GRADUATION PROJECT

Degree in Dentistry

ARTIFCIAL INTELLIGENCE IN DENTAL DIAGNOSIS

Madrid, academic year 2024/2025

Identification number: 125

RESUMEN

Introducción: Con la reciente popularidad de la inteligencia artificial, los investigadores de la odontología han comenzado a entrenar y estudiar numerosos tipos de modelos de inteligencia artificial con el objetivo de proponer herramientas para ayudar a la decisión clínica o como herramienta de estimación de la edad. Para este propósito, a inteligencia artificial podría entrenarse con el método tradicional o sin intervención humana, los métodos tradicionales de Demirjian y Willems son opciones populares para la estimación de la edad, el resultado es confuso respecto de si los métodos son universalmente aplicables, el propósito de esta investigación documental fue explorar el estado actual del entrenamiento de la inteligencia artificial con respecto a la estimación de la edad; **Objetivo**: Evaluar la exactitud de la estimación de la edad basada en inteligencia artificial en pacientes pediátricos comparando la edad predicha por inteligencia artificial obtenida a partir de la ortopantomografía con su edad cronológica; **Material y métodos**: Se utilizó la base de datos PubMed para recopilar estudios relacionados con la estimación de edad basada en inteligencia artificial utilizando ortopantomografía en pacientes pediátricos (edad ≤18 años) publicados en los últimos cinco años (2019-2025); **Resultados**: Un total de doce estudios cumplieron los criterios y se incluyeron en esta investigación documental con una precisión que va del 72,33% al 100% o un error absoluto medio que va de 0,28 a 1,659 años; **Conclusiones**: La inteligencia artificial tuvo un buen desempeño en general en la estimación de la edad entre pacientes pediátricos utilizando ortopantomografía.

PALABRAS CLAVE

Odontología, Inteligencia artificial, Estimación de edad, Ortopantomografía, Odontopediatría.

ABSTRACT

Introduction: With the recent leap and popularity of artificial intelligence, researchers in field of dentistry had started to train and study numerous types of artificial intelligence models with its objective to propose tools to help clinical decision or detection tool, in terms of age estimation, artificial intelligence could be trained with traditional methods or without human interference, whilst the traditional methods of Demirjian and Willems are popular choices for age estimation, the result remains confounded whether traditional methods are universally applicable to all populations, the purpose of this documentary research was to explore the current status of artificial intelligence training regarding age estimation; **Objective**: To evaluate the correctness of artificial intelligence-based age estimation in pediatric patients by comparing artificial intelligence predicted age obtained from orthopantomography to their chronological age; **Material and methods**: PubMed database was used to collect full text studies related to artificial intelligence-based age estimation using orthopantomography in pediatric patients from last five years (2019-2025) , sample age must be equal or younger than eighteen years old; **Results**: A total of twelve studies met the criteria and included in this documentary research with accuracy ranging from 72.33% to 100% or mean absolute error ranging from 0.28~1.659 year; **Conclusions**: Artificial intelligence performed generally well in age estimation among pediatric patients using orthopantomography.

KEYWORDS

Dentistry, Artificial intelligence, Age estimation, Orthopantomography, Pediatric dentistry.

INDEX

1. INTRODUCTION

1.1. Artificial intelligence

The study of artificial intelligence (AI) started as soon as general-purpose computer or personal computers (PC) were available to general public, that is, a PC is able to run multiple types of programs (APP) since it is designed to run any program instead of being specialized or optimized in running one specific program.(1) In the early days of AI use, researchers started to develop AI with deductive reasoning approach: a computer that is able to run logically within a database or a predetermined set of information and be able to make conclusions. An example would be a computer using a calendar to provide an answer as to which month has Christmas holidays; due to the logic of a computer using deductive reasoning approach, the calendar has the information about each holiday, therefore, it should consult the answer in the calendar. Although this process of using deductive reasoning sounds obvious it needs a knowledge base to make the conclusion.(2) Another approach is inductive learning, as humans often make decisions or act on the basis on everyday life experiences rather than on the basis on evidence that has been proved to be correct by countless studies, i.e.: photos of Santa Claus presented to a computer and asking it to determine the current month, in this case the computer uses incomplete data to draw a conclusion. The advantage of using inductive reasoning is that it does not require a complete knowledge base, the disadvantage is the possibility of coming to the wrong conclusion. The process of computer learning from partial evidence or examples to make the prediction is also called machine learning (ML).(2) Nowadays, the current focus of AI research can be broadly classified into these two distinct schools of thought, with the objective of integrating these approaches in a way that reflects how humans act and make decisions in a meaningful way (2), recognizing that the relative influence of these two modes of thought in guiding human action and decision-making differs between individuals.

1.1.1. AI learning methods

AI processes data in primarily two ways, unsupervised learning and supervised learning (2), an example of this would be an scenario in which two groups of students are asked to categorize images of teeth, the supervised learning group receives the help from a professor , who tells them where to look, what to look, to focus on color and shape of the teeth, while the unsupervised group may even categorize those images by features usually not taken into consideration such as inclination, rotation, relative position of the teeth.

There is a minimal difference between these two learning methods: the unsupervised learning method tends to include every feature and characteristics of the data presented, while supervised learning method has guidance during categorization of the data. Humans are closer to unsupervised learning as they store memories of daily life which serves no specific purpose but may be retrieved later if needed. However, in the case of learning how to drive a car is an example of supervised learning, in which the experience is marked as important to achieve a specific goal or purpose.(2)

In terms of AI using these two learning methods, unsupervised learning is useful for reducing data size because it summarizes every characteristic of data set. On the other hand, supervised learning summarizes the data with labels or characteristics previously provided by a human. The superiority of one learning method over another depends on the goal the human user wants to achieve with AI.

1.1.2. Artificial neural network

Among many new methods derived from supervised learning, reinforcement learning has gained much popularity in recent years as it is believed to be closely resembling how biological beings learn,(2) Essentially, it is a combination of a supervised learning method with provided feedback while AI tries to categorize data. Using the example mentioned above of students being asked to categorize photos of teeth while the professor guides them, each time the students made attempt to classify a picture, in this case they would be given a grade; the grade is a form of feedback which may positively reinforce the student's ability to classify various characteristics of teeth.

To mimic how biological beings learn, researchers have developed artificial neural network (ANN) to simulate how nervous system function (3). As the name suggests, ANN is trying to imitate the neural structure of nervous system by using the neuron as the basic unit to process stimuli. Neurons are "connected" to each other: the dendrites receive signals from other neurons, the axon hillock overwatches the signal received and then transmits a new signal via axon terminal (4) In terms of ANN, it could be described as a group of computer nodes that are connected to each other and once a computer receive enough data from others it will send out a new data that might be what the user requested. The ANN learns by modifying the strength of the signal/data received: when a single neuron of ANN is doing the calculation of overall data received, the weaker signal will have less "weight" in the new data (2,5); in other words, the new data will have less information related to the weak signal. For example, if computer one is black, while computer two is white and both send a signal to computer A for calculation, the resulting new data should be grey, and if there is a difference in signal strength between the computers one and two, the resulting color will be grey although the shade may vary in different degrees. ANN typically consists of at least two layers of computers/nodes, an input layer specialized in receiving data like sensory neurons and an output layer specialized in carrying out new data like motor neurons; using the previous example, computer one and two should be located in first layer to receive information while computer A should be located in second layer to generate new data. If more layers are added in between input and output layers, the resulting multilayer ANN structure will also be called “deep learning” (3).

1.2. Age estimation

Age estimation is often used in epidemiology, dentistry, and forensics to determine the age of the subject. In epidemiology it is used to explore the population growth pattern or to assess the impact of environmental factors on growth (6). In dentistry it is often required to know the chronological age (CA) and the dental age (DA) of the patient for a tailored treatment plan since humans have two sets of dentitions. Sometimes, determining DA is necessary to establish the treatment plan for pediatric patients, as in many cases the dentist must prioritize preserving the permanent tooth over the primary tooth if exfoliation is imminent. In the case of delayed teeth eruption, dentists may also require to investigate if DA was lacking behind CA regarding teeth of the patient or in relation to the general condition of the patient. In forensic science, it is mainly used for legal purposes such as identity verification, cases involving individuals who do not remember their age, or those who falsely claim to be a different age (7).

There are various methods to predict age in case the information of CA is not available, in clinical setting dentists perform rough age estimation by checking eruption sequence with radiographs. If further investigation for CA is required, practitioners may employ methods to examine bone/skeletal age or DA, by obtaining radiographs of hand and wrist for skeletal age. The most common one for assessing bone age was Tanner-Whitehouse method (8) which used radiographs with a various combination from bones of hand including radius and ulna, which were then categorized into different stages of development, each categorized radiograph was then graded with a maturation score, the total combined score is used to estimate bone age (9). In case of DA need to be examined, Demirjian (10) or Willems (11) methods are commonly accepted. Demirjian employed the same rationale much like Tanner-Whitehouse method except he used seven teeth including incisors, canine, premolars and molars from the left side of the mandible using orthopantomography (OPG), each tooth is then graded with dental formation from stage 0 then A through H based on the shape of the teeth. Stage 0: there is no sign of calcification. Stage A: the appearance of the first sign of calcification with a cape/inverted cone on the radiograph. Stage B: cusp can be seen and formed a rough outline of occlusal surface. Stage C: a complete occlusal surface formed by enamel can be seen with a cape-like/curved extension toward cervical level. Stage D: crown development surpassed cement-enamel junction (CEJ) and a rough ceiling of pulp horn can be seen in anterior teeth with the shape of umbrella, in posterior teeth a rough ceiling of pulp horn with the shape of hump from camel can be seen. Stage E: the length of the root was shorter than crown, near-vertical walls of pulp chamber can be seen for anterior teeth and premolars, first sign of root bifurcation can be seen as a single point of calcification or in the form of a cape. Stage F: the length of the root was equal or longer than the crown, initial form of open root apex can be seen on all teeth, a sharp triangle formed by the walls of pulp chamber can be observed for anterior teeth and premolars, root bifurcation can be seen developed further downward alongside with root. Stage G: paralleled walls of root canal can be observed; apex was still partially open. Stage H: the closure of the root apex with a uniform width of periodontal ligaments on the root. Once each tooth was graded with dental stage, it was converted into weighted score, then the sum of it can be consulted with maturity graph showing the estimated range of DA (10). The method from Willems (11) is a modified version of Demirjian, it uses the same method to record and grade teeth formation but directly converts dental formation stage into DA instead of doing double conversion like Demirjian method (11).

A more recent method to determine DA was developed by Cameriere et al. (12) in 2006, it used OPG and required the following metrics from each tooth on the left side of mandible except third molar: width of the open apex, length of the tooth, closed apex or complete development of teeth and gender of the patient, the above measurements were then used in a formula developed by Cameriere to calculate the value for dental age. In 2008 Cameriere et al. (13) developed additional method called third molar maturity index (I3M) to determine the probability of the age of patient is over or under eighteen years old, solely using third molar from left mandible of an OPG with similar metrics required for calculation (apex width, length of tooth and gender). The author has participated in several age determination studies including using bones from hand (14) or clavicle (15), the most recent study in which Cameriere was involved was a comparison study of age prediction between AI and manual method (16).

While aforementioned methods for age estimation are commonly accepted, alternative methods include dentin biopsy, dentin racemization, occlusal scheme, assessment of anatomical features of teeth such as coronal height/width. Other indicators assess hard tissues changes such as colors, attrition, size of pulp chamber/secondary dentin, cementum thickness, periodontal shrinkage and root transparency (7).

1.3. Artificial intelligence in dentistry

Due to the advancement of AI and hardware, it can process vast amounts of data in short notice, reduce repetitive tasks which might cause human errors, and operate with minimal human interference, fast enough to provide real-time responses (3).

The current applications of AI in dentistry are: automation of dental record, recommendation of diagnosis and treatment options by training AI to process patient data such as clinical history and radiographs, prediction of future intraoral disease such as caries or gum disease so dental practitioners can employ preventive measures, analysis of radiographs or other dental images to highlight dental structure, location of anatomy landmarks in orthodontics, determination of caries lesions, assessment of bone loss in periodontology, root canal in endodontics, implants in oral surgery or detection of problems that are not easily found by human eyes such as vertical root fracture and prediction of possible failures of dental equipment (3).

Current applications of AI in pediatric dentistry include dental plaque evaluation, detection of fissure sealant, supernumerary teeth and mesiodens, assessment of ectopic eruptions, impacted teeth and CA, classification on tooth developmental stage (3).

1.4. Age estimation using artificial intelligence and orthopantomography

In terms of training AI to do image classification using OPG, researchers needed to standardize each sample by cropping, rotating, adjusting grey scale or resizing. A train-validate-test split of OPGs should then be assigned with the following percentage on each group: 50%-25%-25% (2) or, based on the purpose and objective of the study, each OPG should only exist in one group. With an extremely large sample size (more than thousands) it was recommended to employ higher percentage in training group as estimation errors produced by validation and testing group would be quite low (2).

The computer system perceived image data by converting each pixel from the image into a grid i.e. a 500*500 grid, each empty space on the grid would then be converted with information about the pixel in descriptive form (usually numbers), including but not limited to: color, brightness, connection or disconnection to nearby pixels, nearby pixels pertained the same information as self (3). An image that was converted by computer would be a grid with strings of numbers inside each space if perceived by humans. Converted images then underwent several downsampling (pooling) layers and convolutional layers (3), which the purpose of downsampling was to reduce the size of the image into lower resolution yet still pertain needed information and increase processing efficiency while the purpose of convolutional layer was to generate feature map that contained information from large scope to small scope, i.e. an outline of mandible or inferior alveolar canal, an outline of second molar or shape of mandibular angle, to the apex from the root of premolar or root curvature. The feature maps were used to filter out non-matched images or to propagate images that had close resemblance to feature map to the next layer, the deeper the layers go the smaller the feature maps became, i.e. the image of canine that matched the feature map portraying a canine with straight developing walls from root canal with be propagated to a deeper level that used another feature map with smaller scope which evaluate if the open apex of canine exceeded three pixels. The combination, the number and the order of convolutional layers and downsampling layers used for training AI depend on study design. The final layers (output, fully connected layer) were made for grouping images with similar features together in order for classification to be made (3). AI model that contains downsampling, convolutional and output layers is typically referred as convolutional neural network (CNN) (17).

Depend on study design, researchers could train AI prediction model using data labeled by humans, where the AI was trained under supervision i.e. using Demirjian's method (10); First, images that were assigned to validation and testing groups will be evaluated and categorized through stage 0, A through H, based on visual cues and descriptions of each developmental stage by field experts. An AI prediction model was then prompted to perform the same task: to do classification on teeth development stage based on the validation data (labeled data) previously categorized by experts. Once the AI completed the classification using Demirjian method, the converted sum of weighted score can be consulted through maturity graph to get an estimation on DA.

The AI prediction model can also be trained without the need of a human-made classification. In this case, the AI learns to predict the age of the patient unsupervised (end-to-end). The OPGs assigned to validation and testing group would only contain the image itself and the CA of the patient. AI without labeled data learned by grouping or clustering (3) similar pixels together based on color intensity (brightness or contrast), distance between nearby pixels and position on the grid to. One of the many methods to cluster pixel is to start grouping pixels nearest to each other then gradually combine groups of pixels that were closest to each other. The other method is for AI to assign each pixel to core point, border point or noise in a given space based on density. AI in unsupervised training tried to generate a small-scale feature map that contained color information, border and grain texture at first, and then gradually produced a larger scale feature map that contained tissue, shape, the entire organ or structure. (3)

Nowadays AI training with unsupervised training is deemed rare but still needed because it serves as a method to detect anomalies: unusual relationships between study group and control group.

Training data, as the name suggests, is used for training AI. OPGs belonging to this group can be used multiple times throughout the training process. Validation data was used alongside training data to assist the training of the AI or to obtain an estimation of accuracy or correctness. Test data was used for final evaluation and should be used when the training on AI was deemed satisfactory. OPGs belonging in this group can only be used once to determine the actual accuracy of the trained AI model. During AI training/building sessions OPGs from training group and validation group are used, following the above recommended train-validate-test split, one out of every three OPGs would be from validation group to estimate accuracy. OPGs from train and validation groups can be used/looped multiple times during training or as required by the research objectives.

Once the estimate accuracy is deemed acceptable, the trained AI is then evaluated using OPGs from test group. The performance metrics used, include but are not limited to, the following: accuracy and mean absolute error (MAE).

Accuracy measures the correctness of AI classifications, calculated as the total number of true positive and true negative responses divided by the total number of attempts (18), with a perfect score being 100%. In the context of age estimation, AI determines whether a given OPG belongs to a specific age or age group, making accuracy a straightforward metric for qualitative evaluation.

On the other hand, MAE quantifies the average magnitude of error (5) by measuring the average difference between predicted value and actual value, the perfect score being zero. In age estimation, AI predicts patients’ age based on OPG and the average difference between predicted age and CA is computed. It provided quantitative assessment that often requires comparison with a control group to determine effectiveness.

Other common ML models that are used in AI research for age estimation in dentistry are: gradient boosting decision tree (GBDT), K nearest neighbor (KNN).
GBDT is a hybrid model that contains two types of ML (2). Decision tree is an algorithm that contains numerous nodes that are spread out in a tree or umbrella shape in hierarchical order. Each node contains univariate criteria or multivariate criteria for the data to pass through to the next node: its function is similar to diagnostic flowchart/tree. Gradient boosting/boosted is an algorithm that "learns from bad example": each time/round after AI made prediction, it would pick the poor prediction and then modify the old ML model to a new one, so the prediction made will be closer to the true value in the upcoming round of training (17). When combined the aforementioned two algorithms, in the case of age estimation, a GBDT model would run multiple decision trees at the same time and take into consideration those trees that made poor age predictions, in order to modify/improve it. In the upcoming round the new and improved decision tree can make age prediction closer to CA over the course of multiple rounds of training.
KNN functioned based on one notion: objects that belong to the same group share the same characteristics/features i.e. grass is green and flowers are red, if it is green it must be grass. KNN works by laying out every features in a two-dimensional or three-dimensional (3D) space (17), which may requires multiple spaces to distribute every feature extracted from sample; once all features are allocated/saved in 3D feature spaces or database, the classification begins: new sample will be categorized based on the nearest sample in 3D feature space. In terms of using OPG for age estimation, it can use length of canine as X axis, diameter of the apex in canine as Y axis and genders as Z axis in feature space, once every OPGs has been laid out in 3D space, new OPG introduced can be categorized based on its nearest sample; if the nearest sample belongs to an OPG of a nine years old male patient with canine length of forty pixels, apex width of five pixels, then the newly introduced OPG will be put into the same category of nine years old male.
In 2021, Zaborowicz et al. (19) introduced a series of tooth geometric indicators located in mandible and teeth to train AI, which included but not limited to: ratios of occlusal distance between right canines and right second premolars (X01), ratio between lower right canine length and pulp length (X07), ratio between lower right canine width and pulp chamber width (X11), ratio between distance from occlusal surface of lower right canine to inferior border of mandible and distance from root apex of the same canine to inferior border of mandible (X15)

and ratio between the distance from canine root apex to inferior border of mandible and distance from second premolar root apex to inferior border of mandible (X19), a total of twenty-one custom tooth geometry indicators were used in the study which the results were satisfactory with high accuracies. Said method (19) required researchers to prepare sample/OPGs for AI to learn in validation and testing groups therefore cannot be categorized as end-to-end training.

1.5. Justification

Whilst the traditional methods of Demirjian and Willems are popular choices for age estimation, the result remains confounded whether traditional methods are universally applicable to all populations, studies had shown Demirjian method overestimates CA and might be suited for a specific population (6), one author even argued that Demirjian method was not meant for age estimation originally (20) or Willems method is more inclined to universal population (6,21).

With the recent leap and popularity of AI, researchers have started to train and study numerous types of AI models in hopes to propose a tool to help with clinical decisions or simply a screening tool in terms of age estimation, some train AI in conjunction with Demirjian or Willem methods, some train only with OPG and CA provided. The purpose of this review is to explore the current status of AI training regarding age estimation.

2. OBJETIVE

Primary objective:

To evaluate the correctness of artificial intelligence-based age estimation in pediatric patients by comparing AI-predicted age obtained from orthopantomography to their chronological age.

3. MATERIAL AND METHODS

For the review, the selection process for articles follows the PRISMA guidelines (22), PubMed database is used to collect full text studies on AI-based age estimation using orthopantomography in pediatric patients from last five years (2019-2025). The time frame was chosen for two reasons: first, the search yielded a sufficient number of studies. Second, AI became popular and established its application in recent years. In order to address the objective a research question following PICO framework was used and as follows:

“In pediatric patients (P), does AI diagnosis of age estimation (I) via orthopantomography predicts (O) regarding chronological age (C) correctly?”

The search was performed on PubMed using the following formula:

((ai artificial intelligence[MeSH Terms]) OR (Computer Reasoning OR machine Intelligence OR machine learning OR neural network OR deep learning)) AND (Radiography, Dental[MeSH Terms]) AND (pediatric dentistry[MeSH Terms] OR children[MeSH Terms] OR adolescent[MeSH Terms] OR Infant[MeSH Terms] OR Infant, Newborn[MeSH Terms] OR Child, Preschool[MeSH Terms] OR childhood)

A total of fourteen keywords used, evenly divided between free terms and controlled MeSH terms from the national library of medicine (NLM). The majority of MeSH terms were selected to define pediatric patients.

The initial search was conducted on 5th of November 2024 and final search was completed on the 25th of February 2025, yielding 69 studies after applying the following filter:

- Publication date: last five years
- Text availability: Full text
- Age: Child: birth-18 years
- Sort by: Most recent.

Each abstract was screened with the following inclusion criteria:

- Published within five years (2019-2025)
- Full text available and written in English
- Study and control/testing group consist of pediatric patients (0-18 y.o.) or a specific subset within this range
- Evaluation of AI-based age estimation specifically using orthopantomography (OPG)
- Reported testing group or its subset age group had at least one of the following performance metrics in study: accuracy or mean absolute error (MAE)

Studies were excluded if they met any of the following exclusion criteria:

- Systemic reviews, literature reviews, pilot studies or case studies
- Lack of age information for patients nor in its specific subset
- Focus solely on AI classification of Demirjian method, direct prediction of chronological age(CA) was not made.

After applying these criteria, twelve studies remained. The following data were extracted from each selected study: date of publication, population origin, sample size, sample age, performance metrics (accuracy or MAE that is within eighteen-years-old or younger), and Cohen's kappa.

4. RESULTS

The PubMed database search yielded sixty-nine articles. Title, abstract, tables and figures were examined to their relevance to the research objective.

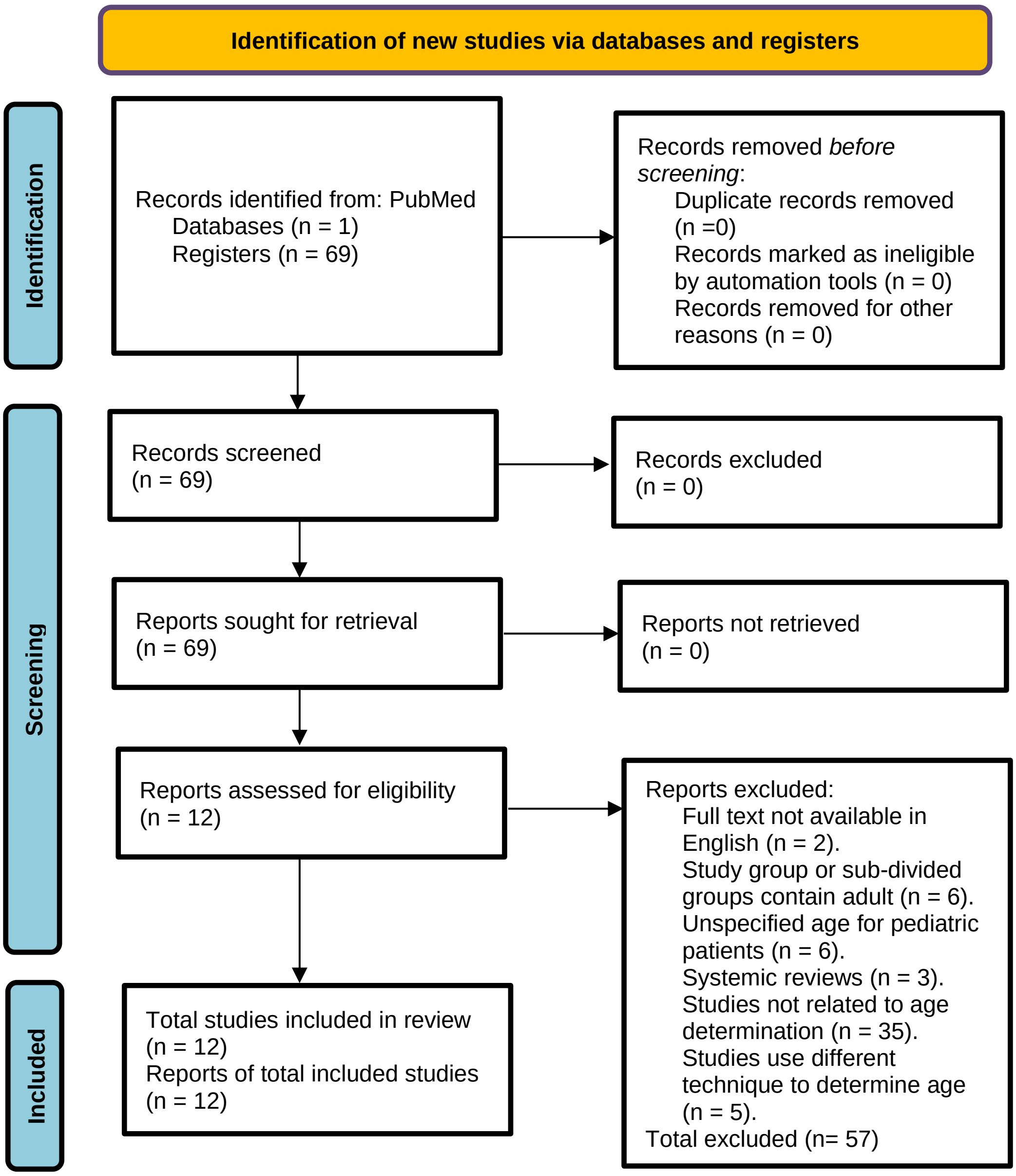
A total of fifty-seven studies were excluded based on the following reasons:

- Full text not available in English
- Subset study groups included adults aged 19 years or older
- Pediatric patient age unspecified
- Systemic reviews
- Studies unrelated to age determination, including research on:
    - Tooth classification and segmentation
    - Detection of caries, mesiodens, missing teeth, anatomical landmarks, gender, caries, cleft palate, osteoporosis, third molars, second premolar, missing teeth, inferior alveolar nerve canal, ectopic eruption
    - Use of DNA markers for CA determination
    - Image generation
    - Prognosis prediction for extraction
- Studies related to age estimation but using image modalities other than OPG, such as:
    - Cone-beam computed tomography (CBCT)
    - Magnetic resonance imaging (MRI)
    - Bitewings radiographs
    - Hand-wrist and knee radiographs

Among the excluded studies, 35 out of 57 were removed due to their irrelevance to the research objective. After final screening, 12 studies were deemed eligible for inclusion.

The following data were extracted: Main author, year of publication, population origin of sample, sample size, sample age, performance metrics of accuracy and MAE, training design.

PRISMA flow diagram (22) regarding article selection can be found below (Figure 1):



PRISMA flow chart (1).

Table 1 Results table

Author and publication year	Population	Weight %	Sample size	Age	Performance metrics		Training design
					Accuracy %	MAE	
Sivri et al. 2024 (23)	Turky	17.7	5898	4~17	-	0.73 year	PANO
Faragalli et al. 2024 (16)	South Africa	5.8	1949	5~14	-	0.69~0.74 year	CAM
Shi et al. 2024 (24)	China	2.0	673	3.07~14.95	-	0.28~1.76 (0.72) year	DEM
Wang et al. 2023 (25)	China	22.9	7649	6~17*	72.33~93.63 %	-	PANO
Wu et al. 2022 (26)	Taiwan	6.1	2052	2.6~17.7	-	0.6916±0.0010~ 0.8801±0.0069 year	DEM
Shen et al. 2022 (27)	China	2.2	748	5~13	-	0.473~0.584 year	DEM, CAM
Shan et al. 2022 (28)	China	4.4	1477	2~17.99	-	0.441~1.659 year	DEM, WIL
Bunyarit et al. 2022 (29)	Malaysia	3.0	1015	5~17.99	94.1%, 92.8%	-	DEM
Sharifonnasabi et al. 2022 (30)	Malaysia	5.8	1922	15~18*	97.43~100%	-	PANO
Zaborowicz et al 2022 (31)	Poland	1.9	619	4~18	-	7.53~15.37 months	CUS
Guo et al. 2021 (32)	China	21.4	7161	5~18*	90.5~95.9%	-	PANO
Galibourg et al. 2021 (33)	France	6.7	2230	4~16*	-	0.729 ±0.025 ~0.812 ±0.028 year	DEM

MAE: mean absolute error, using year as unit of measurement

CAM: AI was trained with Cameriere method

DEM: AI was trained with Demirjian method

WIL: AI was trained with Willems method

PANO: AI was trained only with OPG

CUS: custom geometry indicators measured in OPG.

Sample size: numbers of OPG used in study

* Studies had subset study groups that met inclusion criteria (under or equal to 18 years old) which derived from a larger range of ages, the metrics shown in table only reflect sample equal to or under 18 years old.

The results are presented in Table 1. all 12 studies were comparative studies that evaluate AI-predicted age against CA of pediatric patients (under 18 years old) using OPG.

5. DISCUSION

The results obtained indicate that AI-based age estimation in pediatric patients using OPG demonstrates generally good performance, with accuracy ranging from 72.33% to 100% or MAE ranging from 0.28~1.659 years.

5.1. Training Methodology of AI Models

Only four studies (23,25,30,32) trained AI models using only OPG images, the remaining eight studies incorporated manual classification methods, such as Demirjian, Cameriere or Willems methods.

5.2. Sample Size weighting

Three studies (23,25,32) accounted for largest sample weights, contributing 22.9%, 21.4% and 17.7% respectively or 62% in total, all three studies exclusively trained AI models without integrating additional classification methods (e.g. Demirjian method).

5.3. Trends in AI Training and evaluation metrics

A notable trend in AI training design and performance metrics is the incorporation of manual classification to train AI. Studies that trained without manual classification (25,30,32) primarily reported accuracy and its related metrics (precision, sensitivity, F1 score). In contrast, studies incorporating manual classification to train AI (16,24,26–28,31,33) used MAE and its related metrics (mean square error, root mean square error, R-square). The choice of performance metrics depends on AI training design and researchers' preferences. For instance: Shi et al. (24) reported precision, sensitivity, F1 and MAE but did not include accuracy, Bunyarit et al. (29) reported accuracy yet used Demirjian method to train AI while Sivri et al. (23) reported MAE yet solely used OPG to train AI.

5.4. Performance measured by accuracy

The studies by Wang et al. (25) and Guo et al. (32) highlight that differences in AI model design and evaluation design can influence results of accuracy. Both studies used similar AI model for comparison, yet Wang reported a lower-end accuracy of 72.33%, which was the lowest accuracy of all studies presented in Table 1, whereas Guo reported 90.5%. Wang attributed this discrepancy to differences in training methodology and evaluation: Guo incorporated a binary judgement design (assigning a numeric value of 1 for AI predicted age equal to or older than a specified threshold and a value of 0 for AI predicted age younger than threshold) and

only tasked AI to predict three age thresholds (if predicted age from a given OPG is higher or lower than 14,16 and 18 years-old), whereas Wang's AI model classified patients into four age groups (6-8, 9-11, 12-14, 15-17 years old). These differences in classification and evaluation design are likely to have contributed to the variation in reported accuracy.
Wang et al. (25) expressed the reason of not using manual method to train AI for its limitation of inter and intra-observer errors and the reproducibility of classification technique.

The highest accuracy presented in Table 1. was from Sharifonnasabi et al. (30) ranging from 97.43%-100% with 98.845% in average, the author contributed the result of high accuracy to the combination of two ML models: CNN and KNN. Furthermore, the authors introduced additional 130 OPG which were not used in training with different ethnicity, gender and age, normal dentition, dentition under orthodontic treatment with brackets and wires, mandible underwent surgery or OPG with missing teeth to evaluate the accuracy of the hybrid model which the author claimed it was one of the distinct features of said study.

5.5. Performance by MAE

Shi et al. (24) used a three-step framework for their study: first, they trained AI to detect and segment each tooth in OPG; second, they trained it to classify each tooth with Demirjian method; third, they trained it to estimate dental age. Shi et al. (24) reported the lowest MAE of 0.28 years among studies presented in Table 1, which indicated that the lowest error value between predicted age and CA was obtained in the age group of 5-6 years old. The highest value of MAE among all studies was also seen in the same study (24) with a MAE of 1.76 years in group of 14-15 years old, with an average MAE of 0.72 year in their study. The authors attributed the highest MAE in their study to the almost complete eruption of permanent dentition at around fourteen years of age except for the third molars which were still under development but were not used in their research. (24)
Shan et al. (28) conducted a comparative study between Demirjian method, Willems method, Demirjian method with modified weight scores and numerous ML models trained by Demirjian method in age prediction, hence the presence of high degree of difference in MAE from their study (0.441-1.659 years). The highest performing ML model is GBDT with MAE of 0.441 years; According to the study of Shan et al. (28), five out of ten ML models used had lower MAE than manual methods (Demirjian, Willems).
Shen et al. (27) conducted a comparison study between the performances of three ML models under supervised learning using Demirjian method and Cameriere method; additionally it also compared performance between ML models and traditional methods. All ML models

outperformed traditional methods (Demirjian, Willems) while the best performing ML model in the study was trained with KNN algorithm supervised by Cameriere method demonstrated MAE of 0.473 year.

Zaborowicz et al. (31) continued their previous study (19), using twenty-one custom tooth geometry indicators as reference to train AI model; similar to Cameriere method which also use proportional value instead of fixed values as reference, Zaborowicz (31) used ratio of occlusal distance, ratio of root and pulp and ratio of distance to inferior mandible border. Interestingly, while under no data/sample compression, the most sensitive indicator was ratio between length of second lower premolar and its pulp length (X5) which is similar to the formula developed by Cameriere et al. (12) to determine CA: it used ratio between width of open apex from second lower premolar and length of the tooth had significant contribution to the calculation of CA. The significance of using measurement from lower second premolar to estimate age from these two authors (12,31) may shed a light on future research to focus on lower second premolar.

5.6. Relevant OPG regions for AI in unsupervised learning

In study conducted by Guo et al. (32), the authors compared the performance of accuracy between manual method by Demirjian and three unsupervised ML prediction models, the results shown that the accuracy to differentiate OPGs into three age groups by ML models was higher than manual method. To discover possible factors influencing the high performance of AI; Guo et al. (32) drew a series of heat maps to show where exactly the AI models looked at OPGs during testing phase. AI was found focusing on “empty spaces” (low density areas) such as pulp chamber, periodontal ligament or membrane; the region between nearby teeth and the zone between temporary and permanent teeth: unlike humans, who focused on high density areas such as tooth structure. Another factor might be the possible interference from inter- and intra-observer error due to subjective view from field experts while evaluating developmental stage from Demirjian method. The authors also noted that the Demirjian method included only eight stages, which may not fit the relationship between CA and teeth development perfectly. The study conducted by Guo et al. (32) demonstrated that unsupervised (end-to-end) ML models without manual interference can extract features in OPG that were highly related to CA.

5.7. Limitations and suggestions for future research

This documentary research initially returned sixty-nine articles using a combination of fourteen keywords from a single database yet upon screening only twelve articles met the criteria, it could be a result of poor selection of keywords used or indicates that more databases are needed to yield relevant search result. During data extraction it was found that there were no universal metrics to measure AI performance although there was a pattern to follow either employ accuracy or MAE, no studies presented in this research provided both performance metrics. It is understandable for studies that acquired large sample size preferred to train AI without human interference because it is less time-consuming, no inter- and intra-observer error. Studies that train AI with man-made classification require time for human experts to manually identify tooth stage or measure length or diameter of tooth that were going to be used in validation and testing group, the number of OPGs that need to be manually calibrated increase proportionally as sample size grow; No studies presented in this research compared the performance between end-to-end AI model and AI model trained with manual method. For future studies, it is recommended for researchers to explore difference in performance, time invested and greenhouse gas emission/carbon footprint between AI models trained with man-made classification and AI trained without human interference.

6. CONCLUSIONS

Artificial intelligence performed generally well in age estimation among pediatric patients using orthopantomography from studies published between 2019-2025 with accuracy ranging from 72.33% to 100% or mean absolute error ranging from 0.28~1.659 years. Within limitations of this documentary research, more comparison studies between AI under supervised learning and unsupervised learning is recommended.

7. SUSTAINABILITY

One question remained to be answered: “Is building a dam or water supply system, a worthy investment for the future and sustainability?” the question can also be applied for artificial intelligence. AI technology is being used among daily life, it proves to be efficient, yet it has not been deployed in medical practices. The hardware (data center) and electricity needed to support a functional AI is currently responsible for the increasing 4% of greenhouse gas emission globally (34), one way to improve power consumption of AI is to implement an energy-aware feature on the system so it can predict power consumption according to data usage in order to operate in regulated energy profile. The three pillars of green software (34) which are: efficient coding, the possibility to develop/write the same code with less energy, platform that supports the software is running with minimal energy, can also be utilized in AI system to improve environmental sustainability. With the help of AI tool in age estimation, chronological age can be estimated on the fly without using manual methods, especially useful in scenario requiring large-scale screening. The above recommendations and advantage correspond to sustainable development goals (SDGs) (35) number 3 and 13.

8. REFERENCES

1. Bruce Jacob, Spencer W. Ng, David T. Wang. OVERVIEW: On Memory Systems and Their Design [Internet]. Morgan Kaufmann; 2008. 54 p. Available from: https://doi.org/10.1016/B978-012379751-3.50002-3

2. Aggarwal CC. Artificial Intelligence: A Textbook [Internet]. Cham: Springer International Publishing; 2021 [cited 2024 Nov 10]. Available from: https://link.springer.com/10.1007/978-3-030-72357-6

3. Orhan K, Jagtap R, editors. Artificial Intelligence in Dentistry. 1st ed. 2023. Cham: Springer International Publishing; 2023. 1 p.

4. Pfaff DW, Volkow ND, Rubenstein JL, editors. Neuroscience in the 21st Century: From Basic to Clinical. 3rd ed. 2022. Cham: Springer International Publishing; 2022. 1 p.

5. Shanmuganathan S, Samarasinghe S, editors. Artificial Neural Network Modelling [Internet]. Cham: Springer International Publishing; 2016 [cited 2024 Dec 5]. (Studies in Computational Intelligence; vol. 628). Available from: http://link.springer.com/10.1007/978-3-319-28495-8

6. Esan TA, Yengopal V, Schepartz LA. The Demirjian versus the Willems method for dental age estimation in different populations: A meta-analysis of published studies. PloS One. 2017;12(11):e0186682.

7. Rötzscher K. Forensic and legal dentistry. Cham Heidelberg: Springer; 2014.

8. Tanner JM, Cameron N. Assessment of skeletal maturity and prediction of adult height (TW3 method). 3. ed. London: Saunders; 2001. 110 p.

9. Satoh M. Bone age: assessment methods and clinical applications. Clin Pediatr Endocrinol. 2015;24(4):143–52.

10. Demirjian A, Goldstein H, Tanner JM. A new system of dental age assessment. Hum Biol. 1973 May;45(2):211–27.

11. Willems G, Van Olmen A, Spiessens B, Carels C. Dental age estimation in Belgian children: Demirjian's technique revisited. J Forensic Sci. 2001 Jul;46(4):893–5.

12. Cameriere R, Ferrante L, Cingolani M. Age estimation in children by measurement of open apices in teeth. Int J Legal Med. 2006 Jan;120(1):49–52.

13. Cameriere R, Ferrante L, De Angelis D, Scarpino F, Galli F. The comparison between measurement of open apices of third molars and Demirjian stages to test chronological age of over 18 year olds in living subjects. Int J Legal Med. 2008 Nov;122(6):493–7.

14. Scendoni R, Cingolani M, Giovagnoni A, Fogante M, Fedeli P, Pigolkin YuI, et al. Analysis of carpal bones on MR images for age estimation: First results of a new forensic approach. Forensic Sci Int. 2020 Aug;313:110341.

15. Scendoni R, Ribeiro ILA, Cingolani M, Giovagnoni A, Curzi M, Fedeli P, et al. A new analytical cut-off point for determining 18 years of age using MRI on medial clavicular epiphysis. Leg Med. 2022 Feb;54:102010.

16. Faragalli A, Ferrante L, Angelakopoulos N, Cameriere R, Skrami E. Do machine learning methods solve the main pitfall of linear regression in dental age estimation? Forensic Sci Int. 2025 Feb;367:112353.

17. Huawei Technologies Co., Ltd. Artificial Intelligence Technology [Internet]. Singapore: Springer Nature Singapore; 2023 [cited 2025 Apr 9]. Available from: https://link.springer.com/10.1007/978-981-19-2879-6

18. Hassouna A. Statistics for Clinicians: How Much Should a Doctor Know? [Internet]. Cham: Springer International Publishing; 2023 [cited 2025 Apr 9]. Available from: https://link.springer.com/10.1007/978-3-031-20758-7

19. Zaborowicz K, Biedziak B, Olszewska A, Zaborowicz M. Tooth and Bone Parameters in the Assessment of the Chronological Age of Children and Adolescents Using Neural Modelling Methods. Sensors. 2021 Sep 8;21(18):6008.

20. Liversidge HM. Commentary on “new systems for dental maturity based on seven and four teeth” Demirjian and Goldstein, Annals of Human Biology, 1976,3,411-421. Ann Hum Biol. 2024 Feb;51(1):2401026.

21. Prasad H, Kala N. Accuracy of two dental age estimation methods in the Indian population - A meta-analysis of published studies. J Forensic Odontostomatol. 2019 Dec 30;37(3):2–11.

22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021 Mar 29;n71.

23. Sivri MB, Taheri S, Kırzıoğlu Ercan RG, Yağcı Ü, Golrizkhatami Z. Dental age estimation: A comparative study of convolutional neural network and Demirjian’s method. J Forensic Leg Med. 2024 Apr;103:102679.

24. Shi Y, Ye Z, Guo J, Tang Y, Dong W, Dai J, et al. Deep learning methods for fully automated dental age estimation on orthopantomograms. Clin Oral Investig. 2024 Mar 7;28(3):198.

25. Wang J, Dou J, Han J, Li G, Tao J. A population-based study to assess two convolutional neural networks for dental age estimation. BMC Oral Health. 2023 Feb 17;23(1):109.

26. Wu TJ, Ling Tsai C, Huang YH, Fan TY, Chen YP. Efficacy of machine learning assisted dental age assessment in local population. Leg Med. 2022 Nov;59:102148.

27. Shen S, Yuan X, Wang J, Fan L, Zhao J, Tao J. Evaluation of a machine learning algorithms for predicting the dental age of adolescent based on different preprocessing methods. Front Public Health. 2022 Dec 1;10:1068253.

28. Shan W, Sun Y, Hu L, Qiu J, Huo M, Zhang Z, et al. Boosting algorithm improves the accuracy of juvenile forensic dental age estimation in southern China population. Sci Rep. 2022 Sep 19;12(1):15649.

29. Bunyarit SS, Nambiar P, Naidu M, Asif MK, Poh RYY. Dental age estimation of Malaysian Indian children and adolescents: applicability of Chaillet and Demirjian's modified method using artificial neural network. Ann Hum Biol. 2022 May 19;49(3–4):192–9.

30. Sharifonnasabi F, Jhanjhi NZ, John J, Obeidy P, Band SS, Alinejad-Rokny H, et al. Hybrid HCNN-KNN Model Enhances Age Estimation Accuracy in Orthopantomography. Front Public Health. 2022 May 30;10:879418.

31. Zaborowicz K, Garbowski T, Biedziak B, Zaborowicz M. Robust Estimation of the Chronological Age of Children and Adolescents Using Tooth Geometry Indicators and POD-GP. Int J Environ Res Public Health. 2022 Mar 3;19(5):2952.

32. Guo Y cheng, Han M, Chi Y, Long H, Zhang D, Yang J, et al. Accurate age classification using manual method and deep convolutional neural network based on orthopantomogram images. Int J Legal Med. 2021 Jul;135(4):1589–97.

33. Galibourg A, Cussat-Blanc S, Dumoncel J, Telmon N, Monsarrat P, Maret D. Comparison of different machine learning approaches to predict dental age using Demirjian's staging approach. Int J Legal Med. 2021 Mar;135(2):665–75.

34. Alloghani MA. Artificial Intelligence and Sustainability [Internet]. Cham: Springer Nature Switzerland; 2024 [cited 2025 Apr 10]. (Signals and Communication Technology). Available from: https://link.springer.com/10.1007/978-3-031-45214-7

35. United Nations Department of Economic and Social Affairs. The Sustainable Development Goals Report 2023: Special Edition [Internet]. United Nations; 2023 [cited 2025 Apr 21]. (The Sustainable Development Goals Report). Available from: https://www.un-ilibrary.org/content/books/9789210024914

9. ANNEXES

Abbreviation:

3D: three dimensions

AI: artificial intelligence

ANN: artificial neural network

APP: application

CNN: convolutional neural network

CA: chronological age

CBCT: cone beam computed tomography

CEJ: cement-enamel junction

DA: dental age

GBDT: gradient boosting decision tree

KNN: K nearest neighbors

NLM: national library of medicine

MRI: magnetic resonance imaging

MAE: mean absolute error

OPG: orthopantomography

PC: personal computer

SDGs: sustainable development goals