



OPEN

Age estimation of children and adolescents from mandibles using machine learning

Erika Calvano Kuchler^{1✉}, Paula Patrícia Krohn¹, Eduarda Gaspari Campos Efeiche², Livia Livia Alves Antunes^{3,4}, Allan Abuabara⁴, Bianca Marques de Mattos de Araujo², Flares Baratto-Filho^{2,4}, Svenja Beisel-Memmert¹, Christian Kirschneck¹ & Cristiano Miranda de Araujo^{2,4}

Age estimation is a crucial step in forensic identification, particularly in scenarios where dental structures may be absent. This study aimed to develop and evaluate supervised machine learning models to predict chronological age based on mandibular morphometric measurements in children and adolescents. A sample of lateral cephalometric radiographs from 401 orthodontic patients aged between 6 and 16 years was analysed. Linear and angular mandibular measurements including the total mandibular length (Co-Pog), mandibular ramus height (Co-Go), mandibular body length (Go-Gn), and the gonial angle (Ar-Go-Me) were analysed. Eight supervised machine learning algorithms were trained to predict chronological age based on these measurements and sex. The dataset was split into training (80%) and test (20%) sets, with stratified 5-fold cross-validation to prevent overfitting. Model performance was evaluated using mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R^2), with 95% confidence intervals estimated via bootstrapping. The models based on mandibular morphometric features and sex achieved a minimum MAE of 1.54 years (95% CI: 1.33–1.76) and RMSE of 1.93 (95% CI: 1.66–2.18) on the test set. Cross-validation confirmed model stability, with the Gradient Boosting Regressor achieving the best performance, showing a MAE of 1.21 (95% CI: 1.09–1.32) and R^2 of 0.56 (95% CI: 0.46–0.64). Total mandibular length (Co-Pog) and mandibular ramus height (Co-Go) were the most important predictors. Pairwise comparisons revealed statistically significant differences favoring ensemble methods over linear and simpler tree models. Supervised machine learning models demonstrated promising accuracy for age estimation based on mandibular measurements in growing individuals. Gradient Boosting emerged as the most effective algorithm. However, the generalizability of the models may be influenced by population-specific characteristics and the need for prior knowledge of certain predictor variables. Further external validations are recommended to enhance model applicability across diverse forensic contexts.

Keywords Machine learning, Artificial intelligence, Age determination by skeleton, Mandible

Forensic identification of unknown human remains fundamentally relies on the accurate estimation of biological characteristics, including age and sex. Age estimation, in particular, is critical in forensic medicine for both individual identification and mass disaster victim reconciliation^{1–3}. Beyond postmortem applications, age estimation contributes to the assessment of legal maturity, supporting decisions on whether an individual should be prosecuted as a juvenile or an adult, an important factor that can influence the severity of criminal sentencing⁴.

The mandible is the most frequently recovered bone in human remains. It is sometimes the only bone available for post-mortem investigation⁵ and it is well suited for age estimation, as the mandible exhibits more pronounced development changes than other craniofacial bones⁶. Mandibular dimensions, particularly mandibular length, ramus height, and mandibular angle (gonial angle) have shown strong correlations with age in humans⁷.

¹Department of Orthodontics, Medical Faculty, University Hospital Bonn, Welschnonnenstr. 17, 53111 Bonn, Germany. ²University of Tuiuti of Paraná, Curitiba, Paraná, Brazil. ³Postgraduate Program in Dentistry, Health Institute of Nova Friburgo, Fluminense Federal University, Rio de Janeiro, Brazil. ⁴University of the Joinville Region, Joinville, Santa Catarina, Brazil. ✉email: erika.kuchler@ukbonn.de

Recent advancements in Artificial Intelligence (AI) have introduced novel methodologies to address limitations inherent in conventional forensic age estimation techniques. Machine learning, a subset of AI, offers robust capabilities for nonlinear data analysis⁸ and has significantly improved forensic age estimation by providing faster, more standardized, and objectively quantifiable evaluations. While AI-driven approaches have been applied to mandibular analyses for age estimation, existing studies predominantly focus on dental age methodologies. Although machine learning methods based on dental age stages demonstrate high accuracy, in some crime circumstances, the mandible may be found without teeth due to postmortem tooth loss, intentional avulsion, or environmental factors, making forensic identification more challenging.

Thus, in the present study, we investigated the application of mandibular dimensions with machine learning algorithms for age estimation in children and adolescents.

Methods

Study design

This cross-sectional observational study examined orthodontic records from children and adolescents enrolled in the orthodontic treatment at the Bonn University-Germany. This project was conducted in accordance with the Declaration of Helsinki and approved by the Human Ethics Committee (2024-252-BO). Informed consent was obtained from patients and their legal guardians.

Participants

Orthodontic records, including cephalometric radiographs, were screened for this study. Participants included were aged between 6 and 16 years old. Inclusion criteria consisted of children and adolescents without underlying syndromes or congenital alterations. Individuals presenting one or more teeth missing bilaterally due to agenesis or extraction were excluded from the analysis.

Mandibular size assessment

Digital pre-treatment lateral cephalometric radiographs were analyzed to evaluate mandibular dimensions. Radiographs were imported into OnyxCeph software (version 3.2.180; Image Instruments GmbH, Chemnitz, Germany) as lossless TIF files, calibrated, and digitally assessed. Each cephalogram was oriented using the Frankfort horizontal plane and the midsagittal reference line provided by the software, in order to minimize distortions caused by head inclination or rotation. The following anatomical landmarks were identified on each cephalograms (Fig. 1):

- Gnathion (Gn)—The most inferior and anterior point of the mandible at the midline.
- Menton (Me)—The lowest midline point of the chin.
- Pogonion (Pog)—The most anterior midline point of the chin.
- Gonion (Go)—The most posterior and inferior point of the mandibular angle (gonial angle).
- Condylion (Co)—The most superior point on the head of the mandibular condyle.
- Articulare (Ar)—The intersection point between the posterior border of the mandibular ramus and the base of the skull.

Based on the identified landmarks, the following measurements were recorded:

- Linear distances (mm):
 - Mandibular Ramus height: Co-Go Distance from Condylion (Co) to Gonion (Go).
 - Mandibular body length: Go-Gn - Distance from Gonion (Go) to Gnathion (Gn).
 - Total mandibular length: Co-Pog - Distance from Condylion (Co) to Pogonion (Pog).
- Angular measurement (°):
 - Mandibular angle (gonial angle): Ar-Go-Me—Gonial angle formed by the intersection of lines from Articulare (Ar) to Gonion (Go) and Gonion (Go) to Menton (Me). This angle was also considered a cephalometric indicator of vertical skeletal pattern, thereby incorporating potential vertical morphological influences into the modelling process.

To assess potential confounding effects of sagittal skeletal pattern on age estimation, the sample was classified into skeletal Classes I, II, and III based on the ANB angle, and chronological age was compared among these groups using one-way ANOVA.

Sample size and power test

To assess the adequacy of the adopted sample size for each individual predictor, a post-hoc statistical power analysis was performed using G*Power software (version 3.1.9.6; Heinrich Heine University Düsseldorf, Germany; available at: <https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower>). For continuous predictors, the analysis was conducted under Exact - Correlation: Bivariate normal model, entering the observed Pearson correlation coefficient (r) with chronological age, a error probability, and total sample size. For the dichotomous predictor (sex), the analysis was performed under t tests—Means: Difference between two independent means [two groups], using the observed group means, standard deviations, and sample sizes. Given that this step was intended for exploratory variable selection prior to multivariable modelling, a more liberal significance level ($\alpha = 0.20$) was adopted, as recommended by Hosmer,

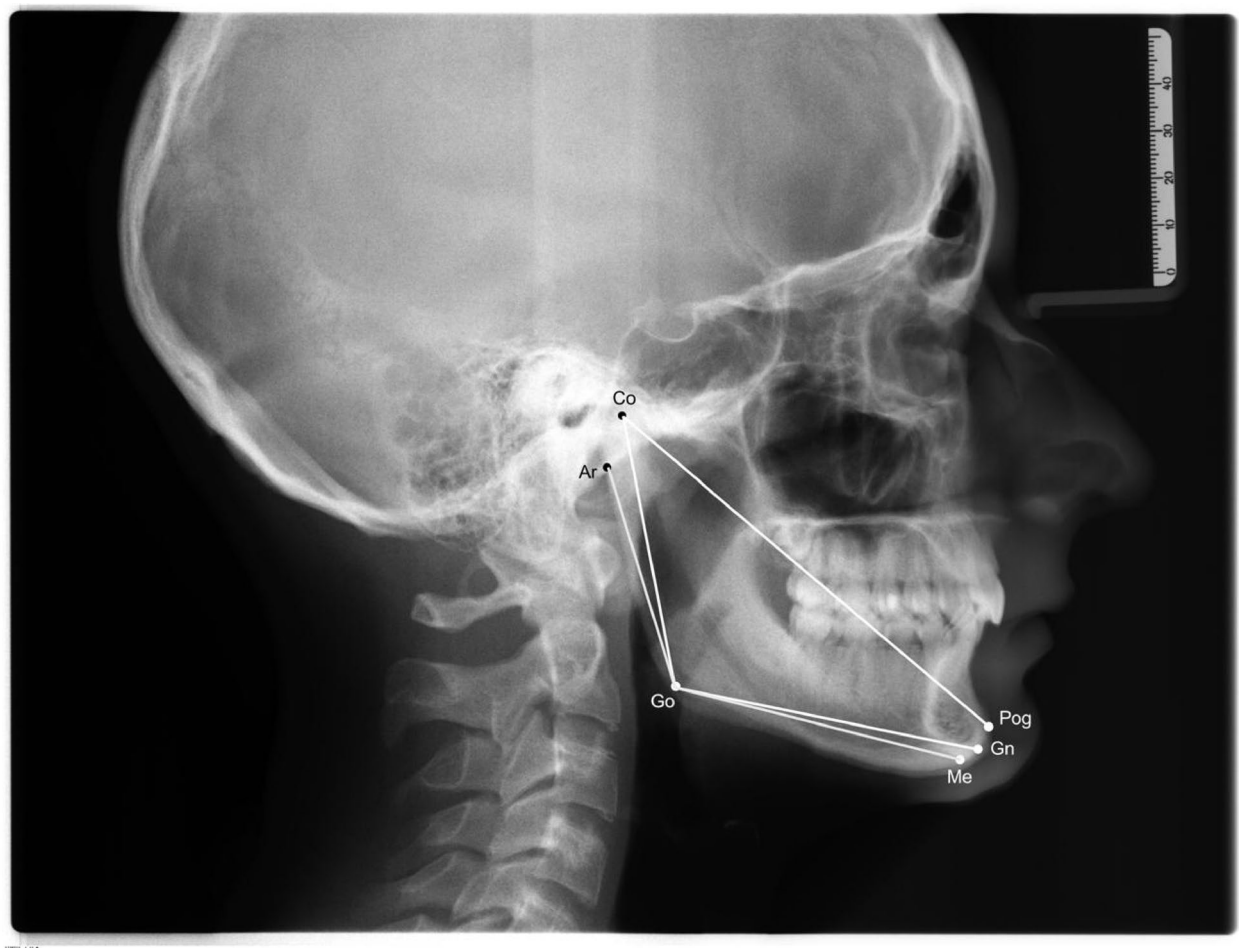


Fig. 1. Lateral cephalometric radiograph showing anatomical landmarks. Linear measurements: Co-Pog, Co-Go, Go-Gn. Angular measurement: Ar-Go-Me. GN = Gnathion, ME = Menton (Me), Pog = Pogonion, Go = Gonion; Co = Condylion Ar = Articulare.

Lemeshow, and Sturdivant (2013)³³, to reduce the likelihood of excluding potentially relevant predictors at the univariate stage.

Feature selection

Predictor variables were selected based on their predictive ability concerning the dependent variable, chronological age. Initially, an exploratory analysis using Pearson correlation test was conducted to examine associations between each independent variable and chronological age. A significance level of 5% ($\alpha=0.05$) was adopted, and variables demonstrating statistically significant correlations ($p<0.05$) were considered relevant and selected for inclusion in predictive models. This approach aimed to identify and retain only those variables that effectively contribute to the explanatory power of the models, thus reducing complexity and enhancing predictive efficiency.

Model development

A predictive model was developed to estimate chronological age based on mandibular size and sex. To ensure a comprehensive evaluation of the relationships present in the dataset—from simple linear associations to more complex, nonlinear patterns—algorithms from four major categories of machine learning were carefully selected: linear models, tree-based models, instance-based methods, and artificial neural networks.

Specifically, eight algorithms were chosen:

- Linear Regression (LR), for its simplicity and interpretability.
- Gradient Boosting Regressor (GB), Random Forest Regressor (RF), Decision Tree Regressor (DT), and Adaboost Regressor (ADA), representing tree-based models known for robustness and strong predictive capabilities.
- Support Vector Regression (SVR) and K-Nearest Neighbors Regressor (KNN), as instance-based methods effective in capturing complex nonlinear relationships.

- Multilayer Perceptron Regressor (MLP), an artificial neural network capable of identifying highly complex data patterns through multiple hidden layers.

This selection aimed to balance model interpretability, predictive performance, and adaptability to diverse data structures, allowing rigorous comparisons among the different methodological approaches^{9–12}.

Training, cross-validation, test and overfitting control

Prior to training the predictive models, the dataset was normalized to enhance numerical stability and improve algorithm learning efficiency. This procedure aims to reduce the influence of varying scales among variables, allowing each feature to contribute equally to model development.

The dataset was randomly split into training (80%) and testing (20%) sets. This approach ensured that model training was conducted on a substantial portion of the data, while an independent subset was reserved to evaluate the model's ability to estimate chronological age in unseen individuals—simulating a real-world application.

To ensure greater generalization capability and minimize the risk of overfitting, stratified five-fold cross-validation was employed. This method involved splitting the data into five subsets, ensuring each subset maintained a representative distribution of the studied variables¹³. Each model was trained and validated repeatedly across these subsets, increasing robustness and reliability of performance metrics obtained.

Hyperparameter optimization

Detailed hyperparameter optimization was performed using the Grid Search method, a systematic and exhaustive search for optimal parameter combinations. Hyperparameters define critical aspects of the algorithms' learning process, directly impacting their capacity to detect patterns within the data¹⁴. Thus, this optimization aimed to identify parameter configurations that maximize the predictive performance of the models.

Model evaluation metrics and feature importance

The accuracy and reliability of predictions made by the developed machine learning models were assessed using the following performance metrics:

- Mean squared error (MSE): Measures the average of the squared differences between predicted and actual chronological ages, giving more weight to larger errors.
- Root mean squared error (RMSE): Calculated as the square root of the MSE, RMSE expresses prediction errors in the same unit (years) as chronological age, making model accuracy easier to interpret.
- Mean absolute error (MAE): Represents the average absolute magnitude of prediction errors, regardless of their direction. This metric provides a straightforward and objective measure of overall predictive accuracy.
- Coefficient of determination (R^2): Indicates the percentage of total variability in chronological age that is explained by each predictive model. Higher values suggest greater explanatory power and more reliable predictions.

To accurately quantify the uncertainty surrounding each performance metric and assess prediction consistency, 95% confidence intervals (CI95%) were calculated using the bootstrap method with 1000 resampling iterations. This approach yielded robust variability estimates and reliable confidence intervals that reflect the models' true performance. For cross-validation results, confidence intervals were derived from the distribution of metrics across different data folds, enabling a comprehensive evaluation of model stability and generalizability.

Visual representations of prediction errors were created using scatter plots, highlighting discrepancies between predicted and actual chronological ages. The prediction error plots were generated by overlaying scatter points on a semi-transparent bivariate kernel density heatmap, graded from navy blue to deep red to indicate the concentration of observations, with higher densities represented by colors closer to red.

Statistical comparisons of MAE values across models were also performed using bootstrap analysis. A statistically significant difference was identified when the computed confidence interval did not include zero, thereby confirming meaningful differences in predictive performance.

To evaluate the relative contribution of each input variable to the model's predictions, feature importance scores were calculated using tools available in the scikit-learn library. These scores provide an estimate of the predictive weight of each variable and are typically derived from models that inherently support this type of analysis. For algorithms that do not natively offer this functionality—specifically K-Nearest Neighbors (KNN), Support Vector Regression (SVR), and Multilayer Perceptron (MLP)—feature importance analysis was not performed^{15–19}.

All statistical analyses and visualizations were performed using the Python programming language in the Google Colaboratory environment. The entire workflow—comprising model construction, training, validation, hyperparameter optimization, and performance evaluation—is illustrated in Fig. 2. To ensure transparency and reproducibility, all scripts developed for the analysis are openly available at <https://doi.org/10.5281/zenodo.15264847>.

Results

Sample characterization

A total of 401 individuals were included, with an age ranging from 6.3 to 16.8 years old, and a mean age of 11.7 ± 2.3 years old; 200 (49.8%) were males and 201 (50.2%) were females (11.4 ± 2.38 years and 11.8 ± 2.37 years respectively). Post-hoc statistical power analysis demonstrated that all predictors assessed in the univariate stage achieved statistical power $\geq 80\%$ for the observed associations, confirming both the adequacy of the sample size and sufficient predictive capacity to support their inclusion in the multivariable modelling. Sagittal skeletal

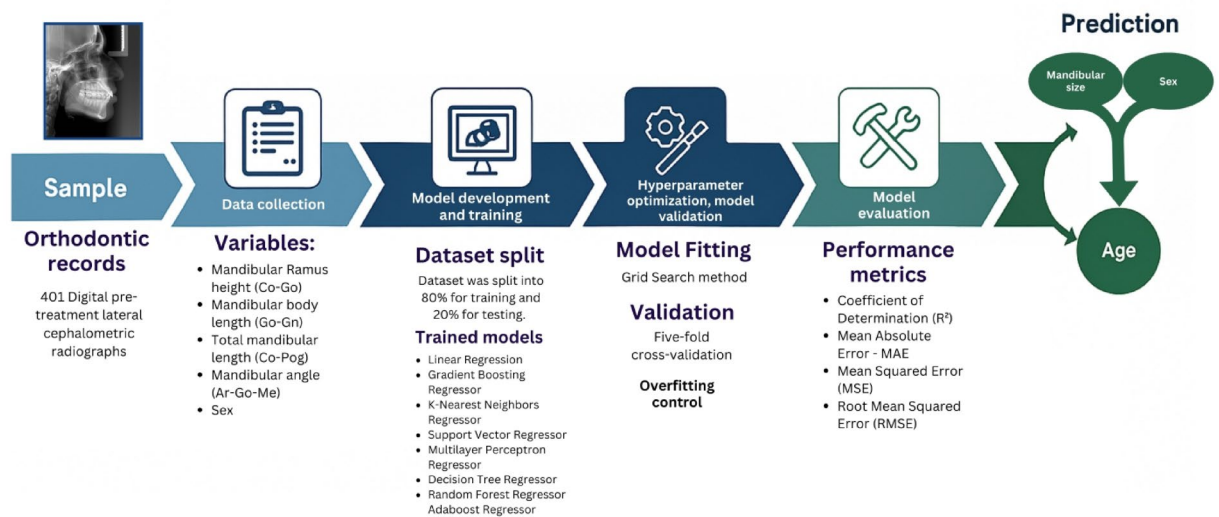


Fig. 2. Flowchart of the main steps of the study.

classification (based on ANB) comprised 177 Class I, 150 Class II, and 74 Class III subjects. A one-way ANOVA comparing chronological age among these groups showed no statistically significant difference ($p=0.488$), indicating a balanced age distribution across sagittal patterns.

Model performance

The predictive models based on mandibular morphometric features and sex demonstrated mean absolute errors (MAE) closer to 1.5 years when compared to actual chronological age. Among the evaluated algorithms, the Gradient Boosting model achieved the lowest MAE on the test set (1.54; 95% CI: 1.33–1.76), followed by the Random Forest, SVM (Support Vector Machine), and KNN (K-Nearest Neighbors) models (Fig. 3). Cross-validation supported this finding, with Gradient Boosting maintaining the lowest MAE (1.21; 95% CI 1.09–1.32) and the highest coefficient of determination ($R^2 = 0.56$; 95% CI 0.46–0.64). However, R^2 values in the test set indicated that only a moderate portion of the variability in chronological age was explained by the included predictors ($R^2 = 0.38$; 95% CI 0.21–0.53) (Table 1).

Pairwise model comparison

Pairwise comparisons of the mean absolute errors revealed distinct patterns of statistical significance among the evaluated models (Fig. 4). The decision tree model showed the poorest performance, with statistically significant differences when compared to the Gradient Boosting and Random Forest models. In contrast, the Gradient Boosting model achieved the best performance, with statistically significant differences compared to the Logistic Regression, AdaBoost, and Decision Tree models.

Feature importance

Among the linear distances evaluated, total mandibular length (Co-Pog) demonstrated the highest predictive importance, followed by mandibular ramus height (Co-Go). The feature importance ranking is presented in Fig. 5.

Discussion

The mandible plays a crucial role in forensic investigations, as its unique anatomical features, dental records, and bone structure can help identify individuals, determine age, and provide insights into cause of death. In age estimation, the mandible can play a crucial role in children and adolescents, not only due to its predictable growth patterns but also because of the sequential development of teeth. The current literature has repeatedly shown the importance of mandibular teeth to estimate age through dental age methods that assess the mineralization stages of developing permanent teeth using radiographic analysis. However, in certain forensic contexts, teeth may be absent from the mandible—whether due to ante-mortem loss, post-mortem damage, or developmental stage—limiting the applicability of traditional dental methods. In such cases, evaluating the use of mandibular dimensions for age estimation in growing individuals becomes particularly relevant. By exploring mandibular measurements and validating their accuracy across diverse populations, researchers and clinicians can enhance forensic identification methods and contribute to more precise age estimation in both living individuals and skeletal remains. Therefore, this study explores the application of mandibular morphometry combined with machine learning algorithms for age estimation.

In recent years, the application of AI in forensic medicine and odontology has grown substantially. A recent systematic review conducted by Singh et al. (2024) compiled studies that employed AI tools for age and

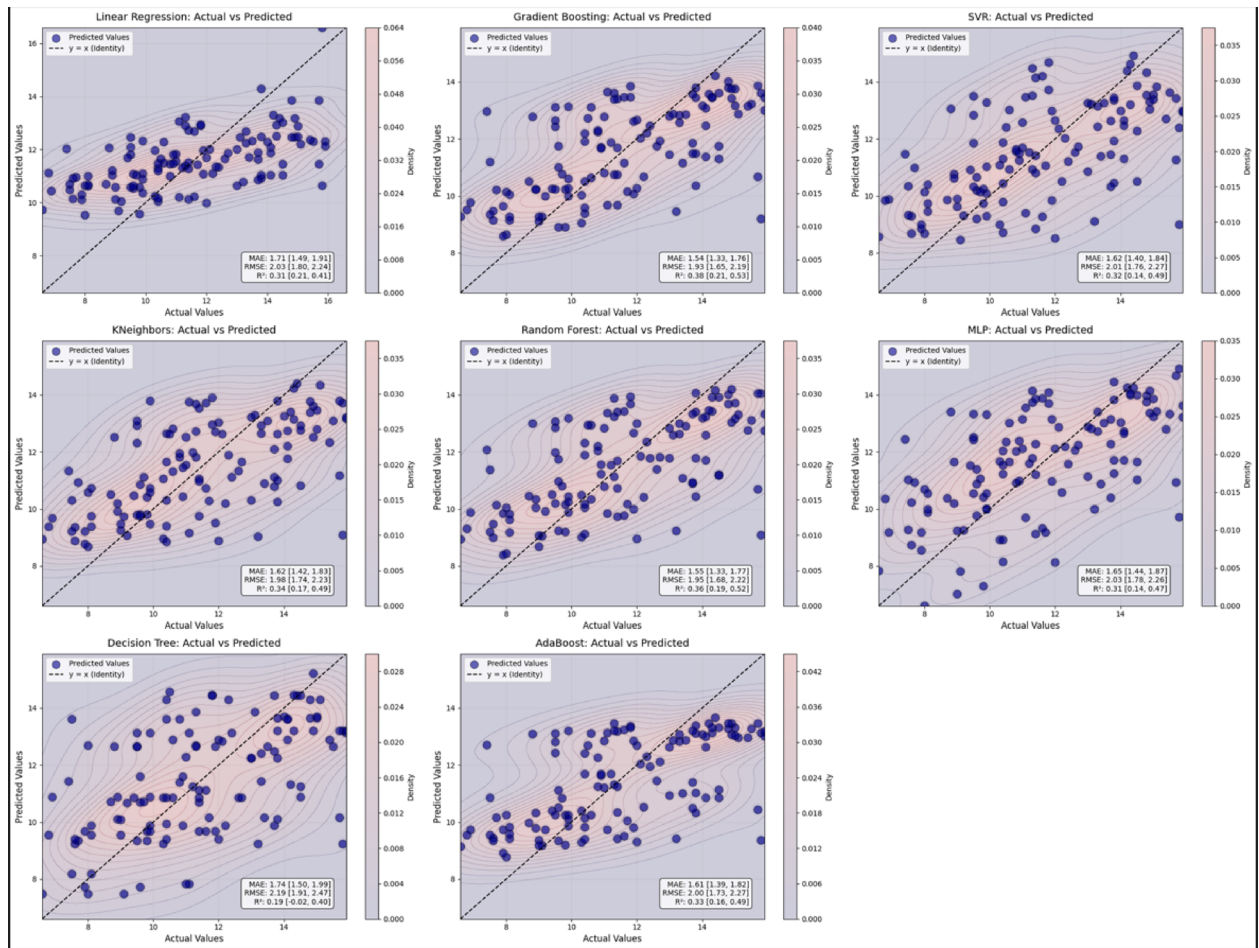


Fig. 3. Scatter plots: Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN).

sex estimation based on maxillofacial radiographs, including panoramic and lateral cephalometric images²⁰. Although the review highlighted promising results, most of the analysed studies focused predominantly on sex determination or broad age group classification, with few studies aiming to precisely estimate chronological age. Furthermore, the majority of the reviewed studies relied on panoramic radiographs, which, despite providing a broad two-dimensional view of the maxillofacial region, present limitations for detailed morphometric analysis when compared to lateral cephalometric radiographs. Another important distinction concerns the age range of the studied populations: whereas previous studies primarily included teenagers and adults, our study focused exclusively on children and adolescents aged between 6 and 16 years, a developmental period characterized by intense skeletal changes. Additionally, while previous studies often incorporated dental parameters—such as the number of teeth or the presence of implants—we based our analysis solely on mandibular morphometric measurements, seeking greater applicability in forensic contexts where dental structures may be absent. It is also noteworthy that the predictive models developed in our study achieved a mean absolute error close to 1.5 years (18 months), which was lower than the 21 months reported by Back et al. using deep learning approaches²¹. Therefore, the use of mandibular morphometric measurements demonstrated the ability to produce more accurate age estimates in children and teenagers, expanding the applicability of forensic identification in challenging scenarios.

In forensic practice, a prediction error of approximately 1.5 years is often deemed clinically acceptable, particularly when age estimation relies on the assessment of a single skeletal indicator. The results obtained in this study are within this range, supporting their potential applicability in real-world forensic scenarios. Liversidge and Marsden (2010), in a study using third molars, reported mean absolute errors ranging from 1.45 to 1.97 years across six evaluated methods²². Previous studies have reported errors of 1.3 and 1.5 years for males and females, respectively, and 1.13 years using a polynomial and Bayesian approach^{23,24}.

The dimensions of the mandible, particularly the height of the mandibular ramus and the gonial angle, have been extensively evaluated in the context of age estimation⁷. Specifically, the height of the mandibular ramus has been previously pointed as having a strong correlation with chronological age^{25,26}. In our study, total mandibular length (Co-Pog) demonstrated the highest predictive importance, followed by the height of the mandibular ramus (Co-Go). This finding is particularly relevant, as the mandibular ramus typically remains intact even in extensively damaged skeletal remains, reinforcing its forensic applicability²⁷. Furthermore, it is well

Model	Optimal hyperparameters	Test data results [CI95%]	Cross-validation results [CI95%]
Gradient boosting	learning_rate: 0.1	MSE = 3.74 [2.73–4.81]	MSE = 2.37 [1.96–2.84]
	max_depth: 4	RMSE = 1.93 [1.65–2.19]	RMSE = 1.54 [1.40–1.69]
	min_samples_split: 2	$R^2 = 0.38$ [0.21–0.53]	$R^2 = 0.56$ [0.46–0.64]
	n_estimators: 100	MAE = 1.54 [1.33–1.76]	MAE = 1.21 [1.09–1.32]
Linear regression	fit_intercept: True	MSE = 4.12 [3.24–5.02]	MSE = 4.28 [2.64–6.07]
	copy_X: True	RMSE = 2.03 [1.80–2.24]	RMSE = 2.06 [1.62–2.46]
	n_jobs: -1	$R^2 = 0.31$ [0.21–0.41]	$R^2 = 0.20$ [-0.13–0.51]
	positive: True	MAE = 1.71 [1.49–1.91]	MAE = 1.56 [1.31–1.78]
Support vector machine	C: 1	MSE = 4.06 [3.09–5.16]	MSE = 2.84 [2.38–3.34]
	kernel: rbf	RMSE = 2.01 [1.76–2.27]	RMSE = 1.68 [1.54–1.83]
	degree: 2	$R^2 = 0.32$ [0.14–0.49]	$R^2 = 0.47$ [0.37–0.55]
		MAE = 1.62 [1.40–1.83]	MAE = 1.33 [1.20–1.47]
K-nearest neighbors	n_neighbors: 9	MSE = 3.94 [3.02–4.98]	MSE = 2.97 [2.45–3.55]
	p: 1	RMSE = 1.98 [1.74–2.23]	RMSE = 1.72 [1.57–1.88]
	weights: uniform	$R^2 = 0.34$ [0.17–0.49]	$R^2 = 0.45$ [0.34–0.54]
		MAE = 1.62 [1.42–1.83]	MAE = 1.37 [1.23–1.51]
Random forest	max_depth: None	MSE = 3.81 [2.81–4.92]	MSE = 2.29 [1.91–2.71]
	max_features: sqrt	RMSE = 1.95 [1.68–2.22]	RMSE = 1.51 [1.38–1.65]
	min_samples_leaf: 4	$R^2 = 0.36$ [0.19–0.52]	$R^2 = 0.57$ [0.48–0.65]
	min_samples_split: 10	MAE = 1.55 [1.33–1.77]	MAE = 1.18 [1.06–1.29]
	n_estimators: 200		
AdaBoost	learning_rate: 0.1	MSE = 4.01 [3.00–5.14]	MSE = 2.58 [2.17–3.02]
	loss: linear	RMSE = 2.00 [1.73–2.27]	RMSE = 1.60 [1.47–1.74]
	n_estimators: 100	$R^2 = 0.33$ [0.16–0.49]	$R^2 = 0.52$ [0.42–0.60]
		MAE = 1.61 [1.39–1.82]	MAE = 1.30 [1.19–1.42]
Decision tree	max_depth: 10	MSE = 4.83 [3.65–6.12]	MSE = 2.94 [2.31–3.65]
	max_features: sqrt	RMSE = 2.19 [1.91–2.47]	RMSE = 1.71 [1.52–1.91]
	min_samples_leaf: 4	$R^2 = 0.19$ [-0.02–0.40]	$R^2 = 0.45$ [0.31–0.57]
	min_samples_split: 10	MAE = 1.74 [1.50–1.99]	MAE = 1.29 [1.14–1.45]
MLP regressor	activation: relu	MSE = 4.11 [3.18–5.12]	MSE = 3.02 [2.52–3.65]
	alpha: 0.001	RMSE = 2.02 [1.78–2.26]	RMSE = 1.73 [1.59–1.91]
	hidden_layer_sizes: (100,)	$R^2 = 0.31$ [0.14–0.47]	$R^2 = 0.44$ [0.31–0.53]
	learning_rate: constant	MAE = 1.65 [1.44–1.87]	MAE = 1.40 [1.27–1.53]
	solver: adam		

Table 1. Summary of the metrics obtained during the cross-validation and testing phases of the regression models used to predict chronological age based on mandibular size and sex, along with their optimal hyperparameters.

established that the gonial angle undergoes morphological changes throughout life: it is initially obtuse during early developmental stages and gradually becomes more acute as the individual matures²⁸. This transformation reflects the dynamics of skeletal growth, as the mandibular ramus tends to increase more in height than the mandibular body in length during growth, resulting in a reduction of the gonial angle²⁹. These biological patterns highlight the potential of mandibular measurements for chronological age estimation, particularly in forensic cases where teeth may be absent or damaged. Thus, the investigation of mandibular morphometry combined with machine learning algorithms may represent a promising alternative for forensic age estimation in children and adolescents.

The application of AI for age prediction based on skeletal features offers significant advantages. AI-based models have demonstrated superior predictive performance and optimize analysis time compared to manual assessments by trained specialists. Integrating these tools into the assessment of bone structures can enhance accuracy, and support clinical decision-making, particularly in contexts with limited availability of specialized professionals³⁰. Given that the mandible often remains intact even in severely compromised bodies³¹, mandibular dimensions analysed using machine learning algorithms can contribute to more precise age estimations in children and adolescents. In forensic practice, specific mandibular measurements can be directly extracted from radiographic images and applied to previously trained models, enabling rapid and standardized chronological age estimates. This approach enhances the speed and consistency of forensic examinations, especially in complex scenarios such as mass disasters or cases involving advanced skeletonization, thereby reducing the subjectivity associated with traditional assessments.

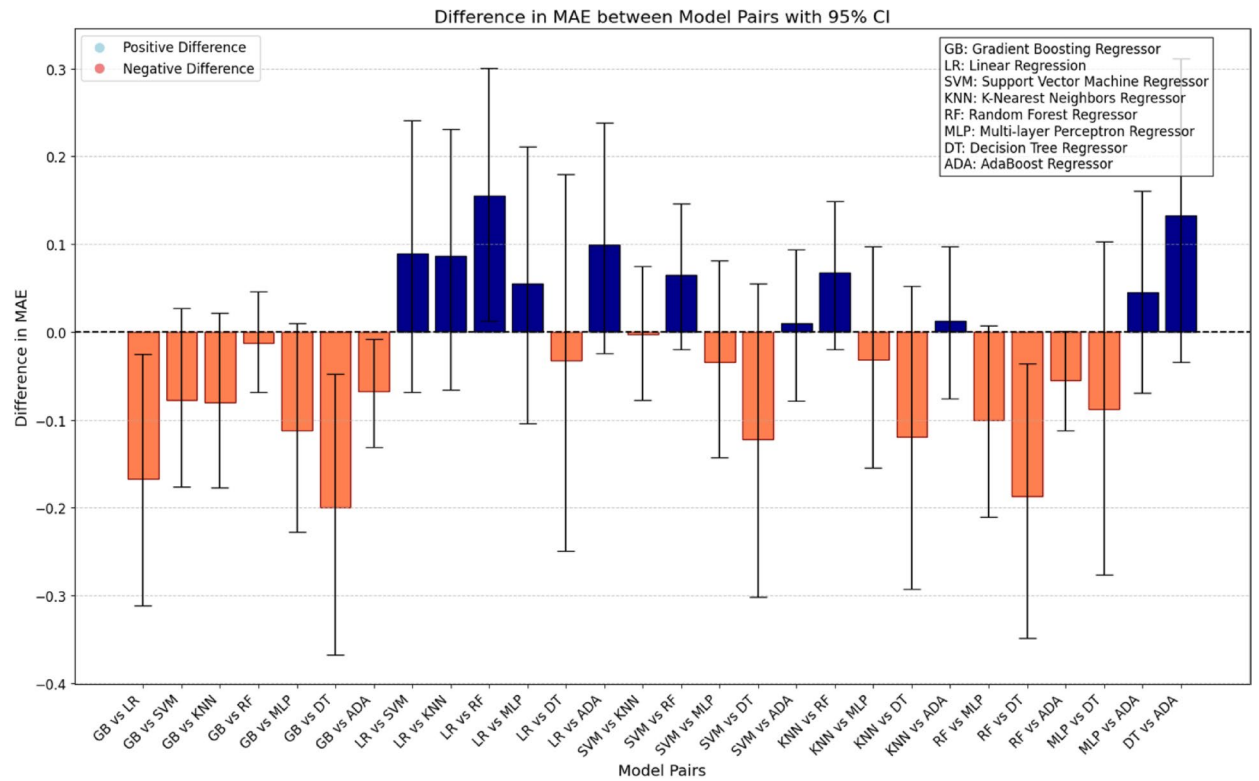


Fig. 4. Difference in mean absolute error (MAE) between model pairs with 95% confidence intervals.

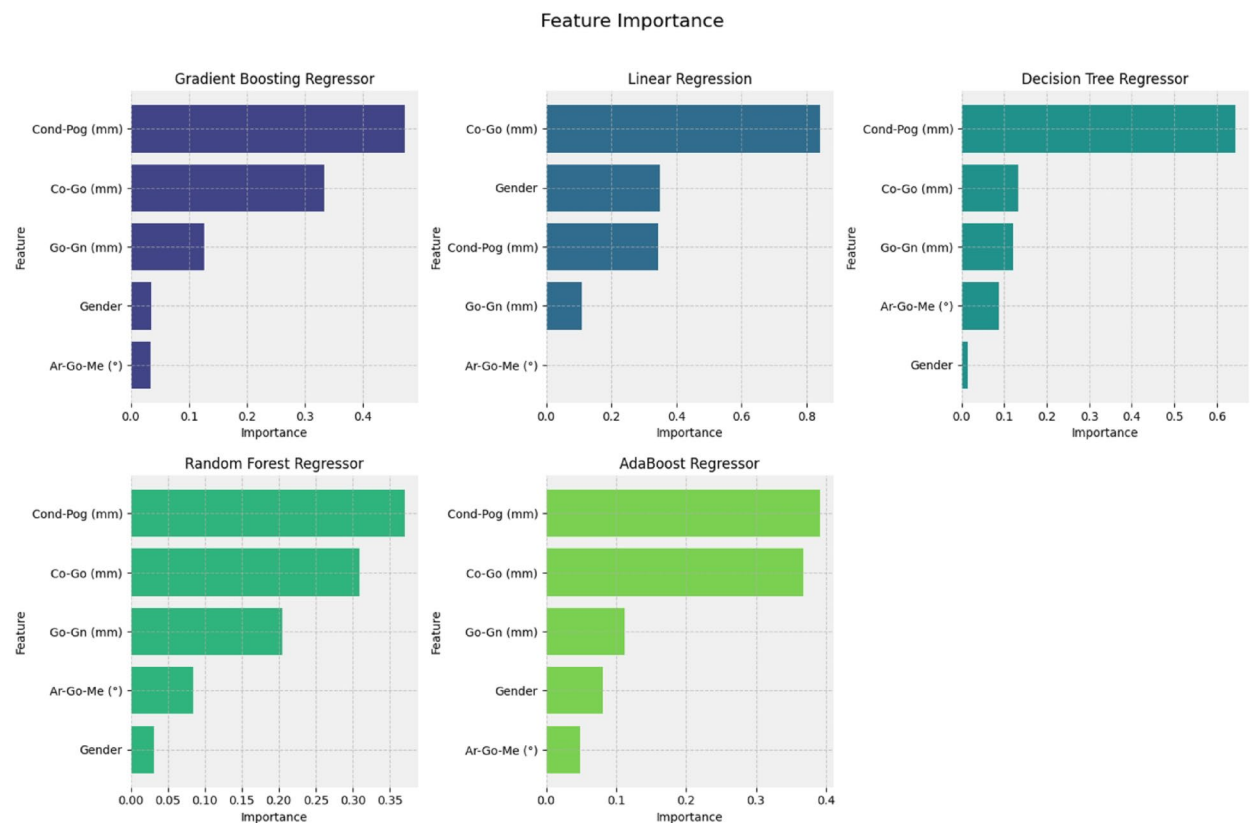


Fig. 5. Feature importance ranking across different machine learning models.

Finally, it is important to acknowledge some limitations of the present study. First, the sample consisted exclusively of German individuals with different skeletal malocclusions, which may limit the generalizability of the results to populations of different ethnic or geographic backgrounds. Additionally, one of the predictors used in the models was the sex of the individuals, which was included due to the distinct growth patterns observed between males and females. However, for practical application in real forensic contexts, the sex must be previously known or determined. Although sex can be reliably identified through methods such as DNA analysis, this additional step involves greater operational complexity and processing time, particularly in emergency forensic scenarios or situations with limited resources. Therefore, it is recommended that future studies validate these models in different populations and also explore alternative strategies for using or predicting the sex variable, aiming to broaden their applicability in forensic practice.

Conclusions

The findings of the present study demonstrated that mandibular dimensions analysed by machine learning algorithms enable precise age estimation in children and adolescents, presenting mean absolute errors close to 1.5 years. Among the evaluated models, Gradient Boosting achieved the best predictive performance, demonstrating robustness and reliability for forensic applications. Additionally, total mandibular length (Co-Pog) and mandibular ramus height (Co-Go) were identified as the most significant predictors, reinforcing the forensic relevance of these anatomical measures. Nevertheless, the generalizability and practical applicability of the proposed predictive models may be influenced by population-specific characteristics and the availability of certain predictor variables in forensic contexts. Therefore, further studies are encouraged to validate these findings in diverse populations and explore alternative predictors to enhance applicability in various forensic scenarios.

Data availability

To ensure transparency and reproducibility, all scripts developed for the analysis are openly available at <https://doi.org/10.5281/zenodo.15264847>.

Received: 2 May 2025; Accepted: 19 September 2025

Published online: 07 October 2025

References

- Kurniawan, A. et al. The applicable dental age estimation methods for children and adolescents in Indonesia. *Int. J. Dent.* <https://doi.org/10.1155/2022/6761476> (2022).
- Lossois, M. & Baccino, E. Forensic age Estimation in migrants: where do we stand? *WIREs. Forensic Sci.* **3**, e1408. <https://doi.org/10.1002/wfs2.1408> (2021).
- Manjrekar, S., Deshpande, S., Katge, F., Jain, R. & Ghorpade, T. Age estimation in children by the measurement of open apices in teeth: A study in the Western Indian population. *Int. J. Dent.* **2022**, 9513501. <https://doi.org/10.1155/2022/9513501> (2022).
- Patil, V. et al. Age assessment through root lengths of mandibular second and third permanent molars using machine learning and artificial neural networks. *J. Imaging.* **9** (2), 33. <https://doi.org/10.3390/jimaging9020033> (2023).
- Ramlal, G., Vevaraju, D., Vemula, A. Y., Swapna, T. & Bindu, P. H. Extrication of DNA from burnt teeth exposed to environment. *J. Clin. Diagn. Res. JCDR.* **11** (8), ZC120–ZC122. <https://doi.org/10.7860/JCDR/2017/26911.10525> (2017).
- Zulkifli, N. A. F. et al. Age estimation from mandibles in malay: A 2D geometric morphometric analysis. *J. Taibah Univ. Med. Sci.* **18** (6), 1435–1445. <https://doi.org/10.1016/j.jtumed.2023.05.020> (2023).
- Leversha, J. et al. Age and gender correlation of gonial angle, Ramus height and bigonial width in dentate subjects in a dental school in Far North Queensland. *J. Clin. Exp. Dent.* **8** (1), e49–e54. <https://doi.org/10.4317/jced.52683> (2016).
- Dhillon, S. K., Ganggayah, M. D., Sinnadurai, S., Lio, P. & Taib, N. A. Theory and practice of integrating machine learning and conventional statistics in medical data analysis. *Diagnostics (Basel Switzerland).* **12** (10), 2526. <https://doi.org/10.3390/diagnostics12102526> (2022).
- Natekin, A. & Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurorobotics.* **7**, 21. <https://doi.org/10.3389/fnbot.2013.00021> (2013).
- Halder, R. K. et al. Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *J. Big Data.* **11**, 113. <https://doi.org/10.1186/s40537-024-00973-y> (2024).
- Bikku, T. Multi-layered deep learning perceptron approach for health risk prediction. *J. Big Data.* **7**, 50. <https://doi.org/10.1186/s40537-020-00316-7> (2020).
- Favoreto, M. W. et al. Prediction based on machine learning of tooth sensitivity for in-office dental bleaching. *J. Dent.* **153**, 105517. <https://doi.org/10.1016/j.jdent.2024.105517> (2025).
- Refaeilzadeh, P., Tang, L. & Liu, H. Cross-validation. In *Encyclopedia of Database Systems*. (eds. Liu, L. & Özsu, M. T.) 532–538 (Springer US, 2009).
- Jiang, X. & Xu, C. Deep learning and machine learning with grid search to predict later occurrence of breast cancer metastasis using clinical data. *J. Clin. Med.* **11** (19), 5772. <https://doi.org/10.3390/jcm11195772> (2022).
- Basso, I. B. et al. Sex prediction through machine learning utilizing mandibular condyles, coronoid processes, and sigmoid notches features. *PLoS One.* **19** (11), e0312824. <https://doi.org/10.1371/journal.pone.0312824> (2024).
- de Araujo, C. M. et al. Sex determination through maxillary dental arch and skeletal base measurements using machine learning. *Head Face Med.* **20** (1), 44. <https://doi.org/10.1186/s13005-024-00446-w> (2024).
- de Araujo, C. M. et al. Predicting the risk of maxillary canine impaction based on maxillary measurements using supervised machine learning. *Orthod. Craniofac. Res.* **28** (1), 207–215. <https://doi.org/10.1111/ocr.12863> (2025).
- Ferraz, A. X. et al. Artificial intelligence model for predicting sexual dimorphism through the hyoid bone in adult patients. *PLoS One.* **19** (11), e0310811. <https://doi.org/10.1371/journal.pone.0310811> (2024).
- Küchler, E. C. et al. Mandibular and dental measurements for sex determination using machine learning. *Sci. Rep.* **14** (1), 9587. <https://doi.org/10.1038/s41598-024-59556-9> (2024).
- Singh, S., Singha, B. & Kumar, S. Artificial intelligence in age and sex determination using maxillofacial radiographs: A systematic review. *J. Forensic Odonto-Stomatol.* **42** (1), 30–37. <https://doi.org/10.5281/zenodo.11088513> (2024).
- De Back, W. et al. Forensic age estimation with Bayesian convolutional neural networks based on panoramic dental X-ray imaging. In *2019 International Conference on Medical Imaging with Deep Learning (MIDL)* (2019).

22. Liversidge, H. M. & Marsden, P. H. Estimating age and the likelihood of having attained 18 years of age using mandibular third molars. *Br. Dent. J.* **209** (8), E13. <https://doi.org/10.1038/sj.bdj.2010.976> (2010).
23. Solari, A. C. & Abramovitch, K. The accuracy and precision of third molar development as an indicator of chronological age in Hispanics. *J. Forensic Sci.* **47** (3), 531–535 (2002).
24. Thevissen, P. W., Fieuws, S. & Willems, G. Human dental age estimation using third molar developmental stages: does a bayesian approach outperform regression models to discriminate between juveniles and adults? *Int. J. Legal Med.* **124** (1), 35–42. <https://doi.org/10.1007/s00414-009-0329-8> (2010).
25. Franklin, D. & Cardini, A. Mandibular morphology as an indicator of human subadult age: interlandmark approaches. *J. Forensic Sci.* **52** (5), 1015–1019. <https://doi.org/10.1111/j.1556-4029.2007.00522.x> (2007).
26. Norris, S. P. Mandibular Ramus height as an indicator of human infant age. *J. Forensic Sci.* **47** (1), 8–11 (2002).
27. Bhuyan, R. et al. Panoramic radiograph as a forensic aid in age and gender estimation: preliminary retrospective study. *J. Oral Maxillofac. Pathol. JOMFP.* **22** (2), 266–270. https://doi.org/10.4103/jomfp.JOMFP_90_17 (2018).
28. Jensen, E. & Palling, M. The gonial angle: A survey. *Am. J. Orthod.* **40** (2), 120–133 (1954).
29. Al-Shamout, R., Alrbata, R., Ammouh, M. & Alhababbah, A. M. Age and gender differences in gonial angle, Ramus height and bigonial width in dentate subjects. *Pak. Oral Dent. J.* **22**(1). (2012).
30. Nguyen, T. et al. High performance for bone age estimation with an artificial intelligence solution. *Diagn. Interv. Imaging.* **104** (7–8), 330–336. <https://doi.org/10.1016/j.diii.2023.04.003> (2023).
31. Krishan, K., Kanchan, T. & Garg, A. K. Dental evidence in forensic Identification - An Overview, methodology and present status. *Open. Dentistry J.* **9**, 250–256. <https://doi.org/10.2174/1874210601509010250> (2015).
32. Faul, F., Erdfelder, E., Buchner, A. & Lang, A. G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods.* **41**(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149> (2009).
33. Hosmer, D. W. Jr., Lemeshow, S. & Sturdivant, R. X. Model-building strategies and methods for logistic regression. In *Applied Logistic Regression* (eds Hosmer, D. W., Lemeshow, S. & Sturdivant, R. X.) (2013). <https://doi.org/10.1002/9781118548387.ch4>

Acknowledgements

Research and Innovation Support Foundation of Santa Catarina State (FAPESC) (Doctoral funding, 19/2024) (AA). National Council for Scientific and Technological Development (CNPq) (Postdoctoral funding, 102385/2024-6) (LAAA). Brazilian Coordination of Higher Education, Ministry of Education (CAPES) (finance code 001).

Author contributions

E.C.K.—conceptualization, supervision, formal analysis, writing-original draft, visualization. P.P.K.—methodology, formal analysis, visualization, writing-original draft. E.G.C.E.—investigation, data curation. L.A.A.A.—resources, investigation, data curation. A.A.—investigation, data curation, visualization. B.M.M.A.—investigation, methodology, writing - original draft. F.B.F.—methodology, validation, project administration. S.B.M.—methodology, formal analysis, resource. C.K.—resources, supervision, funding acquisition, writing-original draft. C.M.A.—conceptualization, supervision, formal analysis, writing-original draft. All authors reviewed the manuscript (writing-review and editing).

Funding

Open Access funding enabled and organized by Projekt DEAL. Research and Innovation Support Foundation of Santa Catarina State (FAPESC) (Doctoral funding, 19/2024) (AA). National Council for Scientific and Technological Development (CNPq) (Postdoctoral funding, 102385/2024-6) (LAAA). Brazilian Coordination of Higher Education, Ministry of Education (CAPES) (finance code 001).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to E.C.K.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025