Genomics Proteomics Bioinformatics 21 (2023) 1059–1065

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

DATABASE

OBIA: An Open Biomedical Imaging Archive

Enhui Jin [1,3,4,#], Dongli Zhao [2,#], Gangao Wu [1,3,4,#], Junwei Zhu [1,3], Zhonghuang Wang [1,3,4], Zhiyao Wei [2], Sisi Zhang [1,3], Anke Wang [1,3], Bixia Tang [1,3], Xu Chen [1,3], Yanling Sun [1,3,*], Zhe Zhang [5,*], Wenming Zhao [1,3,4,*], Yuanguang Meng [2,5,*]

[1] *National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China*
[2] *Chinese People's Liberation Army (PLA) Medical School, Beijing 100853, China*
[3] *CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China*
[4] *University of Chinese Academy of Sciences, Beijing 100049, China*
[5] *Department of Obstetrics and Gynecology, Seventh Medical Center of Chinese PLA General Hospital, Beijing 100700, China*

Received 28 May 2023; revised 26 September 2023; accepted 29 September 2023
Available online 6 October 2023

Handled by Yudong Zhang

KEYWORDS

Open Biomedical Imaging Archive;
Database;
Biomedical imaging;
De-identification;
Quality control

Abstract With the development of artificial intelligence (AI) technologies, **biomedical imaging** data play an important role in scientific research and clinical application, but the available resources are limited. Here we present **Open Biomedical Imaging Archive** (OBIA), a repository for archiving biomedical imaging and related clinical data. OBIA adopts five data objects (Collection, Individual, Study, Series, and Image) for data organization, and accepts the submission of biomedical images of multiple modalities, organs, and diseases. In order to protect personal privacy, OBIA has formulated a unified **de-identification** and **quality control** process. In addition, OBIA provides friendly and intuitive web interfaces for data submission, browsing, and retrieval, as well as image retrieval. As of September 2023, OBIA has housed data for a total of 937 individuals, 4136 studies, 24,701 series, and 1,938,309 images covering 9 modalities and 30 anatomical sites. Collectively, OBIA provides a reliable platform for biomedical imaging data management and offers free open access to all publicly available data to support research activities throughout the world. OBIA can be accessed at https://ngdc.cncb.ac.cn/obia.

[*] Corresponding authors.
E-mail: meng6512@vip.sina.com (Meng Y), zhaowm@big.ac.cn (Zhao W), tj.zhe.zhang@gmail.com (Zhang Z), sunyanling@big.ac.cn (Sun Y).
[#] Equal contribution.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

https://doi.org/10.1016/j.gpb.2023.09.003
1672-0229 © 2023 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

The introduction of advanced imaging technologies has greatly facilitated the development of non-invasive diagnoses. Currently, biomedical images can clearly depict the internal structure (anatomy), morphology, and physiological functions from the molecular scale to the cellular, organ, tissue, lesion, and even the entire organism [1], providing crucial evidence for diagnosis and treatment response assessment [2,3]. Imaging data generated during patient visits has formed a huge accumulation. However, incomplete sharing systems make it challenging for researchers and clinicians to collaborate on utilizing these images to gain significant insights into health and disease [4]. Furthermore, the demand for rapid diagnosis promotes the application of artificial intelligence (AI) in biomedical imaging, and the development of reliable and robust AI algorithms requires sufficiently large and representative image datasets [5,6]. Thus, high-quality biomedical imaging data sharing plays an important role in promoting scientific discoveries and improving diagnostic accuracy.

The National Institutes of Health (NIH) in America sponsors several repositories. The Medical Imaging and Data Resource Center (MIDRC) [7] serves as an open-access platform for COVID-19-related medical images and associated data. Image and Data Archive (IDA) [8], NITRC Image Repository (NITRC-IR) [9], Federal Interagency Traumatic Brain Injury Research (FITBIR) [10], OpenNeuro [11], and National Institute of Mental Health Data Archive (NDA) [12] are dedicated to collecting neuro and brain imaging. The Cancer Imaging Archive (TCIA) [13] and Imaging Data Commons (IDC) [14] are cancer imaging repertories, with TCIA providing images locally and IDC affording collections in the Cancer Research Data Commons (CRDC) cloud environment. Regarding breast cancer, the Cancer Research United Kingdom (UK) funds the OPTIMAM Mammography Image Database (OMI-DB) [15], and the University of Porto in Portugal funds the Breast Cancer Digital Repository (BCDR) [16], providing annotated breast cancer images and clinical details. Most of these repositories support data de-identification and quality control, except NITRC-IR and IDC. Additionally, some universities or institutions provide open-source datasets, such as Open Access Series of Imaging Studies (OASIS) [17], EchoNet-Dynamic [18], Cardiac Acquisitions for Multi-structure Ultrasound Segmentation (CAMUS) project [19], Chest X-ray [20], and Structured Analysis of the Retina (STARE) [21]. In China, the Huazhong University of Science and Technology provides an open resource named integrative Computed Tomography (CT) images and CFs for COVID-19 (iCTCF) [22], which includes CT images and clinical features of patients with pneumonia (including COVID-19 pneumonia). There is still a lack of databases that are dedicated to storing and accepting submissions for various diseases and modalities.

To address this issue, we established the Open Biomedical Imaging Archive (OBIA; https://ngdc.cncb.ac.cn/obia), a repository for archiving biomedical imaging data and related clinical data. As a core database resource in the National Genomics Data Center (NGDC) [23], part of the China National Center for Bioinformation (CNCB; https://ngdc.cncb.ac.cn/), OBIA accepts image submissions from all over the world and provides free open access to all publicly available data to support global research activities. OBIA supports de-identification, management, and quality control of imaging data, providing data services such as browsing, retrieval, and downloading, thus promoting the reuse of existing imaging data and clinical data.

Implementation

Database construction

OBIA is implemented using Spring Boot (a framework easy to create standalone Java applications; https://spring.io/projects/spring-boot) as the backend framework. The frontend user interfaces are developed using Vue.js (an approachable, performant, and versatile framework for building web user interfaces; https://vuejs.org/) and Element UI (a Vue 2.0-based component library for developers, designers, and product managers; https://element.eleme.cn/). The charts on the web page are constructed using ECharts (an open-source JavaScript visualization library; https://echarts.apache.org). All metadata information is stored in MySQL (a free and popular relational database management system; https://www.mysql.com/).

Image retrieval

Deep learning-based methods, such as scale-invariant feature transform (SIFT) [24], local binary patterns (LBP) [25], and histogram of oriented gradient (HOG) [26], demonstrate better performance compared to traditional methods. Deep neural networks excel at extracting superior features for retrieving multimodal medical images of various body organs [27] and enhancing ranking performance in the case of small sample sizes [28].

In OBIA, we leveraged TCIA multi-modal cancer data and utilized EfficientNet [29] as the feature extractor. We trained the model using a triplet network and an attention module to compress the image into a discrete hash value (**Figure 1**). We subsequently convert the trained model into TensorRT format to accelerate inference performance and reduce inference latency. We converted the trained model into TensorRT format to enhance inference performance and reduce latency. To store the hash codes, we employed Faiss, a high-performance similarity search library developed by Facebook AI Research and commonly used in deep learning. We computed image similarity using the hamming distance and returned the most similar images to the query. Our model achieved a mean average precision (MAP) value that surpassed the performance of existing advanced image retrieval models on the TCIA dataset.

Database content and usage

Data model

Imaging data in OBIA are organized into five objects: Collection, Individual, Study, Series, and Image (**Figure 2**). "Collection", bearing an accession number prefixed with "OBIA", provides an overall description for a complete submission. "Individual", possessing an accession number

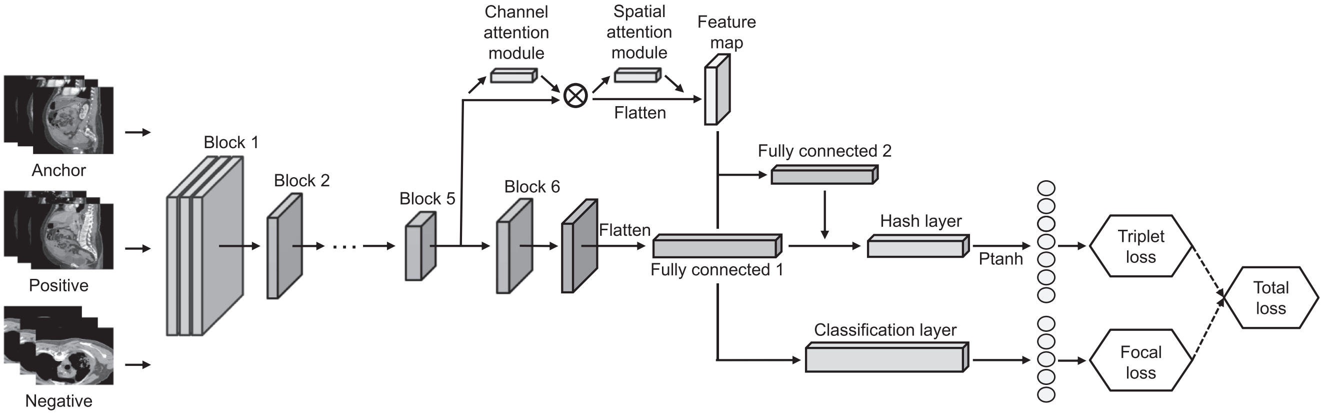

Figure 1 Deep triplet hashing based on attention and layer fusion module
The model uses EfficientNet-B6 as the backbone network and utilizes the CBAM attention module in Block 5 to obtain feature maps. Layer fusion is employed in the fully connected layers, and focal loss and triplet loss are used to generate hash code and class embedding. CBAM, convolutional block attention module.

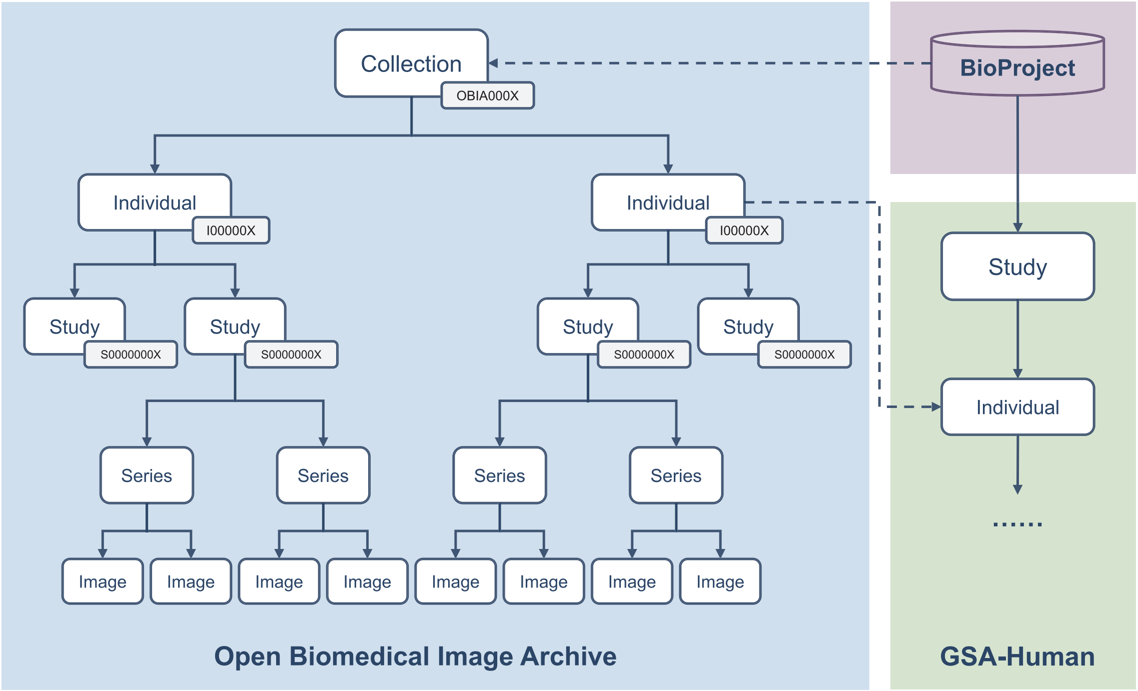

Figure 2 OBIA data model
The Collection and Individual in OBIA can be linked to BioProject and GSA-Human, respectively. The accession numbers for data objects, including Collection, Individual, and Study, are indicated in the gray boxes. Collection accession numbers have a prefix of "OBIA" followed by four consecutive digits, Individual accession numbers have a prefix of "I" followed by six consecutive digits, and Study accession numbers have a prefix of "S" followed by eight consecutive digits. OBIA, Open Biomedical Imaging Archive.

prefixed with "I", defines the characteristics of a human or non-human organism receiving, or registered to receive, healthcare services. "Study", adopting an accession number prefixed with "S", contains descriptive information about radiological examinations performed on an individual. A study may be divided into one or more "Series" according to different logics, such as body part or orientation. "Image" describes the pixel data of a single Digital Imaging and Communications in Medicine (DICOM) file, and an image is related to a single series within a single study. Based on these standardized data objects, OBIA connects the image structure defined by the DICOM standard with actual research projects, realizing data sharing and exchange. Besides, each collection in OBIA is linked to BioProject [30] (https://ngdc.cncb.ac.cn/bioproject/) to provide descriptive metadata about the research project. And the individual in OBIA can be associated with GSA-Human [31] (https://ngdc.cncb.ac.cn/gsa-human/) by individual accession number, if available, which links imaging data with genomic data for researchers to perform multi-omics analysis.

De-identification and quality control

Images may contain protected health information (PHI) and require appropriate handling to minimize the risk of patient privacy breaches. In order to retain as much valuable scientific information as possible while removing PHI, OBIA provides a unified de-identification and quality control mechanism (**Figure 3**) based on the DICOM PS 3.15 Appendix E: Attribute Confidentiality Profile (https://www.dicomstandard.org/). The key elements and rules we adopted include: (1) clean pixel data; (2) clean descriptors; (3) retain longitudinal temporal information modified dates; (4) retain patient characteristics; (5) retain device identity; and (6) retain safe private tags.

OBIA utilizes the Radiological Society of North America (RSNA) MIRC clinical trial processor (CTP) (https://mircwiki.rsna.org/index.php?title=MIRC_CTP) for most of the de-identification work. We constructed a CTP pipeline and developed a common base de-identification script to remove or blank certain standard tags containing or potentially containing PHI. This script also maps local patient IDs to OBIA individual accessions. As for private tags, since they are vendor-specific and scanner-specific and can contain almost anything, we use PyDicom (https://pypi.org/project/pydicom/) to retain attributes that are purely numeric. Some studies determine private element definitions by reading manufacturers' DICOM conformance statements [32]. However, in our practice, the wide variety of image sources made this work too time-consuming, and some conformance statements were not available. In addition to metadata elements, ultrasounds or screen captures usually add some "burned-in" annotations in the pixels data to interpret the images, which may also contain PHI. We provide a filter stage to identify these images.

After the de-identification process is completed, OBIA runs a quality control procedure. Some problematic images are isolated, such as images with a blank header or missing patient ID, corrupted images, other patient images mixed in, *etc*. Submitters can provide relevant information to repair the images or discard them entirely. Duplicate images are removed, leaving only one. Then we use TagSniffer (https://github.com/stlstevemoore/dicom-tag-sniffer) to generate a report for all images. All DICOM elements in the report are carefully reviewed to ensure that they are free of PHI, and certain values (*e.g.*, patient ID, study date) are modified as expected. In addition, we perform a visual inspection of each image series to ensure that no PHI is contained in the pixel values and that the images are visible and uncorrupted.

Data browse and retrieval

OBIA provides user-friendly web interfaces for data query and browsing. Users can browse data of interest by specifying non-image information (*e.g.*, age, gender, and disease) and/or the imaging data extracted from the DICOM header (*e.g.*, modality and anatomical site). Users can also search for data by entering the accession number. OBIA allows users to browse the basic information of collection (title, description, keywords, submitter, data accessibility, *etc.*), individual, study (study description, study age, study date, *etc.*), series (modality, anatomical site, series description, *etc.*), and view thumbnails of images.

OBIA also provides image retrieval functionality designed to find images similar to a query. Users can use this function by clicking the "Image Retrieval" button on the homepage. After uploading the image, the image retrieval model employs the hamming distance to return the top 30 images that are closest to the queried image, serving as the nearest neighbor images. Users have the option to click on these returned images and review their respective image metadata.

Data access and download

OBIA states that the data access policies are set by data submitters. There are two different types of data accessibility: open access and controlled access. Open access means that all data is public for global researchers if released, whereas controlled access means that data can be downloadable only after being authorized by the submitter. OBIA supports online data requests and reviews. Before applying, users need to register and log into OBIA via the Beijing Institute of Genomics (BIG) single sign-on (SSO; https://ngdc.cncb.ac.cn/sso/) system. Applicants shall provide their basic information, specify the scope of data usage, and promise not to attempt to restore the privacy information in the data. The download link will only be provided to the applicant if the data owner approves the request.

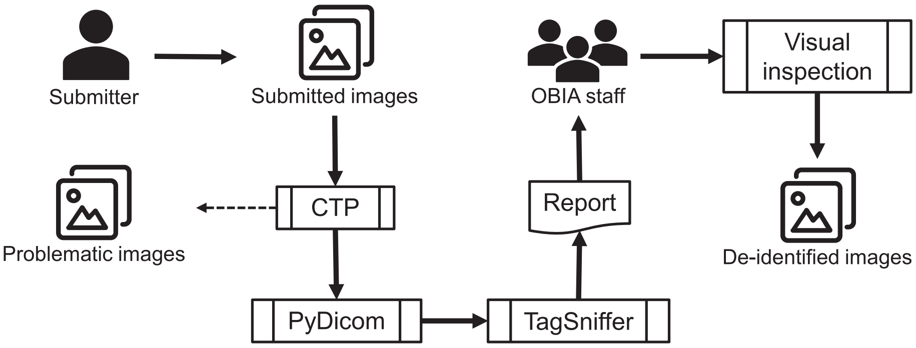

Figure 3 OBIA de-identification and quality control mechanism
Flowchart shows image submission, de-identification, and quality control steps. The de-identification steps include using CTP to process standard tags and PyDicom to handle private tags, and the problematic images will be isolated. The quality control steps include review of reports generated by TagSniffer and visual inspection of image pixels by OBIA staff. CTP, clinical trial processor.

Table 1 Number of individuals, studies, series, and images of each collection

Accession No.	Disease	No. of individuals	No. of studies	No. of series	No. of images
OBIA0001	Endometrial cancer	316	1587	8941	779,171
OBIA0002	Ovarian cancer	202	1548	7151	548,174
OBIA0003	Cervical cancer	117	675	3943	307,101
OBIA0004	Endometrial cancer	302	326	4666	303,863

Note: The data statistics are up to September 2023.

Data submission

OBIA accepts submissions of biomedical imaging data in DICOM format from clinical or specific research projects. To create a submission, users need to log in, fill in the basic information about the collection, and email the necessary clinical information. Image data will be transferred to OBIA offline after de-identification. Users can either use the de-identification process recommended by OBIA or apply their own methods to de-identify the image. Quality control will be conducted for all the data. OBIA will assign a unique accession number to each collection, individual, and study and arrange images into a standard organization. In order to ensure the security of submitted data, backup copies will be stored on physically separate disks. Finally, metadata will be extracted from the image file headers and stored in the database to support data queries.

Data statistics

As of September 2023, OBIA has housed a total of 937 individuals, 4136 studies, 24,701 series, and 1,938,309 images, covering 9 modalities and 30 anatomical sites. Representative imaging modalities are CT, magnetic resonance (MR), and digital radiography (DX) (Table S1). Anatomical sites include the abdomen, breast, chest, head, liver, and pelvis (Table S2). The first batch of collections submitted to OBIA came from the Chinese People's Liberation Army (PLA) General Hospital, including imaging data of three major gynecological tumors: endometrial cancer, ovarian cancer, and cervical cancer. The data were divided into four collections, and **Table 1** shows the number of individuals, studies, series, and images of each collection. In addition, we collected associated clinical metadata, such as demographic data, medical history, family history, diagnosis, pathological types, and treatment methods.

Perspectives and concluding remarks

OBIA is a centralized repository of de-identified biomedical imaging data. Different from existing related databases, OBIA is characterized by publishing imaging data and clinical information from a wide range of imaging modalities in a common DICOM format. Algorithm developers can convert and format as needed, and clinicians and researchers can combine clinical data with images for further analysis. Unlike the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe, China has a Personal Information Protection Law (PIPL), a Data Security Law, and other regulations for medical records and health information, which include elements similar to HIPAA and GDPR. OBIA strictly follows Chinese legal regulations for data processing, quality control, and data sharing, especially for removing PHI. It also promotes data sharing among data submitters and users while ensuring compliance with these laws. As a core database resource within NGDC, OBIA is seamlessly integrated with BioProject and GSA-Human, facilitating the harmonious integration of imaging and genomic data to support multi-omics analysis. In essence, OBIA serves as a dependable platform for sharing clinically relevant imaging data among researchers from diverse institutions, effectively bridging the gap within China's biomedical imaging database landscape.

In the future, we will continue to upgrade the infrastructure of OBIA and increase security protection measures to realize long-term secure storage, management, and access to a large volume of images. At the same time, we will collect more types of biomedical imaging data and gradually increase the corresponding genomic data to expand our data resources. To facilitate data submission and ensure privacy security, we will further optimize the image de-identification process and explore the use of the machine learning-based optical character recognition (OCR) [33] method to remove PHI from image pixels. We will also improve quality control and background automatic review processes to speed up data submission. In compliance with applicable regulations and ethical norms, OBIA's goal is to preserve as much effective image metadata as possible to provide researchers with high-quality imaging data. Furthermore, we plan to develop more intuitive and interactive web interfaces according to users' needs, increase database functions, and integrate related online tools to help analyze biomedical images. In addition, we intend to optimize the image retrieval model to offer users more convenient and precise image retrieval services. Finally, we call for collaborators to collectively build OBIA, submit image data, break down data silos, catalyze new biomedical discoveries, and provide the possibility to create personalized treatments.

Ethical statement

The collection of human imaging data was approved by the Local Ethical Committees in the First Medical Center of the Chinese PLA General Hospital (Approval No. S2022-403). The written informed consent was obtained from the participating patients.

Data availability

OBIA is publicly available at https://ngdc.cncb.ac.cn/obia.

Competing interests

The authors have declared no competing interests.

CRediT authorship contribution statement

Enhui Jin: Investigation, Methodology, Software, Writing – original draft. **Dongli Zhao:** Resources. **Gangao Wu:** Investigation, Methodology, Software, Writing – original draft. **Junwei Zhu:** Software. **Zhonghuang Wang:** Methodology. **Zhiyao Wei:** Resources. **Sisi Zhang:** Methodology. **Anke Wang:** Software, Writing – original draft. **Bixia Tang:** Resources. **Xu Chen:** Resources. **Yanling Sun:** Investigation, Methodology, Writing – review & editing, Project administration. **Zhe Zhang:** Investigation, Methodology, Writing – review & editing. **Wenming Zhao:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition. **Yuanguang Meng:** Conceptualization, Methodology, Writing – review & editing. All authors have read and approved the final manuscript.

Acknowledgments

This work was supported by grants from the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB38050300), the Genomics Data Center Operation and Maintenance of Chinese Academy of Sciences (Grant No. CAS-WX2022SDC-XK05), and the Key Technology Talent Program of the Chinese Academy of Sciences, China.

Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gpb.2023.09.003.

ORCID

ORCID 0009-0000-9916-9508 (Enhui Jin)
ORCID 0000-0003-2316-5927 (Dongli Zhao)
ORCID 0000-0002-3036-5997 (Gangao Wu)
ORCID 0000-0003-4689-3513 (Junwei Zhu)
ORCID 0000-0002-1268-883X (Zhonghuang Wang)
ORCID 0000-0003-4156-3267 (Zhiyao Wei)
ORCID 0000-0002-3852-4796 (Sisi Zhang)
ORCID 0000-0002-2565-2334 (Anke Wang)
ORCID 0000-0002-9357-4411 (Bixia Tang)
ORCID 0000-0001-6102-1751 (Xu Chen)
ORCID 0000-0002-3175-3625 (Yanling Sun)
ORCID 0009-0007-7446-7278 (Zhe Zhang)
ORCID 0000-0002-4396-8287 (Wenming Zhao)
ORCID 0000-0002-4957-3999 (Yuanguang Meng)

References

[1] Wallyn J, Anton N, Akram S, Vandamme TF. Biomedical imaging: principles, technologies, clinical aspects, contrast agents, limitations and future trends in nanomedicines. Pharm Res 2019;36:78.

[2] Li M, Guo R, Zhang K, Lin Z, Yang F, Xu S, et al. Machine learning in electromagnetics with applications to biomedical imaging: a review. IEEE Antennas Propag Mag 2021;63:39–51.

[3] Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical image analysis using convolutional neural networks: a review. J Med Syst 2018;42:226.

[4] Moody A. Perspective: the big picture. Nature 2013;502:S95.

[5] Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, et al. Preparing medical imaging data for machine learning. Radiology 2020;295:4–15.

[6] Sun C, Shrivastava A, Singh S, Gupta A. Revisiting unreasonable effectiveness of data in deep learning era. IEEE Int Conf Comput Vis 2017:843–52.

[7] Baughan N, Whitney H, Drukker K, Sahiner B, Hu TT, Hyun KJG, et al. Sequestration of imaging studies in MIDRC: a multi-institutional data commons. Proc SPIE 2022;12035:91–8.

[8] Crawford KL, Neu SC, Toga AW. The image and data archive at the laboratory of neuro imaging. Neuroimage 2016;124:1080–3.

[9] Kennedy DN, Haselgrove C, Riehl J, Preuss N, Buccigrossi R. The NITRC image repository. Neuroimage 2016;124:1069–73.

[10] Thompson HJ, Vavilala MS, Rivara FP. Common data elements and federal interagency traumatic brain injury research informatics system for TBI research. Annu Rev Nurs Res 2015;33:1–11.

[11] Markiewicz CJ, Gorgolewski KJ, Feingold F, Blair R, Halchenko YO, Miller E, et al. The OpenNeuro resource for sharing of neuroscience data. Elife 2021;10:e71774.

[12] Lee SM, Majumder MA. National institutes of mental health data archive: privacy, consent, and diversity considerations and options for improvement. AJOB Neurosci 2022;13:3–9.

[13] Prior FW, Clark K, Commean P, Freymann J, Jaffe C, Kirby J, et al. TCIA: an information resource to enable open science. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:1282–5.

[14] Fedorov A, Longabaugh WJR, Pot D, Clunie DA, Pieper S, Aerts H, et al. NCI imaging data commons. Cancer Res 2021;81:4188–93.

[15] Halling-Brown MD, Warren LM, Ward D, Lewis E, Mackenzie A, Wallis MG, et al. OPTIMAM mammography image database: a large-scale resource of mammography images and clinical data. Radiol Artif Intell 2021;3:e200103.

[16] Moura DC, Guevara Lopez MA. An evaluation of image descriptors combined with clinical data for breast cancer diagnosis. Int J Comput Assist Radiol Surg 2013;8:561–74.

[17] Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci 2007;19:1498–507.

[18] Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, et al. Video-based AI for beat-to-beat assessment of cardiac function. Nature 2020;580:252–6.

[19] Leclerc S, Smistad E, Pedrosa J, Ostvik A, Cervenansky F, Espinosa F, et al. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans Med Imaging 2019;38:2198–210.

[20] Wang XS, Peng YF, Lu L, Lu ZY, Bagheri M, Summers RM. ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. Proc IEEE Conf Comput Vis Pattern Recognit 2017;2017:3462–71.

[21] Guo S. DPN: detail-preserving network with high resolution representation for efficient segmentation of retinal vessels. J Ambient Intell Humaniz Comput 2023;14:5689–702.

[22] Ning WS, Lei SJ, Yang JJ, Cao YK, Jiang PR, Yang QQ, et al. Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning. Nat Biomed Eng 2020;4:1197–207.

[23] CNCB-NGDC Members and Partners. Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2023. Nucleic Acids Res 2023;51:D18–28.

[24] Lindeberg T. Scale invariant feature transform. Scholarpedia J 2012;7:10491.

[25] Pietikäinen MJS. Local binary patterns. Scholarpedia J 2010;5:9775.

[26] Dalal N, Triggs B. Histograms of oriented gradients for human detection. IEEE Comput Soc Conf Comput Vis Pattern Recognit 2005;1:886–93.

[27] Qayyum A, Anwar SM, Awais M, Majid M. Medical image retrieval using deep convolutional neural network. Neurocomputing 2017;266:8–20.

[28] Fang J, Fu H, Liu J. Deep triplet hashing network for case-based medical image retrieval. Med Image Anal 2021;69:101981.

[29] Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. Proc Mach Learn Res 2019;97:6105–14.

[30] Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, et al. GSA: Genome Sequence Archive. Genomics Proteomics Bioinformatics 2017;15:14–8.

[31] Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, et al. The Genome Sequence Archive Family: toward explosive data growth and diverse data types. Genomics Proteomics Bioinformatics 2021;19:578–83.

[32] Moore SM, Maffitt DR, Smith KE, Kirby JS, Clark KW, Freymann JB, et al. De-identification of medical images with retention of scientific research value. Radiographics 2015;35:727–35.

[33] Monteiro E, Costa C, Oliveira JL. A de-identification pipeline for ultrasound medical images in dicom format. J Med Syst 2017;41:89.