EchoCare

“聆音”超声大模型

TechReport

中国科学院香港创新研究院
人工智能与机器人创新中心
Centre for Artificial Intelligence and Robotics
Hong Kong Institute of Science & Innovation, Chinese Academy of Sciences

EchoCare: A Fully Open and Generalizable Foundation Model for Ultrasound Clinical Applications

Hongyuan Zhang[1], Yuheng Wu[1,2], Mingyang Zhao[3,4,*], Zhiwei Chen[1,5], Rebecca Li[6], Fei Zhu[1,*], Haohan Zhao[1,2], Xiaohua Yuan[7], Meng Yang[8], Chunli Qiu[9], Xiang Cong[9], Haiyan Chen[10], Lina Luan[11], Randolph H.L. Wong[12], Huai Liao[13], Colin A Graham[6], Shi Chang[7], Guowei Tao[9], Dong Yi[1], Zhen Lei[1,4,14], Nassir Navab[15], Sebastien Ourselin[16], Jiebo Luo[1,17], Hongbin Liu[1,14,16], Gaofeng Meng[1,4,14,*]

Abstract

The inherent safety and versatility of ultrasound imaging have made it widely accessible in modern clinical settings for disease diagnosis and health management. Artificial intelligence (AI) that can effectively learn ultrasound representations by integrating multi-source data holds significant promise for advancing clinical care. However, the scarcity of large labeled datasets in real-world clinical environments and the limited generalizability of task-specific models have hindered the development of generalizable clinical AI models for ultrasound applications. In this study, we present EchoCare, a novel ultrasound foundation model for generalist clinical use, developed via self-supervised learning on our curated, publicly available, large-scale unlabeled dataset EchoAtlas. EchoAtlas comprises 4.5 million ultrasound images, sourced from over 20 countries across 5 continents and acquired via a diverse range of distinct imaging devices, thus encompassing global cohorts that are multi-center, multi-device, and multi-ethnic. Unlike prior studies that adopt off-the-shelf vision foundation model architectures, we introduce a hierarchical classifier into EchoCare to enable joint learning of pixel-level and representation-level features, capturing both global anatomical contexts and local ultrasound characteristics. With minimal training, EchoCare outperforms state-of-the-art comparison models across 10 representative downstream ultrasound benchmarks of varying diagnostic difficulties, spanning disease diagnosis, lesion segmentation, organ detection, landmark prediction, quantitative regression, imaging enhancement and report generation. The code and pretrained model are publicly released, rendering EchoCare accessible for fine-tuning and local adaptation, supporting extensibility to additional applications. EchoCare provides a fully open and generalizable foundation model to boost the development of AI technologies for diverse clinical ultrasound applications.

[1]Center for Artificial Intelligence and Robotics, Hong Kong Institute of Science & Innovation, Chinese Academy of Sciences, Hong Kong, China; [2]City University of Hong Kong, Hong Kong, China; [3]State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China; [4]University of Chinese Academy of Sciences, Beijing, China; [5]Division of Electronic Engineering, Faculty of Engineering, The Chinese University of Hong Kong, Hong Kong, China; [6]Accident and Emergency Medicine Academic Unit, The Chinese University of Hong Kong, Hong Kong, China; [7]Xiangya Hospital Central South University, Changsha, China; [8]Hunan Frontline Medical Technology Co., Ltd, Changsha, China; [9]Qilu Hospital of Shandong University, Jinan, China; [10]Zhongshan Hospital of Fudan University, Shanghai, China; [11]Shanghai Geriatric Medical Center, Shanghai, China; [12]Division of Cardiothoracic Surgery, Department of Surgery, The Chinese University of Hong Kong, Hong Kong, China; [13]Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China; [14]State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China; [15]Computer Aided Medical Procedures, Technical University of Munich, Munich, Germany; [16]School of Biomedical Engineering & Imaging Sciences, King's College London, UK; [17]Department of Computer Science, University of Rochester, USA; *Corresponding Authors. Emails: zhaomingyang16@mails.ucas.ac.cn, {fei.zhu,gaofeng.meng}@cair-cas.org.hk,

Contents

1 Introduction

Ultrasound imaging stands as a cornerstone of modern medicine, celebrated for its unique combination of real-time assessment, cost-effectiveness, and inherent safety. This non-invasive and radiation-free modality allows for the dynamic visualization of physiological processes, securing its indispensable role in a wide range of clinical practices [1]. Despite these advantages, ultrasound diagnostic is heavily reliant on the skill of the sonographer and the specialized expertise required to interpret the complex, often subtle, visual information. This inherent complexity, coupled with the ubiquity and versatility of ultrasound, has spurred significant interest in leveraging artificial intelligence (AI) to advance its use. As ultrasound imaging expands to new anatomical regions and clinical applications, there is a growing demand for versatile and generalizable AI models that can adapt to diverse clinical tasks and organs with minimal reliance on new labeled data. Meeting this demand will not only broaden the application of ultrasound analysis but also accelerate the deployment of smart healthcare solutions, making high-quality diagnostics more accessible and efficient.

Recent advances in foundation models (FM) using self-supervised learning have opened new frontiers in medical AI [2, 3, 4, 5, 6, 7]. These models learn general-purpose feature representations directly from raw data, eliminating dependence on extensive expert annotations. Upon completion of pretraining, these models can be effectively adapted to a wide array of downstream clinical tasks with minimal or no additional fine-tuning. This paradigm represents a significant advantage over conventional medical AI approaches, which are typically limited to specific anatomical structures or require extensive retraining when adapted to each new clinical application. However, pretraining of foundation models requires large-scale and diverse datasets, making data acquisition and rigorous curation essential for developing clinically reliable and generalizable systems.

Building on the success of vision foundation models, researchers have started adapting these approaches to ultrasound imaging analysis [8, 9, 10]. Although initial results show promise, several critical challenges could limit their potential clinical impact. First, the scale of available ultrasound datasets remains relatively small, undermining the reliability of models for clinical deployment. Moreover, much of the pretraining data employed in previous studies is private, creating barriers to reproducibility, broader research and application. Second, current collections often focus on narrow anatomical regions, which is insufficient to fully capture the diversity of whole-body regions. This limitation restricts their utility in comprehensive clinical workflows. Third, most approaches rely on off-the-shelf vision foundation model frameworks [11], failing to systematically explore network architecture optimizations tailored to the morphological complexity and spatial hierarchies of anatomical structures. This oversight limits the model's ability to capture anatomical relationships across scales and organs during pre-training. Finally, existing research mainly focuses on a few downstream tasks such as image classification or segmentation, leaving open questions about model capabilities for more diverse clinical applications.

In this work, we introduce EchoCare, a novel foundation model for ultrasound images, accompanied by a systematic investigation of its utility across a diverse spectrum of clinical tasks. EchoCare is pre-trained on EchoAtlas, our newly curated large-scale and openly accessible dataset comprising 4.5 million ultrasound images. Collected from multi-center, multi-device, multi-modality, and multi-ethnic global sources, EchoAtlas ensures diverse data representation. EchoAtlas covers 9 major regions and 52 anatomical organs of the human body, supporting models pretrained on it to generalize effectively across comprehensive whole-body ultrasound clinical applications, as shown in Fig. 1. We have also optimized the architecture of the vision foundation model to better capture hierarchical anatomical structures, from broad ultrasound regions (*e.g.*, abdomen) to specific organs (*e.g.*, liver, kidney), enabling the model to mimic human-like clinical diagnostic reasoning. Extensive evaluations across eight categories of core ultrasound clinical tasks of varying diagnostic difficulties, such as lesion segmentation, organ detection, disease diagnosis, and quantitative regression, reveal that EchoCare significantly outperforms state-of-the-art general-domain foundation models, underscoring the critical need for ultrasound-specific models. Compared with leading ultrasound-focused foundation models, EchoCare also demonstrated superior performance, highlighting the advantages of pretraining on large, diverse data. We will release both EchoCare and the EchoAtlas to promote clinical AI development in ultrasound images upon publication.

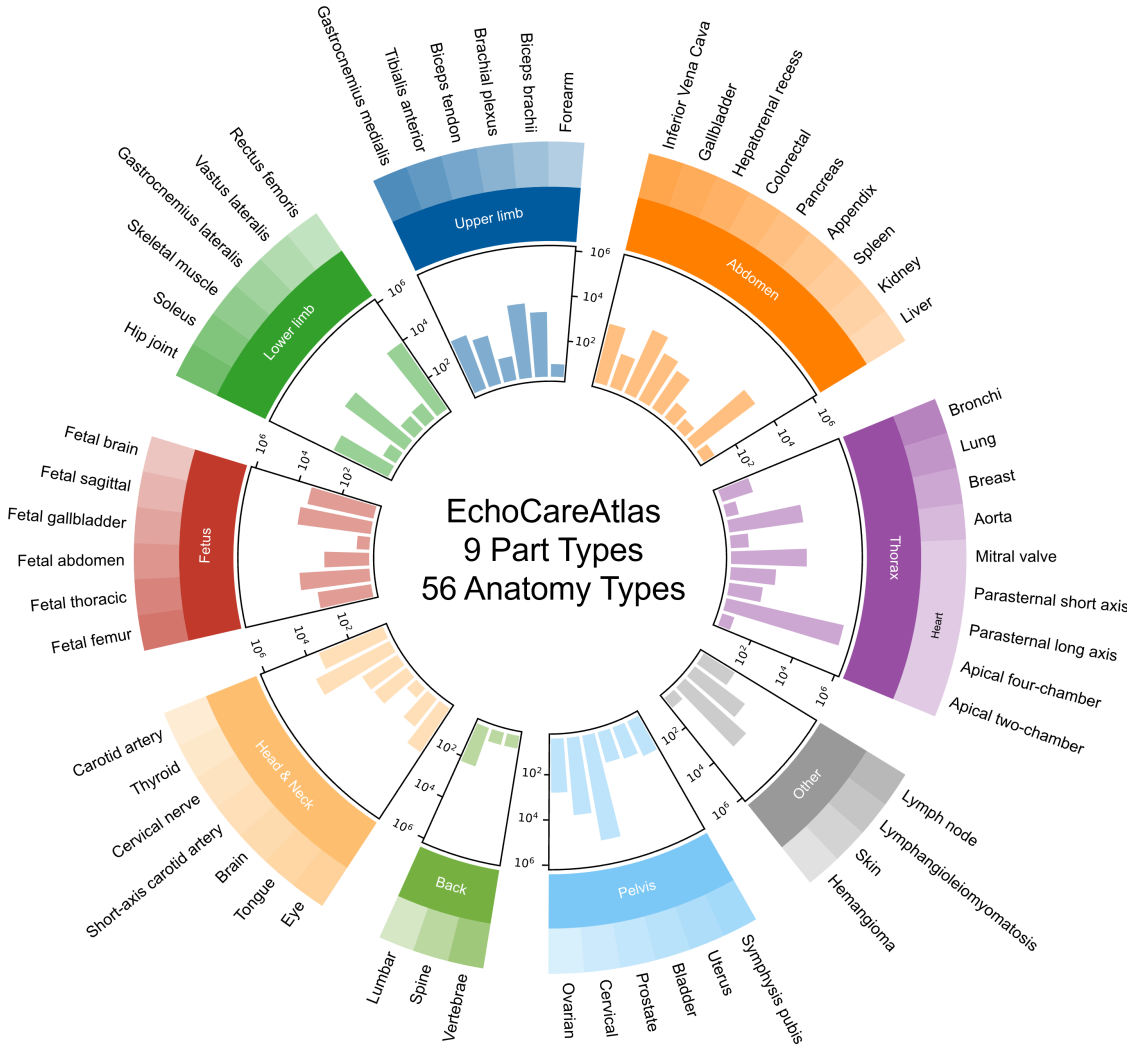

Figure 1: The constructed ontology shows a hierarchy of object types that are used to unify semantic concepts across datasets. Bar plots showing the number of images containing that object type.

2 EchoAtlas

We establish so far the largest public ultrasound image dataset EchoAtlas, integrating 138 ultrasound image datasets from over 20 countries and 5 continents (Fig. 2). Encompassing multiple body organs, scanning devices, imaging modalities, and racial backgrounds, the dataset is designed to ensure data diversity and enhance the generalization of pretrained models across diverse clinical applications. EchoAtlas adheres to rigorous cohort inclusion and exclusion protocols to ensure high quality, including manual removal of sensitive and non-ultrasound images, as well as text cleaning. Using a clinical anatomy system, we generated canonical categorical labels for each image. The dataset's ontology comprises eight representative clinical regions including head, chest, abdomen, limbs, back, fetus, dorsum, pelvis, and an "other" category, with a hierarchical structure spanning 52 meta-object types (*e.g.*, cardiac ventricle) to 56 specific anatomic types (*e.g.*, left cardiac ventricle), mirroring clinical diagnostic workflows. Moreover, an additional manual inspection was performed by randomly sampling 100 images from each class in EchoAtlas to validate correctness. In total,

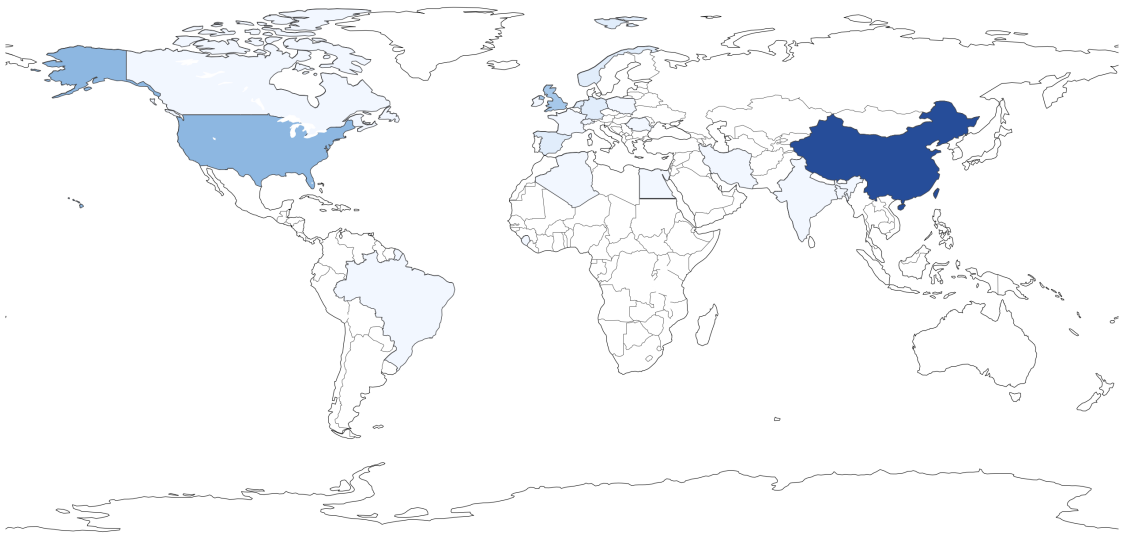

Figure 2: **Worldwide distribution of ultrasound imaging datasets in EchoAtlas**. It illustrates the multi-center, multi-ethnic data collection strategy supporting generalizable AI development for ultrasound clinical applications across diverse populations.

EchoAtlas comprises over 4.5 million distinct image-class tuples, spanning five imaging modalities (B-mode, CEUS, Dropper, M-mode, and Elastography), establishing it as a large-scale, diverse resource for clinical ultrasound care.

2.1 Data Curation

Our data curation process commenced with a systematic search of open academic repositories, including Zenodo [12], Mendeley [13], Stanford AIMI, Figshare, and data/code platforms such as Kaggle [14], GitHub [15], and medical challenge portals (*e.g.*, Grand Challenge [16]). All data collection was concluded by 1 March 2025. Using "ultrasound" as a keyword, we retrieved approximately 13,000 potential datasets for initial screening. The raw dataset underwent a series of exclusion steps (Fig. 3): 1) datasets were filtered to retain common file formats–including image files (*e.g.*, PNG, JPG, BMP) and compressed archives (*e.g.*, ZIP, RAR, TFRecord)–to confirm the presence of ultrasound images; 2) GPT-4o was utilized to extract direct download links from dataset descriptions in excluded text-only candidates; 3) preliminary deduplication was performed by comparing download URLs and computing image hash values for efficiency; and 4) manual curation was implemented to eliminate intra-organ redundancy through fine-grained filtering. Moreover, to mitigate intrinsic anatomical sampling biases and ensure comprehensive coverage, we strategically prioritized underrepresented anatomical structures through targeted efforts: submitting formal access requests to specialized repositories (*e.g.*, EchoNet-Dynamic) and directly contacting authors of ultrasound studies to procure supplementary datasets. Following our rigorous inclusion-exclusion pipeline, we compiled 138 high-quality ultrasound datasets comprising over 4.5 million images, spanning nine major anatomical regions and 32 representative organs.

2.2 Quality Control

An additional quality control and evaluation pipeline was implemented during construction of the EchoAtlas dataset. To ensure data integrity, ultrasound images underwent a rigorous purification workflow: (1)

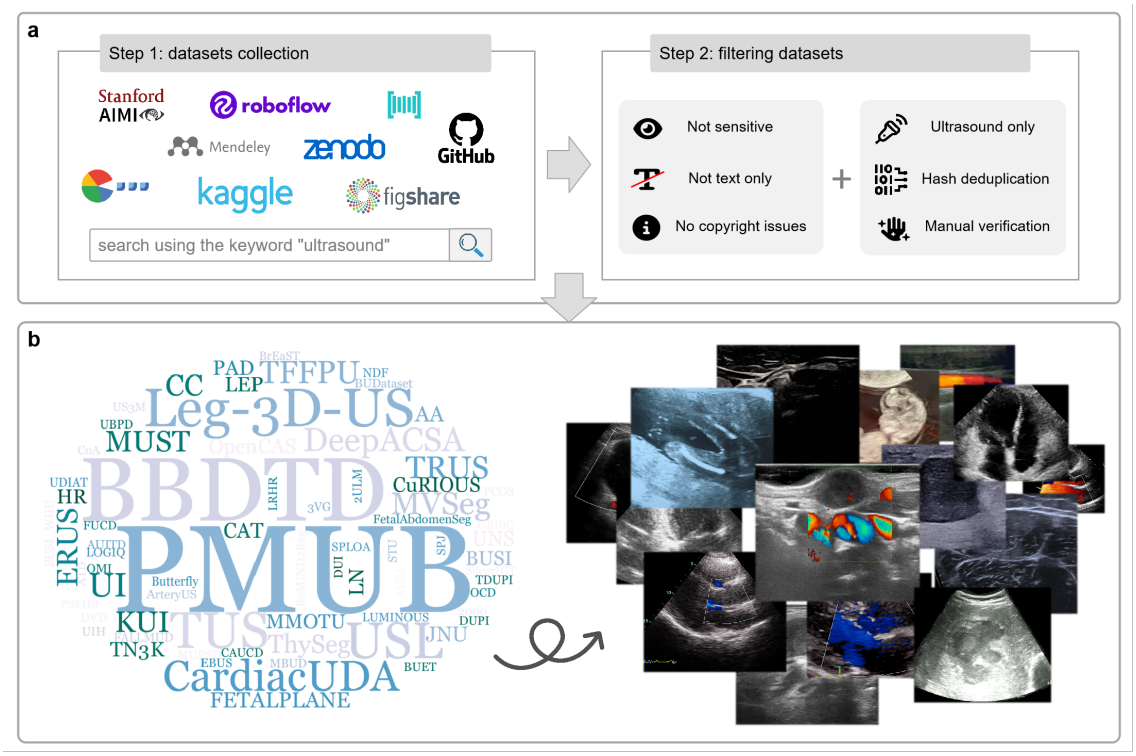

Figure 3: **Overview of the dataset curation pipeline for EchoAtlas. a.** Schematic illustration of the dataset collection and filtering workflow. Logos of major repositories are shown on the left, with keyword-based search ("ultrasound") and exclusion criteria depicted on the right. **b.** Visualization of the curated dataset composition, including a keyword cloud summarizing dataset metadata and representative ultrasound images illustrating anatomical diversity.

Removal of extraneous patient metadata surrounding the image; (2) Discarding of completely empty images or those containing fewer than 1,000 valid (non-zero) pixels; (3) For ultrasound videos, systematic uniform sampling at 10-frame intervals to mitigate redundancy. Post-hoc evaluation of the filtering and deduplication processes was conducted as follows: after data filtering, a random sample of 100 excluded candidates was analyzed, confirming no valid ultrasound images or additional data links. Following deduplication, 100 potential duplicate datasets were manually assessed using a predefined similarity threshold ($\geq 95\%$), verifying their redundancy. These procedures streamlined the dataset from 1,136 to 334 entries by eliminating redundancy. With the inclusion of specialized anatomically balanced datasets, we ultimately curated 138 high-quality ultrasound datasets.

3 EchoCare Model Pre-training

3.1 Model Design

For large-scale visual pretraining on EchoAtlas, we proposed EchoCare, a self-supervised framework for pre-training large vision transformer architectures based on the Masked Image Modeling (MIM) paradigm. Specifically, EchoCare adopts a modular design, comprising an image encoder, an image decoder, and a meta-object classifier (see Fig. 4), each module described in detail below.

The input to EchoCare is a masked image, which is passed along to the image. The image encoder processes the high-resolution image and outputs multi-scale downsampled embeddings. We provide a flexible

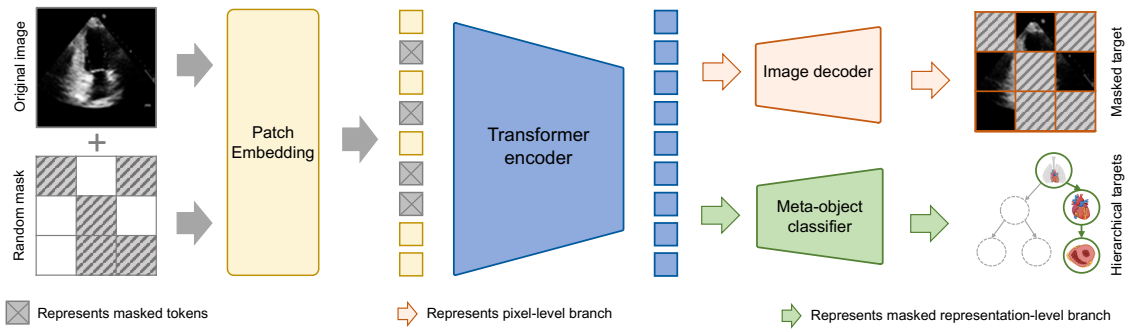

Figure 4: **Flowchart of EchoCare.** EchoCare takes a masked image as input and then outputs the reconstructed ultrasound image. To capture both global anatomical contexts and local ultrasound characteristics, EchoCare also incorporates a novel hierarchical classifier branch.

choice of backbone architectures with Swin Transformer base and large versions. The image decoder outputs a reconstructed image that has the same size as the original image, with a grayscale value between 0 and 1 for each pixel. The meta-object classifier includes input from the image and output object semantics. The output object semantics includes three levels: part, organ and anatomical structure. We follow SimMIM and SwinUNETR to build the image decoder head. The decoder is a transformer that gradually upsamples the image features back to high-resolution pixels. At the last layer, the attention dot product on the pixel embeddings delivers the reconstructed image.

Unified masked pretraining. The input image $x \in \mathrm{R}^{H\times W\times C}$ was split into N image patches $\{x_i^p\}_{i=1}^N$ and then tokenized into $z = [z_1, ..., z_N] \in \mathrm{V}^{h\times w}$ as the output labels of MIM using an image patch embedding layer. At the input layer, 50% image patches were randomly masked, and then the model predicted the visual tokens z_i of the masked patches. Next, we replaced the masked patches with a learnable embedding $e_{[M]} \in \mathrm{R}^D$, making the input corrupted image patches $x^M = \{x_i^p : i \notin M\}_{i=1}^N \cup \{e_{[M]} : i \in M\}_{i=1}^N$ that are fed into the transformer encoder. To optimize the model, we employ a reconstruction loss that aims to minimize the difference between the predicted pixel values of the masked image patches, $\hat{x}_i^p$, and the ground truth pixel values, x_i^p. Specifically, the reconstruction loss is defined as the Mean Absolute Error (MAE) between the predicted and original patches:

$$\mathcal{L}_{\text{MIM}} = \frac{1}{M}\sum_{i\in M}|\hat{x}_i^p - x_i^p|. \tag{1}$$

Hierarchical pretraining. The second pre-trained output (i.e., the meta-objection classifier) is used to further train EchoCare to represent images using hierarchical learning. Therefore, we designed a hierarchical loss for image global representation learning. Specifically, let's assume there are N_p body parts at the first level, which encompass N_o organs at the second level. Based on these N_o organs, there are N_a anatomical structures at the third level. Hence, the meta-object classifier has $N_p + N_o + N_a$ outputs. For each category, if a class is labeled positive, all its ancestor nodes (*i.e*, superclasses) should be labeled positive. And, if a class is labeled negative, all its child nodes (*i.e*, subclasses) should be labeled negative. To ensure the satisfaction of the above hierarchy constraints, we estimate a hierarchy-coherent score vector $P \in [0, 1]^{N_p+N_o+N_a}$. For class i, the updated score vector $p = [p_i] \in [0, 1]$ in P is given as:

$$\begin{cases} p_H = \min(s_u) & \text{if } \hat{l} = 1, \\ 1 - p_H = \min(1 - s_u) = 1 - \max(s_u) & \text{if } \hat{l} = 0. \end{cases} \tag{2}$$

Thus, after getting the hierarchical probabilities, we could maximize the log-likelihood between the probabilities and ground truth classification labels:

$$\mathcal{L}_{\text{HIE}} = \sum -\hat{l} \log(p_H) - (1 - \hat{l}) \log(1 - p_H). \tag{3}$$

Final pretraining loss function. Then, the final pretraining loss function is formulated as we combined the $\mathcal{L}_{\text{PRE}} = \mathcal{L}_{\text{MIM}} + \mathcal{L}_{\text{HIE}}$. Through incorporating $\mathcal{L}_{\text{MIM}}$, we focus on fine-grained details at the pixel level, ensuring that the model captures nuanced features within the data. In parallel, $\mathcal{L}_{\text{HIE}}$ enriches the learning process by providing a broader context through global representations. This dual approach not only improves the model's ability to generalize but also enhances its robustness in various applications.

3.2 Pretraining Protocol

Model architecture. In the EchoCare self-supervised pre-training pipeline we adopt SwinViT-B as the visual backbone. The model is configured with an embedding dimension of 128, stage depths of [2, 2, 18, 2] blocks, and multi-head attention heads of [4, 8, 16, 32]. The patchify stem operates on 2×2 pixels, yielding 128×128 patches at 256×256 input resolution. And, the window is set to 8×8. The encoder maps a $256 \times 256 \times 3$ image to an $8 \times 8 \times 1024$ feature grid, giving a $32\times$ spatial reduction and 1024-dimension final tokens for downstream learning. For image reconstruction decoder head, we used a VAE up-sampling path that progressively doubles spatial resolution until the original 256×256 is restored. Each up-sampling step is implemented as an Upsample layer with scale factor 2 and bilinear interpolation, followed by a 3×3 convolution layer to refine the latent structure. Furthermore, the image hierarchical representation head is implemented with a single fully connected layer that performs meta object classification directly on the pooled 1024-dimension deep representation feature.

Pretraining settings. Image augmentations included random vertical flip ($P = 0.5$), random horizontal flip ($P = 0.5$), and random crop ($P = 0.5$) to convert images to greyscale and weak colour jittering ($P = 0.2$) with specific adjustments to brightness, contrast, saturation and hue. We pretrained EchoCare for one million steps using the pretraining loss of $\mathcal{L}_{\text{PRE}}$ for images. The batch sizes were 256, and EchoCare used an input image with 256×256 pixels and then patched as 2×2 pixels. We used the AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.9$ and $\epsilon = 0.9$ for optimization. We used a cosine learning rate decay scheduler with a peak learning rate of 1.0×10^{-4} and a linear warmup of 10,000 steps. The weight decay was set as 0.05, and the stochastic depth with a rate of 0.1 was used.

Pretraining advantages. Building on EchoAtlas, we pretrained EchoCare (Fig. 4), a novel vision foundation model for ultrasound imaging, and applied it to a suite of clinical tasks. EchoCare employs a modular design based on an extended self-supervised Masked AutoEncoder (MAE) architecture for representation learning, comprising an image encoder to encode input ultrasound image features and two decoders: an image decoder to reconstruct images from sparse patches and an anatomy-classifier decoder for joint learning of hierarchical anatomic features (Fig. 4). Unlike prior medical foundation models that directly adopt off-the-shelf MAE structures or focus solely on local pixel-level prediction, we introduce a novel representation-level prediction branch, the anatomy-classifier, into the MAE framework. This branch learns global and hierarchical anatomical relationships from body regions to organs to anatomic structures, mirroring clinical diagnostic workflows. For example, the anatomy-classifier predicts pathways such as "Thorax→Heart→Apical two-chamber" and "Thorax→Heart→Apical four-chamber". Leveraging the inherent hierarchical organization of the anatomy system, this high-level classification process evolves naturally without human intervention. By integrating local pixel-level and global representation-level features, EchoCare enhances the encoder's ability to interpret ultrasound images, thereby boosting downstream clinical applications. In the following sections, we demonstrate its versatility and generalization to diverse ultrasound clinical tasks.

Table 1: Overview of the 10 curated clinical applications and their corresponding dataset partitions. For each we specify the imaged country, anatomical region, anatomy organ(s), task category (segmentation, detection, classification, landmark localization, regression, enhancement, or report generation), and the exact number of training and testing videos/images.

Name	Challenge	Task type	Country	Anatomy part	Anatomy organ	Train	Test
DDTI [17]	Thyroid node segmentation	Node segmentation	Colombia	Head & Neck	Thyroid nodule	522	123
MusV [18]	Vessel segmentation	Vessel segmentation	China	Head & Neck, Low limb	Carotid and femoral vessels	2,203	911
AbdomenUS	Abdomen organ segmentation	Organ segmentation	China	Abdomen	Liver, Kidney, Pancreas, Bladder, Spleen	3,345	872
Thyroid Cine-Clip [19]	Thyroid node diagnose	Node classification	USA	Head&Neck	Thyroid nodule	157	35
BRA-BUS [20]	BI-RADS category assessment	BI-RADS classficaion	Brazil	Thorax	Breast	1,500	375
SYSU-FLL-CEUS [21]	Liver lesion recognition	Lesion classification	China	Abdomen	Liver	285	68
FOCUS [22]	Thorax and cardiac organ detection	Organ detection	Spain	Fetus	Fetal thorax and cardiac	250	50
BrainLandmark [23]	Brain landmark location	Landmark location	Australia	Fetus	Fetal brain	80	24
CAMUS [24]	Ejection fraction regression	EF regression	France	Thorax	Heart	450	50
USenhance [25]	Low-quality image enhancement	Image enhancement	China	Head & Neck, Abdomen, Thorax	Thyroid, Kidney, Liver, Breast, Carotid artery	1,654	426
USreport [26]	Clinical report generation	Report generation	China	Head & Neck, Abdomen, Thorax	Thyroid, Liver, Breast	1,118	279

4 Evaluation

4.1 Evaluation Methodology

To validate the generalizability of the pretrained foundation model EchoCare, we established 10 external validation tasks spanning representative clinical ultrasound scenarios including lesion segmentation, disease diagnosis, one-shot recognition, and quantitative regression. These tasks leveraged independent datasets covering anatomical regions such as thyroid, venous systems, abdominal organs, and cardiac structures. All external datasets were explicitly excluded from the EchoAtlas pretraining corpus to prevent data leakage and ensure unbiased evaluation of pretraining effects. Below is a detailed breakdown of each clinical validation task and corresponding dataset, organized by task category to highlight translational relevance.

4.2 Validation Tasks and Datasets

Thyroid node segmentation on DDTI dataset (1 classes): The DDTI dataset [17] for thyroid node segmentation comprises 388 patients with B-mode ultrasound scans from the Instituto de Diagnóstico Médico S.A. and National University of Colombia, annotated for nodule lesion segmentation. Images were extracted from thyroid ultrasound video sequences acquired using TOSHIBA Nemio 30 and Nemio MX systems, equipped with 12 MHz convex and linear transducers. Accurate automated segmentation of thyroid nodules enables clinicians to assess morphological features–including size, shape, and margins–to discriminate between benign and malignant lesions, which is critical for early thyroid disease diagnosis. Sub-images were cropped from 42 composite sequences and integrated with single-frame images, yielding a total of 645 ultrasound images with an average resolution of 348×280 pixels and mean mask area of 153.25 pixels. For cross-validation, data were split at the patient level in an 8:2 ratio, resulting in 308:80 patient folds (522:123 images) for training and evaluation.

Artery&vein segmentation on Mus-V dataset (2 classes): The Mus-V dataset [18] for vascular segmentation comprises 3,114 ultrasound images from the Institute of Automation, Chinese Academy of Sciences,

annotated for carotid and femoral vessel segmentation. Images were acquired from 11 healthy volunteers using an Angel Pionner H20 Ultrasound Scanner, capturing carotid and femoral vessels in the arm and neck regions. Accurate arterial-venous segmentation is critical for real-time low-risk vascular interventions–such as those for coronary and peripheral vascular diseases–enabling clinicians to precisely target vessels and minimize the risk of adjacent structure injury. The dataset includes separate annotations for arteries and veins to facilitate vascular analysis and identification, with images sampled from 105 videos (5-160 frames per video) at 400×600 pixel resolution. For evaluation, official train-test splits were used to achieve an 8:2 patient-level division, yielding 2,203:911 images for training and validation.

Abdominal multi-organ segmentation on AbdomenUS dataset (5 classes): Beyond single/two-class segmentation, EchoCare was further validated on multi-organ segmentation to demonstrate its potential in reducing annotation burdens on experts. The AbdomenUS dataset for multi-organ segmentation encompasses 4,217 ultrasound images from BGI Genomics Co., Ltd., acquired from 64 volunteers using the MGIUS-R3 ultrasound system. Images were annotated for at least one of five abdominal organs: 1) liver, 2) pancreas, 3) kidney, 4) bladder, and 5) spleen. This multi-organ annotation framework allows clinicians to systematically evaluate anatomical morphology–including organ shape, positional relationships, and pathological signs–from diverse sonographic perspectives. For model training and validation, data were divided into an 8:2 ratio at the case level, yielding a training set of 51 cases (3,345 B-mode images) and a validation set of 13 cases (872 B-mode images).

Thyroid nodule false positive mitigation on ultrasound cine-clip dataset (2 classes): The thyroid nodule false positive mitigation task leverages the Ultrasound Cine-clip dataset from the Center for Artificial Intelligence in Medicine & Imaging, comprising 192 histopathologically confirmed thyroid nodules (175 benign, 17 malignant) across 167 patients (mean age 56 ± 16 years, 137 female) who underwent cine ultrasound between April 2017 and May 2018. The dataset includes ultrasound cine-clip sequences, radiologist-annotated segmentation, patient demographics, lesion metrics (size/location), and definitive histopathological diagnoses. Given the nonspecific nature of ultrasound findings, which often lead to unnecessary biopsies, AI-driven pre-biopsy triage of benign and malignant nodules holds significant clinical value for reducing false positive cancer classifications. All ultrasound acquisitions were performed using Logiq E9 (GE Healthcare) or Siemens S2000 systems, with images obtained by certified sonographers from supine patients with slightly hyperextended necks. The cine-clips feature 802×1054 pixel resolution. Following official dataset splits, the cohort was partitioned into training (157 cine-clips, 4/5) and validation (35 cine-clips, 1/5) subsets to ensure reproducible evaluation.

BI-RADS category assessment on BRA-BUS dataset (4 classes): The Breast Imaging Reporting & Data System (BI-RADS) category assessment leverages the BRA-BUS dataset, which offers a standardized lexicon and reporting framework for breast ultrasound. BI-RADS facilitates consistent communication of imaging findings among radiologists and clinicians, with final assessments categorized by malignancy likelihood: categories 2 (benign), 3 (probably benign), 4 (suspicious), and 5 (highly suggestive of malignancy), as annotated by senior ultrasonographers. The BRA-BUS dataset comprises 1,875 anonymized images from 1,064 female patients, acquired using four ultrasound systems (GE Logiq 5, GE Logiq 7, Toshiba Aplio 300, GE U-Systems) with linear-array transducers at the National Institute of Cancer (Rio de Janeiro, Brazil). For validation, an official 5-fold cross-validation strategy was employed, combining four folds into the training set (800 patients, 1,500 images) and using the remaining fold for validation (264 patients, 375 images).

Focal liver lesion diagnosis on SYSU-FLL-CEUS dataset (3 classes): The focal liver lesion (FLL) diagnosis task leverages the SYSU-FLL-CEUS dataset, encompassing contrast-enhanced ultrasound data for three pathological types: 186 hepatocellular carcinoma (HCC), 109 hemangioma (HEM), and 58 focal nodular hyperplasia (FNH) cases. Acquired from the First Affiliated Hospital of Sun Yat-sen University using an Aplio SSA-770A ultrasound system (Toshiba Medical Systems), the dataset captures FLLs with

heterogeneous patterns, varying in size, contrast intensity, morphological features, and anatomical location (resolution: 768×576 pixels). Early FLL characterization from ultrasound is critical for timely oncological intervention, as these lesions exhibit diverse imaging phenotypes. The dataset was case- and label-stratified into 8:2 training-evaluation folds to maintain class distribution: the training set includes 150 HCC, 88 HEM, and 47 FNH cases, while the evaluation set contains 36 HCC, 21 HEM, and 11 FNH cases.

Fetal thorax and cardiac detection on FOCUS dataset (2 objects): The FOCUS dataset [22] is designed for fetal thorax and cardiac organ detection, comprising 300 four-chamber view fetal echocardiography ultrasound images from 217 subjects across Hospital Clinic and Hospital Sant Joan de Deu in Barcelona, Spain. This dataset captures the cardiothoracic diameter ratio–a critical biometric for assessing fetal congenital heart disease–via ellipse annotations of cardiac and thoracic regions in every image. All images (230×245 pixels, uniform resolution) feature distinct annotations for fetal cardiac and thoracic structures, varying in size, aspect ratio, and rotational orientation. Following official patient-level splits to prevent data leakage, the dataset was partitioned into 250 training images and 50 evaluation images, maintaining clinical representativeness.

Brain landmark detection on BrainBenchmark dataset (24 landmarks): The brain landmark detection task leverages the BrainBenchmark dataset [23], comprising 104 2D fetal brain ultrasound images acquired at 20–20.6 weeks of gestation. Developed for monitoring neurodevelopmental trajectories, this benchmark captures structural changes from embryonic stages to postnatal development, with images obtained from 70 pregnant women (median age 31 years, range 18–42) via routine mid-trimester scans using a Voluson E10 ultrasound system with a high-frequency transabdominal probe (C2-9). Each image is annotated with 24 anatomical landmarks including 4 skull landmarks, 3 thalamic landmarks, 8 cerebellar perimeter landmarks, 4 cavum landmarks, 3 Sylvian fissure landmarks, and 2 midline edge landmarks. Images were collected from 70 subjects with variable scanning frequencies (8 women scanned three times, 18 women twice, and 44 women once), all without detected abnormalities. For validation, an 8:2 image-level split yielded 80 training and 24 evaluation images to ensure developmental stage representativeness.

Ejection fraction prediction on CAMUS dataset (500 cases): The CAMUS dataset [24] for ejection fraction prediction comprises 500 2D ultrasound sequences, recognized as a standard benchmark for cardiac function assessment. This regression task involves inputting ultrasound frame sequences to predict left ventricular ejection fraction (LVEF), a critical biomarker for evaluating cardiac health and diagnosing heart disease, particularly when derived from four-chamber view acquisitions. Ultrasound sequences were acquired using GE Vivid E95 scanners (GE Vingmed Ultrasound, Horten, Norway) with a GE M5S probe (GE Healthcare, US) at the University Hospital of St Etienne (France). Each sequence includes manual annotations of left ventricular volumes at end-diastole and end-systole, from which ejection fraction is calculated. Following official protocols, the dataset was partitioned into 450 training and 50 validation cases to ensure reproducible evaluation of LVEF prediction models.

Image enhancement based on USenhance dataset (5 organs): The ultrasound image enhancement task [25] leverages the USenhance Challenge 2023 dataset, comprising 2,100 ultrasound images (1,050 unpaired low/high-quality image pairs) across five organs (thyroid, kidney, liver, breast, and carotid artery) from 109 patients. AI-driven enhancement of high-quality ultrasound images from low-fidelity inputs obviates the need for hardware upgrades, driving technological innovation in ultrasound devices and enabling more precise clinical applications. The dataset includes images acquired using diverse imaging systems: thyroid imaging employs the mSonics MU1 (low-end) and Toshiba Aplio 500 (high-end); carotid artery and abdominal imaging use SSUN (low-end) and Toshiba Aplio 500 (high-end); breast imaging utilizes the mSonics MU1 (low-end) and Aixplorer system from SuperSonic Imaging (high-end). All images were resized to a uniform 256×256 pixel resolution. Following an organ-stratified 8:2 split, the dataset was partitioned into 837 training image pairs (232 thyroid, 161 breast, 97 kidney, 119 liver, 228 carotid) and 213 validation image pairs (59 thyroid, 41 breast, 25 kidney, 30 liver, 58 carotid), ensuring clinical representativeness across anatomical structures.

Table 2: **Summary of evaluation metrics**. This table outlines various metrics and their applicability across different tasks, including classification (Cls.), segmentation (Seg.), detection (Det.), localization (Loc.), regression (Reg.), enhancement (Enhance.), and report generation (Rep.).

Metric	Formula	Tasks						
		Cls.	Seg.	Det.	Loc.	Reg.	Enhance.	Rep.
Accurayc (ACC)	$\mathrm{ACC} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}}$	✓						✓
Precision	$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$	✓						✓
Recall	$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$	✓						✓
F_1 score	$\mathrm{F}_1 = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision+Recall}}$	✓						✓
The Area Under the Curve (AUC)	$\mathrm{AUC} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \mathbb{I}(p_i > p_j)}{m \cdot n}$	✓						
Dice Similarity Coefficient (DSC)	$\mathrm{DSC} = \frac{2 \times \lvert \mathrm{X} \cap \mathrm{Y} \rvert}{\lvert \mathrm{X} \rvert + \lvert \mathrm{Y} \rvert}$		✓					
Normalized Surface Dice (NSD)	$\mathrm{NSD} = \frac{\lvert \partial \mathrm{X} \cap \partial \mathrm{Y}_\epsilon \rvert + \lvert \partial \mathrm{Y} \cap \partial \mathrm{X}_\epsilon \rvert}{\lvert \partial X \rvert + \lvert \partial Y \rvert}$		✓					
Average Precision (AP)	$\mathrm{AP} = \frac{1}{\mathrm{N}} \sum_{r_k} \max(\text{Precision at } r_k)$			✓				
$\mathrm{P}_{\mathrm{CTR}}$	$\mathrm{P}_{\mathrm{CTR}} = (1 - \frac{\lvert \mathrm{R}_{\mathrm{true}} - \mathrm{R}_{\mathrm{pred}} \rvert}{\mathrm{R}_{\mathrm{true}}}) \times 100\%$			✓				
Mean Squared Error (MSE)	$\mathrm{MSE} = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \lvert \mathrm{p}_i - \mathrm{g}_i \rvert_2^2$				✓			
Successful Detection Rate (SDR)	$\mathrm{SDR} = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \mathbb{I}(\lvert \mathrm{p}_i - \mathrm{g}_i \rvert \leq \tau)$				✓			
Mean Absolute Error (MAE)	$\mathrm{MAE} = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \lvert y_i - \hat{y}_i \rvert_1^1$					✓		
Natural Image Quality Evaluator (NIQE)	$\mathrm{NIQE} = \sqrt{(\nu_1 - \nu_2)^T (\frac{\sum_1 + \sum_2}{2})^{-1} (\nu_1 - \nu_2)}$						✓	
Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE)	$\mathrm{BRISQUE} = \mathcal{R}(F_{36}(\mathrm{I}))$						✓	
Perception based Image Quality Evaluator (PIQE)	$\mathrm{PIQE} = \frac{1}{\mathrm{M}} \sum_{i=1}^{\mathrm{M}} Q_i$						✓	
Frechet Inception Distance (FID)	$\mathrm{FID} = \lvert\lvert \mu_1 - \mu_2 \rvert\rvert_2^2 + \mathrm{Tr}(\sum_1 + \sum_2 - 2(\sum_1 \sum_2)^{\frac{1}{2}})$						✓	
Bilingual Evaluation Understudy (BLEU)	$\mathrm{BLEU} = \mathrm{BP} \times \exp(\sum_{n=1}^{\mathrm{N}} \omega_n \log p_n)$							✓
Recall-Oriented Understudy for Gisting Evaluation (ROUGE)	$\mathrm{ROUGE} = \frac{\sum_{n=1}^{\mathrm{N}} \mathrm{Recall}_n}{\mathrm{N}}$							✓
Metric for Evaluation of Translation with Explicit ORdering (METEOR)	$\mathrm{METEOR} = \mathrm{F}_{\mathrm{mean}} \times (1 - \mathrm{Penalty})$							✓

Ultrasound text report generation on USreport dataset (3 organs): The USreport dataset is designed for ultrasound text report generation, comprising three independent clinical corpora of ultrasound image-text pairs covering breast, thyroid, and liver examinations. Specifically, it includes 3,534 breast, 2,460 thyroid, and 1,397 liver cases, all sourced from the ultrasonic department database of the PLA General Hospital. AI-driven automated report generation from ultrasound images holds promise to streamline clinical diagnostic workflows. Each report is associated with two representative images selected by clinicians, forming image-text pairs for model training. Following official data splits, an 8:2 train-validation partition was applied: Liver: 1,118 training, 279 validation cases; Breast: 2,827 training, 707 validation cases; Thyroid: 1,968 training, 492 validation cases.

4.3 Evaluation Metrics

In this section, we use a comprehensive array of evaluation metrics to rigorously assess the performance of different models. These metrics span a diverse range of tasks, including classification, segmentation, detection, localization, regression, enhancement, and report generation. Through a detailed analysis of their applicability across these varied tasks, we achieve a nuanced understanding of the models' performance attributes. This systematic approach to evaluation not only facilitates the accurate identification of the strengths and limitations of the models but also provides a robust foundation for future enhancements and optimizations.

Classification. For three diagnostic classification applications: 1) benign-malignant classification of thyroid nodules; 2) breast tumor BI-RADS grading; and 3) diagnosis of focal liver lesions in abdominal ultrasound, we used four metrics to assets model classification performance: Accuracy (ACC), Precision, Recall and F_1 score. Each of these metrics is defined and explained as follows.

- **ACC**: It was employed to assess recognition performance, defined as the ratio of correctly predicted samples to the total number of samples (ranging from 0 to 1). In equation, TP denotes the number of true positives, TN the number of true negatives, FP the number of false positives, and FN the number of false negatives. A higher ACC indicates better overall prediction correctness.

- **Precision**: This metric is used to evaluate the accuracy of the positive predictions made by the model. It is defined as the ratio of true positive samples to the total number of samples predicted as positive. In other words, Precision measures the proportion of correct positive identifications among all positive predictions. A higher Precision indicates a lower rate of false positives, reflecting a model's reliability in identifying positive cases.

- **Recall**: This metric assesses the model's ability to identify all relevant positive cases. It is defined as the ratio of true positive samples to the total number of actual positive samples. Recall measures the proportion of correctly identified positive instances out of all actual positives. A higher Recall indicates that the model effectively captures more true positives, thereby minimizing the number of false negatives.

- **AUC**: The Area Under the Curve (AUC) was utilized to evaluate classification performance, quantifying the model's ability to rank positive instances higher than negative ones. Ranging from 0 to 1. In equation, m and n denote the number of positive and negative samples, respectively; p_i and p_j are the predicted probabilities for positive and negative samples, and $\mathbb{I}$ is the indicator function. An AUC of 1 signifies perfect classification, while 0.5 indicates random performance.

Segmentation. Accurate segmentation in ultrasound images is crucial for enabling sonographers to delineate anatomical structures and identify pathological conditions. To ensure a fair evaluation of the performance of various segmentation models, we selected two widely used metrics: 1) the Dice Similarity Coefficient (DSC) to assess regional segmentation performance; and 2) the Normalized Surface Dice (NSD) to evaluate the accuracy of segmentation margins. Their detailed definitions and explanations are provided below.

- **DSC**: The Dice Similarity Coefficient (DSC) was used to assess segmentation performance, measuring the spatial overlap between predicted and ground-truth region. In equation, X denotes the predicted segmentation mask, and Y denotes the ground-truth mask. Ranging from 0 to 1, with higher values indicating closer correspondence to the ground truth.

- **NSD**: The Normalized Surface Dice (NSD) was used to evaluate segmentation boundary accuracy, quantifying the geometric correspondence between predicted and ground-truth surfaces. In equation, $\partial\mathrm{X}$ and $\partial\mathrm{Y}$ denote the boundaries of the predicted segmentation and ground-truth mask, respectively. $\partial\mathrm{X}_\epsilon$ and $\partial\mathrm{Y}_\epsilon$ represent ϵ-dilated boundaries (with ϵ set to 2 mm in this study), which expand the boundary regions to account for spatial tolerance. Higher NSD indicates more precise surface alignment between the prediction and ground truth, reflecting superior boundary localization performance.

Detection. In clinical ultrasound workflows, reliable detection of organs underpins rapid localization and downstream quantitative analysis. To fairly benchmark competing detection methods in fetal throax&cardiac organ detection task, we selected the community-standard Average Precision (AP) as the performance metric. Also, we included $\mathrm{P}_{\mathrm{CTR}}$ metric to estimate the precision of the Cardiothoracic diameter Ratio (CTR) biometric. Formal definitions and interpretations are provided below.

- **AP**: The AP metric which computes the area under the precision-recall curve, providing a single value that encapsulates the model's precision and recall performance. To ensure an impartial comparison of detection frameworks, we adopt AP under two complementary formulations: 1) box-AP, which evaluates bounding-box overlap via the Intersection-over-Union (IoU) criterion, and 2) mask-AP, which assesses pixel-level IoU between predicted and ground-truth masks.

- $\mathrm{P}_{\mathrm{CTR}}$: This metric was used to estimate the precision of CTR biometric measurement. In equation, $\mathrm{R}_{\mathrm{pred}}$ and $\mathrm{R}_{\mathrm{true}}$ denote the predicted CTR and the ground truth CTR, respectively. The CTR is formulated as $\mathrm{R} = b_C/b_T$, where b_C represents the length of minor axis of the cardiac object, b_T represents that of the thoracic object.

Location. Medical landmark location aims to automatically identify the locations of predefined anatomical points. For fetal landmark location task, we used two ubiquitous metrics prevalent in the medical landmark location domain: 1) Mean Squared Error (MSE), and 2) Successful Detection Rate (SDR). Detailed mathematical formulations and clinical interpretations of each metric are presented as follows.

- **MSE**: The Mean Squared Error (MSE) metric was applied to quantify landmark localization error, measuring the average squared pixel distance between predicted and ground-truth positions. In equation, $\mathbf{p}_i$ denotes the predicted 2D landmark coordinate, $\mathbf{g}_i$ is the ground-truth coordinate, and N is the total number of landmarks. Lower MSE values indicate more precise localization.
- $\mathrm{SDR}(\tau)$: The Successful Detection Rate (SDR) was utilized to evaluate landmark detection accuracy, quantifying the proportion of landmarks localized within a specified tolerance threshold τ. τ was set to 2 pixel, 4 pixel, 10 pixel in this research. Higher SDR values reflect greater reliability in landmark detection, with the threshold parameter τ adjusted to balance clinical tolerance for localization error. This metric is particularly valuable for assessing model consistency across diverse anatomical landmarks.

Regression. Left-ventricular ejection fraction (LVEF) remains the cardinal numeric descriptor of systolic performance and a principal therapeutic lighthouse in heart-failure care. Precision estimation of LVEF is pivotal to the early diagnosis of both congenital and acquired cardiovascular disorders, informs therapeutic decision-making, and enables robust prognostic stratification. To benchmark competing regression models we adopt Mean Absolute Error (MAE), as the single summary metric. Its mathematical formulation and clinical interpretation are detailed below.

- **MAE**: The Mean Absolute Error (MAE) metric was used to measure the average absolute deviation between the predicted LVEF and the expert-derived reference. In equation, $\hat{\mathbf{y}}_i$ and $\mathbf{y}_i$ denote the model-estimated and clinically annotated LVEF values for the i-th subject, respectively. And, the N is the total number of echocardiographic examinations in the testing set.

Enhancement. Ultrasound image enhancement can be formulated as an image generation task that aims to transform low-quality ultrasound images into high-quality ones. However, collecting paired low-quality and high-quality images in clinical settings can be challenging. As a result, the experiments were conducted under unpaired settings using the public USenhance dataset. To enable a comprehensive and reference-free assessment, we curate four blind-quality metrics: Natural Image Quality Evaluator (NIQE) [27], Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [28], Perception based Image Quality Evaluator (PIQE) [29], and Frechet Inception Distance (FID) [30]. Each detailed in the following subsections.

- **NIQE**: The Natural Image Quality Evaluator (NIQE) is based on the construction of a quality-aware collection of statistical features based on a space domain natural scene statistic (NSS) model. In equation, ν_1, ν_2 and $\sum_1$, $\sum_2$ are the mean vectors and covariance matrices of the generated-image Multivariate Gaussian (MVG) model and the high-quality (reference-domain) image's MVG model.
- **BRISQUE**: The Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) utilizes a NSS model and introduces a novel approach to modeling the statistics of pairwise products of neighboring luminance values. The parameters derived from this model provide a quantifiable measure of the naturalness of the image. In equation, I is the input image whose naturalness score is to be computed, $\mathcal{R}(\cdot)$ denotes the Support Vector Regressor (SVR) pre-trained on the LIVE database, while $\mathcal{F}_{36}(\cdot)$ extracts 36 dimension NSS features, including parameters of the symmetric generalized Gaussian distribution (SGGD) and the asymmetric generalized Gaussian distribution (AGGD).

- **PIQE**: The Perception based Image Quality Evaluator (PIQE) is a block-level and no-reference perceptual image-quality assessment method that detects local noise and distortion regions. In equation, Q_i is the perceptual noise-level score of the i-th image block, computed from perceptual features such as brightness, contrast, and edges.

- **FID**: The Frechet Inception Distance (FID) is measures the distributional discrepancy between generated high-quality and reference high-quality ultrasound images in a high-dimensional space through the Frechet distance. In equation, μ_1 and μ_2 denote the 2048-dimensional mean vectors of the reference and generated high-quality ultrasound image features extracted from the Inception-v3 pool3 layer, respectively. $\sum_1$ and $\sum_2$ are the corresponding 2048×2048 covariance matrices. The term $|| \cdot ||_2^2$ represents the Euclidean distance between the two mean vectors, while $\mathrm{Tr}(\cdot)$ is the trace operator.

Report. Automatic generation of reports from medical images alleviates clinicians' documentation workload, allowing them to devote more time to patient care. For the USreport application task, it requires two input ultrasound images and utilizes an AI algorithm to generate a text report. To conduct a comprehensive assessment of the quality of the generated text reports, we selected two sets of metrics: Natural Language Generation (NLG) metrics and Clinical Efficacy (CE) metrics. The NLG metrics comprise established measures, Bilingual Evaluation Understudy (BLEU) [31], Recall-Oriented Understudy for Gisting Evaluation - Longest Common Subsequence (ROUGE-L) [32], and the Metric for Evaluation of Translation with Explicit ORdering (METEOR) [33]. The CE metrics include accuracy, precision, recall, and F_1 score related to essential entities. Each is explained in the following subsections.

- **BLEU**: The Bilingual Evaluation Understudy (BLEU) metric quantifies the quality of machine-translated or automatically generated text by measuring n-gram overlap between system output and one or more reference translations. In equation, BP is the brevity penalty, p_n is the precision of n-grams, and ω_n is the weight. In this study, we set different n-gram lengths, with $n = 1, 2, 3, 4$, to comprehensively assess text similarity.

- **ROUGE-L**: The Recall-Oriented Understudy for Gisting Evaluation (ROUGE) is a metric used to evaluate the quality of summaries by measuring the overlap between the generated summary and reference summaries. In equation, Recall_n represents the recall of n-grams based on the number of matching n-grams between the generated summary and the reference summary, normalized by the total number of n-grams in the reference. N indicates the maximum n-gram length considered in the evaluation. To enhance the evaluation of coherence and fluency in generated summaries, we selected the Longest Common Subsequence (LCS) for our analysis.

- **METEOR**: The Metric for Evaluation of Translation with Explicit ORdering (METEOR) is designed to evaluate the quality of machine translation, integrating both precision and recall. It serves as a robust evaluation metric for assessing the fluency and adequacy of translated content. In equation, $\mathrm{F}_{\mathrm{mean}}$ denotes the weighted F-score, reflecting the degree of alignment between the generated translation and the reference translation. The $\mathrm{Penalty}$ term adjusts for the impact of incoherent matches.

- **CE Metrics**: For the CE metrics, our focus is on extracting key information from the reports rather than assessing text similarity. We identified essential entities for each report based on input from sonographers. Specifically, the entities includes: liver, capsule, echogenicity, vein, kidney, intrahepatic duct, bile duct, gallbladder, margin, pancreas, pancreatic duct, lesion, spleen, nodule. Suppose each dataset comprises a set of m key entities of interest, denoted as $\{1, 2, 3, \ldots, m\}$. If an entity i is mentioned in the report, it is labeled as 1; otherwise, it is labeled as 0. This approach transforms the task into a multi-label classification problem. Then, we could compute accuracy, precision, recall, and F_1 score.

4.4 Statistical Analysis

For all experimental results, performance metrics are reported as mean ± standard deviation across 20 independent trials. For each evaluation task, two-sample t-tests were conducted between the best-performing model and all others, with statistical significance denoted by asterisks (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). A two-sided P-value < 0.05 was considered statistically significant. For the thyroid nodule false positive mitigation binary-classification task, accuracy, sensitivity, and specificity were determined using the optimal cut-off value derived from the ROC curve to maximize the Youden index (sensitivity + specificity − 1). All statistical analyses were performed using Python (version 3.10) and MedCalc (version 22.032). Across all experimental settings, results are visualized via box plots (version 3.9.1) showing quartiles and whiskers at 1.5× interquartile range, based on 20 repeated runs to characterize model performance variability.

4.5 Results on Downstream Ultrasound Applications

We systematically evaluated EchoCare diagnostic performance on 10 clinical applications across 7 task types (Fig. 5 and Fig. 6). These datasets cover tasks ranging from binary diagnosis task to multi-class classification, single-class tumor segmentation to abdominal multi-organ segmentation, as well as ultrasound image enhancement, fetal landmark localization, organ detection and cardiac ejection fraction regression. We compare EchoCare with pervious state-of-the-art (SOTA) task-specific models (w/o FM) and seven representative foundation models: RadImageNet [34], UltraSAM [35], CLIP [36], BiomedCLIP [37], DINO [38], SimMIM [39], USFM [8]. Each model is fully fine-tuned on the task-specific dataset and evaluated with their corresponding metrics. EchoCare consistently outperformed all other models, achieving significant improvements on 10 clinical tasks. These results validate the effectiveness of Echocare. The domain-specific analyses of the experimental results are as follows.

4.5.1 Disease diagnostic classification

Disease diagnostic classification represents a pivotal clinical application of vision foundation models. High-performance ultrasound foundation models can substantially enhance the accuracy of disease lesion classification, mitigate false-positive decisions, and thereby reduce patient anxiety and costs. To demonstrate the utility in clinical decision-making, EchoCare was validated across three distinct diagnostic classification applications: 1) benign-malignant classification of thyroid nodules; 2) breast tumor BI-RADS grading; and 3) diagnosis of focal liver lesions in abdominal ultrasound.

EchoCare achieved leading performance across all the evaluated classification tasks (Fig. 5a-c). Specifically, EchoCare achieved an AUC (Area Under the ROC Curve) of 86.48%±1.19% and an F1-score of 87.45%±1.21% on the thyroid nodule dataset (Fig. 5a); 70.36%±1.01% accuracy and 65.38%±1.06% macro-F1 on breast BI-RADS grading (Fig. 5b); and 87.12%±0.91% accuracy and 83.44%±0.95% macro-F1 for focal liver lesions (Fig. 5c). Compared with the second-best model (USFM [8]), EchoCare outperformed by average margins of 3.35% (AUC) and 4.25% (F1-score) on thyroid nodules, 3.09% (accuracy) and 3.85% (macro-F1) on breast BI-RADS, and 3.45% (accuracy) and 3.98% (macro-F1) on focal liver lesions. These findings highlight EchoCare as a powerful foundation model capable of learning discriminative image representations, demonstrating great potential in distinguishing subtle differences between hepatocellular carcinoma, hemangiomas, and focal nodular hyperplasia. These lesions often exhibit overlapping sonographic appearances, which is a key challenge in achieving accurate manual diagnosis. Collectively, the experimental results confirm that EchoCare serves as a reliable diagnostic auxiliary tool, advancing ultrasound-based disease diagnostic classification and accelerating the clinical decision-making process.

4.5.2 Anatomical segmentation

Accurate segmentation in ultrasound images enables clinicians to characterize morphological features (*e.g.*, size, shape) and detect pathological abnormalities (*e.g.*, neoplastic lesions), which is fundamental for treatment

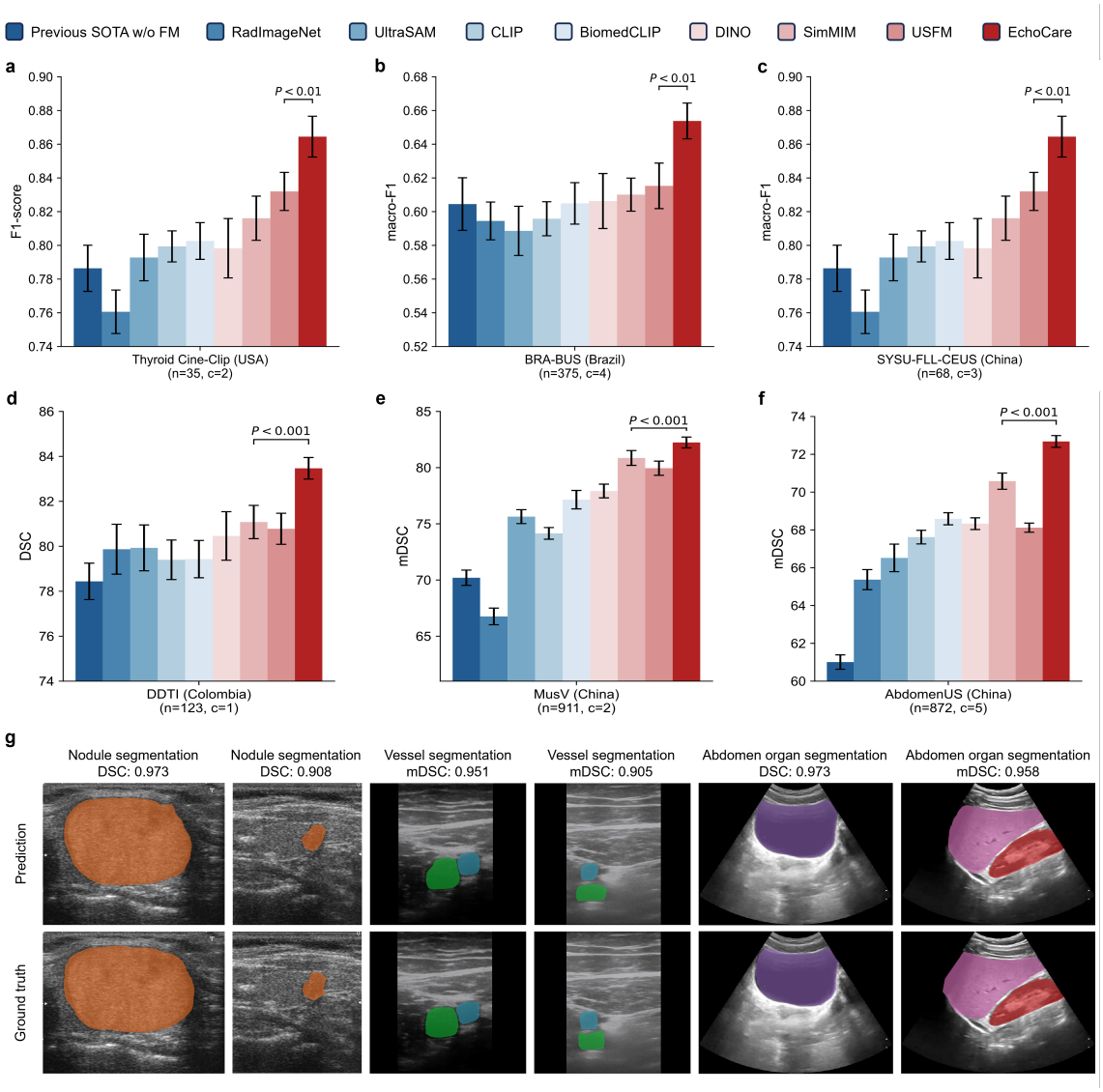

Figure 5: **Evaluation on disease diagnostic classification and anatomical segmentation. a-f.** EchoCare consistently outperforms previous state-of-the-art (SOTA) models (w/o FM) and other existing foundation models (RadImageNet [34], UltraSAM [35], CLIP [36], BiomedCLIP [37], DINO [38], SimMIM [39], USFM [8]) across different classification and segmentation tasks. Specifically, for classification, we evaluate on benign-malignant classification of thyroid nodules (**a**), breast tumor BI-RADS grading (**b**) and diagnosis of focal liver lesions in abdominal ultrasound (**c**). For segmentation, we evaluate on thyroid node segmentation (**d**), arterial-venous vessel segmentation (**e**), and the abdomen multi-organ segmentation (**f**). The two-sided Wilcoxon signed-rank test was used to assess the statistical differences between EchoCare and the second-best model. **g.** Six examples comparing the segmentation results by EchoCare and the ground truth.

planning and prognosis assessment. We evaluated different foundation models on three representative ultrasound images and clinical benchmarks for anatomical segmentation: the DDTI dataset [17] for thyroid node segmentation, the Mus-V dataset [18] for arterial-venous vessel segmentation, and the abdomen multi-organ segmentation.

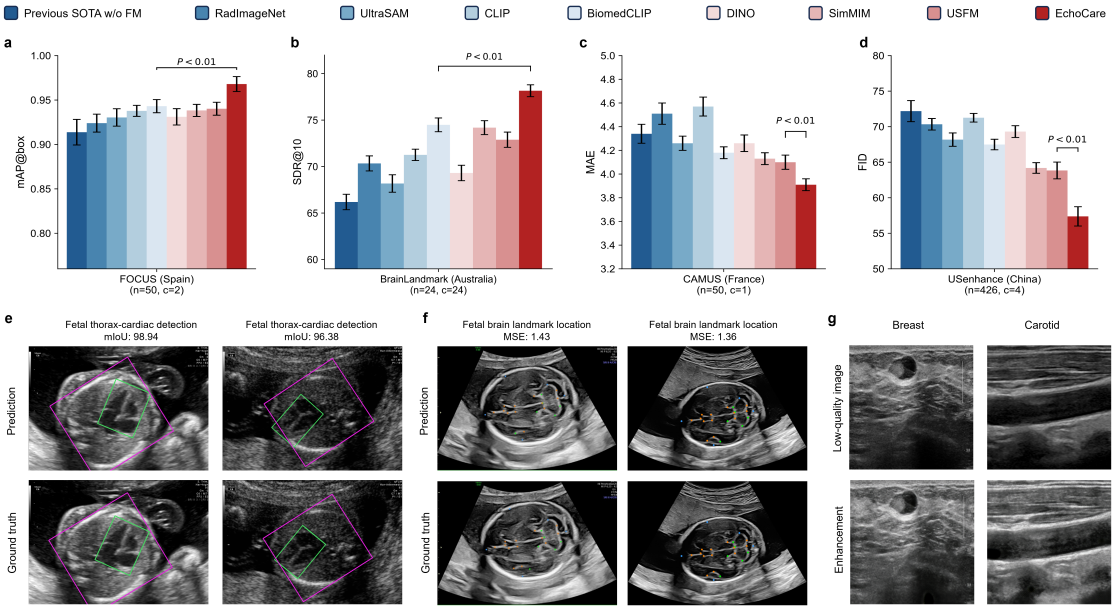

Figure 6: **Evaluation on organ detection, landmark prediction, fraction regression and imaging enhancement. a-d.** EchoCare consistently outperforms pervious SOTA task-specific models (w/o FM) and existing foundation models across different tasks: organ detection (**a**), landmark prediction (**b**), fraction regression (**c**) and imaging enhancement (**d**). The two-sided Wilcoxon signed-rank test was used to assess the statistical differences between EchoCare and the second-best model. **e-f.** Six examples comparing the detection, location and imaging enhancement results by EchoCare and the ground truth.

Compared with existing methods, EchoCare achieved significantly higher performance (Fig. 5d-f), surpassing the next best Dice Similarity Coefficient (DSC) by 2.09% and Normalized Surface Dice (NSD) by 2.26% in the thyroid nodule segmentation task, mDSC by 1.36% and mNSD by 1.03% in the vessel segmentation task (Fig. 5d), mDSC by 2.10% and mNSD by 4.36% in the multi-organ segmentation task (Fig. 5f). In the vascular segmentation task, EchoCare outperforms the second-ranked model (SimMIM) by a remarkable margin (Fig. 5e), which holds clinical significance for real-time vascular interventions (*e.g.*, coronary procedures). Furthermore, EchoCare also surpassed previous SOTA without FM architecture (SwinUNETR [40]) in benchmark evaluations. We showed examples comparing EchoCare segmentation and the ground truth across multiple organs, demonstrating the generalizability of EchoCare (Fig. 5g). The strong performance of EchoCare on segmentation tasks represents a breakthrough for comprehensive abdominal assessments, as clinicians require simultaneous visualization of the liver, pancreas, and kidneys to detect pathological relationships (*e.g.*, liver lesions compressing adjacent organs). Such consistency across single-organ, vascular, and multi-organ tasks underscores EchoCare's capacity to learn generalizable ultrasound features for effective task adaptation.

4.5.3 Fetal cardiac organ detection

Fetal congenital heart disease (CHD) is a leading cause of infant mortality from birth defects, with an incidence reaching up to 8-10 cases per 1,000 live births [22, 41]. Survival rates, requirement for intensive medical care, and risk of developmental disabilities are contingent on the accuracy and timeliness of diagnosis. Thus, early and precise prenatal sonographic diagnosis of CHD has been shown to reduce the risk of perinatal morbidity and mortality. The four-chamber view in fetal echocardiography is a unique and essential tool for assessing CHD. Diagnosis in this view relies on the cardiothoracic diameter ratio (CTR), a biometric defined

as the ratio of thoracic to cardiac short-axis diameters. Therefore, detecting thoracic and cardiac regions from four-chamber echocardiograms is a critical step for CTR analysis and represents a foundational step in CHD diagnosis.

In this study, we evaluated the performance of EchoCare against seven state-of-the-art foundation models and one previous SOTA model (Rotated Faster R-CNN [42]) for fetal thorax and cardiac organ detection using the publicly available FOCUS dataset (300 four-chamber fetal echocardiography ultrasound images). EchoCare outperformed all other models significantly (Fig. 6a). Specifically, it achieved $94.12\%\pm0.71\%$ DSC and $97.26\%\pm0.53\%$ AP (Average Precision) for thoracic object detection, and $93.91\%\pm0.82\%$ DSC and $96.11\%\pm0.64\%$ AP for cardiac object detection, surpassing the second-best model (USFM) by 1.7% mDSC and 2.78% mAP in average detection. Compared to the top ImageNet-based model (Faster-RCNN), EchoCare showed even larger margins (6.22% higher mDSC and 5.42% higher mAP). This outcome underscores that ultrasound, specific pretraining, distinct from natural image pretraining, more effectively captures the domain-specific knowledge inherent to ultrasound imaging. Benefiting from its high detection accuracy, EchoCare also ranked first in CTR measurement accuracy ($94.42\pm1.23\%$), outperforming USFM by 2.78% in P_{CTR}. We provide examples to visualize the detection results of EchoCare and demonstrate the superior performance by comparing with the ground truth (Fig. 5e). These comprehensive results demonstrate that EchoCare has the potential to enhance and accelerate prognosis prediction for CHD in ultrasound clinical practice.

4.5.4 Fetal brain landmark predication

Brain development involves progressive structural changes from early embryonic stages to several months after birth. Identifying fetal brain structures in ultrasound images enables assessment of cortical and subcortical gray matter changes, serving as a valuable tool for detecting developmental abnormalities. However, manual landmark identification is labor-intensive, time-consuming, and prone to intra- and inter-rater inconsistency.

To address this challenge, we evaluated the performance of EchoCare against other pretrained foundation models for predicting fetal brain landmarks using the publicly available BrainBenchmark dataset [23] (104 2D fetal brain ultrasound images acquired at 20-20.6 weeks of gestation from 70 pregnant women). EchoCare outperformed all foundation models significantly (Fig. 6b), achieving a notably lower average MSE (7.71) compared to the second-best model (SimMIM, 8.39). EchoCare also dominated in successful detection rate (SDR) across all pixel thresholds: at $\tau = 2.0$ pixels, it achieved an SDR of 36.27, (surpassing the second-best model SimMIM's 30.24); at $\tau = 4.0$ pixels, it achieved 49.13 (substantially exceeding SimMIM's 42.87); and at $\tau = 10.0$ pixels, it achieved 80.16 (versus BiomedCLIP's 74.49). We also provide examples to visualize the landmark prediction results of EchoCare and compare with the ground truth (Fig. 5f). These results highlight EchoCare's superiority in ultrasound-based landmark prediction, positioning it as a promising tool for automated fetal brain assessment.

4.5.5 Cardiac ejection fraction regression

The assessment of Left Ventricular Ejection Fraction (LVEF) is one of the most important manners in the evaluation of cardiac function. It quantifies the proportion of blood ejected from the left ventricle relative to its total end-diastolic volume. In clinical settings, accurate measurement of LVEF is pivotal to the early diagnosis of both congenital and acquired cardiovascular disorders, informs therapeutic decision-making, and enables robust prognostic stratification.

After observing the superior performance of EchoCare across a range of ultrasound clinical tasks, we further evaluated it on the LVEF regression task using the CAMUS benchmark dataset [24]. This dataset encompasses 2D apical four-chamber and two-chamber view sequences from 500 patients. Model performance is quantified by mean absolute error (MAE) with standard error. EchoCare exhibited superior performance and outperformed the other 10 competing approaches (Fig. 6c). It achieved the lowest MAE of 3.91, surpassing the second-best pretrained model (USFM) by a 19% reduction in MAE. Notably, it significantly outperformed the echo-specific state-of-the-art model (EchoMEM), with a significant 43% reduction in MAE. These

Table 3: **Performance of foundation models for ultrasound report generation on the USData Liver dataset.** Models are fine-tuned for cross-modal representation learning using paired ultrasound images and expert-written reports. Results are reported with mean and std over ten metrics.

Dataset	Organ Type	Metrics	Ultrasound domain		Image-Text domain		ImageNet domain		Ultrasound domain	
			RadImageNet	UltraSAM	CLIP	BiomedCLIP	DINO	SimMIM	USFM	EchoCare
USData	Liver	BELU-1 (↑)	81.14±0.33	80.70±0.34	81.20±0.43	83.24 ±0.22	83.41 ±0.36	83.59 ±0.13	81.34 ±0.66	84.58 ±0.25
		BELU-2 (↑)	76.32±0.47	75.73±0.48	76.43±0.60	79.07 ±0.27	79.22 ±0.46	79.42 ±0.15	76.62 ±0.92	82.05 ±0.20
		BELU-3 (↑)	73.18±0.56	72.48±0.55	73.31±0.69	76.17 ±0.30	76.30 ±0.50	76.48 ±0.16	73.51 ±1.06	80.07 ±0.17
		BELU-4 (↑)	70.75 ±0.63	69.96±0.59	70.89±0.75	73.88 ±0.31	73.97 ±0.53	74.15 ±0.19	71.09 ±1.14	78.47 ±0.14
		METEOR (↑)	50.14 ±0.31	49.73±0.25	50.22±0.36	52.58 ±0.28	52.72 ±0.39	53.00 ±0.16	50.30 ±0.55	51.62 ±0.12
		ROUGE-L (↑)	76.87 ±0.47	76.28±0.45	77.00±0.59	79.80 ±0.32	79.99 ±0.50	80.37 ±0.17	77.17 ±0.89	86.18 ±0.23
		CE@ACC (↑)	42.82±0.91	42.32±0.72	42.36±0.97	47.50 ±2.42	47.64 ±1.67	48.43 ±3.67	43.00 ±1.17	57.82 ±0.12
		CE@PPR (↑)	60.98±1.21	60.74±3.12	62.18±1.45	70.77 ±2.69	72.79 ±2.99	72.26 ±1.87	60.08 ±1.44	87.30 ±0.05
		CE@SEN (↑)	68.76±0.59	68.71±0.81	69.04±0.74	70.94 ±1.50	71.52 ±1.15	71.12 ±2.50	69.07 ±0.59	91.76 ±0.06
		CE@F1 (↑)	64.20±0.68	63.58±0.52	64.91±0.95	69.70 ±1.62	70.53 ±1.36	70.60 ±1.94	63.67 ±0.50	89.30 ±0.01

contributions underscore the potential of EchoCare to advance cardiac LVEF regression and its applicability in real-world clinical workflows.

4.5.6 Low-quality imaging enhancement

High-quality ultrasound imaging is critical for the accurate identification of anatomical structures and disease diagnosis. However, ultrasound examinations using handheld or low-end devices often yield suboptimal images that compromise clinical diagnosis, particularly in resource-limited hospitals or regions. Enhancing such low-quality ultrasound images using AI technologies, for example, through improved contrast, sharpness, and signal-to-noise ratio, alongside noise reduction, could provide a cost-effective alternative to high-end scanners. This approach may also promote the wider adoption of portable ultrasound systems, offering substantial clinical benefits and ultimately improving patient outcomes.

We evaluated EchoCare on the low-quality ultrasound image enhancement task using the USenhance benchmark dataset [25], which encompasses real-world clinical scans from 109 patients across five anatomical regions: thyroid, kidney, liver, breast, and carotid artery. EchoCare was compared with 8 models, including previous SOTA model (EnlightenGAN [43]), ultrasound-based models (RadImageNet [34], UltraSAM [35]), image-text multimodal models (CLIP [36], BiomedCLIP [37]), and self-supervised frameworks (DINO [38], SimMIM [39], USFM [8]). Consistent with previous findings, EchoCare outperformed all competing models across four metrics: NIQE, BRISQUE, PIQE, and FID (Fig. 6d). Specifically, EchoCare achieved mean NIQE, BRISQUE, PIQE, and FID values of $6.35\%\pm1.13\%$, $17.62\%\pm2.15\%$, $30.16\%\pm1.34\%$, and $57.38\%\pm2.36\%$, respectively. These visualizations (Fig. 6g) further demonstrate the superior image quality enhancement ability of EchoCare. These results demonstrate that EchoCare can effectively enhance low-quality ultrasound images, highlighting the potential of AI for practical clinical applications in resource-limited settings.

4.5.7 Clinical report generation

Report generation is essential for healthcare system, providing critical information to clinicians and patients for the diagnosis, prognosis, and treatment planning of a wide range of medical applications. Traditionally, ultrasound reports are written manually by radiologists or sonographers, which is time-consuming and prone to inter-observer variability. Recent advancements in natural language processing and medical image analysis have enabled the development of automated ultrasound report generation systems.

To evaluate the effectiveness of our developed foundation model in ultrasound report generation, we integrate EchoCare into an existing Transformer-based encoder–decoder report generator, where the input is

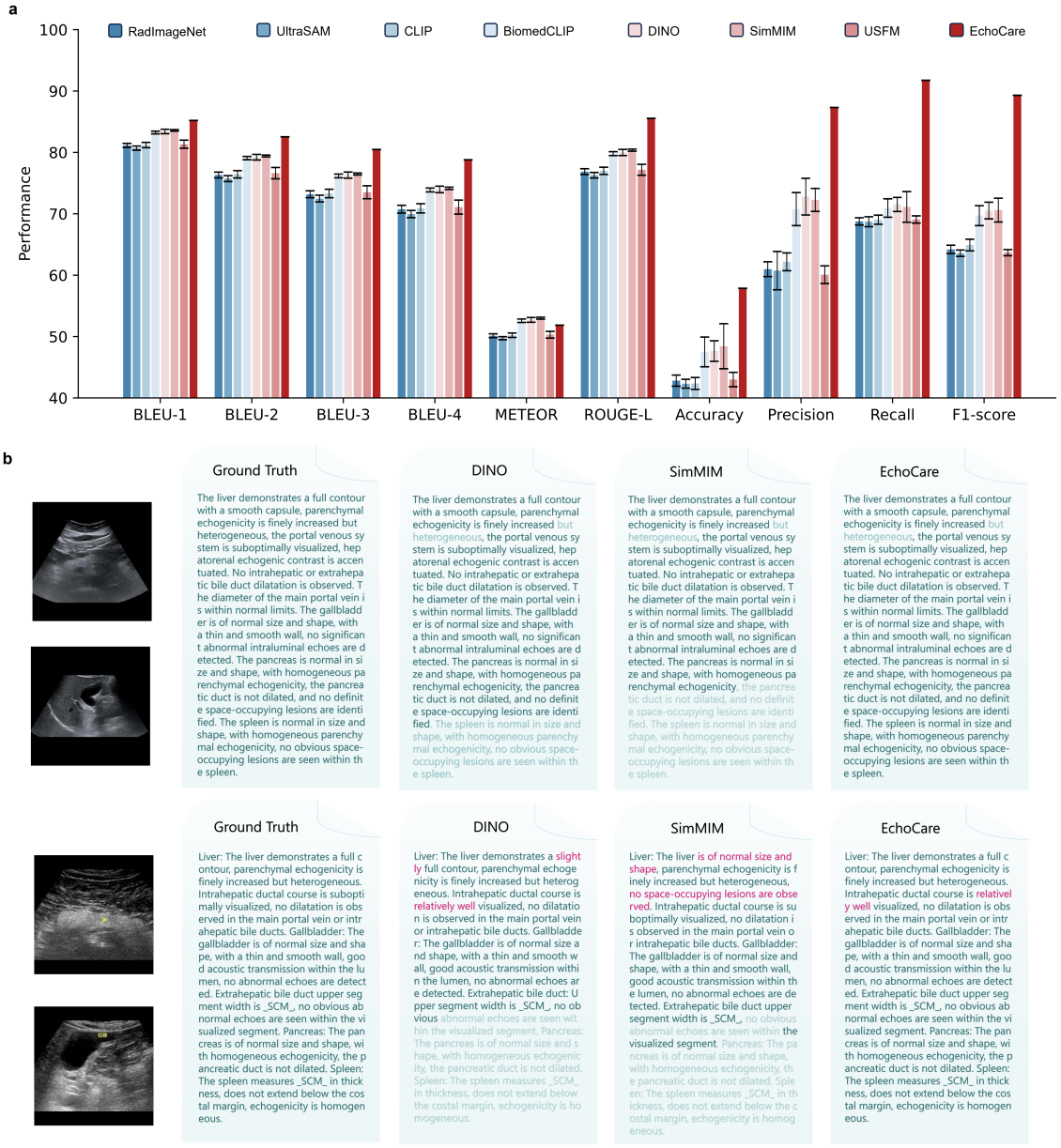

Figure 7: **Clinical report generation on USData [26] Liver dataset. a.** Performance (%) comparison of models on the USData [26] Liver dataset using six language metrics (BLEU-1 to BLEU-4, METEOR, and ROUGE-L) and four classification metrics (Accuracy, Precision, Recall, and F1-score). Error bars denote standard deviation across multiple runs. **b.** Example reports generated by the two strongest baseline models (DINO [38] and SimMIM [39]) and EchoCare, compared against ground truth reports. Deep blue text indicates exact matches, light-colored text denotes missing segments, and vivid purple highlights over-generated content.

the global visual features extracted from ultrasound images. The integrated model is then fine-tuned on the USData Liver dataset [26], which contains paired ultrasound images and corresponding expert-written reports. The experimental results (Fig. 7a) demonstrate that EchoCare achieved the best performance across all ten

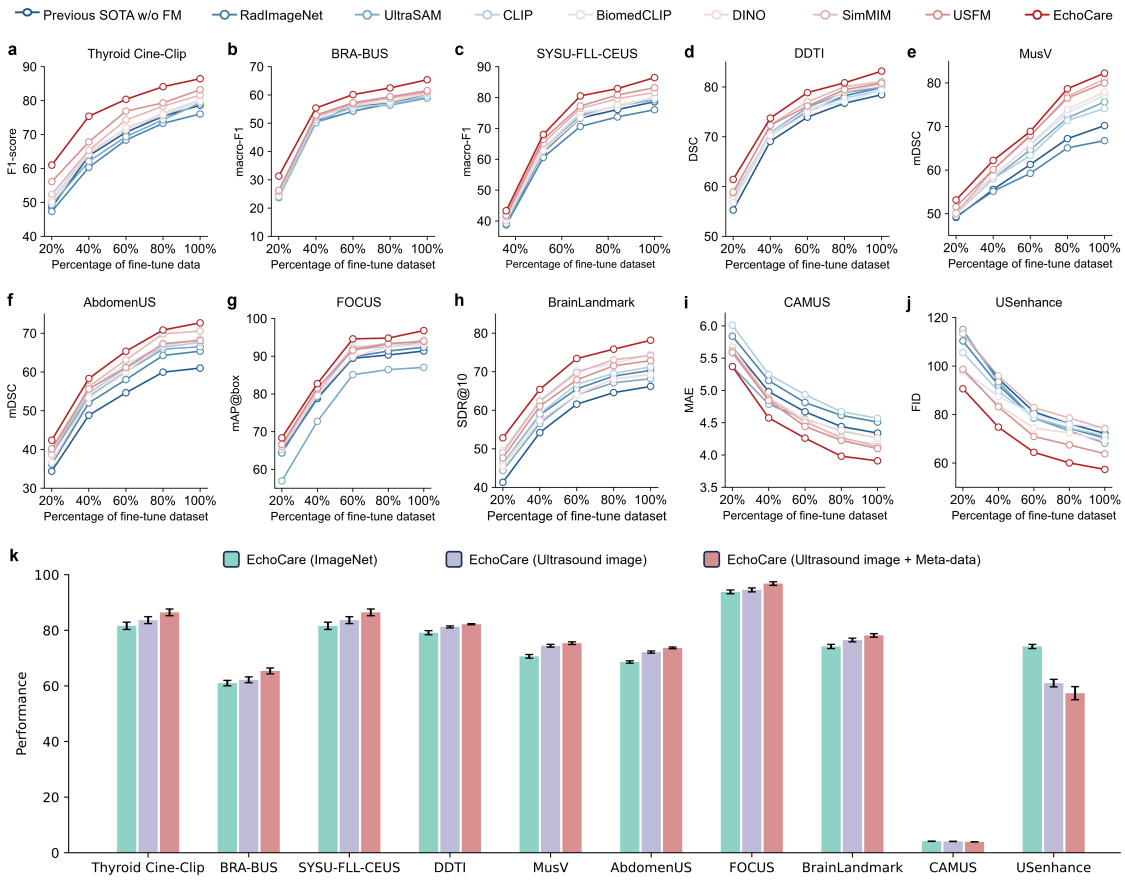

Figure 8: **Label efficiency and further analysis results. a-j.** Comparison between EchoCare and other models (previous SOTA w/o FM and existing foundation models) in label efficiency generalization on ten clinical applications, showing performance at various training data percentages. **k.** Pretrained with large-scale ultrasound images, EchoCare significantly improved performance (%) over models based on natural image pretraining on ten clinical applications. Moreover, with our designed model dual-branch architecture, the performance can be further enhanced.

metrics. Notably, compared with the second-best model (SimMIM [39]), EchoCare outperformed by large margins of 4.32% (BLEU-4), 9.39% (accuracy) and 18.70% (F1-score) (Table. 4). Besides, the examples of generated reports (Fig. 7b) valid the ability of EchoCare for ultrasound report generation, demonstrating the potential of EchoCare in improving the efficiency, consistency, and accessibility of automatic clinical report generation.

5 Discussion

Ultrasound imaging is a crucial tool in modern medicine. This work presents a novel, open-source foundation model named EchoCare to advance general-purpose clinical ultrasound applications. The model is pre-trained on our curated, publicly available dataset of 4.5 million ultrasound images, featuring a highly diverse and balanced distribution of images sourced from over 20 countries and 5 continents. To evaluate EchoCare's clinical utility, we conducted comprehensive validations across a wide range of downstream ultrasound tasks

(lesion segmentation, disease diagnosis, organ detection, landmark prediction, quantitative regression, imaging enhancement and report generation). The results demonstrate the strong effectiveness and generalization capabilities of EchoCare: it consistently outperforms state-of-the-art foundation models such as UltraSAM [35], BiomedCLIP [37], and USFM [8], achieving an average performance gain of 3.1% over the second-best model across all tasks.

The pre-training of our foundation model is powered by EchoAtlas, the largest publicly available ultrasound image dataset to date, featuring over 4.5 million images. The vast scale and diverse, balanced distribution of this dataset are crucial to our success. The core idea behind our data strategy is simple yet powerful: by aggregating numerous public datasets, we can significantly increase data size, expand protocol coverage, and diversify patient populations. This approach allows our model to learn from a broad spectrum of global sources. This extensive pre-training also grants EchoCare a remarkable degree of label efficiency for various downstream clinical tasks (Fig. 8a-j), thereby alleviating the substantial annotation workload for medical experts. For instance, in thyroid nodule segmentation, EchoCare can outperform other models using only 60% of the labeled training data. Furthermore, EchoCare showed consistently high adaptation efficiency, suggesting that EchoCare required less time in adapting to downstream clinical applications, *e.g.*, EchoCare can potentially save about $20\% \sim 40\%$ of the training time required to achieve convergence for the task of disease prediction.

In addition to the substantial size and broad diversity of EchoAtlas, the designed dual-branch architecture further contributes to the superior performance of EchoCare across a wide range of downstream clinical tasks. Unlike previous medical foundation models that relied on standard MAE structures, our enhanced MAE architecture incorporates a unique anatomy-classifier branch. This branch is designed to learn global and hierarchical anatomical relationships, mirroring a clinician's diagnostic process. By integrating this high-level, representation-based learning with the local, pixel-level prediction of the MAE, our model's encoder gains a deeper understanding of ultrasound images. This dual-learning approach significantly boosts the model's ability to interpret images and perform well in a wide range of downstream clinical applications (Fig. 8k).

While EchoCare has demonstrated promising potential of pretrained foundations for ultrasound analysis, several methodological frontiers remain. First, current pretraining exclusively uses image data, omitting clinically actionable text modalities (*e.g.*, ultrasound diagnostic reports). Future iterations will integrate vision-language learning through curated datasets, enabling joint modeling of ultrasound images and associated clinical narratives to expand clinical applications. Second, EchoCare currently treats dynamic modalities (*e.g.*, videos) to static frames, thus failing to utilize the temporal cues essential for applications like cardiac motion analysis or vascular flow assessment. We will extend the architecture to incorporate spatio-temporal transformers, enabling end-to-end training on native video sequences and preserving temporal dynamics. Third, although results across other downstream clinical tasks demonstrate translational potential, rigorous validation is required before clinical adoption, such as real-world integration with clinical decision support systems.

6 Conclusion

In conclusion, we introduce EchoCare, a novel vision foundation model for ultrasound analysis, pretrained on our curated EchoAtlas, which comprises over 4.5 million ultrasound images and is the largest ultrasound dataset to date. By integrating a novel architecture with a massive, diverse dataset, EchoCare establishes an efficient new paradigm for ultrasound image analysis, demonstrating robust adaptability to a broad spectrum of clinical ultrasound tasks and delivering significant performance gains over existing foundation models. Critically, we have made both the EchoCare model and EchoAtlas publicly accessible to accelerate advancements in medical AI, improving clinical decision-making and patient care.

Table 4: **Transfer learning performance of different foundation models on 10 clinical applications.** This evaluation encompassed seven task types: seven task types of segmentation, classification, detection, location, regression, image enhancement and report generation. Results are reported with mean and std.

Task type	Metrics	SOTA w/o FM	Ultrasound domain		Image-Text domain		ImageNet domain		Ultrasound domain	
		SwinUNETR	RadImageNet	UltraSAM	CLIP	BiomedCLIP	DINO	SimMIM	USFM	EchoCare
Node segmentation	DSC (↑)	78.44±0.85	79.87±0.83	79.93±1.02	79.40±1.23	79.43±0.83	80.46±1.10	81.08±0.74	80.78±0.39	83.17±0.19
	NSD (↑)	82.55±0.86	82.90±0.86	83.29±0.92	80.81±0.86	82.56±0.79	83.32±0.43	84.16±0.72	83.11±0.43	86.59±0.21
Vessel segmentation	mDSC (↑)	70.20±0.68	66.76±1.24	75.64±0.62	74.15±1.52	77.15±0.82	77.92±1.61	80.86±0.66	79.95±0.63	82.24±0.48
	mNSD (↑)	84.74±0.57	80.94±1.38	85.68±0.46	85.59±1.28	87.37±0.53	87.23±1.39	88.06±0.58	88.03±0.44	90.53±0.36
Organ segmentation	mDSC (↑)	61.01±0.38	65.37±0.53	66.52±0.73	67.62±0.36	68.59±0.32	68.33±0.31	70.58±0.43	68.12±0.24	72.68±0.31
	mNSD (↑)	70.65±0.38	72.64±0.35	73.75±0.74	74.65±0.40	77.23±0.30	74.12±0.29	74.32±0.44	78.36±0.26	82.84±0.35
		ViT	RadImageNet	UltraSAM	CLIP	BiomedCLIP	DINO	SimMIM	USFM	EchoCare
Node classification	ACC (↑)	81.34±1.29	79.34±1.61	80.16±1.20	80.33±1.23	80.78±0.83	79.64±1.01	82.37±1.04	83.13±1.36	86.48±1.19
	F1-score (↑)	78.64±1.37	76.06±1.29	79.28±1.38	79.94±0.92	80.26±1.09	79.83±1.76	81.61±1.31	83.20±1.13	87.45±1.21
BI-BADS classification	ACC (↑)	65.40±1.26	64.93±1.22	64.17±1.44	66.27±1.31	66.13±1.03	66.58±1.42	66.74±1.66	67.27±1.03	70.36±1.01
	macro-F1 (↑)	60.45±1.56	59.45±1.12	58.86±1.46	59.58±1.01	60.49±1.23	60.63±1.63	61.01±0.98	61.53±1.35	65.38±1.06
Lesion classification	ACC (↑)	82.61±1.41	81.67±0.83	83.69±0.96	82.66±1.03	82.12±1.23	81.93±1.13	83.61±0.82	83.67±0.98	87.12±0.91
	macro-F1 (↑)	78.39±1.14	77.26±0.86	79.37±1.08	78.16±0.93	78.23±1.03	76.73±0.92	79.12±0.71	79.46±0.86	83.44±0.95
		Faster R-CNN	RadImageNet	UltraSAM	CLIP	BiomedCLIP	DINO	SimMIM	USFM	EchoCare
Organ detection	mAP@box (↑)	91.38±1.44	92.40±1.01	87.04±0.98	93.78±0.62	94.31±0.74	93.11±0.92	93.83±0.68	94.02±0.73	96.80±0.64
	CTR (↑)	87.83±1.61	88.45±1.38	84.11±1.45	89.71±1.13	91.17±1.33	90.23±1.37	91.51±1.20	91.64±1.03	94.42±1.23
		ViTPose	RadImageNet	UltraSAM	CLIP	BiomedCLIP	DINO	SimMIM	USFM	EchoCare
Landmark location	MAE (↑)	10.27±0.92	9.51±0.65	9.75±0.76	9.21±0.73	8.89±0.62	9.11±0.91	8.39±0.74	8.87±0.93	7.71±0.79
	SDR@2 (↑)	19.20±0.85	25.64±0.79	25.39±0.83	26.86±0.92	29.15±0.71	27.15±0.96	30.24±0.94	29.52±0.74	36.27±0.75
	SDR@4 (↑)	32.50±0.89	36.85±0.65	34.43±0.77	38.48±0.72	42.34±0.50	38.73±0.61	42.87±0.76	41.29±0.77	49.13±0.82
	SDR@10 (↑)	66.19±0.83	70.34±0.81	68.18±0.94	71.26±0.61	74.49±0.74	69.31±0.83	74.19±0.75	72.89±0.82	80.16±0.64
		EchoNet	RadImageNet	UltraSAM	CLIP	BiomedCLIP	DINO	SimMIM	USFM	EchoCare
EF regression	MAE (↓)	4.34±0.08	4.51±0.09	4.26±0.06	4.57±0.08	4.18±0.05	4.26±0.07	4.13±0.05	4.10±0.06	3.91±0.05
		EnlightenGAN	RadImageNet	UltraSAM	CLIP	BiomedCLIP	DINO	SimMIM	USFM	EchoCare
Image enhancement	NIQE (↓)	6.83±1.49	9.51±0.65	9.75±0.76	9.21±0.73	8.89±0.62	9.11±0.91	8.39±0.74	6.81±1.68	6.35±1.13
	BRISQUE (↓)	20.68±2.72	25.64±0.79	25.39±0.83	26.86±0.92	29.15±0.71	27.15±0.96	30.24±0.94	21.86±2.47	17.62±2.15
	PIQE (↓)	30.29±1.28	36.85±0.65	34.43±0.77	38.48±0.72	42.34±0.50	38.73±0.61	42.87±0.76	41.29±0.77	30.16±1.34
	FID (↓)	62.19±2.48	70.34±0.81	68.18±0.94	71.26±0.61	74.49±0.74	69.31±0.83	74.19±0.75	63.84±3.18	57.38±2.36

Table 5: **The information of dataset assembly.** Our work provides detailed annotations for each ultrasound volume, systematically documenting critical metadata that was absent in prior datasets. This includes specific acquisition equipment (*e.g.*, Philips EPIQ, Siemens ACUSON), ensuring transparency and reproducibility in dataset composition.

Dataset	organs	cases	images	source countries	ultrasound scanners
1. UIC [44, 45]	1	-	2206	GB & US	-
2. CARDIAC [46]	1	-	100	US	Philips iE33/Sonos/EPIQ 5G/EPIQ 7c Siemens ACUSON SC2000
3. CardiacUDA [47]	1	100	24041	CN	Philips HIDACHI
4. BUSI [48]	1	600	1707	EG	GE LOGIQ E9
5. UBIBC [49]	1	-	509	-	-
6. US3M [50]	1	248	1034	CN	Philips & Siemens
7. LRHR [51]	4	-	1523	-	-
8. 2ULM [52]	1	-	683	-	-
9. BrEaST [53]	1	256	248	PL	Hitachi ARIETTA 70 Samsung RS85 Philips Affiniti 70G/EPIQ 5G Esaote
10. MBUD [54]	1	-	250	-	-
11. UDIAT [55]	1	-	163	ES	Siemens ACUSON Sequoia C512
12. OMI [56]	1	-	230	-	-
13. BUET [57]	1	223	261	BD	GE SonixTouch Research
14. BUSI_WHU [58]	1	-	927	CN	-
15. STU [59]	1	-	43	-	-
16. BUDataset [60, 61, 62]	1	232	232	IR	Supersonic Imagine AixPlorer Ultimate
17. LOGIQ [63]	1	192	201	-	GE LOGIQ P7
18. JNU [64]	1	51	4880	CN	Youkey D8
19. HC18 [65]	1	551	999	NL	GE Voluson E8/730
20. DVD [66]	1	-	373	-	-
21. PSFHS [67]	1	-	1358	-	-
22. CAUCD [68]	1	18	129	IN	-
23. HR [69]	1	-	1623	-	-
24. USL [70, 71]	1	7	41111	CH	Siemens
25. S [72]	1	-	314	-	-
26. AULI [73, 74]	1	11468	735	CN	Hitachi Preirus Noblus Aloka a10/a7/3500 Philips EPIQ7/Affiniti50/HD7/IU22 SonoScape S60/S40 GE LOGIQ E9 Mindray Resona 7/DC-8 Esaote MyLab90/MyLab40/MyLab60 Toshiba Apilo500 Siemens S3000/Sequoia512 Supersonic Aixplorer
27. AUITD [75]	1	-	1353	DZ	-
28. OpenCAS [76]	1	-	4109	-	GE Logiq E9
29. TN3K [77, 78]	1	2421	3493	CN	-
30. ThySeg [79]	1	28	7918	DE	Siemens ACUSON NX-3
31. UIH [80]	1	-	321	-	-
32. DUI [81]	1	235	404	ES	Esaote MyLab Class C
33. SPJ [82]	1	-	51	CN	-
34. Butterfly [83]	1	20	1533	-	Butterfly
35. TFFPU [84]	1	-	14802	-	-
36. KUI [85]	1	-	9416	-	SAMSUNG RS85/HS60/RS80A/HS70A
37. NDF [86]	1	-	513	-	-
38. UNS [87]	1	-	5635	-	-
39. UI [88]	1	-	17034	-	-
40. DUPI [89]	2	-	500	-	-
41. TDUPI [90]	2	-	97	-	-
42. PCOS [91]	1	-	13	-	-
43. 2000 [92]	1	-	96	-	-
44. OCD [93]	1	469	648	-	-
45. MMOTU [94]	1	294	1639	CN	Mindray Resona8
46. EBUS [95]	1	4	1021	NO	Olympus EUS Exera EU-C60
47. 3VG [96]	1	100	253	RO	GE Voluson 730 Pro/LOGIQ e
48. FUCD [97]	1	-	704	CN	Verasonics L11-5v/L22-14vX
49. ReMIND2Reg [98]	1	114	952	US	GE N13C5/BK5000 Brainlab AG BK N13C5
50. CuRIOUS [99]	1	23	1752	NO	Sonowand AS

Dataset	organs	cases	images	source countries	ultrasound scanners
51. CAT [100]	1	5	2973	IE	Terason T3000
52. CC [101]	1	58	9028	PL	Philips HD15
53. MUPSD [102]	1	75	1028	US	-
54. TRUS [103]	1	141	12000	GB	Hitachi HI VISION Preirus
55. SPLOA [104]	1	-	900	-	-
56. LN [105]	1	-	1637	-	-
57. PAD [106]	4	579	1922	DE	Toshiba Xario/Aplio XG
58. TUS [107, 108, 109, 110]	1	19	40938	GB	BK
59. AA [111]	1	300	6620	SL	Telemed MicrUs Pro-C60S
60. FetalAbdomenSeg [112]	1	169	1588	BR	Siemens ACUSON GE Voluson 730 Philips EPIQ Elite
61. PMUB [113]	1	1151	509646	US	Hitachi Hi-Vision 5500/Noblus
62. LEP [114]	1	420	3500	CN	Olympus PENTAX Fujifilm Aloka
63. BBDTD [115]	1	-	256423	-	SonoScape E1
64. CoA [116]	1	53	200	CN	GE Vivid7 & Philips IE33
65. MVSeg [117]	1	-	16884	CA & GB	Philips EPIQ/iE33
66. ArteryUS [118]	1	11	1100	-	Mindary UMT-500Plus
67. ERUS [119]	1	77	10006	CN	CANNO
68. UBPD [120]	1	101	1052	CN	Siemens ACUSON NX3 Elite Philips EPIQ5
69. LUMINOUS [121]	1	109	341	US	GE LOGIQ e
70. FETALPLANE [122]	1	1792	7476	ES	GE Voluson E6/S8/S10 Aloka
71. MUST [123]	2	1283	8169	NL	Esaote MyLab Twice
72. DeepACSA [124]	2	77	20000	-	Siemens ACUSON Juniper SuperSonic Imagine Aixplorer Ultimate Esaote MyLab 70
73. FALLMUD [125]	1	-	813	GB	Aloka SSD-5000 PHD
74. ATD [126]	1	1	801	GB	Telemed LS128 Analogic SonixTouch
75. Leg-3D-US [127]	1	28	52795	DE	SuperSonic Imagine Aixplorer
76. SlicerIGT [128]	1	8	706	US	Telemed MicrUs EXT-1H

BD: Bangladesh BR: Brazil CA: Canada CH: Switzerland CN: China DE: Germany DZ: Algeria EG: Egypt ES: Spain FR: France GB: United Kingdom IE: Ireland IN: India IR: Iran NL: Netherlands NO: Norway PL: Poland RO: Romania SL: Sierra Leone US: United States

Table 6: **Download links of the 76 datasets in EchoAtlas.**

Dataset	Download Link
AA [111]	https://zenodo.org/records/12697994
ArteryUS [118]	https://data.mendeley.com/datasets/d4xt63mgjm/1
ATD [126]	https://zenodo.org/records/4989216
AUITD [75]	https://www.kaggle.com/datasets/azouzmaroua/algeria-ultrasound-images-thyroid-dataset-auitd
AULI [73]	https://zenodo.org/records/7272660
BBDTD [115]	https://zenodo.org/records/7081639
BrEaST [53]	https://www.kaggle.com/datasets/magpiesings/breast-lesions-ultrasound-clinical
BUDataset [60, 61, 62]	https://qamebi.com/breast-ultrasound-images-database/
BUET [57]	https://www.kaggle.com/datasets/jarintasnim090/buet-breast-ultrasound-data
BUSI [48]	https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset
BUSI_WHU [58]	https://data.mendeley.com/datasets/k6cpmwybk3/1
Butterfly [83]	https://github.com/jannisborn/covid19l_ultrasound/tree/master/data
CARDIAC [46]	https://humanheart-project.creatis.insa-lyon.fr/database/
CardiacUDA [47]	https://www.kaggle.com/datasets/xiaoweixumedicalai/cardiacudc-dataset
CAUCD [68]	https://www.kaggle.com/datasets/pahunichoudhary/carotid-artery-ultrasound-and-color-doppler
CAT [100]	https://zenodo.org/records/4934835
CC [101]	https://zenodo.org/records/5147854
CoA [116]	https://zenodo.org/records/4960642
CuRIOUS [99]	https://archive.norstore.no/pages/public/datasetDetail.jsf?id=10.11582/2017.00004
DeepACSA [124]	https://zenodo.org/records/5799204
DUI [81]	https://www.kaggle.com/datasets/alfageme/dermatologic-ultrasound-images
DUPI [89]	https://zenodo.org/records/5110223
DVD [66]	https://www.kaggle.com/datasets/yasinelh/ultrasoundvd
EBUS [95]	https://zenodo.org/records/4991954
ERUS [119]	https://arxiv.org/abs/2408.10067
FALLMUD [125]	https://kalisteo.cea.fr/index.php/fallmud/
FetalAbdomenSeg [112]	https://data.mendeley.com/datasets/4gcpm9dsc3/1
FETALPLANE [122]	https://zenodo.org/records/3904280
FUCD [97]	https://zenodo.org/records/14511961
HC18 [65]	https://zenodo.org/records/1322001
HR [69]	https://www.kaggle.com/datasets/priyeshk/carotid-artery-ultrasound-scans-hr-image
JNU [64]	https://figshare.com/articles/dataset/JNU-IFM/14371652
KUI [85]	https://www.kaggle.com/datasets/gurjeetkaurmangat/kidney-ultrasound-images-stone-and-no-stone
LEP [114]	https://zenodo.org/records/8041285
Leg-3D-US [127]	https://www.cs.cit.tum.de/camp/publications/leg-3d-us-dataset/
LN [105]	https://zenodo.org/records/12702916
LOGIQ [63]	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0253202#sec005
LRHR [51]	https://www.kaggle.com/datasets/chirag2466/ultra-lr-hr-ultrasound-image-dataset-for-research
LUMINOUS [121]	https://users.encs.concordia.ca/ impact/luminous-database/

Table 7: (Continued) Download links of the 76 datasets in EchoAtlas.

Dataset	Download Link
MBUD [54]	https://www.kaggle.com/datasets/jarintasnim090/mendeley-dataset
MMOTU [94]	https://github.com/cv516Buaa/MMOTU_DS2Net
MUPSD [102]	https://zenodo.org/records/10475293
MVSeg [117]	https://www.synapse.org/Synapse:syn52383425
NDF [86]	https://www.kaggle.com/datasets/ahmadshtaiyat/new-data-f
OCD [93]	https://www.kaggle.com/datasets/turkertuncer/ovarian-cyst-dataset
OMI [56]	https://www.kaggle.com/datasets/jarintasnim090/omi-dataset
OpenCAS [76]	http://opencas.webarchiv.kit.edu/data/thyroid.zip
PAD [106]	https://zenodo.org/records/7711412
PCOS [91]	https://www.kaggle.com/datasets/anaghachoudhari/pcos-detection-using-ultrasound-images
PMUB [113]	https://www.cancerimagingarchive.net/collection/prostate-mri-us-biopsy/
PSFHS [67]	https://zenodo.org/records/10969427
ReMIND2Reg [98]	https://zenodo.org/records/12700312
S [72]	https://www.kaggle.com/datasets/tunkedsaro/sliver
SlicerIGT [128]	TO DO
SPJ [82]	https://github.com/hidden-ops/NHBS-Net_SPJ_dataset
SPLOA [104]	https://zenodo.org/records/13195053
STU [59]	https://github.com/xbhlk/STU-Hospital
TDUPI [90]	https://zenodo.org/records/5110198
TFFPU [84]	https://www.kaggle.com/datasets/bachaboos/tf-for-pocovid-ultrasound
ThySeg [79]	https://www.cs.cit.tum.de/camp/publications/segthy-dataset/
TN3K [77, 78]	https://github.com/haifangong/TRFE-Net-for-thyroid-nodule-segmentation
TRUS [103]	https://zenodo.org/records/8004388
TUS [107, 108, 109, 110]	https://zenodo.org/records/7740734
UBIBC [49]	https://www.kaggle.com/datasets/vuppalaadithyasairam/ultrasound-breast-images-for-breast-cancer
UBPD [120]	TO DO
UDIAT [55]	https://www.kaggle.com/datasets/jarintasnim090/udiat-data
UIC [44, 45]	https://data.unityimaging.net/
UI [88]	https://www.kaggle.com/datasets/thegna/ultrasound-img
UIH [80]	https://www.kaggle.com/datasets/shengwang1130/ultrasound-image-set-of-hemangioma3
UNS [87]	https://www.kaggle.com/datasets/anupaankarigari/ultrasound-nerve-segmentation
US3M [50]	https://www.kaggle.com/datasets/timesxy/multimodal-breast-ultrasound-dataset-us3m
USL [70, 71]	https://zenodo.org/records/12697994
2000 [92]	https://www.kaggle.com/datasets/shnotweta/2000-images-of-ultrasound-for-pcos
2ULM [52]	https://www.kaggle.com/datasets/drjaveriaamin/ultrasound2ulm
3VG [96]	https://zenodo.org/records/7323401

References

[1] Peter NT Wells. "Ultrasound imaging". In: *Physics in medicine & biology* 51.13 (2006), R83.

[2] Yukun Zhou et al. "A foundation model for generalizable disease detection from retinal images". In: *Nature* 622.7981 (2023), pp. 156–163.

[3] Siyuan Yan et al. "A multimodal vision foundation model for clinical dermatology". In: *Nature Medicine* (2025), pp. 1–12.

[4] Richard J Chen et al. "Towards a general-purpose foundation model for computational pathology". In: *Nature medicine* 30.3 (2024), pp. 850–862.

[5] Xiyue Wang et al. "A pathology foundation model for cancer diagnosis and prognosis prediction". In: *Nature* 634.8035 (2024), pp. 970–978.

[6] DongAo Ma et al. "A fully open AI foundation model applied to chest radiography". In: *Nature* (2025), pp. 1–11.

[7] Jiabo Ma et al. "A generalizable pathology foundation model using a unified knowledge distillation pretraining framework". In: *Nature Biomedical Engineering* (2025), pp. 1–20.

[8] Jing Jiao et al. "Usfm: A universal ultrasound foundation model generalized to tasks and organs towards label efficient image analysis". In: *Medical image analysis* 96 (2024), p. 103202.

[9] Qingbo Kang et al. "URFM: a general Ultrasound Representation Foundation Model for advancing ultrasound image diagnosis". In: *iScience* 28.8 (2025).

[10] Hanwen Xu et al. "A whole-slide foundation model for digital pathology from real-world data". In: *Nature* 630.8015 (2024), pp. 181–188.

[11] Kaiming He et al. "Masked autoencoders are scalable vision learners". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pp. 16000–16009.

[12] "Zenodo". https://zenodo.org/.

[13] "Mendeley". https://www.mendeley.com/.

[14] "Kaggle". https://www.kaggle.com/datasets/bachaboos/tf-for-pocovid-ultrasound.

[15] "Github repository". https://github.com/.

[16] "Grand-challenge platform". https://grand-challenge.org/.

[17] Lina Pedraza et al. "An open access thyroid ultrasound image database". In: *10th International symposium on medical information processing and analysis*. Vol. 9287. SPIE. 2015, pp. 188–193.

[18] Yimeng Geng et al. "Force Sensing Guided Artery-Vein Segmentation via Sequential Ultrasound Images". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2024, pp. 656–666.

[19] Rikiya Yamashita et al. "Toward Reduction in False-Positive Thyroid Nodule Biopsies with a Deep Learning–based Risk-Stratification System Using US Cine-Clip Images". In: *Radiology: Artificial Intelligence* 4 (May 2022). DOI: 10.1148/ryai.210174.

[20] Wilfrido Gómez-Flores, Maria Gregorio-Calas, and Wagner Pereira. "BUS-BRA: A breast ultrasound dataset for assessing computer-aided diagnosis systems". In: *Medical Physics* 51 (Nov. 2023). DOI: 10.1002/mp.16812.

[21] Xiaodan Liang et al. "Recognizing Focal Liver Lesions in CEUS With Dynamically Trained Latent Structured Models". In: *IEEE transactions on medical imaging* 35 (Oct. 2015). DOI: 10.1109/TMI.2015.2492618.

[22] Songxiong Wu et al. *FOCUS: Four-chamber Ultrasound Image Dataset for Fetal Cardiac Biometric Measurement*. Version 1.0. Zenodo, 2025. DOI: 10.5281/zenodo.14597550. URL: https://doi.org/10.5281/zenodo.14597550.

[23] Mariano Cabezas et al. "A benchmark for 2D foetal brain ultrasound analysis". In: *Scientific Data* 11.1 (2024), p. 923.

[24] Sarah Leclerc et al. "Deep learning for segmentation using an open large-scale dataset in 2D echocardiography". In: *IEEE transactions on medical imaging* 38.9 (2019), pp. 2198–2210.

[25] Yi Guo et al. “Ultrasound Image Enhancement challenge 2023”. In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2023*. Zenodo, 2023. DOI: 10.5281/zenodo.7841250. URL: https://doi.org/10.5281/zenodo.7841250.

[26] Jun Li et al. “Ultrasound Report Generation with Cross-Modality Feature Alignment via Unsupervised Guidance”. In: *IEEE Transactions on Medical Imaging* (2024).

[27] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. “Making a “completely blind” image quality analyzer”. In: *IEEE Signal processing letters* 20.3 (2012), pp. 209–212.

[28] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. “No-reference image quality assessment in the spatial domain”. In: *IEEE Transactions on image processing* 21.12 (2012), pp. 4695–4708.

[29] Narasimhan Venkatanath et al. “Blind image quality evaluation using perception based features”. In: *2015 twenty first national conference on communications (NCC)*. IEEE. 2015, pp. 1–6.

[30] Martin Heusel et al. “Gans trained by a two time-scale update rule converge to a local nash equilibrium”. In: *Advances in neural information processing systems* 30 (2017).

[31] Kishore Papineni et al. “Bleu: a method for automatic evaluation of machine translation”. In: *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 2002, pp. 311–318.

[32] Chin-Yew Lin. “Rouge: A package for automatic evaluation of summaries”. In: *Text summarization branches out*. 2004, pp. 74–81.

[33] Satanjeev Banerjee and Alon Lavie. “METEOR: An automatic metric for MT evaluation with improved correlation with human judgments”. In: *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*. 2005, pp. 65–72.

[34] Xueyan Mei et al. “RadImageNet: an open radiologic deep learning research dataset for effective transfer learning”. In: *Radiology: Artificial Intelligence* 4.5 (2022), e210315.

[35] Adrien Meyer et al. “Ultrasam: a foundation model for ultrasound using large open-access segmentation datasets”. In: *arXiv preprint arXiv:2411.16222* (2024).

[36] Alec Radford et al. “Learning transferable visual models from natural language supervision”. In: *International conference on machine learning*. PmLR. 2021, pp. 8748–8763.

[37] Sheng Zhang et al. “A multimodal biomedical foundation model trained from fifteen million image–text pairs”. In: *NEJM AI* 2.1 (2025), AIoa2400640.

[38] Hao Zhang et al. “DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection”. In: *International Conference on Learning Representations*. 2023.

[39] Zhenda Xie et al. “Simmim: A simple framework for masked image modeling”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pp. 9653–9663.

[40] Ali Hatamizadeh et al. “Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images”. In: *International MICCAI brainlesion workshop*. Springer. 2021, pp. 272–284.

[41] Nathalie Jeanne Bravo-Valenzuela, Alberto Borges Peixoto, and Edward Araujo Júnior. “Prenatal diagnosis of congenital heart disease: A review of current knowledge”. In: *Indian heart journal* 70.1 (2018), pp. 150–164.

[42] Shaoqing Ren et al. “Faster R-CNN: Towards real-time object detection with region proposal networks”. In: *IEEE transactions on pattern analysis and machine intelligence* 39.6 (2016), pp. 1137–1149.

[43] Yifan Jiang et al. “Enlightengan: Deep light enhancement without paired supervision”. In: *IEEE transactions on image processing* 30 (2021), pp. 2340–2349.

[44] Grant Duffy et al. “High-Throughput Precision Phenotyping of Left Ventricular Hypertrophy With Cardiovascular Deep Learning”. In: *JAMA Cardiology* 7.4 (Apr. 2022), pp. 386–395. ISSN: 2380-6583. DOI: 10.1001/jamacardio.2021.6059. URL: https://doi.org/10.1001/jamacardio.2021.6059.

[45] Zhe Huang et al. *Fix-A-Step: Semi-supervised Learning from Uncurated Unlabeled Data*. 2023. arXiv: 2208.11870 [cs.LG]. URL: https://arxiv.org/abs/2208.11870.

[46] Alan Cervantes-Guzmán et al. “Robust cardiac segmentation corrected with heuristics”. In: *PLOS ONE* 18 (Oct. 2023), pp. 1–18. DOI: 10.1371/journal.pone.0293560. URL: https://doi.org/10.1371/journal.pone.0293560.

[47] Jiewen Yang et al. *GraphEcho: Graph-Driven Unsupervised Domain Adaptation for Echocardiogram Video Segmentation*. 2023. arXiv: 2309.11145 [cs.CV]. URL: https://arxiv.org/abs/2309.11145.

[48] Walid Al-Dhabyani et al. "Dataset of breast ultrasound images". In: *Data in Brief* 28 (2020), p. 104863. ISSN: 2352-3409. DOI: https://doi.org/10.1016/j.dib.2019.104863. URL: https://www.sciencedirect.com/science/article/pii/S2352340919312181.

[49] "Ultrasound Breast Images for Breast Cancer". https://www.kaggle.com/datasets/vuppalaadithyasairam/ultrasound-breast-images-for-breast-cancer.

[50] Pengfei Yan et al. "TDF-Net: Trusted Dynamic Feature Fusion Network for breast cancer diagnosis using incomplete multimodal ultrasound". In: *Information Fusion* 112 (2024), p. 102592. ISSN: 1566-2535. DOI: https://doi.org/10.1016/j.inffus.2024.102592. URL: https://www.sciencedirect.com/science/article/pii/S1566253524003701.

[51] "Ultra LR-HR Ultrasound Image Dataset for Research". https://www.kaggle.com/datasets/chirag2466/ultra-lr-hr-ultrasound-image-dataset-for-research.

[52] "ultrasound2ulm". https://www.kaggle.com/datasets/drjaveriaamin/ultrasound2ulm.

[53] Anna Pawłowska et al. "Curated benchmark dataset for ultrasound based breast lesion analysis". In: *Scientific Data* 11.1 (2024), p. 148.

[54] "Mendeley Breast Ultrasound Dataset". https://www.kaggle.com/datasets/jarintasnim090/mendeley-dataset.

[55] Moi Hoon Yap et al. "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks". In: *IEEE Journal of Biomedical and Health Informatics* 22.4 (2018), pp. 1218–1226. DOI: 10.1109/JBHI.2017.2731873.

[56] "OMI Breast Ultrasound Dataset". https://www.kaggle.com/datasets/jarintasnim090/omi-dataset.

[57] Jarin Tasnim and Md Kamrul Hasan. "CAM-QUS guided self-tuning modular CNNs with multi-loss functions for fully automated breast lesion classification in ultrasound images". In: *Physics in Medicine & Biology* 69.1 (Dec. 2023), p. 015018. DOI: 10.1088/1361-6560/ad1319. URL: https://dx.doi.org/10.1088/1361-6560/ad1319.

[58] Jin Huang et al. *BUSI_WHU: Breast Cancer Ultrasound Image Dataset*. Version 1. Mendeley Data, 2023.

[59] "STU-Hospital Dataset". https://github.com/xbhlk/STU-Hospital.

[60] Ali Abbasian Ardakani et al. "An open-access breast lesion ultrasound image database: Applicable in artificial intelligence studies". In: *Computers in Biology and Medicine* 152 (2023), p. 106438.

[61] Hessam Hamyoon et al. "Artificial intelligence, BI-RADS evaluation and morphometry: A novel combination to diagnose breast cancer using ultrasonography, results from multi-center cohorts". In: *European Journal of Radiology* 157 (2022), p. 110591.

[62] Hassan Homayoun et al. "Applications of machine-learning algorithms for prediction of benign and malignant breast lesions using ultrasound radiomics signatures: A multi-center study". In: *Biocybernetics and Biomedical Engineering* 42.3 (2022), pp. 921–933.

[63] Yanjun Guo et al. "Segmentation and recognition of breast ultrasound images based on an expanded U-Net". In: *Plos one* 16.6 (2021), e0253202.

[64] Yaosheng Lu et al. "The JNU-IFM dataset for segmenting pubic symphysis-fetal head". In: *Data in Brief* 41 (2022), p. 107904. ISSN: 2352-3409. DOI: https://doi.org/10.1016/j.dib.2022.107904. URL: https://www.sciencedirect.com/science/article/pii/S2352340922001160.

[65] "Automated measurement of fetal head circumference". https://zenodo.org/records/1322001.

[66] "ultrasoundvd". https://www.kaggle.com/datasets/yasinelh/ultrasoundvd.

[67] "Pubic Symphysis-Fetal Head Segmentation and Angle of Progression". https://zenodo.org/records/7851339.

[68] "Carotid Artery Ultrasound and Color Doppler". https://www.kaggle.com/datasets/pahunichoudhary/carotid-artery-ultrasound-and-color-doppler.

[69] "carotid artery ultrasound scans_HR image". https://www.kaggle.com/datasets/priyeshk/carotid-artery-ultrasound-scans-hr-image.

[70] Lorena Petrusca et al. “Hybrid ultrasound/magnetic resonance simultaneous acquisition and image fusion for motion monitoring in the upper abdomen”. In: *Investigative radiology* 48.5 (2013), pp. 333–340.

[71] Valeria De Luca et al. “A Learning-Based Approach for Fast and Robust Vessel Tracking in Long Ultrasound Sequences”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*. Ed. by Kensaku Mori et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 518–525. ISBN: 978-3-642-40811-3.

[72] “Sliver”. https://www.kaggle.com/datasets/tunkedsaro/sliver.

[73] “Annotated Ultrasound Liver images”. https://zenodo.org/records/7272660.

[74] Yiming Xu et al. “Improving artificial intelligence pipeline for liver malignancy diagnosis using ultrasound images and video frames”. In: *Briefings in Bioinformatics* 24.1 (Dec. 2022), bbac569. ISSN: 1477-4054. DOI: 10.1093/bib/bbac569. eprint: https://academic.oup.com/bib/article-pdf/24/1/bbac569/51014628/bbac569.pdf. URL: https://doi.org/10.1093/bib/bbac569.

[75] “Algerian Ultrasound Images Thyroid Dataset: AUITD”. https://www.kaggle.com/datasets/azouzmaroua/algeria-ultrasound-images-thyroid-dataset-auitd.

[76] Tom Wunderling et al. “Comparison of thyroid segmentation techniques for 3D ultrasound”. In: *Medical Imaging 2017: Image Processing*. Vol. 10133. SPIE. 2017, pp. 346–352.

[77] Haifan Gong et al. “Multi-Task Learning For Thyroid Nodule Segmentation With Thyroid Region Prior”. In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. 2021, pp. 257–261. DOI: 10.1109/ISBI48211.2021.9434087.

[78] Haifan Gong et al. “Thyroid Region Prior Guided Attention for Ultrasound Segmentation of Thyroid Nodules”. In: *Computers in Biology and Medicine* 106389 (2022), pp. 1–12.

[79] Markus Krönke et al. “Tracked 3D ultrasound and deep neural network-based thyroid segmentation reduce interobserver variability in thyroid volumetry”. In: *Plos one* 17.7 (2022), e0268550.

[80] “Ultrasound image set of hemangioma3”. https://www.kaggle.com/datasets/shengwang1130/ultrasound-image-set-of-hemangioma3.

[81] Alexandra Laverde-Saad et al. “Discriminative deep learning based benignity/malignancy diagnosis of dermatologic ultrasound skin lesions with pretrained artificial intelligence architecture”. In: *Skin Research and Technology* 28.1 (2022), pp. 35–39.

[82] “NHBS-Net_SPJ_dataset”. https://github.com/hidden-ops/NHBS-Net_SPJ_dataset.

[83] Jannis Born et al. “Accelerating Detection of Lung Pathologies with Explainable Ultrasound Image Analysis”. In: *Applied Sciences* 11.2 (Jan. 2021), p. 672. ISSN: 2076-3417. DOI: 10.3390/app11020672. URL: http://dx.doi.org/10.3390/app11020672.

[84] “TF_FOR_POCOVID_ULTRASOUND”. https://www.kaggle.com/datasets/bachaboos/tf-for-pocovid-ultrasound.

[85] “Kidney Ultrasound Images "Stone" and "No Stone"”. https://www.kaggle.com/datasets/gurjeetkaurmangat/kidney-ultrasound-images-stone-and-no-stone.

[86] “new_data_F_nofiles”. https://www.kaggle.com/datasets/ahmadshtaiyat/new-data-f.

[87] “ultrasound-nerve-segmentation”. https://www.kaggle.com/datasets/anupaankarigari/ultrasound-nerve-segmentation.

[88] “ultrasound_img”. https://www.kaggle.com/datasets/thegna/ultrasound-img.

[89] “dataset for ultrasound pregnancy investigation v0.0.4”. https://zenodo.org/records/5110223.

[90] “test dataset for ultrasound pregnancy investigation”. https://zenodo.org/records/5110198.

[91] “PCOS detection using ultrasound images”. https://www.kaggle.com/datasets/anaghachoudhari/pcos-detection-using-ultrasound-images.

[92] “2000 Ultrasound Images for PCOS Disease Detection”. https://www.kaggle.com/datasets/shnotweta/2000-images-of-ultrasound-for-pcos.

[93] U Rajendra Acharya et al. “Use of nonlinear features for automated characterization of suspicious ovarian tumors using ultrasound images in fuzzy forest framework”. In: *International Journal of Fuzzy Systems* 20 (2018), pp. 1385–1402.

[94] Qi Zhao et al. *MMOTU: A Multi-Modality Ovarian Tumor Ultrasound Image Dataset for Unsupervised Cross-Domain Semantic Segmentation*. 2023. arXiv: 2207.06799 [cs.CV]. URL: https://arxiv.org/abs/2207.06799.

[95] Hanne Sorger et al. "A multimodal image guiding system for Navigated Ultrasound Bronchoscopy (EBUS): A human feasibility study". In: *PloS one* 12.2 (2017), e0171841.

[96] "3vessels + gallbladder". https://zenodo.org/records/7323401.

[97] Zihao Chen et al. "Functional Ultrasound Imaging of Auditory Responses in Comatose Patients". In: *medRxiv* (2024). DOI: 10.1101/2024.12.22.24319283. eprint: https://www.medrxiv.org/content/early/2024/12/26/2024.12.22.24319283.full.pdf. URL: https://www.medrxiv.org/content/early/2024/12/26/2024.12.22.24319283.

[98] Parikshit Juvekar et al. "ReMIND: The Brain Resection Multimodal Imaging Database". In: *medRxiv* (2023). DOI: 10.1101/2023.09.14.23295596. eprint: https://www.medrxiv.org/content/early/2023/09/15/2023.09.14.23295596.full.pdf. URL: https://www.medrxiv.org/content/early/2023/09/15/2023.09.14.23295596.

[99] Yiming Xiao et al. *RESECT: a clinical database of pre-operative MRI and intra-operative ultrasound in low-grade glioma surgeries*. 2017. DOI: 10.11582/2017.00004. URL: https://archive.norstore.no/pages/public/datasetDetail.jsf?id=10.11582/2017.00004.

[100] Ryan Bennett et al. "An ultrasound study of Connemara Irish palatalization and velarization". In: *Journal of the International Phonetic Association* 48.3 (2018), pp. 261–304.

[101] Maciej Stukan et al. "Accuracy of Ultrasonography and Magnetic Resonance Imaging for Preoperative Staging of Cervical Cancer—Analysis of Patients from the Prospective Study on Total Mesometrial Resection". In: *Diagnostics* 11.10 (2021). ISSN: 2075-4418. DOI: 10.3390/diagnostics11101749. URL: https://www.mdpi.com/2075-4418/11/10/1749.

[102] Hongxu Jiang et al. "MicroSegNet: A deep learning approach for prostate segmentation on micro-ultrasound images". In: *Computerized Medical Imaging and Graphics* 112 (2024), p. 102326. ISSN: 0895-6111. DOI: https://doi.org/10.1016/j.compmedimag.2024.102326. URL: https://www.sciencedirect.com/science/article/pii/S089561112400003X.

[103] "MR to Ultrasound Registration for Prostate Challenge - Dataset". https://zenodo.org/records/8004388.

[104] "SPLOA study spleen ultrasound videos". https://zenodo.org/records/13195053.

[105] "lymph node ultrasound image dataset". https://zenodo.org/records/12702916.

[106] Ričards Marcinkevičs et al. "Interpretable and intervenable ultrasonography-based machine learning models for pediatric appendicitis". In: *Medical Image Analysis* 91 (2024), p. 103042.

[107] Qi Li et al. "Trackerless freehand ultrasound with sequence modelling and auxiliary transformation over past and future frames". In: *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2023, pp. 1–5.

[108] Qi Li et al. "Nonrigid Reconstruction of Freehand Ultrasound Without a Tracker". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2024, pp. 689–699.

[109] Qi Li et al. "Long-term dependency for 3d reconstruction of freehand ultrasound without external tracker". In: *IEEE Transactions on Biomedical Engineering* 71.3 (2023), pp. 1033–1042.

[110] Qi Li et al. "Privileged anatomical and protocol discrimination in trackerless 3d ultrasound reconstruction". In: *International Workshop on Advances in Simplifying Medical Ultrasound*. Springer. 2023, pp. 142–151.

[111] "ACOUSLIC-AI : Abdominal Circumference Operator-agnostic UltraSound measurement in Low-Income Countries using Artificial Intelligence". https://zenodo.org/records/12697994.

[112] "Fetal Abdominal Structures Segmentation Dataset Using Ultrasonic Images". https://data.mendeley.com/datasets/4gcpm9dsc3/1.

[113] S Natarajan et al. "Prostate MRI and ultrasound with pathology and coordinates of tracked biopsy (prostate-MRI-US-biopsy)". In: *Cancer Imaging Arch* 10 (2020), p. 7937.

[114] Jiajia Li et al. "Dsmt-net: Dual self-supervised multi-operator transformation for multi-source endoscopic ultrasound diagnosis". In: *IEEE Transactions on Medical Imaging* 43.1 (2023), pp. 64–75.

[115] Malte Mechtenberg et al. "Manual and semi-automatic determination of elbow angle-independent parameters for a model of the biceps brachii distal tendon based on ultrasonic imaging". In: *Plos one* 17.10 (2022), e0275128.

[116] Zhenxing Sun et al. “Diagnostic value of transthoracic echocardiography in patients with coarctation of aorta: the Chinese experience in 53 patients studied between 2008 and 2012 in one major medical center”. In: *PLoS One* 10.6 (2015), e0127399.

[117] Patrick Carnahan. “Towards Patient Specific Mitral Valve Modelling via Dynamic 3D Transesophageal Echocardiography”. In: (2023).

[118] “Common Carotid Artery Ultrasound Images”. https://data.mendeley.com/datasets/d4xt63mgjm/1.

[119] Yuncheng Jiang et al. “Towards a Benchmark for Colorectal Cancer Segmentation in Endorectal Ultrasound Videos: Dataset and Model Development”. In: *arXiv preprint arXiv:2408.10067* (2024).

[120] Yi Ding et al. “MallesNet: A multi-object assistance based network for brachial plexus segmentation in ultrasound images”. In: *Medical Image Analysis* 80 (2022), p. 102511. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2022.102511. URL: https://www.sciencedirect.com/science/article/pii/S136184152200158X.

[121] Clyde J Belasso et al. “LUMINOUS database: lumbar multifidus muscle segmentation from ultrasound images”. In: *BMC Musculoskeletal Disorders* 21 (2020), pp. 1–11.

[122] Xavier P Burgos-Artizzu et al. “Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes”. In: *Scientific Reports* 10.1 (2020), p. 10200.

[123] Francesco Marzola et al. “Deep learning segmentation of transverse musculoskeletal ultrasound images for neuromuscular disease assessment”. In: *Computers in Biology and Medicine* 135 (2021), p. 104623. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2021.104623. URL: https://www.sciencedirect.com/science/article/pii/S0010482521004170.

[124] Paul Ritsche et al. “DeepACSA: automatic segmentation of cross-sectional area in ultrasound images of lower limb muscles using deep learning”. In: *Medicine and Science in Sports and Exercise* (2022).

[125] Hugo Michard et al. “AW-Net: automatic muscle structure analysis on B-mode ultrasound images for injury prevention”. In: *Proceedings of the 12th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. BCB ’21. Gainesville, Florida: Association for Computing Machinery, 2021. ISBN: 9781450384506. DOI: 10.1145/3459930.3469531. URL: https://doi.org/10.1145/3459930.3469531.

[126] D Miguez et al. “A technical note on variable inter-frame interval as a cause of non-physiological experimental artefacts in ultrasound”. In: *Royal Society Open Science* 4.5 (2017), p. 170245.

[127] Vanessa Gonzalez Duque et al. “Ultrasound segmentation analysis via distinct and completed anatomical borders”. In: *International Journal of Computer Assisted Radiology and Surgery* 19.7 (2024), pp. 1419–1427.

[128] Tamas Ungi et al. “Automatic spine ultrasound segmentation for scoliosis visualization and measurement”. In: *IEEE Transactions on Biomedical Engineering* 67.11 (2020), pp. 3234–3241.