



Development of a multi-stage model for intelligent and quantitative appraising of skeletal maturity using cervical vertebrae cone-beam CT images of Chinese girls

Lizhe Xie^{1,2} · Wen Tang^{1,3} · Iman Izadikhah^{1,3} · Zhenqi Zhao⁴ · Yang Zhao⁵ · Hu Li^{2,3} · Bin Yan^{1,2,3} 

Received: 7 August 2021 / Accepted: 17 December 2021
© CARS 2022

Abstract

Purpose Nowadays, the integration of Artificial intelligence algorithms and quantified radiographic imaging-based diagnostic procedures is hailing amplified deliberation particularly in assessment of skeletal maturity. So we intend to formulate a logistic regression model for intelligent and quantitative estimation of Fishman skeletal maturation index (SMI) based on the parameters attained from the cervical vertebrae CBCT images of Chinese girls.

Methods From 709 hand wrist radiographs and CBCT images, 447 samples were randomly selected (called as G1) to build a logistic regression model. The reliability and reproducibility were assessed by the intraclass correlation coefficient (ICC) and weighted Cohen's kappa, followed by Spearman's rank correlation coefficient to identify the parameters significantly associated with the SMI. Two hundred and sixty-two other subjects (named G2) were recruited for external examination of the models by direct visual comparison and the receiver operating characteristic (ROC) curve. In cases of confusion and mispredictions, the model was modified to improve the consistency.

Results Five significant parameters (Chronological age, C3 height (H3)), C4 upper width (UW4), C4 lower width (LW4), and the ratio of posterior height to lower width of C4 (PH4/LW4)) were administered into logistic regression model. Despite total agreement percentage which was 84% (total AUC = 0.92), unsatisfactory performance was noticed for the 6th and 8th stages which were confused with their neighboring stages. After adjustments of the models, the total agreement percentage and AUC were upgraded to 88% and 0.96, respectively.

Conclusion Consistency and fitness evaluation of our models demonstrated adequate prediction percentage and reliability for automated classification of skeletal maturation. The presented constructed logistic regression model has the potential to serve as a maturity evaluation index in clinical craniofacial orthopedics in Chinese girls. The proposed model in this study showed promising strength for being expended in the event of other clinical multi-stage conditions.

Keyword Skeletal maturity · Hand wrist · Cervical vertebra · Cone-beam CT · Artificial intelligent

Lizhe Xie and Wen Tang contributed equally as the first author.

✉ Bin Yan
byan@njmu.edu.cn

¹ Jiangsu Province Key Laboratory of Oral Diseases, Nanjing Medical University, Nanjing, China

² Jiangsu Province Engineering Research Center of Stomatological Translational Medicine, Nanjing Medical University, 136 Hanzhong Street, Gulou District, Nanjing 210029, China

³ Department of Orthodontics, Affiliated Hospital of Stomatology, Nanjing Medical University, Gulou District, 136 Hanzhong Street, Nanjing 210029, China

Introduction

Flawless treatment planning is highly reliant on perception and rating of skeletal maturation rather than chronologic age in growing patients [1, 2]. Accelerated developmental periods apparent in puberty and mandibular growth peaks are critical episodes in craniofacial surgeries and dentofacial orthopedics [3, 4].

⁴ Department of Stomatology, The First People's Hospital of Nantong, Nantong, China

⁵ School of Public Health, Department of Biostatistics, Nanjing Medical University, Jiangning District, Nanjing, China

Amid the biologic indices for skeletal maturity assessment, those that consider wrist X-ray are the most common indicators [5–8]. After introduction of Cone-beam CT (CBCT) in orthodontics, scholars have tried to institute hand–wrist-based prediction models using morphological parameters of cervical vertebrae to evaluate the growth and development status [9, 10]. Previously, measurement of the latter morphological parameters was performed mostly on lateral cephalograms [11–15]. However, as a two-dimensional modality, cephalometric radiographs are inevitably subjected to image superimposition of hard or soft tissues causing them to produce a magnified and distorted projection which is unstable [16]. Thus far, utilizing CBCT is rational for observance of morphologic changes in the cervical vertebrae to assess skeletal maturity [16, 17]. Nevertheless, CBCT enacts an essential role in considerable circumstances such as ectopic teeth, severe maxillofacial malformations, and dental implant surgeries [18]. Considering the great potential of CBCT for being used as the primary diagnostic tool in future [19], eradication of extra radiation doses from wrist X-rays can be instigated initially in the aforementioned circumstances for today's clinic. Despite the remarkable improvements and technological advancements regarding CBCT safety, certain ambiguity and controversy among scholars exists concerning CBCT radiation dose, direct patient benefit and presence of streaking and motion artifacts [20].

As a subfield of Artificial Intelligence (AI), Machine Learning (ML) contrives algorithms in which computers (the machines) adapt and learn from the intrinsic statistical patterns and structures in data to perform vast bioanalytic functions particularly in diagnostic medicine [21, 22]. Previously, regression-based learning methods were concluded as the most frequently used method in the skeletal maturation analyses [23]. Logistic regression is a classification algorithm, used when the value of the target variable is categorical. This method is most commonly used when the data in question have a multi-class output and decide it belongs to one class or another by sorting the data according to the characteristics of a study sample [24, 25]. Although previous studies have taken the advantage of CBCT, yet they relied on multiple linear regression analysis [26, 27] which may not be fully compatible with the wrist bone maturity, since it is clinically divided into discrete or ranked stages. The linear regression has essentially a straightforward functionality on continuous changes which impedes its best performance concerning ordinal predictive models [28]. Contrarily, the logistic regression algorithms not only convey the characteristics of its linear counterpart but also can more accurately and objectively predict and pinpoint the exact maturational stage of wrist bones eliminating concerns about outcome consistency. To date, we have not come across any research

considering this fact and clinical studies suffer from this drawback.

Therefore, the objective of this study is to formulate a clinically applicable model for intelligent and quantitative estimation of hand–wrist maturation based on the parameters derived from the sagittal reconstructed slices of the cervical vertebrae using CBCT images from Chinese girls. We also believe that our proposed method can offer an innovative scheme for other clinical multi-stage problems.

Methods

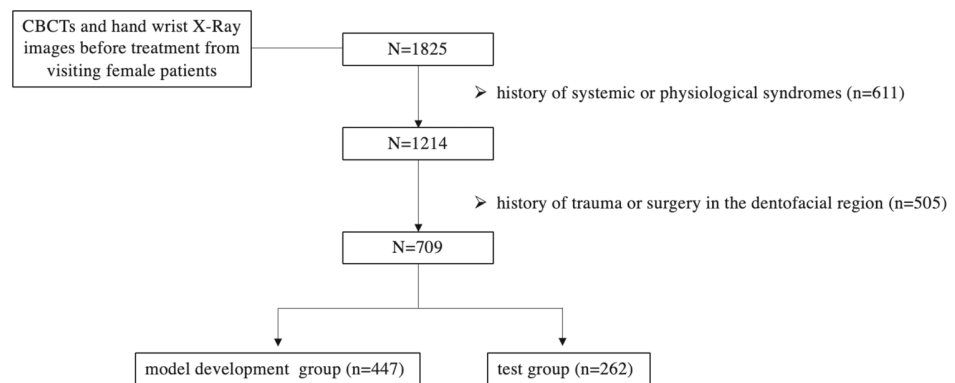
Study population

This retrospective study was granted ethical approval from the institutional review board of Nanjing medical university. A total of 709 sets of existing pretreatment CBCTs and hand wrist X-Ray images from visiting female patients in an age range of 7–19 years (mean age, 11.84 years) was recruited. CBCT data were acquired on the same date as the hand–wrist X-ray. The scans were adopted from the Orthodontic Department of the same institution that provided the approval (registered between January and December 2018) and were taken for various clinical reasons imperative to the patient diagnosis and treatment planning. The samples (1) without history of systemic or physiological syndromes; (2) without history of trauma or surgery in the dentofacial region; (3) having reliable CBCT scans were included (Fig. 1). The images were randomly apportioned into two separate groups, one for developing the models (G1, $n = 447$) and the other serving for examination of the models functioning (G2, $n = 262$) (Table 1).

All the CBCT images were taken by a NewTom 5G computed tomography system (Quantitative Radiology, Verona, Italy) with the exposure parameters of 18×16 cm field of view; 110 kV tube voltage; 1–20 mA (pulsed mode); and 0.3 mm isotropic voxel size. While ensuring maximum intercuspation during imaging, the patients' midsagittal plane was set perpendicular to the horizontal plane, and their Frankfort plane was paralleled with the ground. The skeletal maturation dataset included hand–wrist radiograms taken with a Rotanode DRX-3724HD (Orthopantomograph OP200, Tuusula, Finland) using the exposure parameters of 77 kV, 10 mA, and 8.0 s of exposure duration.

Skeletal maturation assessment

The skeletal maturity of all subjects was assessed using the hand wrist X-ray. The Fishman method was applied to discern subjects in 11 maturational stages of skeletal maturation index (SMI) consistent with the ossified appearance of the digits and wrist bones (Fig. 2) [29]. Three well-trained

Fig. 1 A schematic diagram presenting the selection process of the subjects**Table 1** Sample age (years) characteristics along with the distribution pattern between model development (G1) and test group (G2)

SMI stage	Group (size)	Age (mean, SD)	Age range	<i>P</i> value
1	G1 (<i>n</i> = 25)	8.56 ± 1.42	7–12	0.33
	G2 (<i>n</i> = 18)	8.39 ± 1.14	7–10	
2	G1 (<i>n</i> = 35)	10.29 ± 0.83	8–12	0.36
	G2 (<i>n</i> = 25)	10.44 ± 0.65	9–12	
3	G1 (<i>n</i> = 37)	10.73 ± 0.96	9–13	0.93
	G2 (<i>n</i> = 26)	10.73 ± 1.00	9–13	
4	G1 (<i>n</i> = 35)	11.31 ± 1.05	9–13	0.32
	G2 (<i>n</i> = 22)	11.14 ± 0.99	9–13	
5	G1 (<i>n</i> = 45)	11.56 ± 1.10	9–15	0.99
	G2 (<i>n</i> = 25)	11.40 ± 1.04	9–14	
6	G1 (<i>n</i> = 43)	11.79 ± 1.01	9–15	0.33
	G2 (<i>n</i> = 24)	11.29 ± 0.81	9–13	
7	G1 (<i>n</i> = 46)	12.22 ± 0.92	10–14	0.56
	G2 (<i>n</i> = 28)	12.04 ± 0.92	10–14	
8	G1 (<i>n</i> = 36)	12.22 ± 1.07	10–15	0.82
	G2 (<i>n</i> = 20)	12.10 ± 1.07	10–14	
9	G1 (<i>n</i> = 45)	12.91 ± 1.28	11–18	0.11
	G2 (<i>n</i> = 21)	12.95 ± 0.74	12–15	
10	G1 (<i>n</i> = 50)	13.64 ± 1.22	11–16	0.87
	G2 (<i>n</i> = 23)	13.65 ± 1.15	12–16	
11	G1 (<i>n</i> = 50)	15.00 ± 1.54	12–19	0.54
	G2 (<i>n</i> = 30)	15.30 ± 1.34	12–17	

orthodontist (ZZ, HL, BY with 5, 12, and 17 years of experience) were recruited to meticulously carry out this process to provide a correctness threshold for our multiple ordinal logistic models.

Data acquisition

Mimics software (Version 17.0; Materialise NV, Leuven, Belgium) was employed to execute the measurement process on images from G1 subjects. After importing all the CBCT records (Digital Imaging and Communications in Medicine known as DICOM files), image reorientation was performed based on the Swensen protocol [30] to confirm positional

homogenization. We implemented Multiplanar Reformation (MPR) mode for reconstruction of images in the three different orthogonal planes, thusly each file can portray adjacent slices of the interest areas of the cervical vertebrae in all three planes. Using a formerly introduced technique [27], the anteroposterior axis was set in the axial view defining the deepest posterior point of the C2 vertebral foramen and body midpoint. The vertical axis was coronally set to pass through odontoid process midpoint, and lastly the CBCT-generated sagittal slices of the cervical vertebrae were formed. Figure 3 is demarcating all of the elected landmarks [31], corresponding distances and ratios measured in our investigation by

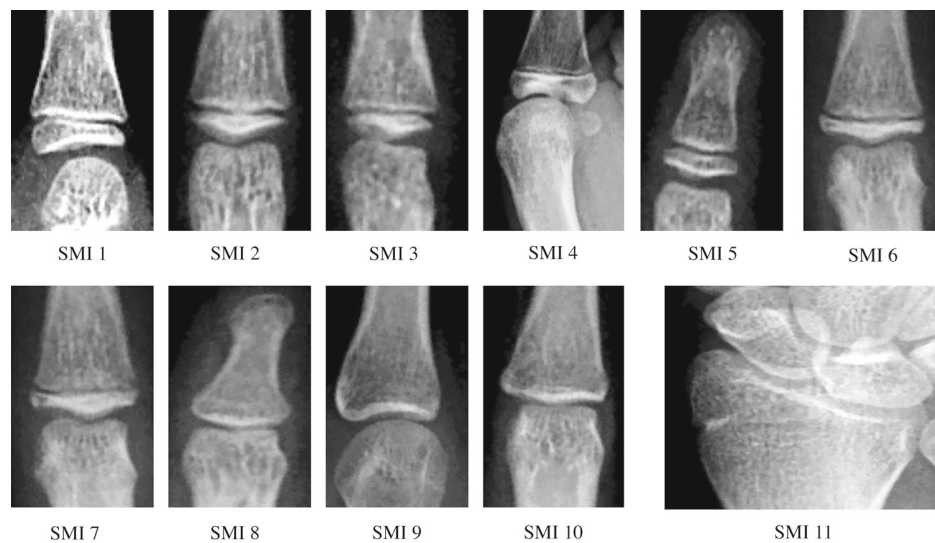


Fig. 2 Description of maturity indicators and their corresponding manifestations. I: *Epiphysis width approximating the diaphysis width*; as seen in third finger distal phalanx (SMI 1), proximal phalanx of third finger (SMI 2), and fifth finger middle phalanx (SMI 3). II: *Ossification*; emergence of abductor sesamoid of the thumb (SMI 4). III: *Capping of*

epiphysis; visible in third finger distal phalanx (SMI 5) and proximal phalanx (SMI 6). Middle phalanx of fifth finger (SMI 7). IV: *Fusion of epiphysis and diaphysis*; observed in third finger distal phalanx (SMI 8), proximal phalanx (SMI 9), and middle phalanx (SMI 10). And lastly, as seen in Radius bone (SMI 11)

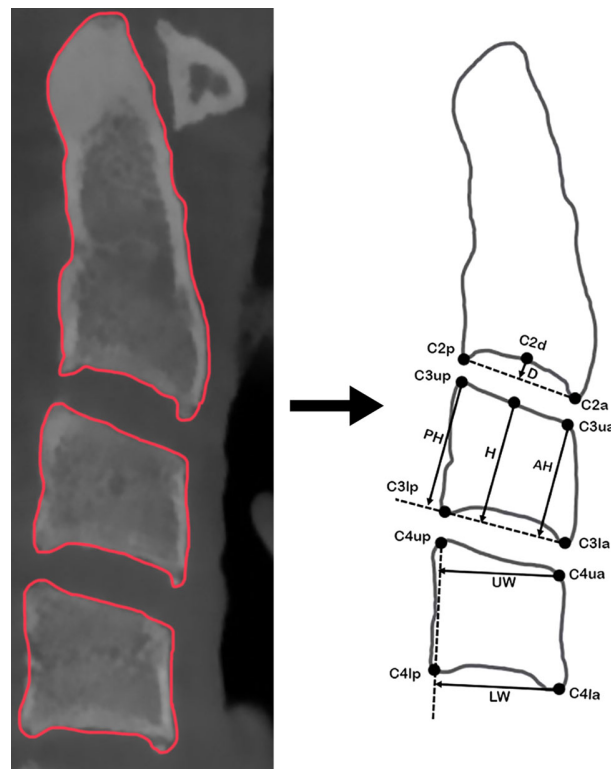


Fig. 3 Illustration of linear measurements performed on each cervical vertebra based on the following landmarks. C2a and C2p (the most anterior and posterior points on the lower border of the body of C2); C3,4la and C3,4lp (the most anterior and posterior points on the lower border

of the body of C3 and C4); C3,4ua and C3,4up (the most superior points of the anterior and posterior borders of the body of C3 and C4); and C3,4um (the middle of the upper border of the body of C3 and C4)

Table 2 Definitions of cervical vertebral variables

Parameter	Description
<i>Vertical distances</i>	
D2	The vertical distance between C2d and the line connecting C2a–C2p
D3,4	The vertical distances between C3,4d and the lines connecting C3la–C3lp and C4la–C4lp
AH3, AH4	The vertical distances between C3,4ua and the lines connecting C3la–C3lp and C4la–C4lp
H3, H4	The vertical distances between C3,4um and the lines connecting C3la–C3lp and C4la–C4lp
PH3, PH4	The vertical distances between C3,4up and the lines connecting C3la–C3lp and C4la–C4lp
UW3, UW4	The vertical distances between C3,4ua and the lines connecting C3up–C3lp and C4up–C4lp
LW3, LW4	The vertical distances between C3,4la and the lines connecting C3up–C3lp and C4up–C4lp
<i>Ratios</i>	
D3,4/AH3,4	Ratio of the distances D3,4 to those of AH3,4
PH3,4/UW3,4	Ratio of the distances PH3,4 to those of UW3,4
PH3,4/LW3,4	Ratio of the distances PH3,4 to those of LW3,4

three different orthodontics (ZZ, HL, BY). All the corresponding parameters are elaborated and defined in Table 2.

Statistical analysis

The attained data were analyzed by SPSS (version 19.0, SPSS, Inc., Chicago, USA), with P value equal to 0.05 as significant difference. The reliability and reproducibility of the attained data from observers were assessed by the intraclass correlation coefficient (ICC) for the parameter measurements and weighted Cohen's kappa tests on data from 200 randomly chosen samples for the SMI assessments. The normal distribution of the data was confirmed by Shapiro–Wilk test. With the presumption of a logistic relationship between the variables, Spearman's rank correlation coefficient was used to identify the parameters significantly associated with the SMI. To ascertain the significant independent predictors of the skeletal maturation level (as the dependent variable) multivariable ordinal logistic regression was used. Selecting the statistically significant parameters was in a stepwise manner assuming 0.05 alpha levels for either entry or removal. The variance inflation factor (VIF) was observed to control the severity of multicollinearity.

The reliability and reproducibility were assessed by the intraclass correlation coefficient (ICC) and weighted Cohen's kappa, followed by Spearman's rank correlation coefficient to identify the parameters significantly associated with the SMI.

Model development

We generated Logistic regression models grounded on the proportional-odds cumulative logit model (Eq. (1)) [32]. Our

novel proposed method yielded individual formulas for each specific stage, posing the advantage of probability determination for each skeletal maturational stage. $P(y = j|x)$ represents the expected probability of each maturational stage in the SMI index ($j = 1, \dots, 11$), whereas a_j denotes the coefficient value as the log-odds, $\beta(x)$ is the constant slope coefficient, and e is referred to as Euler's number ($= 2.71828$). Equation (2) assisted us in determining the linear function of the independent variables.

$$P(y = j|x) = \frac{1}{1 + e^{-a_j + \beta(x)}} - \frac{1}{1 + e^{-a_{j-1} + \beta(x)}} \quad (1)$$

$$\text{logit}P_j = a_j + \beta(x) = g_j(x) \quad (2)$$

For the purpose of clarity, Fig. 4 displays the fate of here developed models after analyzing their performance in form of a flowchart.

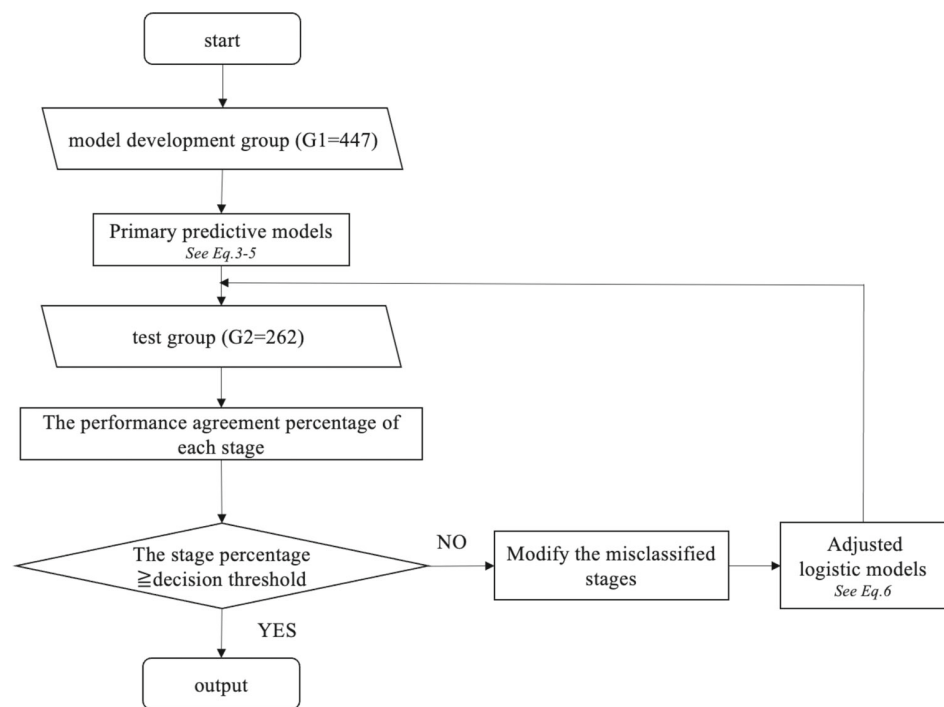
Evaluation

In order to distinguish our model performance from previously introduced methods, a linear regression model was built for estimation of skeletal maturation level after selection of significant parameters in our pilot study based on the principle of "least squares method" [33]:

$$Y = 6.655 + 0.955 \times \text{Age} + 0.406 \times \text{H3} + 1.414 \times \text{H4} + 0.664 \times \text{PH4/UW4} - 0.332 \times \text{LW4}$$

The performance agreement percentage of both linear and logistic models was tested using the data from G2 ($n = 262$). This group was subjected to measurement of the significant predictor variables only and calculation of the maturational

Fig. 4 Flowchart showing the progress of generating logistic regression models based on model development group, and evaluating the agreement percentage of each stage. After performance evaluation of our models corresponding to each stage, misclassified stages were detected when percentage lower than our decision threshold (80%) was observed



stage by the generated models. Comparing with the actual SMI stage, the model's operation was rated by direct comparison as follow:

$$\text{The prediction percentage} = \frac{\text{The number of predicted SMI stages consistent with the gold standard} * 100}{262}$$

After evaluating the performance of our models corresponding to each stage, we will inspect the misclassified stages when percentage lower than our decision threshold (80%) is observed to enhance the agreement percentage of the final model. Followed by a receiver operating characteristic (ROC) curve analysis, less accurate logistic models with unsatisfactory results were appointed for optimization.

Results

Reliability analysis

The ICC range for the measured quantitative data was between 0.947 and 0.998, indicating high inter-rater reliability. The calculated mean and standard deviation (SD) of each parameter is presented in Table 3. The weighted Cohen's kappa test showed high inter-rater reproducibility for the clinical SMI appraisals among the examiners (kappa value = 0.901).

Correlation analysis

Spearman's rank correlation coefficient analysis exhibited 20 parameters significantly correlated by considering the SMI stage as the dependent variable and the measured parameters as the independent variables ($P < 0.05$).

Model development

Logistic model formulation

After application of multiple ordinal logistic regression, significantly correlated parameters with the cervical vertebral maturational stage were narrowed down to chronological age, H3, UW4, LW4, and PH4/LW4 ($P < 0.05$). Equation (3) computed $\beta(x)$ for each of the subjects as a function of their significant predictor variables. The coefficient B exhibited in Table 4 represents a_j . The final model for the prediction of each stage was determined as Eq. (3), which results in six P_j ($j = 1, \dots, 11$) values ranging between 0 and 1, indicative of each stage probability. Followed by calculating each stage

Table 3 Descriptive statistics and correlational analysis results of the 20 parameters

Variables	Mean	Standard deviation	Correlation coefficient	P value
Age (years)	12.09	1.99	.801**	<.001
D2 (mm)	1.43	0.90	.644**	<.001
D3 (mm)	1.12	0.86	.745**	<.001
D4 (mm)	0.93	0.86	.740**	<.001
AH3 (mm)	10.30	5.43	.882**	<.001
AH4 (mm)	9.91	2.28	.860**	<.001
H3 (mm)	11.02	2.33	.872**	<.001
H4 (mm)	11.00	2.21	.892**	<.001
PH3 (mm)	12.34	2.31	.790**	<.001
PH4 (mm)	12.19	2.28	.295**	<.001
UW3 (mm)	13.96	1.65	.187**	<.001
UW4 (mm)	13.62	1.92	.758**	<.001
LW3 (mm)	14.03	1.89	.774**	<.001
LW4 (mm)	14.37	1.87	.188**	<.001
D3/AH3	0.10	0.07	.168**	<.001
D4/AH4	0.09	0.08	.770**	<.001
PH3/UW3	0.92	0.14	.777**	<.001
PH4/UW4	0.90	0.14	.801**	<.001
PH3/LW3	0.87	0.14	.644**	<.001
PH4/LW4	0.85	0.13	.745**	<.001

** Significantly correlated parameters based on $P < 0.01$

probability thru the equation series, the largest estimated P_j value among the 11 results designated the current skeletal maturation stage for each individual ($VIF < 5$; tolerance > 2 , indicating no collinearity between parameters).

To make the final decision about the skeletal maturational stage the highest value ($\arg\max P_j$) will be the most possible HWI stage (y).

$$\beta(x) = 3.769x_1 + 2.678x_2 - 1.719x_3 + 2.16x_4 + 1.006x_5 \quad (3)$$

$$\text{logit}P_j = a + \beta(x) = g_j(x) \quad (4)$$

$$y = \arg \max P_j \quad (j = 1 - 11) \quad (5)$$

Performance analysis

For the linear model, the results (Table 5) showed 58% prediction percentage (Adjusted $R^2 = 0.881$). For the logistic model, the initial performance assessment yielded a total agreement percentage of 84%. As displayed in Fig. 5a, the 6th and 8th stage estimates were remarkably different from their corresponding Fishman SMI stages (63% and 65%, respectively). Upon thorough inspection, it was revealed that these stages were confused with their immediate stages either before or after (stages 5 to 9). Illustrated in the ROC curve (Fig. 5b), the original total AUC was equal to 0.92, indicating

a contenting model fitness. Final decision was set to updating models of stages 5 to 9.

Logistic model updating

Model optimization

Adjusted logistic models were constructed for precisely distinguishing the 5th to 9th stages and improving consistency based on the measurements from 215 subjects in the model development group who were ascertained to be at those specific SMI stages. As displayed in Table 6, Age, H3, PH4/UW4, and LW4 exhibited statistically significant correlations ($P < 0.05$) and were employed to generate 5 newly adjusted logistic regression equations based on Eq. (4).

$$\beta'(x) = 1.345x_1 + 1.769x_2 + 1.226x_3 - 0.621x_4 \quad (6)$$

Performance reevaluation

An increase in the percentage level was observed in the performance of adjusted models; examined on the 120 samples skeletally matured at SMI stages of 5 to 9 in G2. The total agreement was also enhanced as high as 88% (Fig. 6a). Adjunctively, the AUC based on the ROC results for the total predictions was upgraded to 0.94 after performance of the models was reexamined (Fig. 6b).

Table 4 Multivariate Associations Determined by Ordinal Logistic Regression

	B	Standard error of B	Wald statistic	Sig	Odds ratio	Confidence interval (95%)	
						Minimum	Maximum
<i>Thresholds</i>							
[SMI = 1]	− 9.888	0.522	358.553	0.000		− 10.911	− 8.864
[SMI = 2]	− 7.445	0.404	339.552	0.000		− 8.237	− 6.653
[SMI = 3]	− 5.431	0.321	286.123	0.000		− 6.060	− 4.801
[SMI = 4]	− 3.912	0.262	223.013	0.000		− 4.425	− 3.398
[SMI = 5]	− 2.371	0.211	125.943	0.000		− 2.785	− 1.957
[SMI = 6]	− 0.943	0.184	26.270	0.000		− 1.304	− 0.582
[SMI = 7]	0.712	0.185	14.721	0.000		0.348	1.075
[SMI = 8]	2.272	0.218	108.230	0.000		1.844	2.700
[SMI = 9]	4.752	0.318	223.679	0.000		4.129	5.375
[SMI = 10]	8.238	0.473	303.568	0.000		7.311	9.164
<i>Predictors</i>							
Chronological age*	3.769	0.279	182.280	0.000 [†]	43.32131	3.222	4.316
D2mm	0.117	0.206	0.322	0.570	−	− 0.287	0.520
D3mm	1.686	0.959	3.089	0.079	−	− 0.194	3.566
D4	− 0.507	0.317	2.556	0.110	−	− 1.129	0.115
AH3	− 0.102	0.117	0.749	0.387	−	− 0.332	0.129
AH4	0.006	0.513	0.000	0.991	−	− 4.209	4.946
H3*	2.678	0.237	127.841	0.000 [†]	14.56134	2.214	3.143
H4	1.117	0.577	3.745	0.053	−	− 0.014	2.248
PH3	− 2.996	1.947	2.368	0.124	−	− 6.811	0.820
PH4	0.119	2.065	0.003	0.954	−	− 3.928	4.166
UW3	− 0.428	0.395	1.179	0.278	−	− 1.202	0.345
UW4*	− 1.719	0.234	53.733	0.000 [†]	0.179299	− 2.178	− 1.259
LW3	2.423	1.423	2.899	0.089	−	− 0.366	5.212
LW4*	1.006	0.251	16.090	0.000 [†]	2.735742	0.515	1.498
D3/AH3	− 0.368	0.818	0.203	0.653	−	− 1.970	1.234
D4/AH4	2.032	1.161	3.064	0.080	−	− 0.243	4.308
PH3/UW3	2.255	2.037	1.225	0.268	−	− 1.738	6.248
PH4/UW4	− 0.123	2.555	0.002	0.961	−	− 5.131	4.884
PH3/LW3	2.720	1.716	2.513	0.113	−	− 0.643	6.082
PH4/LW4*	2.160	0.193	124.682	0.000 [†]	8.667925	1.781	2.539

B: Standardized regression coefficient, *Significant predictors based on $P < 0.05$, [†]Entered significant predictor variables

Table 5 Multivariate associations determined by linear regression

		Parameter	B	Standard error of B	t	Sig.*	95% CI	
							Lower	Upper
Location	Chronologic age		.955	.178	5.365	.000	.604	1.307
	H3		.406	.173	2.348	.020	2.348	.020
	PH4/UW4		.664	.140	4.741	.000	.387	.940
	LW4		− .332	.113	− 2.954	.004	− .554	− .110
	H4		1.141	.220	5.183	.000	.707	1.576
	Constant		6.655	.089	74.473	.000	6.479	6.832

B: Standardized regression coefficient, *Significant predictors based on $P < 0.05$

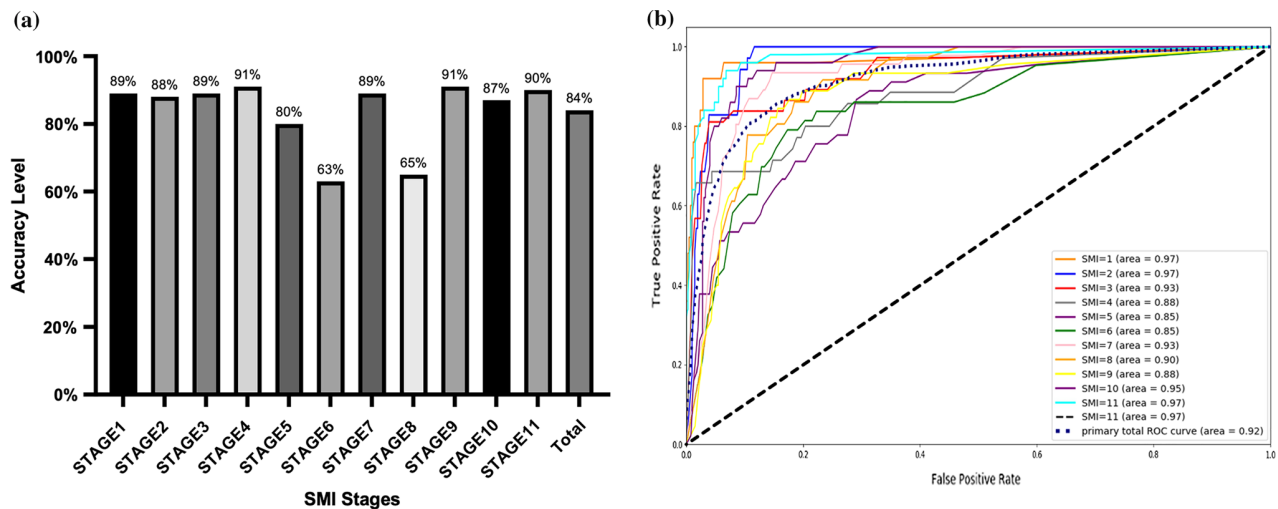


Fig. 5 The prediction percentage of original models. **a** Bar chart: Comparison of model percentage for different stages. **b** Receiver operating characteristic curve: Separate ROC curves for each model and the total performance

Table 6 Secondary Parameter Estimates and Bivariate Associations Determined by Ordinal Logistic Regression

	Parameter	B	Standard error of B	Wald statistic	Sig.*	Odds ratio	95% CI for odds ratio	
							Lower	Upper
Threshold	[SMI = 5]	- 2.605	0.256	103.594	0.000	0.07	- 3.106	- 2.103
	[SMI = 6]	- 0.927	0.193	23.115	0.000	0.40	- 1.305	- 0.549
	[SMI = 7]	0.783	0.192	16.648	0.000	2.19	0.407	1.159
	[SMI = 8]	2.408	0.251	91.997	0.000	11.11	1.916	2.899
Location	Chronologic age	1.345	0.190	50.247	0.000	1.345	0.973	1.717
	H3	1.769	0.233	57.504	0.000	1.769	1.312	2.226
	PH4/UW4	1.226	0.184	44.393	0.000	1.226	0.865	1.587
	LW4	- 0.621	0.187	10.983	0.001	- 0.621	- 0.989	- 0.254

B: Standardized regression coefficient, *Significant predictors based on $P < 0.05$

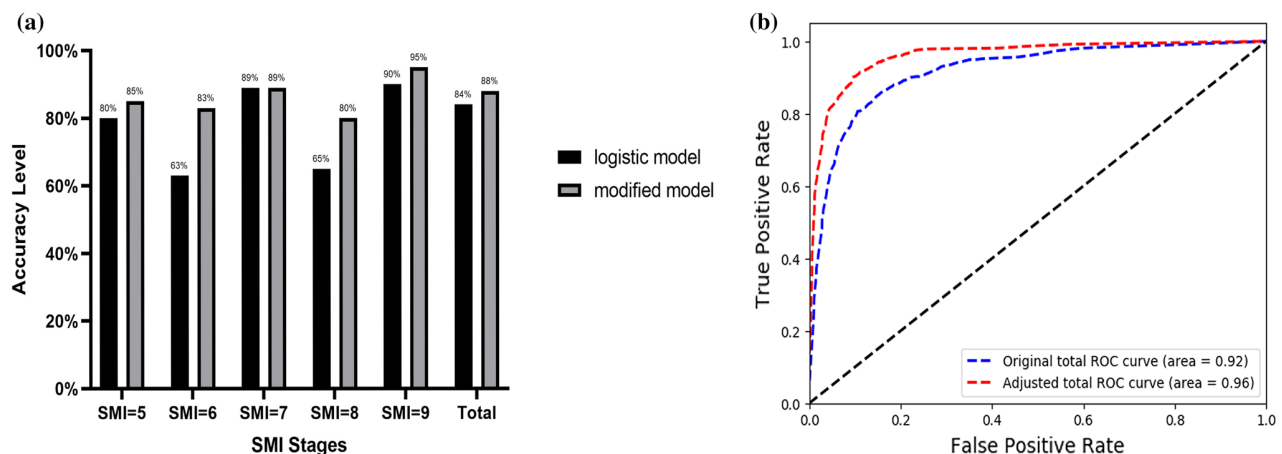


Fig. 6 Performance of the models after adjustment. **a** Bar chart: Comparison of model percentage for different stages before and after optimization. **b** Receiver operating characteristic curve: ROC curves for total model performance before and after adjustment

Discussion

The ossification of hand wrist bones is well known as an indicator of the skeletal maturity level and somatic growth potential of adolescents [34]. CBCT advancements along with emergence of biomathematical modeling and ML in clinical orthodontics motivated us to design a multiple ordinal logistic algorithm to objectively score individuals SMI. Having CBCT images for indicated circumstances gives clinicians an upper hand to avoid an additional X-ray exposure by analyzing cervical vertebral morphological changes as skeletal maturity estimates. Thus, we benefited from the opportunity of having a large sample size of CBCT and hand wrist images to construct an interconnected pattern for eliminating the need for extra exposures in future. The correlation between CVM and SMI have been inspected by different scholars through various approaches [9, 35, 36]. Nevertheless, ML technology can explore the potential correlativity of CVM and SMI, and insignificantly depends on the consistency between two methods. Therefore, the discussion may not have negative effect on our result. In this study, we presented an intelligent innovative multistage model to assess the skeletal maturation based on two-dimensional parameters of cervical vertebrae, derived from CBCT. To understand the interrelationship between somatic growth and sexual maturity, we cannot avoid the fact that secretion of gonadotropin-releasing (GnRH) and sex hormones initiate puberty along with accelerating the jaws growth, especially among females [37–39] leading to a disparity between chronological and skeletal age. In addition to that, due to the effects of gender, socioeconomic, and ethnic factors on skeletal growth and development [40, 41], we only included images from Chinese females.

An essential point regarding our implemented method is that we used the skeletal maturation assessment method proposed by Fishman [29]. This method is declared to be more appropriate as a gold standard due to minimizing the effect of demographic features of the sample [1, 42]. Moreover, to ensure consistency of our results our study population was comprised of only females.

AI is aimed to design machines that mimic our perception and logic, and are capable of performing tasks same as or better than humans. This research also intends to design a scheme that can perform tasks that require human intelligence through ML. ML appliance in clinical medicine has gained a high level of attentiveness as a new approach for empowering diagnostic procedures, prognosis evaluation, computational pharmacology, oral and maxillofacial surgery and orthopedics, and more complex biological processes such as DNA sequencing and epigenomic variation and gene regulation [43–45]. Regarding the skeletal maturation assessment, ML is reported to increase accuracy and decrease variability [46, 47]. Logistic regression is a method

of ML that can assist in exact detection of maturational level by comparing the probability of each possible stage.

Previous proposed regression-based attempts were based on the linear form of this technique [26, 27, 31]. In our study, an acceptable degree of fitness was observed for the linear model (Adjusted R^2 equals 0.881, Table 5). However, aside from its lower reported percentage (58%), this method was limited only to selecting the most correlated parameters with the SMI and not capable of detecting the exact maturational stage rather presenting only a possible range. While, our presented logistic model successfully determined the skeletal maturation level by comparing the probabilities of each 11 individual SMI stages for each subject with an agreement percentage of 84% and the total ROC area under the curve of 0.92, introducing a new approach for clinical assessment of skeletal maturation. This was in strong accordance with Santiago et al. work, whereas the logistic model that combined the parameters of C2, C3, and C4, age, and gender could successfully report an excellent prediction although with smaller sample size [42]. In harmony with that, Miguel-Hurtado's experiment also exhibited that how a classification system based on ML can outclass linear regression in skeletal maturation assessment owing to its aptitude in search space narrowing and granting certitude in clinical decision making [25].

Amasya et al. developed an artificial neural network (ANN) model for cervical vertebral maturation (CVM) analysis with a close agreement between the ANN model and the human observers [13, 48]. In another study, Mutasa et al. introduced a novel deep learning architecture for machine-assisted bone age labeling with promising results [49]. Deep learning driven methods continuously optimize model parameters by labeling a large amount of data without the ability to either mark the actual significant parameters in the analyzed image or finding the drawback loci. This learning process is considered still as "black box" in medicine confronting with culpability and legal consequences [50]. Contrarily, not only our proposed model was merely established based on significant parameters, but also our result could indicate which specific stages are mistakenly predicted or even confused with each other.

Our performance analysis revealed that compared to other stages, the 6th and 8th stages were more difficult to be predicted accurately and were misperceived with their before and next adjacent stages, compelling us to update all our models between 5 and 9th. The underlying reason behind this could be these stages are the transition from the peak period of growth and development to the deceleration period. The age range of these two stages is relatively concentrated (transition to deceleration period) which occurs in a short period of time. On the other hand, the morphological similarities of the bodies of third and fourth cervical vertebrae and their analogous changeover from being rectangular to square form

might have projected an indistinctive amount of growth and ultimately confusion in our model performance. Following that, by reevaluating and addition of a new significant parameter ($PH4/UW4$) our models were optimized. Dissimilar to deep learning and similar to human recognition, the accuracy and reliability of our presented model could be enhanced by further screening and discovering more significant parameters influencing certain stages.

Our presented CBCT-based model is prospective and will have a great contribution to clinical dentistry and also other multi-stage medical conditions in the future; however, since the chronological age acts as an essential parameter, our models cannot be yet employed in the vicinities where subject age is missing such as forensic field. At the same time, the workload of clinicians will be reduced and more objective and consistent evaluation results would be materialized. Although this study benefited from the utilization of CBCT and an accurate measurement process, a clinical indication solely for skeletal maturational estimation cannot be implied. Generally, the establishment of skeletal maturation norms by regression models requires a vast range of data from samples with various demographic features which is not feasible via CBCT considering its restrained clinical indications. Another potential limitation of this study could be the number of observers which is far from Obuchowski's recommendation of having minimum of 10 observers to perform generalize diagnostic tests [13]. Since cervical vertebral ossification is highly relatable to regional and ethnical diversities, the differences between our sample and the general population should not be overlooked and additional studies with a larger sample size to establish externally applicable models are recommended.

Conclusions

Consistency and fitness evaluation of our models demonstrated adequate prediction percentage and reliability of our method for automated classification of skeletal maturation. The presented constructed logistic regression model has the potential to serve as a maturity evaluation index in clinical craniofacial orthopedics in Chinese girls. Intelligent staging of SMI with the assistance of AI facilitates enhancement of the accuracy and objectivity in skeletal maturation evaluation methods. The proposed model in this study showed promising strength for being expended in the event of other clinical multi-stage conditions. Despite optimization of our models, as for any other AI procedure, the role of clinicians' supervision cannot be overruled.

Authors' contributions LX and WT designed the study and performed the experiments. Data collection and manuscript writing were done by II and WT. ZZ and YZ carried out statistical analysis and interpretation

of the results. HL and BY conceived and supervised the whole project. All authors read and approved the final manuscript.

Funding This study is supported by National Natural Science Foundation of China (Grant no. 82071143 and no. 82101079), Key Medical Research Projects of Jiangsu Health Commission (ZDA2020003), Key R&D program of Jiangsu province (BE2018723), Natural Science Foundation of Jiangsu province (BK20180670), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD, 2018-87).

Availability of data and material The data that support the findings of this study are available on request from the corresponding author, [BY]. The data are not publicly available due to being currently used in another ongoing research project.

Code availability Not applicable.

Declarations

Conflicts of interest The authors declare no potential conflicts of interest with respect to the content, authorship and/or publication of this article.

Ethics approval This study was approved by the Institutional ethical committee of Nanjing Medical University (No. PJ2017-045-001).

Consent to participate Written informed consent for utilization of scans in future research projects was obtained from all individuals, their parents or legal tutors prior to taking all the 709 images.

Consent for publication All the authors certify that human research participants provided informed consent for publication of their data and radiographic images in Figs. 2 and 3.

References

1. Flores-Mir C, Nebbe B, Major PW (2004) Use of skeletal maturation based on hand-wrist radiographic analysis as a predictor of facial growth: a systematic review. *Angle Orthod* 74:118–124. [https://doi.org/10.1043/0003-3219\(2004\)074%3c0118:UOSMBO%3e2.0.CO;2](https://doi.org/10.1043/0003-3219(2004)074%3c0118:UOSMBO%3e2.0.CO;2)
2. Wong RWK, Alkhal HA, Rabie ABM (2009) Use of cervical vertebral maturation to determine skeletal age. *Am J Orthod Dentofac Orthop* 136:484.e1–484.e6. <https://doi.org/10.1016/j.ajodo.2007.08.033>
3. Baccetti T, Franchi L, Toth LR, McNamara JA (2000) Treatment timing for twin-block therapy. *Am J Orthod Dentofac Orthop*. <https://doi.org/10.1067/mod.2000.105571>
4. Arat ZM, Rübendüz M, Akgül AA (2003) The displacement of craniofacial reference landmarks during puberty: a comparison of three superimposition methods. *Angle Orthod*. [https://doi.org/10.1043/0003-3219\(2003\)073%3c0374:TDOCRL%3e2.0.CO;2](https://doi.org/10.1043/0003-3219(2003)073%3c0374:TDOCRL%3e2.0.CO;2)
5. Nanda RS (1955) The rates of growth of several facial components measured from serial cephalometric roentgenograms. *Am J Orthod* 41:658–673
6. Hägg U, Taranger J (1982) Maturation indicators and the pubertal growth spurt. *Am J Orthod* 82:299–309
7. Lamparski DG (1972) Skeletal age assessment utilizing cervical vertebrae. Master Science thesis, University of Pittsburgh

8. Greulich W, Pyle S (1959) Radiographic ossification and the adolescent growth spurt. *Am J Orthod* 69:611–619
9. Cunha AC, Cevidanes LHS, Sant'Anna EF, Guedes FR, Luiz RR, McNamara JA, Franchi L, Ruellas ACO (2018) Staging hand–wrist and cervical vertebrae images: a comparison of reproducibility. *Dentomaxillofac Radiol* 47:20170301. <https://doi.org/10.1259/dmfr.20170301>
10. Uysal T, Ramoglu SI, Basciftci FA, Sari Z (2006) Chronologic age and skeletal maturation of the cervical vertebrae and hand–wrist: is there a relationship? *Am J Orthod Dentofac Orthop* 130:622–628. <https://doi.org/10.1016/j.ajodo.2005.01.031>
11. San Román P, Palma JC, Oteo MD, Nevado E (2002) Skeletal maturation determined by cervical vertebrae development. *Eur J Orthod* 24:303–311. <https://doi.org/10.1093/ejo/24.3.303>
12. Kök H, Acilar AM, İzgi MS (2019) Usage and comparison of artificial intelligence algorithms for determination of growth and development by cervical vertebrae stages in orthodontics. *Prog Orthod*. <https://doi.org/10.1186/s40510-019-0295-8>
13. Amasya H, Cesur E, Yıldırım D, Orhan K (2020) Validation of cervical vertebral maturation stages: artificial intelligence vs human observer visual analysis. *Am J Orthod Dentofac Orthop* 158:e173–e179. <https://doi.org/10.1016/j.ajodo.2020.08.014>
14. Dzemidzic V, Sokic E, Tiro A, Nakas E (2015) Computer based assessment of cervical vertebral maturation stages using digital lateral cephalograms. *Acta Inform Med* 23:364–368. <https://doi.org/10.5455/aim.2015.23.364-368>
15. Santiago RC, Cunha AR, Júnior GC, Fernandes N, Campos MJS, Costa LFM, Vitral RWF, Bolognese AM (2014) New software for cervical vertebral geometry assessment and its relationship to skeletal maturation—a pilot study. *Dentomaxillofac Radiol* 43:20130238. <https://doi.org/10.1259/dmfr.20130238>
16. Tekin A, Cesur Aydın K (2019) Comparative determination of skeletal maturity by hand–wrist radiograph, cephalometric radiograph and cone beam computed tomography. *Oral Radiol*. <https://doi.org/10.1007/s11282-019-00408-y>
17. Echevarría-Sánchez G, Arriola-Guillén LE, Malpartida-Carrillo V, Tinedo-López PL, Palti-Menendez R, Guerrero ME (2020) Reliability of cephalograms derived of cone beam computed tomography versus lateral cephalograms to estimate cervical vertebrae maturity in a Peruvian population: a retrospective study. *Int Orthod* 18:258–265. <https://doi.org/10.1016/j.ortho.2020.01.001>
18. Tadinada A, Schneider S, Yadav S (2018) Role of cone beam computed tomography in contemporary orthodontics. *Semin Orthod* 24:407–415. <https://doi.org/10.1053/j.sodo.2018.10.005>
19. Jain S, Choudhary K, Nagi R, Shukla S, Kaur N, Grover D (2019) New evolution of cone-beam computed tomography in dentistry: combining digital technologies. *Imaging Sci Dent* 49:179–190. <https://doi.org/10.5624/isd.2019.49.3.179>
20. Fasbinder DJ, Dennison JB, Heys D, Neiva G (2008) Practical applications of cone-beam computed tomography in orthodontics. *J Am Dent Assoc* 141(Suppl):10S-S14. <https://doi.org/10.1002/jps.3030451211>
21. Currie G, Hawk KE, Rohren E, Vial A, Klein R (2019) Machine learning and deep learning in medical imaging: intelligent imaging. *J Med Imaging Radiat Sci* 50:477–487. <https://doi.org/10.1016/j.jmir.2019.09.005>
22. Schwendicke F, Samek W, Krois J (2020) Artificial intelligence in dentistry: chances and challenges. *J Dent Res* 99:769–774. <https://doi.org/10.1177/0022034520915714>
23. Dallora AL, Anderberg P, Kvist O, Mendes E, Ruiz SD, Berglund JS (2019) Bone age assessment with various machine learning techniques: a systematic literature review and meta-analysis. *PLoS ONE* 14:1–22. <https://doi.org/10.1371/journal.pone.0220242>
24. Gutierrez PA, Perez-Ortiz M, Sanchez-Monedero J, Fernandez-Navarro F, Hervas-Martinez C (2016) Ordinal regression methods: survey and experimental study. *IEEE Trans Knowl Data Eng* 28:127–146. <https://doi.org/10.1109/TKDE.2015.2457911>
25. Miguel-Hurtado O, Guest R, Stevenage SV, Neil GJ, Black S (2016) Comparing machine learning classifiers and linear/logistic regression to explore the relationship between hand dimensions and demographic characteristics. *PLoS ONE* 11:e0165521. <https://doi.org/10.1371/journal.pone.0165521>
26. Byun BR, Il KY, Yamaguchi T, Maki K, Ko CC, Hwang DS, Park SB, Son WS (2015) Quantitative skeletal maturation estimation using cone-beam computed tomography-generated cervical vertebral images: a pilot study in 5- to 18-year-old Japanese children. *Clin Oral Investig* 19:2133–2140. <https://doi.org/10.1007/s00784-015-1415-6>
27. Byun B-R, Kim Y-I, Yamaguchi T, Maki K, Son W-S (2015) Quantitative assessment of cervical vertebral maturation using cone beam computed tomography in Korean girls. *Comput Math Methods Med* 2015:405912. <https://doi.org/10.1155/2015/405912>
28. Tripepi G, Jager KJ, Dekker FW, Zoccali C (2008) Linear and logistic regression analysis. *Kidney Int* 73:806–810. <https://doi.org/10.1038/sj.ki.5002787>
29. Fishman LS (1982) Radiographic evaluation of skeletal maturation. A clinically oriented method based on hand–wrist films. *Angle Orthod* 52:88–112. [https://doi.org/10.1043/0003-3219\(1982\)0522.0.CO;2](https://doi.org/10.1043/0003-3219(1982)0522.0.CO;2)
30. Swennen GRJ, Schutyser F, Hausamen J-E (2006) Three-dimensional cephalometry: a color atlas and manual. Springer, Berlin
31. Chen L, Liu J, Xu T, Long X, Lin J (2010) Quantitative skeletal evaluation based on cervical vertebral maturation: a longitudinal study of adolescents with normal occlusion. *Int J Oral Maxillofac Surg* 39:653–659. <https://doi.org/10.1016/j.ijom.2010.03.026>
32. Grilli L, Rampichini C (2014) Encyclopedia of quality of life and well-being research. Springer, Dordrecht
33. Jaeger B (2006) The method of least squares. In: *Handbook of research on informatics in healthcare and biomedicine*, pp 181–185. IGI Global
34. Hashim H, Mansoor H, Mohamed MH (2018) Assessment of skeletal age using hand–wrist radiographs following Bjork system. *J Int Soc Prev Community Dent* 8:482. https://doi.org/10.4103/jispcd.JISPCD_315_18
35. Ferrillo M, Curci C, Rocuzzo A, Migliario M, Invernizzi M, de Sire A (2021) Reliability of cervical vertebral maturation compared to hand–wrist for skeletal maturation assessment in growing subjects: a systematic review. *J Back Musculoskelet Rehabil*. <https://doi.org/10.3233/BMR-210003>
36. Szemraj A, Wojtaszek-Słomińska A, Racka-Pilszak B (2018) Is the cervical vertebral maturation (CVM) method effective enough to replace the hand–wrist maturation (HWM) method in determining skeletal maturation?—a systematic review. *Eur J Radiol* 102:125–128. <https://doi.org/10.1016/j.ejrad.2018.03.012>
37. Demirjian A, Buschang PH, Tanguay R, Patterson DK (1985) Interrelationships among measures of somatic, skeletal, dental, and sexual maturity. *Am J Orthod* 88:433–438. [https://doi.org/10.1016/0002-9416\(85\)90070-3](https://doi.org/10.1016/0002-9416(85)90070-3)
38. Fishman LS (1979) Chronological versus skeletal age, an evaluation of craniofacial growth. *Angle Orthod* 49:181–189. [https://doi.org/10.1043/0003-3219\(1979\)049%3c0181:CVSAAE%3e2.0.CO;2](https://doi.org/10.1043/0003-3219(1979)049%3c0181:CVSAAE%3e2.0.CO;2)
39. Morris JM, Park JH (2012) Correlation of dental maturity with skeletal maturity from radiographic assessment: a review. *J Clin Pediatr Dent* 36:309–314. <https://doi.org/10.17796/jcpd.36.3.1403471880013622>
40. Cericato GO, Bittencourt MAV, Paranhos LR (2015) Validity of the assessment method of skeletal maturation by cervical vertebrae: a systematic review and meta-analysis. *Dentomaxillofac Radiol*. <https://doi.org/10.1259/dmfr.20140270>

41. Kang ST, Choi SH, Kim KH, Hwang CJ (2020) Evaluation of cephalometric characteristics and skeletal maturation of the cervical vertebrae and hand–wrist in girls with central precocious puberty. *Korean J Orthod* 50:181–187. <https://doi.org/10.4041/kjod.2020.50.3.181>
42. Santiago RC, Cunha AR, Júnior GC, Fernandes N, Campos MJS, Costa LFM, Vitral RWF, Bolognese AM (2014) New software for cervical vertebral geometry assessment and its relationship to skeletal maturation—a pilot study. *Dentomaxillofac Radiol*. <https://doi.org/10.1259/dmfr.20130238>
43. Rajkomar A, Dean J, Kohane I (2019) Machine learning in medicine. *N Engl J Med* 380:1347–1358. <https://doi.org/10.1056/nejmra1814259>
44. Shan T, Tay FR, Gu L (2020) Application of artificial intelligence in dentistry. *J Dent Res*. <https://doi.org/10.1177/0022034520969115>
45. Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, Hoffman MM (2019) Machine learning for integrating data in biology and medicine: principles, practice, and opportunities. *Inf Fusion* 50:71–91. <https://doi.org/10.1016/j.inffus.2018.09.012>
46. Tajmir SH, Lee H, Shailam R, Gale HI, Nguyen JC, Westra SJ, Lim R, Yune S, Gee MS, Do S (2019) Artificial intelligence-assisted interpretation of bone age radiographs improves accuracy and decreases variability. *Skeletal Radiol* 48:275–283. <https://doi.org/10.1007/s00256-018-3033-2>
47. Booz C, Yel I, Wichmann JL, Boettger S, Al Kamali A, Albrecht MH, Martin SS, Lenga L, Huizinga NA, D’Angelo T, Cavallaro M, Vogl TJ, Bodelle B (2020) Artificial intelligence in bone age assessment: accuracy and efficiency of a novel fully automated algorithm compared to the Greulich–Pyle method. *Eur Radiol Exp* 4:6. <https://doi.org/10.1186/s41747-019-0139-9>
48. Amasya H, Yildirim D, Aydogan T, Kemaloglu N, Orhan K (2020) Cervical vertebral maturation assessment on lateral cephalometric radiographs using artificial intelligence: comparison of machine learning classifier models. *Dentomaxillofac Radiol* 49:20190441. <https://doi.org/10.1259/dmfr.20190441>
49. Mutasa S, Chang PD, Ruzal-Shapiro C, Ayyala R (2018) MABAL: a novel deep-learning architecture for machine-assisted bone age labeling. *J Digit Imaging* 31:513–519. <https://doi.org/10.1007/s10278-018-0053-3>
50. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88. <https://doi.org/10.1016/j.media.2017.07.005>

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