



OPEN

DATA DESCRIPTOR

MSHF: A Multi-Source Heterogeneous Fundus (MSHF) Dataset for Image Quality Assessment

Kai Jin¹ , Zhiyuan Gao¹, Xiaoyu Jiang², Yaqi Wang³, Xiaoyu Ma⁴, Yunxiang Li⁵ & Juan Ye¹ 

Image quality assessment (IQA) is significant for current techniques of image-based computer-aided diagnosis, and fundus imaging is the chief modality for screening and diagnosing ophthalmic diseases. However, most of the existing IQA datasets are single-center datasets, disregarding the type of imaging device, eye condition, and imaging environment. In this paper, we collected a multi-source heterogeneous fundus (MSHF) database. The MSHF dataset consisted of 1302 high-resolution normal and pathologic images from color fundus photography (CFP), images of healthy volunteers taken with a portable camera, and ultrawide-field (UWF) images of diabetic retinopathy patients. Dataset diversity was visualized with a spatial scatter plot. Image quality was determined by three ophthalmologists according to its illumination, clarity, contrast and overall quality. To the best of our knowledge, this is one of the largest fundus IQA datasets and we believe this work will be beneficial to the construction of a standardized medical image database.

Background & Summary

Fundus photography is the most widely used imaging modality for screening diabetic retinopathy (DR), glaucoma, age-related macular degeneration, and other eye diseases¹. With the development of artificial intelligence (AI), automatic disease screening via fundus imaging has become a popular topic for researchers and clinical practitioners^{2,3}. Many algorithms have been investigated, and some have already been used in clinical practice⁴⁻⁶. The quality of fundus images is key to the performance of diagnosis models, as an important preliminary step. Therefore, image quality assessment (IQA) is vital for automated systems.

The most reliable method to assess an image quality requires doctors to assess the original images, but it entails a heavy workload. Over the past years, automated IQA has been developed to score the fundus images⁷⁻¹⁰. Once a model is trained, it can produce fast and real-time predictions, improve the workflow and optimize image acquisition, making the whole process more efficient. To train an excellent model, a well-collected dataset is very important.

Several fundus IQA datasets, which are summarized in Table 1, have been established for public use: DRIMDB¹¹, Kaggle DR Image Quality¹², EyeQ Assessment¹³, DeepDRiD¹⁴, etc. Many IQA studies have been carried out on these datasets¹⁵.

Nevertheless, these popular datasets have some drawbacks:

- First, the variety of fundus images is limited. In clinical practice, fundus images have various forms, and different kinds of imaging approaches meet the variable needs of clinical practitioners. For example, color

¹Eye Center, The Second Affiliated Hospital, School of Medicine, Zhejiang University, Zhejiang Provincial Key Laboratory of Ophthalmology, Zhejiang Provincial Clinical Research Center for Eye Diseases, Zhejiang Provincial Engineering Institute on Eye Diseases, Zhejiang, Hangzhou, 310009, China. ²College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China. ³College of Media, Communication University of Zhejiang, Hangzhou, 310018, China. ⁴Institute of Intelligent Media, Communication University of Zhejiang, Hangzhou, 310018, China. ⁵College of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, 310018, China. ✉e-mail: yejuan@zju.edu.cn

Dataset	Year	Number	Quality label	Disease	Annotators
DRIMDB	2014	216 CFP images	'good', 'bad' or 'outlier'	DR, non-DR	NA
Kaggle DR Image Quality	2018	88702 CFP images	'good' or 'poor' for overall quality	DR, non-DR	NA
EyeQ Assessment	2019	28792 CFP images	'good', 'usable' or 'reject' for overall quality	DR, non-DR	2
DeepDRiD	2022	2000 CFP images and 256 UWF images	score criteria 0~10 for artifact, clarity, field definition and 0/1 for overall quality	DR, non-DR	3
MSHF (proposed)	2022	500 CFP images, 500 UWF images and 302 portable camera images	0/1 for illumination, clarity, contrast and overall quality	DR, healthy, glaucoma	3

Table 1. Summarization of publicly available fundus image quality assessment datasets.

fundus photography (CFP) is a widely used kind of fundus image for screening of various ocular disorders. The portable fundus camera is a convenient, hand-held device designed for use in rural areas, and has played an important role in the development of telemedicine. However, the images may lack lesion details, and artifacts are commonly found¹⁶. Ultrawide-field (UWF) imaging is an advanced fundus photography technique, and is becoming more and more popular in clinical scenarios. UWF machines are always costly, though, and cost-effectiveness remains an important consideration. Therefore, an ideal fundus image dataset should consider the above clinical scenarios.

- Second, the criteria to decide the image quality is not very clear. Most datasets only considered the overall quality, making the label rather subjective and not explainable enough. Liu *et al.*¹⁴ provided a solution by using a scoring criteria according to artifact, clarity, field definition and overall quality, making the image quality appear more objective. The detailed quality standard makes the label more persuasive.
- Third, existing fundus IQA datasets are based on DR image datasets, and fundus images of other retinal diseases as well as healthy volunteers were not considered. Some fundus diseases may affect the judgment of image quality due to the lesion of the disease. It's significant for IQA datasets to increase the variety of fundus diseases.

Taking all of these factors into consideration, a public fundus IQA dataset consisting of various forms of fundus images from patients and healthy volunteers with detailed quality labels would be fundamental.

In this paper, we propose a multi-source heterogeneous fundus (MSHF) dataset that contains 500 CFP, 302 portable camera images and 500 UWF images with various source domains from DR and glaucoma patients as well as normal people. For each image, 4 labels are provided: illumination, clarity, contrast and overall quality. Our major contributions can be summarized as follows:

- **Image:** The dataset is composed of sub-databases collected from different devices with diverse appearance patterns, including 500 CFP, 302 portable camera images and 500 UWF images of normal eyes and 2 different eye diseases.
- **Labels:** Each image is labelled by illumination, clarity, contrast and overall quality with 0 or 1, and the dataset has 5208 labels in total.

We believe that the publication of the MSHF dataset will considerably facilitate AI-related fundus IQA research and promote translation from technology to clinical use.

Methods

An overview of the study approach and methodology is presented in Fig. 1.

Data collection. A total of 1302 images were retrospectively collected to form 7 sub-datasets: DR-XJU, DR-ZJU, Glaucoma, Healthy, Local1, Local2 and UWF-mosaic. There are three types of images in these datasets: CFP images, portable camera images and UWF images. These images are from 904 patients, with ages ranging from 21 to 77 years. Written consent was signed by every participant before examinations to inform them that the images would be used for research purpose. Ethical approval for the study was obtained from the Ethics Committee of ZJU-2.

Specifically, DR-XJU, DR-ZJU, Glaucoma and Healthy subsets contained CFP images that were centerfield, including the optic disc and the macular area.

Images of DR-XJU were collected from patients with diabetic retinopathy at the Second Affiliated Hospital of Xi'an Jiaotong University (SAHXJU), captured with a Kowa non-mydratic fundus camera (Kowa Company, Tokyo) with 45 degrees fields of view (FOV) and at 1924 by 1556 pixels.

Images of DR-ZJU, Glaucoma and Healthy were respectively collected from patients with diabetic retinopathy, glaucoma or no disease diagnosed at the Eye Center at the Second Affiliated Hospital of Zhejiang University School of Medicine (SAHZJU). The imaging device was a tabletop TRC-NW8 fundus camera (Top-Con Medical Systems, Tokyo) with 50 degrees FOV and a resolution of 1924 by 1556 pixels.

Local 1 and Local 2 contained portable camera images from healthy volunteers, and the imaging field included centerfield and other locations. These datasets were collected at the Eye Center at SAHZJU, captured

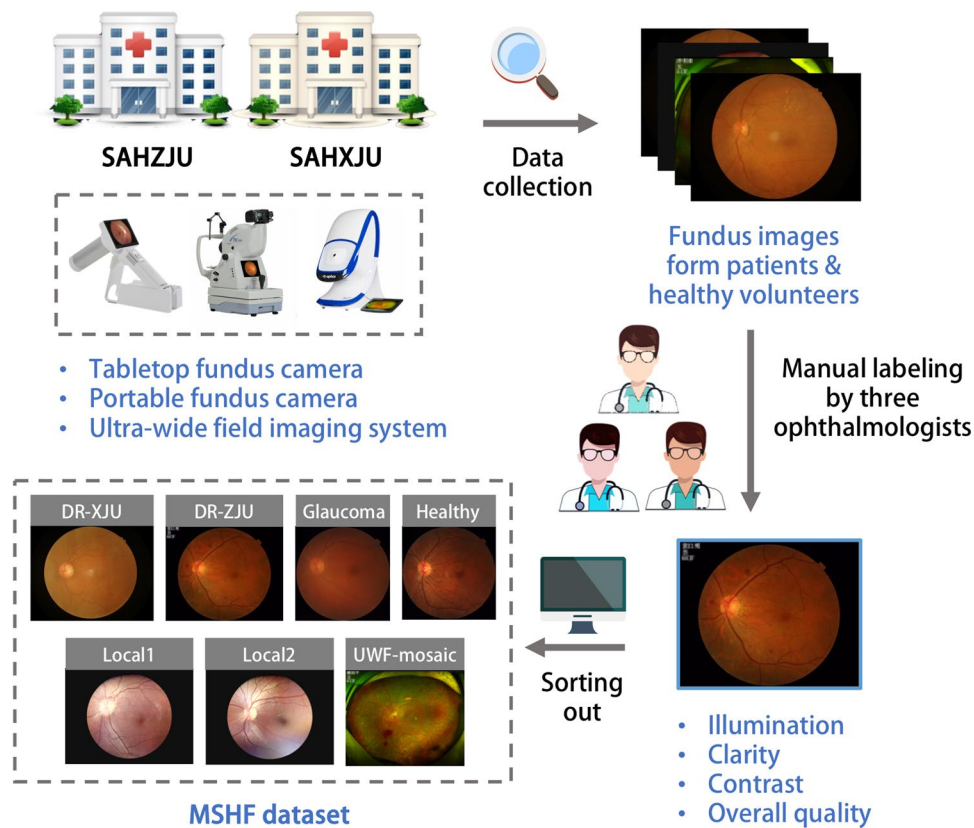


Fig. 1 An overview of the study approach and methodology.

Image type	Device	Name	Number	FOV	Resolution	Disease
CFP images	Kowa Nonmyd fundus camera	DR-XJU	235	45 degrees	1924 × 1556	DR
		DR-ZJU	187	50 degrees	1924 × 1556	DR
	TRC-NW8 fundus camera	Glaucoma	52	50 degrees	1924 × 1556	Glaucoma
		Healthy	26	50 degrees	1924 × 1556	/
Portable camera images	DEC200 portable fundus camera	Local1	199	60 degrees	2560 × 1960	/
		Local2	103	60 degrees	2560 × 1960	/
UWF images	Optos ultra-wide field imaging system	UWF-mosaic	500	200 degrees	1924 × 1556	DR

Table 2. Basic information on the multi-source heterogeneous fundus (MSHF) dataset.

with a DEC200 portable fundus camera (Med-imaging Integrated Solution Inc., Taiwan) with 60 degrees FOV and a resolution of 2560 by 1960 pixels. The difference between Local1 and Local2 was the imaging time period.

UWF-mosaic included UWF images from diabetic retinopathy patients. This dataset was also collected at the Eye Center at SAHZJU, and the capture device was an Optos ultra-wide field imaging system (Optos Plc Fife, Scotland) with 200 degrees FOV and a resolution of 1924 by 1556 pixels.

Detailed descriptions of the MSHF dataset is shown in Table 2. To show the diversity of the images in the MSHF dataset, we converted all images from the RGB color space to the Lab color space, and created a spatial scatter plot to show the distribution, as shown in Fig. 2. There are seven kinds of symbols in the figure, representing the seven subsets.

Quality evaluation. To facilitate the clinical application, the evaluation standard is a generic quality gradation scale that adhered to the generic-but-not-structural principle, as listed in Table 3. The overall quality represents the general impression of the images, and suggests whether or not the image is useable, while the illumination, clarity and contrast are parameters based on the characteristics of human visual system, and indicates the potential aspects to improve the image quality.

Images of the MSHF dataset were labelled by three ophthalmologists according to the principle. If the image was of good quality in a particular category, it was marked as ‘1’, and if not, as ‘0’. The ground truth was decided by the majority rule. Examples of high- and low-quality images of CFP, portable camera and UWF are shown in Fig. 3.

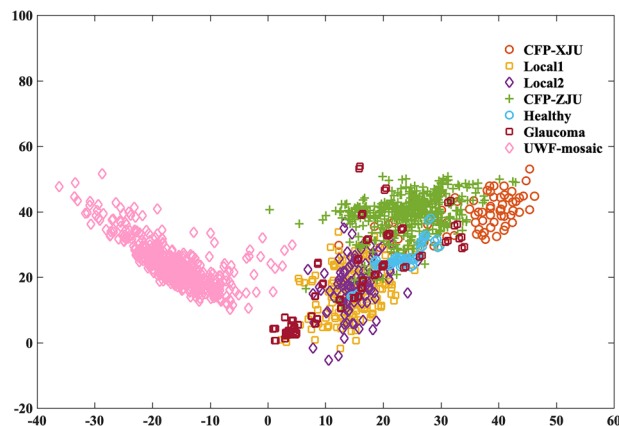


Fig. 2 Spatial scatter plot of all datasets. The horizontal coordinate is the average of the red-green opposite color intensity channel of each image, and the vertical coordinate is the average of the yellow-blue opposite-color intensity channel. Images were converted from the RGB color space to the Lab color space. The ‘a’ channel represents red-green opposite color intensity; if the positive value is higher, then it is redder, otherwise it is greener. The ‘b’ channel represents the strength of yellow-blue opposite colors; if the positive value is higher, then it is more yellow, otherwise it is bluer. The horizontal coordinate is the average ‘a’ channel of each image, and the vertical coordinate is the average ‘b’ channel.

Item	Principle description
Illumination	0: The image is overexposure, underexposure, or with the uneven illumination or color.
	1: The image has good illumination and color.
Clarity	0: The image has noticeable blur in optic disks, vessels, or background.
	1: The image is clear, with no blur in the field.
Contrast	0: Image contrast is low such that the band of pixel intensity is narrow, making it difficult to distinguish the vessel or lesions.
	1: Image has good contrast.
Overall quality	0: The image quality is bad, making it hard to diagnosing.
	1: The image quality is good, and can be used for diagnosing.

Table 3. Generic quality gradation scale.

Dataset division. To make the MSHF dataset applicable for further AI model building, the dataset was manually divided into the training set (80%) and the test set (20%). The training set was used for learning and the test set was used for testing. There was no intersection between the 2 sets, and the variety of images was distributed equally. The 2 sets contained basically equal ratio of good- or poor-quality images. It is worth noting that we offered a possible way of split, and we did not mean to restrict its use. Future researchers can freely use this data set to achieve their research purposes.

Data Records

The MSHF dataset has been uploaded to Figshare in the form of a zipped file¹⁷. The unzipped file folder contains the original fundus photographs and quality evaluation scores. The unzipped file is organized into 2 folders and 2 Microsoft office Excel list, named “Original”, “AI-use”, “MSHF_quality_scores.xlsx” and “Individual_scores.xlsx”, respectively. The “Original” folder contains 3 subfolders, named “CFP”, “Portable_camera” and “UWF-mosaic”. Among them, the “CFP” folder is consisted of “DR-XJU”, “DR-ZJU”, “Glaucoma” and “Healthy”, and the “Portable_camera” folder is consisted of “Local1” and “Local2”. Images in these 7 folders are stored, named and arranged in the same way. The “AI-use” folder contains “train” and “test” subfolders consisting of 1042 images recommended for training and 260 images for testing. Images are named the same as they are in the “Original” folder. The current data split strategy is proposed by our team and might be subject to change for other research purposes. In the file “MSHF_quality_scores.xlsx”, there are 7 sheets corresponding to the 7 subfolders “DR-XJU”, “DR-ZJU”, “Glaucoma”, “Healthy”, “Local1”, “Local2” and “UWF-mosaic”. Five columns are presented in each sheet. The first column represents the name of each image. The subsequent columns indicate the final score of “illumination”, “clarity”, “contrast” and “overall quality” of the image. In the file “Individual_scores.xlsx”, there are scores of each image annotated by three annotators. The score of each item is either 1 or 0.

Technical Validation

Dataset characteristics. There are 1302 fundus images and their corresponding quality labels in the MSHF dataset. These images are acquired from 952 subjects. The mean age of the subjects was 51 years, with a standard deviation of 20.12 years. There were 602 images from female subjects and 700 from males. All the subjects were Asian. Detailed annotation of the dataset is presented in Table 4.

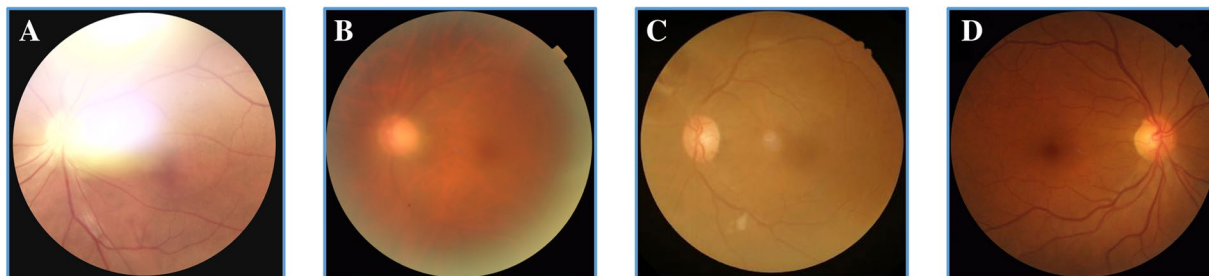


Fig. 3 Examples of images of low and high quality. (A–D) are CFP images, (E–H) are portable camera images, and (I–L) are UWF images. (A), (E) and (I) show the presence of uneven illumination or color. (B), (F) and (J) show the presence of blur that affect the clarity. (C), (G) and (K) show the presence of low contrast. (D), (H), and (L) are images of good quality in every aspect.

Item	Illumination		Clarity		Contrast		Overall quality	
	0	1	0	1	0	1	0	1
DR-XJU	36	199	120	115	78	157	117	118
DR-ZJU	31	156	34	153	6	181	40	147
Glaucoma	45	7	48	4	42	10	50	2
Healthy	2	24	0	26	0	26	0	26
Local1	158	41	94	105	85	114	142	57
Local2	78	25	59	44	41	62	77	26
UWF-mosaic	215	285	163	337	50	450	168	332

Table 4. Dataset annotations.

Images from healthy volunteer are generally in good quality, in terms of every aspect. Images from DR patients have mixed results of overall quality, and good-quality images account for about 60%. However, nearly all images from glaucoma patients are in bad quality. It might be explained that glaucoma patients are generally the elderly, and their pupil cannot be dilated because of the intraocular pressure. The distribution of poor-quality images increases the diversity of the MSHF dataset, as shown in Fig. 2. There are clear differences in the color distributions of the different sub-datasets, and UWF-mosaic differs significantly from other datasets. Glaucoma dataset is also special, because the distribution seems random. LOCAL_1 and LOCAL_2 almost overlap each other, and DR-XJU is a little different from DR-ZJU. The characteristics of the MSHF dataset make it similar to clinical scenarios, and in AI area, the diversity can help to develop robust algorithms.

Inter-annotator consistency. To evaluate the inter-annotator consistency of our dataset, Fleiss Kappa coefficients between the annotators of the four aspects were calculated. The Fleiss Kappa coefficient of ‘contrast’ was 0.786, indicating a substantial agreement, and the results of ‘illumination’, ‘clarity’ and ‘overall quality’ was 0.820, 0.804 and 0.848, suggesting an almost perfect agreement.

Usage Notes

The entire dataset can be downloaded from the link mentioned above. It should be mentioned that the data split strategy was made considering the quality score and the variety of data source, and we did this for the convenience of further artificial intelligence use. For researchers who use the dataset for other purpose, we expect them to cite this paper in their research output and acknowledge the contribution of this dataset in their study.

Code availability

No novel code was used in the construction of MSHF dataset.

Received: 11 November 2022; Accepted: 27 April 2023;

Published online: 17 May 2023

References

1. Poplin, R. *et al.* Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng* **2**, 158–164, <https://doi.org/10.1038/s41551-018-0195-0> (2018).
2. Ting, D. S. W. *et al.* Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol* **103**, 167–175, <https://doi.org/10.1016/j.aopr.2022.100078> (2019).
3. Jin, K. & Ye, J. Artificial intelligence and deep learning in ophthalmology: Current status and future perspectives. *Advances in Ophthalmology Practice and Research* **2**, 100078, <https://doi.org/10.1016/j.aopr.2022.100078> (2022).

4. Ting, D. S. W. *et al.* Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA* **318**, 2211–2223, <https://doi.org/10.1001/jama.2017.18152> (2017).
5. Gulshan, V. *et al.* Performance of a Deep-Learning Algorithm vs Manual Grading for Detecting Diabetic Retinopathy in India. *JAMA Ophthalmology* **137**, 987–993, <https://doi.org/10.1001/jamaophthalmol.2019.2004> (2019).
6. Sayres, R. *et al.* Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy. *Ophthalmology* **126**, 552–564, <https://doi.org/10.1016/j.ophtha.2018.11.016> (2019).
7. Li, Z. *et al.* Development of a deep learning-based image eligibility verification system for detecting and filtering out ineligible fundus images: A multicentre study. *Int J Med Inform* **147**, 104363, <https://doi.org/10.1016/j.ijmedinf.2020.104363> (2021).
8. Karlsson, R. A. *et al.* Automatic fundus image quality assessment on a continuous scale. *Comput Biol Med* **129**, 104114, <https://doi.org/10.1016/j.compbiomed.2020.104114> (2021).
9. Wang, J. *et al.* Automated Explainable Multidimensional Deep Learning Platform of Retinal Images for Retinopathy of Prematurity Screening. *JAMA Netw Open* **4**, e218758, <https://doi.org/10.1001/jamanetworkopen.2021.8758> (2021).
10. Shen, Y. *et al.* Domain-invariant interpretable fundus image quality assessment. *Med Image Anal* **61**, 101654, <https://doi.org/10.1016/j.media.2020.101654> (2020).
11. Sevik, U., Kose, C., Berber, T. & Erdol, H. Identification of suitable fundus images using automated quality assessment methods. *Journal of Biomedical Optics* **19**, 046006, <https://doi.org/10.1117/1.JBO.19.4.046006> (2014).
12. Zhou, K. *et al.* in *Computational Pathology and Ophthalmic Medical Image Analysis Lecture Notes in Computer Science*. Ch. Chapter 29, 245–252, <https://doi.org/10.1007/978-3-030-00949-6> (2018).
13. Fu, H. *et al.* Evaluation of Retinal Image Quality Assessment Networks in Different Color-Spaces. in *MICCAI*, pp 48–56, <https://doi.org/10.48550/arXiv.1907.05345> (2019).
14. Liu, R. *et al.* DeepDRiD: Diabetic Retinopathy-Grading and Image Quality Estimation Challenge. *Patterns (N Y)* **3**, 100512, <https://doi.org/10.1016/j.patter.2022.100512> (2022).
15. Raj, A., Tiwari, A. K. & Martini, M. G. Fundus image quality assessment: survey, challenges, and future scope. *IET Image Processing* **13**, 1211–1224, <https://doi.org/10.1049/iet-ipr.2018.6212> (2019).
16. Rogers, T. W. *et al.* Evaluation of an AI system for the detection of diabetic retinopathy from images captured with a handheld portable fundus camera: the MAILOR AI study. *Eye (Lond)* **35**, 632–638, <https://doi.org/10.1038/s41433-020-0927-8> (2021).
17. Jin, K. *et al.* MSHF: A Multi-Source Heterogeneous Fundus (MSHF) Dataset for Image Quality Assessment, Figshare <https://doi.org/10.6084/m9.figshare.21507564.v1> (2022).

Acknowledgements

This work was financially supported by Natural Science Foundation of China (grant number 82201195), Natural Science Foundation of Zhejiang Province (grant number LQ21H120002), Medical and Health Science and Technology Program of Zhejiang Province (grant number 2021RC064), and Clinical Medical Research Center for Eye Diseases of Zhejiang Province (grant number 2021E50007).

Author contributions

Kai Jin: Conceptualization, Project administration, Manual segmentation, Writing, Reviewing. Zhiyuan Gao: Conceptualization, Methodology, Writing, Visualization, Reviewing. Xiaoyu Jiang: Project conduction, Writing, Reviewing. Yaqi Wang: Project administration, Reviewing. Xiaoyu Ma: Methodology, Writing, Reviewing. Yunxiang Li: Methodology, Reviewing. Juan Ye: Conceptualization, Funding acquisition, Project administration, Methodology, Writing, Reviewing.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Additional information

Correspondence and requests for materials should be addressed to J.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023