scientific data

OPEN

DATA DESCRIPTOR

An annotated heterogeneous ultrasound database

Yuezhe Yang[1,7], Yonglin Chen[2,7], Xingbo Dong[1]✉, Junning Zhang[3], Chihui Long[4], Zhe Jin[1] & Yong Dai[5,6]

Ultrasound is a primary diagnostic tool commonly used to evaluate internal body structures, including organs, blood vessels, the musculoskeletal system, and fetal development. Due to challenges such as operator dependence, noise, limited field of view, difficulty in imaging through bone and air, and variability across different systems, diagnosing abnormalities in ultrasound images is particularly challenging for less experienced clinicians. The development of artificial intelligence (AI) technology could assist in the diagnosis of ultrasound images. However, many databases are created using a single device type and collection site, limiting the generalizability of machine learning models. Therefore, we have collected a large, publicly accessible ultrasound challenge database that is intended to significantly enhance the performance of AI-assisted ultrasound diagnosis. This database is derived from publicly available data on the Internet and comprises a total of 1,833 distinct ultrasound data. It includes 13 different ultrasound image anomalies, and all data have been anonymized. Our data-sharing program aims to support benchmark testing of ultrasound disease diagnosis in multi-center environments.

Background & Summary

Medical ultrasound, particularly ultrasound imaging, has been a well-established field of research and clinical practice since its widespread adoption in the 1960s. It plays a vital role in modern medicine, with billions of examinations performed annually, especially during pregnancy[1]. Its popularity stems from being non-invasive, cost-effective, and capable of providing real-time feedback.

However, there are challenges in ultrasound technology that restrict clinicians' capacity to diagnose diseases precisely under complex circumstances. A major obstacle in ultrasound imaging is the suboptimal image quality, which is frequently marred by noise, visual artifacts, and poor contrast. Variations in operator skills and device setups, like transducer choice and equipment config, also cause image differences and inaccuracies. Additional factors such as respiratory motion, subcutaneous fat in obese patients, obtaining standard sections, and various organ sizes complicate the process[2]. Advanced automated ultrasound image analysis methods are thereby indispensable for attaining more objective and accurate diagnosis, evaluation, and image-guided interventions in the realm of ultrasound clinical applications[3].

Nevertheless, despite the widespread use of medical ultrasound, its image analysis techniques lag behind those of CT and MRI. Although machine learning offers promising improvements in ultrasound image analysis for computer-aided diagnosis, most machine learning research in ultrasound image analysis relies on small databases, typically a few hundred to a few thousand images[4,5]. This limited data size hinders the development of machine learning models and further limits the robustness of developed models or computer-aided diagnoses.

Additionally, many ultrasound databases are generated from a single device type and collected at a single site[6], which limits the diversity of the data. This lack of diversity also diminishes the effectiveness of machine learning models across different environments, making them less reliable in varied clinical settings. Furthermore, much of the current research in ultrasound imaging concentrates on specific tasks like segmentation or diagnosis[7]. This narrow focus means that models are often designed for single-task excellence but may face challenges when integrated into clinical diagnostic systems that need to address comprehensive tasks[6]. Therefore, to advance the

[1]Anhui Provincial International Joint Research Center for Advanced Technology in Medical Imaging, School of Artificial Intelligence, Anhui University, Hefei, 230601, China. [2]School of Electronic and Information Engineering, Anhui Jianzhu University, Hefei, 230601, China. [3]School of Public Health, Anhui University of Science and Technology, Huainan, 232001, China. [4]Department of Radiology, Wuhan Third Hospital/Tongren Hospital of Wuhan University, Wuhan, 430060, China. [5]School of Medicine, Anhui University of Science and Technology, Huainan, 232001, China. [6]The First Hospital, Anhui University of Science and Technology, Huainan, 232001, China. [7]These authors contributed equally: Yuezhe Yang, Yonglin Chen. ✉e-mail: xingbo.dong@ahu.edu.cn

Dataset Name	Task	Number	Data Format	Data Type
Hou *et al.*[10]	Thyroid Nodule	842	Image Set	Clinical Cases
PSFHS[11]	Segmentation	1124	Image/Image Set	Clinical Cases
BUSI[12]	Segmentation	780	Image	Clinical Cases
POCUS[13]	COVID Detection	261	Video/Image	Various Public Sources
Ours	Diagnosis	1833	Video/Image Set	Public Online

Table 1. Comparison of Ultrasound Datasets.

overall progress of machine learning applications in ultrasound imaging, there is an urgent need to develop a heterogeneous public dataset that can be thoroughly explored.

Currently, no such large size and heterogeneous publicly available ultrasound dataset exists[8]. Fortunately, with the development of social media, an increasing number of physicians publicly share anonymized data online to promote discussion and education. Such data is frequently particularly challenging or related to rare diseases, which renders it highly valuable for machine learning. Additionally, the majority of the data is in video format, providing more abundant semantic details, such as crucial temporal information for cardiac examinations[9].

Therefore, with the creator's permission, we extensively gathered a large amount of publicly available ultrasound data online and conducted analysis, screening, and diagnosis. All the data we obtained can only be used for research purposes, with no other uses permitted. We included the URLs of all data sources in the data disclosure section to facilitate easy access. This dataset primarily comes from 31 video creators affiliated with hospitals across 16 provinces in China. The data includes ultrasound videos/images of various organs, such as the heart, thyroid, abdomen, kidneys, and lungs. Online comments, video titles, and content associated with each video were also collected and initially used to generate labels. Subsequently, professional doctors were invited for secondary verification. After excluding all disputed and clearly unusable data, we obtained a dataset with 1,833 valid data samples and 13 disease categories.

Our dataset has several notable features and advantages compared to existing datasets like those of Hou *et al.*[10], PSFHS[11], BUSI[12], and POCUS[13] presented in Table 1. Firstly, in terms of sample size, our dataset consists of 1,833 samples. It surpasses the sample sizes of all the previously reported datasets by a significant margin. This larger sample size allows it to offer a more extensive range of clinical coverage, which is crucial for comprehensive analysis and model training. Secondly, regarding data types, while most resources concentrate only on static images, our dataset stands out as it includes both video and image data. This enables us to conduct richer temporal analysis, as the video data captures changes over time. Moreover, it helps enhance the robustness of the models trained with this dataset, making them more adaptable and reliable in various scenarios. Thirdly, considering the data source, our database comes from public online platforms. This acquisition method brings greater diversity and reflects real-world complexity. The heterogeneous nature of the data has the potential to foster the development of disease diagnosis models that can span different devices and data sources. It also facilitates the training of an ultrasound imaging-assisted diagnostic model that is suitable for diverse clinical environments. Furthermore, the raw data we have released includes a wealth of information, such as patient clinical symptoms, providing significant value for both deep learning models and clinical research. Finally, to verify the robustness of our dataset, we carried out tests using a well-known machine learning model.

Methods

Ethical Approval. All data used in this study were obtained from publicly available sources on the Internet and were anonymized. The acquisition and use of the data were communicated to the creators, and their permission was obtained. All data were securely managed and de-identified to ensure the rights and privacy of the participants. This dataset is secondary and contains no information that could identify individuals.

The study was approved by the Biomedical Ethics Committee of Anhui University (approval no. BECAHU-2024-023) and conducted in accordance with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. As the study utilized publicly available data without involving direct interaction or identifiable personal information, informed consent was waived by the Institutional Review Board.

Data Collection. Our data was collected from publicly available video/image-sets data on the Internet, primarily from the Douyin (DY) platform: www.douyin.com. DY is a short video and music video-sharing mobile application, which was launched in the fall of 2016. It allows users to create and browse short video clips ranging from 15 seconds to one minute in duration. DY boasts over 500 million global monthly active users and more than 250 million daily active users in China[14]. Therefore, all these videos are voluntarily uploaded by the creators in compliance with the regulations of the DY platform. The creators in our study of these videos mainly include researchers and medical imaging professionals who upload the data to exchange knowledge and share case studies. Consequently, the uploaded video data are complex and challenging, including multiple concurrent symptoms and rare diseases, making accurate diagnosis difficult.

Based on Article 10.2 of the DY platform's user service agreement, which states that "The intellectual property rights of the content you upload or publish through Douyin belong to you or the original copyright holder", it implies that for the legal usage of such data, permission from the creators is the sole requirement. We have taken the necessary step of notifying each of the 31 creators and have successfully obtained their permission. Finally, a total of 1,833 video clips and image sets have been collected, and the URLs of all the data sources have been properly recorded.

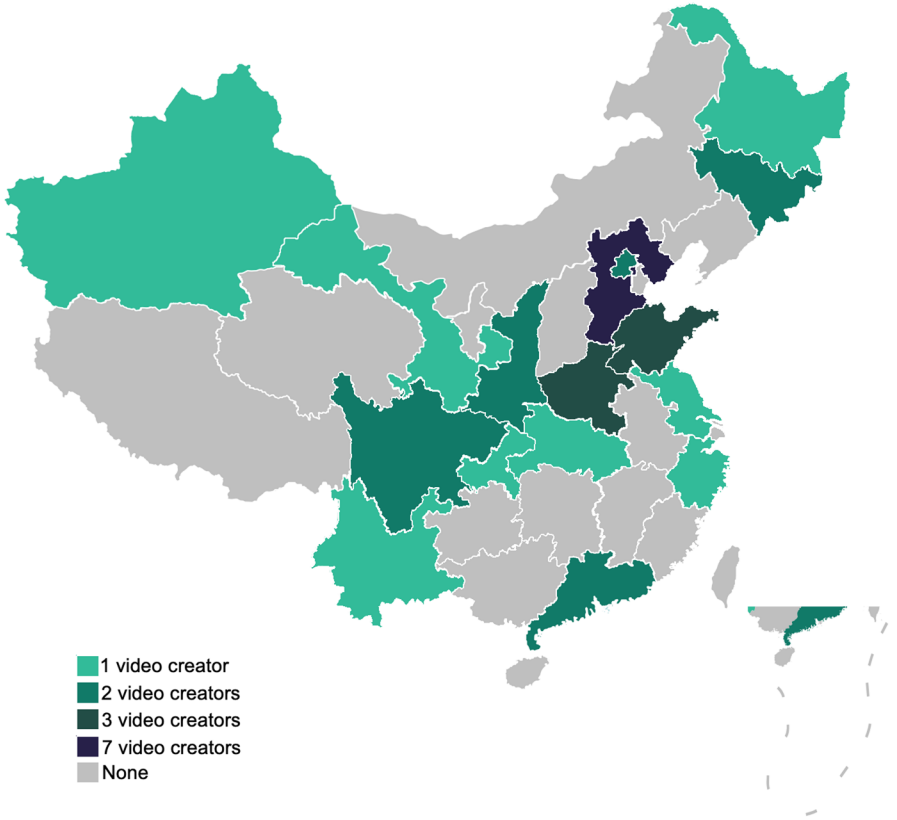

Fig. 1 The Chinese provinces where the dataset is sourced.

Data Characteristics. Our dataset exhibits significant heterogeneity, which is crucial for current machine learning applications[15]. Notably, the 31 video creators in our dataset represent different hospitals across 16 provinces in China, as shown in Fig. 1. Consequently, the ultrasound equipment used varies, and the imaging outcomes are influenced by the operational practices of different physicians during ultrasound examinations. This diversity enhances the generalization capability of deep learning models, overcoming the limitations of data sourced from a single hospital[16]. Moreover, the subjects of the ultrasound examinations are diverse, including newborns, adolescents, pregnant women, and the elderly. The anatomical areas examined also vary, with common sites such as the heart, thyroid, and kidneys being frequently scanned. Importantly, all the ultrasound videos in our dataset contain pathological findings, excluding those from routine check-ups. These pathological images are especially valuable for machine learning, as even a small number of such images can lead to excellent results in disease diagnosis through techniques like transfer learning, thereby providing stronger support for diagnostic assistance[17]. Additionally, unlike most existing medical ultrasound datasets, ours includes a comprehensive set of video/image data. Videos provide rich temporal data, capturing dynamic changes, such as cardiac blood flow throughout the cardiac cycle. As the ultrasound probe is repositioned, the videos also reveal pathological features from different angles. These aspects are challenging to capture in image-only datasets but are crucial for deep learning models, making them extremely valuable[3].

Data Filtering. Since data obtained from publicly available online sources contain significant noise, rigorous data filtering is crucial[18]. The filtering process consists of three steps: **Initial Ultrasound Sorting, Manual Unusable Data Inspection** and **Disagreements Elimination in Annotation**.

Initial Ultrasound Sorting: We found that some creators upload a variety of content. They not only publish videos related to ultrasound examinations but also include daily life videos. Given our focus on ultrasound data, we excluded these daily life videos.

Manual Unusable Data Inspection: Subsequently, we carried out a manual inspection of the remaining videos by carefully reviewing them. The aim was to filter out unusable data. The videos that were excluded during this process had one or more of the following issues:

- They contained a substantial amount of content that had no connection to ultrasound examinations.
- It was difficult to identify symptoms from the information presented in them.
- The ultrasound images within them were extremely unclear, making it hard to extract useful details.
- They lacked proper and strict anonymization of patient information, which could raise ethical and privacy concerns.

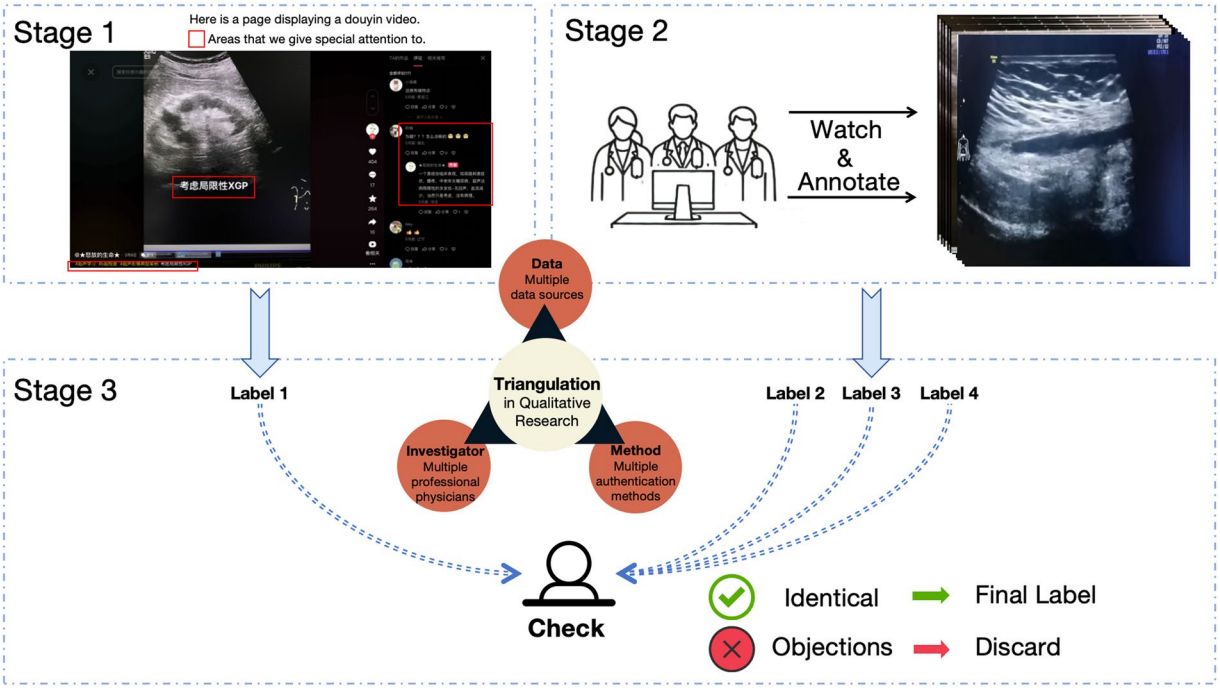

Fig. 2 A flowchart of the annotation process.

Disagreements Elimination in Annotation: During the annotation process, preliminary labels were first assigned to the data, followed by independent review and labeling by three professional clinicians. The following types of data were discarded during this process:

- At least one clinician considered the data ambiguous or unusable.
- At least two clinicians were unable to reach a consensus on the classification or characteristics of a particular piece of data.
- The clinicians' labels did not align with the preliminary labels assigned during the initial annotation phase.
- After implementing this comprehensive and rigorous filtering process, we were successful in obtaining 1,833 pieces of valid data. This valid dataset is composed of 140 image sets and 1,693 videos.

Data Annotation. Proper annotation is crucial to effectively applying the data to machine learning. An overview of our data annotation process can be seen in the Fig. 2. The data annotation process can be divided into three main stages.

In **Stage 1**, we conducted an in-depth analysis on the information available from the videos and the online platform, focusing on three key aspects:

- Comments on the videos were analyzed. Since these videos were posted on a public platform, the comment sections were open to discussion and analysis by online professional physicians and learners. All comments were recorded, as they often contained insights from medical professionals, especially those made or endorsed by the video creators, who directly interact with the patients, making their comments more reliable.
- Descriptions accompanying each video were documented. On DY platform, these descriptions often serve as video titles and frequently include details about pathology or clinical symptoms, which aids in symptom identification and is valuable for machine learning.
- The content of the videos was reviewed. These videos usually contained textual symptom information that was crucial for accurately assigning data labels.
- By combining these three sources, each sample was carefully assigned one reliable label.

In **Stage 2**, three experienced physicians from two different institutions independently annotated the data and provided feedback on whether to include or exclude the data, as well as the corresponding labels. Each physician evaluated the data based on two key questions: **1)** Is this data valuable? **2)** What should the label for this data be? During the annotation process, multiple labels were allowed.

In **Stage 3**, to ensure the validity of our labels, we applied the triangulation method[19]. Any data deemed non-valuable by any physician, or any data for which at least one physician suggested a different label, was excluded. This process served as a quality control measure to maintain consistency and accuracy in the annotations. After this rigorous validation, we categorized the 1,833 data samples into 13 distinct symptoms: Tumor, Cyst, Inflammation, Stone, Nodule, Injury, Calcification, Occupancy, Hernia, Vascular, Polyp, Ectopic, and Anomalies.

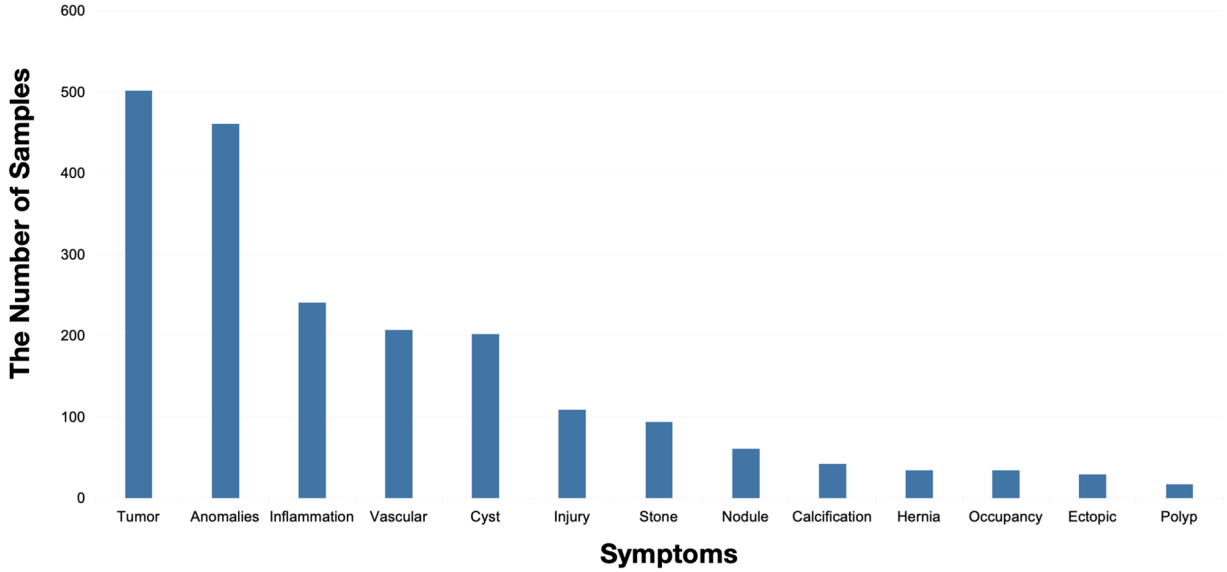

Fig. 3 The data distribution of our ultrasound dataset.

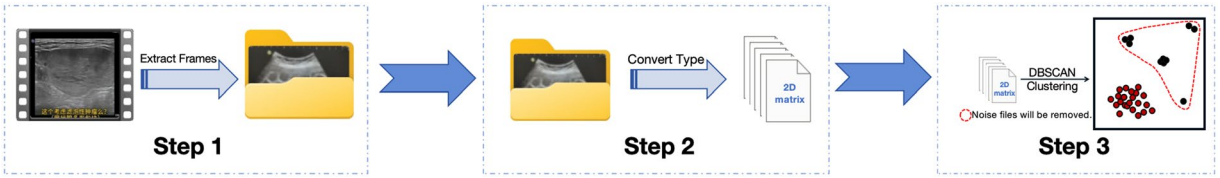

Fig. 4 A flowchart of the data preprocessing.

It is important to note that one sample may exhibit multiple symptoms. For example, Tumor, Cyst, and Stone frequently occur together. In cases where multiple labels were applied to the dataset, we adopted a multi-label strategy. Specifically, if imaging clearly showed the presence of both Tumor and Stone, Tumor was prioritized in the initial label assignment. Only when all clinicians agreed that both conditions were independently present and assigned individual labels, was the sample classified as multi-labeled.

On the other hand, appendicitis often involves the co-occurrence of Inflammation and Stone, with fecaliths commonly regarded as complications of Inflammation. In such cases, the more prominent symptom, Inflammation, was selected as the final label. Appendicitis was treated as a single-label condition because Inflammation is the primary manifestation, while accompanying conditions like fecaliths are considered secondary and not independently present. As a result, such cases were categorized under the single-label strategy.

Furthermore, we observed that the occurrence frequency of the 13 different symptoms in this dataset follows a long-tail distribution, which is in line with expectations. The details are presented in Fig. 3. It is worth noting that although this distribution is commonly seen in nature, it presents a significant challenge for deep learning models[20]. In a nutshell, through the above steps, we ensure that the labels have as little noise as possible. Even if there is some label noise, deep learning models can adapt to such noise[21,22].

Data Preprocessing. Data preprocessing is crucial for deep learning[23]. Our data preprocessing process mainly consists of three major steps, which are illustrated in Fig. 4.

First, we applied a frame extraction approach to the video data[24]. By extracting one frame every ten frames as key frames, we reduced the data volume and converted the videos into image sets, thereby lowering computational costs and reducing model complexity. We also included the original data, as the reinforcement learning method proposed by Huang *et al.*[25] for obtaining key frames can optimize model performance and reduce complexity. Therefore, thoroughly extracting valuable information from video data proved highly beneficial.

Second, in a complete video, there are often segments that display content unrelated to ultrasound or that contain data that cannot be directly utilized. To address this, we conducted Density-Based Spatial Clustering of Applications with Noise (DBSCAN) analysis on all data within each image set[26]. To more effectively eliminate noise, we considered a cluster to be valid only if the number of data samples in that cluster exceeded half of the total number of images in the set. In other words, there was at most one valid cluster, corresponding to the ultrasound images in each image set.

Finally, after unifying the data format and filtering out noisy data, we resized all images to 224 $\times$ 224 pixels and uniformly converted them into `.jpg` files.

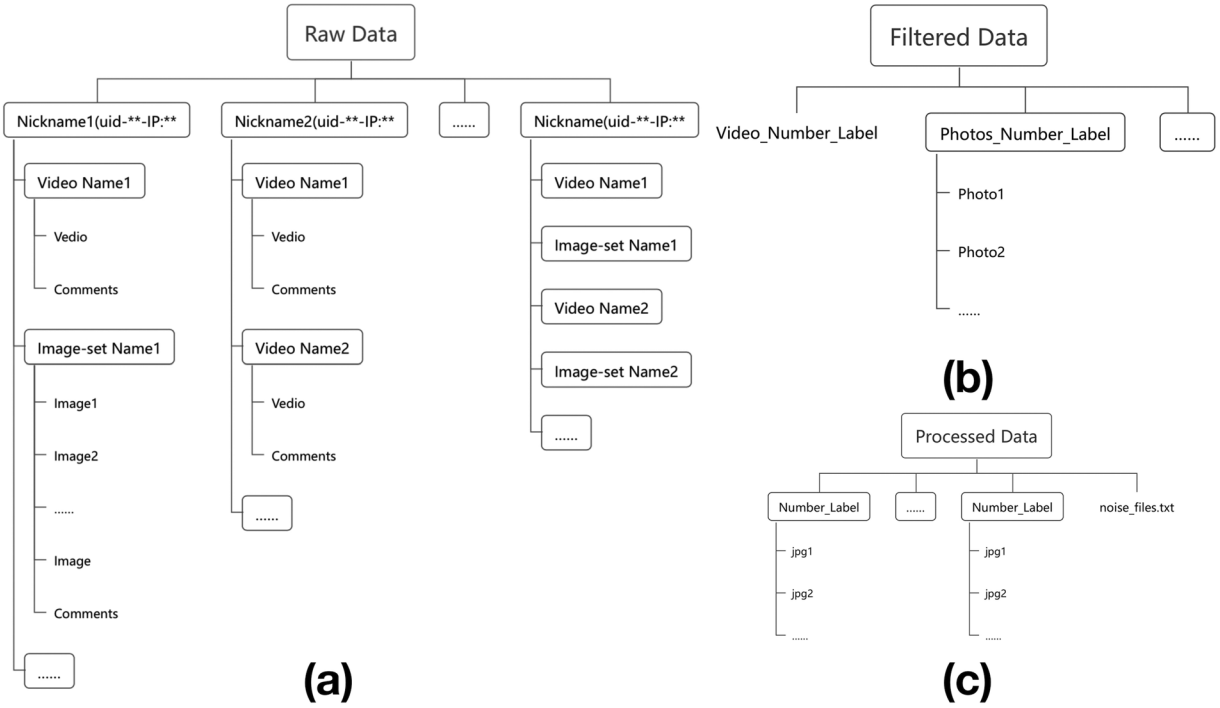

Fig. 5 The structure of the data is visually represented. (**a**) Structure of raw data. (**b**) Structure of filtered data. (**c**) Structure of processed data.

Data Records

All raw data, filtered data, and processed data are provided at Figshare[27].

Raw Data. The raw data consists of the initial, unprocessed data, which includes videos or images before being filtered. The comments associated with each video are included as well. The structure of this data can be seen in Fig. 5(a). This raw data mainly consists of three levels of folders. The first-level folders are named according to the 31 different creators. The second-level folders are named after each video from each creator. The third level contains two or more files, which consist of either a video or multiple images and a table file. The table file contains the online comments associated with each video. Indented entries in the table denote responses to comments. Additionally, whether the creator liked the comment was recorded, and creator comments were highlighted for easy identification. To enhance data robustness, we have renamed all files using encoded formats and saved the mapping between the original and corresponding paths in `rename.xlsx`.

Filtered Data. The naming and formatting of the filtered data have been standardized. The folder structure of this data is illustrated in Fig. 5(b). The main folder contains 1,833 files, with their naming format standardized as `Photos/Video_Number_Label`.

Processed Data. The folder structure of processed data is illustrated in Fig. 5(c). The structure consists of two levels of folders. At the first level, there are 1,833 data samples with a standardized naming format as `Number_Label`. Additionally, the absolute paths of noise data identified through clustering analysis are listed in the `noise_files.txt` file. To facilitate the usage of this dataset, we have also provided code on GitHub for processing different file paths. The second-level folders contain varying numbers of `.jpg` files, each storing resized images in a uniform size, which can be directly used for analysis.

Technical Validation

Data Collection and Annotation. During the data collection and annotation process, we received cross-validation support from medical professionals at Wuhan Third Hospital/Tongren Hospital of Wuhan University, the School of Medicine at Anhui University of Science and Technology, and the First Hospital of Anhui University of Science and Technology. The detailed cross-validation annotation is detailed in the Data annotation section.

Utility Validation for Deep Learning. To illustrate the potential utility of our heterogeneous dataset in deep learning-based ultrasound video diagnosis tasks, we employed the Recurrent Vision Transformer (RViT) model[28], originally designed for video processing task, for disease diagnosis based on the Ultrasound data.

We quantitatively assessed diagnosis performance using three metrics: Top-1 accuracy (ACC), Top-3 accuracy, and Top-5 accuracy. Given the multi-symptom nature of many ultrasound images, we prioritized Top-3

Pretrain	Epoch	*Top1*-ACC	*Top3*-ACC	*Top5*-ACC
✓	10	30.44	69.38	90.00
✗	65	34.14	68.48	85.38

Table 2. We used RViT for case study on our ultrasound database. The pretrain was conducted on the Jester dataset[35].

accuracy as the primary metric, as it aligns with the characteristics of our data and offers a balanced perspective on the network's predictive performance.

Because RViT is limited to single-label classification, we consistently assigned each image a label corresponding to its most prominent feature. This approach enables a fair and consistent evaluation, particularly when using the Top-3 ACC metric, which we hypothesize provides a more robust assessment of prediction outcomes under the multi-symptom conditions present in our dataset. The experimental results Table 2 indicate that, regardless of whether pre-trained weights were utilized, the Top-3 ACC metric exceeded 68.49%, albeit without achieving exceptionally high accuracy. These findings underscore the significant challenges posed by the heterogeneous nature of our dataset and the complexity of applying such data to disease diagnosis tasks. Furthermore, the results reveal notable differences between the diagnosis of intricate ultrasound pathological images and traditional human action recognition tasks.

This case study highlights the challenges and opportunities associated with developing diagnostic models for ultrasound medical imaging. While not a comprehensive evaluation, it provides a meaningful baseline and emphasizes the potential of our dataset to inspire future advancements in machine learning for medical ultrasound diagnostics.

Usage Notes

Deep learning methods have witnessed remarkable progress in handling medical image data. There are several effective techniques that have been developed, including key frame extraction via reinforcement learning[29], more efficient data annotation using active learning[30], learning from noisy data[31], few-shot learning[32], video temporal analysis within a cardiac cycle[33], and enhancing model performance through the incorporation of clinical information[34]. To facilitate a comprehensive exploration of the dataset value, we have supplied the raw data.

The data is sourced from public collections and has been published on social media, with the video creators retaining ownership of the data. Meanwhile, the use of the dataset must comply with the CC BY license. Additionally, all URLs of the video creators mentioned in the database are publicly available at: https://github.com/Bean-Young/AHU-Database/tree/URLs. If there is a need to use data beyond what we have provided, permission must be obtained from the video creators via direct communication. The data presented in this manuscript also forms part of the 2024 iFLYTEK A.I. Developer Competition: https://challenge.xfyun.cn/competition.

Code availability

All the code used in our processing workflow and validation is publicly available at: https://github.com/Bean-Young/AHU-Database. The project documentation provides detailed instructions on how to use the code, facilitating the easy reproduction of our results.

Received: 11 September 2024; Accepted: 13 January 2025;
Published online: 25 January 2025

References

1. Seo, J. & Kim, Y. Ultrasound imaging and beyond: recent advances in medical ultrasound. *Biomed. Eng. Lett.* **7**, 57–58 (2017).
2. Avola, D., Cinque, L., Fagioli, A., Foresti, G. & Mecca, A. Ultrasound medical imaging techniques: a survey. *ACM Comput. Surv.* **54**, 1–38 (2022).
3. Liu, S. *et al.* Deep learning in medical ultrasound analysis: a review. *Engineering* **5**, 261–275 (2019).
4. Yap, M. H. *et al.* Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J. Biomed. Health Inform.* **22**, 1218–1226 (2018).
5. Leclerc, S. *et al.* Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE Trans. Med. Imaging* **38**, 2198–2210 (2019).
6. Brattain, L. J., Telfer, B. A., Dhyani, M., Grajo, J. R. & Samir, A. E. Machine learning for medical ultrasound: status, methods, and future opportunities. *Abdom. Radiol.* **43**, 786–799 (2018).
7. Fiorentino, M. C., Villani, F. P., Di Cosmo, M., Frontoni, E. & Moccia, S. A review on deep-learning algorithms for fetal ultrasound-image analysis. *Med. Image Anal.* **83**, 102629 (2023).
8. Saw, S. N. & Ng, K. H. Current challenges of implementing artificial intelligence in medical imaging. *Physica Medica* **100**, 12–17 (2022).
9. Cikes, M., Tong, L., Sutherland, G. R. & D'hooge, J. Ultrafast cardiac ultrasound imaging. *JACC Cardiovasc. Imaging* **7**, 812–823 (2014).
10. Hou, X. *et al.* An ultrasonography of thyroid nodules dataset with pathological diagnosis annotation for deep learning. *Sci. Data* **11**, 1272 (2024).
11. Chen, G., Bai, J., Ou, Z., Lu, Y. & Wang, H. PSFHS: intrapartum ultrasound image dataset for AI-based segmentation of pubic symphysis and fetal head. *Sci. Data* **11**, 436 (2024).
12. Sulaiman, A. *et al.* Attention based UNet model for breast cancer segmentation using BUSI dataset. *Sci. Rep.* **14**, 22422 (2024).
13. Born, J. *et al.* POCOVID-Net: automatic detection of COVID-19 from a new lung ultrasound imaging dataset (POCUS). Preprint at https://arxiv.org/abs/2004.12084 (2020).
14. Lu, X. & Lu, Z. Fifteen seconds of fame: a qualitative study of Douyin, a short video-sharing mobile application in China. *Lecture Notes Comput. Sci.* **11578**, 233–244 (2019).
15. Kumar, G. *et al.* Data harmonization for heterogeneous datasets: a systematic literature review. *Appl. Sci.* **11**, 8275 (2021).

16. Di Mario, C. Clinical application and image interpretation in intracoronary ultrasound. *Eur. Heart J.* **19**, 207–229 (1998).
17. An, G., Akiba, M., Omodaka, K., Nakazawa, T. & Yokota, H. Hierarchical deep learning models using transfer learning for disease detection and classification based on small number of medical images. *Sci. Rep.* **11**, 4250 (2021).
18. Qian, Y. *et al*. TikTokActions: A TikTok-Derived Video Dataset for Human Action Recognition. *arXiv Preprint* at https://arxiv.org/abs/2402.08875 (2024).
19. Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J. & Neville, A. J. The use of triangulation in qualitative research. *Oncol. Nurs. Forum* **41**, 545–547 (2014).
20. Zhang, Y., Kang, B., Hooi, B., Yan, S. & Feng, J. Deep long-tailed learning: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 10795–10816 (2023).
21. Nettleton, D. F., Orriols-Puig, A. & Fornells, A. A study of the effect of different types of noise on the precision of supervised learning techniques. *Artif. Intell. Rev.* **33**, 275–306 (2010).
22. Karimi, D., Dou, H., Warfield, S. K. & Gholipour, A. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Med. Image Anal.* **65**, 101759 (2020).
23. Wang, S. *et al*. Advances in data preprocessing for biomedical data fusion: an overview of the methods, challenges, and prospects. *Inf. Fusion* **76**, 376–421 (2021).
24. Schultz, R. R. & Stevenson, R. L. Extraction of high-resolution frames from video sequences. *IEEE Trans. Image Process.* **5**, 996–1011 (1996).
25. Huang, R. *et al*. Extracting keyframes of breast ultrasound video using deep reinforcement learning. *Med. Image Anal.* **80**, 102490 (2022).
26. Rehman, S. U., Asghar, S., Fong, S. & Sarasvady, S. DBSCAN: past, present and future. *2014 Fifth Int. Conf. Appl. Digit. Inf. Web Technol.* 232-238 (2014).
27. Yang, Y-Z. *et al*. An ultrasound dataset in the wild for machine learning of disease classification. *figshare* https://doi.org/10.6084/m9.figshare.26889334 (2024).
28. Yang, J. *et al*. Recurring the transformer for video action recognition. *2022 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* 14043–14053 (2022).
29. Mo, C., Hu, K., Mei, S., Chen, Z. & Wang, Z. Keyframe extraction from motion capture sequences with graph-based deep reinforcement learning. *Proc. ACM Multimedia Conf.* 5194–5202 (2021).
30. Amin, S. U., Hussain, A., Kim, B. & Seo, S. Deep learning-based active learning technique for data annotation and improving the overall performance of classification models. *Expert Syst. Appl.* **228**, 120391 (2023).
31. Song, H., Kim, M., Park, D., Shin, Y. & Lee, J. G. Learning from noisy labels with deep neural networks: a survey. *IEEE Trans. Neural Netw. Learn. Syst.* **34**, 8135–8153 (2023).
32. Wang, Y., Yao, Q., Kwok, J. T. & Ni, L. M. Generalizing from a few examples: a survey on few-shot learning. *ACM Comput. Surv.* **53**, 1–34 (2021).
33. Ouyang, D. *et al*. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* **580**, 252–256 (2020).
34. Khader, F. *et al*. Multimodal deep learning for integrating chest radiographs and clinical parameters: a case for transformers. *Radiology* **309**, e230806 (2023).
35. Materzynska, J., Berger, G., Bax, I. & Memisevic, R. The Jester dataset: A large-scale video dataset of human gestures. *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops* **0**, 0–0 (2019).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 62306003).

Author contributions

Y.Y., Z.J., and X.D. conceived of the presented idea; J.Z., C.L., and Y.D. performed and confirmed the labels; Y.C. examined full data anonymization; Y.Y., X.D., and Y.C. collected data; Y.C. hosted the competition; Y.Y. wrote the initial draft; X.D., Y.C., J.Z., C.L., Z.J., and Y.D. completed the critical review and revision of the manuscript and datasets, as well as proofreading.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to X.D.