Chest radiography as a biomarker of ageing: artificial intelligence-based, multi-institutional model development and validation in Japan

Chest radiography as a biomarker of ageing: artificial intelligence-based, multi-institutional model development and validation in Japan

Yasuhito Mitsuyama, Toshimasa Matsumoto, Hiroyuki Tatekawa, Shannon L Walston, Tatsuo Kimura, Akira Yamamoto, Toshio Watanabe, Yukio Miki, Daiju Ueda

Summary

Background Chest radiographs are widely available and cost-effective; however, their usefulness as a biomarker of ageing using multi-institutional data remains underexplored. The aim of this study was to develop a biomarker of ageing from chest radiography and examine the correlation between the biomarker and diseases.

Methods In this retrospective, multi-institutional study, we trained, tuned, and externally tested an artificial intelligence (AI) model to estimate the age of healthy individuals using chest radiographs as a biomarker. For the biomarker modelling phase of the study, we used healthy chest radiographs consecutively collected between May 22, 2008, and Dec 28, 2021, from three institutions in Japan. Data from two institutions were used for training, tuning, and internal testing, and data from the third institution were used for external testing. To evaluate the performance of the AI model in estimating ages, we calculated the correlation coefficient, mean square error, root mean square error, and mean absolute error. The correlation investigation phase of the study included chest radiographs from individuals with a known disease that were consecutively collected between Jan 1, 2018, and Dec 31, 2021, from an additional two institutions in Japan. We investigated the odds ratios (ORs) for various diseases given the difference between the AI-estimated age and chronological age (ie, the difference-age).

Findings We included 101 296 chest radiographs from 70 248 participants across five institutions. In the biomarker modelling phase, the external test dataset from 3467 healthy participants included 8046 radiographs. Between the AI-estimated age and chronological age, the correlation coefficient was 0·95 (99% CI 0·95–0·95), the mean square error was 15·0 years (99% CI 14·0–15·0), the root mean square error was 3·8 years (99% CI 3·8–3·9), and the mean absolute error was 3·0 years (99% CI 3·0–3·1). In the correlation investigation phase, the external test datasets from 34 197 participants with a known disease included 34 197 radiographs. The ORs for difference-age were as follows: 1·04 (99% CI 1·04–1·05) for hypertension; 1·02 (1·01–1·03) for hyperuricaemia; 1·05 (1·03–1·06) for chronic obstructive pulmonary disease; 1·08 (1·06–1·09) for interstitial lung disease; 1·05 (1·03–1·06) for chronic renal failure; 1·04 (1·03–1·06) for atrial fibrillation; 1·03 (1·02–1·04) for osteoporosis; and 1·05 (1·03–1·06) for liver cirrhosis.

Interpretation The AI-estimated age using chest radiographs showed a strong correlation with chronological age in the healthy cohorts. Furthermore, in cohorts of individuals with known diseases, the difference between estimated age and chronological age correlated with various chronic diseases. The use of this biomarker might pave the way for enhanced risk stratification methodologies, individualised therapeutic interventions, and innovative early diagnostic and preventive approaches towards age-associated pathologies.

Funding None.

Lancet Healthy Longev 2023; 4: e478–86

Published Online August 16, 2023 https://doi.org/10.1016/S2666-7568(23)00133-2

See Comment page e446

For the Japanese translation of the abstract see Online for appendix 1

Department of Diagnostic and Interventional Radiology (Y Mitsuyama MD, T Matsumoto PhD, H Tatekawa MD, S L Walston MS, A Yamamoto MD, Prof Y Miki MD, D Ueda MD) **and Department of Premier Preventive Medicine** (T Kimura MD, Prof T Watanabe MD), **Graduate School of Medicine, Osaka Metropolitan University, Osaka, Japan; Center for Health Science Innovation, Osaka Metropolitan University, Osaka, Japan** (T Matsumoto, D Ueda)

Correspondence to: Dr Daiju Ueda, Department of Diagnostic and Interventional Radiology, Graduate School of Medicine, Osaka Metropolitan University, Osaka 545-8585, Japan
ai.labo.ocu@gmail.com

Introduction

Research on ageing and longevity is becoming increasingly important as the global population ages. Ageing is a complex change that is strongly associated with several diseases,[1,2] but its effects on different individuals are highly heterogeneous. At the same age, some individuals are already frail and require assistance in activities of daily living, whereas others are independent and do not display substantial physiological deterioration until a very old age. In response to the need to better understand the biological ageing process and the determinants of healthy ageing, various molecular and phenotypic biomarkers of ageing have been proposed.[3,4] One such phenotypic biomarker, chest radiography, is simple, commonly used, objective, reproducible, and minimally invasive.[5,6] Furthermore, chest radiographs have the potential to assess complex ageing-associated changes because they can show not only the shape of features within the body, but also details of internal organs and bones. These features make chest radiographs a simple and quantitative potential biomarker

Research in context

Evidence before this study

We evaluated the current state of knowledge on signs of ageing on chest radiography by searching PubMed, MEDLINE, and Web of Science for articles published in English between database inception and Jan 31, 2023, using the keywords "artificial intelligence", "AI", "deep learning", "convolutional neural network", "age", "aging", "chest x-ray", "chest radiography", and "chest radiograph". Several studies have previously used artificial intelligence (AI) to estimate age from chest radiographs, but none reported development of a biomarker of ageing using chest radiographs from multi-institutional data.

Added value of this study

We created a new biomarker of ageing that estimates age from chest radiographs, based on an AI model developed from a cohort of healthy individuals. The AI model developed in this study exhibited excellent correlation coefficients, root mean square errors, and mean absolute errors on a test dataset obtained from an institution independent from the datasets used for training. Additionally, we found that the difference between chest radiography-based apparent age and chronological age was associated with several diseases, especially chronic diseases.

Implications of all the available evidence

Chest radiographs are a valuable source of marks of ageing, which can be investigated using AI. Not only can the AI model accurately estimate chronological age in the healthy cohort, but various diseases also increase the difference between chest radiography-based apparent age and chronological age. Chest radiographs contain intrinsic information that AI models can leverage to assess relationships between disease and physical ageing.

of ageing that might provide a means to clarify the relationship between ageing and various diseases.

Deep learning is a subfield of artificial intelligence (AI) that has propelled substantial advancements in the field of medical imaging.[7,8] Deep learning can automatically extract features from training data (whereas conventional machine learning needs the features to be manually extracted),[7,8] which is why deep learning is particularly advantageous for tasks involving complex or unknown features. Given that the relationship between ageing and chest radiographs is not yet fully understood, deep learning could possibly help to elucidate this connection. Although some AI research has explored the use of chest radiographs to estimate age, these studies have primarily relied on cohorts of patients known to have disease or were single-centre studies and, as such, do not provide conclusive evidence for the use of chest radiographs as a biomarker of ageing.[9–13]

We aimed to develop a deep learning-based AI model to estimate age from chest radiographs using a multi-institutional cohort of healthy participants. Furthermore, we aimed to apply the developed model to chest radiographs of a cohort of patients with known disease to investigate the relationship between AI-estimated age and each disease, thereby exploring the usefulness of AI-estimated age using chest radiographs as a biomarker.

Methods

Study design

This retrospective, multi-institutional study took place over two major phases: first, biomarker modelling, and second, correlation investigation. In the biomarker modelling phase, we used chest radiographs, which are taken at periodic health check-ups in Japan, to develop a biomarker of ageing. In the correlation investigation phase, we evaluated the correlation between AI-estimated age and disease. In the first phase, a training, tuning, and external testing study was conducted on an AI model to estimate chronological age using digital chest radiographs from a cohort of healthy participants from three institutions in Japan (the MedCity21 clinic, Osaka [institution A], the Bell-land General Hospital, Sakai [institution B], and the Kashiwara Municipal Hospital, Kashiwara [institution C]). Additionally, the regions of interest for age estimation were visualised using a saliency map for each chest radiograph. In the second phase, an external testing study using images collected from a cohort of patients with known disease from two additional institutions in Japan (the Osaka Metropolitan University Hospital, Osaka [institution D] and the Morimoto General Hospital, Osaka [institution E]) was conducted to investigate whether the difference between the AI-estimated age and chronological age (ie, difference-age) was associated with the disease. This study was conducted in accordance with the Declaration of Helsinki and the protocol was reviewed and approved by the Ethical Committee of the Osaka Metropolitan University Graduate School of Medicine (approval number 2022-151). The requirement for informed consent was waived by the committee because the images used in this study had been previously acquired during daily clinical practice. This manuscript adheres to TRIPOD reporting guidelines.[14]

Participants, examination, and image acquisition

For the biomarker modelling phase, screening chest radiographs with a posteroanterior view in the standing position were retrospectively collected from periodic health check-ups at institution A (from April 16, 2014, to Dec 28, 2020), institution B (from Aug 15, 2012, to Sept 10, 2021), and institution C (from May 22, 2008, to Dec 28, 2021). If a participant underwent chest radiography more than once in the data collection period, all examinations were included. Disease registration was done for all participants on the day of the examination at

each hospital. All participants with previous disease registrations, including on the day of examination, were excluded according to the biomarker definition.[3,4] The disease registry was established in accordance with the International Classification of Diseases and Related Health Problems, tenth revision.[15] No information on race or ethnicity was collected.

For the correlation investigation phase, chest radiographs with a posteroanterior view in the standing position were consecutively collected at institution D (from Jan 1, 2018, to Dec 31, 2020) and institution E (from Jan 1, 2018, to Dec 31, 2021). If a participant underwent chest radiography more than once in the data collection period of the correlation investigation phase, only the first examination was included.

Ground truth labelling

For the biomarker modelling phase, chest radiographs were labelled using the chronological age of participants. For the correlation investigation phase, chest radiographs were labelled with chronological age and any disease present at the time of chest radiography using the International Classification of Diseases and Related Health Problems, tenth revision.[15] For instance, acute diseases that had already resolved and conditions suspected of being a particular disease but concluded not to be after thorough examination were not labelled.

Data partition

In the biomarker modelling phase, all labelled chest radiographs from institutions A and B were randomly assigned on a participant basis (ie, if a participant had more than one radiograph, they were all assigned together) in an 8:1:1 ratio into training, tuning, and internal test datasets for each institution, whereas institution C was used as an external test dataset. In the correlation investigation phase, institutions D and E were used as external test datasets. The training dataset was used to train the AI model and the tuning dataset was used to internally tune the AI model. The internal test dataset

	Biomarker modelling phase							Correlation investigation phase	
	Training datasets		Tuning datasets		Internal test datasets		External test dataset	External test datasets	
	Institution A	Institution B	Institution A	Institution B	Institution A	Institution B	Institution C	Institution D	Institution E
Total number of examinations	37746	9471	4808	1179	4732	1117	8046	22184	12013
Total patients examined	21925	3771	2975	471	2971	471	3467	22184	12013
Male	10178 (46·4%)	2048 (54·3%)	1402 (47·1%)	247 (52·4%)	1374 (46·2%)	246 (52·2%)	1940 (56·0%)	11102 (50·0%)	6444 (53·6%)
Female	11747 (53·6%)	1723 (45·7%)	1573 (52·9%)	224 (47·6%)	1597 (53·8%)	225 (47·8%)	1527 (44·0%)	11082 (50·0%)	5569 (46·4%)
Age, years	49 (42–57)	64 (50–70)	49 (41–57)	65 (51–71)	49 (42–58)	62 (48–70)	56 (45–66)	70 (52–80)	68 (53–76)
Participants without disease	21925	3771	2975	471	2971	471	3467	0	0
Participants with disease	0	0	0	0	0	0	0	22184	12013
Hypertension	0	0	0	0	0	0	0	9557 (43·1%)	4596 (38·3%)
Dyslipidaemia	0	0	0	0	0	0	0	6284 (28·3%)	2268 (18·9%)
Hyperuricaemia	0	0	0	0	0	0	0	2972 (13·4%)	757 (6·3%)
Diabetes	0	0	0	0	0	0	0	8108 (36·5%)	3639 (30·3%)
Malignant tumour	0	0	0	0	0	0	0	9861 (44·5%)	3915 (32·6%)
Chronic obstructive pulmonary disease	0	0	0	0	0	0	0	884 (4·0%)	643 (5·4%)
Asthma	0	0	0	0	0	0	0	1511 (6·8%)	1128 (9·4%)
Interstitial lung disease	0	0	0	0	0	0	0	1213 (5·5%)	454 (3·8%)
Inactive tuberculosis	0	0	0	0	0	0	0	440 (2·0%)	608 (5·1%)
Non-tuberculosis mycobacteria	0	0	0	0	0	0	0	224 (1·0%)	85 (0·7%)
Heart failure	0	0	0	0	0	0	0	3548 (16·0%)	2907 (24·2%)
Chronic ischaemic heart disease	0	0	0	0	0	0	0	983 (4·4%)	1030 (8·6%)
Atrial fibrillation	0	0	0	0	0	0	0	1825 (8·2%)	1524 (12·7%)
Cerebral infarction	0	0	0	0	0	0	0	1239 (5·6%)	2516 (20·9%)
Liver cirrhosis	0	0	0	0	0	0	0	1274 (5·7%)	537 (4·5%)
Spinal stenosis	0	0	0	0	0	0	0	1302 (5·9%)	621 (5·2%)
Osteoporosis	0	0	0	0	0	0	0	4084 (18·4%)	983 (8·2%)
Anaemia	0	0	0	0	0	0	0	5578 (25·1%)	1899 (15·8%)
Chronic renal failure	0	0	0	0	0	0	0	2242 (10·1%)	422 (3·5%)

Data are n, n (%), or median (IQR).

Table 1: **Demographic characteristics of the training, tuning, internal test, and external test datasets**

was an independent dataset collected from the same institutions as training and tuning datasets, and was not used for training and tuning.[16] Conversely, the external test dataset was collected from a different facility and used exclusively for performance evaluation.

Model development

An AI model was developed to calculate the model-estimated age using ConvNeXt.[17] Root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE) were compared as the loss function. The AI model was trained and tuned using the training and tuning datasets, respectively, starting with the initial parameters from the ImageNet-pretrained model and allowing the overall parameters to be updated. During training, the AI model identified which features could predict chronological age from posteroanterior radiographs. The model that achieved the smallest loss function value on the tuning dataset (within 300 epochs) was selected as the best-performing model. The longest side of the image was downscaled to 320 pixels while maintaining the aspect ratio, and the width along the shorter side was then padded black to 320 pixels.[18] We performed data expansion with TrivialAugment Wide.[19] Every development process was performed using the PyTorch framework (version 2.0.1). Detailed processes for the development of the AI model and machine environment are shown in appendix 2 (p 2). An outline of the model is shown in appendix 2 (p 3), and the source code is available online.[20]

See Online for appendix 2

	Internal test datasets		External test dataset
	Institution A	Institution B	Institution C
Overall			
Correlation coefficient	0·96 (0·95–0·96)	0·97 (0·96–0·97)	0·95 (0·95–0·95)
Mean square error, years	9·9 (9·3–11·0)	11·0 (9·9–13·0)	15·0 (14·0–15·0)
Root mean square error, years	3·1 (3·0–3·3)	3·3 (3·1–3·5)	3·8 (3·8–3·9)
Mean absolute error, years	2·5 (2·4–2·5)	2·6 (2·4–2·8)	3·0 (3·0–3·1)
Age range 21–40 years			
Mean square error, years	11·0 (9·3–14·0)	10·0 (7·0–14·0)	14·0 (12·0–15·0)
Root mean square error, years	3·3 (3·1–3·7)	3·2 (2·6–3·7)	3·7 (3·4–3·9)
Mean absolute error, years	2·6 (2·4–2·7)	2·5 (2·1–3·0)	2·8 (2·7–3·0)
Age range 41–60 years			
Mean square error, years	8·7 (8·1–9·3)	13·0 (11·0–16·0)	14·0 (13·0–15·0)
Root mean square error, years	3·0 (2·8–3·1)	3·6 (3·3–4·0)	3·8 (3·6–3·9)
Mean absolute error, years	2·3 (2·2–2·4)	2·9 (2·6–3·2)	3·0 (2·9–3·1)
Age range 61–80 years			
Mean square error, years	12·0 (11·0–13·0)	9·2 (8·1–11·0)	15·0 (14·0–16·0)
Root mean square error, years	3·5 (3·3–3·7)	3·0 (2·8–3·2)	3·9 (3·8–4·0)
Mean absolute error, years	2·8 (2·6–2·9)	2·4 (2·2–2·6)	3·1 (3·0–3·2)
Age range 81–100 years			
Mean square error, years	..	57·0 (33·0–89·0)	72·0 (55·0–91·0)
Root mean square error, years	..	7·5 (5·7–9·4)	8·5 (7·4–9·6)
Mean absolute error, years	..	7·2 (5·6–9·0)	7·9 (6·8–8·9)

All data except correlation coefficient are years (99% CI).

Table 2: **Results of the biomarker modelling phase**

Model test

In the biomarker modelling phase, the regression performance of the AI model was assessed using the best-performing model as identified by the tuning dataset. The internal test datasets (from institutions A and B) and the external test dataset (from institution C) were used to assess the accuracy of the AI model in estimating chronological age from chest radiographs. Saliency maps were generated for each radiograph in the external test dataset (institution C) to indicate the region of interest used in the calculation of the AI-estimated age. Shapley Additive exPlanations (SHAP), a framework that applies the Shapley value (an importance value calculated based on game theory) to machine learning, was used to generate saliency maps for each radiograph.[21] For a computer vision task, SHAP computes the contribution of a group of pixels to a particular prediction. After stratification of the saliency maps of the chest radiographs into 20-year chronological age groups, each map was averaged to produce an average saliency image for each age group. The location of hot (higher SHAP values) or cold (lower SHAP values) regions seen in saliency maps was evaluated by radiologists (YMit and HT).

In the correlation investigation phase, external test datasets (from institutions D and E) were used to investigate the relationship between disease and difference-age. To show the region of interest of the AI model, we also generated saliency maps for one external dataset (from institution D) with SHAP.

Statistical analysis

In the biomarker modelling phase, the correlation coefficient, RMSE, MSE, and MAE between the model-estimated age and chronological age were calculated to evaluate the regression performance of the AI model. Calibration was done with linear regression to adjust model prediction.

In the correlation investigation phase, the correlation coefficient, RMSE, MSE, and MAE between the model-estimated age and chronological age were also calculated. To examine the strength of the relationship between each disease and difference-age, logistic regression analysis was performed with each disease as the objective variable and difference-age and chronological age as predictors. Chronological age was included as a predictor in the model to adjust for differences in disease frequency by age. Additionally, we performed a multivariate analysis incorporating chronic diseases as predictors (excluding the particular disease from predictors when it was the outcome variable) to handle the interactions between diseases. In the subanalysis, we incorporated sex as a predictor and calculated the odds ratios (ORs) for malignant tumours in each organ. We re-examined the

data from institutions D and E after excluding participants who fell outside the age range of institutions A and B, used to develop the model (ie, participants younger than 18 years and older than 94 years).

All analyses were performed using R (version 4.0.0). These statistical inferences were performed with a two-sided significance level of 1%, and the performance metrics were presented with 99% CIs.

Role of the funding source

There was no funding source for this study.

Results

In the biomarker modelling phase, after exclusion of 14889 participants with existing conditions, 67099 radiographs from 36051 healthy participants who underwent health check-ups at institutions A, B, and C were included. In the correlation investigation phase, after exclusion of 138554 participants for whom chest radiographs were not available, 34197 radiographs from 34197 participants from institutions D and E were included. Detailed demographics for each dataset are shown in table 1. The eligibility flowchart for each dataset is shown in appendix 2 (p 4) and the age distribution of the data collected from the five institutions is shown in appendix 2 (pp 5–6). Radiological devices and technical parameters used for chest radiographs at each institution (A, B, C, D, and E) are summarised in appendix 2 (p 9).

The best-performing model was obtained at 200 epochs with the lowest loss value using the RMSE for the loss function. The correlation coefficient, RMSE, MSE, and MAE of the external test dataset (from institution C), in the biomarker modelling phase, are shown in table 2. The RMSE, MSE, and MAE of the external test dataset (from institution C) separated by age group are shown in table 2 and scatter plots and Bland-Altman plots are shown in figure 1. The model regression metrics based on sex are shown in appendix 2 (p 10). Calibrated results were nearly identical to the original results (appendix 2 pp 7, 11), so in the interest of simplicity, the non-calibrated model was chosen for the correlation investigation phase. Averaged saliency maps show higher SHAP values (hot regions) at the top of the mediastinum and lower SHAP values (cold regions) at the bilateral lower lung fields. Saliency maps separated by age group are shown in figure 2.

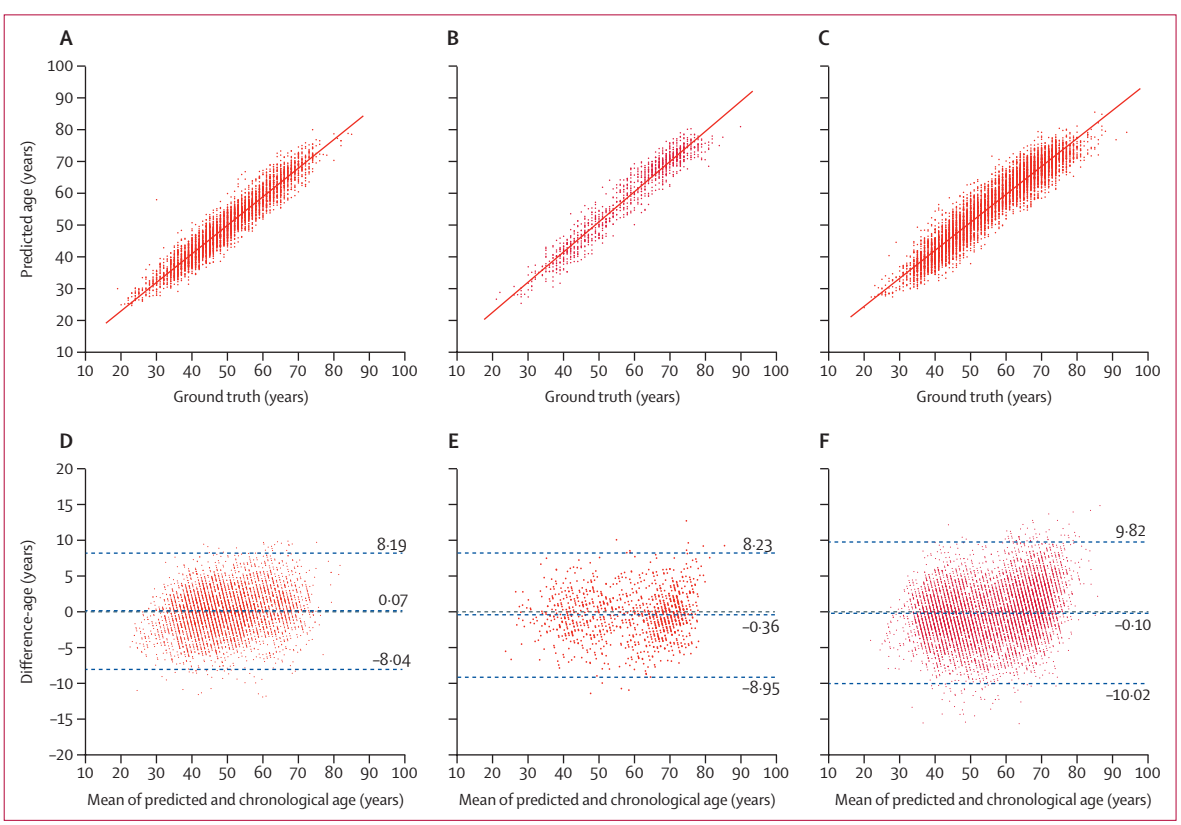

Figure 1: **Model performance results**

Scatterplots (each including a regression line) of ground truth and predicted values of the artificial intelligence model for the internal dataset from institution A (A); the internal dataset from institution B (B); and the external test dataset from institution C (C). Pearson's correlation coefficients are 0·96 (99% CI 0·95–0·96) for the internal test dataset from institution A, 0·97 (0·96–0·97) for the internal test dataset from institution B, and 0·95 (0·95–0·95) for the external test dataset from institution C. Bland-Altman plots containing three dotted lines, with the middle line representing the mean difference value, and the upper and lower lines representing the 99% CIs for the internal dataset from institution A (D); the internal dataset from institution B (E); and the external dataset from institution C (F).

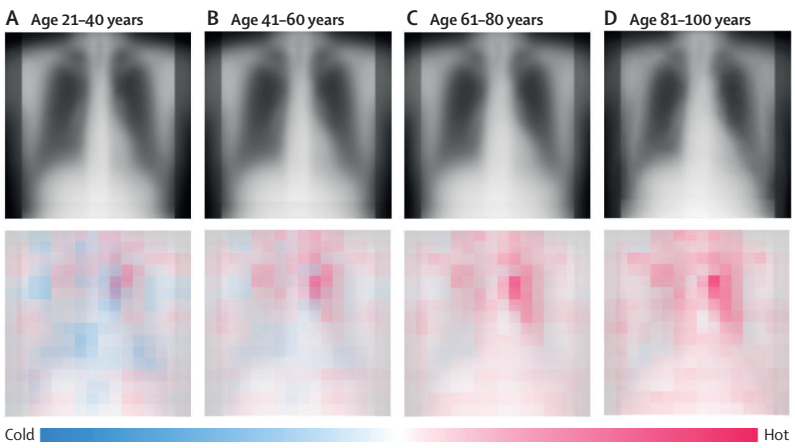

Figure 2: **Saliency maps from the external test dataset**
Average saliency images of each 20-year chronological age group in the external test dataset from institution C. The top panels show averaged chest radiographs for all participants in the group and the bottom panels show averaged saliency maps. Hot areas in the saliency maps indicate characteristics of increasing age in chest radiographs; cold areas indicate characteristics of decreasing age.

	Without chronic diseases as predictors of difference-age		With chronic diseases as predictors of difference-age	
	Odds ratio	p value	Odds ratio	p value
Hypertension	1·06 (1·05–1·06)	<0·0001	1·04 (1·04–1·05)	<0·0001
Dyslipidaemia	1·03 (1·02–1·04)	<0·0001	1·01 (1·00–1·02)	0·041
Hyperuricaemia	1·06 (1·05–1·07)	<0·0001	1·02 (1·01–1·03)	<0·0001
Diabetes	1·03 (1·02–1·04)	<0·0001	1·01 (1·00–1·01)	0·031
Malignant tumour	1·06 (1·05–1·07)	<0·0001	1·05 (1·05–1·06)	<0·0001
Pneumonia	0·97 (0·96–0·99)	0·0002	0·98 (0·96–1·00)	0·0050
Chronic obstructive pulmonary disease	1·06 (1·05–1·08)	<0·0001	1·05 (1·03–1·06)	<0·0001
Asthma	1·01 (1·00–1·02)	0·085	1·00 (0·98–1·01)	0·33
Interstitial lung disease	1·10 (1·08–1·11)	<0·0001	1·08 (1·06–1·09)	<0·0001
Inactive tuberculosis	1·03 (1·01–1·05)	<0·0001	1·01 (0·99–1·03)	0·26
Non-tuberculosis mycobacteria	1·04 (1·01–1·08)	0·0020	1·03 (0·99–1·07)	0·040
Heart failure	1·02 (1·01–1·03)	<0·0001	1·00 (0·99–1·01)	0·76
Acute myocardial infarction	0·98 (0·94–1·03)	0·30	0·98 (0·93–1·03)	0·40
Chronic ischaemic heart disease	1·03 (1·02–1·05)	<0·0001	1·02 (1·00–1·03)	0·0050
Atrial fibrillation	1·05 (1·04–1·06)	<0·0001	1·04 (1·03–1·06)	<0·0001
Cerebral infarction	1·01 (1·00–1·02)	0·098	0·99 (0·98–1·01)	0·17
Liver cirrhosis	1·07 (1·06–1·09)	<0·0001	1·05 (1·03–1·06)	<0·0001
Spinal stenosis	1·04 (1·03–1·06)	<0·0001	1·04 (1·02–1·06)	<0·0001
Osteoporosis	1·04 (1·03–1·05)	<0·0001	1·03 (1·02–1·04)	<0·0001
Anaemia	1·04 (1·03–1·05)	<0·0001	1·02 (1·01–1·03)	<0·0001
Acute renal failure	1·04 (0·97–1·10)	0·17	1·01 (0·95–1·08)	0·60
Chronic renal failure	1·07 (1·06–1·09)	<0·0001	1·05 (1·03–1·06)	<0·0001
Urinary tract infection	1·00 (0·97–1·04)	0·79	0·99 (0·96–1·03)	0·62
Enteritis	0·99 (0·92–1·06)	0·61	0·99 (0·92–1·06)	0·76
Cholecystitis	0·97 (0·91–1·03)	0·25	0·99 (0·93–1·05)	0·59

Data are odds ratio (99% CI).

Table 3: **Odds ratio summary for external test datasets D and E (cohorts of patients with known diseases)**

In the correlation investigation phase, the difference-age was correlated with various chronic diseases in the external datasets. No statistically significant differences were observed for most acute diseases. ORs for multivariate analysis with and without chronic diseases as predictors of difference-age are shown in table 3. The correlation between difference-age and each disease is shown in figure 3. A summary of the ORs for correlations between difference-age and diseases by institution is shown in appendix 2 (p 12). The correlation coefficient, RMSE, MSE, and MAE metrics from the correlation investigation phase are shown in appendix 2 (p 13). Results of the subanalysis done with sex as a predictor are shown in appendix 2 (p 14) and the ORs of malignant tumours for each organ are shown in appendix 2 (p 15). Results after excluding participants who fell outside the age range of the biomarker modelling phase are shown in appendix 2 (p 16). Averaged saliency maps for each disease are shown in appendix 2 (p 8).

Discussion

In this study, we developed an AI model that shows that chest radiographs can be used as a biomarker of ageing. The AI model's results for the correlation coefficient, RMSE, MSE, and MAE for the external dataset of the healthy cohort suggest that the AI-estimated age correlates very strongly with chronological age. Application of the AI model to participant data showed an association between signs of ageing in chest radiographs and several chronic diseases, but not with acute diseases.

To the best of our knowledge, this is the first multi-institutional study to create and validate a deep learning-based regression model for ageing using chest radiographs from healthy participants. In addition, we investigated the association between various diseases and AI-estimated age to assess its usefulness as a biomarker. In contrast to our study, for which the model was created from a cohort of healthy individuals, previous similar studies were modelled on chest radiographs from cohorts that included participants with a known disease.[9,11–13] One study used a cohort that excluded individuals with abnormal findings on chest radiographs, but this criterion would not have guaranteed exclusion of pathological states.[13] Furthermore, deep learning models that are trained and tested on a dataset from a single institution are susceptible to overfitting.[8,22] To prevent overfitting and confirm robustness and generalisability, we collected data for the test set from several different facilities.[22,23] Additionally, no previous studies have identified the relationship between chest radiographs as a biomarker of ageing and various diseases.

In the biomarker modelling phase, we found that saliency maps of the AI findings on the chest radiographs from the healthy cohort showed a variety of age-related findings. The hottest region of the averaged saliency

maps was at the top of the mediastinum and the coldest region was at the bilateral lower lung fields. These hot sites seem to coincide with important features (eg, calcification and tortuosity of the aortic arch) for age estimation identified by previous studies estimating age from chest radiographs, whether conducted manually[24,25] or via deep learning.[9,11–13] In five of six previous studies, the region of interest was on the aortic torch.[9,11–13,24] Our saliency maps—separated by age group—showed an increasing accumulation of hot regions in the superior mediastinum with ageing. Although there are differences in methodology and cohort, this consistent result supports the reliability of the saliency maps from our model. Similarly, the saliency maps showed that the bilateral lower lung fields were increasingly colder for younger ages. These findings agree with those of previous studies examining age-related changes in chest radiographs, which reported emphysematous changes and reticular shadows in the lungs (especially in the lower lung fields).[24,25]

In the correlation investigation phase, we found correlations between our AI-determined model-estimated age and a variety of chronic diseases. The OR in the cohort of patients with a known disease showed a correlation with hypertension, hyperuricaemia, chronic obstructive pulmonary disease, interstitial lung disease, chronic renal failure, liver cirrhosis, osteoporosis, and atrial fibrillation. Conversely, little correlation with acute diseases was observed. This result means that our AI captures chronic changes rather than acute changes in chest radiography, which is reasonable because ageing results from chronic changes that accumulate over time. The relationship between chronic diseases and difference-age can be inferred from several perspectives, such as that of arteriosclerotic changes, changes in the lung field, and changes in the shape of the bones and heart. From the perspective of arteriosclerotic changes, aortic calcification occurs not only with ageing, but also with atherosclerosis caused by hypertension, hyperuricaemia, diabetes, and dyslipidaemia.[26] One study showed that calcification of the aortic arch is associated with hypertension and chronic renal failure,[27] whereas other research suggests that hyperuricaemia is associated with calcification of the abdominal aorta[28] and elevated uric acid concentrations are associated with calcification of the aortic arch.[29] These observations are in line with our results, which focus on the calcification of the aorta. From the perspective of changes in the lung field, we found that the lung fields (especially the lower lung fields) had a substantial effect on the accurate estimation of ageing through saliency maps. The estimated age could reasonably be higher for interstitial lung disease and chronic obstructive pulmonary disease, which produce reticular shadows and emphysematous changes known to be related to ageing.[24,25] Surprisingly, our model also showed a strong correlation between liver cirrhosis and difference-age. Findings associated with liver cirrhosis on chest radiographs have not been definitively reported. However, liver cirrhosis is known to cause lung field changes representative of hepatopulmonary syndrome and portopulmonary hypertension, such as dilation of the capillaries in the lung field and main pulmonary artery diameter.[30,31] Other changes that might occur include the development of collateral blood vessels in the thorax and pleural effusions. Another possibility is that patients with cirrhosis have cardiac dysfunction due to increased cardiac output and heart rate, as well as decreased systemic vascular resistance (cirrhotic cardiomyopathy).[32] Cardiac dysfunction might be accompanied by cardiac enlargement, pulmonary oedema, or pleural effusion, which might be reflected in the chest radiographs, and these lung field changes could have been captured by the AI model. Although this association suggests that liver cirrhosis is detectable from chest radiographs, further studies are needed to confirm this possibility. The analysis also showed high ORs for osteoporosis and atrial fibrillation, which could potentially be linked to age-related changes in bone and heart shape. Previous studies support the use of AI models for osteoporosis detection from chest radiographs.[33] Atrial fibrillation has also been reported to cause atrial enlargement,[34] and AI has been shown to be

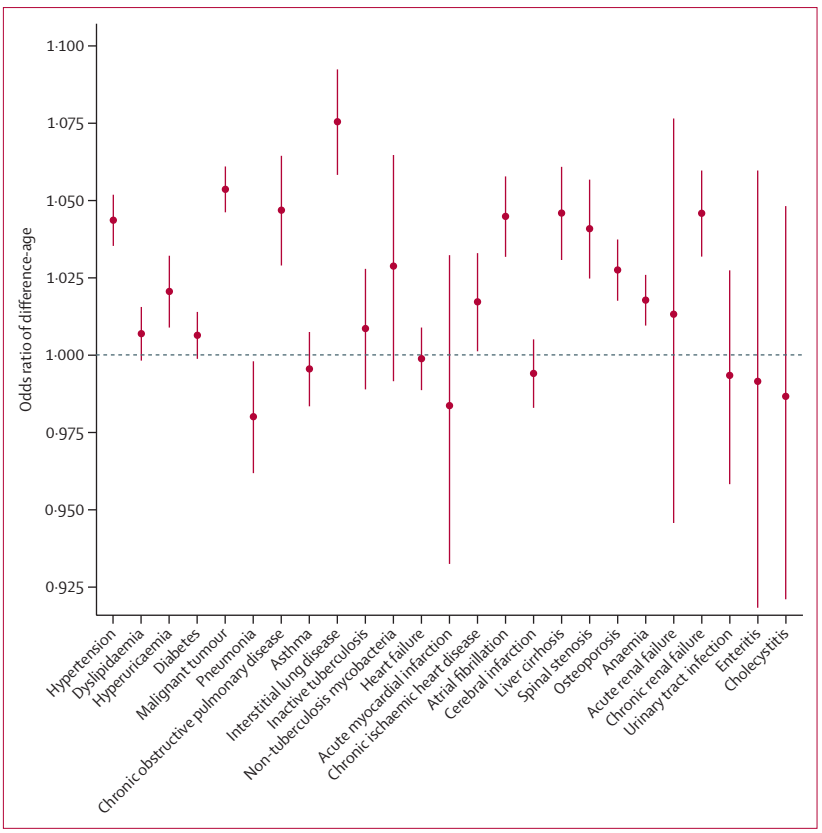

Figure 3: **Correlation between difference-age and each disease**
Odds ratio of difference-age of each disease in the external datasets from institutions D and E. Circles represent means and lines represent 99% CIs.

capable of the detection of atrial fibrillation in chest radiographs.[35] These findings suggest that our model might have detected features of osteoporosis and atrial fibrillation in chest radiographs.

This study had several limitations. First, further validation using a prospective dataset is desirable to confirm causality in addition to correlation. Second, we did not compare our AI model with established biological age markers (eg, frailty index and epigenetic clock). An early diagnosis of frailty is important to improve quality of life and promote healthy ageing, and this comparison might provide additional information about our model. Third, we generated saliency maps using downsampled images. Finally, this AI model was developed and validated with radiographs collected in Japan, and hence its applicability to different racial and ethnic populations should be examined.

In conclusion, we created an AI model that uses chest radiographs as biomarkers of ageing and found that the AI-estimated age of patients with a known disease was higher than their chronological age. Our AI model could potentially serve as an indicator for age-related diseases and help to track the efficacy of healthy ageing interventions. The application of the model in these areas remains speculative and necessitates further research for validation.

Contributors

YMit, TK, and TW acquired the data. YMit, TM, and DU developed the model. YMit, HT, and DU analysed and interpreted the data. YMit, TM, HT, and DU drafted the manuscript. SLW, AY, and YMik revised and approved the manuscript. YMit and DU accessed and verified all the underlying data. All authors had access to all the data reported in the study and accept responsibility for the decision to submit for publication.

Declaration of interests

We declare no competing interests.

Data sharing

The study protocol and metadata are available upon request to the corresponding author. However, chest radiographs are not available at this time, as these have been withheld by the hospitals that participated in the trials in order to protect participant privacy. The code used to develop our model is available at https://github.com/Medical-AI-Lab/Nervus/. Additionally, the web application is available at https://huggingface.co/spaces/MedicalAILabo/Xp-age/. The source code of SHAP is available at https://github.com/slundberg/shap.

Acknowledgments

We are grateful to the MedCity21 clinic, Morimoto Hospital, Kashiwara Municipal Hospital, and Bell-land General Hospital for providing the data for this study.

References

1 Weinert BT, Timiras PS. Invited review: theories of aging. *J Appl Physiol* 2003; **95:** 1706–16.
2 Higgins-Chen AT, Thrush KL, Levine ME. Aging biomarkers and the brain. *Semin Cell Dev Biol* 2021; **116:** 180–93.
3 Jylhävä J, Pedersen NL, Hägg S. Biological age predictors. *EBioMedicine* 2017; **21:** 29–36.
4 Wagner KH, Cameron-Smith D, Wessner B, Franzke B. Biomarkers of aging: from function to molecular biology. *Nutrients* 2016; **8:** 338.
5 Mettler FA Jr, Bhargavan M, Faulkner K, et al. Radiologic and nuclear medicine studies in the United States and worldwide: frequency, radiation dose, and comparison with other radiation sources—1950–2007. *Radiology* 2009; **253:** 520–31.
6 Raoof S, Feigin D, Sung A, Raoof S, Irugulpati L, Rosenow EC 3rd. Interpretation of plain chest roentgenogram. *Chest* 2012; **141:** 545–58.
7 Hinton G. Deep learning—a technology with the potential to transform health care. *JAMA* 2018; **320:** 1101–02.
8 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521:** 436–44.
9 Raghu VK, Weiss J, Hoffmann U, Aerts HJWL, Lu MT. Deep learning to estimate biological age from chest radiographs. *JACC Cardiovasc Imaging* 2021; **14:** 2226–36.
10 Lu MT, Ivanov A, Mayrhofer T, Hosny A, Aerts HJWL, Hoffmann U. Deep learning to assess long-term mortality from chest radiographs. *JAMA Netw Open* 2019; **2:** e197416.
11 Karargyris A, Kashyap S, Wu JT, Sharma A, Moradi M, Syeda-Mahmood T. Age prediction using a large chest x-ray dataset. *arXiv* 2019; published online March 9. https://doi.org/10.48550/arXiv.1903.06542 (preprint).
12 Ieki H, Ito K, Saji M, et al. Deep learning-based age estimation from chest x-rays indicates cardiovascular prognosis. *Commun Med* 2022; **2:** 159.
13 Yang CY, Pan YJ, Chou Y, et al. Using deep neural networks for predicting age and sex in healthy adult chest radiographs. *J Clin Med* 2021; **10:** 4431.
14 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015; **350:** g7594.
15 WHO. The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. Geneva: World Health Organization, 1993.
16 Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health* 2019; **1:** e271–97.
17 Liu Z, Mao H, Wu C-Y, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. 2022 IEEE/CVF conference on computer vision and pattern recognition. 11976–86.
18 Sabottke CF, Spieler BM. The effect of image resolution on deep learning in radiography. *Radiol Artif Intell* 2020; **2:** e190015.
19 Müller SG, Hutter F. TrivialAugment: tuning-free yet state-of-the-art data augmentation. 2021 IEEE/CVF international conference on computer vision. 774–82.
20 Matsumoto T, Walston SL, Miki Y, Ueda D. Nervus: a comprehensive deep learning classification, regression, and prognostication tool for both medical image and clinical data analysis. *arXiv* 2022; published online Dec 12. https://doi.org/10.48550/arXiv.2212.1113 (preprint).
21 Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. 31st international conference on neural information processing systems, 2017. 4768–77.
22 Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology* 2018; **286:** 800–09.
23 Bluemke DA, Moy L, Bredella MA, et al. Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers—from the Radiology Editorial Board. *Radiology* 2020; **294:** 487–89.
24 Ensor RE, Fleg JL, Kim YC, de Leon EF, Goldman SM. Longitudinal chest x-ray changes in normal men. *J Gerontol* 1983; **38:** 307–14.
25 Leelakanok N, Piyavisetpat N. CT features of normal lung change in asymptomatic elderly patients. *Asian Biomed* 2015; **9:** 613–23.
26 Herrington W, Lacey B, Sherliker P, Armitage J, Lewington S. Epidemiology of atherosclerosis and the potential to reduce the global burden of atherothrombotic disease. *Circ Res* 2016; **118:** 535–46.
27 Yamada S, Oshima M, Watanabe Y, Miyake H. Arterial location-specific calcification at the carotid artery and aortic arch for chronic kidney disease, diabetes mellitus, hypertension, and dyslipidemia. *Calcif Tissue Int* 2014; **95:** 267–74.
28 Li YW, Chen WL. Clinical relevance of serum uric acid and abdominal aortic-calcification in a national survey. *Clin Cardiol* 2020; **43:** 1194–201.
29 Yan W, Sun G, Luo A, et al. Serum uric acid is independently associated with aortic arch calcification in a cross-sectional study of middle-aged and elderly women. *Nutr Metab Cardiovasc Dis* 2020; **30:** 932–38.
30 Machicao VI, Balakrishnan M, Fallon MB. Pulmonary complications in chronic liver disease. *Hepatology* 2014; **59:** 1627–37.

31 Ishikawa T, Egusa M, Kawamoto D, et al. Screening for portopulmonary hypertension using computed tomography-based measurements of the main pulmonary artery and ascending aorta diameters in patients with portal hypertension. *Hepatol Res* 2022; **52:** 255–68.

32 Møller S, Henriksen JH. Cirrhotic cardiomyopathy: a pathophysiological review of circulatory dysfunction in liver disease. *Heart* 2002; **87:** 9–15.

33 Jang M, Kim M, Bae SJ, Lee SH, Koh JM, Kim N. Opportunistic osteoporosis screening using chest radiographs with deep learning: development and external validation with a cohort dataset. *J Bone Miner Res* 2022; **37:** 369–77.

34 Sanfilippo AJ, Abascal VM, Sheehan M, et al. Atrial enlargement as a consequence of atrial fibrillation. A prospective echocardiographic study. *Circulation* 1990; **82:** 792–97.

35 Matsumoto T, Ehara S, Walston SL, Mitsuyama Y, Miki Y, Ueda D. Artificial intelligence-based detection of atrial fibrillation from chest radiographs. *Eur Radiol* 2022; **32:** 5890–97.