Identification of age-dependent features of human bronchi using explainable artificial intelligence

Hiroaki Ikushima [1,2] and Kazuhiro Usui [1]

[1]Department of Respiratory Medicine, NTT Medical Center Tokyo, Tokyo, Japan. [2]Department of Respiratory Medicine, The University of Tokyo Hospital, Tokyo, Japan.

Corresponding author: Hiroaki Ikushima (hikushima-tky@umin.ac.jp)

Abstract

Background Ageing induces functional and structural alterations in organs, and age-dependent parameters have been identified in various medical data sources. However, there is currently no specific clinical test to quantitatively evaluate age-related changes in bronchi. This study aimed to identify age-dependent bronchial features using explainable artificial intelligence for bronchoscopy images.

Methods The present study included 11 374 bronchoscopy images, divided into training and test datasets based on the time axis. We constructed convolutional neural network (CNN) models and evaluated these models using the correlation coefficient between the chronological age and the "bronchial age" calculated from bronchoscopy images. We employed gradient-weighted class activation mapping (Grad-CAM) to identify age-dependent bronchial features that the model focuses on. We assessed the universality of our model by comparing the distribution of bronchial age for each respiratory disease or smoking history.

Results We constructed deep-learning models using four representative CNN architectures to calculate bronchial age. Although the bronchial age showed a significant correlation with chronological age in each CNN architecture, EfficientNetB3 achieved the highest Pearson's correlation coefficient (0.9617). The application of Grad-CAM to the EfficientNetB3-based model revealed that the model predominantly attended to bronchial bifurcation sites, regardless of whether the model accurately predicted chronological age or exhibited discrepancies. There were no significant differences in the discrepancy between the bronchial age and chronological age among different respiratory diseases or according to smoking history.

Conclusion Bronchial bifurcation sites are universally important age-dependent features in bronchi, regardless of the type of respiratory disease or smoking history.

Introduction

Ageing is a natural process that causes changes in the function and structure of organs. Various medical data sources have identified age-dependent parameters in different organs [1, 2]. For example, in terms of lung function, the forced expiratory volume in 1 s (FEV_1), a parameter obtained by spirometry, shows age-dependent changes [3], and bone mineral density also exhibits age-dependent changes in terms of the bone function [4]. Recently, machine-learning techniques have been used to identify age-dependent features from various medical images, including head magnetic resonance images [5] and chest radiography images [6]. The structure of the trachea and bronchi also shows age-dependent changes, including increases in the bronchial diameter and anatomical dead space [7, 8]. However, there is currently no specific test to detect and quantitatively evaluate these age-dependent changes in the trachea and bronchi in clinical practice.

Recent advances have made it possible to apply machine learning to medical data for diagnostics, treatment strategy decision-making and prognostic prediction [9]. However, machine-learning models with

a large number of parameters (*e.g.* deep-learning models) often suffer from a black-box problem, where it is difficult for humans to understand the decision-making process of the model's output. To address this issue, explainable artificial intelligence (XAI) research has been advancing rapidly [10]. XAI enables the prediction reasoning of machine-learning models to be explained, making it possible to describe important features that may not have been identified by humans in a way that humans can understand.

In this study, we aimed to identify age-dependent features of the trachea and bronchi that can be quantitatively observed through bronchoscopy. We developed a convolutional neural network (CNN) model to estimate patient age from bronchial images obtained by bronchoscopy. To identify age-dependent features, we then applied XAI methods to the model and evaluated the elements on which the model was focused.

Materials and methods
Data acquisition and preparation
A total of 11 374 bronchoscopy images of 336 consecutive patients who underwent video bronchoscopy at our hospital between January 2020 and February 2023 were obtained from our registry. Bronchoscopes used include Olympus BF type 260, 1T260, UC260FW, XP290, P290, Q290 and 1TH1200. The study protocol was approved by the institutional review board of the NTT Medical Center Tokyo with a waiver for the requirement of opt-in consent from patients (number 000200000816-01). We manually excluded images taken with methods other than white light, images taken outside the body, images obtained with dye and images with airway stents. The remaining images were converted into 400×400-pixel JPEG format, and the extra black margins were removed. The images were divided into a training dataset and a test dataset along the time axis (training dataset: January 2020 to January 2022, test dataset: February 2022 to February 2023). Patients who underwent bronchoscopy in both the training and test periods were excluded from the test cohort to eliminate any potential for information leakage.

Development, training, evaluation and explanation of deep-learning models
We used PyTorch (version 2.0.0) as a library in the Python (version 3.9.16) programming language to develop our deep-learning models. We evaluated four CNN architectures (DenseNet 121, MobileNet V3, EfficientNet B1 and EfficientNet B3 [11–13]). For transfer learning, these models with pre-trained weights on ImageNet were downloaded from the torchvision.models subpackage. The output layer containing 1000 fully connected nodes was replaced with a single final neuron so that the model outputs a single numerical value for the predicted age as a regression model (figure 1). We used the mean squared error as our loss function and trained the models using the AdamW optimiser [14]. We adopted a group k-fold cross-validation method during model training, wherein the training dataset was divided into five groups ensuring that images of each patient were not split across multiple groups. We applied image augmentation during training to improve the generalisability of our model and avoid overfitting. The images used for the training were augmented with a random rotation (up to 10 degrees) and a random perspective transformation (distortion scale=0.1). The trained CNN models were applied to the test dataset to measure the "bronchial age" of each image. We used gradient-weighted class activation mapping (Grad-CAM) to visualise the areas of interest of our models [15].

Statistical analysis
For the statistical analysis, we used the SciPy library (version 1.10.1). We used Pearson's correlation coefficient to analyse correlations and t-test or ANOVA to test for differences in measured bronchial age. We considered p-values <0.05 to indicate statistical significance.

Results
Patient characteristics and image information
A total of 11 374 bronchoscopy images were taken in our hospital between January 2020 and February 2023 (figure 2). Images taken with methods other than white light (endobronchial ultrasonography or narrow band imaging: 121 images), images taken outside the body (28 images), images with dye (virtual-assisted lung mapping: 25 images) and images with airway stents (five images) were excluded. Thus, 11 195 images were included in the analyses. To prevent information leakage, the images were divided into a training dataset (January 2020 to January 2022) and a test dataset (February 2022 to February 2023) along the time axis. Two patients who underwent bronchoscopy in both the training and test periods were excluded from the test cohort. The patient characteristics are shown in table 1. The training dataset and test dataset contained 6413 and 4669 images, respectively. The training cohort and test cohort contained 212 and 122 patients, respectively. ∼70% of the patients in each cohort had a smoking history of >1 pack-year.
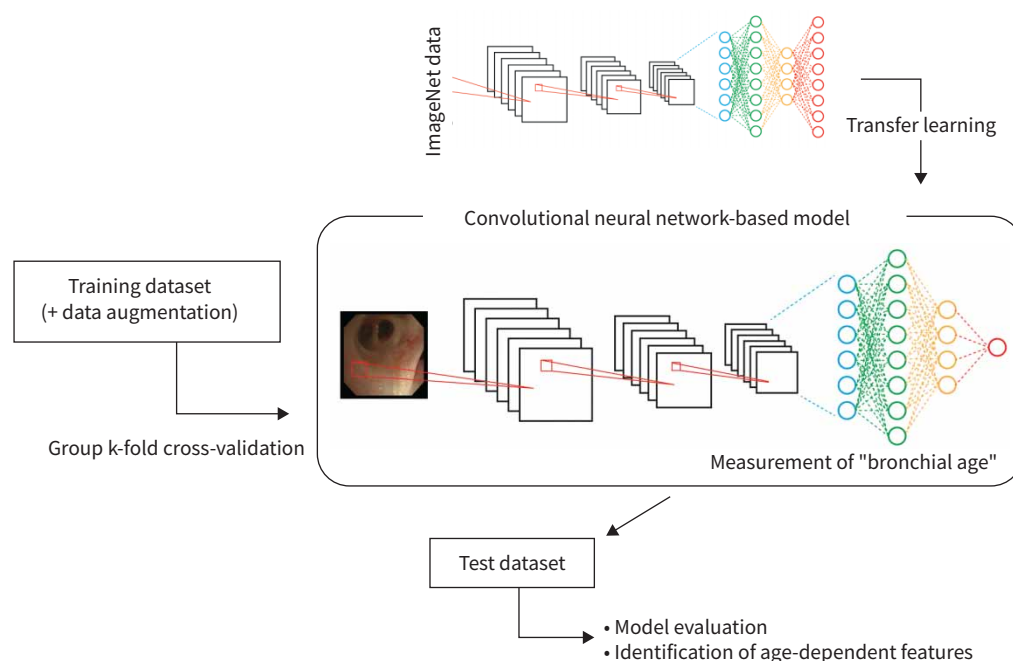
FIGURE 1 Overall study framework. The weights of the models were initialised with pre-trained weights on ImageNet data, followed by transfer learning for further training. The output layer, comprising 1000 fully connected nodes, was replaced with a single final neuron. Our bronchoscopy images were partitioned into training and test datasets along the time axis. Various models with distinct architectures were trained independently. The trained deep-learning models were applied to the test dataset to evaluate these models.

Model training and evaluation

CNNs are a type of deep-learning algorithm that is primarily used for image classification and object recognition tasks. First, we tried to construct a CNN that could estimate patient age from bronchial images obtained by bronchoscopy. We adopted four representative CNN architectures as candidates: DenseNet 121, MobileNet V3, EfficientNet B1 and EfficientNet B3. We applied transfer learning to train these CNNs. After training the models using the training dataset, we next estimated patient age from images in the test dataset (figure 3). In each neural network architecture, the age estimated from bronchoscopy images (bronchial age) showed a significant correlation with chronological age. Among the four architectures, EfficientNet B3 yielded the highest Pearson's correlation coefficient (0.9617, 95% CI 0.9595–0.9638). Therefore, the EfficientNet B3-based model was used for the subsequent analyses.

Interpretation of the deep-learning model and identification of age-dependent features

Next, we applied an XAI method to our trained model to identify age-dependent features of the trachea and bronchi, using Grad-CAM to evaluate the elements on which the model focuses for measuring bronchial age. We visualised the results using a heatmap, which revealed that the model predominantly attended to bronchial bifurcation sites, both in images where it accurately predicted chronological age and in those in which it exhibited discrepancies (figure 4). Furthermore, the model focused attention on bronchial bifurcation sites, regardless of whether its output was higher or lower than the chronological age. These findings indicate that bronchial bifurcation sites are crucial landmarks for the quantification of age-related changes in the trachea and bronchi.

Relationship between bronchial age and lung age

FEV_1, which is measured by spirometry, is typically used as an indicator of age-related changes in lung function. Conversely, the measured FEV_1 value can be used to calculate the corresponding age, which is referred to as the lung age. Thus, we next investigated the relationship between the bronchial age, which is measured from bronchoscopy images using the deep-learning model that we constructed, and the lung age, which is measured by spirometry. Among 122 patients in the test cohort, we analysed 101 patients who underwent spirometry before bronchoscopy. We defined the median bronchial age measured from each
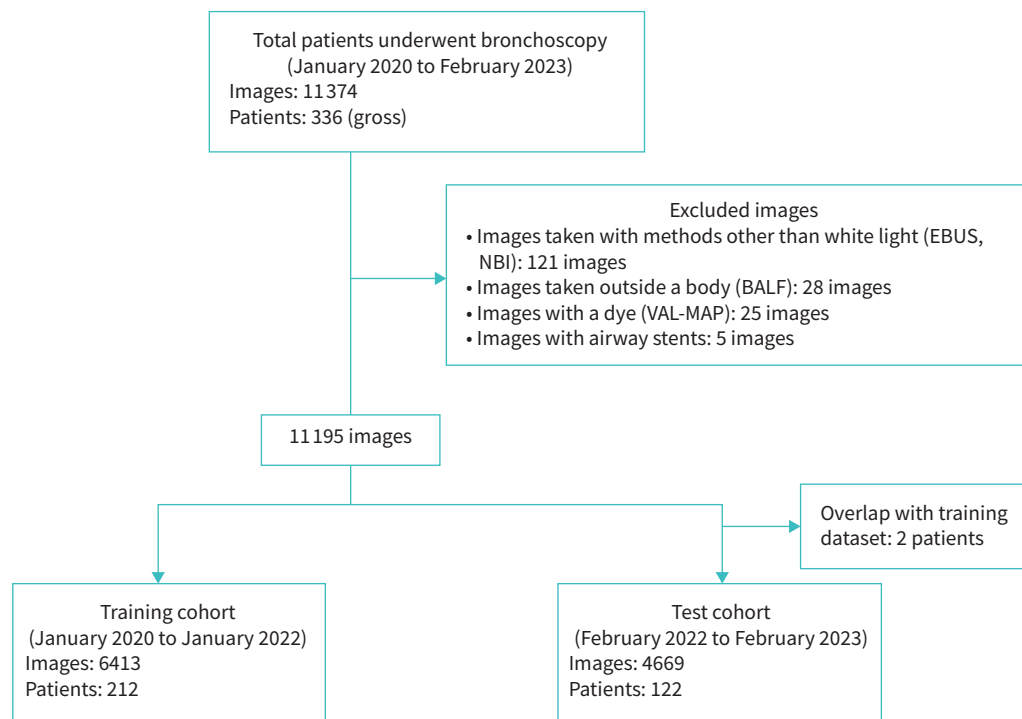
FIGURE 2 Flow diagram of the bronchoscopy images. Bronchoscopy images of patients who received the procedure at our hospital during the period of January 2020 to February 2023 were obtained from our registry. These images were divided into training and test datasets along the time axis (training dataset: January 2020 to January 2022, test dataset: February 2022 to February 2023). Patients who underwent bronchoscopy in both the training and test periods were excluded from the test cohort to eliminate any potential for information leakage. EBUS: endobronchial ultrasound; NBI: narrow-band imaging; BALF: bronchoalveolar lavage fluid; VAL-MAP: virtual-assisted lung mapping.

patient's bronchoscopy images as the patient's bronchial age. A scatter diagram showed a positive correlation between lung age and bronchial age (supplementary figure S1), although the correlation coefficient was lower than that between chronological age and bronchial age. These findings suggest that

TABLE 1 Clinical characteristics of the patients in the training and test cohorts		
	Training cohort	**Test cohort**
Images	6413	4669
Patients	212	122
Chronological age, years	69 (23–93)	72 (21–95)
Male	137 (64.6)	75 (61.5)
Smoking history		
⩽1 pack-year	62 (29.2)	39 (32.0)
>1 pack-year	150 (70.8)	83 (68.0)
Disease		
Lung cancer	125 (59.0)	64 (52.5)
Interstitial pneumonia	15 (7.1)	25 (20.5)
Infection	31 (14.6)	20 (16.4)
Allergy	2 (0.9)	2 (1.6)
Other	39 (18.4)	11 (9.0)
Location of disease		
Right	95 (44.8)	56 (45.9)
Left	72 (34.0)	30 (24.6)
Both lungs	45 (21.2)	36 (29.5)

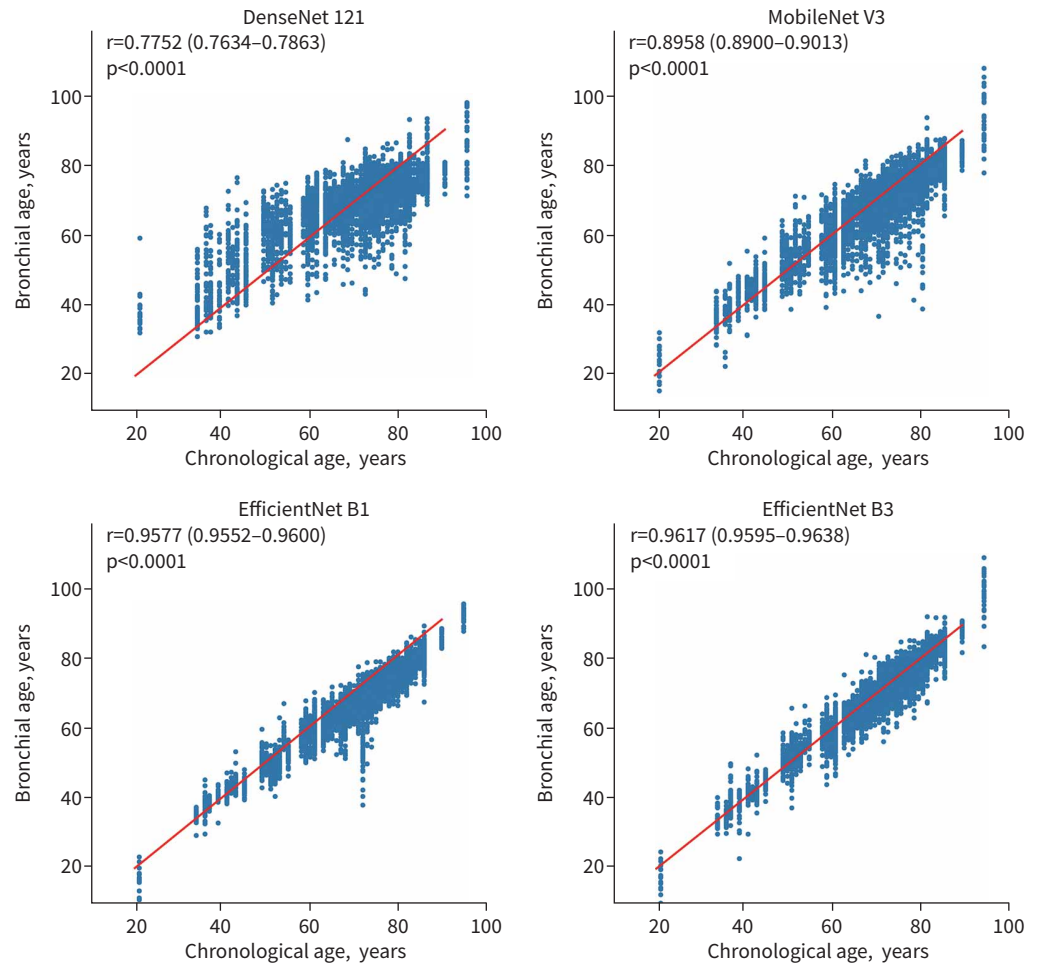Data are presented as n, median (range) or n (%).

FIGURE 3 Accuracy of the trained deep-learning models. The actual chronological age and the predicted bronchial age of each image in the test dataset were plotted on scatter plots. The bronchial ages were calculated utilising four trained models. Pearson's correlation coefficient, along with the 95% confidence interval, were indicated for each model. Red lines signify that the bronchial age is equal to the chronological age.

while there was a certain relationship between the age-related changes in the ventilation function, as represented by FEV_1, and the age-related changes in the bronchial structure, as represented by bronchial bifurcation sites, many other factors are involved in a complex manner.

Universal measurement of bronchial age regardless of respiratory disease or smoking history

Next, we compared the distribution of bronchial age measured by our deep-learning model for each respiratory disease. We classified 122 cases in the test cohort into five groups based on their definitive diagnoses: lung cancer, interstitial pneumonia, infection, allergy and others (table 1). While we observed slight differences in the measured bronchial age between groups (p=0.0357), we did not find significant differences in the discrepancy between bronchial age and chronological age (p=0.1258; figure 5a). These results suggest that our constructed deep-learning model can measure bronchial age accurately across a wide range of respiratory diseases and is not limited to a specific disease.

We also investigated whether smoking history affects the calculation of bronchial age by our deep-learning model. We classified 122 cases in the test cohort into two groups based on whether they had a smoking history of >1 pack-year (table 1) and compared the bronchial age between the two groups (figure 5b). There was no significant difference in bronchial age between the two groups (p=0.3017), and we did not observe a significant difference in the discrepancies between bronchial age and chronological age (p=0.9540). These findings suggest that our constructed model can accurately measure bronchial age regardless of smoking history.
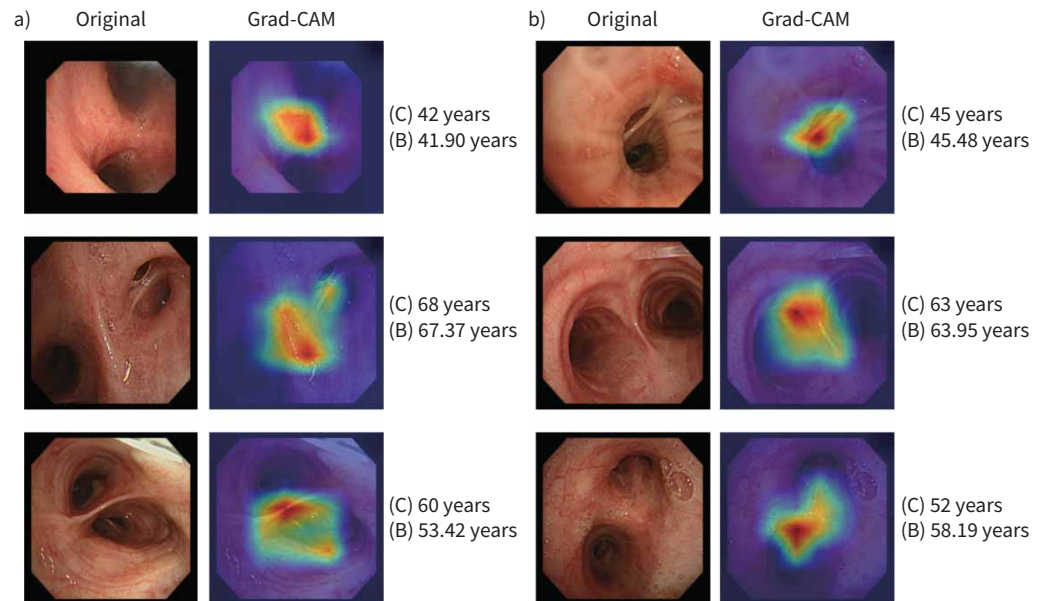
FIGURE 4 Visualisation of age-dependent features of the trachea and bronchi identified by our model. Examples of the original bronchoscopy images and their corresponding heatmap visualisation using gradient-weighted class activation mapping (Grad-CAM). The actual chronological age (C) and predicted bronchial age (B) of each image are also indicated. The examples include images with both a) lower and b) higher bronchial age than chronological age. They also include images where the model accurately predicted the chronological age (upper four cases) and those where it exhibited discrepancies (lower two cases).

Our analyses demonstrated that the constructed deep-learning model can estimate bronchial age from bronchoscopy images regardless of respiratory disease type or smoking history, suggesting the universal importance of bronchial bifurcation sites as age-dependent features.

Discussion

In this study, we constructed a deep-learning model to estimate the chronological age from bronchial images obtained by bronchoscopy, and the model uncovered the importance of bifurcation sites as age-dependent features in the trachea and bronchi. Although changes in the tracheal bifurcation angle during the growth period in children have been reported [16], to the best of our knowledge, there have been no reports on the quantitative relationship between bronchial bifurcation sites and ageing in adults.

Bronchoscopy has become increasingly important in the field of precision medicine (*e.g.* for lung cancer and interstitial pneumonia). Although this practice has led to the increased generation and curation of huge bronchoscopy imaging data, at the present time the analysis and interpretation of these data is largely immature. One possible answer for this problem is the introduction of artificial intelligence for the analysis of bronchoscopy data. Recent research on the analysis of images by artificial intelligence has progressed in the areas such as radiographs [17], pathological slides [18] and ophthalmic images [19]. Although some studies have developed anatomical interpretation models or bronchoscopic navigation systems that are aided by artificial intelligence [20, 21], the analysis of bronchoscopic findings by artificial intelligence is still immature. As a next step, our study expands the possibility of deep learning in the analysis of bronchoscopy images by providing strong evidence that it can be used to extract age-dependent features from bronchoscopy images.

Numerous studies have reported on the relationship between ageing and the function or structure of organs [2]. Furthermore, through these studies, feature extraction has been performed to capture the age-dependent changes in the function and structure of each organ. For example, the balance between bone resorption and formation changes with ageing, resulting in age-related changes in the bone function and structure. Bone mineral density has been utilised as an age-dependent feature to capture these changes [22]. With the development of machine learning, some studies have used machine learning to extract age-dependent features. A deep-learning model that can estimate the age of adults from chest
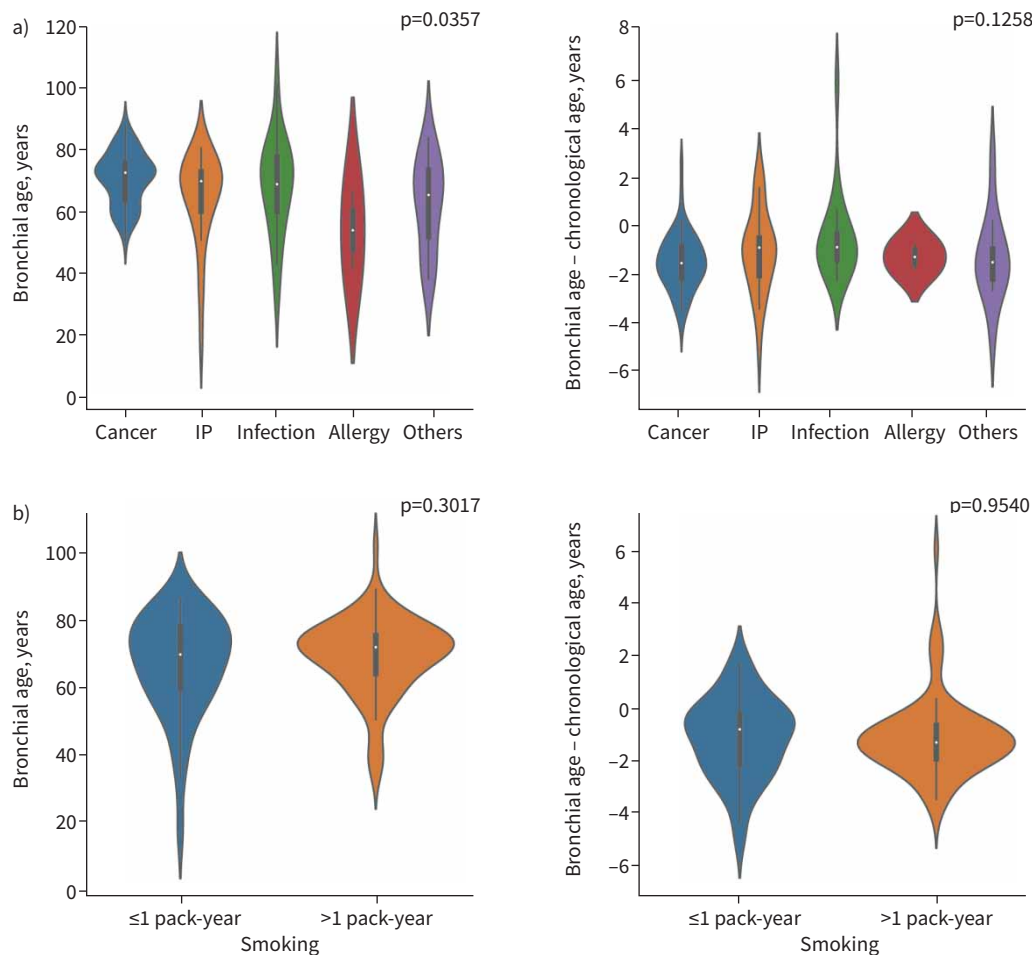
FIGURE 5 Differences in the distribution of bronchial age according to type of respiratory disease or smoking history. a) Violin plots show the distribution of bronchial age or discrepancies between bronchial age and chronological age in patients with lung cancer, interstitial pneumonia (IP), infection, allergy or other conditions. ANOVA was used to test for differences between the different respiratory diseases. b) The test cohort was divided into two groups based on their smoking history of >1 pack-year, and the bronchial age or difference between bronchial age and chronological age were compared between the two groups. The t-test was used to test for differences between the two groups.

radiography images has been reported [6]. In this study, the model mainly focused on the top of the mediastinum, irrespective of the pathology present in the chest radiography images, suggesting that tortuosity and calcification of the aorta are important age-dependent features in chest radiographs. Similar studies have been performed for magnetic resonance imaging of the brain and 12-lead electrocardiography [5, 23]. This is the first deep-learning model to calculate bronchial age, and the application of an XAI method to our model identified bronchial bifurcation sites as age-dependent features of the trachea and bronchi. Machine-learning models without human specification of features sometimes extract features that humans have not anticipated. Bifurcation sites may be one such feature in the ageing of the trachea and bronchi. The significance of bronchial bifurcation sites was demonstrated, irrespective of their position or imaging direction. Thus, it is assumed that our model captures features of the mucosal surface of bronchial bifurcation sites rather than structural characteristics (*e.g.* the angle of bifurcation sites). One plausible biological interpretation is that atrophy of the mucosa at bronchial bifurcation sites reflects age-dependent alterations in the trachea and bronchi. During the development of the respiratory system, repeated budding starting from the lung bud creates an intricate network of bronchi [24]. In addition, bifurcation sites are subjected to various stresses, including mechanical stress [25]. Given these facts, it is reasonable that bronchial bifurcation sites are susceptible to age-related deterioration and are age-dependent features. Further studies are needed to definitely identify which parameters of bronchial bifurcation sites are quantitatively captured. For example, synthetic histology generated by a conditional generative adversarial

network identified histological parameters associated with molecular state of tumours [26]. Such algorithms can be applied to bronchoscopy images.

The discrepancy between bronchial age and chronological age showed no significant difference among different respiratory diseases or according to smoking history, suggesting the universal importance of bronchial bifurcation sites as age-dependent features. Our training dataset includes patients with a wide variety of respiratory diseases, and it can therefore be considered that age-dependent features that are not specific to particular lung diseases were extracted in the training. Conversely, our training dataset did not include bronchial images of "normal" humans. Therefore, our analysis does not cover comparisons of normal and abnormal bronchi, and our model does not reflect the normal status of the trachea and bronchi.

CNNs are commonly used for image recognition tasks in deep learning. In addition, recently, Transformer, which has demonstrated high performance in natural language processing, has been applied to image recognition [27]. Transformer-based models have achieved state-of-the-art results in various tasks, such as image classification [28]. In our study, we also attempted to construct Transformer-based models such as Vision Transformer and Swin Transformer [29], as well as CNNs. However, these Transformer-based models did not yield a higher correlation coefficient than EfficientNet.

The present study was associated with several limitations. First, the study was performed using data from a single institution, and our model did not reflect differences in bronchoscopes or operators. However, we applied image augmentation during training to improve the generalisability of our model and avoid overfitting. Second, we collected image data retrospectively. Although a prospective validation study might be needed to examine the generality of our model, our images were divided into training and test datasets along the time axis. Finally, as mentioned earlier, the training dataset consisted of images of patients with some respiratory disease. In previous studies of other age-related markers, such as lung age (FEV_1), bone age (bone mineral density) and vascular age (pulse wave velocity), datasets from healthy volunteers were used to create reference values. In a clinical setting, bronchoscopy is only performed for individuals with suspected respiratory disease. Collecting bronchoscopy images from healthy volunteers to create a dataset of healthy individuals is ethically challenging due to the invasiveness of bronchoscopy. Furthermore, to the best of our knowledge, there is no public dataset of bronchoscopy images, unlike chest radiographs. Therefore, we divided our images into training and test datasets as an alternative method.

In conclusion, we constructed a deep-learning model that can estimate the chronological age from images obtained by bronchoscopy, and the application of an XAI method to our model revealed the importance of bifurcation sites as age-dependent features in the human adult trachea and bronchi. Although deep-learning methodology has been applied to bronchoscopy for location guidance [20, 21], our results suggest that an analysis based on deep-learning models can also be applied to biological evaluation of bronchoscopy images in the era of digital medicine. Further studies with various conditional datasets will be required to analyse the pathological significance of bronchial bifurcation sites.

References

1   López-Otín C, Blasco MA, Partridge L, et al. The hallmarks of aging. Cell 2013; 153: 1194–1217.

2   Kennedy BK, Berger SL, Brunet A, et al. Geroscience: linking aging to chronic disease. Cell 2014; 159: 709–713.

3   Sasaki E, Nakamura M, Kida K, et al. Reference values for spirogram and blood gas analysis in Japanese non-smoking healthy adults. J Jpn Respir Soc 2001; 39: 383–399.

4   Seeman E, Delmas PD. Bone quality – the material and structural basis of bone strength and fragility. N Engl J Med 2006; 354: 2250–2261.

5   Wang J, Knol MJ, Tiulpin A, et al. Gray matter age prediction as a biomarker for risk of dementia. Proc Natl Acad Sci USA 2019; 116: 21213–21218.

6   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.

7   Hannallah MS, Benumof JL, Ruttimann UE. The relationship between left mainstem bronchial diameter and patient size. *J Cardiothorac Vasc Anesth* 1995; 9: 119–121.

8   Sharma G, Goodwin J. Effect of aging on respiratory system physiology and immunology. *Clin Interv Aging* 2006; 1: 253–260.

9   Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. *N Engl J Med* 2023; 388: 1201–1208.

10   Gunning D, Stefik M, Choi J, *et al.* XAI – explainable artificial intelligence. *Sci Robot* 2019; 4: eaay7120.

11   Huang G, Liu Z, van der Maaten L, *et al.* Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017; pp. 4700–4708.

12   Howard A, Sandler M, Chen B, *et al.* Searching for MobileNetV3. *Proceedings of the IEEE/CVF International Conference on Computer Vision* 2019; pp. 1314–1324.

13   Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research 2019; 97: 6105–6114.

14   Loshchilov I, Hutter F. Decoupled weight decay regularization. *arXiv* 2019; preprint [https://doi.org/10.48550/arXiv.1711.05101].

15   Selvaraju RR, Cogswell M, Das A, *et al.* Grad-CAM: visual explanations from deep networks *via* gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision 2017; pp. 618–626.

16   Herek D, Herek O, Ufuk F. Tracheobronchial angle measurements in children: an anthropometric retrospective study with multislice computed tomography. *Clin Exp Otorhinolaryngol* 2017; 10: 188–192.

17   Nam JG, Hwang EJ, Kim J, *et al.* AI improves nodule detection on chest radiographs in a health screening population: a randomized controlled trial. *Radiology* 2023; 307: e221894.

18   Liu MY, Chen TY, Williamson DFK, *et al.* AI-based pathology predicts origins for cancers of unknown primary. *Nature* 2021; 594: 106–110.

19   Abràmoff MD, Cunningham B, Patel B, *et al.* Foundational considerations for artificial intelligence using ophthalmic images. *Ophthalmology* 2022; 129: e14–e32.

20   Yoo JY, Kang SY, Park JS, *et al.* Deep learning for anatomical interpretation of video bronchoscopy images. *Sci Rep* 2021; 11: 23765.

21   Sganga J, Eng D, Graetzel C, *et al.* Autonomous driving in the lung using deep learning for localization. *arXiv* 2019; preprint [https://doi.org/10.48550/arXiv.1907.08136].

22   Lane NE. Epidemiology, etiology, and diagnosis of osteoporosis. *Am J Obstet Gynecol* 2006; 194: S3–S11.

23   Attia ZI, Friedman PA, Noseworthy PA, *et al.* Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circ Arrhythm Electrophysiol* 2019; 12: e007284.

24   Morrisey EE, Hogan BL. Preparing for the first breath: genetic and cellular mechanisms in lung development. *Dev Cell* 2010; 18: 8–23.

25   Salzar RS, Thubrikar MJ, Eppink RT. Pressure-induced mechanical stress in the carotid artery bifurcation: a possible correlation to atherosclerosis. *J Biomech* 1995; 28: 1333–1340.

26   Dolezal JM, Wolk R, Hieromnimon HM, *et al.* Deep learning generates synthetic cancer histology for explainability and education. *NPJ Precis Oncol* 2023; 7: 49.

27   Henry EU, Emebob O, Omonhinmin CA. Vision transformers in medical imaging: a review. *arXiv* 2022; preprint [https://doi.org/10.48550/arXiv.2211.10043].

28   Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16×16 words: transformers for image recognition at scale. *arXiv* 2020; preprint [https://doi.org/10.48550/arXiv.2010.11929].

29   Liu Z, Lin Y, Cao Y, *et al.* Swin transformer: hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision 2021; pp. 10012–10022.