2019) Development and validation of deep learning-based algorithms for the estimation of chronological age using panoramic dental x-ray images
68. Huang G, Liu Z, Maaten L v d, Weinberger KQ (2017) Densely connected convolutional networks, pp 2261–2269. Published. https://doi.org/10.1109/CVPR.2017.243
69. Al-Khateeb H, Epiphaniou G, Reviczky A, Karadimas P, Heidari H (2018) Proactive threat detection for connected cars using recursive Bayesian estimation. IEEE Sensors J 18(12):4822–4831. https://doi.org/10.1109/JSEN.2017.2782751