Kuen Wai MA, Jiajun ZHU, Hai Ming WONG, 2026. A framework to integrate deep learning and an ethnicity-specific reference data set for reliable dental age estimation from digital panoramic radiographs. *ENGINEERING Information Technology & Electronic Engineering,* 26(8):260013.
https://doi.org/10.1631/ENG.ITEE.2026.0013

A framework to integrate deep learning and an ethnicity-specific reference data set for reliable dental age estimation from digital panoramic radiographs

Key words: Deep neural networks (DNNs); Dental age estimation; Panoramic radiographs; Southern Chinese; Tooth development

Corresponding author: Jiajun ZHU
E-mail: zhujj@zju.edu.cn

Corresponding author: Hai Ming WONG
E-mail: wonghmg@hku.hk
ORCID: https://orcid.org/0000-0003-3411-6442

Motivation

1. Dental age estimation is a critical component across various sectors, including clinical practice, forensic investigations, and pediatric dentistry.
2. Tooth development stage (TDS) classifications are one of the conventional approaches in dental age estimation, determining an individual's chronological age based on the development of his/her teeth.
3. However, TDS classifications are often labor-intensive, relying on manual annotations by trained experts, and several studies have reported that Demirjian's dental age assessment method is less accurate when applied to other ethnic groups.

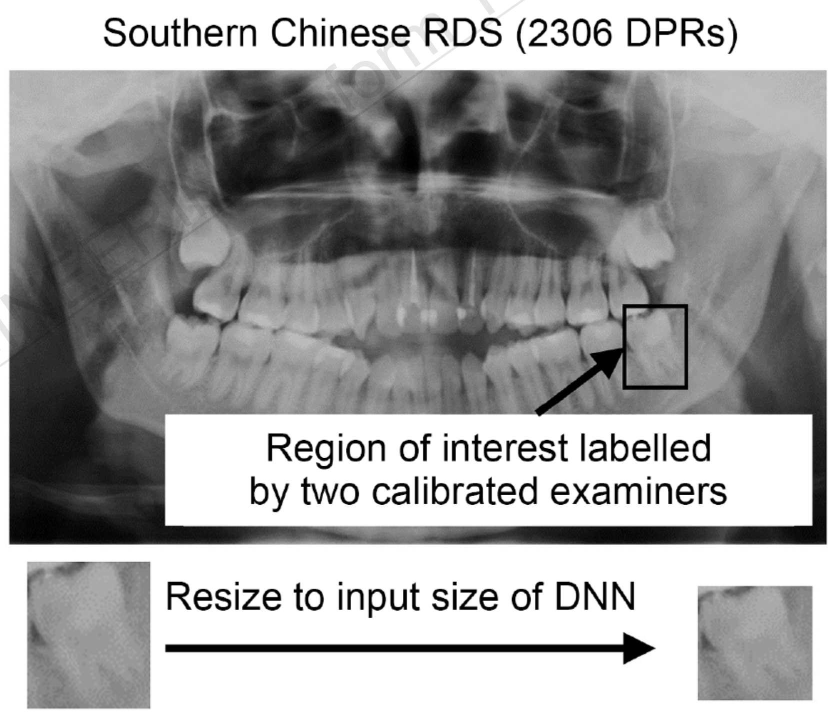
Main idea

1. Leveraging the power of artificial intelligence, applications of deep neural networks (DNNs) have demonstrated remarkable capabilities in image recognition, natural language processing, and medical imaging analysis.
2. Despite significant advancements in automated dental age estimation facilitated by DNN applications, challenges remain in achieving precise and reliable results across diverse populations.
3. By combining this ethnicity-specific reference data set (RDS) with state-of-the-art DNN technology, this study aims to develop an accurate and reliable framework (DNN-IMDA) for automated estimation of Demirjian's TDSs on digital panoramic radiographs (DPRs), thereby enabling reliable dental age estimation for southern Chinese individuals.

Method

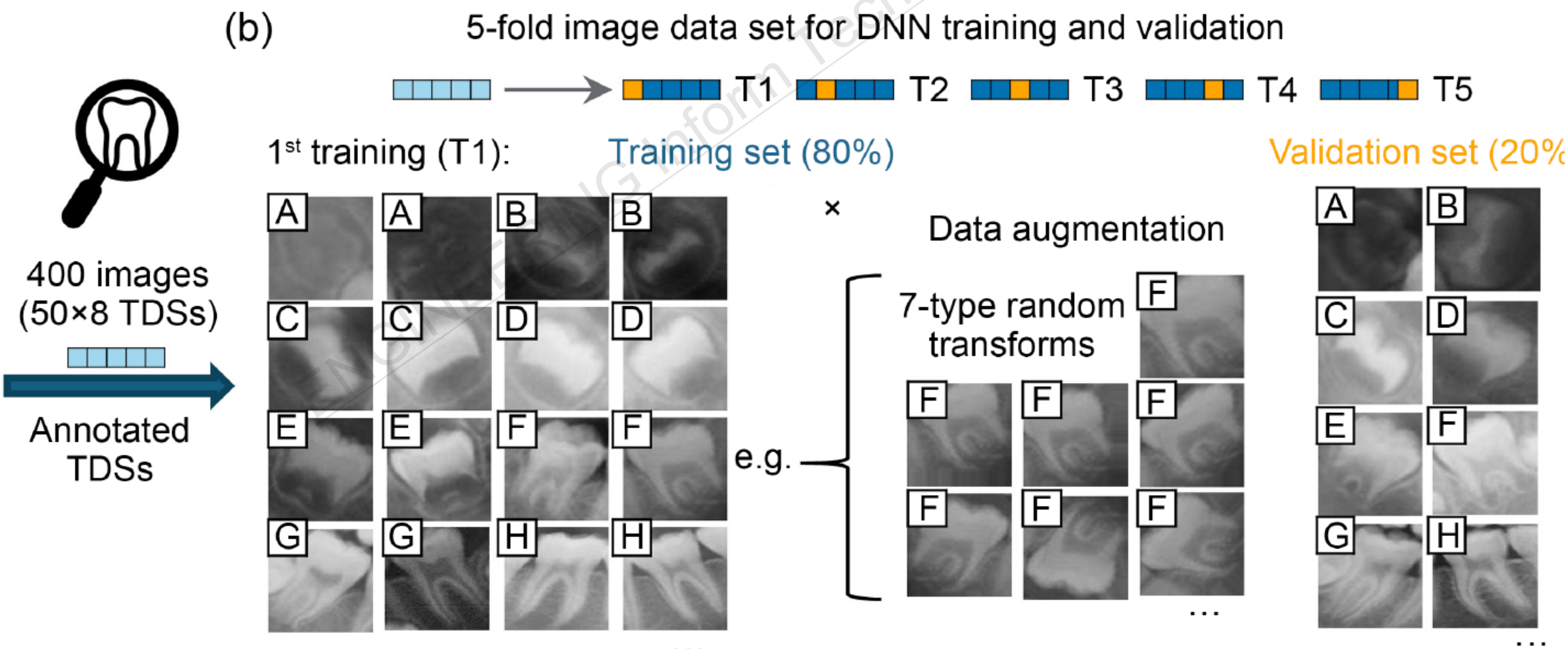
1. Southern Chinese RDS is developed from 2306 DPRs collected from the Department of Paediatric Dentistry & Orthodontics, Prince Philip Dental Hospital, Hong Kong, China, covering individuals aged 2–25 years with balanced gender distribution.

(a)



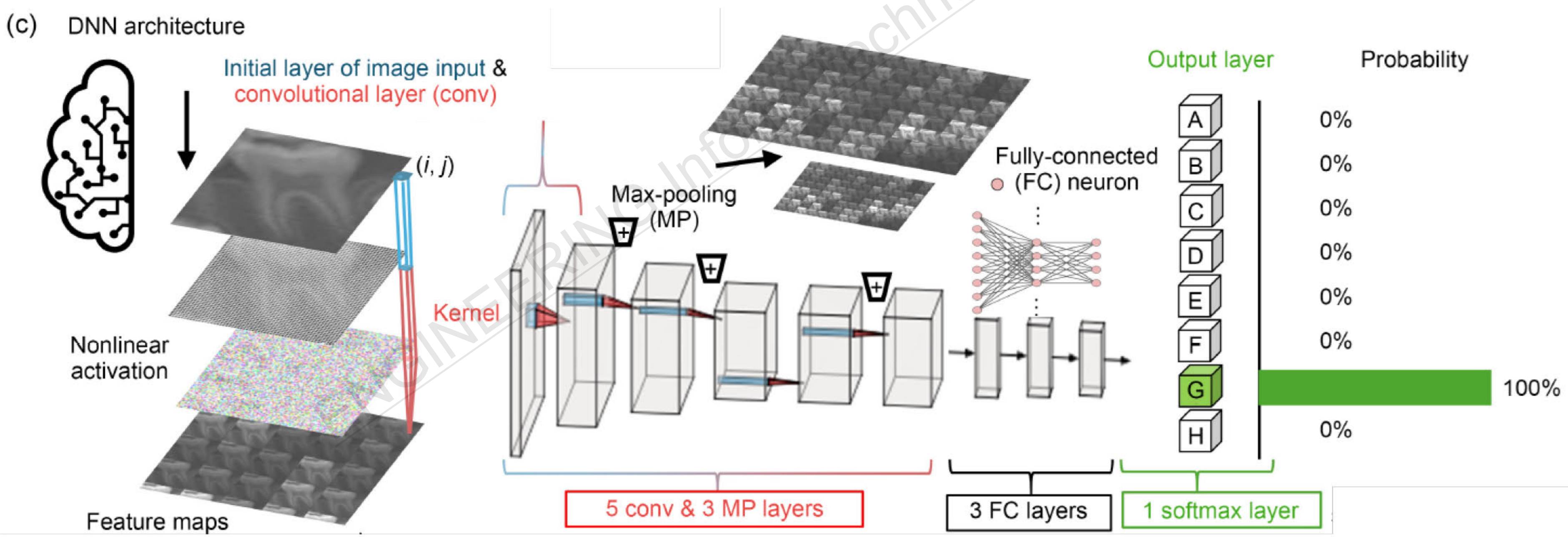
Method (Cont'd)

2. Regions of interest containing molars of the radiographs are manually labelled by two calibrated examiners. A total of 400 molar images (50×8 TDSs) with annotated TDS are randomly selected to construct an image data set for DNN training and validation.



Method (Cont'd)

3. Eight different DNN architectures are evaluated for their performance in TDS classification. The transfer learning approach involves extracting all layers except the final three classification layers.



Method (Cont'd)

Table 2 Summary of the eight architectures of DNNs applied in the study for tooth development stage classification

Architecture	Depth	Input size	Input value range	Number of parameters ($\times 10^6$)	Activation function	Pooling	Reference
AlexNet	8	227×227×3	[0, 255]	61.0	ReLU	Max-pooling	Krizhevsky et al. (2017)
DenseNet-201	201	224×224×3	[0, 255]	20.0	ReLU	Average pooling	Huang et al. (2017)
GoogLeNet	22	224×224×3	[0, 255]	7.0	ReLU	Average pooling	Szegedy et al. (2015)
Inception-v3	48	299×299×3	[0, 255]	23.9	ReLU	Average pooling	Szegedy et al. (2016)
ResNet-18	18	224×224×3	[0, 255]	11.7	ReLU	Max-pooling	He et al. (2016)
ResNet-50	50	224×224×3	[0, 255]	25.6	ReLU	Max-pooling	He et al. (2016)
VGG-16	16	224×224×3	[0, 255]	138	ReLU	Max-pooling	Simonyan and Zisserman (2015)
Xception	71	299×299×3	[0, 255]	22.9	ReLU	Average pooling	Chollet (2017)

Method (Cont'd)

4. The calculation of integrated mean dental age (IMDA): For each individual, the corresponding mean age and SE for each scored TDS are extracted from the southern Chinese RDS using the following systematic approach. Each TDS is weighed inversely proportional to the SE of its mean age:

$$w_i = \frac{1}{\mathrm{SE}_i^2}, \qquad (10)$$

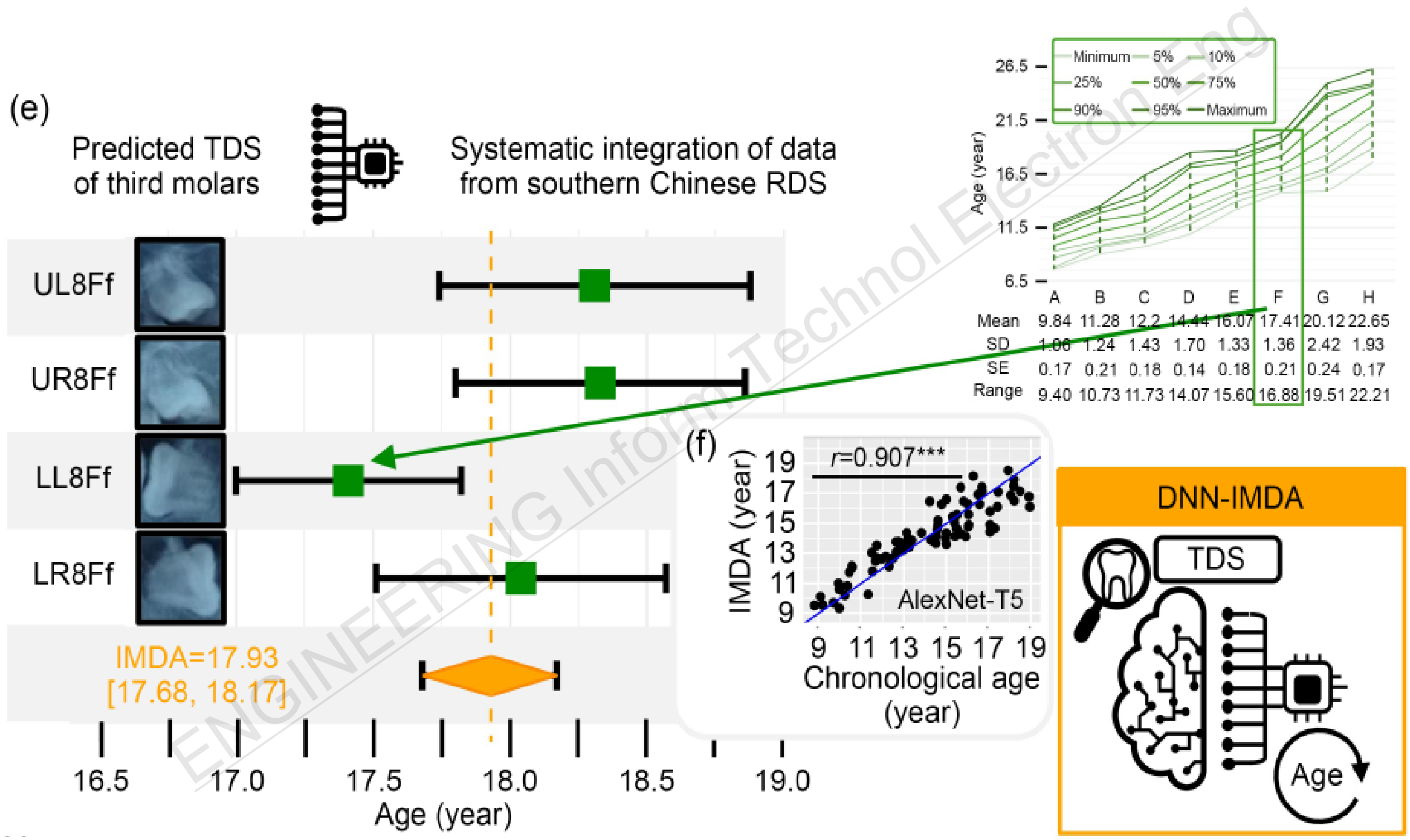
where w_i is the weight for tooth i and SE_i is the SE of the mean age for that specific TDS. IMDA is calculated by

$$\mathrm{IMDA} = \sum_{i=1}^{m} w_i \cdot \bar{x}_i \Big/ \sum_{i=1}^{m} w_i, \qquad (11)$$

where m is the number of developing teeth (stages A–G) present in the individual, and $\bar{x}_i$ is the mean age for TDS of tooth i from the southern Chinese RDS. The 95% CI for IMDA is calculated using the SE of the weighted mean:

$$\mathrm{IMDA} \pm 1.96 \cdot \sqrt{1 \Big/ \sum_{i=1}^{m} w_i}. \qquad (12)$$

Method (Cont'd)



Major results

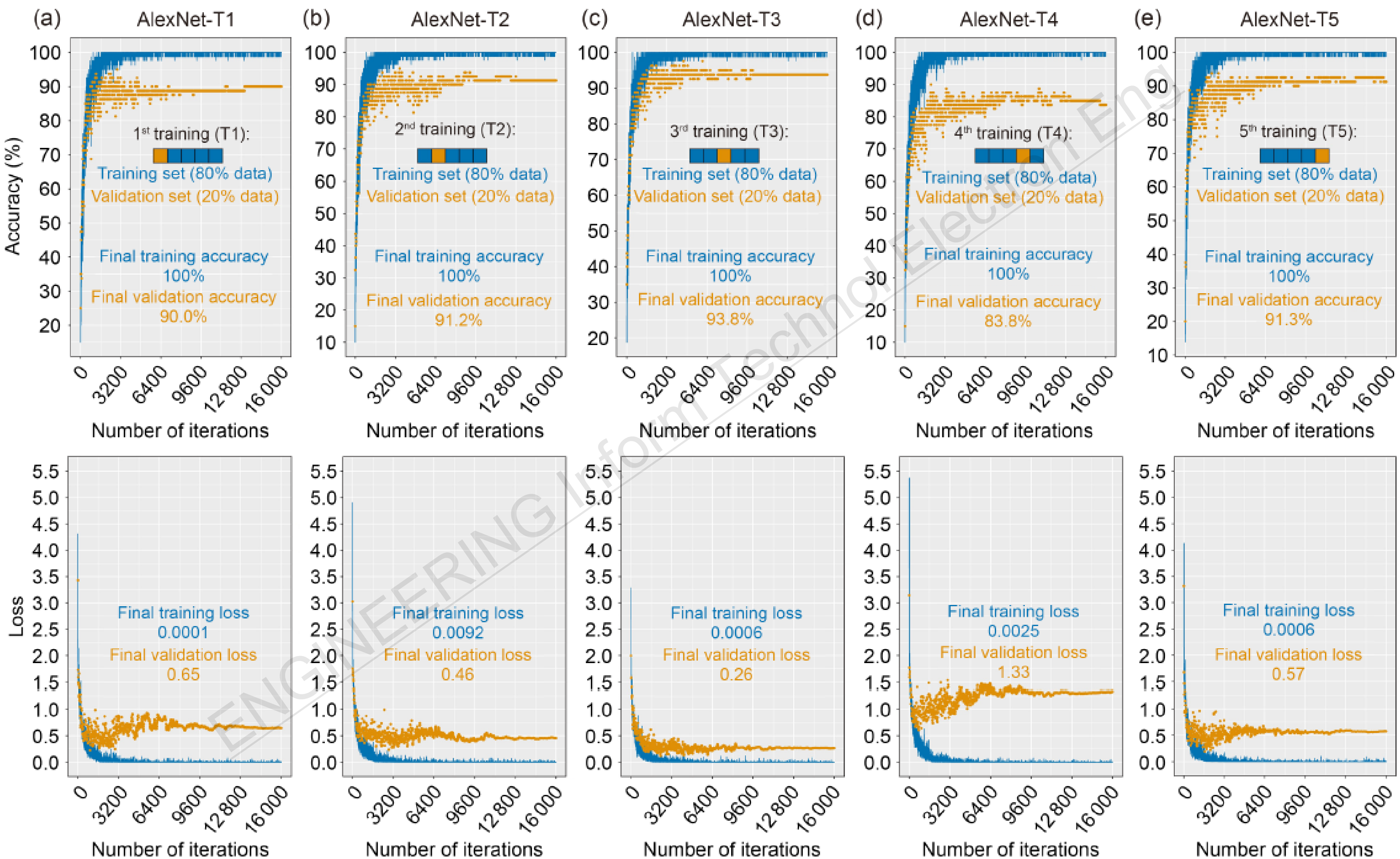


Fig. 2 Training and validation performances of AlexNet models across 5-fold cross-validation training (T1–T5) after 16 000 iterations in 4000 epochs: (a) AlexNet-T1; (b) AlexNet-T2; (c) AlexNet-T3; (d) AlexNet-T4; (e) AlexNet-T5 (blue: metrics of model performance in training; orange: metrics of model performance in validation; accuracy: classification accuracy; loss: prediction loss for 8-stage prediction)

Major results (Cont’d)

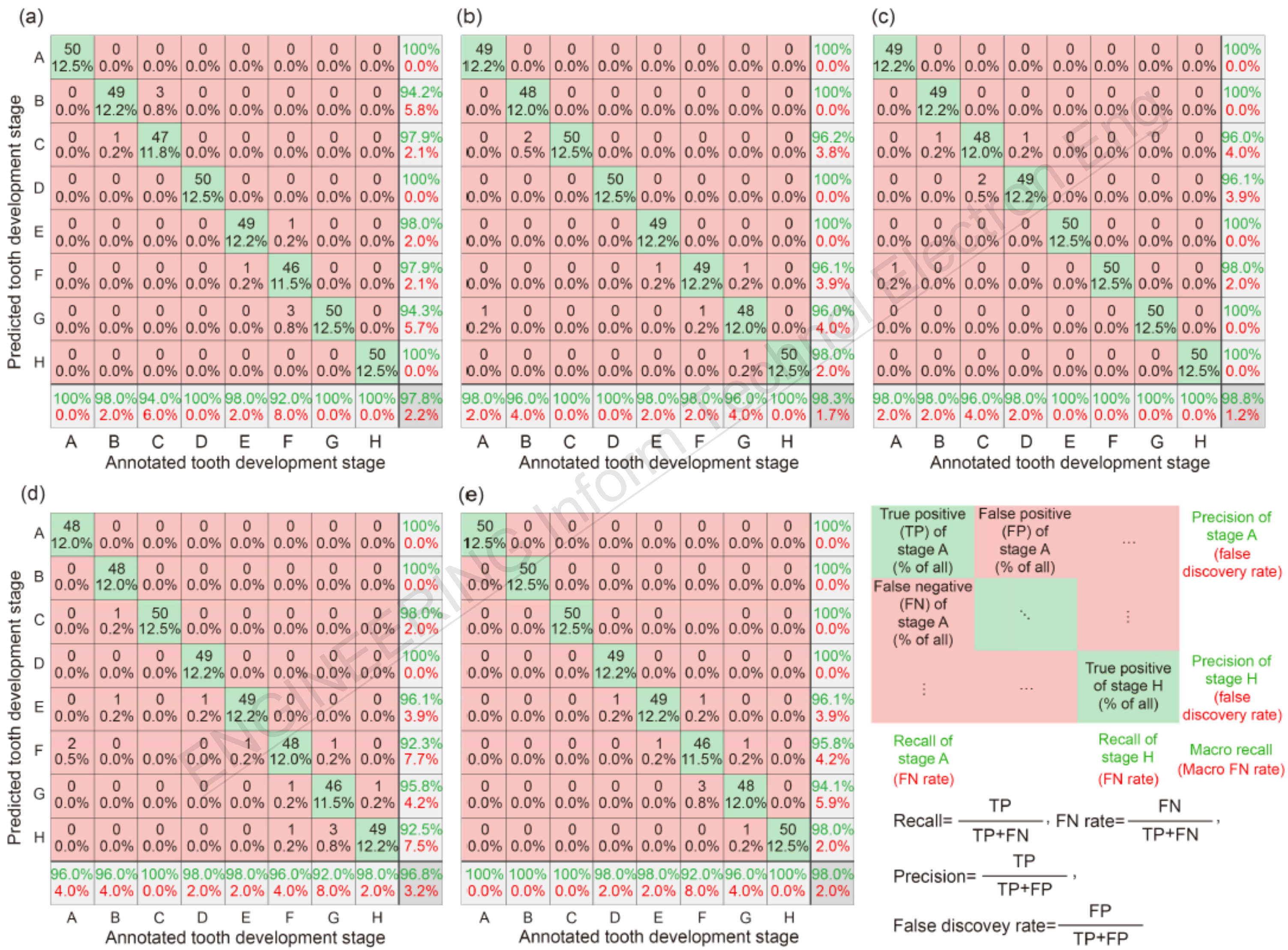
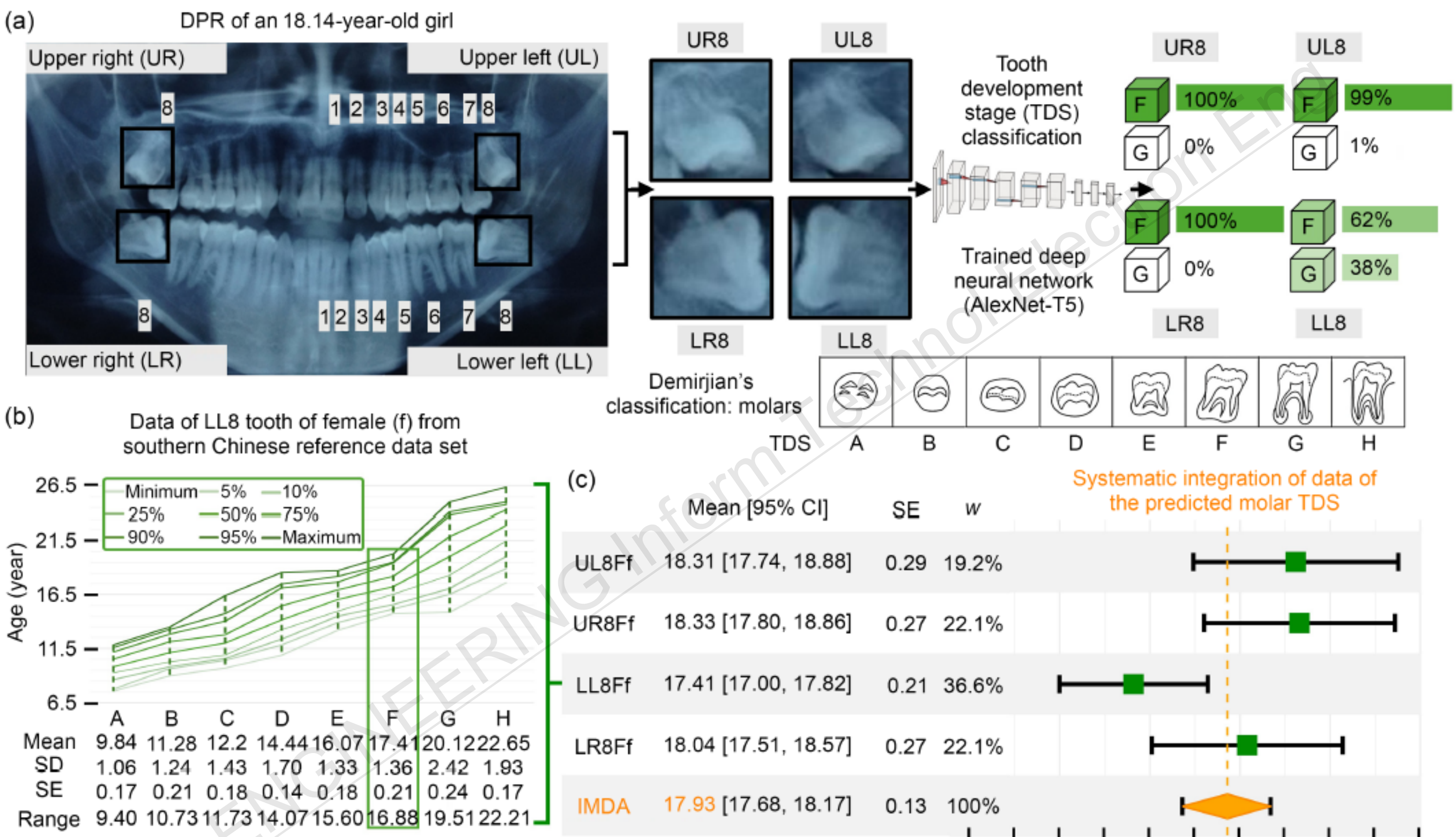


Fig. 3 Confusion matrices for classification performance of the AlexNet models in each molar TDS classification: (a) AlexNet-T1; (b) AlexNet-T2; (c) AlexNet-T3; (d) AlexNet-T4; (e) AlexNet-T5 (green diagonal cells: the counts that are correctly classified for that class; red off-diagonal cells: the counts that are misclassified)

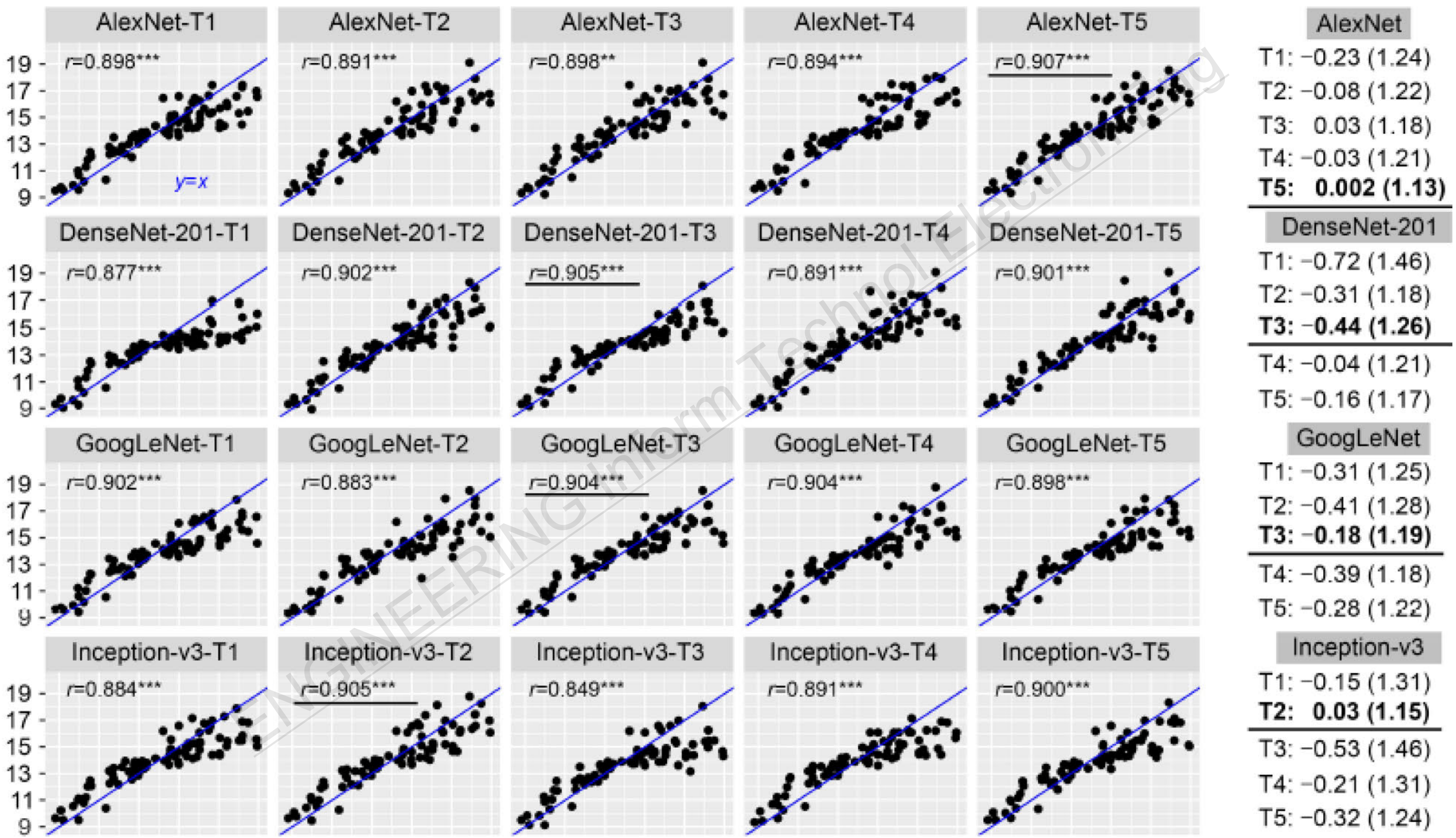
Major results (Cont'd)



DPRs=digital panoramic radiographs; SD=standard deviation; SE=standard error; CI=confidence interval; $w=1/SE^2$; IMDA=integrated mean dental age. Teeth in stage H are excluded from the systematic integration of data.

Fig. 4 IMDA calculation example for an 18.14-year-old girl using the DNN-IMDA framework integrating the trained DNNs and the southern Chinese RDS from her DPR: (a) TDS classification of the third molars (UL8, UR8, LL8, and LR8) using the trained DNN and AlexNet-T5 (green bar is the predicted probability of TDSs (stage F or G)); (b) statistical data of LL8Ff from the southern Chinese RDS (see Table S17 in the supplementary materials); (c) systematic integration of data of the predicted molar TDS (error bar in the forest plot indicates the 95% CI of each predicted molar TDS and IMDA; orange dashed line: IMDA; green box: mean age of the predicted molar TDS)

Major results (Cont’d)



Major results (Cont'd)

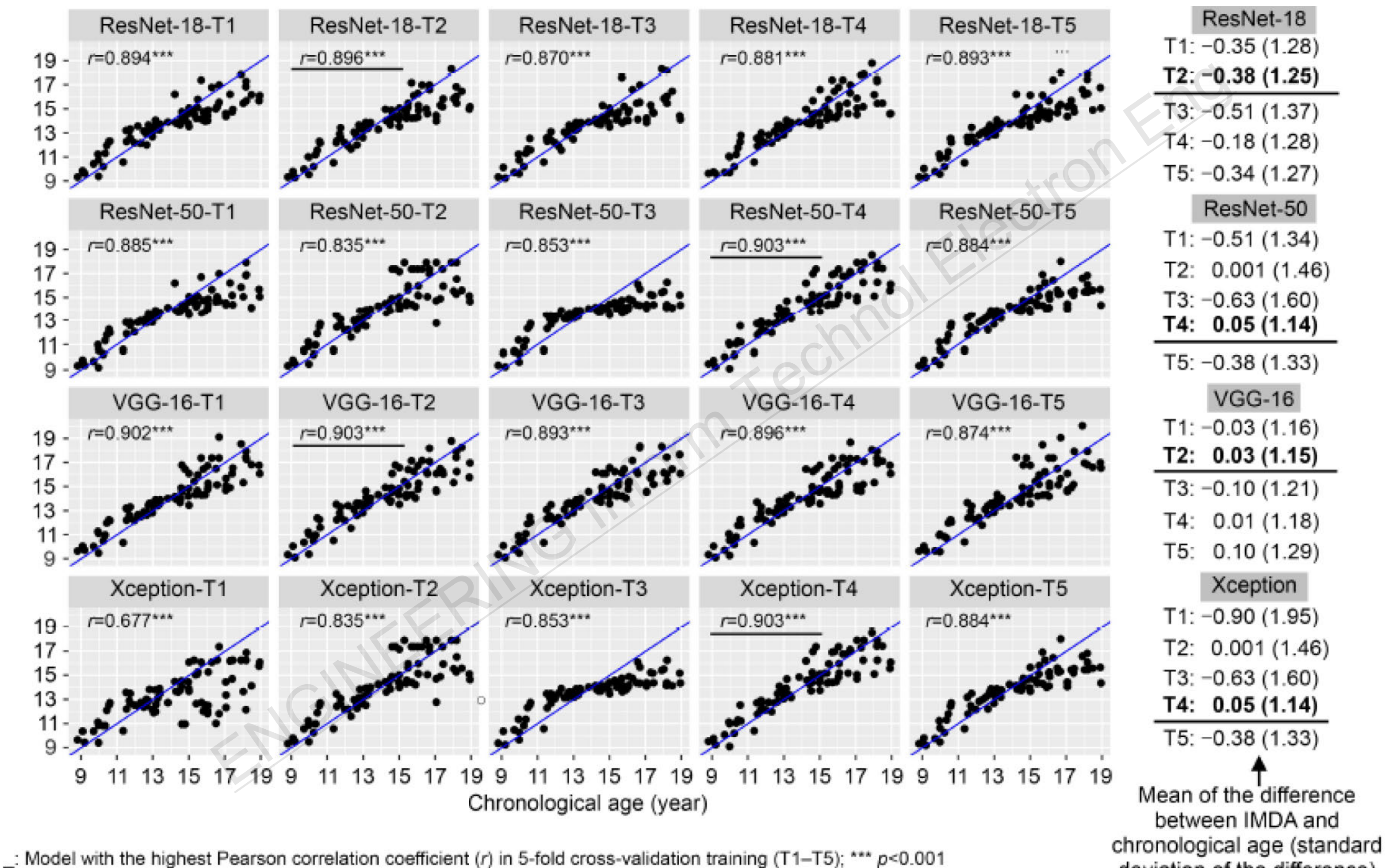


Fig. 5 Scatterplots of IMDA obtained from the 40 trained models (8 deep neural network×5 cross-validation folds) against chronological age (blue line represents the line of identity with *y*=*x* and IMDA=chronological age)

Conclusions

1. Among the architectures, DenseNet-201 achieves the highest mean validation accuracy of 93.0%±2.1%, with macro recall and macro F1 scores of 0.985±0.008 and 0.984±0.008, respectively, and the lowest validation loss (0.39±0.23), demonstrating excellent performance in molar TDS classification.
2. In this study, the DNN-predicted TDS is mapped to the data from the most complete southern Chinese RDS. The results of the IMDA method show strong agreement with the gold standard (chronological age). The maximum Pearson correlation coefficients between IMDA and chronological age in 5-fold cross-validation training are found to exceed 0.90 (p<0.001), except for ResNet-18 (r=0.896, p<0.001).
3. The lowest mean difference and highest mean correlation coefficient between the estimated IMDA from the best-performing DNN (AlexNet) and chronological age are −0.063 years (−3.3 weeks) and r=0.898 (p<0.001), respectively.

Brief bio of the lead author

Hai Ming WONG is a Clinical Professor and the Undergraduate Programme Director in Paediatric Dentistry at the University of Hong Kong (HKU), and the Clinic Manager for Paediatric Dentistry and Orthodontics at the Prince Philip Dental Hospital. She has also served as Assistant Dean (Research and Innovation) and Discipline Coordinator of Paediatric Dentistry at the Faculty of Dentistry, HKU.