ORIGINAL ARTICLE

Machine learning and deep learning enabled age estimation on medial clavicle CT images

Lirong Qiu[1] · Anjie Liu[1,2] · Xinhua Dai[3] · Guangfeng Liu[1] · Zhao Peng[4] · Mengjun Zhan[1] · Junhong Liu[1] · Yufan Gui[1] · Haozhe Zhu[5] · Hu Chen[5] · Zhenhua Deng[1] · Fei Fan[1]

Abstract

The medial clavicle epiphysis is a crucial indicator for bone age estimation (BAE) after hand maturation. This study aimed to develop machine learning (ML) and deep learning (DL) models for BAE based on medial clavicle CT images and evaluate the performance on normal and variant clavicles. This study retrospectively collected 1049 patients (mean± SD: 22.50±4.34 years) and split them into normal training and test sets, and variant training and test sets. An additional 53 variant clavicles were incorporated into the variant test set. The development stages of normal MCE were used to build a linear model and support vector machine (SVM) for BAE. The CT slices of MCE were automatically segmented and used to train DL models for automated BAE. Comparisons were performed by linear versus ML versus DL, and normal versus variant clavicles. Mean absolute error (MAE) and classification accuracy was the primary parameter of comparison. For BAE, the SVM had the best MAE of 1.73 years, followed by the commonly-used CNNs (1.77–1.93 years), the linear model (1.94 years), and the hybrid neural network CoAt Net (2.01 years). In DL models, SE Net 18 was the best-performing DL model with similar results to SVM in the normal test set and achieved an MAE of 2.08 years in the external variant test. For age classification, all the models exhibit superior performance in the classification of 18-, 20-, 21-, and 22-year thresholds with limited value in the 16-year threshold. Both ML and DL models produce desirable performance in BAE based on medial clavicle CT.

Keywords Age determination by skeleton · Forensic anthropology · Clavicle CT · Machine learning · Deep learning

---

Lirong Qiu, Anjie Liu, and Xinhua Dai have contributed equally to this study.

Zhenhua Deng and Fei Fan have contributed equally to this study.

✉ Zhenhua Deng
dengzhenhua@scu.edu.cn

✉ Fei Fan
fanfei@scu.edu.cn

[1] West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu 610041, People's Republic of China

[2] University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China

[3] Department of Laboratory Medicine, West China Hospital, Sichuan University, Chengdu 610041, People's Republic of China

[4] Department of Radiology, West China Hospital, Sichuan University, Chengdu 610041, People's Republic of China

[5] College of Computer Science, Sichuan University, Chengdu 610064, People's Republic of China

Introduction

Living age estimation is a crucial issue in forensic science, competitive sports, protecting children's rights, and illegal immigrants. The hand is one of the most reliable methods to estimate age in children up to 14–15 years [1]. After hand maturation, the medial clavicle epiphysis (MCE) is considered the optimal choice for age estimation due to its ongoing development, strong correlation with age, and easy accessibility. Previous studies have demonstrated that MCE development continues until around 30 years old [1–3]. Thin-slice CT is recommended and commonly used to observe the developmental stages of MCE, as it provides clear visualization without overlapping [4].

Diverse qualitative or quantitative methods have been proposed for age estimation. While, Schmeling et al.'s staging method [5] and Kellinghaus et al.'s substage method [6] are the most widely used approaches and have been approved in regions such as Europe, Asia, and Africa. Except for the linear regression used by Shedge et al. [7], most studies

have employed general descriptive analysis to investigate the age distribution across different stages for MCE-based age estimation. The minimum age principle, which is based on descriptive statistics, is utilized for age classification [8]. The minimum age principle is the youngest age of each stage in the reference study. If the minimum age of a stage is over the 18-year threshold, the occurrence of this stage predicts that the individual is older than 18 years. Most studies [3, 7, 9, 10] have demonstrated that individuals with grades 3c and above were all over 18 years old, highlighting the crucial role of MCE in classifying the 18-year threshold. However, the descriptive analysis is susceptible to bias from sample distribution [11].

Recently, machine learning (ML) has emerged as a powerful tool for age estimation. These ML approaches can establish nonlinear models with strong prediction ability [12]. The support vector machine (SVM) is the most used ML algorithm and showed excellent performance in age estimation and classification [11, 13–15]. Based on the handcrafted features, the SVM algorithm can effectively address the limitation of linear regression in handling non-linear relationships between age and bone features, thus enabling the exploration of complex nonlinear patterns [11].

Despite the advancements in machine learning, handcrafted features are still a crucial component of ML models. The shape variation of MCE, which cannot be analyzed manually, may affect the accuracy of age estimation. Deep learning (DL) can automatically identify and extract multi-level imaging features and achieve end-to-end learning. Several DL algorithms have been successfully applied in the field of age estimation, such as U-Net, Res Net, and VGG. To date, only a few studies used DL for analyzing the medial clavicle [16, 17], with their primary objective being to classify adult age from clavicle X-ray and the localization of MCE on CT. And a low accuracy was noticed in age classification. To our knowledge, no artificial intelligence study has performed automated clavicle feature extraction for age estimation.

This study aimed to explore the potential of ML in enhancing age estimation within the manual staging method and develop several DL models for automated age estimation. The models' performance for age estimation and classification was compared and discussed in detail.

Materials and methods

Data acquisition

This retrospective study initially comprised 1049 patients aged 14.00–29.99 years undergoing chest CT scans for clinical routine at the West China Hospital of Sichuan University between October 2016 and June 2022. Patients who underwent chest CT examination in a supine position with raised arms were included in this study as recommended by Tozakidou et al. [18]. Patients who showed any pathology of the clavicle (e.g., fracture, infection, surgical fixation, or tumor) were excluded. Additionally, patients with malignant tumors, developmental disorders, endocrine disease, or those taking drugs or treatment affecting skeletal development were also excluded. Sex, date of birth, and examination data were collected from the medical database. Age was determined by the following formula: chronological age = (date of examination − date of birth) /365.25, rounded to two decimal places. Figure 1 shows the data acquisition flowchart. This retrospective study was approved by the Ethics Committee of Sichuan University. And informed consent was waived because of the retrospective nature. All methods were carried out following the relevant guidelines and regulations.

The CT scans were acquired using different scanners, including the uCT780 (United Imaging Healthcare, Shanghai, China), the SIEMENS SOMATOM Definition (Siemens, Forchheim, German), the Phillips Brilliance (Phillips, Amsterdam, Netherlands), and the GE Medical System Revolution (GE Healthcare, Milwaukee, WI, USA). All CT scans had a slice thickness of 1.00mm, a window width of 2600Hu, and a window level of 800Hu. The CT slices containing MCE were extracted manually and stored.

According to the age distribution and the normal or variant MCE, male and female samples were randomly selected as normal training and test sets and variant training and test sets. An additional 53 independent patients with clavicle variation were also collected and incorporated into the variant test set.

Manual methods

Only normal MCEs without shape variation were analyzed in this part. After multiplanar recombination, the axial and coronal views of MCE were examined to evaluate the degree of MCE development using standard staging procedures (Fig. 2) introduced by Schmeling et al. [5] and Kellinghaus et al. [6]. Both sides of the MCE were analyzed. A subset of 50 samples was randomly selected to evaluate the intra- and inter-observer agreement by using the weighted Kappa analysis. For intra-observer agreement, the subset was reevaluated by the first forensic examiner after one month, who has 6 years of experience in forensic age estimation from radiographs. For inter-observer agreement, the subset was evaluated by the second forensic examiner, who has 4 years of experience in this field. The machine learning models were trained and tested based on the staging results of the first forensic examiner. The grading results were assigned with a value of 1–9 for grades. The grades of both sides were used as independent variables to develop the linear and SVM model for age estimation.
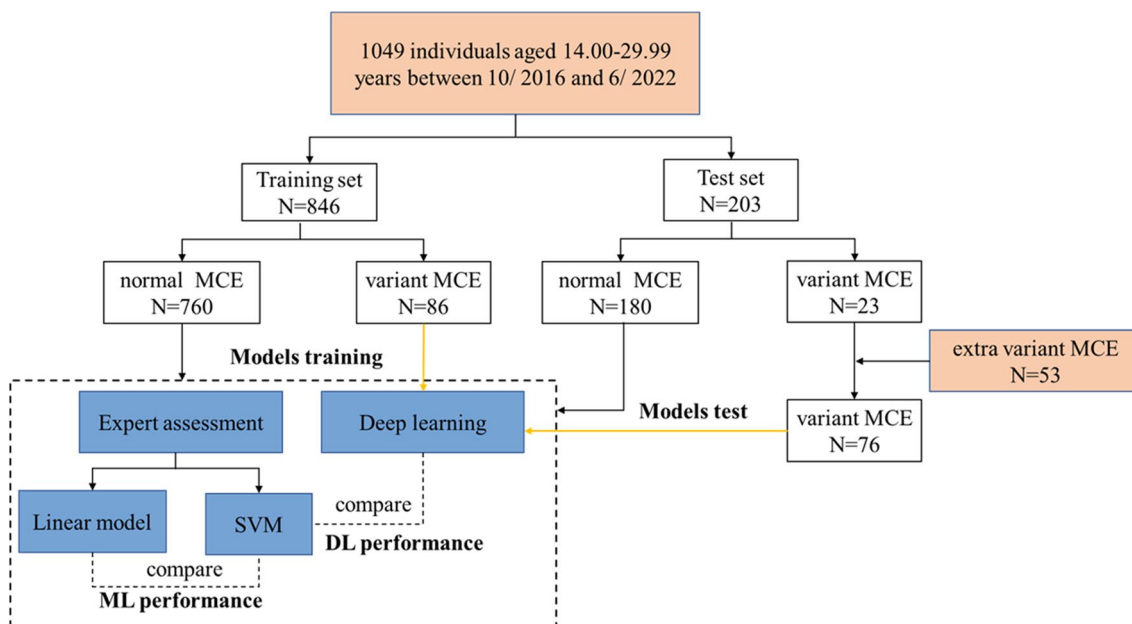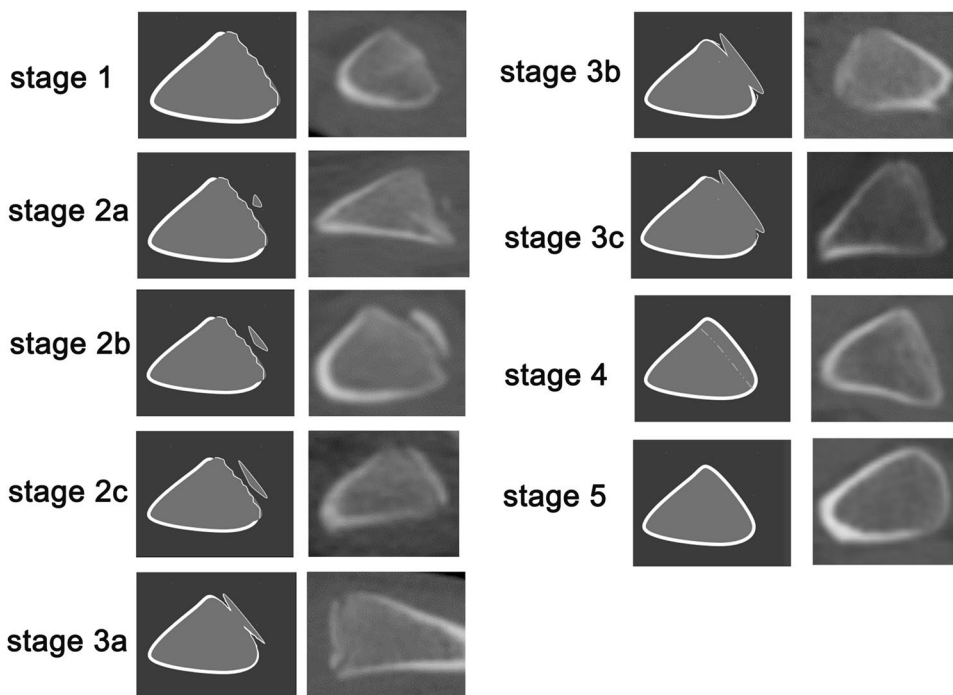
Fig. 1 Flowchart of data acquisition and division for the model training and test

**Fig. 2** Schematic drawings and CT images of the development stages of medial clavicle epiphysis

DL models training

Since the CT images were acquired from various device manufacturers, all images underwent intensity normalization to mitigate discrepancies. The CT slices containing MCE were subjected to filtering and standardized to 20 slices automatically by using the nearest neighbor interpolation algorithm, ensuring consistency in the number of input images for the network (Fig. 3a).

Since MCE is relatively small in chest CT scans, automatic segmentation of MCE is performed to enhance the learning ability of the network. A total of 600 CT slices from 30 samples were used to train a U-Net to segment MCE
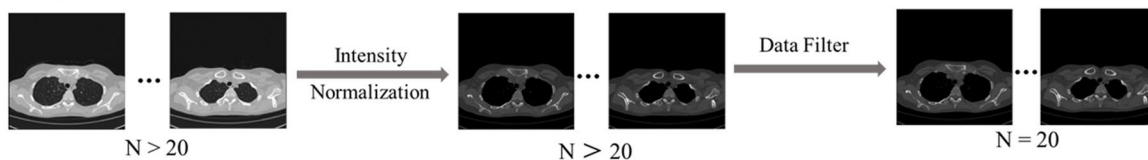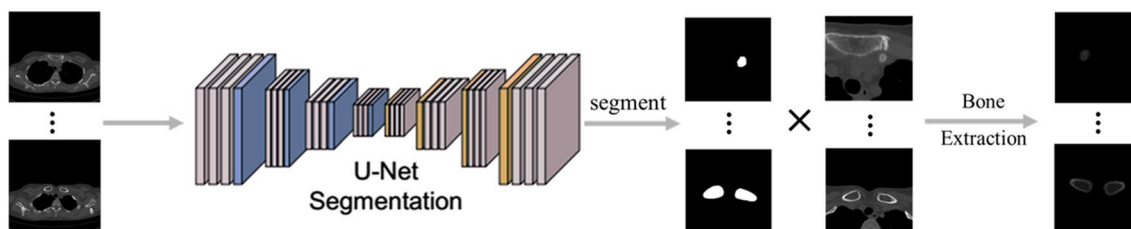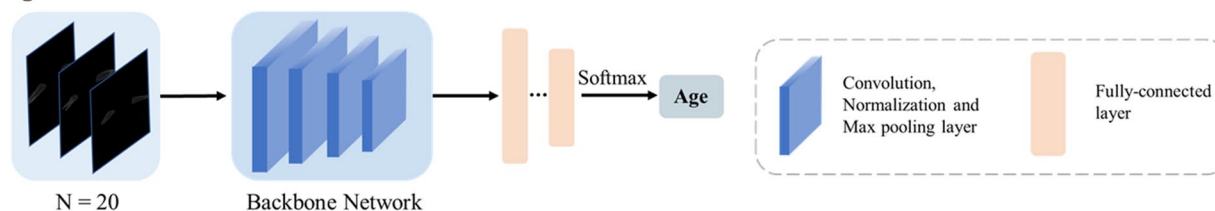
Fig. 3 Overview of the automatic age estimation framework

automatically (Fig. 3b). Finally, the image size was resized to 512×512.

In this study, we trained and compared six DL models, including five traditional CNN models (VGG-19, Res Net 18 and 34, SE Net 18, and SK Net), and a hybrid neural network CoAt Net [19]. The DL network training process is shown in Fig. 3c. Five-fold cross-validation was performed, and datasets were randomly split into training and validation sets in a 4:1 ratio for each fold. All the DL models were trained under the PyCharm platform using an Nvidia GeForce RTX 2080 8GB GPU. The networks are trained with a batch size of two and optimized with an AdamW optimizer, while the weight decay parameter is set to 0.3. The initial learning rate is set to 0.0005 and it is attenuated by a factor of 0.3 after every 30 training cycles for a total of 60 training cycles.

Statistical analysis

The performance of all methods was evaluated on the same independent test set. The best DL model was used for the patients with shape variation of MCE. The Pearson correlation coefficient (r) between estimated age and chronological age was analyzed. The mean absolute error

± standard deviation (MAE±SD), mean error (ME), and root mean square error (RMSE) were compared.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} | EA_i - CA_i | \tag{1}$$

$$ME = \frac{1}{n} \sum_{i=1}^{n} \left(EA_i - CA_i \right) \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(EA_i - CA_i \right)^2} \tag{3}$$

where EA means the estimated age, CA means the chronological age.

The accuracy, sensitivity, and specificity of age classification were evaluated for the age thresholds of 16, 18, 20, and 22 years (The important legal age limits in China's criminal law and Civil Code) and the age threshold of 21 years (this age threshold is internationally of exceptional relevance). The receiver operating characteristic (ROC) curve was constructed and the corresponding area under the curve (AUC) was obtained to evaluate the classification performance of models.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (4)$$

$$sensitivity = \frac{TP}{TP + FN} \qquad (5)$$

$$specificity = \frac{TN}{TN + FP} \qquad (6)$$

where *TP* represents true positive, which is the number of correctly classified samples as reaching or exceeding a specific legal age. *TN* represents true negative, which is the number of correctly classified samples as being below a specific legal age. *FP* represents false positive, which is the number of falsely classified samples as reaching or exceeding a specific legal age. *FN* represents false negative, which is the number of falsely classified samples as being below a specific legal age. Accuracy values reflect the classification accuracy in all samples. Sensitivity values reflect the classification accuracy in the individuals aged ≥X years. Specificity values reflect the classification accuracy in the individuals aged < X years. X means the age thresholds of 16, 18, 20, 21, and 22.

The observer agreements were analyzed using Cohen's kappa coefficients. The Wilcoxon test was employed to determine differences between left and right MCE. The sex differences were compared by the Mann-Whitney U test. The statistical analysis was performed using IBM SPSS, version 22.0, and GraphPad Prism, vision 8.0. Significance was set at the 5% level.

Results

Dataset characteristics

The detailed characteristics of patients are listed in Table 1. The mean age and standard deviation (mean ± SD) of the1049 patients was 22.50±4.34 years, comprising 500 males (mean ± SD: 22.37±4.24 years) and 549 females (mean ± SD: 22.62±4.43 years). Among them, 109 samples with clavicle variation were excluded from the manual analysis but included in the DL (mean ± SD: 22.81±4.61 years). The mean age of the additional 53 variant patients was 24.63±4.08 years.

Manual staging results

The manual method yielded a κ value of 0.879 (95% confidence interval: 0.762–0.975) for intra-observer agreement and 0.855 (95% confidence interval: 0.736–0.950) for inter-observer agreement, indicating good levels of consistency among observers. The side differences were observed in 135 out of the 940 normal MCE samples, and the Wilcoxon test revealed a significant difference in the development of bilateral MCE (p=0.006). The Mann-Whitney U test indicated no statistically significant differences between males and females, except at stage 4 (p=0.004) (Table 2).

Table 2 and Figure E1 show the age distribution of developmental stages by sex. The results revealed a significant positive association between the developmental stages of the

Table 1 Frequency distribution of training and test samples by age and sex

Age(year)	Normal Training Set		Normal Test Set		Variant Training Set		Variant Test Set		Total
	Female	Male	Female	Male	Female	Male	Female	Male	
14.00~14.99	17	14	5	3	1	5	2	0	47
15.00~15.99	18	12	6	3	1	3	1	0	44
16.00~16.99	19	19	2	6	0	6	2	1	55
17.00~17.99	18	22	8	3	0	4	0	0	55
18.00~18.99	27	14	8	1	1	4	2	1	58
19.00~19.99	22	25	5	7	2	2	2	1	66
20.00~20.99	34	27	6	9	1	4	2	2	85
21.00~21.99	38	35	9	7	0	0	3	3	95
22.00~22.99	22	21	7	4	0	6	0	3	63
23.00~23.99	32	29	9	3	2	6	3	2	86
24.00~24.99	24	27	9	3	5	3	2	2	75
25.00~25.99	26	19	2	6	2	3	9	5	72
26.00~26.99	26	29	7	6	2	2	3	5	80
27.00~27.99	28	25	7	6	0	6	3	3	78
28.00~28.99	32	19	7	4	0	3	3	3	71
29.00~29.99	25	15	10	2	7	5	6	2	72
total	408	352	107	73	24	62	43	33	1102

Table 2 Descriptive statistics for age by sex of all ossification stages expressed in years

Stage	Female			Male			p value*
	n	Mean±SD	Min-Max	n	Mean±SD	Min-Max	
1	43	15.74±1.41	14.09-18.76	46	15.89±1.20	14.12-18.89	0.443
2a	27	16.85±1.55	14.43-19.79	25	17.26±1.47	14.50-19.87	0.453
2b	26	17.61±1.91	14.44-21.58	17	18.05±1.37	16.22-20.14	0.274
2c	42	19.00±2.10	15.27-22.72	14	18.71±1.25	16.99-20.75	0.677
3a	55	20.71±2.09	16.70-25.81	40	20.04±1.40	16.54-22.35	0.168
3b	36	21.09±2.07	17.78-26.50	36	20.97±1.49	17.83-24.49	0.901
3c	114	23.37±2.21	18.41-29.88	104	23.44±2.27	18.24-29.04	0.777
4	163	27.10±2.05	20.14-29.93	142	26.41±2.16	19.17-29.88	0.004
5	8	28.26±1.05	26.65-29.66	2	26.31±1.97	24.92-27.70	0.178

*Sex differences analyzed by the Mann-Whitney U test

Table 3 The performance of different models for age estimation in the normal test set
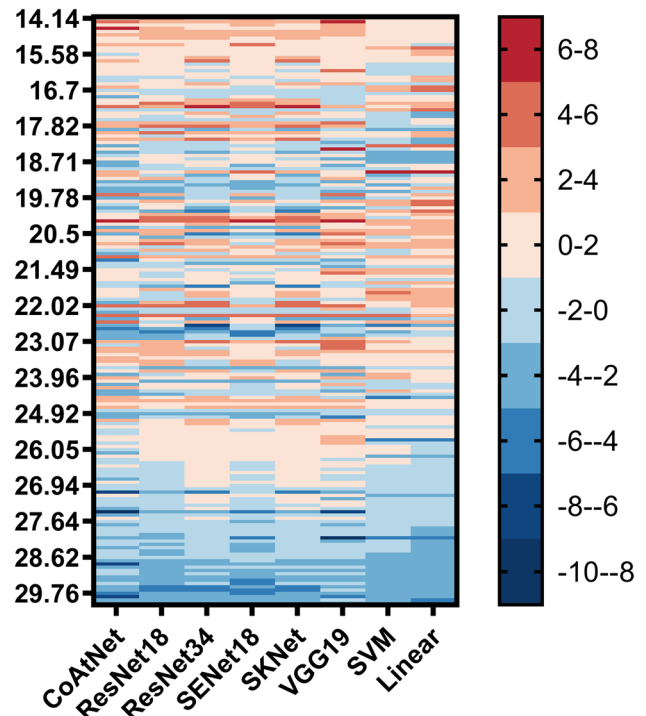
Models	r	MAE±SD	RMSE	ME	%AE<2 years
CoAt Net	0.820	2.01±1.63	2.59	-0.20	100/180
Res Net 18	0.858	1.85±1.32	2.27	0.03	106/180
Res Net 34	0.846	1.93±1.45	2.41	0.20	110/180
SE Net 18	0.862	1.77±1.39	2.25	-0.01	118/180
SK Net	0.834	1.93±1.59	2.50	0.20	107/180
VGG-19	0.864	1.83±1.41	2.31	0.28	111/180
SVM	0.874	1.73±1.28	2.15	0.03	117/180
Linear	0.849	1.94±1.31	2.34	0.08	101/180

Abbreviations: %AE<2 years, the percentage of absolute error within 2 years between the estimated age and chronological age; *MAE*, mean absolute error; *ME*, mean error; *RMSE*, root mean square error; *SVM*, support vector machine; *Linear*, linear model

Fig. 4 The heatmap depicting estimation errors of all eight models in the normal test set. The color of each cell represents the error of a single sample (Y-axis) predicted by a specific model (X-axis). The red indicates overestimation and the blue indicates underestimation. The higher error of estimation is associated with darker colorings

MCE and chronological age. The youngest age of individuals with stage 3c and above was older than 18 years old in both sexes, indicating that 3c and above can aid in determining the 18-year threshold.

Estimation performance of deep learning versus manual method

Based on manual stages, the traditional linear model was constructed for age estimation (age= =12.958+0.834×right stage+0.768×left stage). An SVM model (kernel='poly', C=1, degree=2) was also built for age estimation with the hyper-parameters optimized through grid search. The performance of both SVM and DL models for age estimation was evaluated on a normal test set, as shown in Table 3. Results indicate that the SVM model, based on manual feature extraction, can improve accuracy compared to traditional linear models. Among all DL models, SE Net 18 demonstrated the highest accuracy and was most similar to that of the SVM model. The attention heatmaps generated via Grad-CAM were used to visualize the region of interest

(ROI) of the trained SE Net 18 (Figure E2). The heatmaps of SE Net 18 revealed that the network exhibited attention not only towards the medial clavicle end but also demonstrated a higher focus on the cortical bone region of the clavicle.

The heatmap of estimation errors (Fig. 4) indicates that the phenomenon of 'attraction of the middle' was not prominent in this study. Only a few age groups at either end exhibited obvious overestimation or underestimation, with overestimation primarily observed among those aged 27 years or

older. Additionally, it is visually apparent from the heatmap that most samples had an estimation error within ±2 years.

Classification performance of deep learning versus manual method

Table 4 and Figure 5 demonstrate that clavicle CT has limited value in the classification of the 16-year threshold due to its poor specificity in this study. This lower specificity may be caused by the limited samples under 16 years old in this study. While, it exhibits superior performance in the classification of 18-, 20-, 21-, and 22-year threshold. A comparable performance of age classification was noticed in all models. The confusion matrix of age classification in different age groups is shown in Figure E3 and the ROC curves are shown in Figure E4.

Performance of deep learning in variation test

Since the manual method was incapable of analyzing variant clavicles, DL models were trained with such variants to evaluate their ability on variant clavicles (Fig. 6). The models were also tested on an external variation test including 76 samples with variant clavicles. The results obtained from SE Net 18 showed that r was 0.751, MAE was 2.08 ± 1.62 years, RMSE was 2.63 years, and ME was -0.33 years in age estimation. Although the accuracy was lower than that of normal samples, it also indicated that DL models could be used to analyze variant samples. The performance of age classification is shown in Table 5 and Figure E5. Despite the limited number and uneven distribution of the variant clavicles, the age classification accuracy in those samples remains relatively high.

Discussion

The medial clavicle epiphysis is a crucial indicator for age estimation in late adolescence and early adulthood, particularly in the determination of the 18-year threshold. Consistent with previous studies [3, 7, 9, 10], this study confirms that stage 3c can be used to determine the 18-year threshold. Additionally, we employed ML and DL techniques to establish the age estimation models which have a superior performance compared to traditional linear model.

The staging methods proposed by Schmeling et al. [5] and Kellinghaus et al. [6] were commonly used for age estimation from MCE. Thin-slice CT of 1mm or less can more clearly observe the epiphyseal line and avoid visual deception by the partial volume effect [20]. Consistent with previous studies, this study found that stage 3c and above may serve as a reliable indicator for determining the 18-year threshold (Table E1) [1, 3, 6, 7, 9, 10]. While

Table 4 The performance of different models for age classification in the normal test set

Models	16-year threshold				18-year threshold				20-year threshold				21-year threshold				22-year threshold			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
CoAt Net	0.906	0.963	0.353	.911	0.856	0.903	0.667	.921	0.844	0.862	0.807	.936	0.850	0.852	0.847	.921	0.833	0.826	0.841	.925
Res Net 18	0.906	1.000	0.000	.918	0.894	0.910	0.833	.923	0.894	0.935	0.807	.957	0.822	0.861	0.764	.923	0.856	0.902	0.807	.939
Res Net 34	0.900	0.963	0.294	.902	0.872	0.910	0.722	.927	0.878	0.902	0.825	.933	0.844	0.889	0.778	.927	0.822	0.891	0.750	.936
SE Net 18	0.906	0.988	0.118	.918	0.900	0.938	0.750	.940	0.867	0.911	0.772	.942	0.856	0.926	0.750	.940	0.861	0.891	0.830	.952
SK Net	0.911	1.000	0.059	.897	0.894	0.924	0.778	.916	0.889	0.919	0.825	.954	0.822	0.880	0.736	.916	0.828	0.880	0.773	.927
VGG-19	0.894	0.926	0.588	.931	0.872	0.910	0.722	.936	0.900	0.943	0.807	.958	0.867	0.944	0.750	.936	0.856	0.946	0.761	.949
SVM	0.928	0.963	0.588	.953	0.906	0.931	0.806	.937	0.911	0.902	0.930	.964	0.872	0.926	0.792	.937	0.839	0.902	0.773	.928
Linear	0.928	0.963	0.588	.959	0.894	0.944	0.694	.938	0.911	0.992	0.737	.965	0.856	0.935	0.736	.938	0.800	0.946	0.648	.929

Abbreviations: *ACC*, accuracy; *AUC*, area under the ROC curve; *SEN*, sensitivity; *SPE*, specificity; *SVM*, support vector machine; *Linear*, linear model

[a] Sensitivity values are equal to the classification accuracy in the individuals ≥ X years.

[b] Specificity values are equal to the classification accuracy in the individuals < X years.

X represents the legal age limits of 16, 18, 20, 21, or 22, respectively.
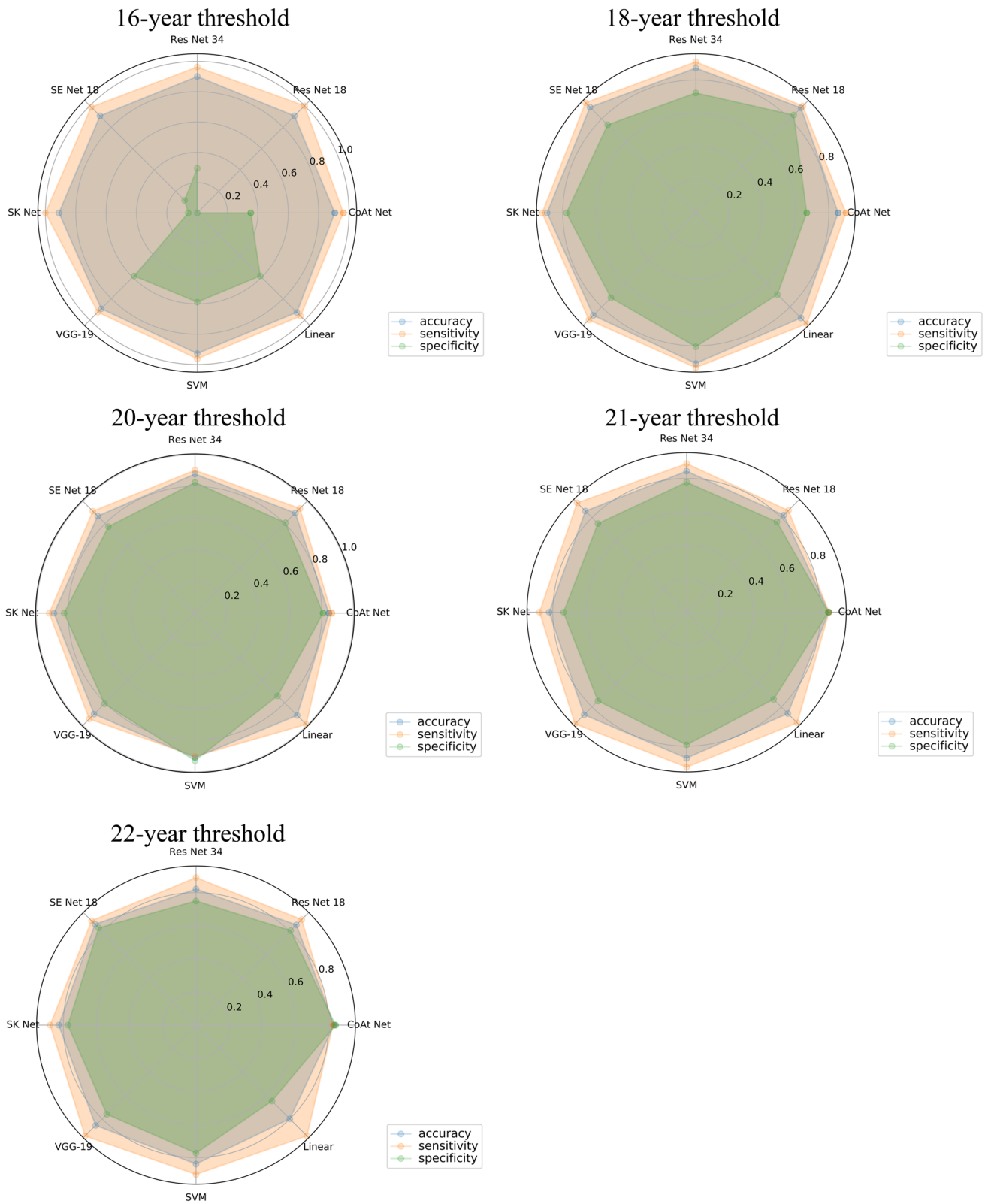
Fig. 5 Radar chart of classification accuracy of all the eight models in the normal test set. Each axis arranged radially represents a model. Orange represents sensitivity. Blue represents accuracy. Green represents specificity

Fig. 6 Examples of variant clavicles and their estimation error. **a–c** Medial metaphyseal concavities; **d–e**. depressed metaphyseal; **f** Epiphyseal cartilage was ossified before the ossification center was completely ossified. CA, chronological age; EA, estimated age
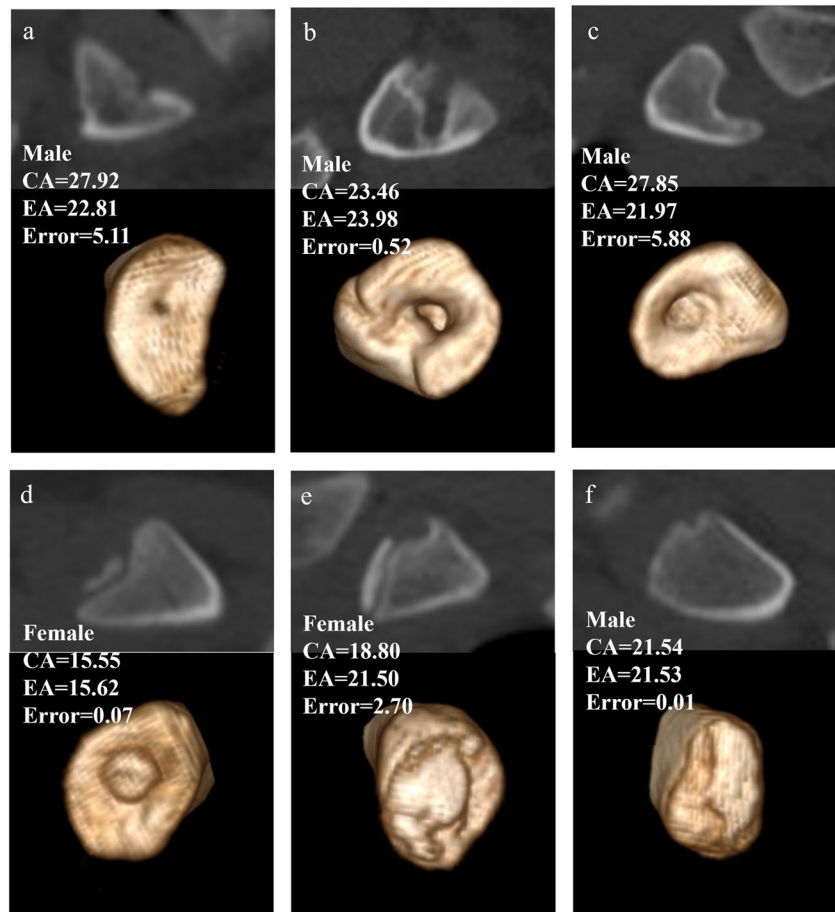
Table 5 The results of age classification on the 76 variant test samples

Age thresholds	ACC	SEN	SPE	AUC
16-year threshold	0.970 (74/76)	1.000 (73/73)	0.330 (1/3)	.941
18-year threshold	0.921 (70/76)	0.960 (67/70)	0.500 (3/6)	.938
20-year threshold	0816 (62/76)	0.910 (58/64)	0.420 (5/12)	.882
21-year threshold	0.855(65/76)	0.920 (5/60)	0.625 (6/16)	.900
22-year threshold	0.882 (67/76)	0.930 (50/54)	0.770 (17/22)	.887

Abbreviations: *ACC*, accuracy; *AUC*, the area under the ROC curve; *SEN*, sensitivity; *SPE*, specificity

[a]Sensitivity values are equal to the classification accuracy in the individuals ≥ X years

[b]Specificity values are equal to the classification accuracy in the individuals < X years

X represent the legal age limits of 16, 18, 20, 21, or 22, respectively

previous research has indicated that stage 5 typically emerges around age 25 in both sexes [7, 10, 20, 21]. Similarly, we found that the earliest observation of stage 5 was at age 26 in females and age 24 in males. However, this study only found 10 samples exhibiting stage 5, indicating that the epiphyseal line may disappear after the age of 30 when using thin-slice CT imaging.

Previous studies have generally found nonsignificant differences between the two sides of MCE [9, 21–23], but our study revealed a significant difference as determined by the Wilcoxon test. This discrepancy may be attributed to the shape variation of MCE, ethnicity variations, and the quantitative measures of substages. The difference between males and females is statistically insignificant except for stage 4 in this study, which is consistent with previous studies [1, 6, 7, 20] that most of the stages exhibit no sex differences. Conversely, some studies have identified significant differences between the sexes. Since the insignificant difference between sexes and significant difference in bilateral stages, the bilateral stages were used

for regression analysis. Males and females were analyzed together to ensure an adequate sample size for DL analysis. De Tobel et al. [24] also suggested incorporating both clavicles in the age estimation model.

The statistical analysis and the presentation of the conclusion are fundamental parts of age estimation [25]. Previous studies have used general descriptive analysis to estimate age and determine the 18-year threshold. However, this method is susceptible to mimicking the age structure of the reference population [11]. Shedge et al. [7] trained a linear regression model for age estimation from MCE on 350 samples and tested it on 50 samples with an MAE of 1.32 years. Due to variations in samples and CT machines across different studies, it is impossible to fairly compare the accuracy of age estimation. In this study, SVM outperforms the linear model (p= 0.0004) with an MAE of 1.73 years in the same dataset. The lower performance of the linear model may be attributed to its unsuitability for ordinal variables and its inability to accurately depict the relationship between age and epiphyseal development [11]. The possible reasons for this slightly lower accuracy compared to Shedge et al. include: The sample size in this study is nearly three times that of Shedge et al.'s [7], allowing for a more comprehensive analysis of MCE development but introducing more uncertain factors; a 1.0 mm CT was utilized which may be inadequate in displaying MCE compared to the 0.6 mm used by Shedge et al.

Although SVM can improve the accuracy of age estimation, it should be noted that traditional ML models are still based on manual staging which is subjective, time-consuming, and complex. Deep learning can automatically extract and integrate multiple age-related features from multi-slice clavicle CT scans. Deep learning has been applied for age estimation for years and achieved good results [26–30]. Previous automated bone age estimation mainly focused on the wrist, knee, and pelvis. Due to the relatively small size of the clavicle in chest CT, automatic age estimation from MCE requires segmentation to enable the DL model to focus on MCE and disregard extraneous noise. The sole investigation of clavicle CT involved the automated localization of medial clavicular epiphysis (MCE), achieving an accuracy rate of 88% in the test set [16]. Differing from solely automated localization, our study implemented automated segmentation and age estimation.

Both traditional CNNs and a hybrid model CoAt Net were chosen as the backbones for automated age estimation. The results show that the traditional CNN model outperforms CoAt Net. Regrettably, CoAt Net did not achieve satisfactory evaluation performance as previous studies suggested [19, 31]. One possible reason is that the large capacity of CoAt Net requires a substantial image to capture global relationships. However, this study only

had slightly over a thousand samples, which may not fully optimize the model parameters and thus fail to train the model effectively. In contrast, CNNs have relatively modest data size demands and concentrate on local features, making them more appropriate for age estimation from medical images. This study is the first application of DL techniques in achieving end-to-end age estimation from clavicle CT scans, with accuracy levels comparable to those obtained through traditional ML methods. And the attention heatmaps revealed a higher focus on the cortical bone region of the clavicle in this study. In the future, further exploration can be performed for deep learning-based age estimation using either MCE or cortical bone alone to assess their respective contributions in this process.

Variant MCE has been observed in many studies, with approximately 10–20% of samples exhibiting non-assessable shape variants [20, 21, 32–35]. In this study, we have identified 109 samples of variant MCE (10.39%) that cannot be analyzed using the current staging method. This normal morphological variation poses a bottleneck in accurately evaluating MCE for age estimation using the traditional manual staging method. The morphological variants were excluded in the age analysis in previous studies [24]. In this study, automated age estimation was also applied to the variant clavicles and demonstrated the effective handling of DL for such cases, albeit with a slightly lower accuracy compared to normal samples.

There are limitations in this study. First, the age distribution is unbalanced due to its retrospective nature. Future large-scale studies with different ethnicities are still needed. And more advanced methods such as EasyEnsemble or deep learning techniques are expected to be applied for data balancing in the future. Additionally, it is recommended to gather more diverse MCE data for training DL networks to improve their ability in variable samples. Second, CT was used to represent the MCE in this study, which has a health risk from ionizing radiation exposure. In recent years, some studies have been carried out on age estimation from clavicular MRI [34, 36, 37]. This automated analysis of clavicle CT is also expected to be transferred to clavicle MRI in the future. Third, this study represents the first attempt at automated age estimation from clavicular CT scans. To achieve accurate results, a comprehensive approach that incorporates multiple joints is necessary. Previous studies have combined MRI of the MCE, wrist, and teeth to estimate age and have found that using all three joints produces better results than relying on a single joint [38].

In conclusion, this study objectively evaluated and compared the performance of the linear model, ML, and DL for age estimation. The results demonstrate a significant improvement in accuracy using ML and DL. This study can be a robust computer-assisted tool to aid the experts in age estimation.

Declarations

References

1. Pattamapaspong N, Madla C, Mekjaidee K, Namwongprom S (2015) Age estimation of a Thai population based on maturation of the medial clavicular epiphysis using computed tomography. Forensic Sci Int 246(123):e1–e5. https://doi.org/10.1016/j.forsciint.2014.10.044

2. Houpert T, Rérolle C, Savall F, Telmon N, Saint-Martin P (2016) Is a CT-scan of the medial clavicle epiphysis a good exam to attest to the 18-year threshold in forensic age estimation? Forensic Sci Int 260:103.e1-.e3. https://doi.org/10.1016/j.forsciint.2015.12.007

3. Torimitsu S, Makino Y, Saitoh H et al (2019) Age estimation based on maturation of the medial clavicular epiphysis in a Japanese population using multidetector computed tomography. Leg Med (Tokyo) 37:28–32. https://doi.org/10.1016/j.legalmed.2018.12.003

4. Wittschieber D, Ottow C, Vieth V et al (2015) Projection radiography of the clavicle: still recommendable for forensic age diagnostics in living individuals? Int J Legal Med 129:187–193. https://doi.org/10.1007/s00414-014-1067-0

5. Schmeling A, Schulz R, Reisinger W, Muhler M, Wernecke KD, Geserick G (2004) Studies on the time frame for ossification of the medial clavicular epiphyseal cartilage in conventional radiography. Int J Legal Med 118:5–8. https://doi.org/10.1007/s00414-003-0404-5

6. Kellinghaus M, Schulz R, Vieth V, Schmidt S, Pfeiffer H, Schmeling A (2010) Enhanced possibilities to make statements on the ossification status of the medial clavicular epiphysis using an amplified staging scheme in evaluating thin-slice CT scans. Int J Legal Med 124:321–325. https://doi.org/10.1007/s00414-010-0448-2

7. Shedge R, Kanchan T, Garg PK et al (2020) Computed tomographic analysis of medial clavicular epiphyseal fusion for age estimation in Indian population. Leg Med (Tokyo) 46:101735. https://doi.org/10.1016/j.legalmed.2020.101735

8. De Tobel J, Bauwens J, Parmentier GIL et al (2020) Magnetic resonance imaging for forensic age estimation in living children and young adults: a systematic review. Pediatr Radiol 50:1691–1708. https://doi.org/10.1007/s00247-020-04709-x

9. Ekizoglu O, Hocaoglu E, Inci E, Can IO, Aksoy S, Sayin I (2015) Estimation of forensic age using substages of ossification of the medial clavicle in living individuals. Int J Legal Med 129:1259–1264. https://doi.org/10.1007/s00414-015-1234-y

10. Gurses MS, Inanir NT, Gokalp G, Fedakar R, Tobcu E, Ocakoglu G (2016) Evaluation of age estimation in forensic medicine by examination of medial clavicular ossification from thin-slice computed tomography images. Int J Legal Med 130:1343–1352. https://doi.org/10.1007/s00414-016-1408-2

11. Fan F, Dong X, Wu X et al (2020) An evaluation of statistical models for age estimation and the assessment of the 18-year threshold using conventional pelvic radiographs. Forensic Sci Int 314:110350. https://doi.org/10.1016/j.forsciint.2020.110350

12. Peng J, Han H, Yi Y, Huang H, Xie L (2022) Machine learning and deep learning modeling and simulation for predicting PM2.5 concentrations. Chemosphere 308:136353. https://doi.org/10.1016/j.chemosphere.2022.136353

13. Fan F, Tu M, Li R et al (2020) Age estimation by multidetector computed tomography of cranial sutures in Chinese male adults. Am J Phys Anthropol 171:550–558. https://doi.org/10.1002/ajpa.23998

14. Xu C, Qu H, Wang G et al (2015) A novel strategy for forensic age prediction by DNA methylation and support vector regression model. Sci Rep 5:17788. https://doi.org/10.1038/srep17788

15. Shen S, Liu Z, Wang J, Fan L, Ji F, Tao J (2021) Machine learning assisted Cameriere method for dental age estimation. BMC Oral Health 21:641. https://doi.org/10.1186/s12903-021-01996-0

16. Wesp P, Sabel BO, Mittermeier A et al (2023) Automated localization of the medial clavicular epiphyseal cartilages using an object detection network: a step towards deep learning-based forensic age assessment. Int J Legal Med. https://doi.org/10.1007/s00414-023-02958-7

17. Kengkard P, Choovuthayakorn J, Mahakkanukrauh C et al (2023) Convolutional neural network of age-related trends digital radiographs of medial clavicle in a Thai population: a preliminary study. Anat Cell Biol 56:86–93. https://doi.org/10.5115/acb.22.205

18. Tozakidou M, Meister RL, Well L et al (2021) CT of the medial clavicular epiphysis for forensic age estimation: hands up? Int J Legal Med 135:1581–1587. https://doi.org/10.1007/s00414-021-02516-z

19. Dai Z, Liu H, Le QV, Tan M. (2021) CoAtNet: Marrying Convolution and Attention for All Data Sizes. pp. arXiv:2106.04803.

20. Kellinghaus M, Schulz R, Vieth V, Schmidt S, Schmeling A (2010) Forensic age estimation in living subjects based on the ossification status of the medial clavicular epiphysis as revealed by thin-slice multidetector computed tomography. Int J Legal Med 124:149–154. https://doi.org/10.1007/s00414-009-0398-8

21. Wittschieber D, Schulz R, Vieth V et al (2014) The value of substages and thin slices for the assessment of the medial clavicular epiphysis: a prospective multi-center CT study. Forensic Sci Med Pathol 10:163–169. https://doi.org/10.1007/s12024-013-9511-x

22. Franklin D, Flavel A (2015) CT evaluation of timing for ossification of the medial clavicular epiphysis in a contemporary Western Australian population. Int J Legal Med 129:583–594. https://doi.org/10.1007/s00414-014-1116-8

23. Ekizoglu O, Hocaoglu E, Inci E et al (2015) Forensic age estimation by the Schmeling method: computed tomography analysis of the medial clavicular epiphysis. Int J Legal Med 129:203–210. https://doi.org/10.1007/s00414-014-1121-y

24. De Tobel J, Hillewig E, van Wijk M et al (2020) Staging clavicular development on MRI: pitfalls and suggestions for age estimation. J Magn Reson Imaging 51:377–388. https://doi.org/10.1002/jmri.26889

25. Mostad P, Schmeling A, Tamsen F (2022) Mathematically optimal decisions in forensic age assessment. Int J Legal Med 136:765–776. https://doi.org/10.1007/s00414-021-02749-y

26. Larson DB, Chen MC, Lungren MP, Halabi SS, Stence NV, Langlotz CP (2018) Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. Radiology 287:313–322. https://doi.org/10.1148/radiol.2017170236

27. Lee H, Tajmir S, Lee J et al (2017) Fully automated deep learning system for bone age assessment. J Digit Imaging 30:427–441. https://doi.org/10.1007/s10278-017-9955-8

28. Spampinato C, Palazzo S, Giordano D, Aldinucci M, Leonardi R (2017) Deep learning for automated skeletal bone age assessment in X-ray images. Med Image Anal 36:41–51. https://doi.org/10.1016/j.media.2016.10.010

29. Fan F, Ke W, Dai X et al (2023) Semi-supervised automatic dental age and sex estimation using a hybrid transformer model. Int J Legal Med 137:721–731. https://doi.org/10.1007/s00414-023-02956-9

30. Lu T, Diao Y-r, Tang X-e et al (2023) Deep learning enables automatic adult age estimation based on CT reconstruction images of the costal cartilage. Eur Radiol. https://doi.org/10.1007/s00330-023-09761-3

31. Chen Y, Wan Y, Pan F (2023) Enhancing multi-disease diagnosis of chest X-rays with advanced deep-learning networks in real-world data. J Digit Imaging:1–16. https://doi.org/10.1007/s10278-023-00801-4

32. Rudolf E, Kramer J, Schmidt S, Vieth V, Winkler I, Schmeling A (2019) Anatomic shape variants of extremitas sternalis claviculae as collected from sternoclavicular thin-slice CT-studies of 2820 male borderline-adults. Int J Legal Med 133:1517–1528. https://doi.org/10.1007/s00414-019-02065-6

33. Scharte P, Vieth V, Schulz R et al (2017) Comparison of imaging planes during CT-based evaluation of clavicular ossification: a multi-center study. Int J Legal Med 131:1391–1397. https://doi.org/10.1007/s00414-017-1615-5

34. Vieth V, Schulz R, Brinkmeier P, Dvorak J, Schmeling A (2014) Age estimation in U-20 football players using 3.0 tesla MRI of the clavicle. Forensic Sci Int 241:118–122. https://doi.org/10.1016/j.forsciint.2014.05.008

35. Schmidt S, Ottow C, Pfeiffer H et al (2017) Magnetic resonance imaging-based evaluation of ossification of the medial clavicular epiphysis in forensic age assessment. Int J Legal Med 131:1665–1673. https://doi.org/10.1007/s00414-017-1676-5

36. Hillewig E, Degroote J, Van der Paelt T et al (2013) Magnetic resonance imaging of the sternal extremity of the clavicle in forensic age estimation: towards more sound age estimates. Int J Legal Med 127:677–689. https://doi.org/10.1007/s00414-012-0798-z

37. Widek T, De Tobel J, Ehammer T, Genet P (2023) Forensic age estimation in males by MRI based on the medial epiphysis of the clavicle. Int J Legal Med 137:679–689. https://doi.org/10.1007/s00414-022-02924-9

38. Štern D, Payer C, Giuliani N, Urschler M (2019) Automatic age estimation and majority age classification from multi-factorial MRI data. IEEE J Biomed Health Inform 23:1392–1403. https://doi.org/10.1109/JBHI.2018.2869606