Robust AI-driven Identification and Quantification of Medical Imaging biomarkers: Pulmonary Fibrosis Segmentation Challenge 2025 (PuFiSeg)

Sheng Zhang, Shiyi Wang, Tao Wang, Zhilin Yue, Helin Zhou, Yang Nan, Yingying Fang, Jiaheng Wang, Junzhi Ning, Zhenxuan Zhang, Xia Chen, Felder Federico N, Simon Walsh, Guang Yang

Department of Bioengineering and Imperial-X, Imperial College London
National Heart and Lung Institute, Imperial College London
Royal Brompton Hospital, London, UK
Tianjin Fourth Central Hospital, Tianjin city, China
pufiseg2025@gmail.com; g.yang@imperial.ac.uk

Abstract

Pulmonary fibrosis is a group of severe lung conditions that manifest as lung stiffening and scarring, and leads to the mortality on par with breast cancer, which presents a severe public health concern in the world. The manual delineation of pulmonary fibrosis regions is unduly burdensome, since there are substantially huge challenges to identify the pulmonary fibrosis regions accurately, e.g., high-level domain expertise, inter/intra-observer variability, blurring boundary, hundreds of slices per scan and slice by slice consistency constraint in 3D HRCT space. Currently, no effort has been yet made to conduct the automatic modelling of pulmonary fibrosis areas. To explore this brand new domain, 463 3D HRCT cases were collected and included in this challenge. The pulmonary fibrosis regions were meticulously annotated slice by slice by three experienced radiologists. Competitors are encouraged to devise automatic pulmonary fibrosis segmentation models with high robustness and generalization abilities. This challenge is an open-call challenge and new submissions are allowed after the conference.

1 Organizers

Sheng Zhang, Shiyi Wang, Tao Wang, Zhilin Yue, Helin Zhou, Yang Nan, Yingying Fang, Jiaheng Wang, Junzhi Ning, Zhenxuan Zhang, Guang Yang are with the Department of Bioengineering and Imperial-X, Imperial College London, London, UK.

Felder Federico N, Simon Walsh, Guang Yang are also with Royal Brompton Hospital, London, UK. Xia Chen is with Tianjin Fourth Central Hospital, Tianjin city, China.

Main contacts and e-mails: Sheng Zhang, pufiseg2025@gmail.com; Tao Wang, ortonwangtao@gmail.com; Guang Yang, g.yang@imperial.ac.uk.

2 Tasks

To prosper the development and application of AI techniques into the domain of medical image analysis, we design two tracks in this challenge, which is the first challenge focusing on pulmonary fibrosis segmentation.

Track 1: Competitors are encouraged to develop automatic full-supervision segmentation models of pulmonary fibrosis with high robustness and generalization abilities.

Track 2: It is well-known that medical images acquired by different devices, parameter settings, imaging protocols, demographics and clinical sites lead to the huge challenge of significant data distribution discrepancy, where a model trained on the source domain performs badly while being directly applied to the target domain. Fortunately, Unsupervised Domain Adaptation (UDA) can transfer knowledge from a labeled source domain to an unlabeled target domain robustly. Thus, this track will aim at the practical UDA issue in the medical domain.

3 Expected number of participants

We expect at least 400 participants to attend this challenge, which is estimated based on previous challenges, e.g., ATM22 (315 participants) and PARSE22 (469 participants).

4 Publication and future plans

Challenge results will be summarised into a high-impact journal paper (e.g., in the Lancet Digital Health), describing the latest techniques in pulmonary fibrosis segmentation and novel quantitative imaging biomarkers. The challenge paper will invite participants from the top-five teams as co-authors. After the acceptance of the challenge paper, the challenge data and results will be free to use for the public.

5 Duration and resource requirements

The challenge will be held on CodaLab and take half day to conclude and share experience in the satellite event of IJCAI2025 conference. The site for this challenge will be set up once the proposal is accepted.

6 Participation rules

Registration Guidance:

(1) Participants need to download and sign the registration form carefully. Note: registering by official institutional email. Sending an email to pufiseg2025@gmail.com with the real names, institute and one signed registration form for

Tasks 1 and 2.

(2) Registering an account on the CodaLab platform using the same official institutional email, selecting "register" button from "Participate" tab in the challenge website. Failure to do so will lead to the rejected application.

(3) Only after steps (1) and (2), the dataset download link will be given to approved participants through email.

(4) Participants are not allowed to use multiple accounts to participate in the challenge, and participants in a team are also not allowed to make individual submissions.

Submission:

(1) In the validation phase, the algorithm output (segmented masks) will be submitted to CodaLab for evaluation. Each team will be allowed to submit at most three times per day.

(2) In the test phase, each team can only submit one successful docker image and a qualified short paper which describes the method (4-12 pages, LNCS format) to the organizers' email (pufiseg2025@gmail.com). Link to submission instructions.

(3) Only fully-automated segmentation methods based on the released training set only are acceptable. Any types of manual procedures (e.g., manual annotation, manual correction) are not allowed.

(4) The top-3 teams will be requested to release their codes/models for reproducibility. The 3+ ranked teams are encouraged to release their codes.

(5) All participants should agree that their short papers can be used by the organizers for a challenge paper.

(6) If registered participants do not make complete submissions (Docker container and methodology paper), they (and their groups) will be listed in the dishonest list and will be forbidden to join future PuFiSeg challenges.

(7) Redistribution, transfer, and sharing of data are prohibited. Participants can only utilise the data independently. After the submission of our work, the challenge data will be accessible again.

(8) All data in this challenge can be only used for academic purposes. If you used it in your research, please cite the following papers [Zhang *et al.*, 2024; Nan *et al.*, 2024; Fang *et al.*, 2024].

(9) For a fair comparison, participants are not allowed to use any additional data and pre-trained models.

7 Challenge schedule

2025/04/15: Challenge website opening.
2025/04/20: Challenge registration opening.
2025/06/05: Registration end and training cases released.
2025/06/20: Validation submission opening.
2025/07/30: Test data released and test phase opening.
2025/08/16: All submissions closed.

8 Datasets

Track 1: It contains 262 3D HRCT volumes annotated by 3 experts with more than 5-year clinical experience. The data is mainly selected from the OSIC data, the test data won't be released, and only for off-line evaluation. OSIC dataset is an open-access global dataset on pulmonary diseases.

Track 2: It contains 201 3D HRCT volumes annotated by 3 experts with more than 5-year clinical experience. All the volumes are from the Tianjin Fourth Central Hospital, Tianjin city, China, which holds a significant different data distribution from **Track 1**. High-quality, highly accurate digital imaging biomarkers in medical images can thus be an excellent way to improve clinicians' diagnosis, prognosis and prediction of response to treatment. All data complies with all applicable privacy laws, regulations, consents and related restrictions.

Table 1: Data distribution details of PuFiSeg2025.

Track 1	Train	Validation	Test
No. scans	120	52	90
Sizes	$s1 \times w1 \times w1$	$s1 \times w1 \times w1$	$s1 \times w1 \times w1$
Track 2	Train (unlabelled)	Validation	Test
No. scans	91	40	70
Sizes	$s2 \times w2 \times w2$	$s2 \times w2 \times w2$	$s2 \times w2 \times w2$

$s1 \in [397, 764], w1 \in \{512, 768\}$. $s2 \in [366, 894], w2 = 512$.

9 Evaluation Metrics

The evaluation will be conducted on the CodaLab platform, with the assessment docker provided by the organizers' group. The evaluation codes can be viewed here. The segmentation performance is evaluated in four aspects as follow:

(1) Intersection over Union (IoU): Measuring the overlap ratio between the semantic prediction and ground truth.

(2) Precision: Measuring how many of the predicted positive samples are actually correct.

(3) Running time: Low time-consuming is more efficient in the practical clinical environment.

(4) Maximum used GPU memory: Low memory usage indicates the good application potential in more widely used computing platform.

9.1 Ranking Method

The ranking scheme includes the following steps:

(1) Calculating the IoU, Precision, Maximum used GPU memory, and Running time for all cases.

(2) Ranking the IoU, Precision, Maximum used GPU memory, and Running time separately.

(3) Averaging these rankings with weights. (Tie if the rankings are equal.)

The ranking criteria of validation phase: Score = (IoU + Precision) * 0.5. The ranking criteria of test phase:

$$\begin{aligned} Score = Rank[(IoU + Precision) * 0.5] * 0.7 \\ + Rank[Time] * 0.15 + Rank[Memory] * 0.15 \end{aligned} \quad (1)$$

10 Awards

Successful participation certificates (digital) will be awarded to all teams that have valid test scores in the challenge leaderboard and complete short paper submissions reviewed by the organizing committee. Top-1 team will receive a free flight ticket to IJCAI2026 conference (Up to $1200 USD for flight). The second team will receive $500 US dollars with digital certificates (for all members of the team). The third team will receive $300 US dollars with digital certificates (for all members of the team).

References

[Fang *et al.*, 2024] Yingying Fang, Shuang Wu, Sheng Zhang, Chaoyan Huang, Tieyong Zeng, Xiaodan Xing, Simon Walsh, and Guang Yang. Dynamic multimodal information bottleneck for multimodality classification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7696–7706, 2024.

[Nan *et al.*, 2024] Yang Nan, Xiaodan Xing, Shiyi Wang, Zeyu Tang, Federico N Felder, Sheng Zhang, Roberta Eufrasia Ledda, Xiaoliu Ding, Ruiqi Yu, Weiping Liu, et al. Hunting imaging biomarkers in pulmonary fibrosis: Benchmarks of the aiib23 challenge. *Medical Image Analysis*, 97:103253, 2024.

[Zhang *et al.*, 2024] Sheng Zhang, Yingying Fang, Yang Nan, Shiyi Wang, Weiping Ding, Yew-Soon Ong, Alejandro F Frangi, Witold Pedrycz, Simon Walsh, and Guang Yang. Fuzzy attention-based border rendering orthogonal network for lung organ segmentation. *IEEE Transactions on Fuzzy Systems*, 2024.