AI in thoracic imaging

Akash

SR, pulmonary medicine, PGIMER

Overview:

- What is Artificial intelligence?

- How might it be helpful in medicine?

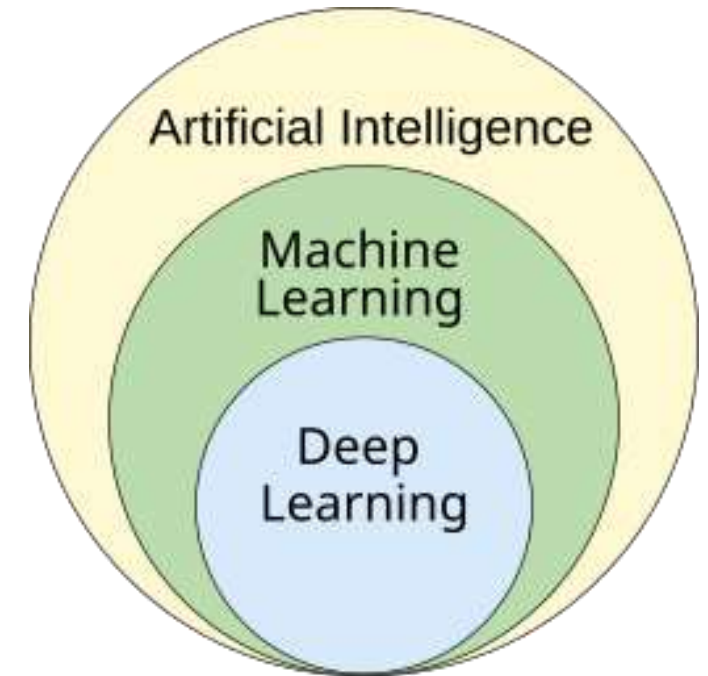- What are the current advances in thoracic imaging with the help of AI?

- How accurate is it?

- Is it an adjunctive method or will it replace diagnosticians?

Artificial intelligence

- AI can be defined as the ability of computers to perform task that normally requires human intelligence.

- Machine learning is a subfield of AI in which statistical models are used to learn patterns from data in order to accomplish a specific task

- Artificial neural networks (ANNs) are loosely modelled on the human brain and consist of multiple layers of 'neurons' that successively process input data until the output layer is reached. Deep neural networks are a more complex network with usually more than 10 intermediate layers (DNN).

Gonem S, Janssens W, Das N, Topalovic M. Applications of artificial intelligence and machine learning in respiratory medicine. Thorax. 2020 May 14;75(8):695–701

AI in medical field

Drug development and design

Development of precision therapeutic approaches

Predictive models for specific diseases: PTE, lung cancer, coronary artery diseases

Augmentation of precision in robotic surgery and invasive therapeutic procedures

Interpretation of histopathological imaging

Interpretation of radiological imaging

Aiding in medical decision making (holistic approach with all the information)

Fields of pulmonary medicine AI is used in

- Thoracic imaging for lung nodule detection and ILD classification

- Prediction of survival or disease progression from histopathological imaging

- Development of risk prediction models for pulmonary thromboembolism and pulmonary hypertension

- Composite interpretation of pulmonary function tests to predict risk of disease progression or exacerbation in obstructive and restrictive lung diseases
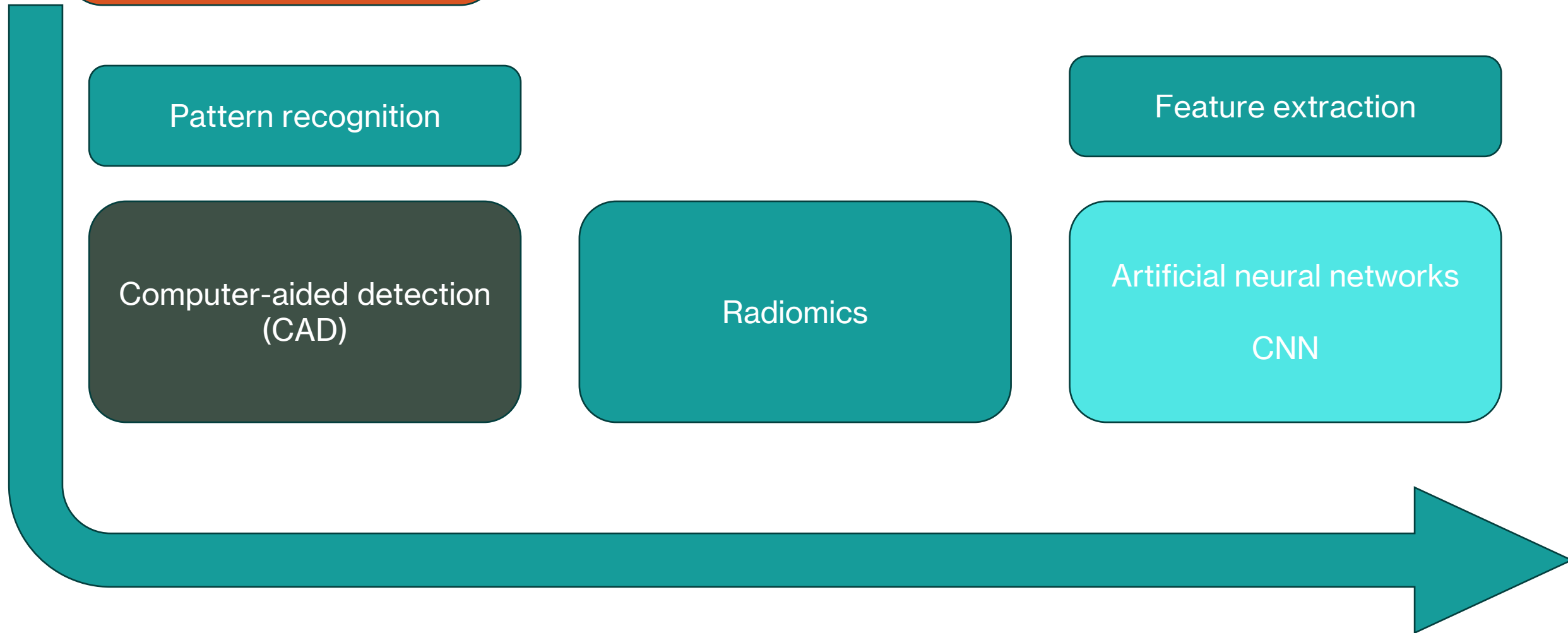
Computer aided detection (CADe and CADx)

- The computer is fed data regarding 'what is abnormal' (training)

- On subsequent instances, on being fed raw non-labelled data (images) computer predicts parts of the new data that might be 'abnormal' according to its training

- Points out such areas of suspected abnormality to radiologist

- Radiologist reviews those areas and decide whether they are actually abnormal or not

- **Principle**: sophisticated pattern recognition
- **Prerequisite**: thousands of normal and abnormal images labelled
- **Steps:** removes bugs, adjusts for variabilities of exposure
- **Output:** detects abnormalities in imagings likely to be missed by human eyes
- **Reliability**: radiologist/pathologist takes the final decision regarding the accuracy of detection
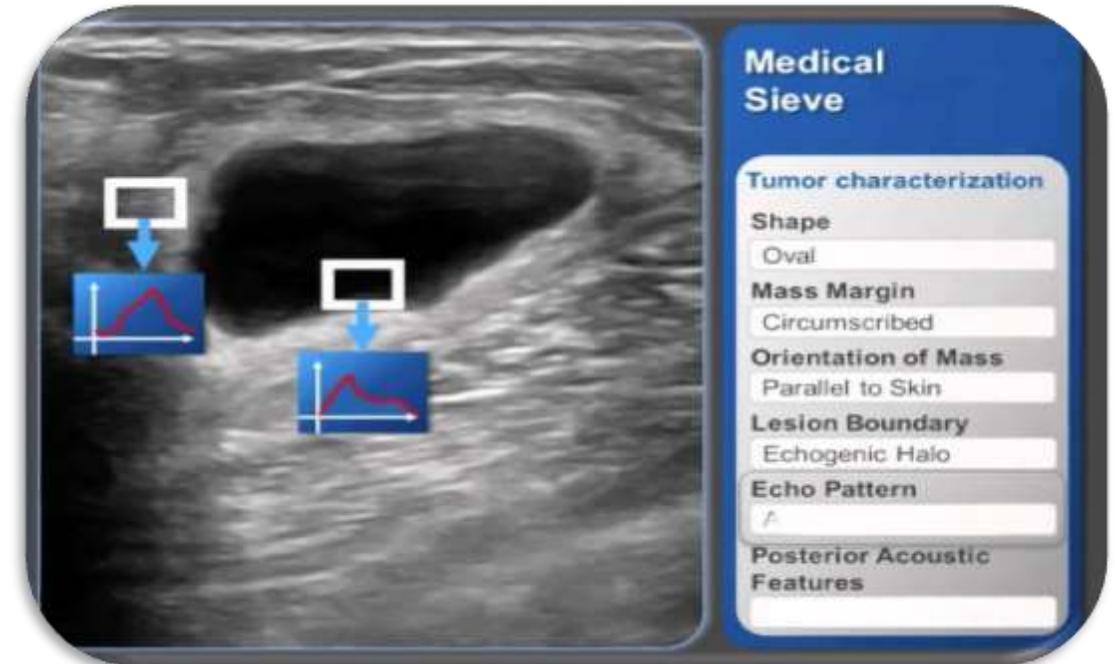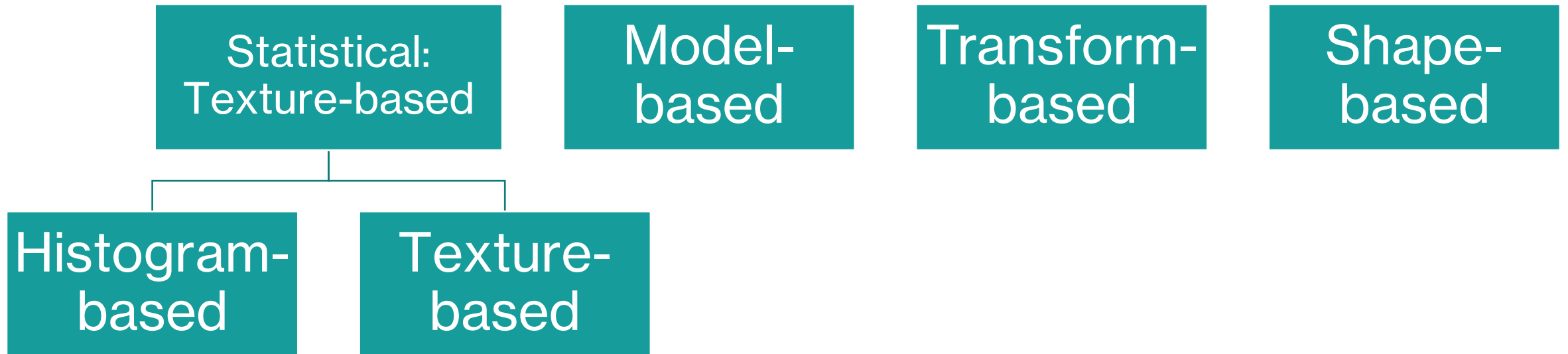
Radiomics:

- Radiomics generally aims to extract quantitative, and ideally reproducible, information from diagnostic images, including complex patterns that are difficult to recognize or quantify by the human eye.

- It can identify with considerable accuracy, the change in size, shape and amount of heterogeneity in a tumor that may be missed by human eyes.

- In a large cohort of patients, it has the ability to identify previously unknown trends of disease progression, evolution and response to treatment, which would be impossible to be identified by human brain just due to the large size of the data set and innumerability of the variables.

Terminology Used in Radiomics and AI	
Artificial intelligence	Wide-ranging branch of computer science, generating complex software that perform tasks that would typically have required human intelligence, by sensing and responding to a feature of their environment.
CAD (Computer Aided Detection or Diagnosis)	Technology combining elements of artificial intelligence with radiological and pathology image processing. Its aim is to assist in the detection and/or diagnosis of diseases, improving the accuracy of radiologists with a reduction in time in the interpretation of images.
Radiomics	Method that extracts a large number of quantitative features from radiographic medical images using data-characterization algorithms, to help in disease diagnosis and prognosis.
Machine Learning	Field in artificial intelligence studying computer algorithms that improve automatically through experience, by building a model based on sample data, known as "training data", in order to make predictions or decisions. Supervised learning: The computer receives example inputs and their foreseen outputs. Its goal is to learn a general and reproducible function that links inputs to outputs. Unsupervised learning: The computer receives no labels to the learning algorithm for previously undetected patterns in a data set, leaving it on its own to find structure in its input.
Convolutional neural networks	Class of deep neural networks, which have the particularity of being fully connected networks. It gives them the advantage of understanding the hierarchical pattern in data and assembling more complex patterns using smaller and simpler patterns.
Voxel	Single sample, or data point, on a regularly spaced, three-dimensional grid. In CT scans, the values of voxels are Hounsfield units. A voxel is a 3D pixel.
ROI (Region of Interest)	Image areas containing the information relevant to image processing.
Skew of histogram	Measure of the asymmetry of attenuation distribution. The lung normal attenuation histogram is skewed to the left. There is a decreased leftward skewness in IPF.
Kurtosis of histogram	Measurement of how sharp an attenuation distribution curve is. Kurtosis is abnormally low in idiopathic pulmonary fibrosis (IPF).
Threshold measurement	Total count of pixels/voxels above or below a specific attenuation value that determines a relative volume. Threshold measures in emphysema quantifies the extent of emphysema according to a specific index of -950 Hounsfield units (HU).
Texture analysis	Statistical methods that evaluate spatial relationship between voxels in an ROI, in order to characterize textural features of the parenchyma and give information about heterogeneity.

Radiomics features (how they interpret images)

Statistical:
Texture-based

Model-
based

Transform-
based

Shape-
based

Histogram-
based

Texture-
based

Histogram-based approach

- These are mainly calculations of variances in terms of grayscale intensity, deviation (skewedness) from the mean and change in that intensity at different areas at specified distances from ROI (region of interest)

- These gives the idea of heterogeneity inside a tumor, how varied the pixels and voxels are inside that tumor

- A high degree of heterogeneity has been shown to be associated with higher level of aggressiveness of a tumor, progression of disease or absence of response to treatment
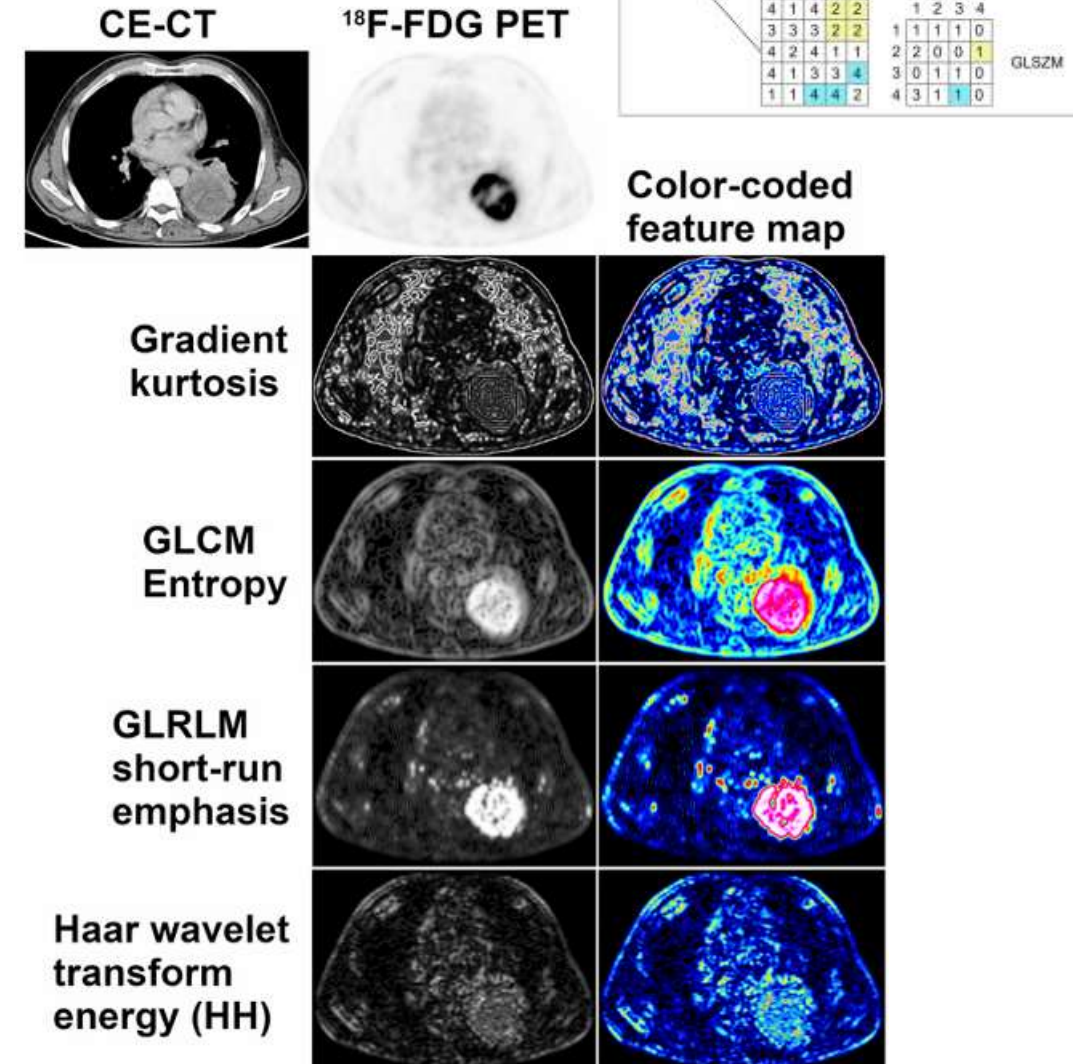
Model-based approach

- This feature analyzes gray-level special information to characterize objects or shapes

- A parameterized model of texture generation is calculated and fitted to the ROI, and its estimated parameters are used as radiomic features

- Fractal analysis yields features in fractal dimensions which reflects the rate of addition of structural details with increasing magnification, scale and resolution reflecting increasing complexity

- Lacunarity detects lack of rotational or translational invariance therefore giving an idea regarding increasing inhomogeneity

Transform-based approach

- Fourier, Gabor, and Haar wavelet transforms, analyze gray-level patterns in a different space

- Wavelet de composition of an image is possible by applying a pair of so-called quadrature mirror filters, a high-pass and a low-pass filter

- High-pass filtering in both directions captures diagonal details

- High-pass filtering followed by low-pass filtering captures vertical edges

- Low-pass filtering followed by high-pass filtering captures horizontal edges, and low-pass filtering in both directions captures the lowest frequencies, at different scales

- Wavelet transformation can be used not only for generation of radiomic features but also for image segmentation or as a pre-processing step to texture analysis

Shape-based approach

- Describe geometric properties of ROIs

- Features include compactness and sphericity

- These describe how the shape of an ROI differs from that of a circle (for 2D analyses) or a sphere (for 3D analyses), and density

- This relies on the construction of a minimum oriented bounding box (or rectangle for 2D analyses) enclosing the ROI
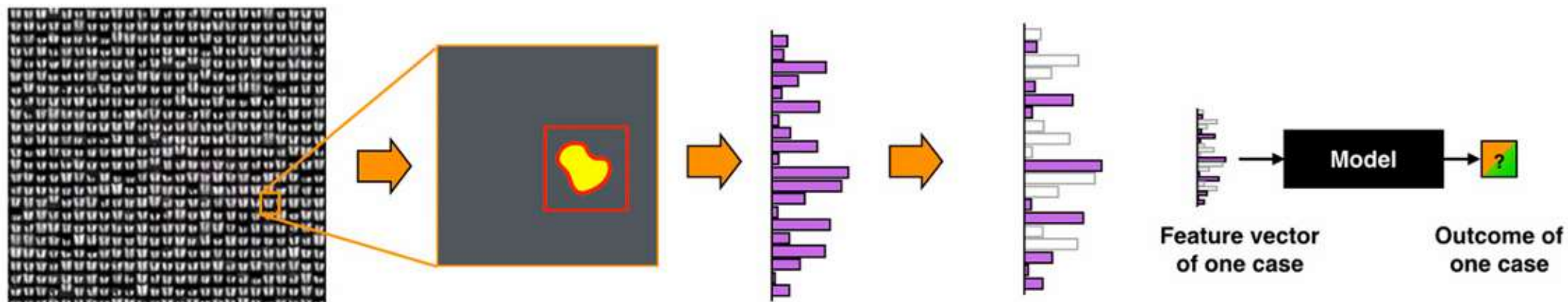
Training data | Select region of interest and segmentation of lesion | Calculate feature candidates | Select features | Predict

Feature vector of one case

Model

Outcome of one case

Select features to represent variability in the data well

Select features to serve prediction well

Construct features to optimally represent data for prediction

Option 1

Option 2

Option 3

Application in clinical pulmonary medicine

Lung nodule characterization

Lung cancer detection

Interstitial lung diseases

Infection: tuberculosis

Pulmonary vascular diseases

Obstructive airway diseases

Pleural diseases

Segmentation

Image acquisition

TRADITIONAL RADIOMICS : HANDCRAFTED FEATURES PROCESSING

Histogram   Shape features

Textural features   Wavelets   Deep features

Feature extraction

Feature selection

Classification

DEEP LEARNING

INPUT

OUPUT

Input layer

Hidden layers : *convolutional layers* for feature extraction
+ *pooling layers* for feature aggregation

Output layer

How does deep neural networks work?

- Deep neural networks are developed in a way to resemble the human neuronal connections.

- In human beings, neurons are interconnected, and these connections improve the analysis of inputs and refine the outcome in response to that input

- Deep neural networks work in a similar way

- There are input layers and output layers

- In between these two layers there are numerous hidden layers neural interconnections

- The dataset given as input is analyzed and according to the training provided and self-training, the dataset is analyzed through the meshwork of 'neurons' to give a result

Deep Neural Network

out layer   hidden layer 1   hidden layer 2   hidden layer 3

output layer
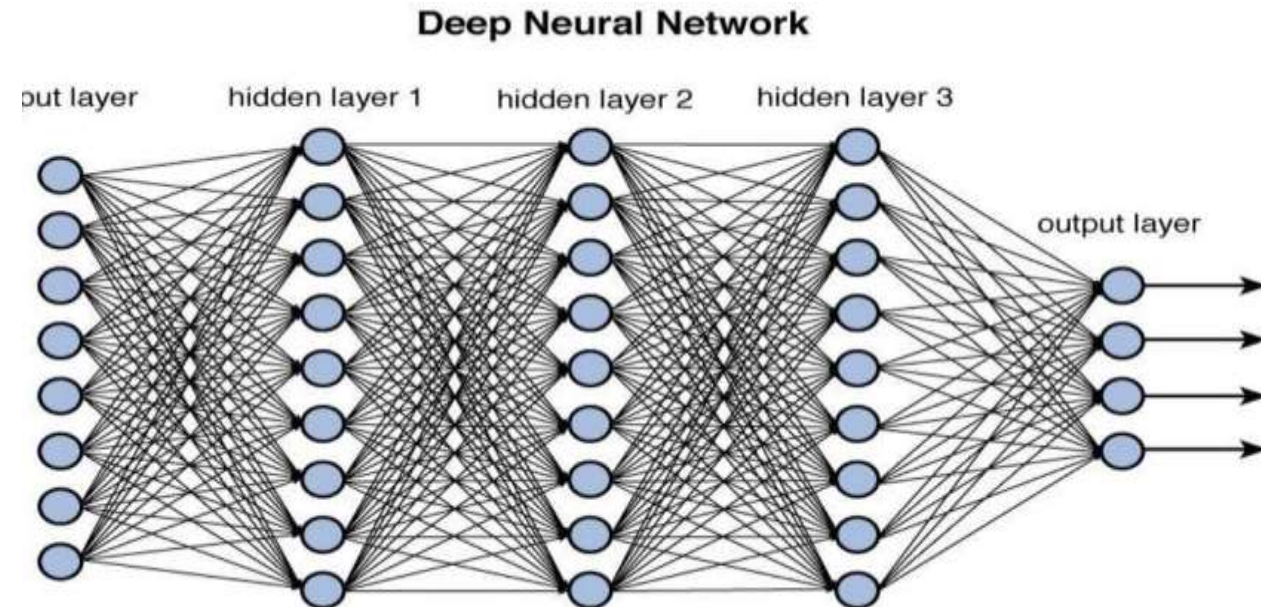
Figure 12.2 Deep network architecture with multiple layers.

Convolutional neural network

- Different types of artificial neural networks are used for different purposes

- Convolutional neural networks are especially useful for classifying and identifying specific features in a large dataset of images
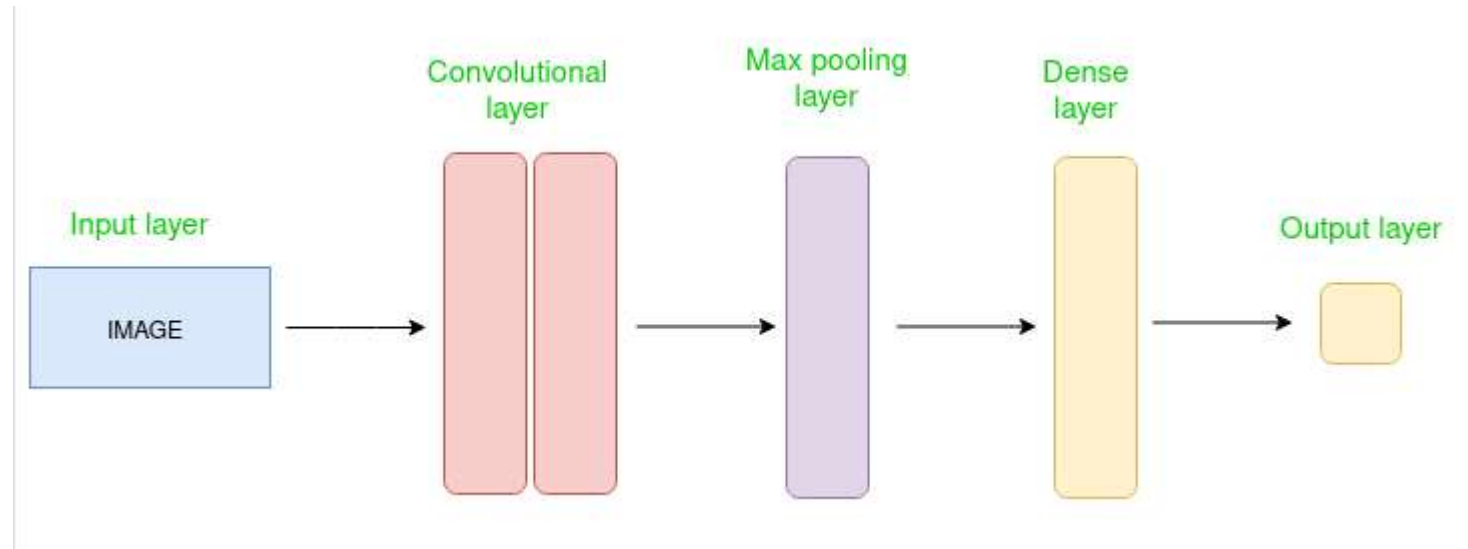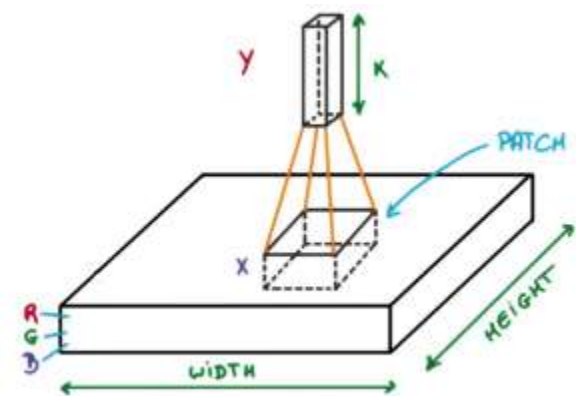
The Architecture of Convolutional Neural Networks

Input   Convolution   Pooling   Fully Connected   Output

Feature Extraction

Classification

RELU RELU RELU RELU RELU RELU
POOL POOL POOL
CONV CONV CONV CONV CONV CONV
FC

car
truck
airplane
ship
horse
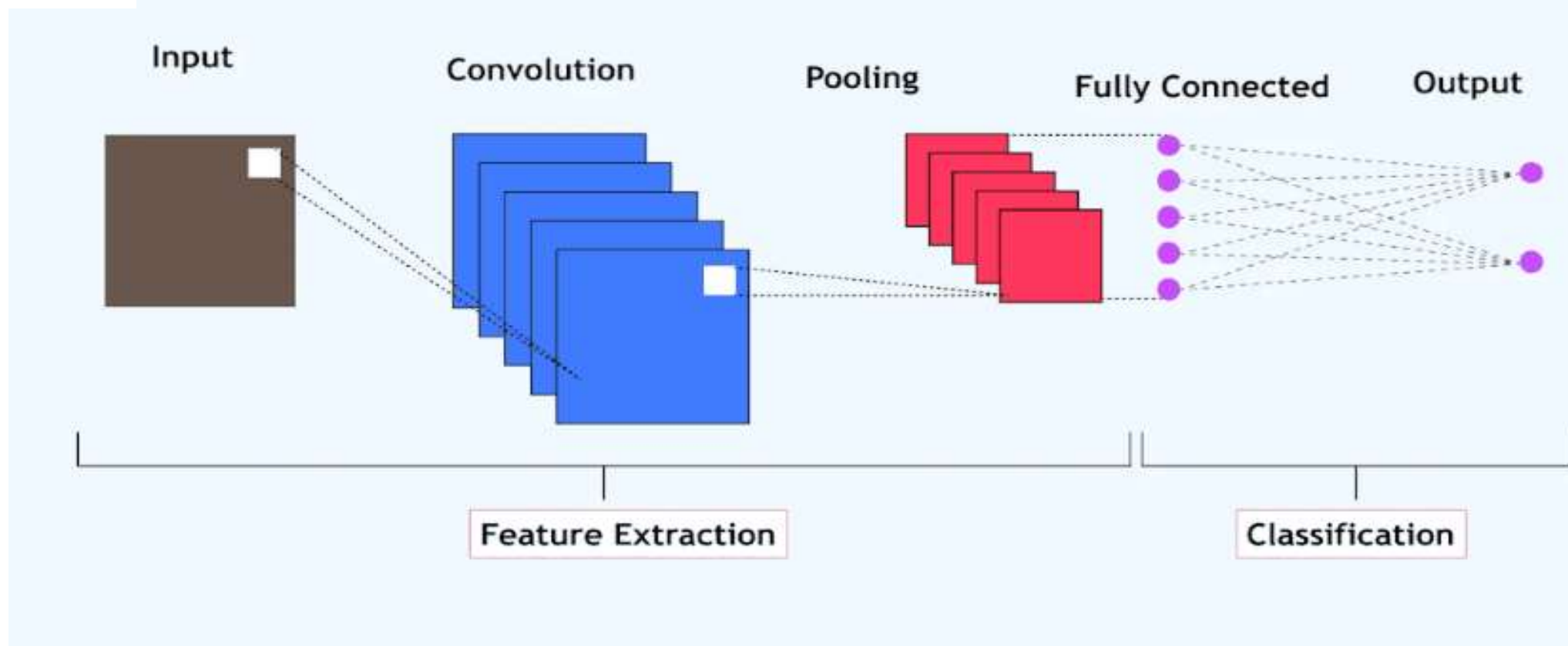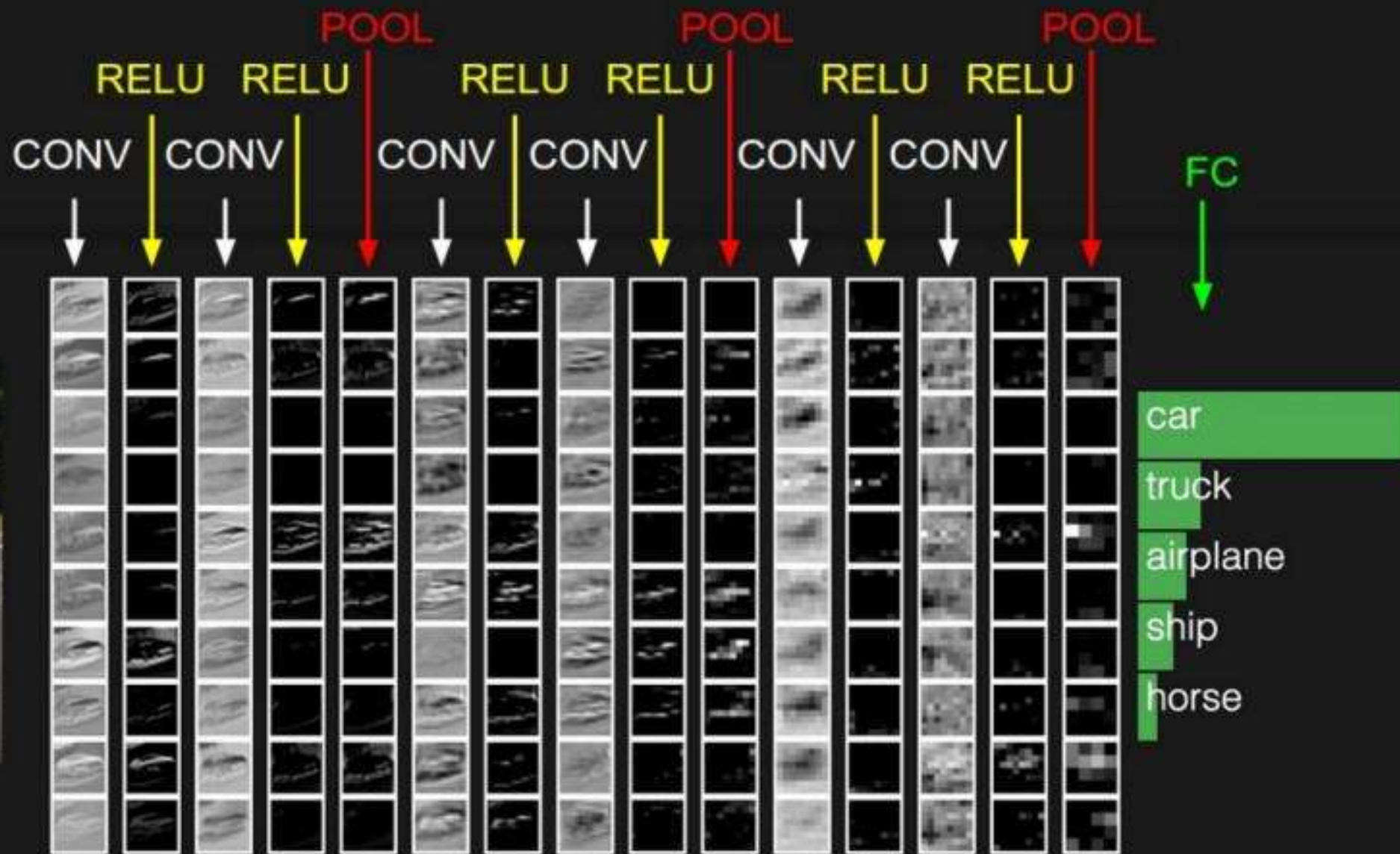
- Advantages of CNN

1. Good at detecting patterns and features in images, videos, and audio signals.

2. Robust to translation, rotation, and scaling invariance.

3. End-to-end training, no need for manual feature extraction.

4. Can handle large amounts of data and achieve high accuracy.

- Disadvantages of CNN:

1. Computationally expensive to train and require a lot of memory.

2. Can be prone to overfitting if not enough data or proper regularization is used.

3. Requires large amounts of labeled data.

4. Interpretability is limited, it's hard to understand what the network has learned.

Pulmonary nodules and lung cancer:

- Traditional clinico-radiological approaches use pre-existent risk factors and mostly size and progression of size of a pulmonary nodule to predict its malignant potential and prognostication

- It requires repeat imaging at specific intervals

- It is still an imperfect method

- Radiomics not only detects and analyzes size and shape of nodules to predict the malignant potential, also additionally predicts histological types and prediction and evaluation of treatment responses

In a 2017 study, CT images of 76 patients with lung nodules were collected and image segmentation was done (biopsy reports of the patients were known)

760 radionomic signatures were identified and they were tested between the two groups (benign and malignant nodules) to see which of them were significantly different between the two groups

Among those that were different (P<0.05), best 4 features were selected and applied in the radionomic analysis of the same dataset (as comparators, all 760 features and a randomly selected set of 4 features were also applied)

Chen CH, Chih Han Chang, Chih Yen Tu, Liao WC, Wu BR, Kuei Ru Chou, et al. Radiomic features analysis in computed tomography images of lung nodule classification. PLOS ONE. 2018 Feb 5;13(2):e0192002–2.

Radiomic features analysis in computed tomography images of lung nodule classification

Chia-Hung Chen[1©], Chih-Kun Chang[2©], Chih-Yen Tu[1], Wei-Chih Liao[1], Bing-Ru Wu[1], Kuei-Ting Chou[3], Yu-Rou Chiou[4], Shih-Neng Yang[3,4], Geoffrey Zhang[5], Tzung-Chi Huang[3,4,6]*

1 Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, China Medical University Hospital, Taichung, Taiwan, 2 Department of Medical Imaging, Chang Bing Show Chwan Memorial Hospital, Changhua, Taiwan, 3 Department of Biomedical Imaging and Radiological Science, China Medical University, Taichung, Taiwan, 4 Artificial Intelligence Center for Medical Diagnosis, China Medical University Hospital, Taichung, Taiwan, 5 Department of Radiation Oncology, Moffitt Cancer Center, 12902 USF Magnolia Drive, Tampa, FL, United States of America, 6 Department of Bioinformatics and Medical Engineering, Asia University, Taichung, Taiwan

Fig 1. Radiomics analysis workflow. First, the clinical CT images of malignant and benign pulmonary nodules were collected. Second, image segmentation was used to delineate the pulmonary nodules. Next, the image features were extracted by the automated high-throughput feature analysis algorithm. Finally, the statistical analysis was applied and the sequential forward search was used for feature selection for the classification of lung nodules.

- When the classifier produced results, the accuracy for the selected 4 radiomic features were 84%

- Sensitivity and specificity (biopsy being the gold standard) were 92.85% and 72.73% respectively

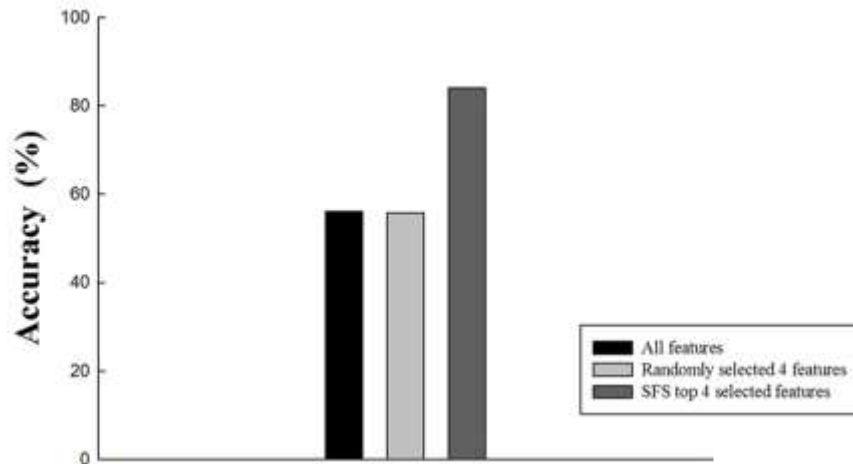- Accuracy for the randomly selected groups were about 55%
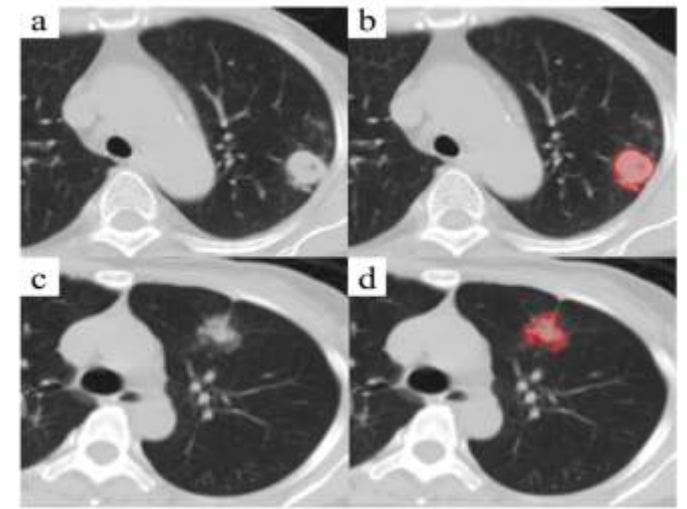

a   b
c   d

Table 3. Radiomics feature list that had significant difference (p<0.05) between malignant and benign groups.

Category of feature	Filter associated	Feature name	#
Intensity based features			5
	N/A	minI, maxI, meanI, Kurtosis, 130	
Wavelet based features			4
	LLH	min	
	LHH	min	
	HLH	contrast	
	HHL	lcl homo	
Laws features			65
	EEL	uniformity	
	EES	max, SD, RMS, energy	
	ELL	Kurtosis, energy, entropy	
	ELS	max, mean, SD, RMS	
	ESE	max, mean, SD, Coeff Vari, RMS, contrast, lcl homo	
	ESL	max, SD, RMS	
	ESS	max, SD, RMS	
	LEL	Kurtosis	
	LES	max, SD, RMS, energy	
	LLE	Peak, mean, Kurtosis	
	LLL	Kurtosis, energy, uniformity	
	LLS	min	
	LSE	max, SD, RMS	
	LSL	min, SD, Skewness, energy	
	LSS	max, SD, Coeff Vari, RMS, energy	
	SES	max, SD, RMS	
	SLL	energy	
	SLS	SD, RMS	
	SSE	max, SD, RMS	
	SSE	max, SD, RMS	
	SSL	Skewness, CV, energy	
	SSS	max, SD, Skewness, RMS	
Run-length based features			2
	N/A	LGRE, SRLGE	

Table 2. Radiomics feature characteristics.

Radiomics feature	# of features
Intensity & shape based features	33
LoG based features	96
Wavelet based features	128
Laws features	432
Co-occurrence features	26
Run-length based features	11
GLSZM based features	11
NGTDM based features	5
Fractal Dimension features	8

Accuracy (%) bar chart:
- All features
- Randomly selected 4 features
- SFS top 4 selected features

- In 2022, a study was conducted to compare the accuracy, sensitivity and specificity of radiomics, CNN and experts' manual analysis in separating benign and malignant nodules on HRCT thorax.

- They retrospectively collected data of 720 patient (each with one nodule and surgical biopsy/trans-thoracic biopsy/transbronchial biopsy-proven disease)

- They randomly divided the patients' data into two groups (7:3)- training set and testing set for CNN model

- For radiomics model used 42 dedicated hand-crafted features and 104 widely used features and 3 random forest prediction models were made – RF with radiomics, EF with clinical features and RF with both

- 2 blinded radiologists were asked to classify the nodules on the basis of their observation (represents manual analysis of the images)

The diagnostic and prognostic value of radiomics and deep learning technologies for patients with solid pulmonary nodules in chest CT images

Rui Zhang[1†], Ying Wei[2†], Feng Shi[2], Jing Ren[1], Qing Zhou[2], Weimin Li[1*] and Bojiang Chen[1*]

Fig. 1 The framework of the deep learning model. A The pre-trained model. B The proposed model "1 × 1 × 1 Conv" refers to the convolutional layer with 1 × 1 × 1 kernel. Dice loss refers to Dice similarity coefficient between the nodule mask and class activation mapping (attention map)
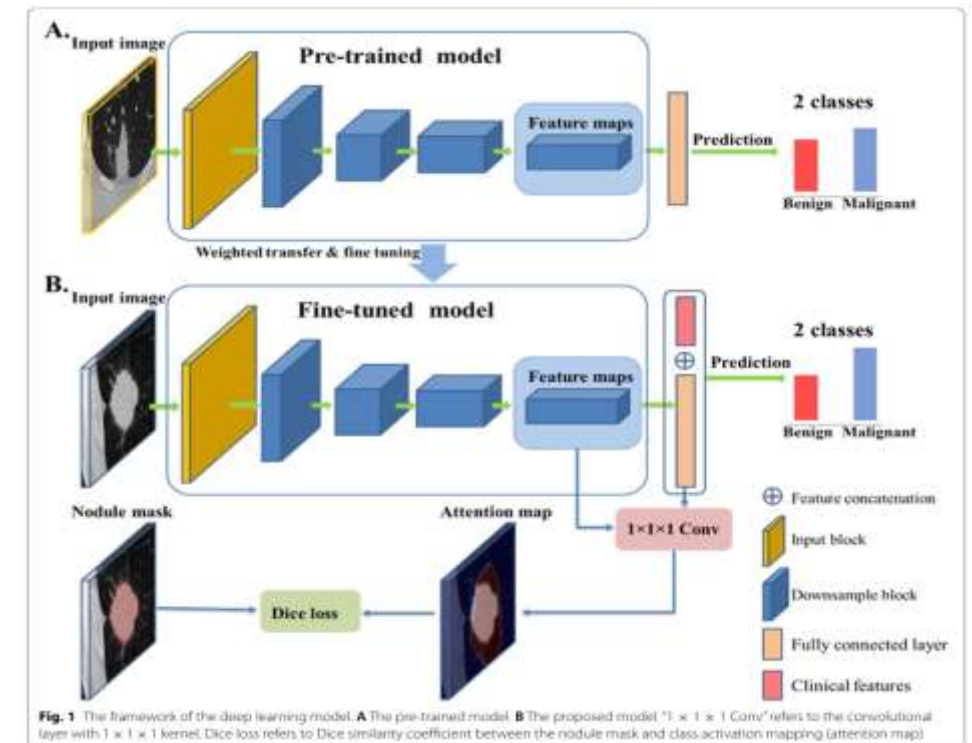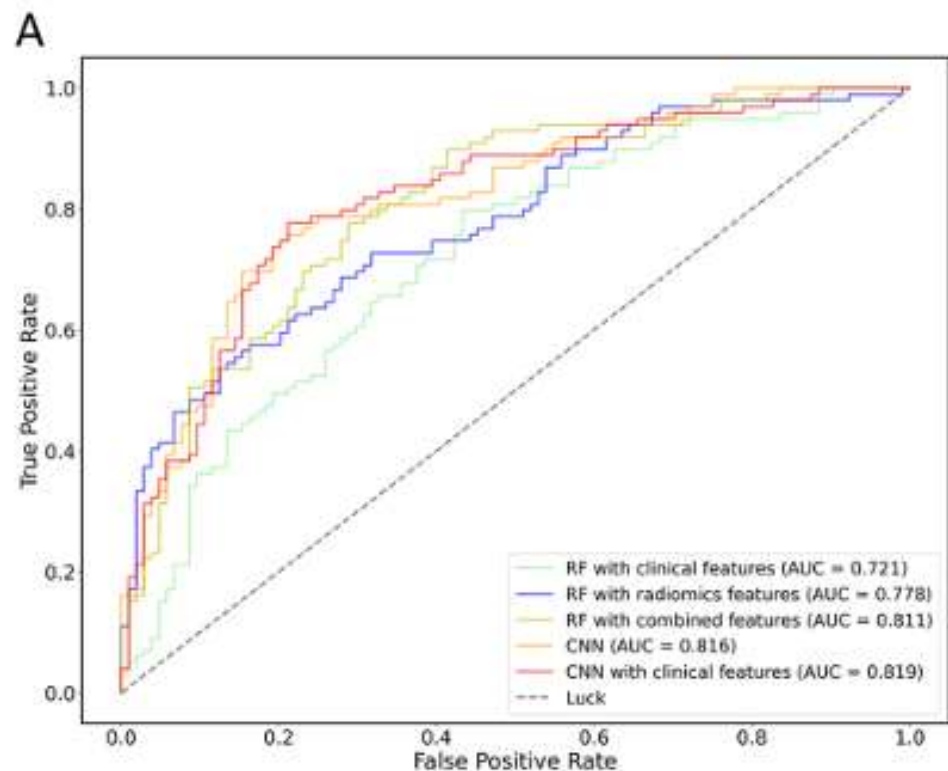
Zhang R, Wei Y, Shi F, Ren J, Zhou Q, Li W, et al. The diagnostic and prognostic value of radiomics and deep learning technologies for patients with solid pulmonary nodules in chest CT images. BMC Cancer. 2022 Nov 1;22(1).

A — ROC curves with legend:
- RF with clinical features (AUC = 0.721)
- RF with radiomics features (AUC = 0.778)
- RF with combined features (AUC = 0.811)
- CNN (AUC = 0.816)
- CNN with clinical features (AUC = 0.819)
- Luck

B — Calibration plot with legend:
- RF + Clinical (loss=0.215)
- RF + Radiomics (loss=0.195)
- RF + Combined (loss=0.184)
- CNN (loss=0.183)
- CNN + Clinical (loss=0.177)
- Perfectly calibrated

CNN with clinical features had the highest AUC for ROC

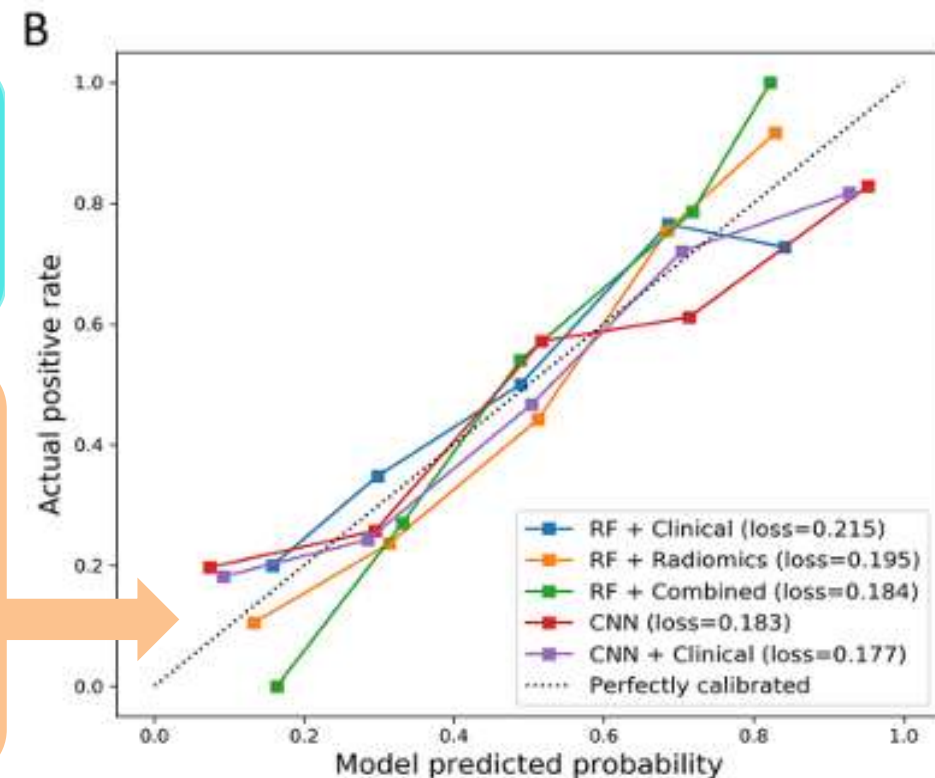It also had minimal loss

Best accuracy: CNN+clinical

Table 2 Predictive performance of different classification models and junior radiologists in the testing set

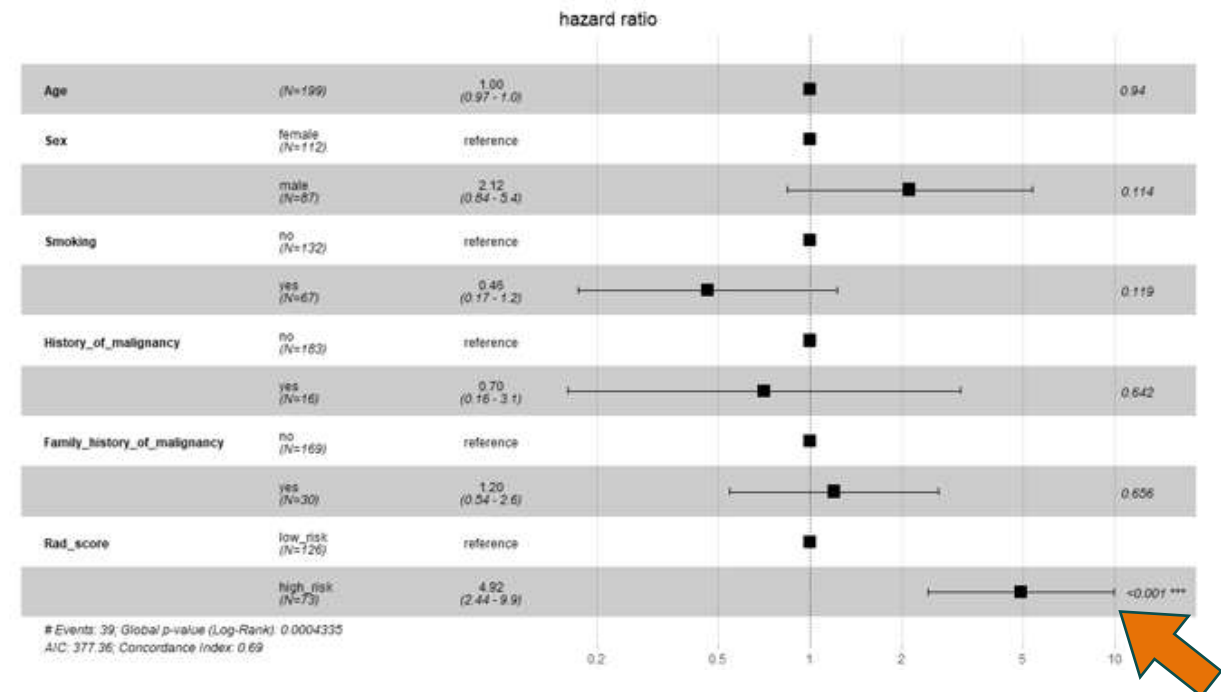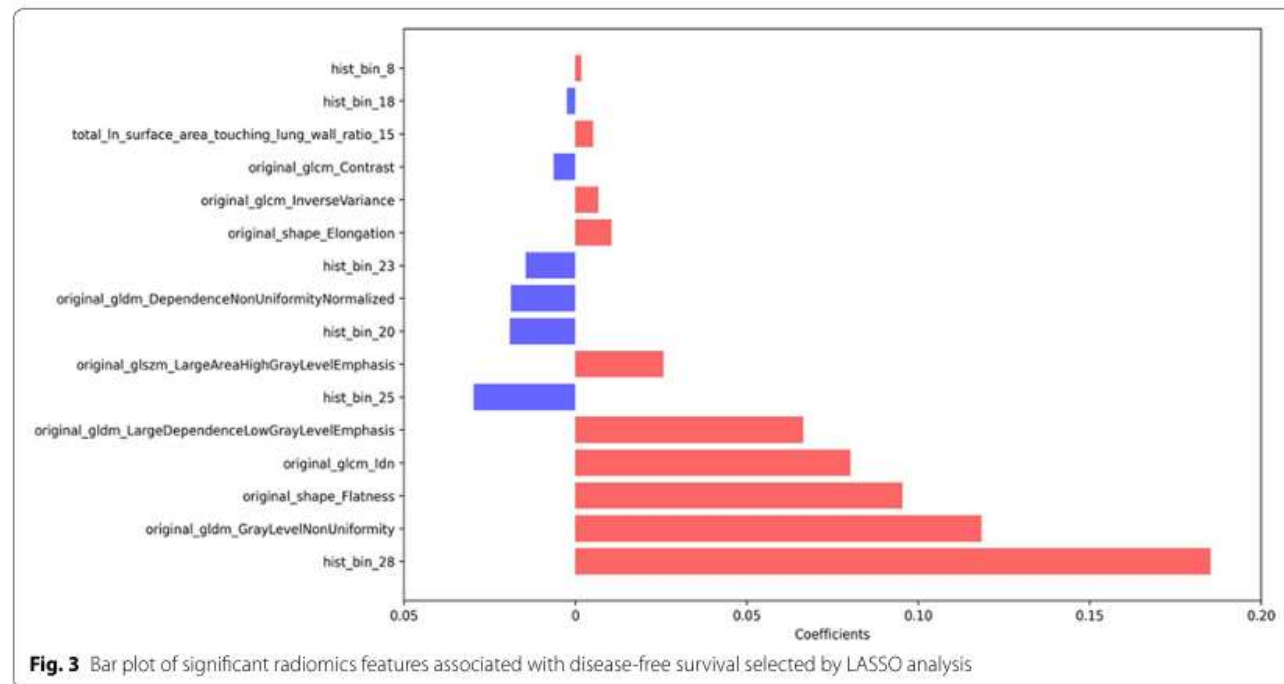| Model or radiologist	Sensitivity	Specificity	Accuracy	AUC
RF + Clinical	0.535 [0.437, 0.633]	0.740 [0.656, 0.824]	0.640 [0.574, 0.706]	0.721 [0.651, 0.791]*
RF + Radiomics	0.747 [0.661, 0.833]	0.606 [0.512, 0.700]	0.675 [0.611, 0.739]	0.778 [0.738, 0.858]
RF + Combined	0.616 [0.520, 0.712]	**0.788 [0.709, 0.867]**	0.704 [0.641, 0.767]	0.811 [0.713, 0.839]
CNN	0.758 [0.674, 0.842]	**0.788 [0.709, 0.867]**	0.773 [0.715, 0.831]	0.816 [0.758, 0.875]
CNN + Clinical	0.778 [0.696, 0.860]	**0.788 [0.709, 0.867]**	**0.783 [0.726, 0.840]**	**0.819 [0.760, 0.877]**
Radiologist 1	0.778 [0.696, 0.860]	0.452 [0.356, 0.548]	0.611 [0.544, 0.678]	0.615 [0.538, 0.692]
Radiologist 2	**0.990 [0.970, 1.000]**	0.519 [0.423, 0.615]	0.749 [0.689, 0.808]	0.755 [0.688, 0.821]

*Significant difference was found between the CNN model with clinical features and RF with clinical features by Delong test ($p < 0.05$)

Abbreviations: RF Random forest, CNN Convolutional neural network, AUC Area under the receiver operating characteristic curves

- Survival data from 295 patients were collected (all of whom underwent surgical resection for adenocarcinoma) and analyzed

- LASSO analysis found 16 radiomics features that were associated with survival

- These 16 radiomic signatures were used to calculate the Rad-score and patients were classified on the basis of a cut off of 0.183 (based on X-tile) into high or low risk groups

- The Kaplan-Meier survival analysis showed that DFS between the low-risk and high-risk groups were statistically different (P<0.011)

Fig. 3 Bar plot of significant radiomics features associated with disease-free survival selected by LASSO analysis

- **The Kaplan-Meier survival curves in the training set.**

- **Survival analysis among stage I patients with solid adenocarcinomas**

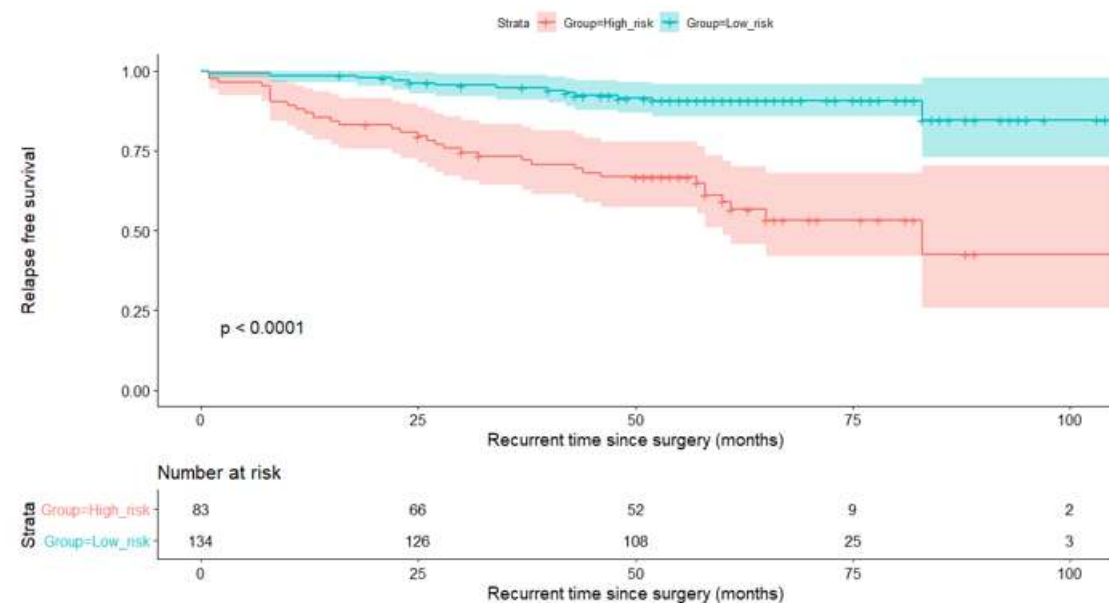**Fig. S2.** Kaplan-Meier survival curves in the training set based on radiomics features.
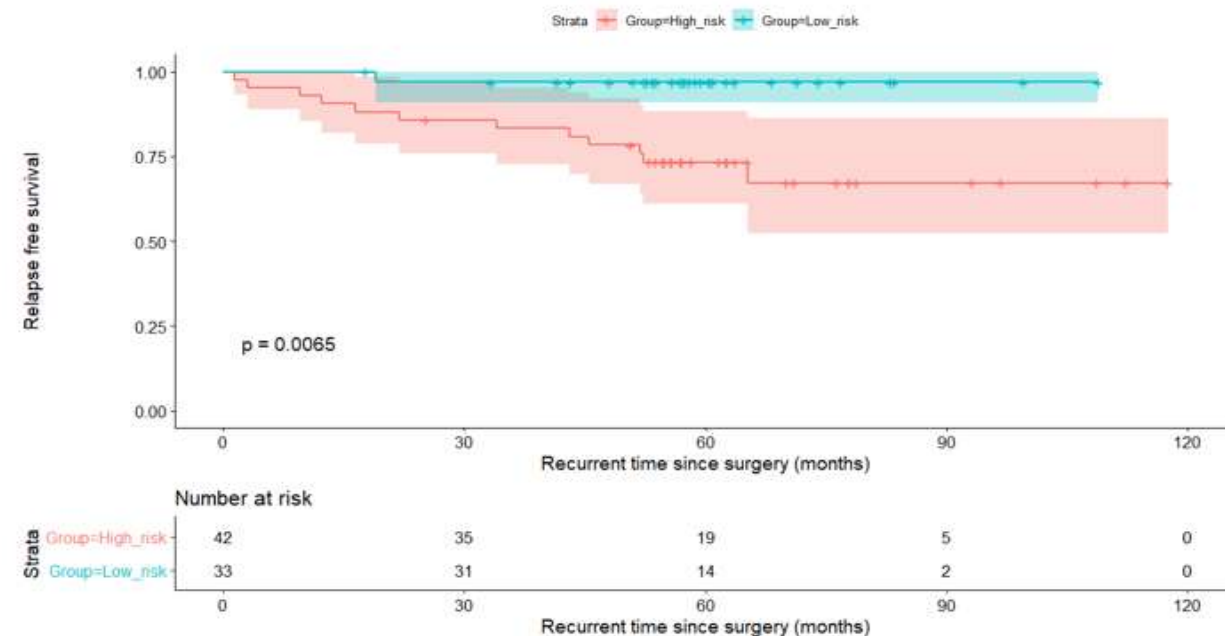
- CNN model was also used for survival analysis

- It was not able to predict survival as effectively as the radiomics model
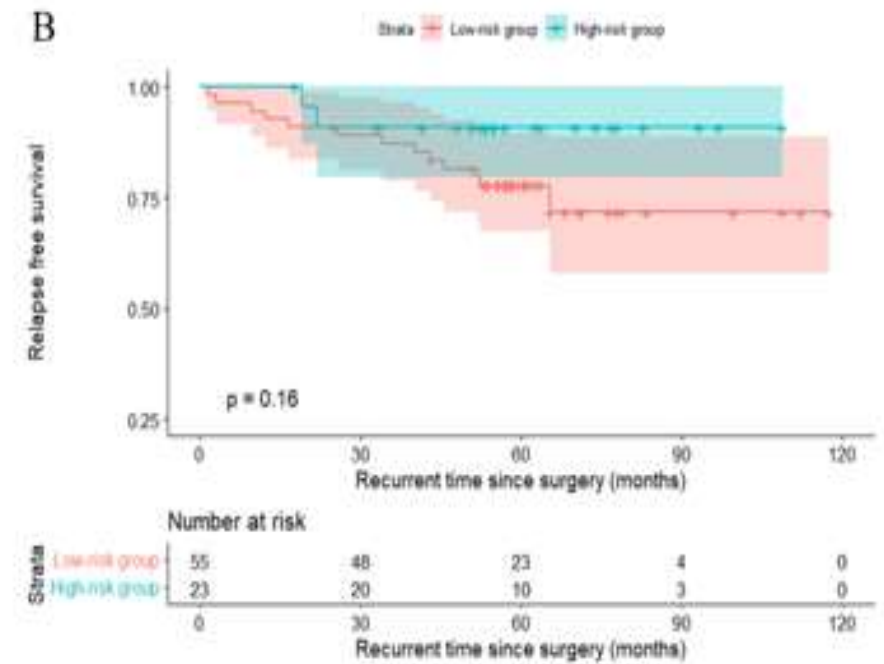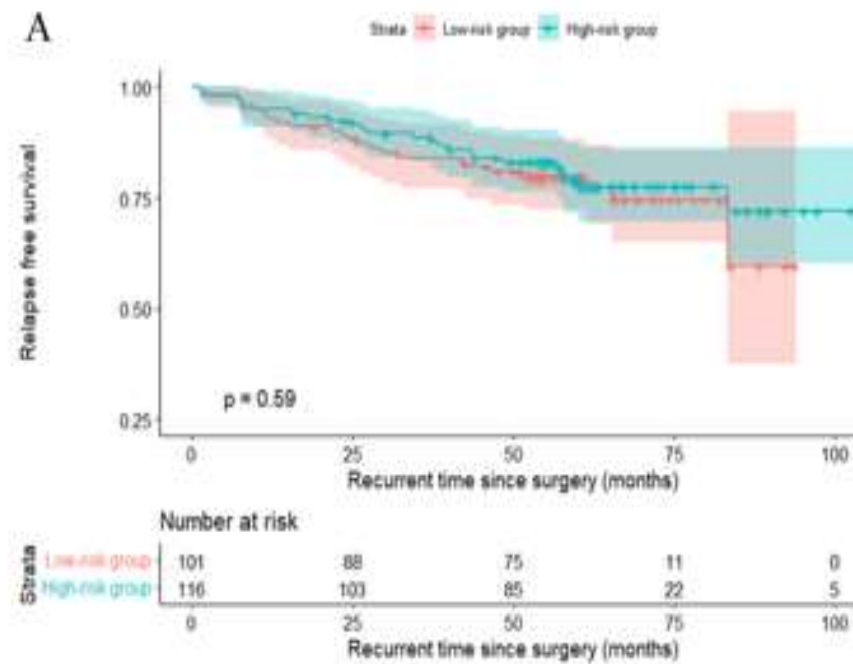
- The study showed that radiomics features can be successfully applied to differentiate patients with benign nodules from those with malignant nodules on the basis of HRCT thorax

- Both radiomics and CNN, in addition to clinical features outperformed the radiologists (however, it was mentioned they were not as experienced as many other radiologists might be, and were also arbitrarily chosen)

- Radiomics features may be helpful in predicting relapse free survival

- CNN was not as effective as radiomics in predicting outcome of patients with lung nodules

- A 2023 systematic review and metaanalysis selected 20 studies that evaluated radiomics and deep learning to differentiate malignant lung nodules from benign ones

- All the studies were retrospective

- The authors calculated the positive and negative likelihood ratio and diagnostic odds ratio for radiomics and deep learning techniques in terms of their ability to correctly classify a nodule

- They also evaluated the studies with QUARADS 2 score and the radiomics used with RQS 2

The classification of benign and malignant lung nodules based on CT radiomics: a systematic review, quality score assessment, and meta-analysis

Fandong Zhu[1] (iD), Chen Yang[1], Jiajun Zou[1], Weili Ma[1], Yuguo Wei[2] and Zhenhua Zhao[1] (iD)

Fig. 1. Study flow diagram.

Zhu F, Yang C, Zou J, Ma W, Wei Y, Zhao Z. The classification of benign and malignant lung nodules based on CT radiomics: a systematic review, quality score assessment, and meta-analysis. Acta Radiologica. 2023 Oct 10;64(12):3074–84.
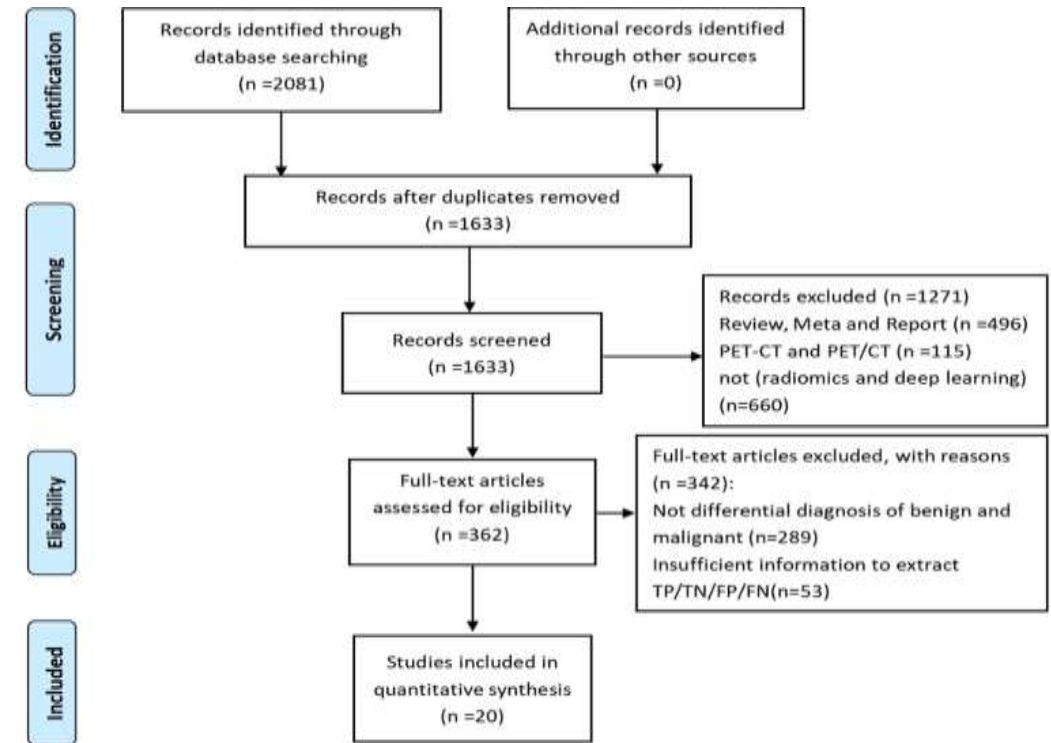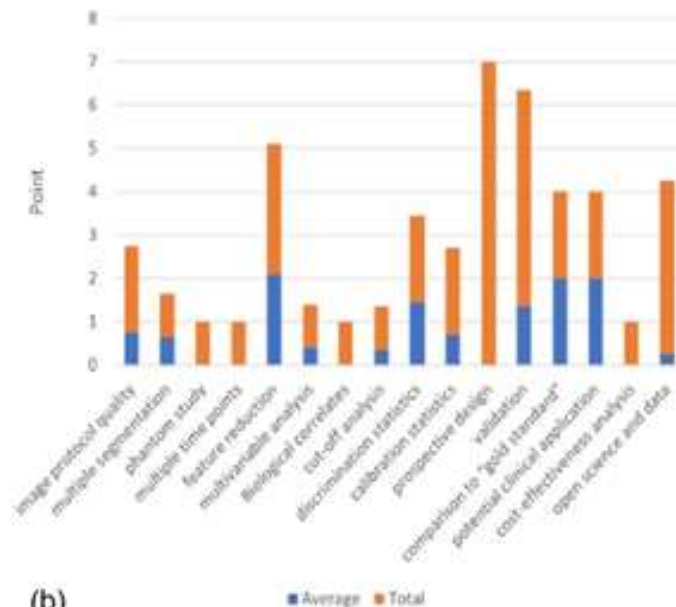
Table I. Characteristics of the included studies.

Study and year	Country	Method	Segmentation	Nodules	No. of patients	Malignant lesions	Benign lesions	TP	FP	FN	TN
Beig et al., 2019 (22)	USA	Radiomics	Internal and around	LA vs. granulomas	145	72	73	53	24	19	49
Chen et al., 2020 (23)	China	Radiomics	Internal	LA vs. granulomas	64	35	29	25	5	10	24
Dennie et al., 2016 (24)	Canada	Texture	Internal	Malignant lung nodules vs. granulomas	55	31	24	23	0	8	24
Feng et al., 2020 (25)	China	Deep learning	Internal	LA vs. TBG	192	141	51	128	20	13	31
Heuvelmans et al., 2021 (26)	Netherlands	Deep learning	Internal	Malignant lung nodules vs. benign lung nodules	2016	205	1811	203	2	1411	400
Hu et al., 2020 (27)	China	texture	Internal	Malignant lung nodules vs. benign lung nodules	30	15	15	13	6	2	9
Kamiya et al., 2014 (28)	Japan	Histograms	Internal	Malignant lung nodules vs. benign lung nodules	93	72	21	41	2	31	19
Liu et al., 2020 (29)	China	Radiomics	Internal	LA vs. benign lung nodules	63	39	24	33	4	6	20
Mao et al., 2019 (30)	China	Radiomics	Internal	Malignant lung nodules vs. benign lung nodules	98	21	77	17	6	4	71
Ni et al., 2021 (31)	China	Radiomics	Internal	Malignant lung nodules vs. PSP	60	35	25	28	4	7	21
Suo et al., 2016 (32)	China	Texture	Internal and around	Malignant lung nodules vs. Inflammatory lesions	48	28	20	23	4	5	16
Uthoff et al., 2019 (33)	USA	Machine learning	Internal and around	Malignant lung nodules vs. benign lung nodules	100	50	50	50	0	0	50
Uthoff et al., 2019 (34)	USA	Radiomics	Internal and around	NSCLC vs. HPN	71	40	31	34	5	6	26
Wan et al., 2020 (35)	China Taiwan	Artificial intelligence	Internal	Malignant lung nodules vs. benign lung nodules	75	47	28	44	5	3	23
Wang et al., 2021 (36)	China Taiwan	Radiomics	Internal	LA vs. SPCH	62	49	13	45	1	4	12
Wei et al., 2020 (37)	China	Mathematical diagnosis model	Internal	NSCLC vs. granulomas	61	40	21	25	6	15	15
Wei et al., 2021 (38)	China	Radiomics	Internal	Malignant lung nodules. vs. TBG	78	41	37	36	37	5	37
Yang et al., 2018 (39)	China	Radiomics	Internal	LA vs. granulomas	91	63	28	52	7	11	21
Zhang et al., 2019 (40)	China	Radiomics	Internal	LA vs. FOP	226	109	117	93	12	16	105
Zhao et al., 2021 (41)	China	Radiomics	Internal	LA vs. PC	128	66	62	50	11	12	55

FN, false negative; FOP, focal organising pneumonia; FP, false positive; HPN, histoplasmosis pulmonary nodules; LA, lung adenocarcinoma; NSCLC, non-small-cell lung cancer; PC, pulmonary cryptococosis; PSP, pulmonary sclerosing pneumocytoma; SPCH, solitary pulmonary capillary haemangioma; TBG, tuberculous granuloma; TN, true negative; TP, true positive.

Fig. 2. Methodological quality was evaluated by using the RQS tool. (a) The proportion of studies with different RQS percentage scores. (b) Average scores of each RQS item (orange bars = total points of each item, blue bars = actual points). RQS, radiomics quality score.
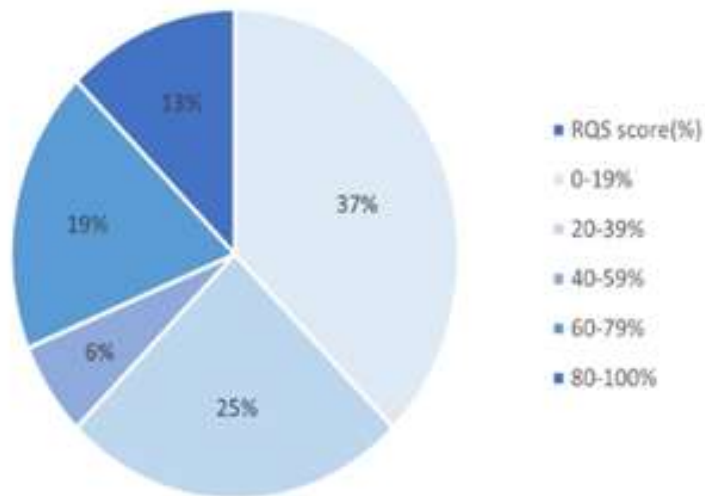
RQS (2) was considerably low in most of studies. Highest score was merely 17 (47% of total score)

- The studies included in this analysis were of sufficient quality to meet the requirements.
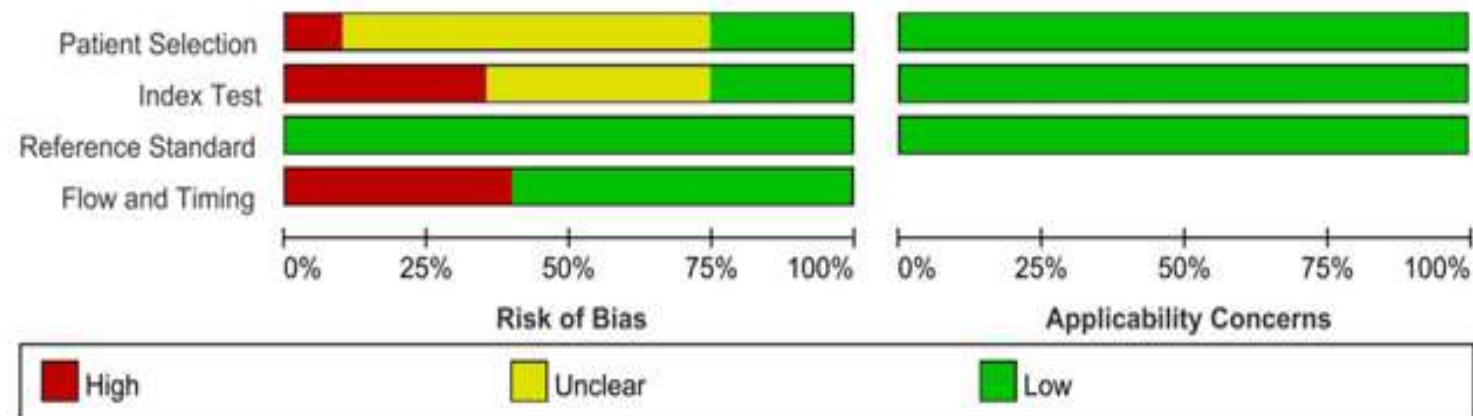- The absence of publication bias was demonstrated in Deeks' funnel plot asymmetry test (P=0.96)

Table 2. Overview of RQS items and mode of the respective score in the reviewed studies.

RQS checkpoint	RQS item number and name	Description and (points)	Average score
First	Item 1: image protocol quality	Well-documented protocol (+1) AND/OR publicly available protocol (+1)	0.75
Second	Item 2: multiple segmentation	Testing feature robustness to segmentation variability, e.g. different physicians/algorithms/software (+1)	0.65
	Item 3: phantom study	Testing feature robustness to scanner variability, e.g. different vendors/scanners (+1)	0
	Item 4: multiple time points	Testing feature robustness to temporal variability, e.g. organ movement/expansion/shrinkage (+1)	0
Third	Item 5: feature reduction	Either feature reduction OR adjustment for multiple testing is implemented (+3); otherwise (−3)	2.10
	Item 6: multivariable analysis	Non-radiomic feature are included in/considered for model building (+1)	0.40
	Item 7: biological correlates	Detecting and discussing correlation of biology and radiomic features (+1)	0
	Item 8: cutoff analysis	Determining risk groups by either median, pre-defined cutoff, or continuous risk variable (+1)	0.35
	Item 9: discrimination statistics	Discrimination statistic and its statistical significance are reported (+1); a resampling technique is also applied (+1)	1.45
	Item 10: calibration statistics	Calibration statistic and its statistical significance are reported (+1); a resampling technique is also applied (+1)	0.70
	Item 11: prospective design	Prospective validation of a radiomics signature in an appropriate trial (+7)	0
	Item 12: validation	Validation is missing (−5) OR internal validation (+2) OR external validation on single dataset from one institute (+3) OR external validation on two datasets from two distinct institutes (+4) OR validation of a previously published signature (+4) validation is based on three or more datasets from distinct institutes (+5)	1.35
	Item 13: comparison to "gold standard"	Evaluating model's agreement with/superiority to the current "gold standard" (+2)	2.00
	Item 14: potential clinical application	Discussing model applicability in a clinical setting (+2)	2.00
	Item 15: cost-effectiveness analysis	Performing a cost-effectiveness of the clinical application (+1)	0
	Item 16: open science and data	Open source scans (+1) AND/OR open source segmentations (+1) AND/OR open source code (+1) AND/OR open source representative features and segmentations (+1)	0.25

Radiomics: the bridge between medical imaging and personalized medicine

Philippe Lambin ✉, Ralph T.H. Leijenaar, Timo M. Deist, Jurgen Peerlings, Evelyn E.C. de Jong, Janita van Timmeren, Sebastian Sanduleanu, Ruben T.H.M. Larue, Aniek J.G. Even, Arthur Jochems, Yvonka van Wijk, Henry Woodruff, Johan van Soest, Tim Lustberg, Erik Roelofs, Wouter van Elmpt, Andre Dekker, Felix M. Mottaghy, Joachim E. Wildberger & Sean Walsh
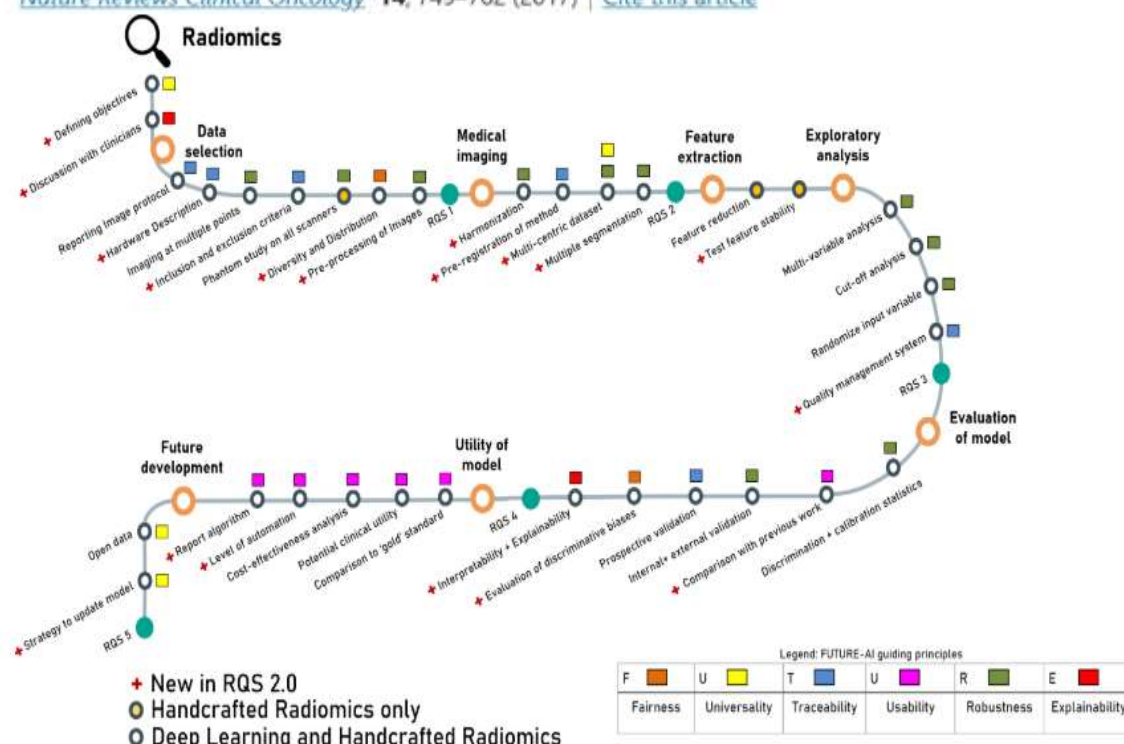
Distinction between benign and malignant pulmonary nodule

Sensitivity- 81%   Specificity- 86%   DOR- 27.00   AUC-0.91
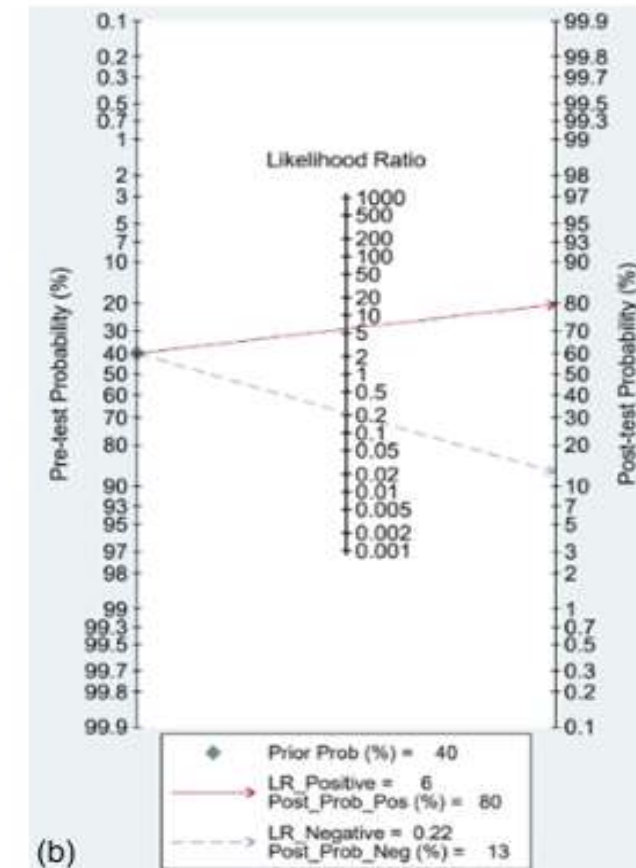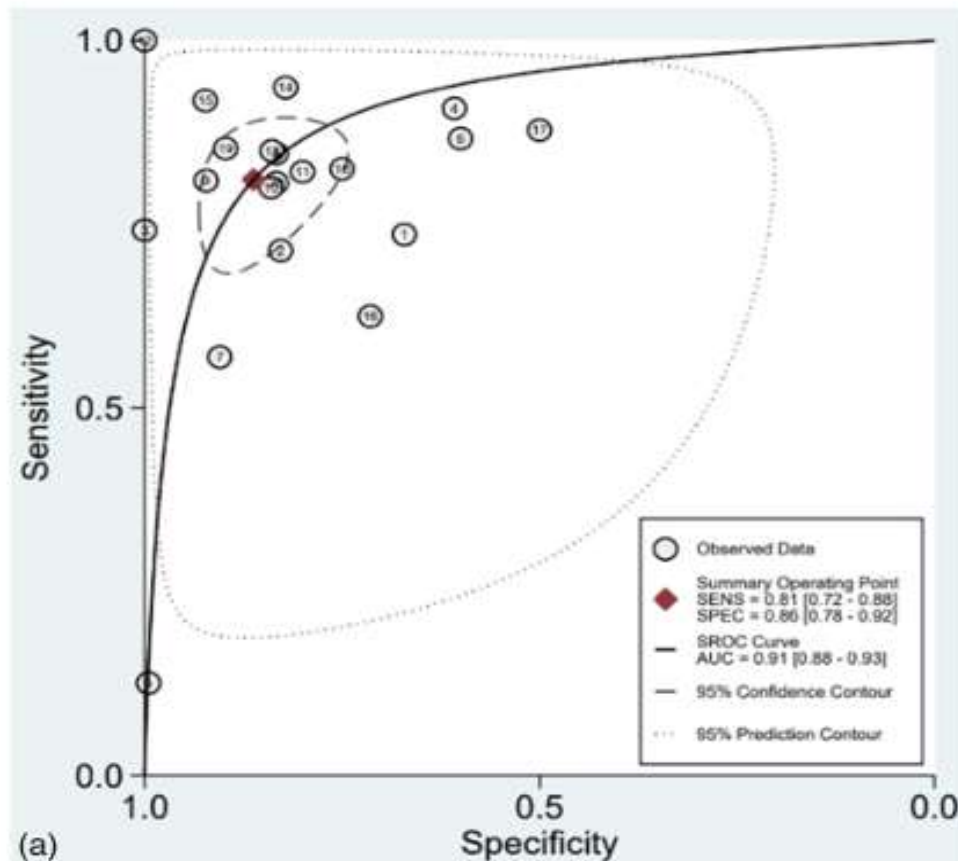
Diagnostic odds ratio- odds of finding malignancy in actual malignant nodules/ odds of finding malignancy when malignancy is not present

Table 3. Meta-regression results of radiomics for the diagnosis of benign and malignant lung nodules.

Covariates	Subgroups	No. of studies	P	Sensitivity	P1	Specificity	P2
CT scanning equipment	GE	10	0.59	0.84 (0.75–0.94)	0.45	0.87 (0.77–0.96)	0.23
	Others	10		0.77 (0.65–0.90)		0.86 (0.76–0.95)	
CT slice thickness (cm)	>2	9	0.64	0.79 (0.67–0.91)	0.09	0.89 (0.81–0.97)	0.52
	≤2	11		0.83 (0.73–0.92)		0.83 (0.73–0.93)	
Modeling method	Logistic	9	0.71	0.84 (0.75–0.94)	0.41	0.85 (0.74–0.96)	0.10
	Others	11		0.78 (0.66–0.89)		0.87 (0.79–0.96)	
RQS	≥13	14	0.07	0.85 (0.78–0.92)	0.88	0.81 (0.72–0.90)	0.00*
	<13	6		0.69 (0.51–0.87)		0.94 (0.88–1.00)	
Clinical factor	Yes	8	0.15	0.81 (0.68–0.93)	0.15	0.79 (0.65–0.92)	0.01*
	Unclear	12		0.82 (0.72–0.92)		0.90 (0.84–0.97)	
Segmentation method	Manual	12	0.18	0.75 (0.64–0.86)	0.00*	0.88 (0.80–0.96)	0.46
	Automatic	8		0.88 (0.80–0.96)		0.83 (0.71–0.96)	
Enhanced scanning	Yes	9	0.27	0.73 (0.60–0.87)	0.01*	0.89 (0.80–0.97)	0.43
	Unclear	11		0.86 (0.78–0.94)		0.84 (0.74–0.94)	
Geographical location	Mainland China	12	0.04†	0.83 (0.74–0.92)	0.39	0.79 (0.68–0.89)	0.00*
	Other	8		0.78 (0.65–0.92)		0.94 (0.89–0.99)	
No. of patients	≥90	10	0.81	0.79 (0.67–0.90)	0.06	0.88 (0.79–0.96)	0.31
	<90	10		0.83 (0.73–0.93)		0.85 (0.74–0.95)	
ROI	Tumor	14	0.82	0.80 (0.70–0.90)	0.13	0.86 (0.77–0.94)	0.20
	Peritumor	6		0.84 (0.71–0.96)		0.87 (0.76–0.99)	
Tumor size (mm)	≥15	9	0.37	0.82 (0.72–0.93)	0.25	0.80 (0.68–0.93)	0.01*
	<15	11		0.80 (0.69–0.91)		0.90 (0.83–0.97)	
Published year	≥2020	11	0.39	0.80 (0.69–0.91)	0.10	0.83 (0.72–0.93)	0.02†
	<2020	9		0.82 (0.72–0.93)		0.90 (0.82–0.97)	

Values in parentheses are 95% CI.

*P < 0.01.

†P < 0.05.

CT, computed tomography; ROI, region of interest; RQS, radiomics quality score.

Higher QRS reduced specificity

Inclusion of clinical factors reduced specificity

Larger tumor reduced specificity

- This study showed that pooled sensitivity and specificity as well as diagnostic odd ratio were considerably high with radiomics in differentiating benign and malignant lung nodules

- Some of the studies included machine learning and deep learning as well

- Both hand-picked radiomics signatures (trained by humans) and radiomic signature chosen by artificial intelligence were used in different studies

- These gave rise to significant heterogeneity

- Question still remains whether it should be completely left to AI without any need for human verification

- A common concern regarding AI-detection of lung nodules is that whether AI would be able to correctly identify a nodule on the background of chronic lung disease

- A 2022 study addressed this concern

- The study was aimed to determine whether AI was non-inferior in identifying lung nodules on the background of ILD/COPD/pulmonary oedema/pulmonary embolism when compared to radiologists

- They kept a study population of patients who had nodules reported on CT report and a control population with no reported nodule (to assess the False Positive rate in presence of background lung disease).

Abadia AF, Yacoub B, Stringer N, Snoddy M, Kocher M, Schoepf UJ, et al. Diagnostic Accuracy and Performance of Artificial Intelligence in Detecting Lung Nodules in Patients With Complex Lung Disease: A Noninferiority Study. Journal of Thoracic Imaging [Internet]. 2022 May 1 [cited 2022 Sep 23];37(3):154–61.

STUDY POPULATION

ILD, COPD, RB, EDEMA, PE

Presence of at least one suspicious lung nodule on radiology reports of chest CT scans

ILD 21
COPD 21
RB 21
EDEMA 19
PE 21

CONTROL POPULATION

ILD, COPD, RB, EDEMA, PE

Absence of lung nodules on radiology reports of chest CT scans

RB EXCLUDED 10

ILD 10
COPD 10
EDEMA 10
PE 10
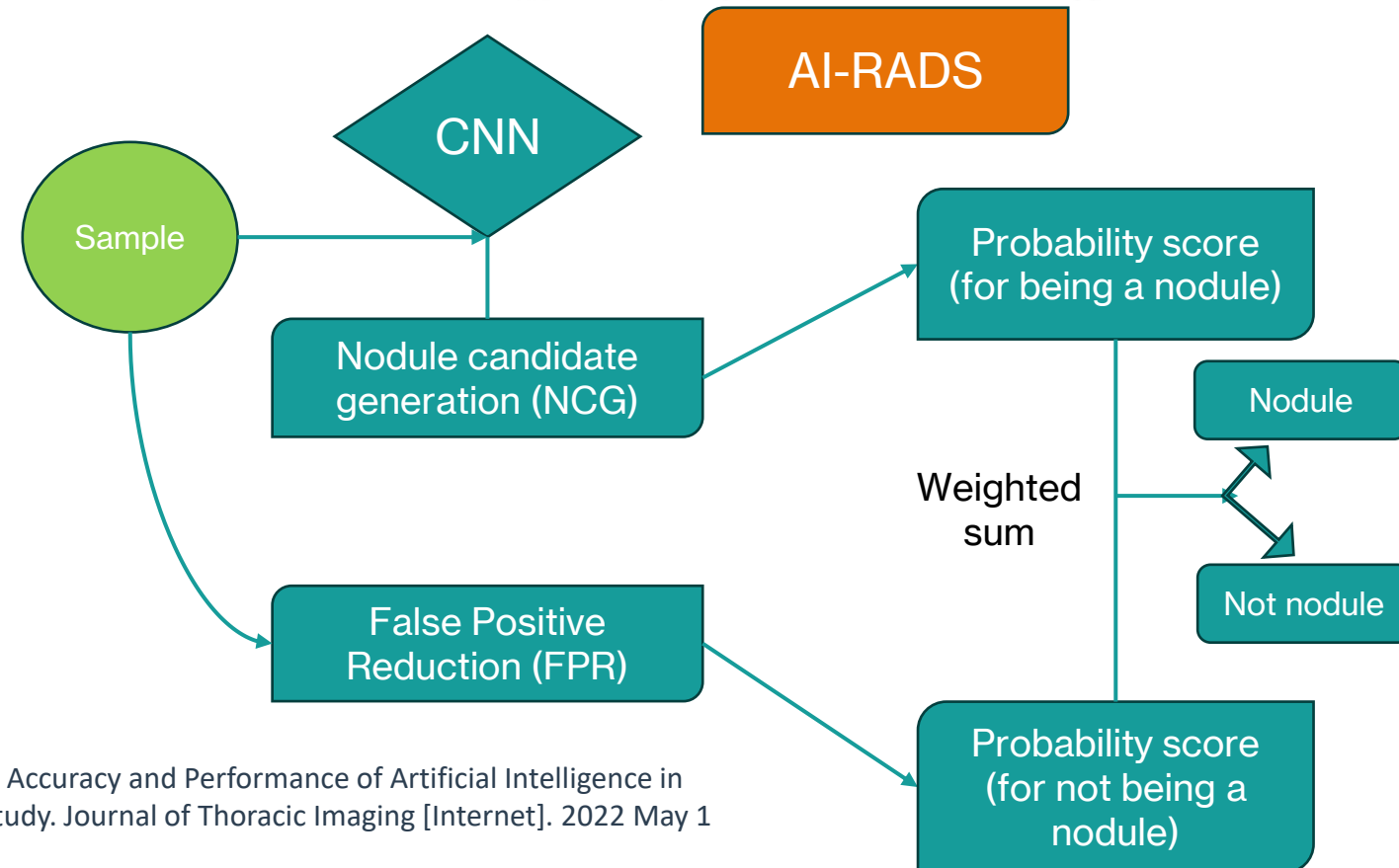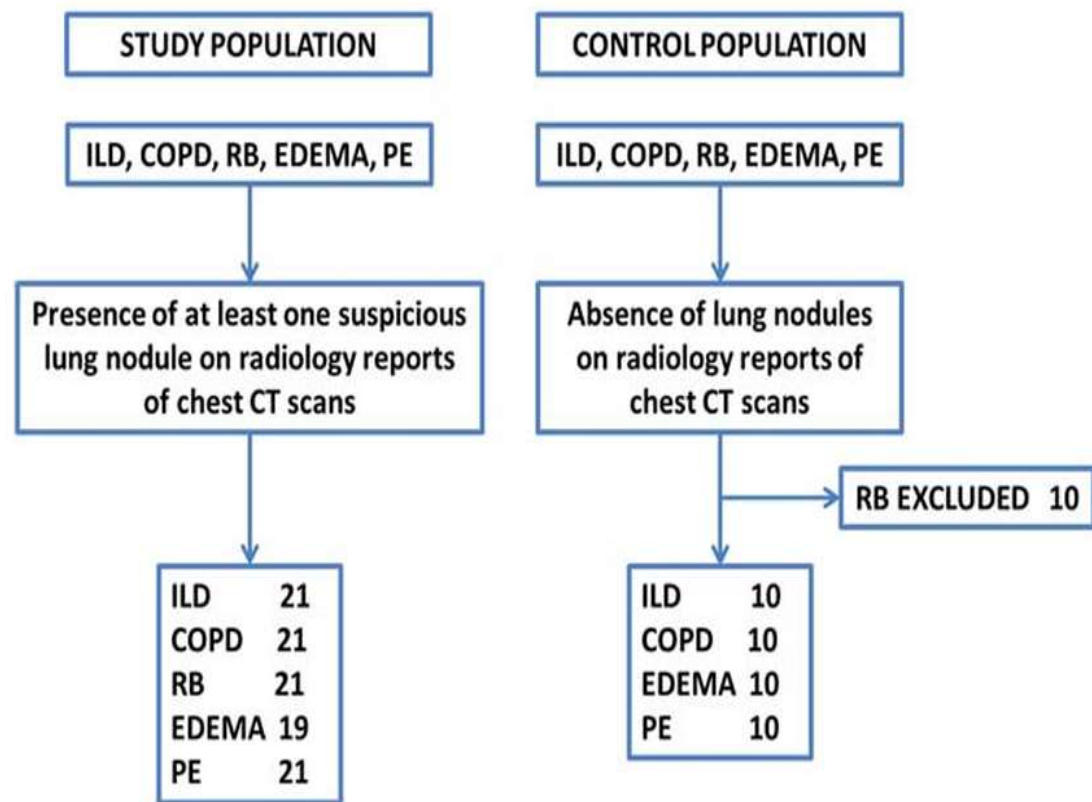
TABLE 1. Baseline Characteristics

	Study Population (n = 103)	Control Population (n = 40)
Age (y)	62.6 ± 12.9	58.4 ± 16.4
No. men	55 (53.4)	16 (40.0)
Condition		
ILD	21 (20.4)	10 (25.0)
COPD	21 (20.4)	10 (25.0)
RB	21 (20.4)	NA
EDEMA	19 (18.4)	10 (25.0)
PE	21 (20.4)	10 (25.0)
BMI	24.9 (21.8-30.6)	29.9 (25.6-34.6)
Smoking (pack years)	15.0 (0.0-38.8)	12.0 (0.0-30.0)
Current smoker	26 (25.2)	8 (20.0)
DM	26 (25.2)	5 (12.5)
HTN	67 (65.0)	23 (57.5)
TB exposure	3 (2.9)	1 (2.5)
Asbestos exposure	7 (6.8)	0
Family Hx of lung cancer	7 (6.8)	3 (7.5)
$CTDI_{vol}$ (mGy)	8.0 (4.3-11.4)	8.2 (4.9-14.7)
DLP (mGy·cm)	290.6 (133.3-468.7)	244.7 (127.9-477.5)

LESIONS	Lobe	Volume [mm³]	Max. 2D Ø [mm]	Max. 3D Ø [mm]
L1	Right Middle Lobe	362.2	10.8	11.4
L2	Left Upper Lobe	36.3	8.4	8.5
L3	Right Upper Lobe	31.6	5.5	5.5
L4	Left Upper Lobe	33.0	5.3	5.4
Tumor Burden			30.0	

Not for diagnosis *

L1: 10.8 mm

L1 10.8 mm

TABLE 2. Results From the AI-RAD and Expert Nodule Assessment of the *Study Population*

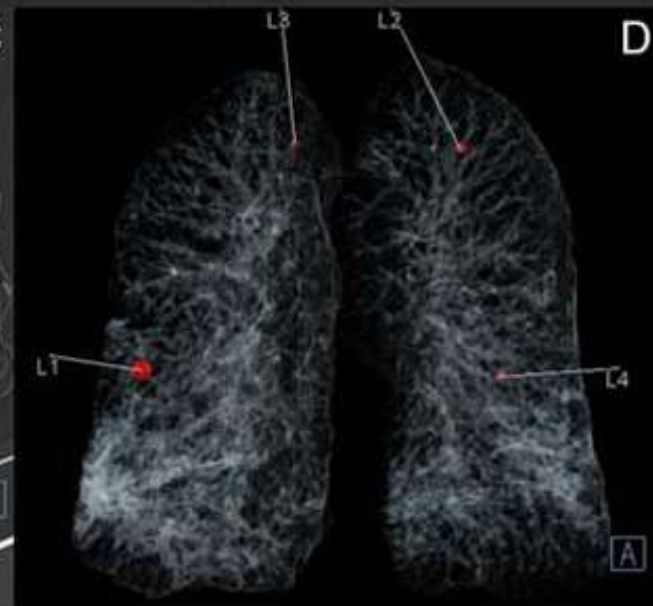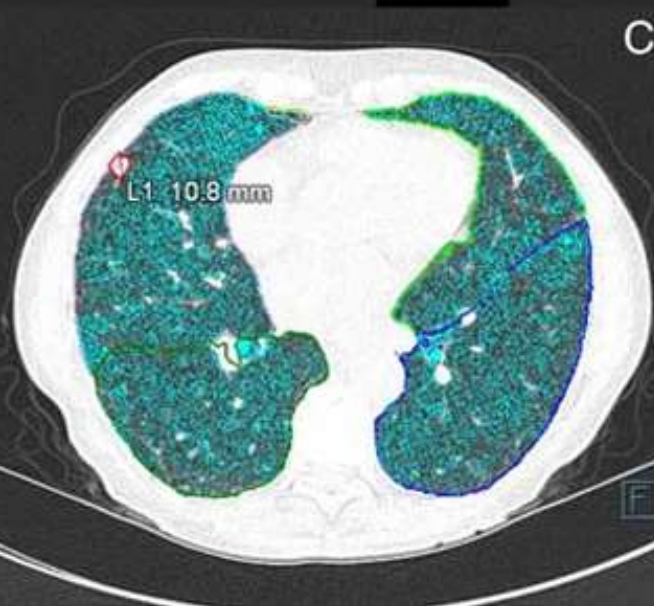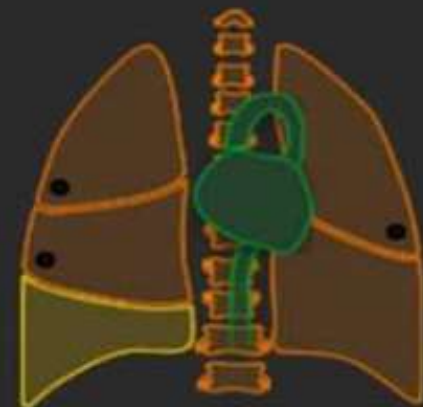|	All Conditions	ILD	COPD	RB	EDEMA	PE
Nodules						
Missed by expert (TP)	37 (8.4)	5 (5.8)	6 (6.9)	14 (17.1)	3 (4.2)	9 (7.9)
Missed by AI (FN)	129 (29.3)	27 (31.4)	17 (19.5)	23 (28.0)	26 (36.1)	36 (31.6)
Detected by expert and AI (TP)	233 (52.8)	48 (55.8)	51 (58.6)	32 (39.0)	37 (51.4)	65 (57.0)
AI FP	38 (8.6)	5 (5.8)	13 (14.9)	10 (12.2)	6 (8.3)	4 (3.5)
Wrong location by AI*	4 (0.9)	1 (1.2)	0	3 (3.7)	0	0
Total detected by AI	312 (70.7)	59 (68.6)	70 (80.5)	59 (72.0)	46 (63.9)	78 (68.4)
Total detected by expert	366 (83.0)	76 (88.4)	68 (78.2)	58 (70.7)	63 (87.5)	101 (88.6)
Total nodules†	441	86	87	82	72	114
AI—median size (mm)	8.4 (6.3-11.6)	8.4 (6.9-11.5)	7.7 (6.0-10.7)	7.6 (5.4-10.2)	10.4 (7.2-13.8)	9.1 (6.5-13.9)
Expert—median size (mm)	7.1 (5.3-10.5)	6.9 (5.5-10.2)	6.0 (4.9-8.1)	7.1 (4.9-9.2)	8.4 (5.8-10.3)	8.2 (5.5-18.6)
Accuracy						
AI sensitivity	67.7%	66.3%	77.0%	66.7%		
AI PPV (positive predictive value)	87.7%	91.4%	81.4%	82.1%		

Nodule values are reported as total number and percentages, n (%). Nonparametric continuous variables are expre~~
quartile ranges.
*The nodule was detected correctly but reported in the wrong lobe.
†Nodule addition of total detected by the expert, total by AI-RAD but missed by expert, and the number of FP yiel~~
on the nodules are based on this total.
AI indicates Artifical Intelligence; COPD, Chronic Obstructive Pulmonary Disease; FN, False Negative; FP, False P~~
PE, Pulmonary Embolism; RB, Respiratory Bronchiolitis; TP, True Positive.

In comparison to the expert radiologist, the AI showed non-inferior efficacy in detecting nodule and classifying them correctly (Nodule present vs nodule absent)

Control population

TABLE 4. AI-RAD's Nodules Present Versus Absent Test Results

	The Truth			
			Sensitivity	96.1%
	Has Nodules	No Nodules	Specificity	77.5%
AI score				
Has nodules	99 (TP)	9 (FP)	PPV	91.7%
No nodules	4 (FN)	31 (TN)	NPV	88.6%

TABLE 3. Results From the AI-RAD Assessment of the *Control Population*

	All Conditions	ILD	COPD	EDEMA	PE
FP	7	1	2	1	3
TN	33	9	8	9	7
FP size (mm)*	15.2 ± 10.1	11.0	6.3; 16.2	23.50	28.4; 6.9; 9.6
AI specificity	82.5%	90.0%	80.0%	90.0%	70.0%

RB cases were not included in this assessment as all of them had presence of lung nodules mentioned in the radiology reports.

*Given the small amount of FP found, it was decided to report the average size (and SD) for "All Conditions," and the size of each of the
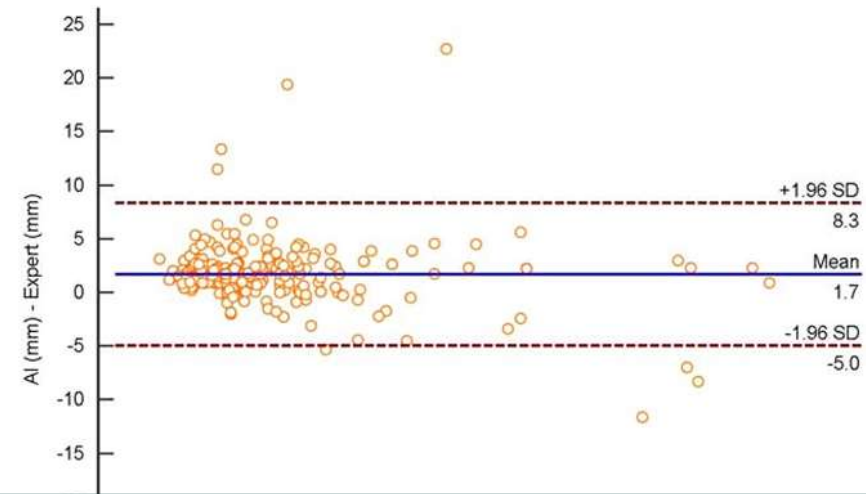
- AI showed good correlation in measuring the nodules as compared with measurements made by the radiologist
- Time taken was shorter in case of AI
- On revisiting 20 random cases of this cohort, radiologists claimed interpreting with AI gave them more confidence regarding the diagnosis
- They detected nodules with greater accuracy and in shorter time (when taking help of AI)

Total detected by AI	185 (80.4)	29 (85.3)	38 (80.9)	35 (79.5)	34 (66.7)	49 (90.7)
Total in reports	182 (79.1)	25 (73.5)	37 (78.7)	37 (84.1)	44 (86.3)	39 (72.2)
Total nodules*	230	34	47	44	51	54
Accuracy						
% of nodules reported detected by AI†	75.8%	80.0%	75.7%	75.7%	63.6%	87.2%
AI Sensitivity	89.4%	96.7%	90.5%	85.4%	79.5%	96.1%

Nodule values are reported as total number and percentages, n (%). At our institution, radiologists only reported the largest 1 to 3 nodules on most reports; therefore, we assessed those nodules reported versus the largest three nodules (or less) found by AI-RAD (validated as TP by our expert radiologist).

*Obtained by adding the total number of nodules found in the reports and the total number of nodules not reported that were found by AI-RAD.

†Total number of nodules found by AI-RAD that were also in the reports divided by the total number of nodules reported across all cases.

- Prediction of lung cancer in patients with high-risk status has been an important prospective field of research in AI technologies

- Only thoracic imagine can not predict risk of lung cancer in healthy subjects

- Multiple AI models have been used to predict the risk of lung cancer in such patients, but they take into account multiple factors such as age, smoking status, spirometry values and family history just like the traditional models of risk prediction which used regression analysis for development

AI imaging in interstitial lung disease

- Radiological diagnosis

- Quantification on basis of CT imaging

- Prognostication

- Assessment of disease progression and response to treatment

- In 2021 a study used CXRs and CT thorax from chronic fibrosing ILD patients and trained a DCNN AI to predict the presence of absence of fibrosing ILD on the basis of CXR alone

- The deep convolutional network-based AI showed an AUC of 0.91 for differentiating CF-ILD from normal CXRs successfully (all CF-ILD confirmed by MDD) and were generally better than the clinicians in doing so

Deep-learning algorithm to detect fibrosing interstitial lung disease on chest radiographs

Hirotaka Nishikiori [1], Koji Kuronuma[1], Kenichi Hirota[2], Naoya Yama[3], Tomohiro Suzuki[4], Maki Onodera[3], Koichi Onodera[3], Kimiyuki Ikeda[1], Yuki Mori[1], Yuichiro Asai[1], Yuzo Takagi[5], Seiwa Honda[4], Hirofumi Ohnishi[6], Masamitsu Hatakenaka[3], Hiroki Takahashi[1] and Hirofumi Chiba[1]
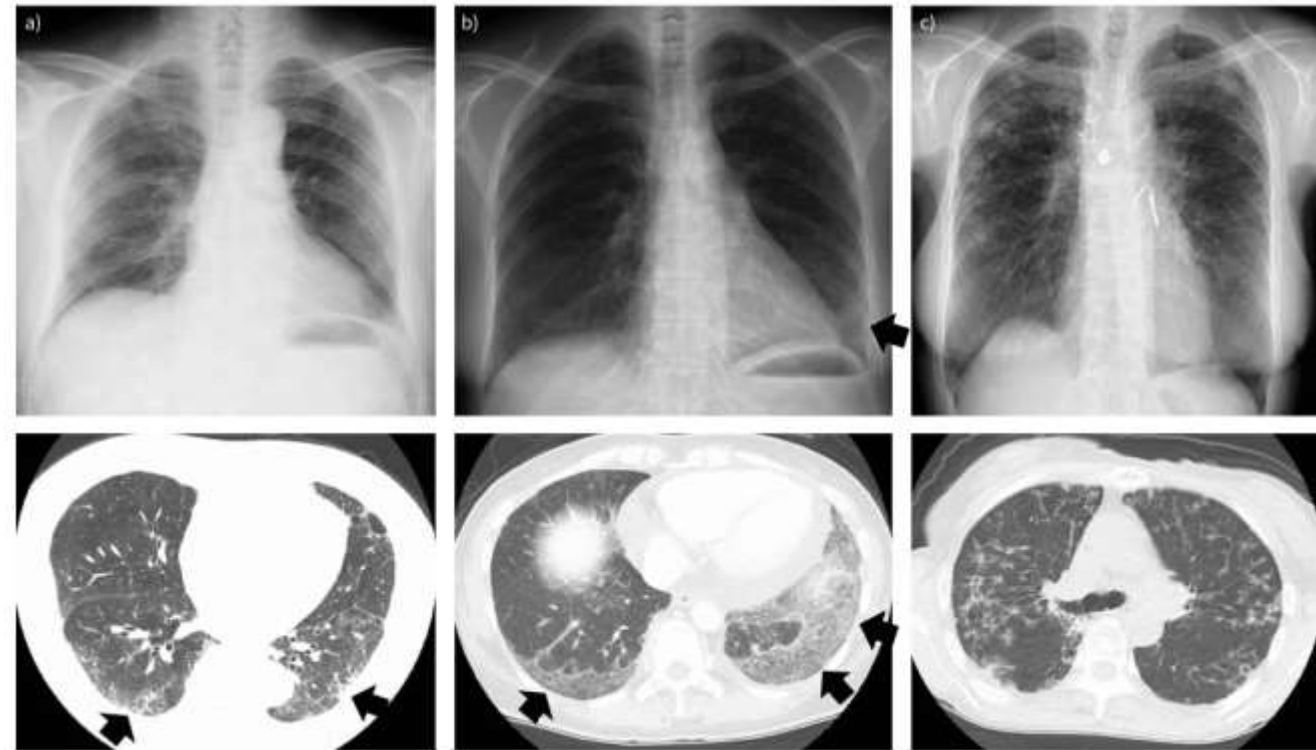
Nishikiori H, Kuronuma K, Hirota K, Yama N, Suzuki T, Onodera M, et al. Deep-learning algorithm to detect fibrosing interstitial lung disease on chest radiographs. The European Respiratory Journal [Internet]. 2023 Feb 1 [cited 2023 Sep 5];61(2):2102269.

- In 2021 a study used CXRs and CT thorax from chronic fibrosing ILD patients and trained a DCNN AI to predict the presence of absence of fibrosing ILD on the basis of CXR alone

- The deep convolutional network-based AI showed an AUC of 0.91 for differentiating CF-ILD from normal CXRs successfully (all CF-ILD confirmed by MDD) and were generally better than the clinicians in doing so

Nishikiori H, Kuronuma K, Hirota K, Yama N, Suzuki T, Onodera M, et al. Deep-learning algorithm to detect fibrosing interstitial lung disease on chest radiographs. The European Respiratory Journal [Internet]. 2023 Feb 1 [cited 2023 Sep 5];61(2):2102269.
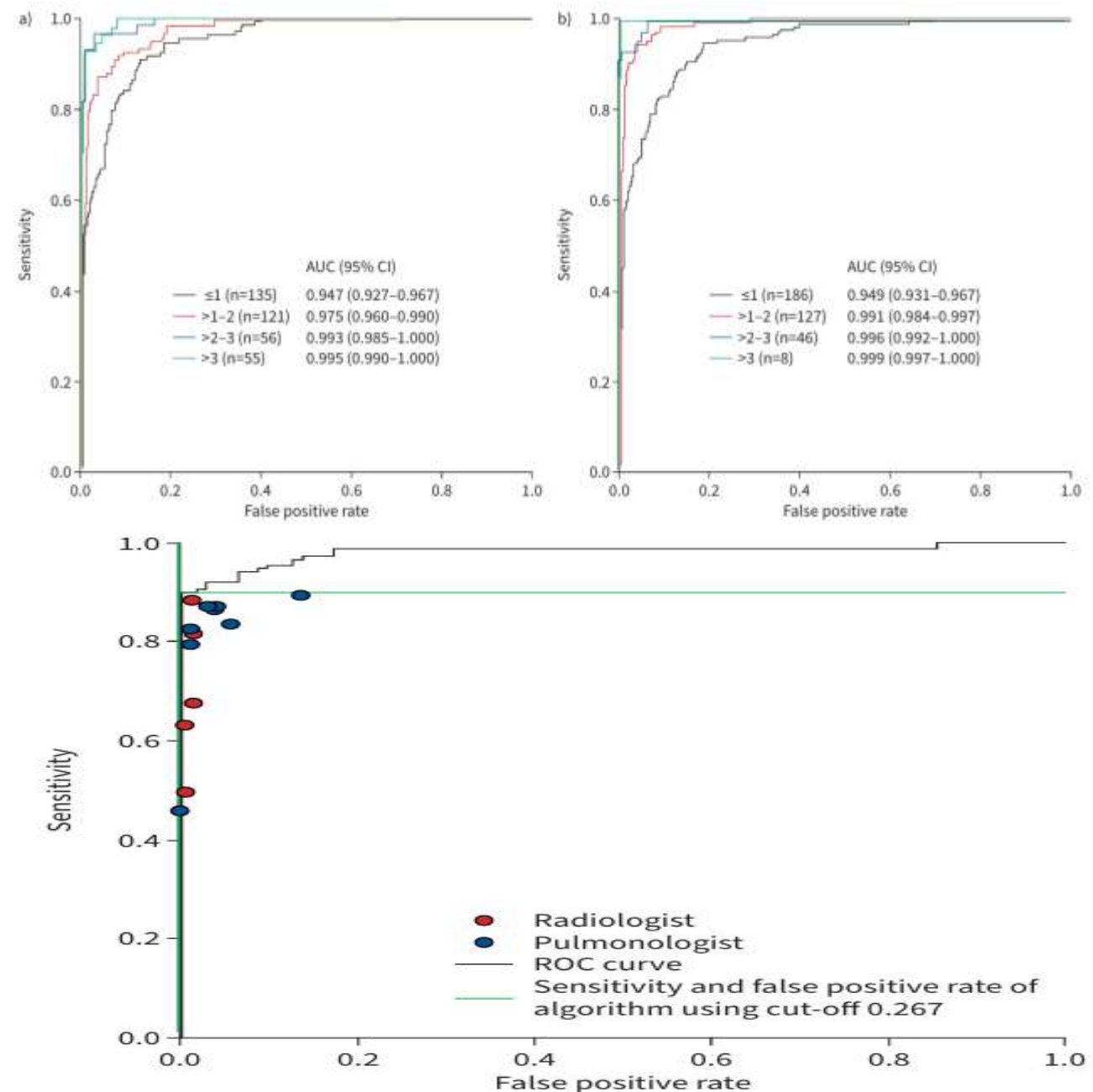
- In 2018, a case-cohort study compared the accuracy of classification of fibrotic interstitial lung diseases diagnosed by AI (deep learning) with that by 91 expert radiologists.

- They trained the AI system with a dataset of 1157 HRCTs, finally comprising of 420096 images (each a montage of four slices) classified according to the 2011 ATS guidelines on diagnosis of pulmonary fibrosis

- The AI was then asked to interpret HRCTs of a test set of 150 individuals (test set B).

- The accuracy of the algorithm on test set A was 73·3%, while that of the radiologists was 70.7%, and the algorithm completed the task in 2.3 seconds

Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. The Lancet Respiratory Medicine. 2018 Nov;6(11):837–45.

- Interobserver agreement between the algorithm and the radiologist's majority opinion was good (κw=0·69), outperforming 56 (62%) of 91 thoracic radiologists.

- The algorithm provided equally prognostic discrimination between usual interstitial pneumonia and non-usual interstitial pneumonia diagnoses (hazard ratio 2·88, 95% CI 1·79–4·61, p<0·0001) compared with the majority opinion of the thoracic radiologists (2·74, 1·67–4·48, p<0·0001).

- The study showed that deep learning algorithms may be a reproducible and accurate, and at the same time time-saving aid or alternative to manual evaluation of CT scan for diagnosis of fibrotic interstitial lung diseases (IPF).
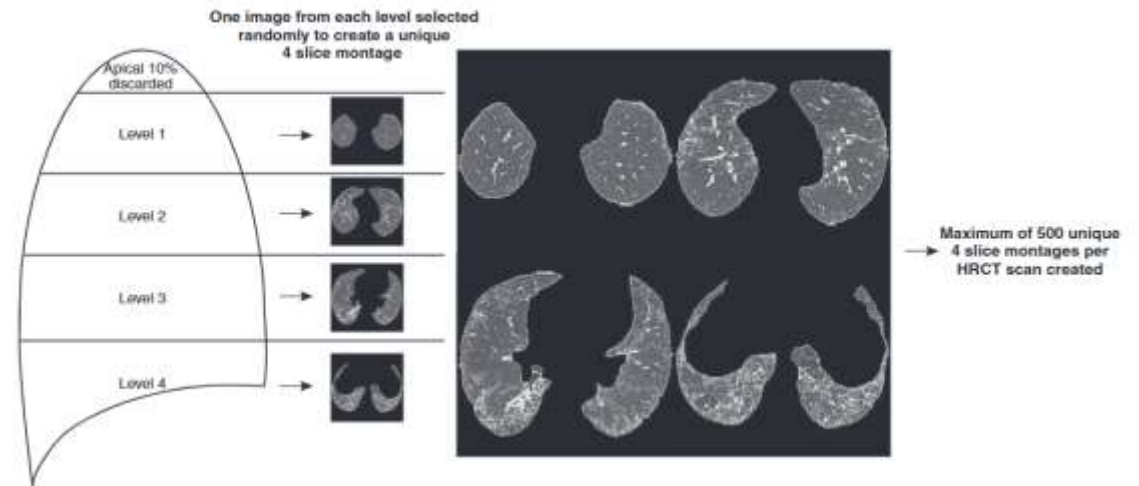
- A 2022 study took patients from an Australian IPF-registry (a good number of registered patients were found to have non-IPF disease) and applied an AI algorithm to them (SOFIA-systemic objective fibrotic image analysis algorithm, based on CNN)

- They compared the diagnosis made by SOFIA and compared them with the diagnosis made by two expert radiologists

- They also assessed whether the diagnosis made by the AI correlated with the disease outcome of the patients (followed up until they had transplant/death or disease progression at 12 months as detected by fall in FVC/DLCO

Simon L. F. Walsh[1], John A. Mackintosh[2], Lucio Calandriello[3], Mario Silva[4], Nicola Sverzellati[4], Anna Rita Larici[3], Stephen M. Humphries[5], David A. Lynch[5], Helen E. Jo[6], Ian Glaspole[7], Christopher Grainge[8], Nicole Goh[9,10,11], Peter M. A. Hopkins[2,12], Yuben Moodley[13], Paul N. Reynolds[14], Christopher Zappala[15], Gregory Keir[16], Wendy A. Cooper[17,18], Annabelle M. Mahar[17], Samantha Ellis[19], Athol U. Wells[1,20], and Tamera J. Corte[6]

[1]National Heart and Lung Institute, Imperial College London, London, United Kingdom; [2]Queensland Lung Transplant Service, The Prince Charles Hospital, Brisbane, Queensland, Australia; [3]Dipartimento di Diagnostica per immagini, Radioterapia, Oncologia ed Ematologia, Fondazione Policlinico Universitario A. Gemelli, Istituto di Ricovero e Cura a Carattere Scientifico, Rome, Italy; [4]Scienze Radiologiche, Dipartimento di Medicina e Chirurgia, Università di Parma, Parma, Italy; [5]Department of Radiology, National Jewish Health, Denver, Colorado; [6]Respiratory Medicine, Royal Prince Alfred Hospital, New South Wales, Australia; [7]Department of Allergy and Respiratory Medicine, Alfred Hospital, Melbourne, Victoria, Australia; [8]Department of Respiratory Medicine, New Lambton Heights, John Hunter Hospital, New South Wales, Australia; [9]Department of Respiratory and Sleep Medicine, Austin Health, Melbourne, Victoria, Australia; [10]Institute for Breathing and Sleep, Melbourne, Victoria, Australia; [11]University of Melbourne, Melbourne, Victoria, Australia; [12]Faculty of Medicine, University of Queensland, Brisbane, Queensland, Australia; [13]School of Medicine & Pharmacology, University of Western Australia, Perth, Western Australia, Australia; [14]Royal Adelaide Hospital Chest Clinic, Adelaide, South Australia, Australia; [15]Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia; [16]Department of Respiratory Medicine, Princess Alexandra Hospital, Brisbane, Queensland, Australia; [17]Tissue Pathology and Diagnostic Oncology, New South Wales Health Pathology, Royal Prince Alfred Hospital, Sydney, New South Wales, Australia; [18]School of Medicine, University of Sydney, Sydney, New South Wales, Australia; [19]Department of Radiology, Alfred Health, Melbourne, Victoria, Australia; and [20]Interstitial Lung Disease Unit, Royal Brompton Hospital, London, United Kingdom

One image from each level selected randomly to create a unique 4 slice montage

Apical 10% discarded

Level 1
Level 2
Level 3
Level 4

Maximum of 500 unique 4 slice montages per HRCT scan created

Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. The Lancet Respiratory Medicine. 2018 Nov;6(11):837–45.

SOFIA was trained during a study in 2019 by 420096 4-slice montages of IPF patients

SOFIA classified the cases as having probability IPF : definite UIP- 0.985 probable UIP- 0.011; indeterminate- 0.002; alternative diagnosis- 0.002

The 2 radiologists evaluated the HRCTs for probabilities of IPF diagnosis (2018 ATS guidelines)
UIP- 75%; probable UIP- 25% indeterminate for UIP- 0%; and alternative diagnosis- 0%

By use of PIOPED model classified them

By use of PIOPED model classified them

- UIP not included in the differential, 0–4%;
- Low probability of UIP, 5–29%;
- Intermediate probability of UIP, 30–69%;
- High probability of UIP, 70–94%;
Pathognomonic for UIP, 95–100%

- UIP not included in the differential, 0–4%;
- Low probability of UIP, 5–29%;
- Intermediate probability of UIP, 30–69%;
- High probability of UIP, 70–94%;
Pathognomonic for UIP, 95–100%

Compared the results with the actual follow up data of the registry

Only SOFIA-PIOPED probability of UIP was significantly related with transplant-free-survival in the patients;
The relation remained significant even after being adjusted for ILD extent, age and sex;

Number at risk

sofia_pioped_uip = 0	19		2	0
sofia_pioped_uip = 1	11		4	0
sofia_pioped_uip = 2	105		1	1
sofia_pioped_uip = 3	42		0	0
sofia_pioped_uip = 4	42	12	3	0	0

——— sofia_pioped_uip = 0   — — — - sofia_pioped_uip = 1
· · · · · · · · · · sofia_pioped_uip = 2   — · — · — sofia_pioped_uip = 3   – · · — · · — · · – sofia_pioped_uip = 4
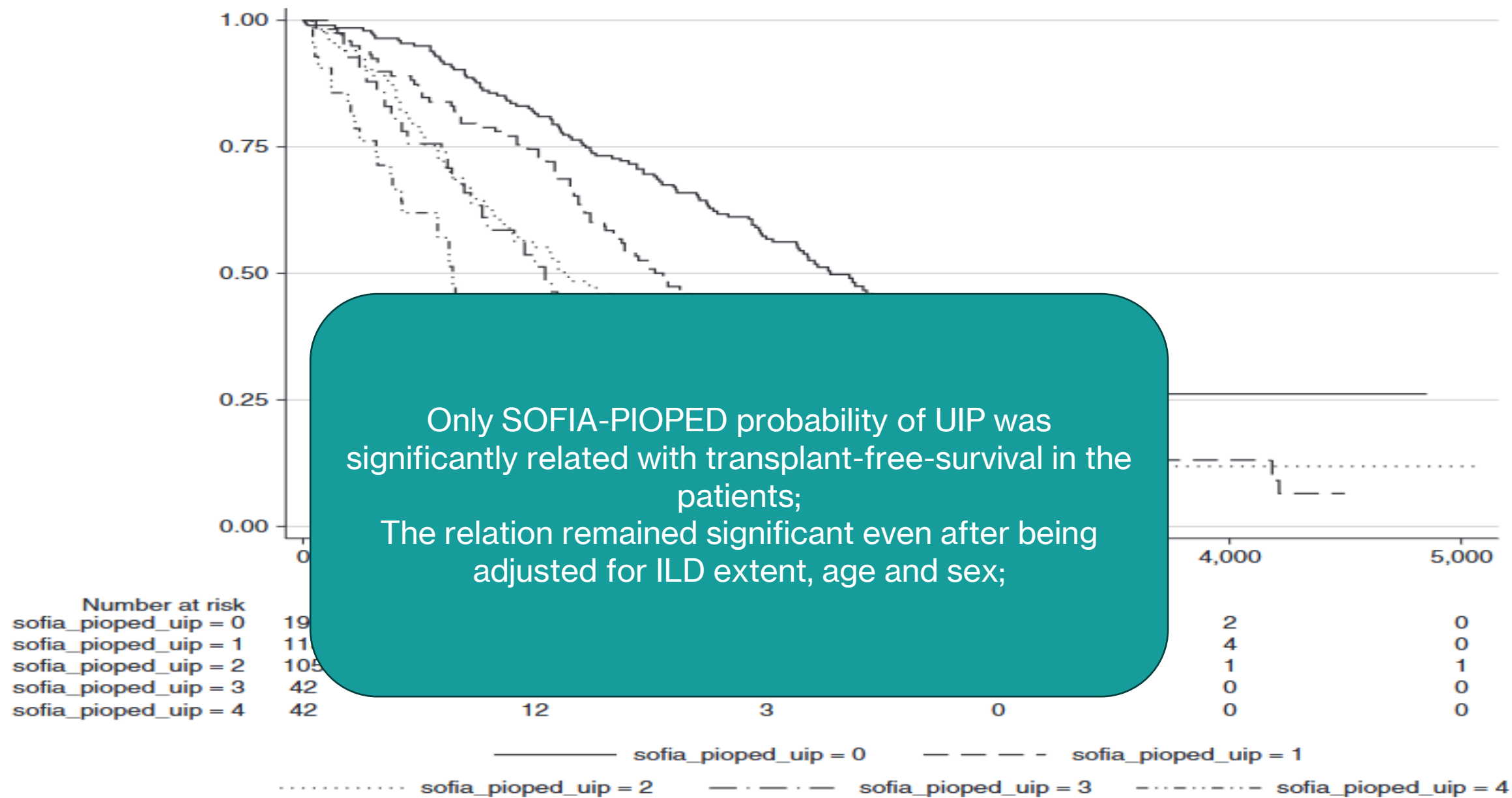
Figure 3. Kaplan-Meier analysis of survival differences between patients assigned to SOFIA-PIOPED (Systematic Objective Fibrotic Imaging Analysis Algorithm–Prospective Investigation of Pulmonary Embolism Diagnosis) usual interstitial pneumonia categories.

	HR	P Value	CI 95%
SOFIA PIOPED UIP probability categories	1.52	<0.0001	1.38-1.67
%Predicted FVC (n=356) *	0.08	<0.0001	0.04-0.16
SOFIA PIOPED UIP probability categories	1.33	<0.0001	1.20-1.48
%Predicted DLco (n=313) *	0.02	<0.0001	0.01-0.05
SOFIA PIOPED UIP probability categories	1.33	<0.0001	1.19-1.47
CPI (n=309) *			

Table E2. Cox p... function. *Limited... direction.

Table 7. Cox P... Pneumonia Prob... Biopsy (n = 86)

Variable (n = 83*)	HR	P Value	95% CI
SOFIA-PIOPED UIP probability categories	1.75	<0.0001	1.37–2.25
Guideline histological pattern	1.29	0.109	0.94–1.78
Total ILD extent (1% increments)	1.01	0.237	0.99–1.02

Radiologists' indeterminate group

- SOFIA-PIOPED classifications were significantly related to disease progression at 12 months (FVC, DLCO)
- SOFIA-PIOPED classification were more significantly related with survival than SLB results
- It re-classified a number of patients radiologists had put in indeterminate class (21 of 83)

Increasing SOFIA-PIOPED probability category was associated with a 2.37-fold increased likelihood of progressive disease at 12 months

- This study is pivotal in showing that certain AI algorithms can aid in not only accurately diagnose but also predict mortality in ILDs

- Prevents misclassification and delays in treatment (antifibrotic)

However,
- The study had very few patients undergoing biopsy

- Radiologists' personal opinions rather than consensus of a number of radiologists were used

- The agreement between radiologists and SOFIA was fair at best (weighted kappa-0.35)

- Others have attempted to differentiate IPF from non-IPF on the basis of wedges of CT images evaluated by AI algorithm, but only found that a higher disease severity is needed to improve diagnostic accuracy

- In line with the previous studies, it has been seen that even with robust training-dataset, diagnostic accuracy of AI remains around 65% unless clinical data are also involved in the classification process when the accuracy goes up to >80%

Deep Learning of Computed Tomography Virtual Wedge Resection for Prediction of Histologic Usual Interstitial Pneumonitis

Hiram Shaish[1], Firas S. Ahmed[1], David Lederer[2], Belinda D'Souza[1], Paul Armenta[1], Mary Salvatore[1], Anjali Saqi[3], Sophia Huang[1], Sachin Jambawalikar[1], and Simukayi Mutasa[1]

[1]Department of Radiology and [3]Department of Pathology, Columbia University Medical Center, New York, New York; and [2]Regeneron Pharmaceuticals, Tarrytown, New York

ORCID IDs: 0000-0002-9914-528X (H.S.); 0000-0001-5258-0228 (D.L.).

ROC Curve for Model
Area Under the Curve = 0.7417
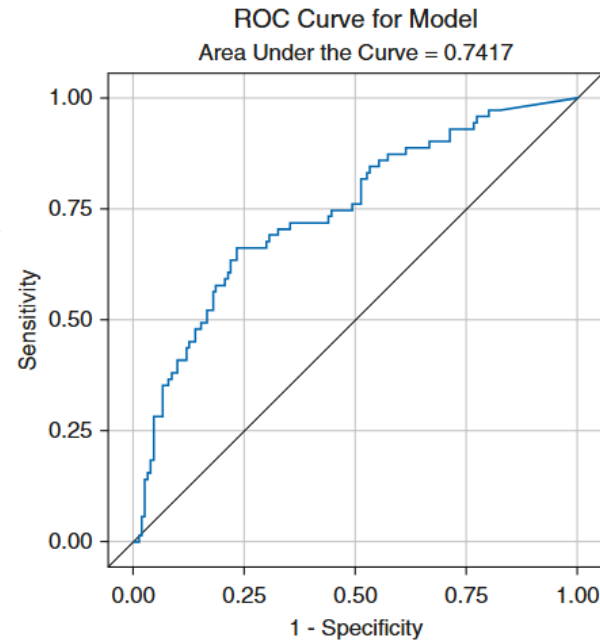
Sensitivity 74%   Specificity 58%

Testing Pathology Dx of UIP	CNN-predicted UIP		
	No [n (%)]	Yes [n (%)]	Total [n (%)]
No	35 (44.87)	21 (26.92)	56 (71.79)
Yes	4 (5.13)	18 (23.08)	22 (28.21)
Total	39 (50.00)	39 (50.00)	78 (100.00)

- Shaish H, Ahmed FS, Lederer DJ, D'Souza B, Armenta PM, Salvatore MM, et al. Deep Learning of Computed Tomography Virtual Wedge Resection for Prediction of Histologic Usual Interstitial Pneumonitis. Annals of the American Thoracic Society. 2021 Jan 1;18(1):51–9.
- Furukawa T, Oyama S, Yokota H, Yasuhiro Kondoh, Kataoka K, Takeshi Johkoh, et al. A comprehensible machine learning tool to differentially diagnose idiopathic pulmonary fibrosis from other chronic interstitial lung diseases. Respirology. 2022 Jun 13;27(9):739–46.

- Others have attempted to differentiate IPF from non-IPF on the basis of wedges of CT images evaluated by AI algorithm, but only found that a higher disease severity is needed to improve diagnostic accuracy

- In line with the previous studies, it has been seen that even with robust training-dataset, diagnostic accuracy of AI remains around 65% unless clinical data are also involved in the classification process when the accuracy goes up to >80%

A comprehensible machine learning tool to differentially diagnose idiopathic pulmonary fibrosis from other chronic interstitial lung diseases

Taiki Furukawa[1,2,3] | Shintaro Oyama[2,3] | Hideo Yokota[2,4] | Yasuhiro Kondoh[5] | Kensuke Kataoka[5] | Takeshi Johkoh[6] | Junya Fukuoka[7] | Naozumi Hashimoto[1] | Koji Sakamoto[1] | Yoshimune Shiratori[3] | Yoshinori Hasegawa[8]
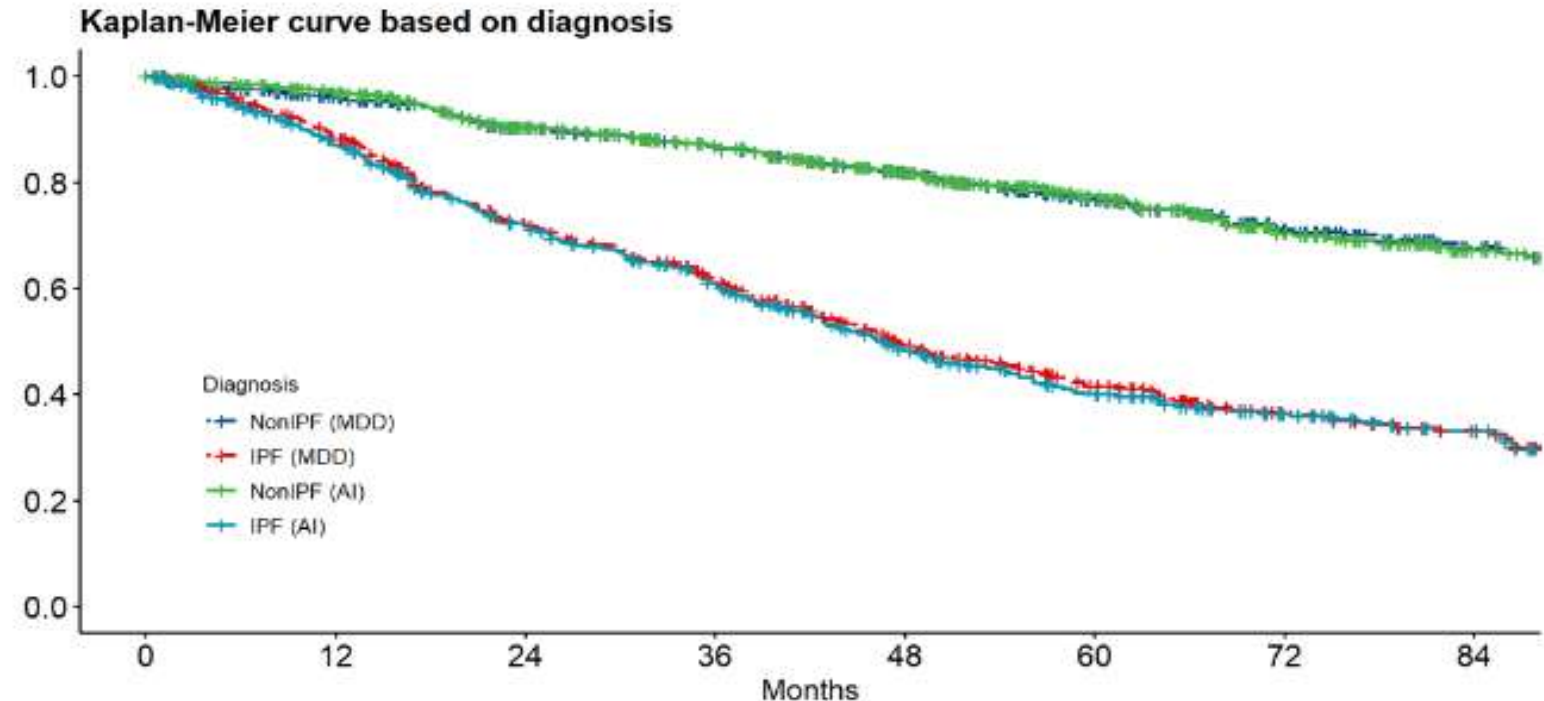
Kaplan-Meier curve based on diagnosis

Diagnosis
- NonIPF (MDD)
- IPF (MDD)
- NonIPF (AI)
- IPF (AI)

- Shaish H, Ahmed FS, Lederer DJ, D'Souza B, Armenta PM, Salvatore MM, et al. Deep Learning of Computed Tomography Virtual Wedge Resection for Prediction of Histologic Usual Interstitial Pneumonitis. Annals of the American Thoracic Society. 2021 Jan 1;18(1):51–9.
- Furukawa T, Oyama S, Yokota H, Yasuhiro Kondoh, Kataoka K, Takeshi Johkoh, et al. A comprehensible machine learning tool to differentially diagnose idiopathic pulmonary fibrosis from other chronic interstitial lung diseases. Respirology. 2022 Jun 13;27(9):739–46.
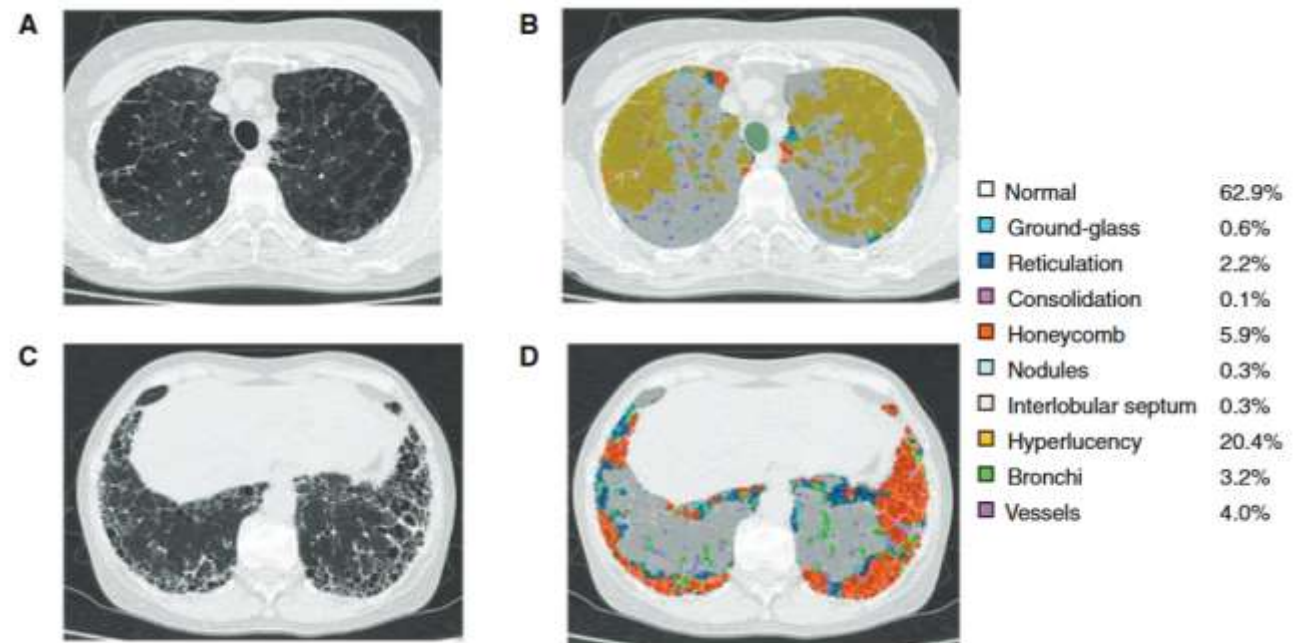
- A 2022 study evaluated an AI-based quantitative CT-assessment tool called AIQCT (Kiyoto U+ Fujifilm) for its correlation with visual diagnosis by radiologist and pulmonologists, with lung function tests and with prognosis

- AIQCT automatically gives data regarding multiple parameters in the CT categorized under lung extraction, airway extraction, pulmonary vessel extraction, and lung parenchyma segmentation: 10 patterns were automatically generated based on various values of these 4 segments

Novel Artificial Intelligence-based Technology for Chest Computed Tomography Analysis of Idiopathic Pulmonary Fibrosis

Tomohiro Handa[1,2], Kiminobu Tanizawa[1], Tsuyoshi Oguma[1], Ryuji Uozumi[3], Kizuku Watanabe[1], Naoya Tanabe[1], Takafumi Niwamoto[1], Hiroshi Shima[1], Ryobu Mori[1], Tomomi W. Nobashi[4], Ryo Sakamoto[4], Takeshi Kubo[4], Atsuko Kurosaki[5], Kazuma Kishi[6], Yuji Nakamoto[4], and Toyohiro Hirai[1]

[1]Department of Respiratory Medicine, [2]Department of Advanced Medicine for Respiratory Failure, [3]Department of Biomedical Statistics and Bioinformatics, and [4]Department of Diagnostic Imaging and Nuclear Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan; [5]Department of Diagnostic Radiology, Fukujuji Hospital, Kiyose, Tokyo, Japan; and [6]Department of Respiratory Medicine, Graduate School of Medicine, Toho University, Tokyo, Japan

ORCID IDs: 0000-0002-3378-6412 (T. Handa); 0000-0002-5719-0744 (K.T.); 0000-0002-9546-9869 (R.U.); 0000-0002-7481-0212 (N.T.); 0000-0003-2595-7450 (T.N.); 0000-0001-9781-7321 (T.W.N.); 0000-0003-1310-9403 (R.S.); 0000-0002-3397-4472 (T.K.); 0000-0002-2803-1488 (A.K.); 0000-0002-6801-0147 (K.K.); 0000-0001-5783-8048 (Y.N.).

Normal	62.9%
Ground-glass	0.6%
Reticulation	2.2%
Consolidation	0.1%
Honeycomb	5.9%
Nodules	0.3%
Interlobular septum	0.3%
Hyperlucency	20.4%
Bronchi	3.2%
Vessels	4.0%

Figure 1. Visualization and quantification of lung parenchymal lesions using the artificial intelligence-based quantitative computed tomographic system. (A–D) Representative high-resolution CT scans (A and C) and corresponding overlayed images (B and D) are shown.

Handa T, Kiminobu Tanizawa, Tsuyoshi Oguma, Ryuji Uozumi, Watanabe K, Tanabe N, et al. Novel Artificial Intelligence-based Technology for Chest Computed Tomography Analysis of Idiopathic Pulmonary Fibrosis. Annals of the American Thoracic Society. 2022 Mar 1;19(3):399–406.

- It was found that for most of the features there was moderate to strong correlation with visual observation

- The radiological features of ILD as determined by AIQCT correlated well with the spirometric findings

- Especially the bronchial volume and the label "normal lung volume", when combined with GAP staging, showed a significant correlation with mortality

- It may reflect the previously established correlation between traction bronchiectasis and poor prognosis in fibrotic ILDs

- Use of antifibrotic drugs, reticulations and bronchial volume were independently associated with disease progression/survival
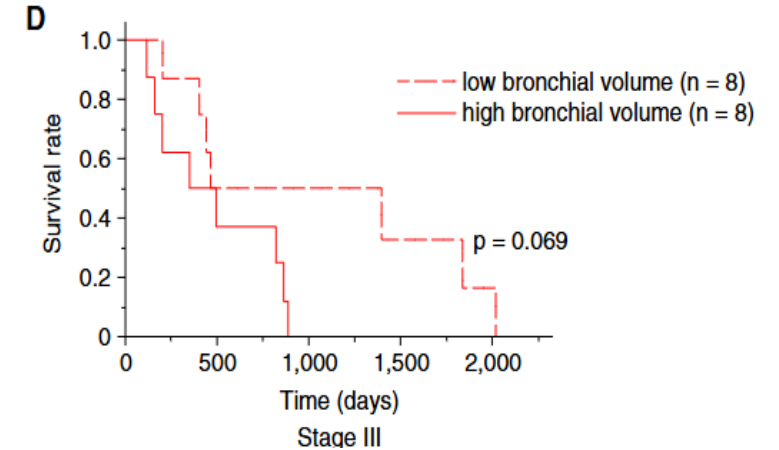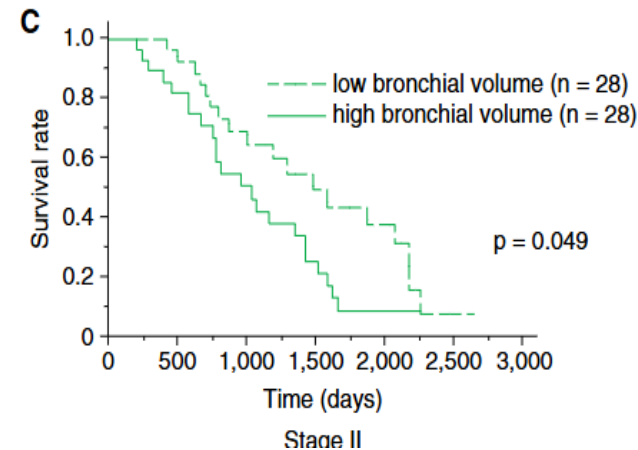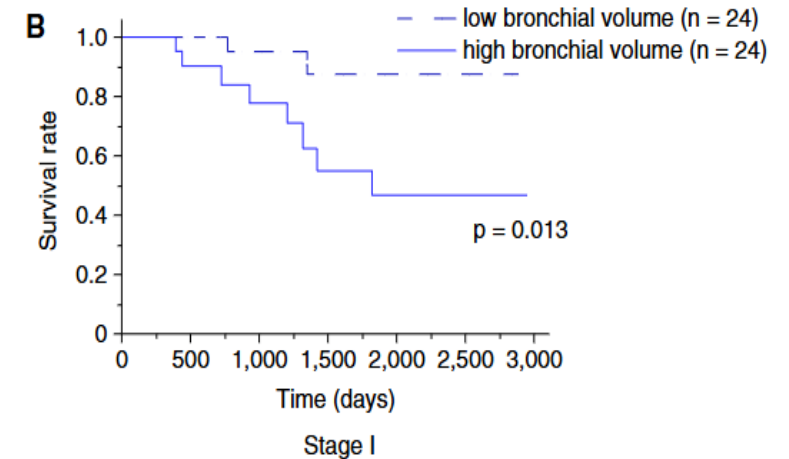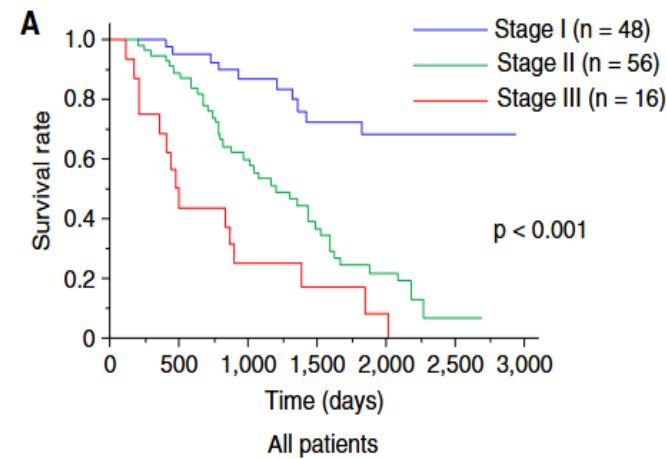
- But it must be kept in mind that AI was not always correct in classification, even in training set

- It was not refined enough to differentiate between small features in CT scan

Original labeling (28 patterns)	Number of labeled images used for training	Final labeling (10 patterns)
Normal lung	816	Normal lung
Borderline between normal and hyperlucency	507	Normal lung
Faint ground-glass opacity	496	Normal lung
Borderline dilatation of bronchioles	183	Normal lung
Ground-glass opacity	372	Ground-glass opacity
Centrilobular ground-glass opacity	108	Ground-glass opacity
Nodular ground-glass opacity	69	Ground-glass opacity
Reticulation	274	Reticulation
Fine reticulation	213	Reticulation
Consolidation	269	Consolidation
Pleural effusion	59	Consolidation
Pleural thickening	47	Consolidation
Honeycombing	178	Honeycombing
Tree-in-bud	168	Nodules
Small nodules (not centrilobular)	210	Nodules
Centrilobular nodules	90	Nodules
Interlobular septal thickening	409	Interlobular septum
Hyperlucency	225	Hyperlucency
Cyst	179	Hyperlucency
Centrilobular emphysema	235	Hyperlucency
Panlobular emphysema	112	Hyperlucency
Cavity surrounded by infiltration	51	Hyperlucency

Detection of tuberculosis:

- Detection of tuberculosis on basis of chest X ray (screening)

- Patterns of tubercular involvement in radiographs/CT

- Assessment of severity of tuberculosis on radiological basis

- A 2023 systematic review different studies evaluat intelligence in detecting images

- It found ree of se spec tudies u up used, im tec s C con n smea B

- The sensitivity and specific between 0.58-0.99 and 0.57-0 respectively

Pooled sensitivity- 94%

High degree of heterogeneity

Imaging modality CXR

Imaging modality CXR

Sputum smear gold standard
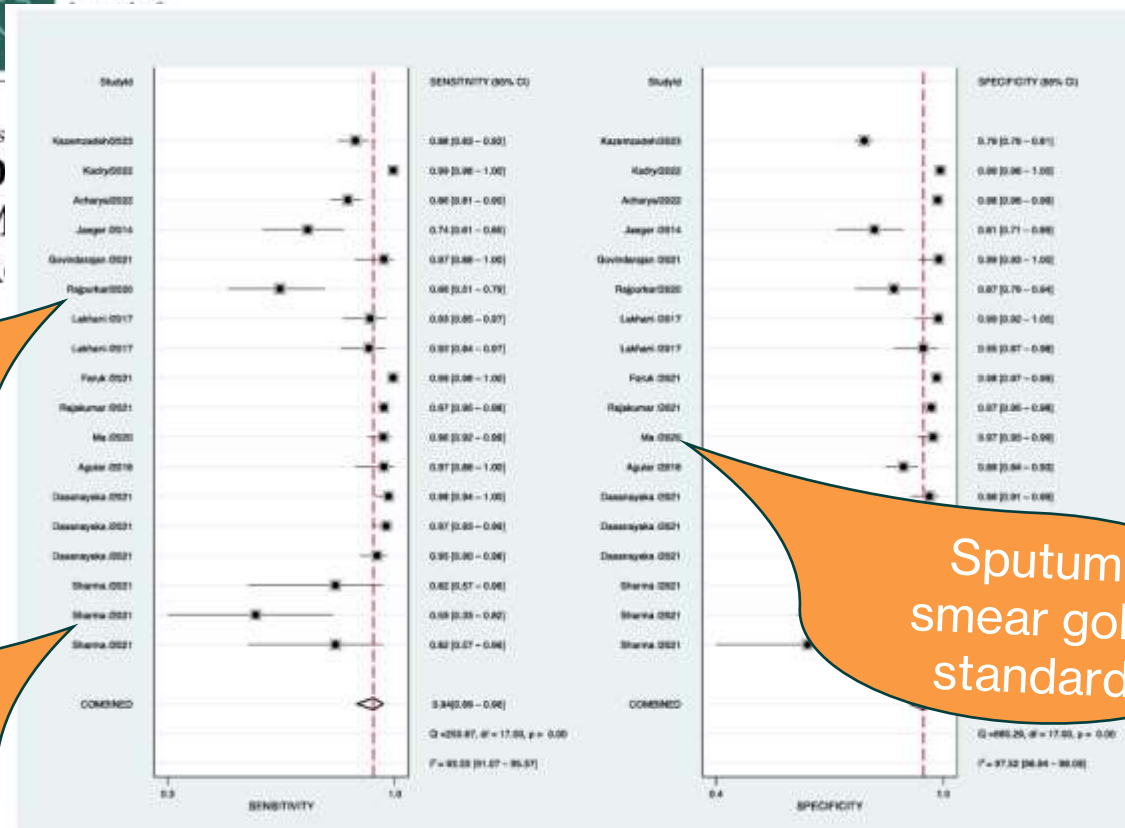
Risk of bias was high (and uncertain) in large number of studies

Zhan Y, Wang Y, Zhang W, Ying B, Wang C. Diagnostic Accuracy of the Artificial Intelligence Methods in Medical Imaging for Pulmonary Tuberculosis: A Systematic Review and Meta-Analysis. Journal of Clinical Medicine. 2022 Dec 30;12(1):303.

Table 2. Subgroup analysis based on different standards.

Studies	Sensitivity (95%CI)	Specificity (95%CI)	DOR (95%CI)	AUC (95%CI)
All (23)	0.91(0.89–0.93)	0.65(0.55–0.75)	20(13–29)	0.91(0.89–0.94)
Study Design				
Prospective (12)	0.91(0.87–0.94)	0.48(0.34–0.62)	9(4–20)	0.85(0.82–0.88)
				92)
				86)
				80)
				93)
				97)
				82)
AI type				
Deep learning (13)	0.91(0.89–0.92)	0.62(0.48–0.74)	16(10–23)	0.91(0.88–0.93)
Machine learning (9)	0.93(0.85–0.97)	0.61(0.46–0.75)	21(11–42)	0.87(0.83–0.89)

Abbreviation: DOR, diagnostic odds ratio; AUC, area under curve.

- Sensitivity was higher when CT scan was used as imaging modality
- Specificity was better when it was compared with human reader, sputum smear which have poor sensitivity
- Specificity was poorer when highly sensitive tests such as GeneXpert was used as golden standard
- Accuracy differed according to the AI models used: deep learning and machine learning had better sensitivity but poor specificity

- In 2021, a retrospective cohort study evaluated a fully automated CNN-based system to identify patients with imaging features suggestive of TB as well as their severity of involvement

COMPUTED TOMOGRAPHY

A fully automatic artificial intelligence–based CT image analysis system for accurate detection, diagnosis, and quantitative severity evaluation of pulmonary tuberculosis

Chenggong Yan[1,2] · Lingfeng Wang[3,4] · Jie Lin[1,5] · Jun Xu[6] · Tianjing Zhang[4] · Jin Qi[3] · Xiangying Li[7] · Wei Ni[8] · Guangyao Wu[2,9] · Jianbin Huang[1] · Yikai Xu[1] · Henry C. Woodruff[2,10] · Philippe Lambin[2,10]

Yan C, Wang L, Lin J, Xu J, Zhang T, Qi J, et al. A fully automatic artificial intelligence–based CT image analysis system for accurate detection, diagnosis, and quantitative severity evaluation of pulmonary tuberculosis. European Radiology. 2021 Nov 29;32(4):2188–99.
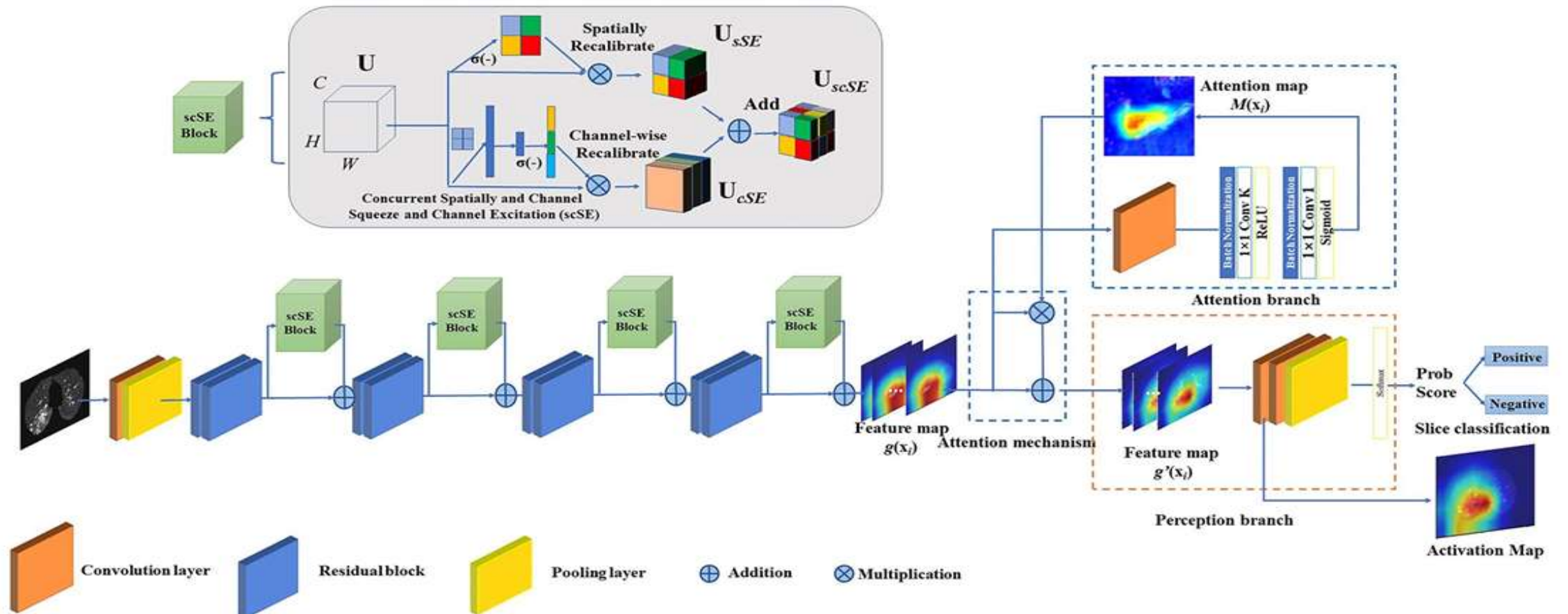
- They assessed how well the AI-detected TB cases and severity correlated with two independent radiologists

- The AI differentiated normal from abnormal CT, detected the sites of involvement, predicted the activity of the disease and commented on disease severity

Dataset	Detected candidate regions	True +ve	False +ve	PPV
Test 2	563	518	45	92%
Test 3	502	440	62	87.6%
Test 4	931	869	62	93.3%

Ground truth: consensus by two radiologists

a Severe | Non-severe

TB score(RUL): 0.15
Subjective score (RUL):1
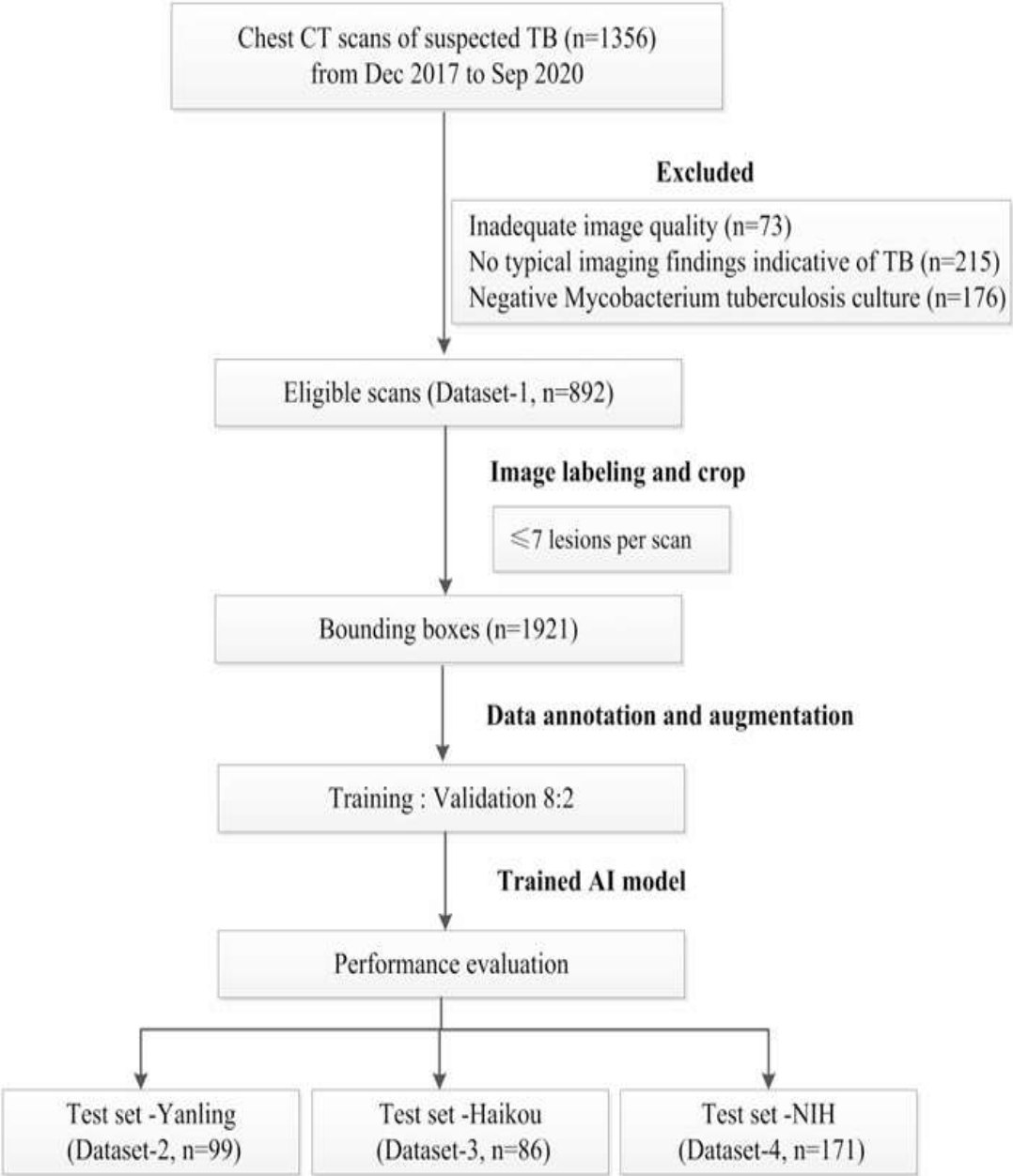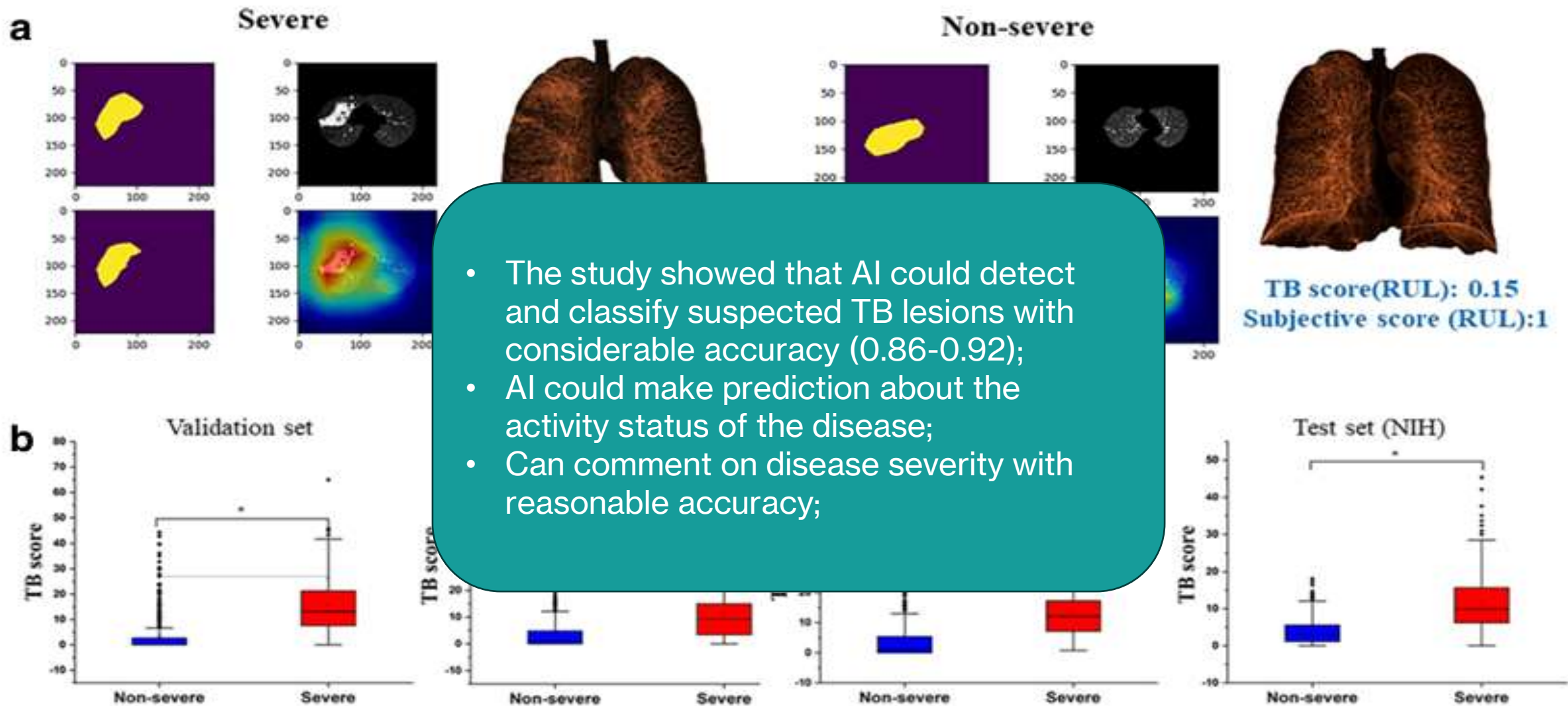
- The study showed that AI could detect and classify suspected TB lesions with considerable accuracy (0.86-0.92);
- AI could make prediction about the activity status of the disease;
- Can comment on disease severity with reasonable accuracy;

b Validation set | Test set (NIH)

TB score — Non-severe | Severe

Lung TB score: Ratio of lesion volumetric summation to that of the corresponding lung lobes
Severe -LTS >2 in any lobe; Non-severe-LTS <2

Pulmonary vascular diseases:

- Pulmonary hypertension

- Pulmonary embolism

- Problem with early diagnosis of pulmonary hypertension through imaging modality is that the affected pulmonary arterioles are not visible properly with the best resolution of images

- Artificial intelligence is largely based on the power of the imaging modality to clearly delineate the region of interest

- Imaging solely is not currently developed enough to utilize AI in diagnosing and predicting pulmonary hypertension in patients

- But as RHC is an invasive procedure and any advance in diagnosing PH with imaging should be explored

- In 2016 a study used MRI of the MPA and heart along with computational models that take into account windkessel effect and applies them through random forest decision tree to diagnose pulmonary hypertension

- They evaluated the diagnostic accuracy of different combinations of imaging and computational data with the RHC being the gold standard

Diagnosis of pulmonary hypertension from magnetic resonance imaging–based computational models and decision tree analysis

Angela Lungu,[1,2] Andrew J. Swift,[1,2] David Capener,[1] David Kiely,[2,3] Rod Hose,[1,2] Jim M. Wild[1,2]

[1]Cardiovascular Science Department, University of Sheffield, Sheffield, South Yorkshire, United Kingdom; [2]Insigneo Institute for in silico Medicine, University of Sheffield, Sheffield, South Yorkshire, United Kingdom; [3]Sheffield Pulmonary Vascular Disease Unit, Royal Hallamshire Hospital, Sheffield, South Yorkshire, United Kingdom

Lungu A, Swift AJ, Capener D, Kiely DG, Hose R, Wild JM. Diagnosis of Pulmonary Hypertension from Magnetic Resonance Imaging–Based Computational Models and Decision Tree Analysis. Pulmonary circulation. 2016 Jun 1;6(2):181–90.
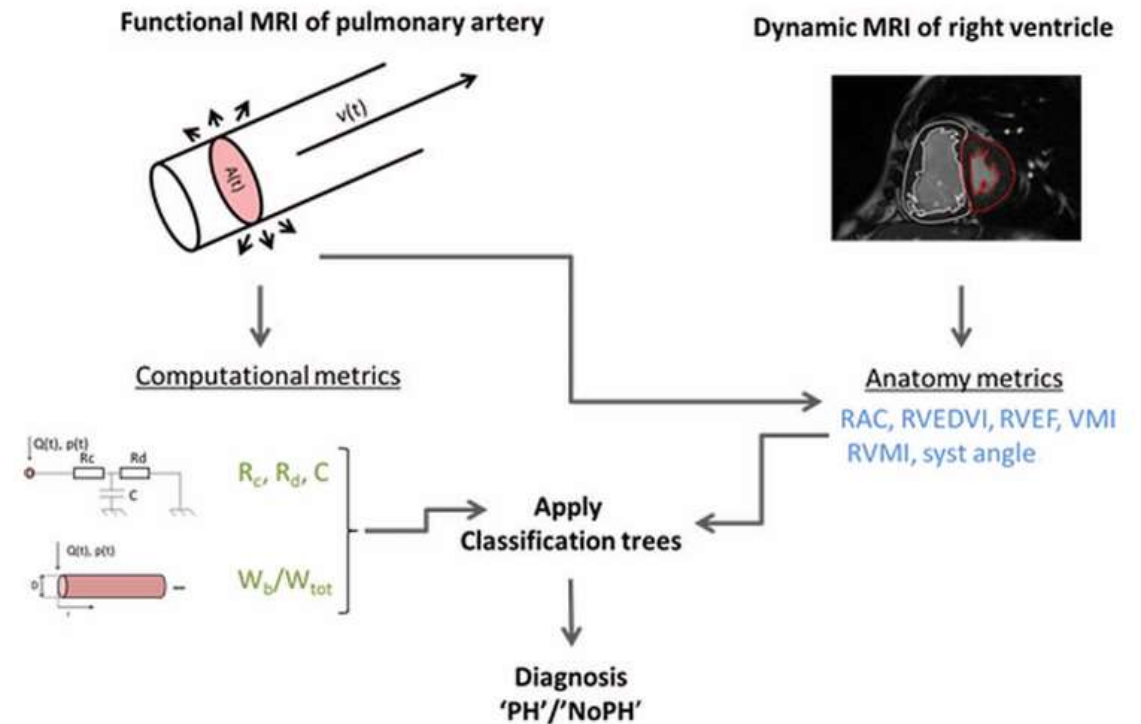
Table 2. Noninvasive metrics' individual accuracies evaluated for a cutoff value corresponding to maximum Youden index

| Noninvasive PH markers (data-driven threshold)	AUC	Misclassification error	Sensitivity	Specificity	Threshold
1D model					
W_b/W_{tot}	0.88	0.25	0.68	1	0.35
0D model					
R_d, mmHg s/mL	0.85	0.25	0.72	0.87	0.45
R_c, mmHg s/mL	0.67	0.38	0.56	0.86	0.01
C, mL/mmHg	0.83	0.25	0.72	0.87	0.88
PA imaging					
RAC, %	0.81	0.29	0.67	0.87	16
CMR					
RVMI, g/m^2	0.81	0.26	0.68	0.93	16.01
RVEDVI, mL/m^2	0.64	0.51	0.36	0.93	0.94
VMI	0.73	0.26	0.74	0.73	0.27
RVEF, %	0.74	0.38	0.56	0.87	43
Systolic septal angle, degrees	0.81	0.35	0.56	1	164

Note: PH: pulmonary hypertension; AUC: area under the curve; 1D: one-dimensional; W_b/W_{tot}: ratio of backward to total wave power; 0D: zero-dimensional (Windkessel); R_d: distal resistance; R_c: characteristic resistance; C: total pulmonary compliance; RAC: relative area change of main pulmonary artery (PA); CMR: cardiac magnetic resonance; RVEDVI: right ventricle end-diastolic volume index; RVEF: right ventricle ejection fraction; VMI: ventricular mass index; RVMI: right ventricle mass index.

Two computational models were included-

1D model: used Wb/Wtot (backward wave to total wave proportion)

0D model: included distal resistance, characteristic resistance and total pulmonary compliance

MRI data had good specificity but poor sensitivity alone
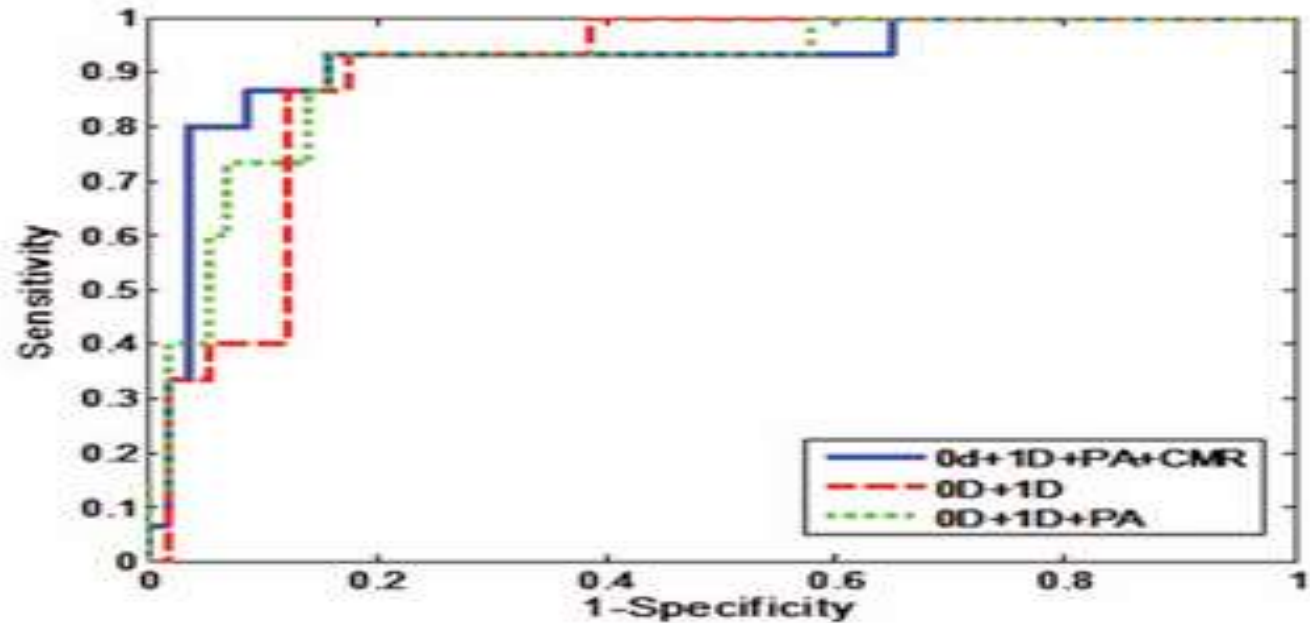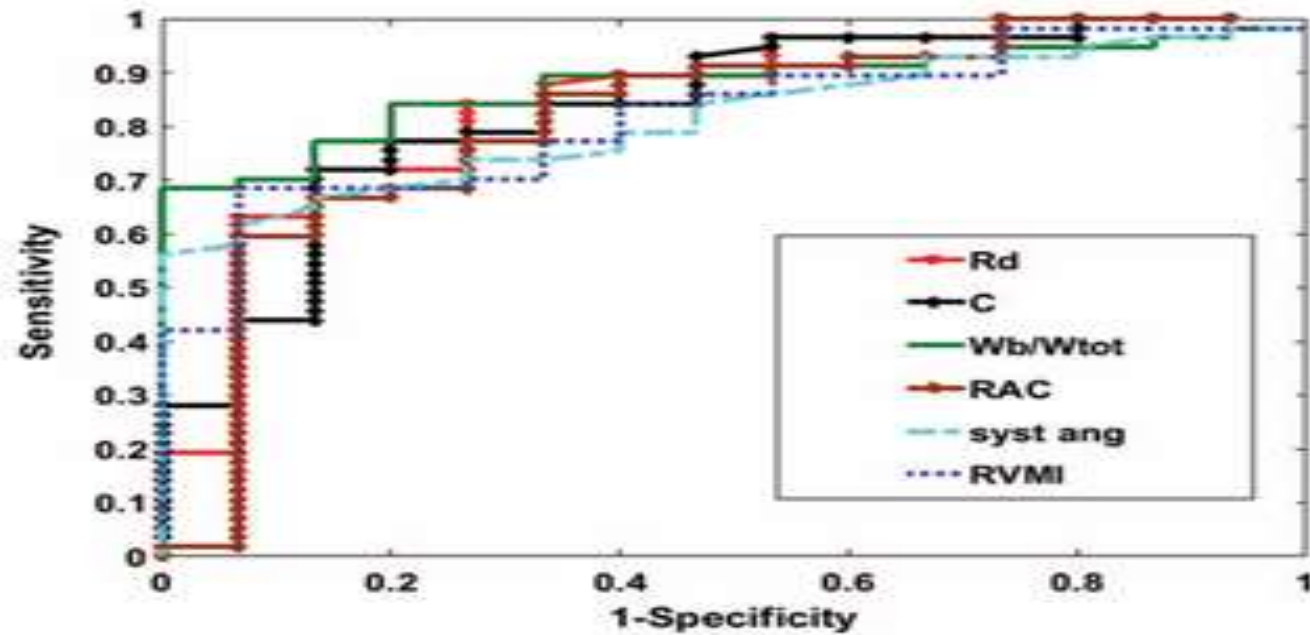
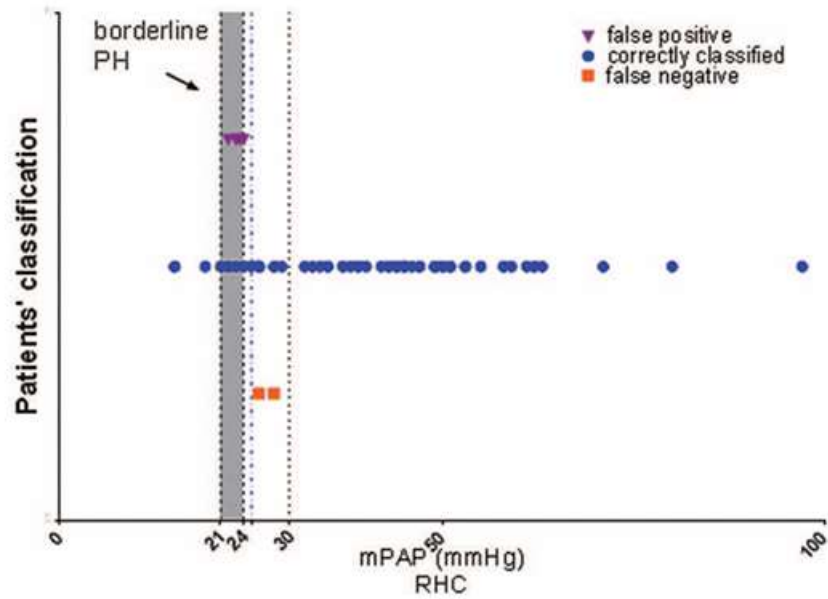It showed that the 0D+1D+PA+CMR together had the best overall accuracy

Table 3. Classification accuracies for the coupled pulmonary hypertension metrics models

LOOCV	AUC	Misclassification error	Sensitivity	Specificity
0D + 1D	0.89	0.21	0.88	0.47
0D + 1D + PA	0.9	0.13	0.93	0.67
0D + 1D + PA + CMR (all)	0.89	0.14	0.97	0.47
0D + 1D + PA + CMR	0.91	0.08	0.97	0.73

Note: LOOCV: leave-one-out cross validation; AUC: area under the curve. For all models, threshold is not applicable. Zero-dimensional (0D) metrics derived from the 0D Windkessel model. One-dimensional (1D) metrics derived from the 1D wave model. Pulmonary artery (PA) metrics derived from the two-dimensional images of the main PA. Cardiac magnetic resonance (CMR) metrics derived solely from measurements on the cardiac images, with AUC > 0.8. CMR (all) includes all measured metrics derived from the cardiac images.

Only 6 out of the 72 patients were misclassified by the model
All the patients who were false negative or false positive had mPAP very close to 25 mmHg
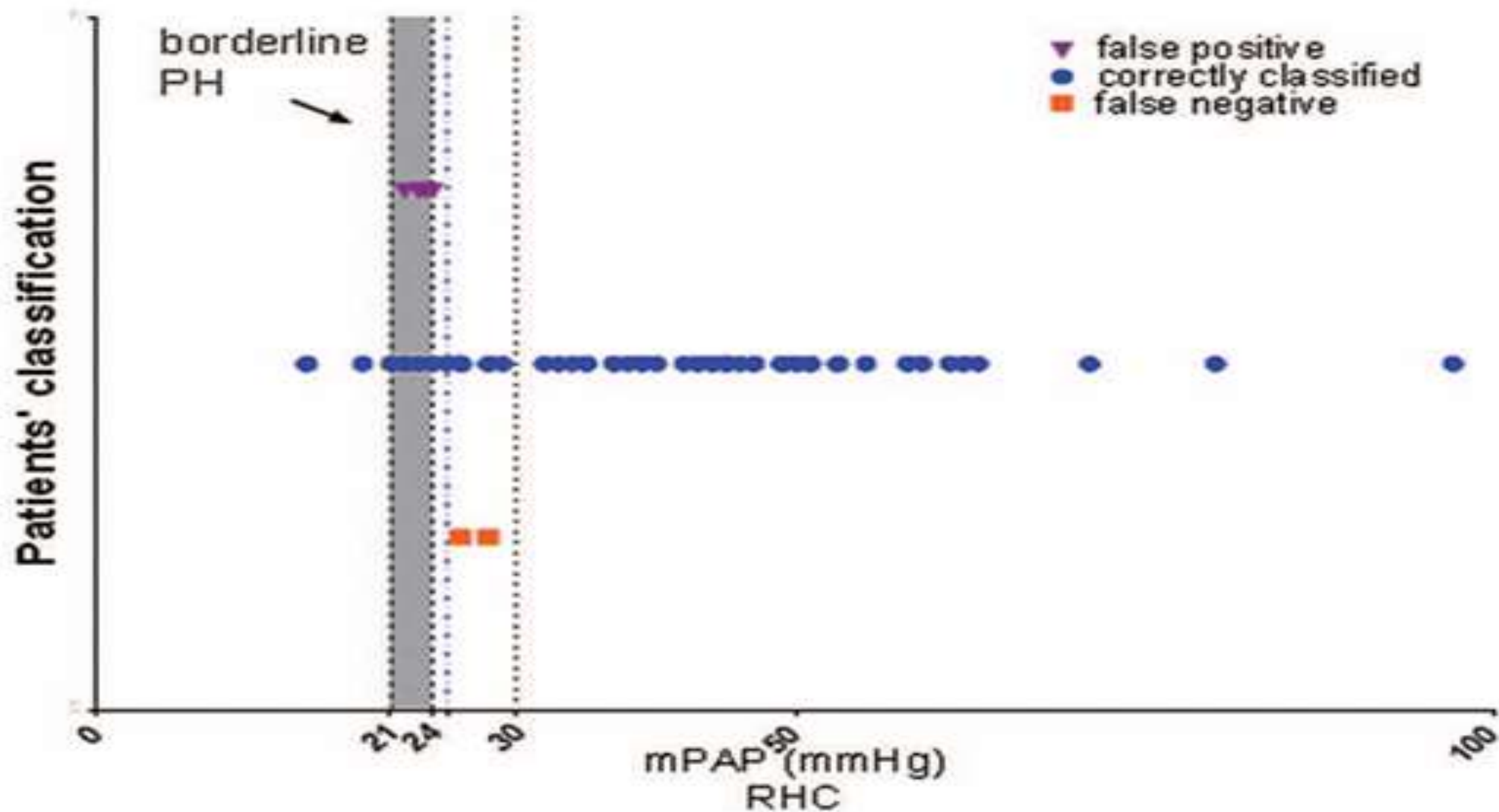
Pulmonary embolism

	Study type / country	Population	Intervention (ML algorithm)	Comparison	Outcome
Gawlitza J et al. (2021)	Retrospective study, unicenter N = 2045 PE prevalence 33 %–43 % Germany	Emergency room Age > 18 Suspected PE, Blood testing CT	LR SVM RF AdaBoost	Cluster analysis	Potential predictors to further increase pre-test probability.
Liu H et al. (2021)	Retrospective study, unicenter N = 573 PE prevalence 41 % China	Inpatients Digital medical database Age < 45	LR DT NN SVM RF	167 inpatients with deep venous thrombosis (DVT) and/or pulmonary embolism (PE) and 406 patients without DVT or PE	To develop and externally validate a new prediction model for young-middle-aged people using machine learning methods
Ryan L (2022)	Retrospective study unicenter N = 63,798 PE prevalence: 0.5 % USA	Inpatients population, electronic health record data from medical and surgical inpatients	LR NN XGBoost	Non PE encounters (%) n = 60,297 PE encounters (%) n = 309	To identify patients at risk of PE before the clinical detection of onset in an inpatient population.
Hou L (2021)	Retrospective study unicenter N = 3619 PE prevalence: 16.9 %–19.1 % China (Sanghai)	Inpatients that received pulmonary computed tomography imaging Age: older than 18 years	LR SVM RF GBDT	Compared the models with representative baseline algorithms, and investigated their predictability and feature interpretation	To establish a novel PE risk prediction model based on ML methods and to evaluate the predictive performance of the model and the contribution of variables to the predictive performance.
Shen J (2022)	Retrospective study multicenter Training model: N = 331,268; PE prevalence: 3.3 % External validation model N = 1,660,715; PE prevalence: 3.7 % USA	Inpatients 12 medical institutions for training model 32 medical institutions for external validation model	XGBoost	External validation of the model from each of 32 medical institutions (total n = 1,660,715; 3.7 % PE positive) without retraining	To conduct the first large-scale external validation of a machine learning-based PE prediction model which uses EHR data from the first three hours of a patient's hospital stay to predict the occurrence of PE within the next 10 days of the inpatient stay.

Sensitivity, specificity, accuracy and AUC for machine learning algorithms.

	ML model	Sensitivity	Specificity	Accuracy	AUC (CI 95 %)
Gawlitza J et al. (2021)	LR	NA	NA	>90 %	0.977 (NA)
	SVM	NA	NA	>90 %	0.945 (NA)
	RF	NA	NA	>90 %	0.947 (NA)
	AdaBoost	NA	NA	>90 %	<0.900 (NA)
Liu H et al. (2021)	LR	61 %	99 %	88 %	0.837 (0.756–0.919)
	DT	59 %	99 %	87 %	0.799 (0.667–0.931)
	NN	46 %	97 %	82 %	0.841 (0.756–0.925)
	SVM	56 %	97 %	85 %	0.875 (0.806–0.944)
	RF	59 %	99 %	87 %	0.850 (0.793–0.907)
Ryan L (2022)	LR	81 %	35 %	NA	0.67 (NA)
	NN	81 %	44 %	NA	0.74 (NA)
	XGBoost	81 %	70 %	NA	0.85 (NA)
Houl L (2021)	LR	68.1 %	66.1 %	66.5 %	0.716 (NA)
	SVM	75 %	65.1 %	67 %,	0.743 (NA)
	RF	71.5 %,	72.7 %,	72.5 %,	0.791 (NA)
	GBDT	63.9 %,	81.1 %	77.8 %	0.799 (NA)
Shen J (2022)	XGBoost	80 %	85 %	NA	0.88 (0.79–0.93)

ML: Machine Learning, PE: Pulmonary Embolism, LR: Logistic Regression, SVM: Support-vector machines, RF: Random Forest; AdaBoost: Adaptative Boosting, GDBT: Gradient Descent Boosting Tree, XGBoost: Extreme Gradient Boosting, DT: decision tree, NN: Feed-forward Neural Network, AUC: area under the curve. NA: not applicable.

- A retrospective study aimed at external validation of a machine-learning algorithm (XGBoost) for prediction of pulmonary embolism in hospitalized patient with high risk of PTE

- The algorithm was trained with a dataset including 330,000 patients
- From 12 institutions

- It was then applied to over 1.7 million patients admitted in 32 different hospitals to predict the occurrence of pulmonary embolism in them

- Gold standard was hospital data documenting PTE with specific ICD-10 code and acquiring thrombolytics and anticoagulants for its treatment

Full Length Article

Massive external validation of a machine learning algorithm to predict pulmonary embolism in hospitalized patients

Jieru Shen, Satish Casie Chetty *, Sepideh Shokouhi, Jenish Maharjan, Yevheniy Chuba, Jacob Calvert, Qingqing Mao

Dascena, Inc., Houston, TX, United States

Shen J, Casie Chetty S, Shokouhi S, Maharjan J, Chuba Y, Calvert J, et al. Massive external validation of a machine learning algorithm to predict pulmonary embolism in hospitalized patients. Thrombosis Research. 2022 Aug;216:14–21.
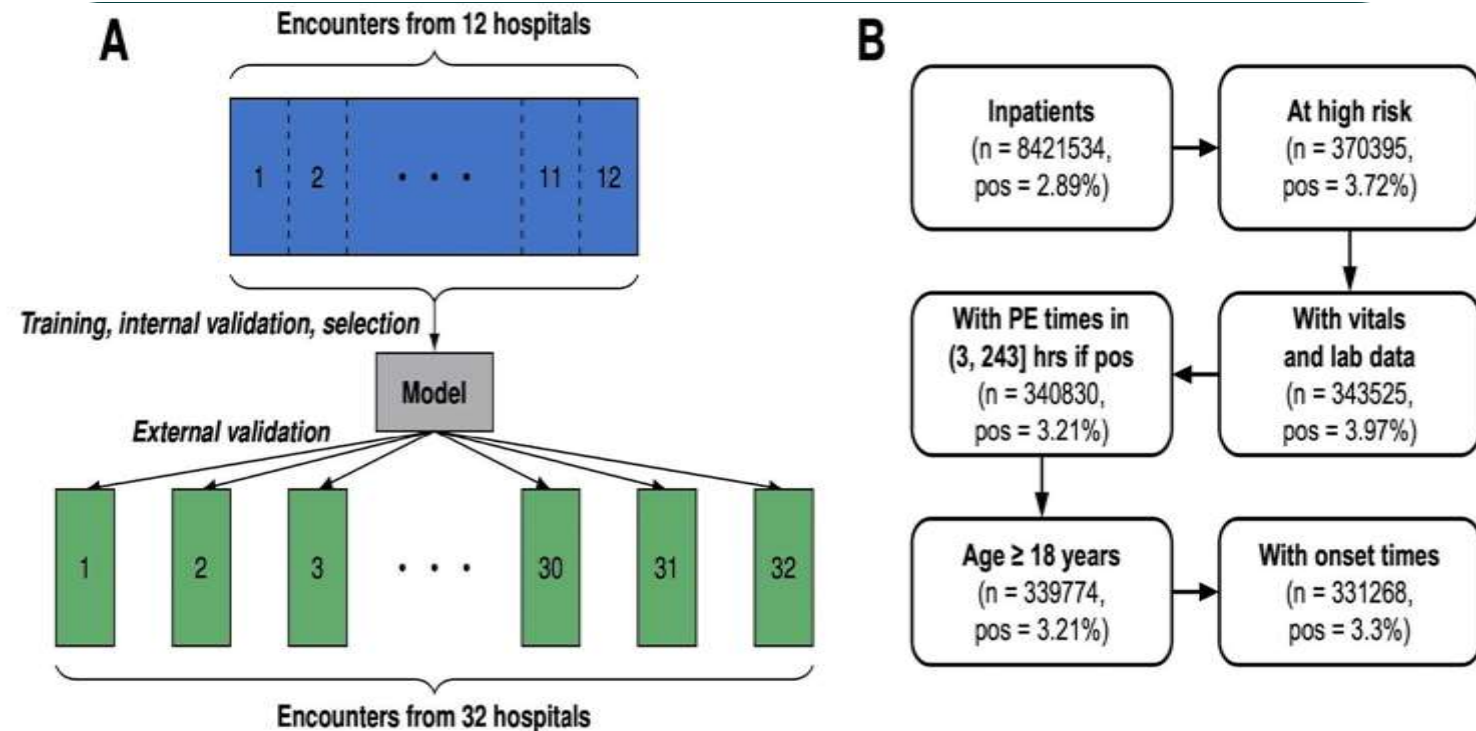
- The machine learning algorithm was found to have moderately good sensitivity in predicting PTE in hospitalized patient

- It still did miss out on some of the PTE cases

- This can be attributable to the lower PTE rates in the training cohort (3.3%) than in the test cohort (3.7%)

- AUROC varied between 0.79 and 0.93 (across the hospitals)
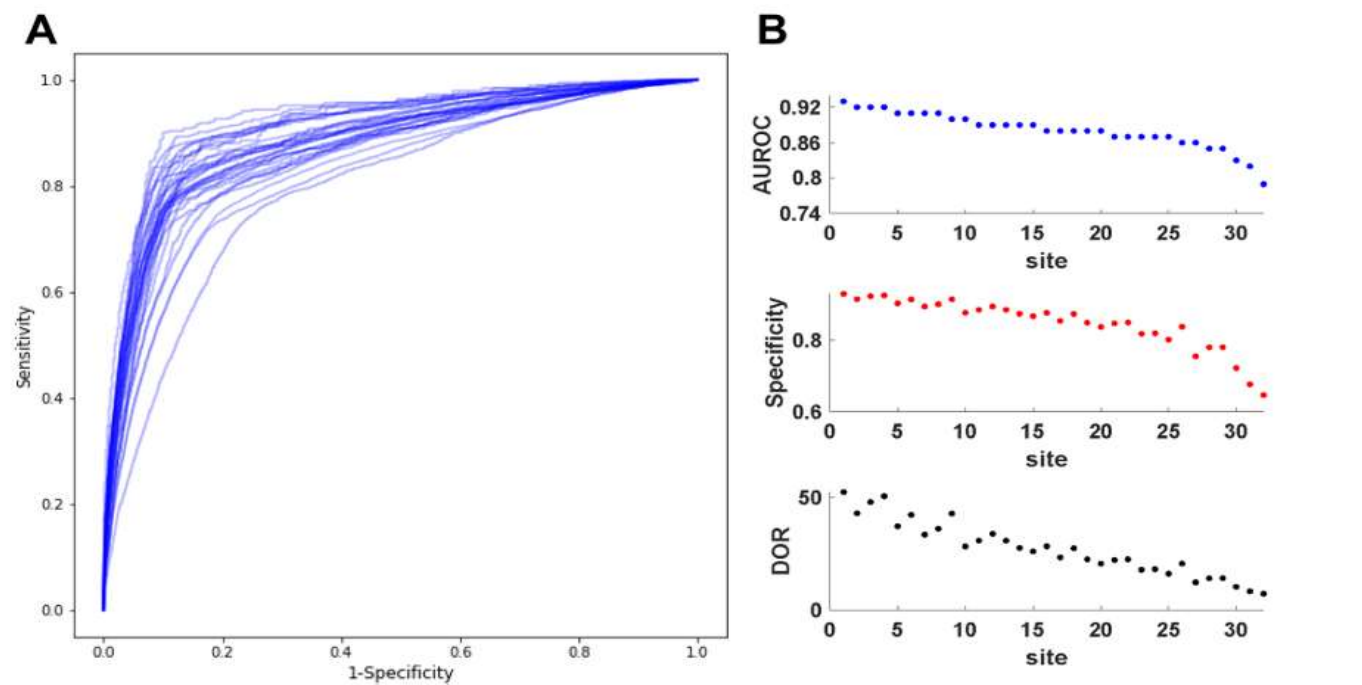- Prior VTE and mean blood pressures contributed most to the AI's decision in predicting PTE

Fig. 2. Performance of the eXtreme Gradient Boosting (XGB) model on the test datasets from 32 institutions. (A) Receiver operator characteristic curves for the 32 test sets. (B) Area under the receiver operating characteristic (AUROC) values for the 32 test sets, in descending order, with corresponding specificity and diagnostic o
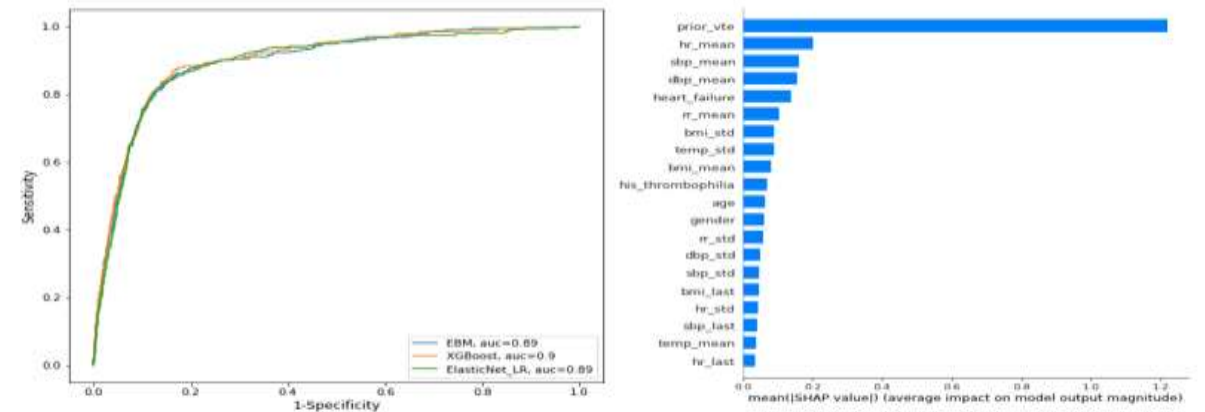
Suppl. Figure 2. Evaluation of model performance. (A) Comparison of performance between Explainable Boosting Machine (EBM), eXtreme Gradient Boosting (XGB), and Elastic Net Logistic Regression (LR) models in terms of AUROC on a hold-out test subset of the training dataset. (B) The top 20 features for the XGB model. Abbreviations: auc — area under the curve, vte — venous thromboembolism, hr — heart rate, sbp — systolic blood pressure, dbp — diastolic blood pressure, rr — respiratory rate, bmi — body mass index.

Obstructive airway disease:

- Diagnosis of obstructive airway diseases are mostly straight-forward and involves inexpensive and accurate investigations which are easily accessible

- Use of expensive imaging in addition to artificial intelligence is not always necessary

- There are few instances where it may be helpful

- Early diagnosis of COPD (before it is identified my spirometry)

- Phenotypical classification of asthma

Study characteristics.

Auther	Country	Model type	Sample size	AUC	Sensitivity	Specificity	Other performance indicators
Chen et al 2022 [17]	China	CNN(ResNet10) + MIL	4823(Developmental set:4552 (NLST:3715),External test set:271)	1.00/ 0.86	0.97/ 0.0.66	1.00/0.83	ACC:1.00/0.78
Dorosti et al 2023[18]	Germany	CNN(DenseNet121)	78(Developmental set:60,Test set:18)	0.86	NA	NA	
Du et al 2020 [19]	China	CNN(AlexNet)	280(10-fold cross-validation)	0.92	NA	NA	ACC:0.89 TP:63 FP:5 FN:27 TN:185
Erdem et al 2023[20]	Türkiye	CNN(AlexNet)	802	1.00	NA	NA	F1-score:1.00 Precision:1.00 Recall:1.00
Gonzá lez et al 2018[21]	Spain	CNN	9983(Developmental set:8983,Test set:1000)(COPD Gene)	0.86	NA	NA	TP:348 FP:116 FN:111 TN:425
Guan et al 2024 [22]	China	CNN(ResNet50)	1024(Developmental set:726,Test set:298)	0.90/ 0.70	0.85/0.63	0.85/0.73	ACC:0.85/0.64 NPV:0.53/0.23 PPV:0.96/0.94
		DT		0.94/ 0.73	0.91/0.84	0.84/0.58	ACC:0.89/0.81 NPV:0.64/0.36 PPV:0.97/0.93
Ho et al 2021 [23]	Korea	CNN	596(5-fold cross-validation)	0.94	0.88	0.94	ACC:0.89 Precision:0.83 F1-Score:0.85
Li(1) et al 2022 [24]	China	GCN	600(Developmental set:500,Test set:100)(DLCST)	0.81	NA	NA	ACC:0.77 F1-Score:0.78 Precision:0.80 TP:40 FP:10 FN:13 TN:37
Li(2) et al 2022 [25]	China	LR	322 (Developmental set:257,Test set:65)	0.97	0.98	0.92	ACC:0.96 Precision:0.98 F1-Score:0.98
Moslemi et al 2022[26]	Canada/ Germany	SVM	95	NA	0.87	0.71	ACC:0.80 F1-Score:0.81
Ramalho et al 2014[27]	Brazil	ELM	72	NA	NA	NA	ACC:0.96
Savadjiev et al 2021[28]	Canada	CNN(ResNet50)	274	0.89	NA	NA	Precision:0.66 Recall:0.85
Silvia et al 2024 [29]	Germany	SSL	9861(Developmental set:1549,Test set:2312)(COPD Gene)	0.85	NA	NA	AUPRC:0.76
Sørensen et al 2012[30]	Denmark	KNN	300(Developmental set:200,Test set:100)	0.71	NA	NA	
Sun et al 2022 [31]	China	CNN(ResNet18) + MIL	2013(Developmental set:1393,External test set:620(NLST))	0.93/ 0.87	0.81/0.80	0.93/0.84	NPV:0.89/0.98 PPV:0.87/0.33 F1-Score:0.89/0.46
Tang et al 2020 [32]	Canada	CNN(ResNet152)	4742(Developmental set:2589 (PanCan),External test set:2153 (ECLIPSE))	0.89	NA	NA	NPV:0.76 PPV:0.85 F1-Score:0.76 Precision:0.80
Wu(1) et al 2023[33]	China	CNN(ResNet26)	581(Developmental set:380,External test set:201)	NA	0.93/0.92	0.97/0.80	ACC:0.95/0.85 NPV:0.93/0.93 PPV:0.97/0.76
Wu(2) et al 2023[34]	China	CNN(VGG-16) + MIL	561(Developmental set:360,External test set:201)	NA	0.95/0.77	0.97/0.89	ACC:0.96/0.83 NPV:0.95/0.80 PPV:0.97/0.88
Xu et al 2020 [35]	China	CNN(AlexNet) + PCA + MIL	280(10-fold cross-validation)	0.99	0.99	0.99	ACC:0.99 F1-Score:0.99
Xue et al 2023 [36]	China	CNN(ResNet50) + MIL	1060(Developmental set:800,External test set:260)	0.95/ 0.87	0.92/0.77	0.92/0.83	ACC:0.95/0.87
Zhang et al 2022[37]	China	CNN(DenseNet201)	599(Developmental set:373,External test set:226)	0.99/ 0.90	0.95/0.81	0.91/0.84	ACC:0.93/0.82 F1-Score:0.95/ 0.85
Zhu et al 2024 [38]	China	VAE + MLP	2317(Developmental set:1853,Test set:464)	0.97	NA	NA	ACC:0.89 Precision:0.91 Recall:0.89 F1-Score:0.89

- Multiple observational studies have used machine learning, deep learning (most commonly CNN) to diagnose COPD from CT scan of thorax

- Most of the studies have found that there is a good accuracy with which the AI diagnose COPD from the CT imaging

Pooled sensitivity- 86 % (95 %CI, 78–91 %).
Specificity - 87 % (95 %CI, 83–91 %).
PLR - 6.75 (95 %CI, 4.67–9.74).
NLR - 0.16 (95 %CI, 0.10–0.26).
DOR - 41.05 (95 %CI, 19.10–88.25).
AUC- 93 % (95 %CI, 90–95 %).

Subgroup analysis.

Group	Number	Sensitivity	Specificity	Diagnostic Odds Ratio	AUC
AItype Deep Learning	7	82 % (95 %CI, 76–87 %)	87 % (95 %CI, 79–91 %)	29 (95 % CI, 17–50)	91 % (95 % CI, 88–93 %)
Meachine Learning	5	93 % (95 %CI, 85–97 %)	84 % (95 %CI, 79–88 %)	75 (95 % CI, 27–209)	89 % (95 % CI, 86–91 %)
Deep learning with MIL	3	87 % (95 %CI, 61–96 %)	89 % (95 %CI, 78–95 %)	54 (95 % CI, 6–504)	94 % (95 % CI, 92–96 %)
Area China	12	87 % (95 %CI, 76–93 %)	89 % (95 %CI, 82–93 %)	51 (95 % CI,18–151)	94 % (95 % CI, 91–96 %)
Not China	3	82 % (95 %CI, 73–89 %)	84 % (95 %CI, 76–90 %)	24 (95 % CI, 9–68)	90 % (95 % CI, 87–92 %)
Data Set Testing set	15	90 % (95 %CI, 84–94 %)	92 % (95 %CI, 88–95 %)	99 (95 % CI, 43–231)	96 % (95 % CI, 94–98 %)
External Validation Set	6	80 % (95 %CI, 73–86 %)	84 % (95 %CI, 81–86 %)	20 (95 % CI, 14–30)	85 % (95 % CI, 82–88 %)
Corresponding Testing set	6	91 % (95 %CI, 86–95 %)	94 % (95 %CI, 92–95 %)	152 (95 % CI, 85–271)	96 % (95 % CI, 94–97 %)

Wu Q, Guo H, Li R, Han J. Deep learning and machine learning in CT-based COPD diagnosis: Systematic review and meta-analysis. International Journal of Medical Informatics [Internet]. 2025 Jan 30;196:105812.

- Studies have used emphysema as a measure of COPD severity

- Radiomics features that have been associated with emphysema have been identified and utilized by machine learning to diagnose and stage COPD

- Character of airways as seen on CT has been also included for training machine and its output utilized in diagnosing and classifying COPD

- However, they don't specifically mention diagnosing early COPD (before overt imaging features appear) and impact on treatment and outcome

Table 3 The preprocessing, feature extraction, and feature selection method of deep learning method in COPD identification and stage

Team	Reference	Preprocessing	Feature extraction	Feature selection
Mets et al.	[64]	The segmentation of lung and airway	Three quantitative CT biomarkers (emphysema, air trapping, and bronchial wall thickness)	-
González et al.	[66]	Join four views into a single montage	CNN features	-
Cheplygina et al.	[67]	3D Region of interest (ROI) from CT image	Gaussian scale space features	-
Sathiya et al.	[68]	Gray Scale	Gray Level Co-occurrence Matrix	-
Xu et al.	[70]	The segmentation of lung from CT image	CNN features (AlexNet)	Principle component analysis
Tang et al.	[72]	Lung mask generation, spatial normalisation	CNN features (ResNet-152)	-
Hasenstab et al.	[74]	Co-registration, lung segmentation	Emphysema and air trapping feature	-
Li et al.	[76]	Volume of Interest segmentation from CT	1395 radiomics features	Variance threshold, Select K Best method, and least absolute shrinkage and selection operator (LASSO)
Yang et al.	[80]	Lung region segmentation	1316 radiomics features	LASSO
Yang et al.	[81]	Lung parenchyma segmentation	1316 radiomics features	Generalized linear model and LASSO
Puchakayala et al.	[86]	Segmentation of lung and airways	Demographics features, emphysema feature, lung and airway radiomics features	-

- In 2023, a new approach for imaging in COPD was proposed

- 3D reconstruction of airway and parenchyma obtained from CT images were evaluated by deep CNN to comment on whether the patient had COPD or not

- The assumption was that subtle narrowing of airways, irregularities on surfaces of the lungs would differentiate patients afflicted with COPD from the healthy controls

- The method showed a good accuracy and was superior to the other artificial networks used for this purpose

Deep CNN for COPD identification by Multi-View snapshot integration of 3D airway tree and lung field

Yanan Wu [a,b], Ran Du [a], Jie Feng [c], Shouliang Qi [a,b,*], Haowen Pang [a,b], Shuyue Xia [d], Wei Qian [e]

[a] College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China
[b] Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Northeastern University, Shenyang, China
[c] School of Chemical Equipment, Shenyang University of Technology, Liaoyang, China
[d] Respiratory Department, Central Hospital Affiliated to Shenyang Medical College, Shenyang, China
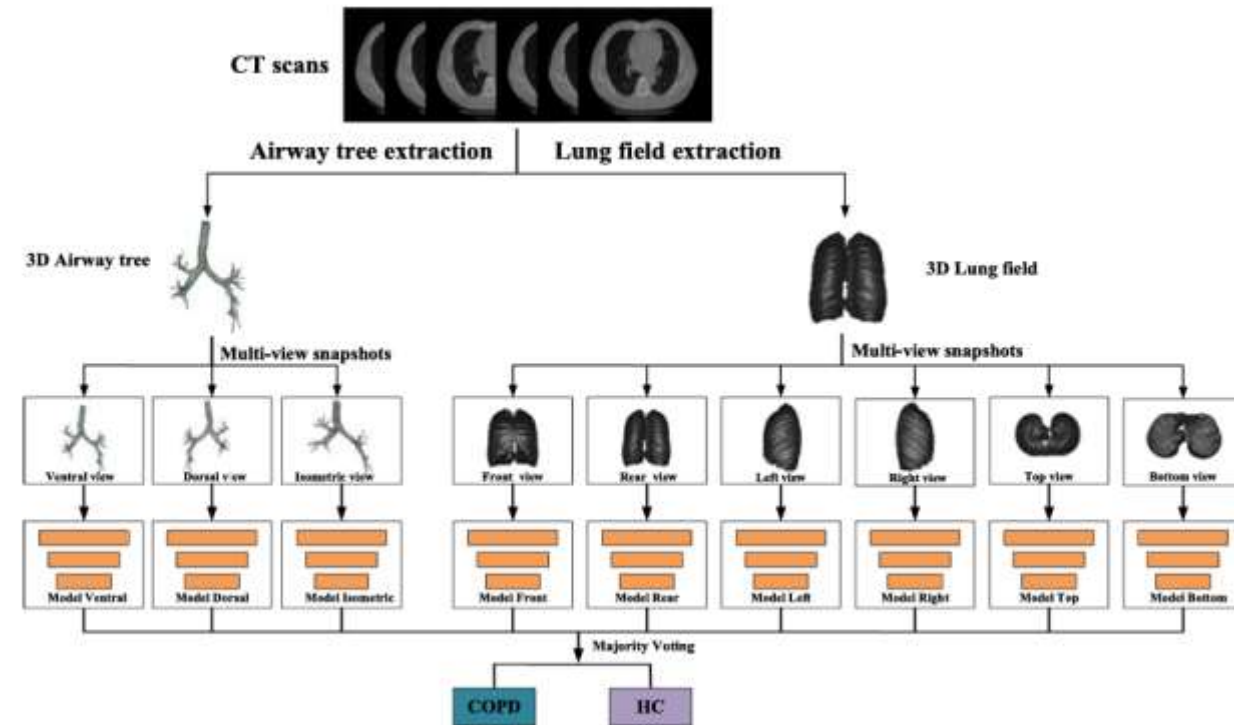
Fig. 1. Workflow of the multi-view-based majority voting convolutional neural network.

Wu Y, Du R, Feng J, Qi S, Pang H, Xia S, et al. Deep CNN for COPD identification by Multi-View snapshot integration of 3D airway tree and lung field. Biomedical Signal Processing and Control. 2023 Jan 1;79:104162–2.

View: Ventral **View: Dorsal** **View: Isometric**

View: Front **View: Rear** **View: Left**

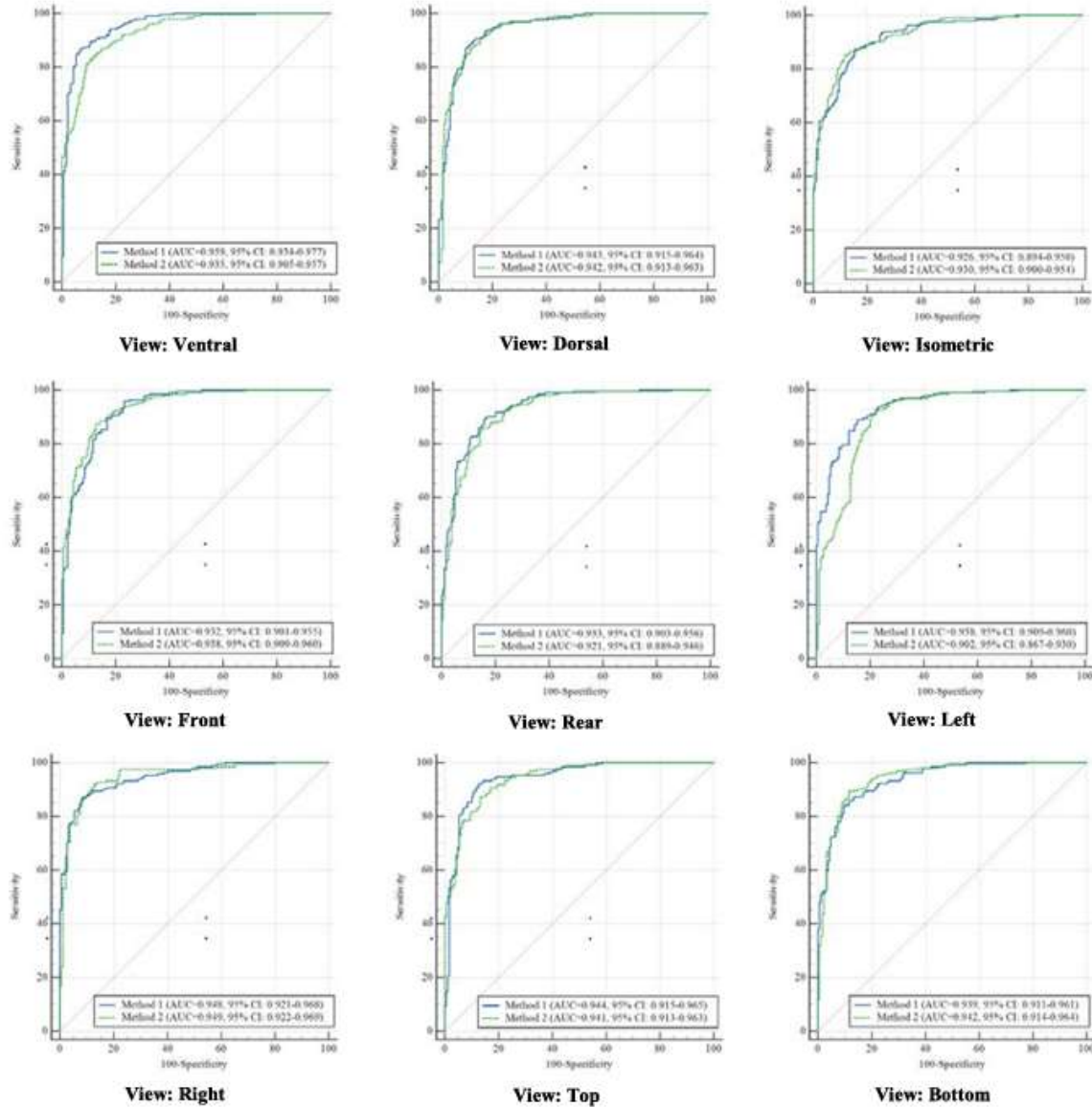**View: Right** **View: Top** **View: Bottom**

Fig. 5. ROC of nine single-view using ResNet-26 model by different airway segmentation and lung field segmentation methods on Dataset 1. *In the view of the airway tree, Method 1 means the Airway 1 method and Method 2 means the Airway 2 method. For the view of the lung field, Method 1 means the Lung field 1 method and Method 2 means the Lung field 2 method.

Table 5

Comparison between our method and the state-of-the-art CNN method on Dataset 1.

Methods	ACC	SEN	SPE	PPV	NPV
VGG16	0.898	0.866	0.938	0.942	0.855
Inception_v3	0.926	0.901	0.955	0.958	0.895
ResNet50	0.889	0.849	0.940	0.947	0.832
DenseNet121	0.929	0.901	0.961	0.963	0.895
Our method	**0.947**	**0.929**	**0.967**	**0.968**	**0.926**

Comparison of our method and state-of-the-art methods.

Methods	ACC	SEN	SPE	PPV	NPV
BOA-CNN [40]	0.881	0.864	0.901	0.905	0.858
Resnet50 [42]	0.889	0.849	0.940	0.947	0.832
González [31]	0.824	0.895	0.753	0.783	0.877
Our method	**0.947**	**0.929**	**0.967**	**0.968**	**0.926**

The views evaluated did not affect the outcome significantly;
In terms of accuracy, this method was superior from all the AI techniques previously used

Asthma and pleural effusion: scarce data

- Although artificial intelligence has been evaluated in diagnosing asthma and monitoring its response to treatment, there has been scarce or no data regarding using AI to radiology for this purpose

- Similarly, use of AI in imaging has been very limited in case of pleural effusion.

- Studies that considered pleural abnormalities have only done so in the context of malignant infiltration of the pleura as part of evaluation for extent/staging of bronchogenic carcinoma

Concerns regarding AI:

- Can not be relied upon without final verification by human

- Known to misclassify images that has been slightly altered (even by few pixels)

- Known to be biased in testing if the training dataset has bias

- Problem of generalizability: may not work on imaging from older machines already installed in some centers

- 'BLACK-BOX' – we do not often know how the neural network is interpreting data and making predictions→ difficult to make changes without retraining

Future possibilities:

- Thermal imaging scan to assess asthma response to medication

- Detection of airway abnormalities from 3D reconstructed images of CT thorax (emphysema, bronchiectasis, congenital airway malformations)

- Better prediction models for suspecting PTE in patients clinico-radiologically

- Delineation of haemodynamics in pulmonary hypertension non-invasively

AI in medical imaging in India: where are we now?

- Role of artificial intelligence in medical field is blooming in India

- The market for AI-powered machines (largely robotics) is rapidly growing

- However, AI-imaging for lung cancer, ILD and TB detection is not widely applied

- The cost and time required for integrating AI systems with existing PACS and RIS are considerably higher than training radio-technologists in identifying gross abnormalities and suspicious lesions on radiology

- Large studies in AI-imaging are also lacking to build confidence and interest regarding its application in diagnostics and even imaging

Conclusion:

- Artificial intelligence in thoracic imaging have made great strides in areas such as early detection of malignant nodules, predicting malignant potential of nodules and survival of patients

- It has shown great potential in classifying and predicting survival in interstitial lung diseases

- In diagnosing pulmonary hypertension and pulmonary embolism it has shown promise, but more studies are required

- In diagnosing and classifying COPD according to severity, it has established its efficacy but whether the increased cost is justified requires farther evaluation

- In asthma and pleural diseases, more studies are required to find its place

- Although AI has been found to be reasonably accurate in diagnosing certain diseases, it requires a large dataset for accuracy, to avoid overfitting and to minimize false negatives

- The benefit of AI imaging lies in the rapidity with which it can diagnose a large number of patients

- Almost all the studies show that the outputs of AI systems had to be validated by human radiologists and while doing it a considerable number of false positive and false negatives have been identified

- Artificial intelligence should be considered as an aid to the diagnosticians rather than an alternative or a replacement

- Artificial intelligence is a rapidly expanding modality in field of medicine

- There has been a considerable gap between this technology and the clinicians who are supposed to use it

- Composite approaches by clinician and AI may not only increase the diagnostic accuracy of both as well as dramatically reduce time to diagnosis

- Clinicians should be more accepting of it and more research in this field should be aimed at validating different AI methods as well as cost-benefit analysis

Thank you