

# Foundation Model for Cardiac MRI Reconstruction: Meeting the Real-world Challenge of Multi-center, Multi-vendor, and Multiple Diseases: Structured description of the challenge design

## CHALLENGE ORGANIZATION

### Title

Use the title to convey the essential information on the challenge mission.

Foundation Model for Cardiac MRI Reconstruction: Meeting the Real-world Challenge of Multi-center, Multi-vendor, and Multiple Diseases

### Challenge acronym

Preferable, provide a short acronym of the challenge (if any).

CMRxRecon2025

### Challenge abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

Cardiac MRI (CMR) has become an essential tool for diagnosing and evaluating cardiovascular diseases, offering multi-parametric, high-resolution anatomical and functional data. CMR reconstruction from highly under-sampled k-space data has gained significant attention in recent research. Numerous AI-based image reconstruction algorithms have shown potential in improving imaging performance and patient experience in recovering high-quality images from aggressively undersampled k-space measurements. However, the complexity and diversity of CMR scans in real-world applications, involving various image contrasts, sampling trajectories, equipment vendors, anatomical structures, and disease types, present a great challenge for existing AI-based reconstruction methods, which are usually developed for only one or a few specific scanning settings. In practice, there are often inevitable distribution mismatches between the training data and target data, due to the diversities listed above. Therefore, building and validating universal and robust reconstruction models for handling these diversities remain to be critical technical challenges for multi-parametric CMR imaging.

In our past MICCAI "CMRxRecon" challenge series (CMRxRecon2023 and CMRxRecon2024), we have released fully sampled k-space data for multi-parametric cardiac imaging from a total of 600 individuals in a single medical center, and provided relevant technical infrastructure as well as a baseline model for CMR reconstruction. However, this dataset did not cover multi-center, multi-vendor, and multiple diseases, which are the most valued targets in real-world clinical applications. Therefore, in this proposed challenge, we aim to make an important leap towards real-world clinical scenarios by extending the challenge scope in two directions:

1) To evaluate the robustness and generalization performance of reconstruction foundation models in more than 5 centers and 10 scanners, all those data are unseen during training stage.

2) To evaluate the clinical performance of reconstruction foundation models under no less than 5 cardiovascular diseases (e.g., hypertrophic cardiomyopathy, dilated cardiomyopathy, myocardial infarction, coronary artery disease, arrhythmias), all those data are unseen during training stage.

A total of 600 new volunteers with multi-parametric CMR imaging (200 cases for training, 100 validation cases and 300 test cases) will be included in the proposed CMRxRecon2025 challenge, with no less than 100 cases per center. We have preliminarily set up the Synapse platform for CMRxRecon2025 and uploaded 1% of the data for review (<https://www.synapse.org/#!Synapse:syn59814210/wiki/628454>). For Task 1, we will release data from two centers as the training set. We will retain data from three undisclosed centers as the validation set and test set, to better evaluate the model's generalization performance across different centers. For Task 2, we will only release data from healthy volunteers as the training set and retain all five disease types as the validation set and test set, to better evaluate the model's generalization performance on previously unseen diseases.

### Challenge keywords

List the primary keywords that characterize the challenge.

Cardiac MRI, image reconstruction, foundation model, multi-center, multiple diseases, generalization, trustworthiness

### Year

2025

### Novelty of the challenge

Briefly describe the novelty of the challenge.

N/A

### Task description and application scenarios

Briefly describe the application scenarios for the tasks in the challenge.

N/A

## FURTHER INFORMATION FOR CONFERENCE ORGANIZERS

### Workshop

If the challenge is part of a workshop, please indicate the workshop.

This challenge will be jointly hosted by the CMRxRecon Committee and Society for Cardiovascular Magnetic Resonance (SCMR) society. It could potentially be part of the Statistical Atlases and Computational Modeling of the Heart (STACOM) workshop. We have been collaborating with STACOM for three consecutive years.

### Duration

How long does the challenge take?

Half day.

In case you selected half or full day, please explain why you need a long slot for your challenge.

N/A

### Expected number of participants

Please explain the basis of your estimate (e.g. numbers from previous challenges) and/or provide a list of potential participants and indicate if they have already confirmed their willingness to contribute.

We expect full participation from 40-60 teams at MICCAI 2025. This is estimated from the previous CMRxRecon2023 challenge, in which more than 200 teams registered the challenge and 26 teams completed all stages of the challenge, including the final submission of the Docker container and the conference paper. According to the post-event survey conducted last year, all the 26 participating teams who filled out the questionnaire expressed their willingness to continue participating in the next challenge. We expect a considerable increase of participants this year, due to the fast-growing research communities of CMR, fast imaging, generative AI, and foundation models. This anticipation is also based on the growing popularity of the CMRxRecon2023 dataset (278 downloads, <https://www.synapse.org/#!/Synapse:syn51471091/wiki/622170>) and the CMRxRecon2024 dataset (more than 100 downloads, and the number is still growing, <https://www.synapse.org/#!/Synapse:syn54951257/wiki/627141>).

### Publication and future plans

Please indicate if you plan to coordinate a publication of the challenge results.

After the challenge, we will summarize the results in a challenge paper and submit it to a high-impact journal (e.g., Nature Medicine, Medical Image Analysis, IEEE Transactions on Medical Imaging, Journal of Cardiovascular Magnetic Resonance).

### Space and hardware requirements

Organizers of on-site challenges must provide a fair computing environment for all participants. For instance, algorithms should run on the same computing platform provided to all.

Participants are expected to train models in their local computational environments and to submit docker containers on our Synapse platform (<https://www.synapse.org/#!/Synapse:syn59814210/wiki/628454>). Training and validation data are available for registered teams to download. A leaderboard will be maintained on the Synapse platform during the validation phase. For the fairness of final evaluation, participants are required to submit their docker containers on Synapse platform and we will use the organizers' servers for testing. We will also require the teams to report their computational cost and optionally carbon footprint for model development. Regarding the computer configuration we plan to use for the testing phase, it is as follows:

- OS: Linux (Ubuntu 20.04.5 LTS, CENTOS7 or AlmaOS8)
- CPU: 2.5GHz, 20 cores;
- RAM: 64 GB; GPU: Tesla V100 (32 GB VRAM, single GPU);
- GPU Driver Version: 470;
- CUDA Version: 11.4;
- Time Limitation: 20 hours/team for each task.

For the on-site day, we will take a hybrid form to guarantee the participation of all the teams and audiences. We will reveal the challenge outcome and invite the winning teams to present their methods. In addition, to promote open dialogue and collaboration with clinical communities, device manufacturers, and device regulatory bodies, we are in active negotiation with SCMR society to host a joint workshop and invite representatives from the top three teams to present and exchange ideas.

# TASK 1: CMR reconstruction foundation model for multi-center evaluation

## SUMMARY

### Abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

CMR has emerged as a critical tool in the diagnosis and assessment of cardiovascular diseases, providing high-resolution anatomical and functional information. However, the process of reconstructing CMR data is complex and computationally intensive, especially in multi-center clinical settings. Differences in scanner configurations, imaging protocols, sampling trajectories, and patient populations across centers contribute to data heterogeneity, posing significant challenges for achieving consistent and reliable evaluations. Addressing these challenges necessitates the development of a foundation model for CMR reconstruction that can accommodate the variability inherent in multi-center data.

The objective of this challenge is to develop a foundation model that 1) provide high-quality image reconstruction for highly-accelerated (undersampled) MRI acquisitions; 2) being able to process multiple parametric measurements, views, and scanning protocols across multiple centers using a single foundation model.

The development of a foundation pre-trained reconstruction model is essential to tackle the diverse range of cardiac imaging applications across different centers and scanners. Traditional reconstruction methods often require specialized algorithms and settings for different imaging centers, making them less flexible and time-consuming to implement. By contrast, a foundation pre-trained model can offer a unified framework that can handle various imaging settings across different centers, allowing for faster and more robust reconstructions across different CMR protocols. This approach not only improves efficiency but also promotes consistency and standardization in multi-center CMR reconstruction.

The CMRxRecon2025 Challenge aims to foster advancements in multi-center CMR reconstruction by providing a platform for researchers to develop and evaluate reconstruction methods. The dataset will include multi-center k-space data, consisting of multi-parametric imaging across more than 5 centers all over the world, to cover diverse populations. The dataset also includes imaging of different anatomical views such as long-axis (including 2-chamber, 3-chamber, and 4-chamber) and short-axis. Additionally, we have preliminarily set up the Synapse platform for CMRxRecon2025 and uploaded 1% of the data for review (<https://www.synapse.org/#!Synapse:syn59814210/wiki/628454>).

The challenge encourages participants to develop innovative approaches, especially as the use of foundation pre-trained models, to tackle the complexities of multi-center CMR imaging. The envisioned technical impact of this challenge is the development of easy-to-deploy, reliable, generalizable, and data-efficient reconstruction methods that can enhance both the patient experience and the diagnostic quality of multi-center CMR images in heterogeneous clinical and research environments, enabling more accurate and comprehensive cardiac assessment worldwide.

### Keywords

List the primary keywords that characterize the task.

Multi-center evaluation, cardiac image reconstruction, foundation model

## ORGANIZATION

### Organizers

a) Provide information on the organizing team (names and affiliations).

Chengyan Wang

Human Phenome Institute, Fudan University, China

Michael Markl

SCMR president, Northwestern Univ Chicago, USA

Claudia Prieto

SCMR 2025 programme chair, Kings College London, UK and Pontificia Universidad Católica de Chile, Santiago, Chile

Chen Qin

Department of Electrical and Electronic Engineering & I-X, Imperial College London, UK

Shuo Wang

Digital Medical Research Center, School of Basic Medical Sciences, Fudan University, China

Daniel Kim

SCMR Science committee chair, Northwestern Univ Chicago, USA

Alistair Young,

Kings College London, UK

Jun Lyu

Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

He Wang, Hao Li, Zhensen Chen

Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, China

Guang Yang, Fanwen Wang

Department of Bioengineering/Imperial-X, Imperial College London, UK

Xiaobo Qu, Zi Wang

Department of Electronic Science, Xiamen University, Xiamen, China

Cheng Ouyang

Department of Engineering Science, University of Oxford, UK

Jiayu Zhu

Shanghai United Imaging Healthcare Advanced Technology Research Institute Co., Ltd., China

Yajing Zhang

Clinical Science Manager, MR Business Unit, Philips Healthcare Suzhou, China

Jing Qin

School of Nursing, The Hong Kong Polytechnic University, Hong Kong

Xiahai Zhuang

School of Data Science, Fudan University, China

Yan Li

Department of Radiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Sha Hua

Department of Cardiovascular Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, China

Lianming Wu

Department of Radiology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

Ziqiang Xu

Shanghai Fuying Medical Technology Co., Ltd., China

Ying-Hua Chu

Siemens Healthineers Ltd., China

Weibo Chen

Philips Healthcare, China

b) Provide information on the primary contact person.

Chen Qin, c.qin15@imperial.ac.uk

Department of Electrical and Electronic Engineering & I-X, Imperial College London, UK

### Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place. Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One-time event with fixed submission deadline.

### Challenge venue and platform

a) Report the event (e.g. conference) that is associated with the challenge (if any).

28th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2025), and the Annual Scientific Sessions of the Society for Cardiovascular Magnetic Resonance (SCMR).

b) Report the platform (e.g. grand-challenge.org) used to run the challenge.

synapse.org

c) Provide the URL for the challenge website (if any).

Website for CMRxRecon2023: <https://cmrxrecon.github.io>. Synapse platform for CMRxRecon2023:

<https://www.synapse.org/#!/Synapse:syn51471091/wiki/622170> Website for CMRxRecon2024:

<https://cmrxrecon.github.io/2024/Home.html> Synapse platform for CMRxRecon2024:

<https://www.synapse.org/#!/Synapse:syn54951257/wiki/627141> The new website for this year is under active construction. The new Synapse platform for CMRxRecon2025 (under construction):

<https://www.synapse.org/#!/Synapse:syn59814210/wiki/>

### Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Fully automatic.

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

It should be restricted to the data provided by the previous CMRxRecon challenge as well as data from the 'fastMRI' challenge (the most related public dataset), under the terms and conditions associated with the data usage.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

They may participate but are not eligible for awards and will not listed in the leaderboard.

d) Define the award policy. In particular, provide details with respect to challenge prizes.

The top 5 winners receive monetary awards. We are in active negotiation with the sponsor about the value (approximately \$2000 in total).

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

All submissions will be reported in the leaderboard. Participating teams can opt out of publication of their results in the leaderboard.

Prize-winning methods will be announced publicly as part of a scientific session at the MICCAI annual meeting. In

addition, we also plan to collaborate with SCMR to host a joint workshop and invite representatives from the top three teams to present and exchange ideas.

f) Define the publication policy. In particular, provide details on ...

- ... who of the participating teams/the participating teams' members qualifies as author
- ... whether the participating teams may publish their own results separately, and (if so)
- ... whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

Participating teams with a valid submission can nominate their team members as co-authors for the challenge paper. We reserve the right to exclude teams if they break the challenge rules. Participating teams can publish their own results but after a 3-month embargo period.

### Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

Docker container will be accepted as submission through the Synapse platform. Submission details will be published at the time point of challenge announcement.

b) Provide information on the possibility for participating teams to evaluate their algorithms before submitting final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

Participating teams are allowed to make 3 formal submissions per task. Only the last run submission is officially counted to rank challenge results. Before the final submission on the test set, participants can test their docker containers on the validation dataset to avoid submission errors.

### Challenge schedule

Provide a timetable for the challenge. Preferably, this should include

- the release date(s) of the training cases (if any)
- the registration date/period
- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

[Apr 1, 2025] website opens for registration, release training and validation images

[Apr 10, 2025] submission system opens for validation

[Aug 1, 2025] submission system opens for testing

[Sept 1, 2025] registration and docker submission deadline

[Oct 8, 2025] release final results during the MICCAI annual meeting

[Jan 31, 2026] challenge summary and invited talk from the winners during the SCMR annual meeting

### Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

We have received ethics approval from the local ethics committee of School of Basic Medical Sciences, Fudan University granted on 17/11/2021, No. 2021-Y060.

### Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)
- CC BY-ND (Attribution-NoDerivs)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
- CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

CC BY

### Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

We have released the source code for evaluating and ranking the results at Github:

<https://github.com/CmrXRecon/CMRxRecon>

b) In an analogous manner, provide information on the accessibility of the participating teams' code.

In the past CMRxRecon2023 challenge, all top ten teams have made their code and method details public, and they have allowed us to consolidate all their information:

<https://github.com/CmrXRecon/CMRxRecon/tree/main/ModelPool>.

### Conflicts of interest

Provide information related to conflicts of interest. In particular provide information related to sponsoring/funding of the challenge. Also, state explicitly who had/will have access to the test case labels and when.

We declare no conflicts of interest. Test images will only be accessible to the challenge organizers.

## MISSION OF THE CHALLENGE

## Field(s) of application

State the main field(s) of application that the participating algorithms target.

Examples:

- Diagnosis
- Education
- Intervention assistance
- Intervention follow-up
- Intervention planning
- Prognosis
- Research
- Screening
- Training
- Cross-phase

Research, Diagnosis.

## Task category(ies)

State the task category(ies)

Examples:

- Classification
- Detection
- Localization
- Modeling
- Prediction
- Reconstruction
- Registration
- Retrieval
- Segmentation
- Tracking

Reconstruction.

## Cohorts

We distinguish between the target cohort and the challenge cohort. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on ex vivo data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final

biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding sex or age (target cohort).

a) Describe the target cohort, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

The target cohort is patients requiring multi-parametric CMR exams, especially patients with arrhythmia and those who cannot adhere to standard imaging protocols. The results of this study will directly accelerate clinical scanning speed, increase exam throughput, and reduce medical costs.

b) Describe the challenge cohort, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

The proposed challenge will release multi-channel k-space data from at least 5 medical centers, more than 10 scanners, including mainstream device vendors (GE, Philips, Siemens, and United imaging), 1.5T, 3.0T and 5.0T field strengths, more than 5 types of sequences (e.g., dark blood, tagging, phase-contrast), including anatomical views including long-axis, short-axis, outflow tract, and aortic (cross-sectional and sagittal views), and at least 5 diseases such as hypertrophic cardiomyopathy, dilated cardiomyopathy, myocardial infarction, coronary artery disease, and arrhythmias. A total of 600 volunteers with multi-parametric CMR imaging (200 cases for training, 100 validation cases and 300 test cases) will be included, with no less than 100 cases per center. The dataset will include multi-parametric k-space data, consist of cardiac cine, mapping, tagging, phase-contrast, and dark-blood imaging. The dataset also includes imaging of different anatomical views like long-axis (2-chamber, 3-chamber, and 4-chamber), short-axis, outflow tract, and aortic (cross-sectional and sagittal views). During the training phase, we will release data from two centers as the training set. We will retain data from three undisclosed centers as the validation set and test set, to better evaluate the model's generalization performance across different centers. Additionally, we have preliminarily set up the Synapse platform for CMRxRecon2025 and uploaded 1% of the data for review (<https://www.synapse.org/#!/Synapse:syn59814210/wiki/628454>).

### Imaging modality(ies)

Specify the imaging technique(s) applied in the challenge.

Magnetic Resonance Imaging

### Context information

Provide additional information given along with the images. The information may correspond ...

a) ... directly to the image data (e.g. tumor volume).

none

b) ... to the patient in general (e.g. sex, medical history).

none

### Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

Multi-parametric CMR data is acquired from the heart and aorta. During the training phase, we will release data from two centers as the training set. We will retain data from three undisclosed centers as the validation set and test set, respectively, to better evaluate the model's generalization performance across different centers.

b) Describe the algorithm target, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

The algorithm target is to assess the reconstruction performance on multi-parametric CMR imaging based on a pre-trained reconstruction foundation model. In the challenge cohort, we recruit volunteers to be scanned with different k-space trajectories and accelerating factors.

### Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.
- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

Peak signal-to-noise ratio (PSNR)[1], structural similarity index measure (SSIM)[1] and normalized mean squared error (NMSE)[2] between reconstructed images and ground truth images (fully sampled data).

1. When evaluating SSIM, we will narrow down the assessment field-of-view to the region where the heart is located, to avoid interference from the background area.
2. During the testing and ranking phase, we will invite three radiologists to independently score the top five teams ranked by SSIM. The radiologists will assess several image characteristics using a five-point Likert scale, which includes signal loss, sharpness of heart contour, image artifacts (indicating the extent to which artifacts affect image quality), and clinical utility (ranging from 1 (non-diagnostic) to 5 (excellent)). Each radiologist's score for signal loss, sharpness of heart contour, image artifacts, and clinical utility carries a weight of 25%. The final score will be the weighted sum of these four ratings. We will average the scores from the three radiologists to determine the teams' final scores, which will then be ranked from highest to lowest.

Reference:

[1] Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM[C]//2010 20th international conference on pattern recognition. IEEE, 2010: 2366-2369.

[2] Hameed A, Abotiheen M H A, Abdulzahra R. Quality measurement of blurred images using NMSE and SSIM metrics in HSV and RGB color spaces[J]. Physics Journal, 2015, 1: 105-111.

## DATA SETS

### Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

The proposed challenge will release multi-channel k-space data from more than 10 scanners, including mainstream device vendors (GE, Philips, Siemens, and United imaging), 1.5T, 3.0T and 5.0T field strengths.

b) Describe relevant details on the imaging process/data acquisition for each acquisition device (e.g. image acquisition protocol(s)).

We follow the recommendations of CMR exams reported in the previous publication (doi: 10.1007/s43657-02100018x, 10.1007/s43657-021-00018-x). We use 'TrueFISP' for cine, PC and tagging, and 'FLASH' for mapping and dark-blood imaging. The collected imaging planes include long-axis (2-chamber, 3-chamber, and 4-chamber), short-axis, outflow tract, and aortic (cross-sectional and sagittal views). Typically 5-15 slices are acquired. For mapping, signal data were collected at the end of the diastole with ECG triggering. The cardiac cycle is segmented into 15–25 phases with a temporal resolution of around 50 ms. Typical geometrical parameters include: spatial resolution  $2.0 \times 2.0 \text{ mm}^2$ , slice thickness 8.0 mm, and slice gap 4.0 mm.

c) Specify the center(s)/institute(s) in which the data was acquired and/or the data providing platform/source (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

The following centers have already contacted us and can support our data collection. Additionally, we are actively reaching out to more international medical centers in the hope of collecting raw data from more sources.

Zhangjiang Imaging Center, Fudan University, China

Department of Radiology, Imperial College London, UK

Department of Radiology, Zhongshan Hospital, Fudan University, China

Department of Electronic Science, Xiamen University, Xiamen, China

Department of Radiology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

Department of Cardiovascular Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, China

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

Data are acquired with a team of radiographers and clinical advisors as appropriate.

### Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

Training cases include fully sampled k-space data and auto-calibration lines (ACS, 24 lines) will be provided in '.mat' format.

Validation cases include under-sampled k-space data, sampling trajectories, and auto-calibration lines (ACS, 24 lines) with various acceleration factors in '.mat' format.

Test cases include fully sampled k-space data, under-sampled k-space data, sampling trajectories and auto-calibration lines (ACS, 24 lines). Test cases will not be released before the challenge ends.

b) State the total number of training, validation and test cases.

1,200 cases in total (600 cases from previous CMRxRecon challenge datasets from single center and 200 new cases from multi-centers for training, 100 validation cases (multi-centers) and 300 test cases (multi-centers)). During the training phase, we will release data from two centers as the training set. We will retain data from three undisclosed centers as the validation set and test set, to better evaluate the model's generalization performance across different centers.

c) Explain why a total number of cases and the specific proportion of training, validation and test cases was chosen.

This will be the largest publicly available dataset in the field of CMR reconstruction. The sample size of training dataset is reasonable for training deep learning reconstruction networks for CMR imaging according to previous literature and our experience. The sample size of validation and test dataset would well represent the distribution of the data cohort and achieve robust evaluation of the models according to last year's CMRxRecon challenge results, in which 60 cases were used for validation and 120 cases for testing

(<https://www.synapse.org/#!Synapse:syn51471091/wiki/622548>). The results of the test set and validation set have shown very high consistency in both scores and team rankings (the top 3 are exactly the same).

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

This dataset will include a diverse range of ethnicities, disease types, and age distributions. We will aim for an age and gender balance between the training and test data.

e) Challenge organizers are encouraged to (partly) use unseen, unpublished data for their challenges. Describe if new data will be used for the challenge and state the number of cases along with the proportion of new data.

N/A

### Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

N.A.

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

N.A.

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

N.A.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

N.A.

### Data pre-processing method(s)

Describe the method(s) used for pre-processing the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

The raw k-space data exported from the scanner will be processed and transformed to '.mat' format using the script provided by our vendor. A readme file will be provided to describe the content and usage of the data.

### Sources of error

a) Describe the most relevant possible error sources related to the image annotation. If possible, estimate the magnitude (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

N.A.

b) In an analogous manner, describe and quantify other relevant sources of error.

N.A.

## ASSESSMENT METHODS

### Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

For cardiac cine, PSNR, SSIM and NMSE between reconstructed images and ground truth images (fully sampled data) will be used for performance evaluation.

1. When evaluating SSIM, we will narrow down the assessment field-of-view to the region where the heart is located, to avoid interference from the background area.

2. During the testing and ranking phase, we will invite three radiologists to independently score the top five teams ranked by SSIM. The scoring will cover three aspects: image quality, image artifacts, and clinical utility. We will consider both the radiologists' scores and the SSIM results to generate a comprehensive ranking.

For cardiac mapping, Root Mean Squared Error (RMSE) of T1 and T2 measurements in the LV myocardium region between reconstructed images and ground truth images (fully sampled data) will be used for performance evaluation.

T1 and T2 calculations are listed below:

A 3-parameter fitting model was used to calculate T1 of each signal time course:

$$S(TI) = A - B e^{(-TI/T_1^*)}$$

where TI represents the inversion time, and A, B, and T1\* were three parameters to be estimated. T1 was determined from the resulting A, B, and T1\* by applying the equation:

$$T_1 = T_1^* (B/A - 1)$$

The T2 value is calculated with the equation:

$$S(t) = A \exp(-t/T_2)$$

where  $S(t)$  represents the signal intensity,  $A$  is the scale factor and  $t$  denote T2 preparation time.

b) Justify why the metric(s) was/were chosen, preferably with reference to the biomedical application.

We keep these metrics aligned with the 'fastMRI' challenge evaluation metrics for fair evaluations.

### Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

We will perform automated segmentation of the images (using a pre-trained cardiac segmentation model) and segment the heart according to AHA guidelines for measurement and statistics. Finally, we will combine all the measurement results and provide an average score.

For cardiac cine, PSNR, SSIM and NMSE between reconstructed images and ground truth images (fully sampled data) will be used for performance evaluation.

1. When evaluating SSIM, we will narrow down the assessment field-of-view to the region where the heart is located, to avoid interference from the background area.

2. During the testing and ranking phase, we will invite three radiologists to independently score the top five teams ranked by SSIM. The scoring will cover three aspects: image quality, image artifacts, and clinical utility. We will consider both the radiologists' scores and the SSIM results to generate a comprehensive ranking.

For cardiac mapping, Root Mean Squared Error (RMSE) of T1 and T2 measurements in the LV myocardium region between reconstructed images and ground truth images (fully sampled data) will be used for performance evaluation.

b) Describe the method(s) used to manage submissions with missing results on test cases.

Participating teams are required to submit docker containers and process all the cases in the test set on our server. For the cases without valid output, we will assign it to the lowest value of metric.

c) Justify why the described ranking scheme(s) was/were used.

PSNR, SSIM and RMSE are commonly used metrics and proven valid in almost all existing related works and previous challenges. This strategy is consistent with the 'fastMRI' challenge and our previous 'CMRxRecon' challenge last year.

### Statistical analyses

a) Provide details for the statistical methods used in the scope of the challenge analysis. This may include

- description of the missing data handling,
- details about the assessment of variability of rankings,
- description of any method used to assess whether the data met the assumptions, required for the particular statistical approach, or
- indication of any software product that was used for all data analysis methods.

Please see the description of metrics and ranking methods above. We will incorporate the Wilcoxon signed rank test to conduct statistical analysis on all challenge results.

b) Justify why the described statistical method(s) was/were used.

Please see the description of metrics and ranking methods above.

### **Further analyses**

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

N/A

## **TASK 2: CMR reconstruction foundation model for multiple diseases evaluation**

### **SUMMARY**

#### **Abstract**

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

CMR is a critical tool for diagnosing a variety of cardiovascular diseases, including coronary artery disease, cardiomyopathy, congenital heart defects, and heart valve disorders. However, the complexity and variability of these diseases present significant challenges for accurate and efficient image reconstruction, especially for small models whose training data does not sufficiently cover the wide range of cardiac conditions. The primary difficulty arises from the substantial anatomical and physiological variations among patients, which can lead to domain shifts that complicate the reconstruction process. Conventional AI-based reconstruction methods, which rely on domain-specific algorithms, struggle to adapt to these variations.

The goal of this research is to develop a robust foundation model capable of reliably reconstructing CMR images and evaluate the performance on different diseases, employing both uniform and variable-density sampling at acceleration factors ranging from 3x to 20x. The model is designed to generalize across multiple diseases (e.g., hypertrophic cardiomyopathy, dilated cardiomyopathy, myocardial infarction, coronary artery disease, arrhythmias), making it a versatile tool in cardiovascular diagnostics.

From a technical standpoint, the challenge encourages the development of innovative deep learning approaches that enhance data efficiency, generalizability, trustworthiness, and image fidelity in CMR reconstruction. From a clinical perspective, the challenge aims to improve patient care by enabling quicker and more reliable diagnoses of cardiovascular diseases through advanced accelerated CMR imaging techniques.

By addressing the significant variations in physiological structures due to multiple diseases and overcoming the challenges posed by domain shifts in training data, this task has the potential to make substantial advancements in the field of cardiovascular diagnostics.

#### **Keywords**

List the primary keywords that characterize the task.

Cardiac image reconstruction, multiple diseases, foundation model, trustworthy

### **ORGANIZATION**

#### **Organizers**

a) Provide information on the organizing team (names and affiliations).

Chengyan Wang

Human Phenome Institute, Fudan University, China

Michael Markl

SCMR president, Northwestern Univ Chicago, USA

Claudia Prieto

SCMR 2025 programme chair, Kings College London, UK and Pontificia Universidad Católica de Chile, Santiago, Chile

Chen Qin

Department of Electrical and Electronic Engineering & I-X, Imperial College London, UK

Shuo Wang

Digital Medical Research Center, School of Basic Medical Sciences, Fudan University, China

Daniel Kim

SCMR Science committee chair, Northwestern Univ Chicago, USA

Alistair Young,

Kings College London, UK

Jun Lyu

Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

He Wang, Hao Li, Zhensen Chen

Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, China

Guang Yang, Fanwen Wang

Department of Bioengineering/Imperial-X, Imperial College London, UK

Xiaobo Qu, Zi Wang

Department of Electronic Science, Xiamen University, Xiamen, China

Cheng Ouyang

Department of Engineering Science, University of Oxford, UK

Jiayu Zhu

Shanghai United Imaging Healthcare Advanced Technology Research Institute Co., Ltd., China

Yajing Zhang

Clinical Science Manager, MR Business Unit, Philips Healthcare Suzhou, China

Xiahai Zhuang

School of Data Science, Fudan University, China

Yan Li

Department of Radiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Sha Hua

Department of Cardiovascular Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, China

Lianming Wu

Department of Radiology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

Ziqiang Xu,

Shanghai Fuying Medical Technology Co., Ltd., China

Ying-Hua Chu

Siemens Healthineers Ltd., China

Weibo Chen

Philips Healthcare, China

b) Provide information on the primary contact person.

Hao Li, h\_li@fudan.edu.cn

Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, China

### Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place. Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One-time event with fixed submission deadline.

### Challenge venue and platform

a) Report the event (e.g. conference) that is associated with the challenge (if any).

28th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2025), and the Annual Scientific Sessions of the Society for Cardiovascular Magnetic Resonance (SCMR)

b) Report the platform (e.g. grand-challenge.org) used to run the challenge.

synapse.org

c) Provide the URL for the challenge website (if any).

Website for CMRxRecon2023: <https://cmrxrecon.github.io>. Synapse platform for CMRxRecon2023:

<https://www.synapse.org/#!/Synapse:syn51471091/wiki/622170> Website for CMRxRecon2024:

<https://cmrxrecon.github.io/2024/Home.html> Synapse platform for CMRxRecon2024:

<https://www.synapse.org/#!/Synapse:syn54951257/wiki/627141> The new website for this year is under active construction. The new Synapse platform for CMRxRecon2025 (under construction):

<https://www.synapse.org/#!/Synapse:syn59814210/wiki/>

## Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

**Fully automatic.**

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

It should be restricted to the data provided by the previous CMRxRecon challenge as well as data from the 'fastMRI' challenge (the most related public dataset), under the terms and conditions associated with the data usage.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

**They may participate but are not eligible for awards and will not listed in the leaderboard.**

d) Define the award policy. In particular, provide details with respect to challenge prizes.

**The top 5 winners receive monetary awards. We are in active negotiation with the sponsor about the value (approximately \$2000 in total).**

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

All submissions will be reported in the leaderboard. Participating teams can opt out of publication of their results in the leaderboard.

Prize-winning methods will be announced publicly as part of a scientific session at the MICCAI annual meeting. In addition, we also plan to collaborate with SCMR to host a joint workshop and invite representatives from the top three teams to present and exchange ideas.

f) Define the publication policy. In particular, provide details on ...

- ... who of the participating teams/the participating teams' members qualifies as author
- ... whether the participating teams may publish their own results separately, and (if so)
- ... whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

Participating teams with a valid submission can nominate their team members as co-authors for the challenge paper. We reserve the right to exclude teams if they break the challenge rules. Participating teams can publish their own results but after a 3-month embargo period.

## Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

**Docker container will be accepted as submission through the Synapse platform. Submission details will be published at the time point of challenge announcement.**

b) Provide information on the possibility for participating teams to evaluate their algorithms before submitting final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

**Participating teams are allowed to make 3 formal submissions per task. Only the last run submission is officially counted to rank challenge results. Before the final submission on the test set, participants can test their docker containers on the validation dataset to avoid submission errors.**

### Challenge schedule

Provide a timetable for the challenge. Preferably, this should include

- the release date(s) of the training cases (if any)
- the registration date/period
- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

[Apr 1, 2025] website opens for registration, release training and validation images

[Apr 10, 2025] submission system opens for validation

[Aug 1, 2025] submission system opens for testing

[Sept 1, 2025] registration and docker submission deadline

[Oct 8, 2025] release final results during the MICCAI annual meeting

[Jan 31, 2026] challenge summary and invited talk from the winners during the SCMR annual meeting

### Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

**We have received ethics approval from the local ethics committee of School of Basic Medical Sciences, Fudan University granted on 17/11/2021, No. 2021-Y060.**

### Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)
- CC BY-ND (Attribution-NoDerivs)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
- CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

CC BY

### Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

We have released the source code for evaluating and ranking the results at Github:

<https://github.com/CmrRxRecon/CMRxRecon>

b) In an analogous manner, provide information on the accessibility of the participating teams' code.

In the past CMRxRecon2023 challenge, all top ten teams have made their code and method details public, and they have allowed us to consolidate all their information:

<https://github.com/CmrRxRecon/CMRxRecon/tree/main/ModelPool>.

### Conflicts of interest

Provide information related to conflicts of interest. In particular provide information related to sponsoring/funding of the challenge. Also, state explicitly who had/will have access to the test case labels and when.

We declare no conflicts of interest. Test images will only be accessible to the challenge organizers.

## MISSION OF THE CHALLENGE

### Field(s) of application

State the main field(s) of application that the participating algorithms target.

Examples:

- Diagnosis
- Education
- Intervention assistance
- Intervention follow-up
- Intervention planning
- Prognosis

- Research
- Screening
- Training
- Cross-phase

Research, Diagnosis.

### Task category(ies)

State the task category(ies)

Examples:

- Classification
- Detection
- Localization
- Modeling
- Prediction
- Reconstruction
- Registration
- Retrieval
- Segmentation
- Tracking

Reconstruction.

### Cohorts

We distinguish between the target cohort and the challenge cohort. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on ex vivo data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding sex or age (target cohort).

a) Describe the target cohort, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

The target cohort is patients requiring multi-parametric CMR exams, especially patients with arrhythmia and those who cannot adhere to standard imaging protocols. The results of this study will directly accelerate clinical scanning speed, increase exam throughput, and reduce medical costs.

b) Describe the challenge cohort, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

The proposed challenge will release multi-channel k-space data from at least 5 medical centers, more than 10 scanners, including mainstream device vendors (GE, Philips, Siemens, and United imaging), 1.5T, 3.0T and 5.0T field strengths, more than 5 types of sequences (e.g., dark blood, tagging, phase-contrast), including anatomical

views including long-axis, short-axis, outflow tract, and aortic (cross-sectional and sagittal views), and at least 5 diseases such as hypertrophic cardiomyopathy, dilated cardiomyopathy, myocardial infarction, coronary artery disease, and arrhythmias. A total of 600 volunteers with multi-parametric CMR imaging (200 cases for training, 100 validation cases and 300 test cases) will be included, with no less than 100 cases per center. The dataset will include multi-parametric k-space data, consist of cardiac cine, mapping, tagging, phase-contrast, and dark-blood imaging. The dataset also includes imaging of different anatomical views like long-axis (2-chamber, 3-chamber, and 4-chamber), short-axis, outflow tract, and aortic (cross-sectional and sagittal views). Additionally, we have preliminarily set up the Synapse platform for CMRxRecon2025 and uploaded 1% of the data for review (<https://www.synapse.org/#!/Synapse:syn59814210/wiki/628454>).

### Imaging modality(ies)

Specify the imaging technique(s) applied in the challenge.

Magnetic Resonance Imaging

### Context information

Provide additional information given along with the images. The information may correspond ...

a) ... directly to the image data (e.g. tumor volume).

none

b) ... to the patient in general (e.g. sex, medical history).

none

### Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

Multi-parametric CMR data is acquired from the heart and aorta. During the training phase, we will release data from two centers as the training set. We will retain data from three undisclosed centers as the validation set and test set, respectively, to better evaluate the model's generalization performance across different centers.

b) Describe the algorithm target, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

The algorithm target is to assess the reconstruction performance on various k-space sampling based on a universal pre-trained reconstruction foundation model. In the challenge cohort, we recruit volunteers to be scanned with different k-space trajectories and accelerating factors.

### Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.
- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

PSNR, SSIM and NMSE between reconstructed images and ground truth images (fully sampled data).

1. When evaluating SSIM, we will narrow down the assessment field-of-view to the region where the heart is located, to avoid interference from the background area.
2. During the testing and ranking phase, we will invite three radiologists to independently score the top five teams ranked by SSIM. The radiologists will assess several image characteristics using a five-point Likert scale, which includes signal loss, sharpness of heart contour, image artifacts (indicating the extent to which artifacts affect image quality), and clinical utility (ranging from 1 (non-diagnostic) to 5 (excellent)). Each radiologist's score for signal loss, sharpness of heart contour, image artifacts, and clinical utility carries a weight of 25%. The final score will be the weighted sum of these four ratings. We will average the scores from the three radiologists to determine the teams' final scores, which will then be ranked from highest to lowest.

## DATA SETS

### Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

The proposed challenge will release multi-channel k-space data from more than 10 scanners, including mainstream device vendors (GE, Philips, Siemens, and United imaging), 1.5T, 3.0T and 5.0T field strengths.

b) Describe relevant details on the imaging process/data acquisition for each acquisition device (e.g. image acquisition protocol(s)).

We follow the recommendations of CMR exams reported in the previous publication (doi:

10.1007/s43657-02100018x, 10.1007/s43657-021-00018-x). We use 'TrueFISP' for cine, PC and tagging, and 'FLASH' for mapping and dark-blood imaging. The collected imaging planes include long-axis (2-chamber, 3-chamber, and 4-chamber), short-axis, outflow tract, and aortic (cross-sectional and sagittal views). Typically 5-15 slices are acquired. For mapping, signal data were collected at the end of the diastole with ECG triggering. The cardiac cycle is segmented into 15–25 phases with a temporal resolution of around 50 ms. Typical geometrical parameters include: spatial resolution 2.0×2.0 mm<sup>2</sup>, slice thickness 8.0 mm, and slice gap 4.0 mm.

c) Specify the center(s)/institute(s) in which the data was acquired and/or the data providing platform/source (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

The following centers have already contacted us and can support our data collection. Additionally, we are actively reaching out to more international medical centers in the hope of collecting raw data from more sources.

Zhangjiang Imaging Center, Fudan University, China

Department of Radiology, Imperial College London, UK

Department of Radiology, Zhongshan Hospital, Fudan University, China

Department of Electronic Science, Xiamen University, Xiamen, China

Department of Radiology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

Department of Cardiovascular Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, China

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

Data are acquired with a team of radiographers and clinical advisors as appropriate.

### Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

Training cases include fully sampled k-space data and auto-calibration lines (ACS, 24 lines) will be provided in '.mat' format.

Validation cases include under-sampled k-space data, sampling trajectories, and auto-calibration lines (ACS, 24 lines) with various acceleration factors in '.mat' format.

Test cases include fully sampled k-space data, under-sampled k-space data, sampling trajectories and auto-calibration lines (ACS, 24 lines). Test cases will not be released before the challenge ends.

b) State the total number of training, validation and test cases.

1,200 cases in total (600 cases from previous CMRxRecon challenge datasets from single center and 200 new cases from multi-centers for training, 100 validation cases (multi-centers) and 300 test cases (multi-centers)). In the training phase, we will only release data from healthy volunteers as the training set and retain all five disease types as the validation set and test set, to better evaluate the model's generalization performance on previously unseen diseases.

c) Explain why a total number of cases and the specific proportion of training, validation and test cases was chosen.

This will be the largest publicly available dataset in the field. The sample size of training dataset is reasonable for training deep learning reconstruction networks for CMR imaging according to previous literature and our experience. The sample size of validation and test dataset would well represent the distribution of the data cohort and achieve robust evaluation of the models according to last year's 'CMRxRecon' challenge results, in which 60 cases were used for validation and 120 cases for testing

(<https://www.synapse.org/#!Synapse:syn51471091/wiki/622548>). The results of the test set and validation set have shown very high consistency in both scores and team rankings (the top 3 are exactly the same).

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

This dataset will include a diverse range of ethnicities, disease types, and age distributions. We will aim for an age and gender balance between the training and test data.

e) Challenge organizers are encouraged to (partly) use unseen, unpublished data for their challenges. Describe if new data will be used for the challenge and state the number of cases along with the proportion of new data.

N/A

### Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

N.A.

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

N.A.

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

N.A.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

N.A.

### Data pre-processing method(s)

Describe the method(s) used for pre-processing the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

The raw k-space data exported from the scanner will be processed and transformed to '.mat' format using the script provided by our vendor. A readme file will be provided to describe the content and usage of the data.

### Sources of error

a) Describe the most relevant possible error sources related to the image annotation. If possible, estimate the magnitude (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

N.A.

b) In an analogous manner, describe and quantify other relevant sources of error.

N.A.

## ASSESSMENT METHODS

### Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

For cardiac cine, PSNR, SSIM and NMSE between reconstructed images and ground truth images (fully sampled data) will be used for performance evaluation.

1. When evaluating SSIM, we will narrow down the assessment field-of-view to the region where the heart is located, to avoid interference from the background area.
2. During the testing and ranking phase, we will invite three radiologists to independently score the top five teams ranked by SSIM. The scoring will cover three aspects: image quality, image artifacts, and clinical utility. We will consider both the radiologists' scores and the SSIM results to generate a comprehensive ranking.

For cardiac mapping, Root Mean Squared Error (RMSE) of T1 and T2 measurements in the LV myocardium region between reconstructed images and ground truth images (fully sampled data) will be used for performance evaluation.

T1 and T2 calculations are listed below:

A 3-parameter fitting model was used to calculate T1 of each signal time course:

$$S(TI)=A-B e^{(-TI/T_1^{*})}$$

where TI represents the inversion time, and A, B, and  $T_1^{*}$  were three parameters to be estimated. T1 was determined from the resulting A, B, and  $T_1^{*}$  by applying the equation:

$$T_1=T_1^{*} (B/A-1)$$

The T2 value is calculated with the equation:

$$S(t)=A \exp(-t/T_2)$$

where S(t) represents the signal intensity, A is the scale factor and t denote T2 preparation time.

b) Justify why the metric(s) was/were chosen, preferably with reference to the biomedical application.

We keep these metrics aligned with the 'fastMRI' challenge evaluation metrics for fair evaluations.

### Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

We will perform automated segmentation of the images (using a pre-trained cardiac segmentation model) and segment the heart according to AHA guidelines for measurement and statistics. Finally, we will combine all the measurement results and provide an average score.

For cardiac cine, PSNR, SSIM and NMSE between reconstructed images and ground truth images (fully sampled data) will be used for performance evaluation.

1. When evaluating SSIM, we will narrow down the assessment field-of-view to the region where the heart is located, to avoid interference from the background area.
2. During the testing and ranking phase, we will invite three radiologists to independently score the top five teams ranked by SSIM. The radiologists will assess several image characteristics using a five-point Likert scale, which includes signal loss, sharpness of heart contour, image artifacts (indicating the extent to which artifacts affect image quality), and clinical utility (ranging from 1 (non-diagnostic) to 5 (excellent)). Each radiologist's score for

signal loss, sharpness of heart contour, image artifacts, and clinical utility carries a weight of 25%. The final score will be the weighted sum of these four ratings. We will average the scores from the three radiologists to determine the teams' final scores, which will then be ranked from highest to lowest.

For cardiac mapping, Root Mean Squared Error (RMSE) of T1 and T2 measurements in the LV myocardium region between reconstructed images and ground truth images (fully sampled data) will be used for performance evaluation.

b) Describe the method(s) used to manage submissions with missing results on test cases.

Participating teams are required to submit docker containers and process all the cases in the test set on our server. For the cases without valid output, we will assign it to the lowest value of metric.

c) Justify why the described ranking scheme(s) was/were used.

PSNR, SSIM and RMSE are commonly used metrics and proven valid in almost all existing related works and previous challenges. This strategy is consistent with the 'fastMRI' challenge and our previous 'CMRxRecon' challenge last year.

### Statistical analyses

a) Provide details for the statistical methods used in the scope of the challenge analysis. This may include

- description of the missing data handling,
- details about the assessment of variability of rankings,
- description of any method used to assess whether the data met the assumptions, required for the particular statistical approach, or
- indication of any software product that was used for all data analysis methods.

Please see the description of metrics and ranking methods above. We will incorporate the Wilcoxon signed rank test to conduct statistical analysis on all challenge results.

b) Justify why the described statistical method(s) was/were used.

Please see the description of metrics and ranking methods above.

### Further analyses

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

N/A

## ADDITIONAL POINTS

### References

Please include any reference important for the challenge design, for example publications on the data, the annotation process or the chosen metrics as well as DOIs referring to data or code.

#### Reference of the CMR imaging acquisition protocol:

1. Wang C, Lyu J, Wang S, et al. CMRxRecon: An open cardiac MRI dataset for the competition of accelerated image reconstruction. arXiv preprint arXiv:2309.10836, 2023.
2. Lyu J, Qin C, Wang S, et al. The state-of-the-art in Cardiac MRI Reconstruction: Results of the CMRxRecon Challenge in MICCAI 2023. arXiv preprint arXiv:2404.01082, 2023.
3. Wang C, Li Y, Lv J, et al. Recommendation for Cardiac Magnetic Resonance Imaging-Based Phenotypic Study: Imaging Part. Phenomics. 2021, 1(4): 151-170. <https://doi.org/10.1007/s43657-021-00018-x>

#### Reference for previously developed reconstruction algorithms:

1. Wang C, Jang J, Neisius U, et al. Black blood myocardial T2 mapping. Magnetic resonance in medicine. 2019, 81(1): 153-166.
2. Lyu J, Li G, Wang C, et al. Region-focused multi-view transformer-based generative adversarial network for cardiac cine MRI reconstruction. Medical Image Analysis, 2023: 102760.
3. Qin C, Schlemper J, Caballero J, et al. Convolutional recurrent neural networks for dynamic MR image reconstruction. IEEE transactions on medical imaging, 2018, 38(1): 280-290.
4. Qin C, Duan J, Hammernik K, et al. Complementary time-frequency domain networks for dynamic parallel MR image reconstruction. Magnetic Resonance in Medicine, 2021, 86(6): 3274-3291.
5. Lyu J, Tian Y, Cai Q, et al. Adaptive channel-modulated personalized federated learning for magnetic resonance image reconstruction. Computers in Biology and Medicine, 2023, 165: 107330.
6. Lyu J, Tong X, Wang C. Parallel Imaging With a Combination of SENSE and Generative Adversarial Networks (GAN). Quantitative Imaging in Medicine and Surgery. 2020, 10(12): 2260-2273.
7. Lyu J, Sui B, Wang C, et al. DuDoCAF: Dual-Domain Cross-Attention Fusion with Recurrent Transformer for Fast Multi-contrast MR Imaging. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2022: 474-484.
8. Ouyang C, Schlemper K, et al. Generalizing Deep Learning MRI Reconstruction across Different Domains, arXiv preprint arXiv: 1902.10815, 2019.

#### Further comments

Further comments from the organizers.

No