Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

M2-US: A Large-scale Multi-domain and Multi-task Ultrasound Image Dataset

FANGYIJIE WANG

fangyijie.wang@ucdconnect.ie

University College Dublin https://orcid.org/0009-0003-0427-368X

Tanya Akumu

University of Barcelona

Vien Ngoc Dang

University of Barcelona

Amelia Jiménez-Sánchez

University of Barcelona

Jiancheng Yang

Aalto University https://orcid.org/0000-0003-4455-7145

Guénolé Silvestre

University College Dublin

Karim Lekadir

University of Barcelona

Jieyun Bai

jieyun.bai@ieee.org

Jinan University

Kathleen M. Curran

University College Dublin

Research Article

Keywords: Ultrasound, Multi-domain, Multi-task, Medical Image Analysis

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11823363/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

M2–US: A Large-scale Multi-domain and Multi-task Ultrasound Image Dataset

Fangyijie Wang[1,3⋆], Tanya Akumu[2], Vien Ngoc Dang[2], Amelia Jiménez-Sánchez[2], Jiancheng Yang[4], Guénolé Silvestre[1,5] Karim Lekadir[2,6], Jieyun Bai[7,8⋆], and Kathleen M. Curran[1,3]

[1] Research Ireland Centre for Research Training in Machine Learning
[2] Artificial Intelligence in Medicine Lab (BCN-AIM), Department of Mathematics and Computer Science, Universitat de Barcelona, Barcelona, Spain
[3] School of Medicine, University College Dublin, Dublin, Ireland
[4] Health Intelligence Lab, Aalto University, Espoo, Finland
[5] School of Computer Science, University College Dublin, Dublin, Ireland
[6] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
[7] Department of Cardiovascular Surgery, The First Affiliated Hospital of Jinan University, Jinan University, Guangzhou, China
[8] Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand

Abstract. Ultrasound imaging is widely used across clinical specialties. Artificial Intelligence methods have shown remarkable advantages in ultrasound image analysis in recent decades. However, the development of generalizable Artificial Intelligence methods remains limited by the heterogeneity of ultrasound data and the task-specific design of most existing datasets. Here, we present M2–US, a large-scale **multi-domain** benchmark for unified **multi-task ultrasound** image analysis. The dataset consists of 67,620 ultrasound images from public datasets and retrospectively collected clinical cohorts. It covers more than 30 anatomical structures and pathological targets across maternal–fetal, cardiac, vascular, abdominal, pulmonary, breast, thyroid, gynecological, and other clinical applications. M2–US supports four major prediction tasks: segmentation, classification, object detection, and landmark regression, organized into 26 subtasks. The dataset provides predefined training, validation, and test partitions, together with task-specific annotations and standardized metadata linking ultrasound images to their corresponding segmentation masks, class labels, bounding boxes, and landmark coordinates. Baseline experiments using Feature Pyramid Network (FPN) and DINOv3 models show the usability of the dataset and evaluation pipeline across heterogeneous ultrasound tasks. To our knowledge, M2–US is the first large-scale ultrasound benchmark comprising data acquired using approximately 24 scanner models from more than 10 manufacturers and covering over 30 anatomical structures and pathological targets. The benchmark supports four major downstream tasks across segmentation, classification, detection, and regression. Together with our evaluation pipeline, the benchmark enables systematic evaluation of task-specific models, multi-task learning methods, foundation model adapta-

⋆ Corresponding authors: fangyijie.wang@ucdconnect.ie, baijieyun@jnu.edu.cn

tion, as well as cross-domain generalization in ultrasound image analysis. The data and code are publicly available at `https://m2-us.github.io/`.

Background & Summary

Ultrasound is widely used across a broad range of clinical applications because it provides real-time imaging without producing radiation and is comparatively affordable, portable, and accessible. It supports routine assessment in obstetrics and gynecology, cardiology [37], vascular imaging [29], thyroid examination [40], and image-guided interventions [24,41]. Depending on the clinical application, ultrasound images can be used to delineate anatomical structures [21], characterize pathological findings [12], localize lesions [35], identify standard imaging planes [42], and estimate clinically relevant biometric measurements [42,12]. Because of these characteristics, ultrasound is important in both hospitals and community healthcare, point-of-care applications, and resource-constrained settings [25,16,31].

However, ultrasound imaging exhibits substantial variability across operators, devices, acquisition protocols, anatomical targets, and patient populations. Image appearance can be affected by probe orientation, scanning trajectory, applied pressure, acoustic window [38,18], imaging depth, gain, focus [13], and other acquisition parameters [26,14,43]. Additional variation arises from scanner manufacturers, probe types, transducer frequencies, reconstruction pipelines, clinical sites, and patient populations [8,9,32]. Ultrasound images are also affected by modality-specific characteristics such as speckle noise, acoustic shadowing, reverberation, signal dropout, low soft-tissue contrast, and incomplete anatomical boundaries [33]. Consequently, models developed using data from a single anatomical target, institution, or scanner may not readily generalize to other clinical settings.

A growing number of public ultrasound datasets and biomedical imaging challenges have supported the development and standardized evaluation of ultrasound image analysis methods (Table 1). These resources cover a range of clinical applications, including fetal biometry, intrapartum assessment, fetal heart analysis, carotid plaque assessment, thyroid nodules, breast lesions, liver disease, and other anatomical structures [27,15,6,3,4,7,28,36,11]. For example, HC18 [15] focuses on fetal-head segmentation and head-circumference estimation, while PSFHS [6] and IUGC [3,7] provide fetal-head and pubic-symphysis annotations for intrapartum ultrasound analysis. FUGC [4] targets cervical ultrasound analysis, whereas FETUS [1] and CSV focus on fetal-heart and carotid-plaque assessment, respectively. Other datasets provide annotations for breast lesions, thyroid nodules, liver lesions, fetal abdominal structures, and bone surfaces [36,2,19,39].

Despite this progress, many existing ultrasound datasets remain focused on a single anatomical structure, clinical condition, or annotation type. More recent resources have started to increase anatomical diversity. UUSIC25 [23], for example, combines seven anatomical targets and supports segmentation and classification, while UUSIVC2026 [17] extends this setting to 14 organs. Other initiatives

have broadened the methodological scope: FoundUS [5] focuses on ultrasound biometry and regression across multiple parameters, whereas TUS-REC2025 [20] addresses three-dimensional freehand ultrasound reconstruction. These developments reflect two complementary trends in ultrasound dataset construction: increasing anatomical diversity and increasing diversity of annotation and prediction types. Nevertheless, datasets that simultaneously combine multiple clinical domains, anatomical targets, geographical origins, acquisition systems, and fundamentally different annotation types remain limited.

To address this limitation, we introduce the **M2–US** dataset, a large-scale multi-domain and multi-task ultrasound imaging resource assembled from publicly available datasets and retrospectively collected clinical data with documented provenance. M2–US contains 67,620 ultrasound images covering more than 30 anatomical structures and pathological targets across maternal–fetal, cardiac, vascular, abdominal, pulmonary, breast, thyroid, ovarian, uterine, and other clinically relevant applications. The anatomical coverage of M2–US is summarized schematically in Fig. 1, illustrating the maternal-fetal and general clinical ultrasound applications represented in the dataset. The dataset integrates data acquired across multiple countries and approximately 24 ultrasound scanner models from more than 10 manufacturers, thereby incorporating substantial geographical and acquisition heterogeneity.

M2–US provides annotations corresponding to four major forms of ultrasound image analysis: **segmentation, classification, object detection, and landmark-based regression**. The dataset is organized into **26 subtasks**, comprising 12 segmentation, nine classification, three detection, and three landmark-based regression tasks. The annotations therefore include pixel-wise masks for anatomical structures and pathological regions, image-level categorical labels, bounding boxes for target localization, and anatomical landmarks for ultrasound biometry. This heterogeneous annotation structure enables the same resource to support task-specific model development, multi-task learning, representation learning, foundation-model adaptation, and studies of cross-domain generalization.

In this paper, we document M2–US's source datasets, clinical cohorts, anatomical and geographical composition, annotation formats, data harmonization, dataset organization, and quality-control procedures. By consolidating heterogeneous ultrasound data within a common structure, M2–US aims to facilitate the development and evaluation of ultrasound AI methods across multiple organs, clinical applications, and prediction tasks. Furthermore, this dataset provides a reproducible resource for investigating the impact of heterogeneity in ultrasound imaging environments on model transferability and robustness. To support reproducible benchmarking, we also provide baseline implementations and evaluation code for the supported prediction tasks.

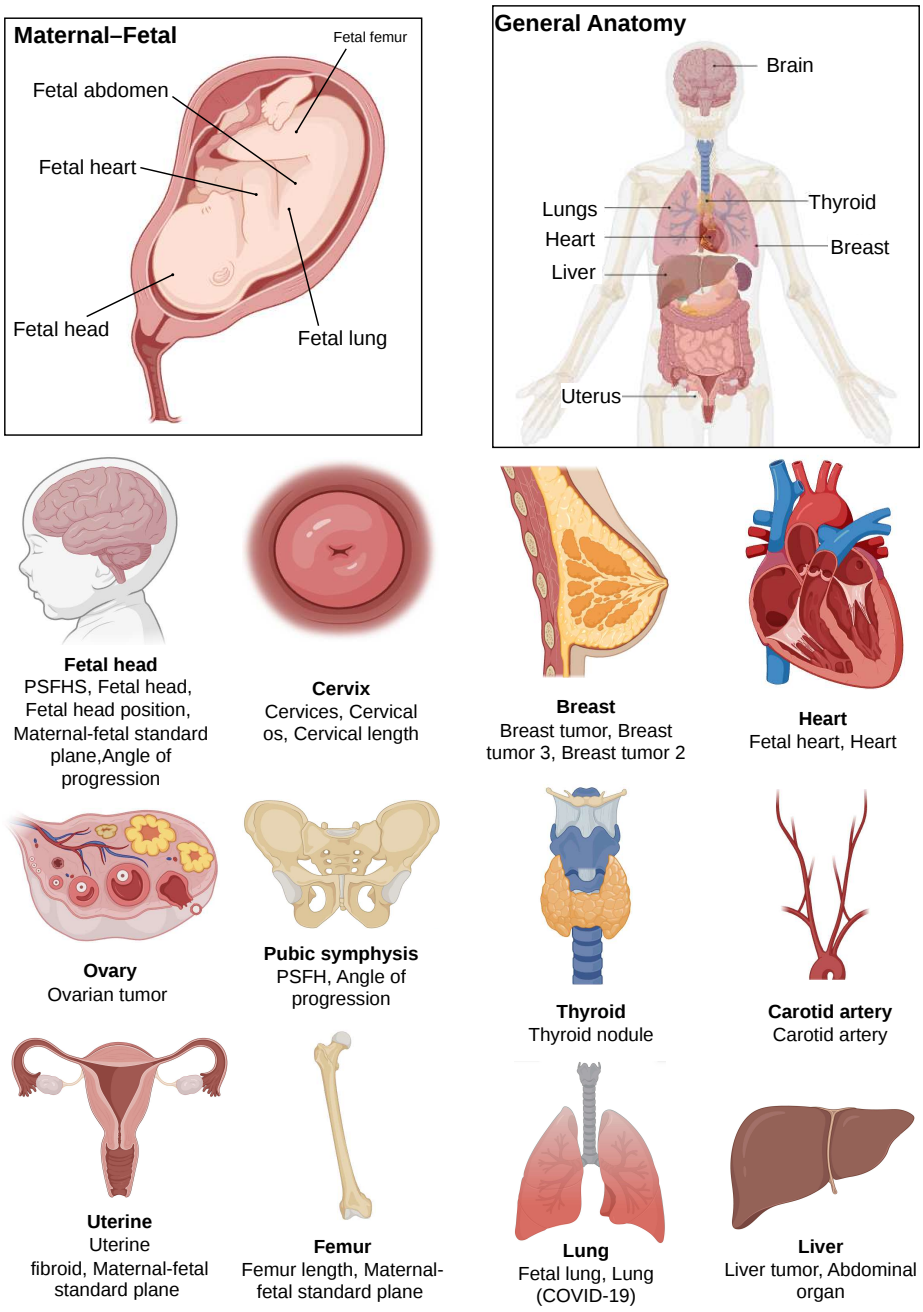

Fig. 1: Anatomical coverage of the M2–US dataset. M2–US comprises ultrasound data spanning maternal–fetal and general clinical applications, including fetal, gynecological, cardiovascular, pulmonary, breast, thyroid, hepatic, vascular, and other anatomical structures and regions. The illustrations provide a schematic overview of the anatomical targets represented in the dataset, with the corresponding M2–US dataset/task categories listed beneath each anatomy.

Table 1: Summary of existing ultrasound datasets and challenges.

Dataset	Clinical domain	Annotation type	Images (videos)	Countries / centres	Public
US	Fetal biometry	Segmentation	284	UK / 1 centre	No
A-AFMA	Amniotic fluid	Landmark detection	–	UK / 1 centre	No
FUGC	Cervix	Segmentation	890	Spain / 1 centre	Yes
PSFHS	Fetal head and pubic symphysis	Segmentation	5,101	China / 3 centres	Yes
IUGC	Fetal head and pubic symphysis	Segmentation	774 images / 68,106 videos	China / 3 centres	Yes
ACOUSLIC-AI	Fetal abdomen	Segmentation	592	Sierra Leone, Tanzania, Netherlands / 3 centres	Yes
FETAL_PLANE_DB	Fetal standard planes	Classification	12,400	Spain / 2 centres	Yes
MFUSPAC	Fetal ultrasound	Classification	450	Malawi, Egypt, Uganda, Ghana, Algeria / 5 centres	Yes
FASSD	Fetal abdomen	Segmentation	1,500	Brazil / 1 centre	Yes
HC18	Fetal head	Segmentation	1,334	Netherlands / 1 centre	Yes
DFHI	Fetal head	Segmentation; object detection	3,832	Spain and Netherlands / 3 centres	Yes
CSV	Carotid plaque	Segmentation; classification	3,000	China / 12 centres	Yes
FETUS	Fetal heart	Segmentation; classification	5,000	China / 7 centres	Yes
UUSIC25	Breast, thyroid, liver, kidney, fetal head, heart, and appendix	Segmentation; classification	12,002	11 data sources	Yes
TUS-REC2025	Left and right forearms	Reconstruction	170 videos	UK / 1 centre	Yes
UUSIVC2026	14 organs	Segmentation; classification	13,502	3+ centres	Yes
FoundUS	26 ultrasound parameters	Regression	20,000+	7+ data sources	Yes
UltraBones100k	Fibula, tibia, and foot bone surfaces	Segmentation	>100,000	1 centre	Yes
Liver lesion	Liver lesions	Segmentation	519	China / 1 centre	No
NAFLD	Steatosis and fibrosis	Classification	10,352	Saudi Arabia / 2 centres	Yes
ThyroidXL	Thyroid nodule	Classification; detection; segmentation	11,635	Vietnam / 1 centre	Yes
Annotated US	13 anatomical categories	Classification	1,833	China / 1 centre	Yes
BUS-UCLM	Breast lesions	Classification; detection; segmentation	683	Spain / 1 centre	Yes
Brain tumor	Brain tumor	Segmentation	1,856	1 centre	Yes
TN5000	Thyroid nodule	Detection; classification	5,000	China / 1 centre	Yes
M2–US	**Multi-organ, multi-disease ultrasound**	**Segmentation; classification; detection; landmark regression**	**67,602**	**12 countries / multi-centre**	**Yes**

Methods

Ethics approval and informed consent

All clinical data included in M2–US were collected in accordance with the relevant institutional and national ethical requirements. For previously published public datasets, the ethical approvals and consent procedures were those reported by the original data providers. These datasets were reused and redistributed in accordance with their applicable licences and terms of access. All private clinical data were retrospectively collected under protocols reviewed and approved by the relevant Institutional Review Boards (IRBs) or Ethics Committees. For the segmentation tasks, the fetal head and pubic symphysis data were approved by the Institutional Review Board of Guangzhou Women and Children's Medical Center, Guangzhou, China (No. 405A01); the cervix data by the Institutional Review Board of Zhujiang Hospital, Southern Medical University, Guangzhou, China (No. 2023-SYJS-023); the cervical os data and the fetal head by the Institutional Review Board of the First Affiliated Hospital of Jinan University, Guangzhou, China (No. JNUKY-2022-019); the fetal lung data by the same institution (No. JNUKY-2023-050); and the cardiac data by the Institutional Review Board of Guangzhou Women and Children's Medical Center (No. 279A01). For the classification tasks, the fetal head position and fetal breech position data were collected under approval from the Institutional Review Board of the First Affiliated Hospital of Jinan University (No. JNUKY-2022-019), while the fetal lung classification data were collected under approval No. JNUKY-2023-050 from the same institution. The private thyroid nodule detection data were approved by the Institutional Review Board of the First Affiliated Hospital of Naval Military Medical University, Shanghai, China (No. CHEC2025-077). For the regression tasks, the angle of progression, cervical length, and femur length data were collected under approval from the Institutional Review Board of the First Affiliated Hospital of Jinan University (No. JNUKY-2022-019). All private images were de-identified before their inclusion in the benchmark. The requirement for individual informed consent was waived by the respective ethics committees where applicable because of the retrospective study design and the use of de-identified medical imaging data.

Data Sources and Dataset Composition

The dataset was assembled by integrating multiple publicly available ultrasound datasets with additional private clinical datasets from participating institutions. These sources were selected to provide broad coverage of ultrasound image-analysis tasks, anatomical targets, and clinical applications. The final resource contains 67,620 samples across four task categories: segmentation (28,249), classification (26,280), detection (7,229), and regression (6,230). The distribution of samples across source datasets, clinical domains, and task categories is shown in Fig. 2, highlighting the multi-domain and multi-task composition of the benchmark. Public datasets were obtained from their original repositories and used

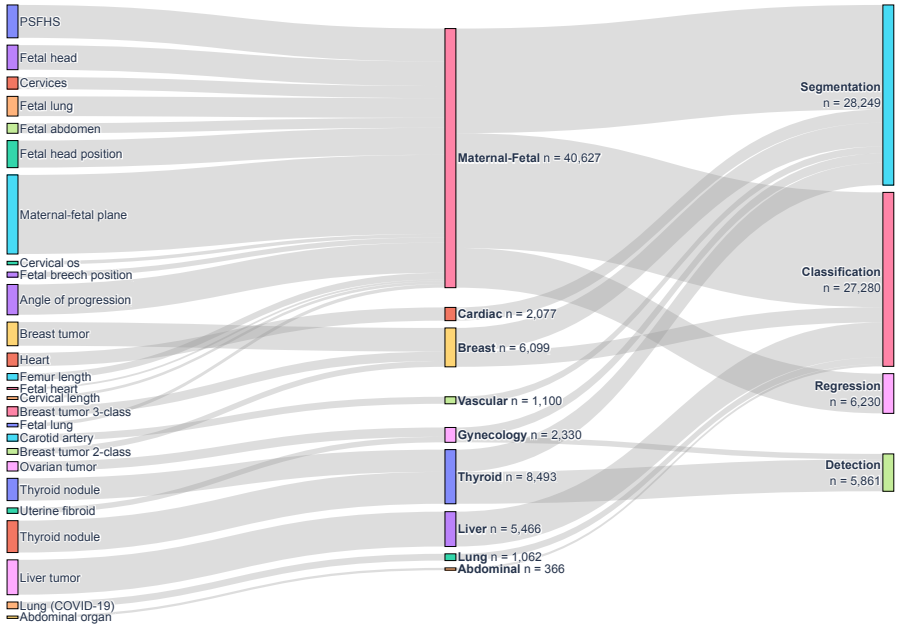

Fig. 2: Overview of the M2–US dataset composition across clinical domains and learning tasks. The diagram illustrates the distribution of ultrasound samples from individual datasets to their corresponding clinical domains and downstream tasks, including segmentation, classification, detection, and regression. Flow width is proportional to the number of samples associated with each dataset–task combination, with the total number of samples shown for each clinical domain and task.

according to the corresponding licenses, whereas private datasets were incorporated under institutional ethical approvals and data-governance requirements. For each source, task-relevant images and annotations were extracted and organized within a unified dataset structure. Table 1 provides the task-level sample distributions, source information, and licensing details.

The dataset covers over 30 anatomical structures. The segmentation task includes targets such as the fetal head, pubic symphysis, cervix, fetal heart, fetal abdomen, carotid artery, ovarian and breast tumors, thyroid nodules, and cardiac structures. The classification task includes fetal position, standard-plane recognition, and disease or lesion classification tasks. The detection and regression components cover lesion localization, anatomical structure detection, and biometric measurements such as angle of progression, cervical length, and femur length.

Table 2: Geographical origin and ultrasound devices associated with the datasets included in the benchmark. The "Test included" column indicates whether the corresponding task is included in the held-out test set.

Targets	Task ID	Country	Ultrasound device	Test included
Segmentation				
PSFH	head_symphysis_multi	China	Voluson P8, Esaote MyLab, Lian-med ObEye, Youkey Q7	Yes
Cervices	cervix_multi	China	Voluson E8	Yes
Cervical os	cervix	China	Esaote M8	Yes
Fetal head	fetal_head	Netherlands, Spain, Malawi, Egypt, Uganda, Ghana, Algeria	Voluson E6, Voluson 730, Voluson S8, Voluson E8, Voluson P8, Voluson S10, Mindray DC-N2, Aloka, Siemens, Edan DUS 60	No
Fetal lung	lung	China	Voluson E10, E6, E8, P8	Yes
Carotid artery	carotid_artery	–	Mindray UMT-500Plus	No
Fetal heart	fetal_heart	China	–	No
Fetal abdomen	fetal_abdomen_multi	Brazil	Siemens Acuson, Voluson 730, Philips EPIQ Elite	No
Ovarian tumor	ovary_tumor	China	Mindray Resona8	No
Breast tumor	breast_lesion	China, Spain, Brazil, Mexico, Egypt	LOGIQ E9, LOGIQ E9 Agile, GE Logiq 5, Logiq 7, U-Systems, Toshiba Aplio 300, Siemens ACUSON S2000TM	No
Thyroid nodule	thyroid_nodule	China	GE Logiq E9, ARIETTA 850, RESONA 70B	No
Heart	cardiac_multi	France	GE Vivid E95	Yes
Classification				
Fetal head position	fetal_head_pos_cls	China	Lian-med ObEye	Yes
Fetal breech position	fetal_sacral_pos_cls	China	Lian-med ObEye	Yes
Lung (COVID-19)	lung_disease_3cls	Uganda	Clarius C3 and C5	No
Maternal-fetal standard plane	fetal_plane_cls	Spain	Voluson E6, S8, S10, Aloka	No
Breast tumor 3	breast_3cls	Mexico, Egypt	Siemens ACUSON S2000TM, LOGIQ E9, Agile	No
Breast tumor 2	breast_2cls	China	–	No
Liver tumor	liver_lesion_2cls	–	Logiq7, Logiq9	No
Abdominal organ	organ_cls	–	–	No
Fetal lung	lung_2cls	China	Voluson E10, E6, E8, P8	Yes
Detection				
Thyroid nodule	thyroid_nodule_det	China	GE Logiq E9 and S7	Yes
Uterine fibroid	uterine_fibroid_det	–	–	Yes
Regression				
Angle of progression	IUGC	China	Voluson P8, Esaote MyLab, Lian-med ObEye, Youkey Q7	Yes
Cervical length	FUGC	China	Voluson E8	Yes
Femur length	fetal_femur	Spain	Voluson E6, S8, S10, Aloka	Yes

Data Acquisition

Public datasets incorporated into M2–US were obtained from their original repositories and were used according to their published data descriptions and licensing conditions. Detailed acquisition protocols, cohort characteristics, and inclusion criteria are available in the corresponding original publications.

Annotation Protocol and Label Definition

For publicly available datasets, M2–US keeps the original annotations released with each source dataset, including segmentation masks, classification labels, landmarks, and biometric measurements, as applicable. The corresponding annotation procedures and clinical definitions are described in the original publications. For the private datasets introduced in this study, annotations were generated and reviewed as described below.

Data Harmonization

Data from the different source datasets were harmonized into a common task-oriented structure while preserving the original image content and task-specific annotations. Each sample was assigned a standardized task category, subtask, image path, and target definition. Source labels were mapped to consistent class encodings within each task, and all task annotations were represented through a shared metadata schema in a CSV file.

For segmentation tasks, the original pixel-wise masks were linked to their corresponding ultrasound images. For classification tasks, source labels were converted into encoded categorical targets. For detection tasks, bounding-box annotations were represented using the original image coordinates through `x_min`, `y_min`, `x_max`, and `y_max` fields, with source XML annotations retained. For regression tasks, we stored anatomical landmarks as 2D coordinate pairs in the original image coordinate system.

The released images were not globally resized, intensity-normalized, or converted to a single file format, in order to preserve the characteristics of the original datasets and acquisition settings. Instead, our data harmonization focused on standardizing dataset organization, metadata fields, subtask identifiers, label representation, and file path naming conventions. Model-specific preprocessing, such as image resizing and normalization, was applied only during the baseline experiments and is described separately in the Section 1.

Dataset Partitioning

We partitioned the dataset into training, validation, and test sets for each task to support standardized model development and evaluation. The complete dataset contains 67,620 samples. Across all tasks, 40,388 samples were assigned to the training set, 6,743 to the validation set, and 21,857 to the test set, corresponding approximately to a 60:10:30 split. Table 1 summarizes the sample distribution

for each target and task, together with the corresponding data source and license information.

Our data partitioning was performed at the image level rather than the patient level because patient identifiers were not consistently available across all datasets. Consequently, for datasets containing multiple images from the same examination or patient, complete patient-level separation between partitions could not be guaranteed. This predefined partitioning provides consistent training, model selection, and evaluation sets across the benchmark tasks.

Data Anonymization & privacy

We de-identified all ultrasound images before including them in the dataset to protect participant privacy and comply with institutional data-governance requirements. The de-identification procedure removed on-screen text and graphical overlays which contain potentially identifiable information, including patient names, identification numbers, examination dates, and other clinical identifiers. We also removed identifying information from the associated metadata before data integration and public release. Only de-identified images and task-relevant annotations were retained in the final dataset.

Data Records

Data repository and organization

The dataset will be deposited in the Zenodo repository under the title "A Large-scale Multi-domain and Multi-task Ultrasound Image Dataset" (DOI: 10.5281/zenodo.23558392). The public release will contain the complete training, validation, and test partitions. All files will be publicly accessible without an application procedure or other access restrictions.

The deposited dataset is organized into three top-level directories: `train`, `val`, and `test`. These directories define the predefined training, validation, and test partitions, respectively. Each partition contains four task-specific directories, namely `Segmentation`, `Classification`, `Detection`, and `Regression`, together with a `csv_files` directory containing the corresponding metadata and annotations:

```
M2-US/
├── train/
│   ├── Segmentation/
│   │   └── [Task Name]/
│   ├── Classification/
│   │   └── [Task Name]/
│   ├── Detection/
│   │   └── [Task Name]/
│   ├── Regression/
│   │   └── [Task Name]/
```

Table 3: The details of M2–US datasets. The dataset is constructed from multiple public datasets and supplemented with a limited amount of private clinical data. All datasets included in M2–US are distributed under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence, including both previously public datasets and newly released clinical data.

Targets	**Task ID**	**Training**	**Validation**	**Testing**	**Sum**	**Source (DOI)**
Segmentation		15991	2674	9584	28249	
PSFH	head_symphysis_multi	3059	511	1584	5154	10.5281/zenodo.7851338
Cervices	cervix_multi	261	44	1595	1900	10.5281/zenodo.16893174
Cervical os	cervix	327	55	164	546	10.5281/zenodo.16893174
Fetal head	fetal_head	2311	386	1157	3854	10.5281/zenodo.1322000;10.5281/zenodo.7757789
Fetal lung	lung	1852	310	927	3089	Private
Carotid artery	carotid_artery	659	110	331	1100	10.17632/d4xt63mgjm.1
Fetal heart	fetal_heart	177	30	93	300	10.5281/zenodo.14597549
Fetal abdomen	fetal_abdomen_multi	952	159	447	1558	10.17632/4gcpm9dsc3.1
Ovarian tumor	ovary_tumor	881	147	441	1469	10.6084/m9.figshare.25058690
Breast tumor	breast_lesion	2219	372	1118	3709	10.1016/j.dib.2019.104863;10.17632/7fvgj4jsp7.3;10.17632/k6cpmwybk3.1;10.1002/mp.16812
Thyroid nodule	thyroid_nodule	2094	350	1049	3493	10.1109/ISBI48211.2021.9434087
Heart	cardiac_multi	1199	200	678	2077	10.1109/TMI.2019.2900516
Classification		16362	2727	8191	27280	
Fetal head position	fetal_head_pos_cls	2546	423	1274	4243	Private
Fetal breech position	fetal_sacral_pos_cls	483	77	242	802	Private
Lung (COVID-19)	lung_disease_3cls	636	107	319	1062	10.17632/hb3p34ytvx.2
Maternal-fetal standard plane	fetal_plane_cls	7439	1240	3721	12400	10.5281/zenodo.3904279
Breast tumor 3	breast_3cls	876	147	440	1463	10.1016/j.dib.2019.104863;10.17632/7fvgj4jsp7.3
Breast tumor 2	breast_2cls	555	93	279	927	10.17632/k6cpmwybk3.1
Liver tumor	liver_lesion_2cls	3279	547	1640	5466	10.6084/m9.figshare.29755535
Abdominal organ	organ_cls	219	37	110	366	https://github.com/ftsvd/USAnotAI
Fetal lung	lung_2cls	329	56	166	551	Private
Detection		3515	587	1759	5861	
Thyroid nodule	thyroid_nodule_det	2999	500	1501	5000	10.1038/s41597-025-05757-4
Uterine fibroid	uterine_fibroid_det	516	87	258	861	10.17632/n2zcmcypgb.2
Regression		3702	617	1911	6230	
Angle of progression	IUGC	2818	470	1465	4753	10.5281/zenodo.17355569
Cervical length	FUGC	260	44	134	438	10.5281/zenodo.16893173
Femur length	fetal_femur	624	103	312	1039	Private
Total		39570	6605	21445	67620	

```
    └── csv_files/
        └── [Task Name].csv   # Complete annotations

├── val/
    ├── Segmentation/
        └── [Task Name]/
    ├── Classification/
        └── [Task Name]/
    ├── Detection/
        └── [Task Name]/
    ├── Regression/
        └── [Task Name]/
    └── csv_files/
        └── [Task Name].csv   # Complete annotations

└── test/
    ├── Segmentation/
        └── [Task Name]/
    ├── Classification/
        └── [Task Name]/
    ├── Detection/
        └── [Task Name]/
    ├── Regression/
        └── [Task Name]/
    └── csv_files/
        └── [Task Name].csv   # Complete annotations
```

where `[Task Name]` is the full name of each task.

The task directories are subdivided according to the individual challenge tasks and source datasets. The `csv_files` directories contain one metadata table for each task–source combination. The current release structure contains 45 CSV files in both the training and validation partitions and 42 CSV files in the test partition. The CSV filenames begin with `Seg`, `Cls`, `Det`, or `Reg`, indicating segmentation, classification, detection, or regression, respectively. The remainder of each filename identifies the associated source dataset or data subset.

Image and annotation formats

Ultrasound images are provided in their distributed raster formats, including PNG (`.png`), JPEG (`.jpg` and `.jpeg`), and bitmap (`.bmp`). Segmentation labels are stored as raster masks and linked to their corresponding images through the metadata tables. Detection annotations are additionally retained as XML files where supplied by the source dataset; the corresponding bounding-box coordinates are also represented directly in the CSV tables. Classification labels and regression landmarks are stored in the CSV files.

The file paths recorded in the metadata tables are relative to the corresponding `csv_files` directory. Consequently, the directory hierarchy can be downloaded and used without replacing the stored paths, provided that its internal structure is preserved.

Metadata tables

All four task types use a common set of metadata fields. Task-specific columns encode the corresponding segmentation, classification, detection, or regression targets. The CSV fields are defined in Table 4:

Table 4: Definitions of the fields contained in the metadata CSV files.

Field	Applicable tasks	Description
`image_path`	All	Relative path from the `csv_files` directory to the input ultrasound image.
`height`	All	Height of the image in pixels.
`width`	All	Width of the image in pixels.
`task_name`	All	General task category: segmentation, classification, detection, or regression.
`task_id`	All	Machine-readable identifier for the specific challenge task.
`num_classes`	All	Target cardinality defined for the task. For classification, segmentation, and detection tasks, this field denotes the number of target classes. For landmark-regression tasks, it denotes the number of annotated landmarks.
`mask_path`	Segmentation and detection	Relative path to the segmentation mask or source XML annotation file.
`mask`	Classification	Encoded class label assigned to the image. Despite the field name, this column contains a classification target rather than an image mask.
`x_min`, `y_min`, `x_max`, `y_max`	Detection	Coordinates defining the bounding box in the coordinate system of the original image.
`point_1_xy`, `point_2_xy`, `point_3_xy`	Regression	Two-dimensional landmark coordinates represented as `[x, y]` pairs. The number of point columns depends on the regression task.

For segmentation tasks, `image_path` and `mask_path` associate each ultrasound image with its pixel-wise annotation. For classification tasks, `image_path` identifies the image and mask contains its encoded categorical label. For detection tasks, the bounding-box coordinates are provided explicitly through `x_min`, `y_min`, `x_max`, and `y_max`, while `mask_path` retains the link to the correspond-

ing XML annotation when available. For regression tasks, each `point_n_xy` field contains the coordinates of one annotated anatomical landmark.

The complete release contains 26 subtasks: 12 segmentation tasks, 9 classification tasks, 2 detection tasks, and 3 regression tasks. The definitions of the subtasks, class encodings, segmentation-mask pixel values, and anatomical landmarks are provided in Table 2 and 3.

Data access

All three data partitions, including the annotations associated with the `test` partition, are publicly available through Zenodo. No registration, data-use application, or other access restriction will be imposed.

Technical Validation

Multi–task benchmark setup

To validate the usability of M2–US for multi-task model development, we trained two multi-task architectures representing different modeling paradigms. The first architecture has a conventional feature pyramid network (FPN)–based architecture [22], while the second architecture uses a pretrained DINOv3 backbone [30] for feature extraction. Both models were optimized using AdamW for 50 epochs with a batch size of 20 snd an initial learning rate of $10^{-}4$. We trained models once under the same experimental setting on an NVIDIA 4090 GPU. The training objectives comprised multiclass Dice loss for segmentation, cross-entropy loss for classification, and mean squared error loss for landmark regression. For detection task, we used a customized objective function which combines binary cross-entropy loss for foreground confidence and L_1 loss for bounding-box regression, weighted by 1 and 8, respectively. Images are resized to 256×256 and normalized channel-wise using the ImageNet mean values $(0.485, 0.456, 0.406)$ and standard deviations $(0.229, 0.224, 0.225)$.

We used the training set for optimization, the validation set for model selection, and the ranking-test partition only for final evaluation. Segmentation was evaluated using the Dice similarity coefficient (DSC) and Hausdorff distance (HD); classification was evaluated using the area under the receiver operating characteristic curve (AUC), F1-score (F1), and Matthews correlation coefficient (MCC); detection was evaluated using intersection over union (IoU); and landmark-based regression was evaluated using mean radial error (MRE). These metrics characterize complementary aspects of model performance for pixel-wise delineation, image-level recognition, object localization, and anatomical landmark localization, respectively.

Baseline experiments

We evaluated two baseline configurations on all M2–US tasks: an FPN-based model [22] with EfficientNet [34] backbone and a DINOv3-based model [30]

using the ViT encoder [10] with task-specific prediction heads. For the FPN baseline, all network layers were fine-tuned during training. For the DINOv3 baseline, only the final six encoder layers were fine-tuned together with the task-specific prediction heads, while the remaining pretrained encoder layers were frozen. Both models were trained and selected using the same predefined data partitions and optimization protocol. Table 5 reports performance on the test set using the primary evaluation metric for each task. These experiments were intended to verify the usability of the dataset and evaluation pipeline rather than establish state-of-the-art performance.

Table 5: Overall performance of two baselines, FPN and DINOv3, on M2–US.

Parent Task	Task ID	Primary Metric	FPN (EfficientNet)	DINOv3 (ViT)
Seg	head_symphysis_multi	DSC	0.7228 ± 0.2197	0.8598 ± 0.0721
	cervix_multi		0.8427 ± 0.1461	0.5922 ± 0.1116
	cervix		0.6588 ± 0.2493	0.8398 ± 0.1862
	fetal_head		0.8732 ± 0.1518	0.9620 ± 0.0466
	lung		0.7584 ± 0.2241	0.7964 ± 0.1368
	carotid_artery		0.8510 ± 0.1752	0.9294 ± 0.0640
	fetal_heart		0.8215 ± 0.1724	0.8267 ± 0.1192
	fetal_abdomen_multi		0.8114 ± 0.1773	0.6804 ± 0.1543
	ovary_tumor		0.7016 ± 0.2408	0.6785 ± 0.2813
	breast_lesion		0.7501 ± 0.2279	0.6923 ± 0.2592
	thyroid_nodule		0.6987 ± 0.2553	0.6087 ± 0.2629
	cardiac_multi		0.3882 ± 0.2422	0.5922 ± 0.1116
Cls	fetal_head_pos_cls	Macro-F1	0.7975	0.8041
	fetal_sacral_pos_cls		0.7332	0.7438
	lung_disease_3cls		0.5803	0.7838
	fetal_plane_cls		0.9201	0.8011
	breast_3cls		0.8093	0.6923
	breast_2cls		0.8307	0.5798
	liver_lesion_2cls		0.8283	0.8277
	organ_cls		0.8295	0.9503
	lung_2cls		0.8582	0.5675
Det	thyroid_nodule_det	IoU	0.2140 ± 0.2325	0.4329 ± 0.3137
	uterine_fibroid_det		0.0795 ± 0.1285	0.2677 ± 0.3021
Reg	IUGC	MRE	24.8426 ± 7.9509	52.9702 ± 13.8030
	FUGC		8.5527 ± 15.5407	25.20 ± 19.5352
	fetal_femur		38.1056 ± 60.2084	108.3840 ± 74.9840

Both baseline models can be trained and evaluated across the heterogeneous task set, although performance varied between tasks and anatomical structures. The DINOv3 baseline achieved higher performance on several segmentation and detection tasks, while the FPN baseline achieved better results on some classification and regression tasks. Therefore, these results highlight the diversity and difficulty of M2–US. Together with the released baseline code and evaluation pipeline, we report this table to provide a reproducible reference benchmark for future method development and comparison.

Usage Notes

Annotations in M2–US are task–specific. Users should consult the CSV files to identify the task, target type, and labels associated with each image. The original directory structure should be preserved because the recorded image and annotation paths are relative to the corresponding `csv_files` directory.

We recommend retaining the predefined `train`, `val`, and `test` partitions for benchmarking. The `test` partition was used to rank submissions in the ISBI 2026 FMC_UIA challenge [5] and should be reserved for final evaluation rather than model development or hyperparameter selection. Comparisons with challenge results require the same task definitions, evaluation metrics, and score-aggregation procedure. Alternative partitions may be used but should be described clearly. Because patient, examination, and acquisition-sequence identifiers are not provided, patient– or examination–level independence cannot be verified for newly constructed partitions.

M2–US combines data from multiple sources and consequently varies in anatomical region, acquisition setting, image characteristics, and annotation type. Users should document all preprocessing operations and report performance by task. Users can reproduce the FMC_UIA challenge ranking by using the official evaluation protocol.

M2–US can support single– and multi–task learning, transfer learning, data-efficient learning, and cross-dataset generalization studies. However, it is not representative of all ultrasound systems, clinical populations, or diagnostic applications. Models trained on M2–US require independent external validation and should not be used directly for clinical decision–making. Dataloaders, baseline implementations, and evaluation code are available in the GitHub repository to support reproducible use of the dataset.

Users should cite both this Data Descriptor and the Zenodo dataset record when reusing the data.

Code availability

The reproducible experiment codebase is available at `https://github.com/m2-us/experiments`. The code for data preprocessing and data loading is available at `https://github.com/m2-us/m2-us`.

Data availability

The dataset is publicly available at `https://m2-us.github.io/` and `https://doi.org/10.5281/zenodo.23558392`. License: CC BY 4.0.

Acknowledgements

This work was funded by Taighde Éireann – Research Ireland through the Research Ireland Centre for Research Training in Machine Learning

(18/CRT/6183). This research was supported by a grant from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (AIMIX project - Grant Agreement No. 101044779). AJS has received funding under the Horizon Europe grant 101213369 (DVPS).

Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

Ethics statement

All private clinical data included in this study were retrospectively collected under protocols reviewed and approved by the relevant Institutional Review Boards (IRBs) or Ethics Committees. Fetal head and pubic symphysis segmentation data were collected with approval from the Institutional Review Board of Guangzhou Women and Children's Medical Center, Guangzhou, China (No. 405A01), and the cardiac data were collected under approval No. 279A01 from the same institution. Cervix segmentation data were approved by the Institutional Review Board of Zhujiang Hospital, Southern Medical University, Guangzhou, China (No. 2023-SYJS-023). Data collected at the First Affiliated Hospital of Jinan University, Guangzhou, China, including cervical os and fetal head segmentation data, fetal head position and breech position classification data, and angle of progression, cervical length, and femur length regression data, were collected under approval No. JNUKY-2022-019. Fetal lung data used for both segmentation and classification were collected under approval No. JNUKY-2023-050 from the same institution. The private thyroid nodule detection data were approved by the Institutional Review Board of the First Affiliated Hospital of Naval Military Medical University, Shanghai, China (No. CHEC2025-077). All data were processed in accordance with the requirements of the respective institutional approvals and were de-identified before inclusion in the dataset.

References

1. Website. `http://119.29.231.17:90/index.html`, accessed: 6 October 2026
2. Alshagathrh, F., Alzubaidi, M., Alswat, K., Aldhebaib, A., Alahmadi, B., Alkubeyyer, M., Alosaimi, A., Alsadoon, A., Alkhamash, M., Schneider, J., Househ, M.: Large annotated ultrasound dataset of non-alcoholic fatty liver from saudi hospitals for analysis and applications. Data in Brief **58**, 111266 (2025). `https://doi.org/10.1016/j.dib.2024.111266`
3. Bai, J., Tang, Y., Liu, X., Hu, J., Li, Y., Chen, X., Wang, Y., Ma, C., Li, Y., Guo, B., Jiao, J., Huang, Y., Wang, K., Li, L., Ma, Y., Han, X., Shao, H., Yang, Z., Liu, Q., Hu, Y., Kuang, J., Song, S., Krishna, A., Khan, Z.A., Li, Z., Zhang, Z., Zhang, H., Cheng, Y., Zhang, X., Chen, X., Yan, H., Tong, L., Du, B., Deng, B., Chen, Y., Peng, Z., Rezaei, S., Gan, J., Cai, W., Wang, F., Curran, K.M., Silvestre, G.,

Khobo, I., Lu, Y., Ni, D., Huang, Y., Yaqub, M., Ma, J., Lekadir, K., Li, S.: Iugc: A benchmark of landmark detection in end-to-end intrapartum ultrasound biometry. Medical Image Analysis **110**, 103960 (2026). https://doi.org/10.1016/j.media.2026.103960

4. Bai, J., Tang, Y., Zhou, Z., Islam, M., Tabassum, M., Almar-Munoz, E., Liu, H., Meng, H., Lv, N., Deng, B., Chen, Y., Peng, Z., Xiao, Y., Xiao, L., Tran, N.K., Phan-Le, D.P., Nguyen, H.D., Liu, X., Hu, J., Huang, M., Liang, J., Feng, C., Zhang, X., Tong, L., Du, B., Pham, H.H., Nguyen, T.H., Xu, M., Jiang, J., Zhang, J., Liu, Y., Hasan, M.K., Gan, J., Liang, Z., Cai, W., Huang, Y., Luo, G., Yaqub, M., Lekadir, K.: Fugc: Benchmarking semi-supervised learning methods for cervical segmentation. IEEE Transactions on Medical Imaging **45**(6), 2950–2960 (2026). https://doi.org/10.1109/TMI.2026.3666364
5. Bai, J., Yaqub, M., Ma, J., Lekadir, K., Gan, J., Cai, W., Ni, D., Li, S.: Foundation model challenge for ultrasound biometry (Apr 2026). https://doi.org/10.5281/zenodo.19736827
6. Bai, J., Zhou, Z., Ou, Z., Koehler, G., Stock, R., Maier-Hein, K., Elbatel, M., Martí, R., Li, X., Qiu, Y., Gou, P., Chen, G., Zhao, L., Zhang, J., Dai, Y., Wang, F., Silvestre, G., Curran, K., Sun, H., Xu, J., Cai, P., Jiang, L., Lan, L., Ni, D., Zhong, M., Chen, G., Campello, V.M., Lu, Y., Lekadir, K.: Psfhs challenge report: Pubic symphysis and fetal head segmentation from intrapartum ultrasound images. Medical Image Analysis **99**, 103353 (2025). https://doi.org/10.1016/j.media.2024.103353
7. Bai, J., Zhou, Z., Tang, Y., Gan, J., Liang, Z., Fan, J., Mcguire, L.B., Clarke, J.L., Cai, W., Spurway, J., Tan, Y., Wang, S., Shen, W., Yu, W., Li, Y., Zhang, P., Jiang, W., Li, Y., Al Nasi, S.M.A.B., Abzhanov, A., Saeed, N., Yaqub, M., Xia, Z., Li, H., Lan, L., Ramesh, J., Bacher, V., Eid, M., Kalabizadeh, H., Rupprecht, C., Namburete, A.I., Yeung, P.H., Wyburd, M.K., Dinsdale, N.K., Serikbey, A., Li, J., Chen, S.L., Hu, Z., Liu, N., Deng, Y., Hu, W., Tan, C., Zhang, W., Nhi, M.T., Koehler, G., Stock, R., Maier-Hein, K., Elbatel, M., Li, X., Slimani, S., Campello, V.M., Ohene-Botwe, B., Khobo, I., Huang, Y., Han, Z., Hou, H., Qiu, D., Zheng, Z., Luo, G., Ni, D., Lu, Y., Lekadir, K., Li, S.: Beyond benchmarks of iugc: Rethinking requirements of deep learning method for intrapartum ultrasound biometry from fetal ultrasound videos. Medical Image Analysis **111**, 104043 (2026). https://doi.org/10.1016/j.media.2026.104043
8. Balinisteanu, A., Duchenne, J., Puvrez, A., Wouters, L., Bézy, S., Youssef, A., Minten, L., Bekhuis, Y., Van Langenhoven, L., Papangelopoulou, K., Cieplucha, A., Cattapan, I., Tostes, P., Bogaert, J., Vinereanu, D., Thomas, J.D., Badano, L.P., Voigt, J.U.: Vendor differences in 2d-speckle tracking global longitudinal strain: an update on a 10-year standardization effort. European Heart Journal - Cardiovascular Imaging **26**(8), 1360–1373 (Aug 2025)
9. Blaivas, M., Blaivas, L.N., Tsung, J.W.: Deep learning pitfall: Impact of novel ultrasound equipment introduction on algorithm performance and the realities of domain adaptation. Journal of Ultrasound in Medicine **41**(4), 855–863 (2022). https://doi.org/10.1002/jum.15765
10. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021)
11. Duong, V.H., Vu, H., Phan, H.D., Nguyen, D.Q., Pham, D.H., Le, Q.T., Nguyen, B.S., Do, T.D., Dinh, V.S., Nguyen, T.C., Pham, H.H., Ngo, D.H.: ThyroidXL:

Advancing thyroid nodule diagnosis with an Expert-Labeled, Pathology-Validated dataset. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. pp. 616–626. Springer Nature Switzerland (2026)
12. Fiorentino, M.C., Villani, F.P., Di Cosmo, M., Frontoni, E., Moccia, S.: A review on deep-learning algorithms for fetal ultrasound-image analysis. Medical Image Analysis **83**, 102629 (2023). https://doi.org/10.1016/j.media.2022.102629
13. Gunabushanam, G., Scoutt, L.M.: Ultrasound image optimization for the interventional radiologist. Techniques in Vascular and Interventional Radiology **24**(3), 100766 (Sep 2021)
14. Guo, X., Alsharid, M., Zhao, H., Wang, Y., Lander, J., Papageorghiou, A.T., Noble, J.A.: A visually grounded language model for fetal ultrasound understanding. Nature Biomedical Engineering **10**(8), 1629–1645 (Aug 2026)
15. van den Heuvel, T.L.A., de Bruijn, D., de Korte, C.L., Ginneken, B.v.: Automated measurement of fetal head circumference using 2d ultrasound images. PLOS ONE **13**(8), 1–20 (08 2018). https://doi.org/10.1371/journal.pone.0200412
16. Holmes, J.F., Kelley, K.M., Wootton-Gorges, S.L., Utter, G.H., Abramson, L.P., Rose, J.S., Tancredi, D.J., Kuppermann, N.: Effect of abdominal ultrasound on clinical care, outcomes, and resource use among children with blunt torso trauma: A randomized clinical trial. JAMA **317**(22), 2290–2296 (Jun 2017)
17. Huang, Q., Lin, Z., Han, L., Zhang, T., Wang, X., Zhou, Y., Zhang, Y., Xu, D., Bao, L., Mann, R., Wu, S., Ni, D., Shao, L., Tan, T.: Universal ultrasound image & video analysis challenge: Multi-organ classification and segmentation across b-mode and contrast-enhanced ultrasound (Apr 2026). https://doi.org/10.5281/zenodo.19729665
18. Ihnatsenka, B., Boezaart, A.: Ultrasound: Basic understanding and learning the language. International Journal of Shoulder Surgery **4**, 55 (2010)
19. Li, J., Zhu, L., Shen, G., Zhao, B., Hu, Y., Zhang, H., Wang, W., Wang, Q.: Liver lesion segmentation in ultrasound: A benchmark and a baseline network. Computerized Medical Imaging and Graphics **123**, 102523 (2025). https://doi.org/10.1016/j.compmedimag.2025.102523
20. Li, Q., Huang, Y., Saeed, S.U., Barratt, D.C., Clarkson, M.J., Vercauteren, T., Hu, Y.: Trackerless 3d freehand ultrasound reconstruction challenge 2025 (tus-rec2025) - train dataset (Apr 2025). https://doi.org/10.5281/zenodo.15224704
21. Li, Y., Xu, Q., Zhang, Y., He, X., Zhang, Q., Yao, Y., Tesem, F.B., Chen, X., Wang, R., Chen, Z., et al.: Uniultra: Interactive parameter-efficient sam2 for universal ultrasound segmentation. IEEE Transactions on Multimedia (2026)
22. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 936–944 (2017). https://doi.org/10.1109/CVPR.2017.106
23. Lin, Z., Han, L., Zhang, T., Wang, X., Zhou, Y., Zhang, Y., Xu, D., Bao, L., Mann, R., Li, S., Wu, S., Ni, D., Tan, T.: Universal ultrasound image challenge: Multi-organ classification and segmentation (uusic25) (Mar 2025). https://doi.org/10.5281/zenodo.15094669
24. Maani, F., Saeed, N., Saleem, T.J., Farooq, Z., Alasmawi, H., Diehl, W., Mohammad, A., Waring, G., Valappil, S., Bricker, L., Yaqub, M.: FetalCLIP: a visual-language foundation model for fetal ultrasound image analysis. npj Digital Medicine (Jun 2026)
25. McLario, D.J., Sivitz, A.B.: Point-of-Care ultrasound in pediatric clinical care. JAMA Pediatrics **169**(6), 594–600 (Jun 2015)

26. Noble, J., Boukerroui, D.: Ultrasound image segmentation: a survey. IEEE Transactions on Medical Imaging **25**(8), 987–1010 (2006). https://doi.org/10.1109/TMI.2006.877092
27. Rueda, S., Fathima, S., Knight, C.L., Yaqub, M., Papageorghiou, A.T., Rahmatullah, B., Foi, A., Maggioni, M., Pepe, A., Tohka, J., Stebbing, R.V., McManigle, J.E., Ciurte, A., Bresson, X., Cuadra, M.B., Sun, C., Ponomarev, G.V., Gelfand, M.S., Kazanov, M.D., Wang, C.W., Chen, H.C., Peng, C.W., Hung, C.M., Noble, J.A.: Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: A grand challenge. IEEE Transactions on Medical Imaging **33**(4), 797–813 (2014). https://doi.org/10.1109/TMI.2013.2276943
28. Sappia, M.S., de Korte, C.L., van Ginneken, B., Ninalga, D., Kondo, S., Kasai, S., Hirasawa, K., Akumu, T., Martín-Isla, C., Lekadir, K., Campello, V.M., Fabila, J., Beverdam, A., van Dillen, J., Neff, C., Murphy, K.: Acouslic-ai challenge report: Fetal abdominal circumference measurement on blind-sweep ultrasound data from low-income countries. Medical Image Analysis **105**, 103640 (2025). https://doi.org/10.1016/j.media.2025.103640
29. Seddiki, R., Mirault, T., Sitruk, J., Mohamedi, N., Messas, E., Pernot, M., Baranger, J., Goudot, G.: Advancements in noncontrast ultrasound imaging for Low-Velocity flow: A technical review and clinical applications in vascular medicine. Ultrasound in Medicine & Biology **51**(7), 1035–1042 (Jul 2025)
30. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., Massa, F., Haziza, D., Wehrstedt, L., Wang, J., Darcet, T., Moutakanni, T., Sentana, L., Roberts, C., Vedaldi, A., Tolan, J., Brandt, J., Couprie, C., Mairal, J., Jégou, H., Labatut, P., Bojanowski, P.: DINOv3 (2025), https://arxiv.org/abs/2508.10104
31. Sorensen, B., Hunskaar, S.: Point-of-care ultrasound in primary care: a systematic review of generalist performed point-of-care ultrasound in unselected populations. The Ultrasound Journal **11**(1), 31 (Nov 2019)
32. Soylu, U., Chandrasekeran, V., Czarnota, G.J., Oelze, M.L.: Migration of deep learning models across ultrasound scanners. IEEE Transactions on Biomedical Engineering **72**(11), 3210–3220 (2025). https://doi.org/10.1109/TBME.2025.3564567
33. Speranza, G., Mischkewitz, S., Al-Noor, F., Kainz, B.: Value of clinical review for AI-guided deep vein thrombosis diagnosis with ultrasound imaging by non-expert operators. npj Digital Medicine **8**(1), 135 (Mar 2025)
34. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PmLR (2019)
35. Tao, X., Cao, Y., Jiang, Y., Wu, X., Yan, D., Xue, W., Zhuang, S., Yang, X., Huang, R., Zhang, J., Ni, D.: Enhancing lesion detection in automated breast ultrasound using unsupervised multi-view contrastive learning with 3D DETR. Medical Image Analysis **101**, 103466 (Apr 2025)
36. Vallez, N., Bueno, G., Deniz, O., Rienda, M.A., Pastor, C.: BUS-UCLM: Breast ultrasound lesion segmentation dataset. Scientific Data **12**(1), 242 (Feb 2025)
37. Villemain, O., Baranger, J., Friedberg, M.K., Papadacci, C., Dizeux, A., Messas, E., Tanter, M., Pernot, M., Mertens, L.: Ultrafast ultrasound imaging in pediatric and adult cardiology: Techniques, applications, and perspectives. JACC: Cardiovascular Imaging **13**(8), 1771–1791 (Aug 2020)
38. Wells, P.N.T.: Ultrasound imaging. Physics in Medicine & Biology **51**(13), R83 (jun 2006). https://doi.org/10.1088/0031-9155/51/13/R06

39. Wu, L., Cavalcanti, N.A., Seibold, M., Loggia, G., Reissner, L., Hein, J., Beeler, S., Viehöfer, A., Wirth, S., Calvet, L., Fürnstahl, P.: Ultrabones100k: A reliable automated labeling method and large-scale dataset for ultrasound-based bone surface extraction. Computers in Biology and Medicine **194**, 110435 (2025). https://doi.org/10.1016/j.compbiomed.2025.110435
40. Zhang, J., Liu, S., Luo, X., Lyu, G., Shen, L.: Thyro-lmd: A benchmark dataset and sample-driven data loading, attention, and regularization for long-tailed multi-label thyroid ultrasound diagnosis. IEEE Transactions on Medical Imaging **45**(7), 3986–4001 (2026). https://doi.org/10.1109/TMI.2026.3690144
41. Zhang, Y., Yan, Y., Wang, K., Cai, M., Xiang, Y., Guo, Y., Tu, P., Ying, T., Chen, X.: A navigation-guided 3D breast ultrasound scanning and reconstruction system for automated multi-lesion spatial localization and diagnosis. Medical Image Analysis **110**, 103965 (May 2026)
42. Zhang, Y., Yang, X., Ji, C., Hu, X., Cao, Y., Chen, C., Sui, H., Li, B., Zhen, C., Huang, W., Deng, X., Yin, L., Ni, D.: Deep learning model for real-time nuchal translucency assessment at prenatal us. Radiology: Artificial Intelligence **7**(4), e240498 (2025). https://doi.org/10.1148/ryai.240498, pMID: 40304583
43. Zhao, H., Zheng, Q., Teng, C., Yasrab, R., Drukker, L., Papageorghiou, A.T., Noble, J.A.: Memory-based unsupervised video clinical quality assessment with multi-modality data in fetal ultrasound. Medical Image Analysis **90**, 102977 (2023). https://doi.org/10.1016/j.media.2023.102977