This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Oculo: A multilabel dataset for identification of ocular abnormalities from ultrasound images

Sneha Kumari[1,2], Payal Shah[2], Vaibhav Ganatra[3], Meenakshi Mahesh[2], and Mohit Jain[1⋆]

[1] Microsoft Research, India
[2] Sankara Eye Hospital, India
[3] University of Toronto, Canada
payalnaresh@sankaraeye.com, mohja@microsoft.com

Abstract. Ocular ultrasound is a critical imaging modality for assessing vitreoretinal pathologies, such as retinal detachment, vitreous hemorrhage, and intraocular tumors. However, interpretation requires specialized expertise and exhibits substantial inter-rater variability, limiting access in resource-constrained settings. Although prior studies have applied deep learning (DL) to ocular ultrasound analysis, progress has been constrained by the lack of public datasets and standardized benchmarks. In this work, we introduce *Oculo*, a publicly available multi-label dataset comprising 1,630 B-scan images from 1,242 patients, annotated for five ophthalmic abnormalities. We systematically benchmark the performance of four DL backbones and three domain-specific foundation models, across two task settings and three training strategies. We find that EfficientNet-B0 [27] with full-finetuning achieves the best mean performance (F1 = 0.77) across pathology labels. *Oculo* provides a foundation for standardized evaluation and future methodological advances in automated ocular ultrasound analysis. Dataset and source code available here: https://sri-kanchi-kamakoti-medical-trust.github.io/Oculo/

Keywords: Deep Learning · Foundation Model · Vitreoretinal health

1 Introduction

Ocular ultrasound is the preferred imaging modality for assessing and monitoring vitreoretinal health, especially when media opacities in the eye obscure direct visualization by slit lamp or ophthalmoscope. It is widely used to diagnose vitreous hemorrhage and retinal detachments; detect, measure, and monitor ocular tumors such as choroidal melanomas and retinoblastomas; and identify intraocular foreign bodies following trauma. However, interpretation of ocular ultrasounds is challenging and requires specialized training to differentiate between subtle pathological features. Typically, specialists are accessible only in tertiary clinical settings and specialized ophthalmic centers, thereby limiting access in primary care and resource-constrained settings, and even in emergency rooms [3, 10, 17,

⋆ Corresponding author.

25, 28, 29]. Additionally, interpretation exhibits significant inter-rater variability [4]. Automated ocular ultrasound analysis is therefore needed to improve access and standardization in vitreoretinal care.

Prior studies have explored deep learning (DL) models for interpreting ocular ultrasound to detect various ophthalmic disorders (Table 1), including vitreous opacities [6], retinal detachment [1, 4], contour abnormalities [13], and intraocular tumors [30]. However, these works rely on private institutional datasets. This lack of public datasets and code limits reproducibility and hinders objective evaluation of novel advancements in the field. With the emergence of self-supervised learning, domain-specific foundation models have been developed in ultrasonography (e.g., OpenUS [31]) and ophthalmology (e.g., VisionFM [18]). Due to the unavailability of public datasets, VisionFM is pretrained on proprietary data, and OpenUS does not include ocular ultrasound. Recently, the ERDES (Eye Retinal DEtachment ultraSound) dataset [17] was released as the first public ocular ultrasound dataset, but its annotations are limited to retinal detachment. Beyond ocular ultrasound, AI has been increasingly adopted across ophthalmology [7, 15] for surgical workflow understanding [14, 22], complication detection [23], clinical decision support [24], and patient education [19], highlighting the growing role of AI throughout the ophthalmic care continuum.

In this work, we introduce *Oculo*, a multi-label dataset for identifying ophthalmic abnormalities from ocular ultrasound (also called B-scan) images. The dataset comprises 1,630 B-scan images from 1,242 patients, annotated for five abnormalities, including vitreous dot echo (508 cases), retinal detachment (186), mass lesion (115), and contour abnormalities (92 cases). Next, we benchmark existing DL models across architectures, task-settings, pretraining regimes, and finetuning strategies. Our best-performing approach, based on EfficientNet-B0 [27], achieves a mean F1 score of 0.92 (single-task setting) and 0.77 (multi-task setting), across labels. Specifically, we make the following contributions:

1. We introduce *Oculo*, the first publicly available ocular ultrasound dataset annotated for five ophthalmic abnormalities, spanning vitreous, retinal and choroidal conditions.
2. We benchmark four DL backbones on *Oculo* under single-task and multi-task settings. We observe that EfficientNet-B0 [27] with full-finetuning achieves the best overall performance.
3. We also evaluate three domain-specific foundation models - VisionFM [18] (ophthalmology) and USFM [11], OpenUS [31] (ultrasonography) - across three training strategies (full-finetuning, LP $\rightarrow$ FT[12], and linear probing). VisionFM and USFM demonstrate strong transferability, with linear probes achieving performance comparable to fully-finetuned EfficientNet-B0 model.

2 *Oculo*: An ocular ultrasound dataset

Data Collection. The dataset was collected at Sankara Eye Hospital, a tertiary eye hospital in India, comprising 15 ophthalmologists (including 3 vitreoretinal

Table 1: Summary of prior work on deep learning-based ocular ultrasound analysis. Notably, *Oculo* is the first publicly available ocular ultrasound dataset with annotations for multiple ophthalmic abnormalities.

Study	Public dataset	Ophthalmic Abnormalities considered				
		Vitreous Dot Echo	Vitreous Detachment	Retinal Detachment	Abnormal Contour	Mass Lesion
Feng et al. [6]	✗	✓	✗	✗	✗	✗
Chen et al. [4]	✗	✓	✓	✓	✗	✗
Adithya et al. [1]	✗	✓	✓	✓	✗	✗
Li et al. [13]	✗	✓	✓	✓	✓	✗
Wang et al. [30]	✗	✓	✗	✓	✓	✓
Navard et al. [17]	✓	✗	✗	✓	✗	✗
Oculo (ours)	✓	✓	✓	✓	✓	✓

specialists), over 10 optometrists, and a daily patient volume $\geq$500. Scans were acquired over $\sim$4 years (May 2022 - January 2026). Informed consent was obtained from all patients.

The study protocol was approved by the hospital's Institutional Review Board (#SEH/BLR/EC/2024/256).

B-scan images were captured using the Appasamy Marvel II [2] ultrasound system. The device uses a 12 MHz transducer in B-scan mode to visualize the posterior segment, including the vitreous cavity, retina, optic nerve head, and surrounding tissues. For image acquisition, patients were seated with eyes closed. The probe was placed over the eyelid after a sterile water-soluble coupling gel was applied to prevent signal attenuation. Real-time visualization on a connected display enabled optimal probe positioning, after which still frames were captured. Multiple orientations were captured per patient to ensure adequate visualization of pathology.

Images were exported in PNG format (resolution 900×900) to prevent compression artifacts. During offline quality control, images with obscured fields of view (due to improper probe placement or orientation) were excluded. The final dataset consists of 1,630 images from 1,242 patients.

Data Annotation. The dataset was annotated for eight binary labels corresponding to distinct ophthalmic abnormalities (summarized in Table. 2). Two senior ophthalmologists with expertise in ocular ultrasound interpretation independently performed the annotations. Prior to labeling, they jointly annotated a calibration set to establish consensus-based guidelines. Inter-rater agreement was high, with a mean Cohen's κ of 0.86 (96.7% agreement) across labels, indicating reliable annotation quality. All annotations were subsequently reviewed by a senior retina specialist with 16 years of clinical experience, who adjudicated ambiguous cases to finalize the ground truth labels.

Dataset Characteristics. Of the 1,630 B-scan images, 854 (52.4%) show no abnormality, 415 (25.5%) contain one abnormality, 228 (14.0%) contain two, 107 (6.6%) contain three, and 26 (1.6%) contain four or more concurrent abnor-

Table 2: Description and prevalence of pathologies annotated in *Oculo*. The three rare labels (grayed) are released but excluded from evaluation.

Pathology	Description	Prev.
Vitreous Dot Echo	Scattered bright specks in the vitreous cavity; commonly seen in vitreous hemorrhage or vitritis	508 (31.2%)
Membranous Echo	Presence of thin strands in the vitreous cavity	339 (20.8%)
Retinal Detachment	High-amplitude membranous echo tethered to the optic nerve head	186 (11.4%)
Mass Lesion	Solid well-defined bright mass on posterior globe wall; may cause posterior acoustic shadowing (e.g., retinoblastoma)	115 (7.1%)
Abnormal Contour	Deformed globe wall morphology; typically a bulge or flattening of the sclera (e.g., posterior staphyloma, coloboma)	92 (5.6%)
Choroidal Detachment	Thick, smooth, dome-shaped or biconvex membrane	29 (1.8%)
Phthisis Bulbi	Small, irregularly shaped globe with loss of anatomical landmarks.	19 (1.2%)
Posterior Vitreous Detachment	Thin, linear membrane separated from the retinal surface; may (not) be separated from the optic disc	17 (1%)

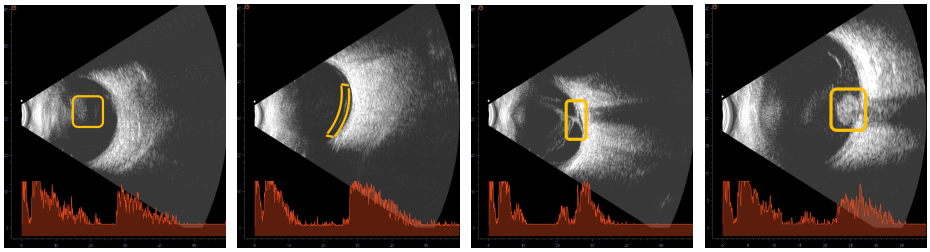
malities. The maximum observed co-occurrence is six abnormalities in a single scan. The prevalence of each label is summarized in Table 2, and representative examples are shown in Figure 1.

The dataset exhibits significant class imbalance. Vitreous Dot Echo is the most prevalent label (31.2%), whereas Posterior Vitreous Detachment (PVD), Choroidal Detachment (CD), and Phthisis Bulbi (PB) have limited positive samples - 17 (1%), 29 (1.8%), and 19 (1.2%), respectively. CD and PB are inherently uncommon conditions. Although PVD is common in general population, only complete PVD was labeled positive in our dataset, as partial PVD is difficult to reliably identify from static images. Partial PVD cases were labeled as membraneous echo (ME) when appropriate. Owing to their low prevalence, PVD, CD, and PB are excluded from our evaluation experiments. Nevertheless, all labeled images are released to support future research on rare conditions.

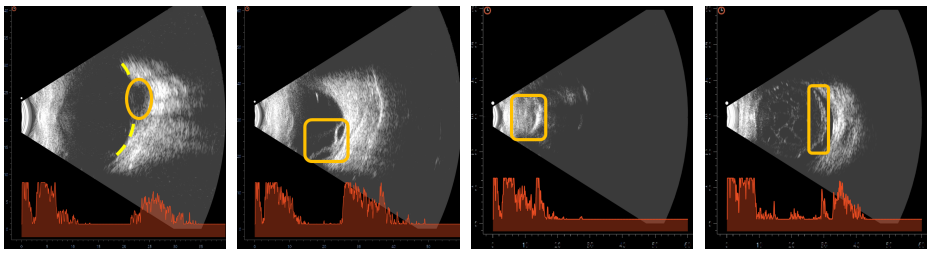
3 Benchmarking Deep Learning (DL) Models on *Oculo*

We evaluate multiple DL models on the *Oculo* dataset.

Models. We benchmark four image classification architectures: EfficientNet-B0 [27], ResNet50 [9], VGG-19-BN [26], and ViT-B-16 [5]. ResNet and Vision Transformers (ViT) are standard architectures for image classification, while EfficientNet and VGG represent parameter-efficient and overparameterized variants, respectively. Additionally, given the recent surge in foundation models, we evaluate three foundation models (FMs): VisionFM [18], USFM [11], and OpenUS [31]. USFM and OpenUS are pretrained on large corpora of ultrasound images spanning multiple anatomical regions (excluding ocular scans), thereby incorporating modality-specific priors. In contrast, VisionFM is an ophthalmic FM trained on diverse ophthalmic images, including fundus photography, optical



(a) VDE (b) ME (c) RD (d) Mass Lesion



(e) AC (f) CD (g) PB (h) PVD

Fig. 1: Representative samples for pathological labels annotated in the *Oculo* dataset. VDE: Vitreous Dot Echo, ME: Membranous Echo, RD: Retinal Detachment, AC: Abnormal Contour, CD: Choroidal Detachment, PB: Phthisis Bulbi, PVD: Posterior Vitreous Detachment.

coherence tomography (OCT), *ocular ultrasounds*, slit-lamp biomicroscopy, and external eye images, thereby incorporating anatomical priors of the eye.

Task Settings. Models are evaluated in both single-task and multi-task settings. In the single task setting, separate models are trained for each pathology label. In the multi-task setting, a single model is trained for all labels, i.e., a shared encoder backbone is trained with label-specific linear heads, and the overall loss is computed as the average of individual losses.

Training Strategies. All models are initialized with pretrained weights. CNNs and ViTs use ImageNet [21] initialization, whereas FMs use their pretrained weights. We evaluate three training paradigms: full-finetuning, linear probing, and linear-probe-then-finetune [12]. All paradigms use a randomly initialized linear head atop representations extracted from the backbone. In full-finetuning, all model parameters are updated simultaneously from the onset of training. However, to mitigate the risk of pretrained feature distortion due to noisy gradients from the linear head [12], we also experiment with the linear-probe-then-finetune paradigm (LP $\rightarrow$ FT). In this setup, the backbone is initially frozen while only training the linear head. This is followed by full finetuning. In linear probing, the backbone is always frozen and only the linear head is trained. Hence, linear probing is applied only to FMs, since they already possess prior information about the task.

Table 3: Comparison of average performance across architectures in single-task and multi-task settings. Best performance in the single-task setting is highlighted in **bold**, and best performance in the multi-task setting is underlined.

Model	Setting	Performance Metrics					
		Acc.	Prec.	Rec.	Spec.	F1	AUROC
EfficientNet-B0 [27]	Single-Task	95.2	**96.2**	**88.6**	**98.4**	**0.92**	**0.98**
	Multi-Task	93.2	80.4	74.8	96.8	0.77	0.96
ResNet50 [9]	Single-Task	94.2	86.7	78.8	97.8	0.83	0.94
	Multi-Task	89.8	59.4	72.3	95.8	0.65	0.90
VGG-19-BN [26]	Single-Task	**97.3**	90.6	81.4	98.2	0.86	**0.98**
	Multi-Task	90.4	77.0	67.8	96.6	0.70	0.93
ViT-B-16 [5]	Single-Task	90.7	76.7	52.9	96.3	0.61	0.94
	Multi-Task	92.9	85.4	72.9	96.3	0.78	0.95

Evaluation. Performance is assessed per label using binary classification metrics: accuracy, precision, recall (sensitivity), specificity, F1 score, and AUROC. Mean performance across labels is used for model performance comparison.

The dataset is randomly split into training, validation, and test sets (75 / 10 / 15). All models are trained and evaluated on identical splits. Since the dataset contains multiple images per patient, the dataset is split at the patient level to prevent data leakage. Average results from 5 independent runs are reported. As mentioned earlier, PVD, CD, and PB classes were excluded, due to limited sample sizes; results are reported on the remaining five labels.

We provide all implementation details - data splits, learning rate schedules, hyperparameter configs, data augmentation, etc. in our code repository.

4 Results

Performance of DL models.

Table 3 compares the average performance of the four DL architectures under single-task and multi-task configurations. CNN backbones demonstrate strong single-task performance, with EfficientNet-B0 achieving the highest mean F1 score of 0.92 (accuracy 95.2%) across labels. This finding is consistent with prior studies [1, 4, 6, 13, 30]. However, CNN performance declines significantly in the multi-task setting (EfficientNet-B0 F1 score drops to 0.77). In contrast, the transformer-based ViT-B-16 exhibits the opposite trend: multi-task training improves its performance significantly, with the F1 score increasing from 0.61 (single-task) to 0.78 (multi-task), achieving the best multi-task performance. We attribute its weaker single-task results to the lack of image-specific inductive biases, which makes it suboptimal in limited-data settings [5]. Multi-task training likely provides richer supervision to ViT, enabling the model to leverage cross-label correlations and learn more robust representations.

Finally, the reduced multi-task performance of other backbones indicates that jointly optimizing multiple pathological labels is a challenging task.

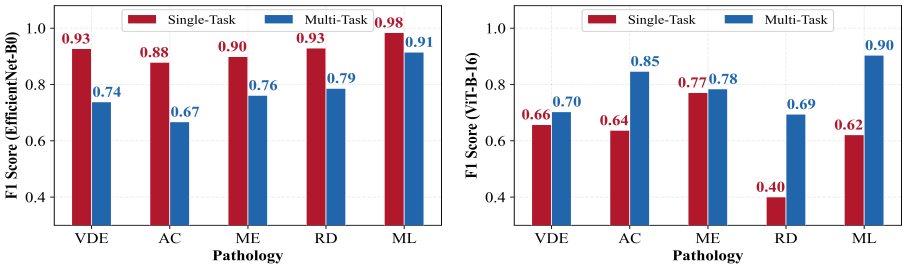


Fig. 2: Comparison of single-task and multi-task performance of EfficientNet-B0 (left) and ViT-B-16 (right) across individual pathology labels. VDE: Vitreous Dot Echo, AC: Abnormal Contour, ME: Membranous Echo, RD: Retinal Detachment, ML: Mass Lesion.

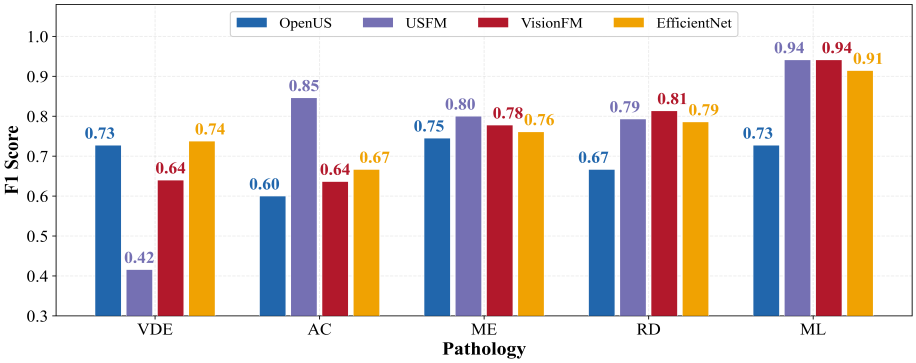


Fig. 3: Comparison of Foundation Models (linear probe) and EfficientNet-B0 (full fine-tuning) performance across individual pathology labels.

Figure 2 shows per-label performance for EfficientNet-B0 and ViT-B-16. Across settings, EfficientNet-B0 performs better on texture-dependent pathologies (e.g., VDE, ME) than on shape-dependent ones (e.g., AC). In contrast, ViT-B-16 achieves higher performance on shape-related tasks than on texture-based tasks. (Note: We do not analyze ViT-B-16 in the single-task setting due to its poor performance under data-constrained training.) These findings align with previous studies on inductive biases: CNNs exhibit a texture bias [8], whereas ViTs show a shape bias [16]. Both models perform well on mass lesions as they present characteristic shape (mushroom- or button-shaped) and texture (hyperechoic blobs).

Performance of Foundation Models.

Table 4 summarizes the performance of foundation models in ultrasonography (OpenUS [31], USFM [11]) and ophthalmology (VisionFM [18]). These models demonstrate strong transferability to *Oculo*, with USFM (F1: 0.76) and VisionFM (F1: 0.76) closely matching the performance of the best task-specific

Table 4: Comparison of fine-tuning strategies across models. We choose the best-performing dedicated backbone, EfficientNet-B0, as baseline.

Model	**Training Strategy**	**Performance Metrics**					
		Acc.	Prec.	Rec.	Spec.	F1	AUROC
OpenUS [31]	Full FT	87.7	73.5	39.3	94.4	0.46	0.77
	Linear Probe	92.1	83.1	62.0	96.3	0.69	0.93
	LP → FT	91.1	76.6	62.1	94.1	0.66	0.91
USFM [11]	Full FT	91.8	83.4	70.0	96.2	0.75	0.93
	Linear Probe	92.2	**87.8**	70.6	**98.4**	0.76	0.95
	LP → FT	90.4	77.8	66.8	95.9	0.71	0.92
VisionFM [18]	Full FT	88.2	73.5	41.3	97.3	0.52	0.90
	Linear Probe	92.9	86.1	70.2	97.7	0.76	0.94
	LP → FT	91.4	82.2	67.5	97.1	0.73	0.89
EfficientNet-B0 [27]	Full FT	**93.2**	80.4	**74.8**	96.8	**0.77**	**0.96**
	LP → FT	91.3	74.5	65.5	95.8	0.69	0.93

model, EfficientNet-B0 (F1: 0.77). While VisionFM's performance is expected given its pretraining on ocular ultrasound data, the performance disparity between the two general-purpose ultrasound models (USFM and OpenUS) is notable. Although neither is pretrained on ocular scans, USFM (F1: 0.76) significantly outperforms OpenUS (F1: 0.69). This difference can be attributed to the scale of their pretraining datasets: USFM was pretrained on 2M images, compared to 300K for OpenUS. These findings provide empirical evidence that large-scale pretraining can compensate for the lack of domain-specific data and generalize effectively to unseen medical domains.

Figure 3 shows per-label performance of linear probes trained on frozen foundation model representations. USFM and VisionFM achieve strong performance on Mass Lesions (F1: 0.94). While per-label performance depends on multiple factors (including pretraining dataset size, architecture, and hyperparameters), interestingly, USFM exhibits divergent performance on VDE and AC. We hypothesize that this behavior stems from its frequency-domain pretraining objective, which utilizes a focal frequency loss, that effectively emphasizes low spatial frequencies to capture macro-structural morphology (like in AC). Consequently, the model tends to categorize high-frequency specular data as inherent noise, leading to poor performance on VDE.

Table 4 also compares different training strategies. Interestingly, across all models, we observe that full finetuning reduces average model performance (F1 score) compared to training linear probes on frozen model representations. Potential explanations include domain shift, class imbalance, and the modest size of our dataset relative to the foundation models. Notably, Ruffini et al. [20] similarly observed that finetuning biomedical foundation models on small, imbalanced datasets can underperform linear probing.

5 Conclusion

In this work, we presented *Oculo*, a publicly available multi-label dataset for automated analysis of ocular ultrasound images, and established benchmark results across diverse deep learning and foundation model architectures. By addressing the lack of open datasets, our work enables reproducible evaluation and objective comparison of emerging methods for ocular ultrasound interpretation. We anticipate that *Oculo* will catalyze future research, enabling the development of improved representation learning strategies for ocular ultrasonography, more robust multi-label modeling approaches, and the incorporation of ocular ultrasound data into future ophthalmic and ultrasonography foundation models. A limitation of *Oculo* is that it is currently single-center and single-device; we believe its public release will enable future studies on generalization across sites, populations, and devices. Broadly, progress in automated ocular ultrasound analysis has the potential to improve accessibility, standardization, and scalability of retinal screening, particularly in resource-constrained and non-specialist settings.

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

References

1. Adithya, V.K., Baskaran, P., Aruna, S., Mohankumar, A., Hubschman, J.P., Shukla, A.G., Venkatesh, R.: Development and validation of an offline deep learning algorithm to detect vitreoretinal abnormalities on ocular ultrasound. Indian Journal of Ophthalmology **70**(4) (2022)
2. Appasamy Associates: Marvel ii a/b scan with ubm ultrasound scanner. https://www.appasamy.com/ultrasound-scanner/ab-scan-with-ubm, accessed: 27 Feb 2026
3. Baker, N., Amini, R., Situ-LaCasse, E.H., Acuña, J., Nuño, T., Stolz, U., Adhikari, S.: Can emergency physicians accurately distinguish retinal detachment from posterior vitreous detachment with point-of-care ocular ultrasound? Am. J. Emerg. Med. **36**(5), 774–776 (May 2018)
4. Chen, D., Yu, Y., Zhou, Y., Peng, B., Wang, Y., Hu, S., Tian, M., Wan, S., Gao, Y., Wang, Y., Yan, Y., Wu, L., Yao, L., Zheng, B., Wang, Y., Huang, Y., Chen, X., Yu, H., Yang, Y.: A deep learning model for screening multiple abnormal findings in ophthalmic ultrasonography (with video). Transl. Vis. Sci. Technol. **10**(4), 22 (Apr 2021)
5. 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)
6. Feng, L., Zhang, Y., Wei, W., Qiu, H., Shi, M.: Applying deep learning to recognize the properties of vitreous opacity in ophthalmic ultrasound images. Eye **38**(2), 380–385 (Feb 2024). https://doi.org/10.1038/s41433-023-02705-7

7. Gairola, S., Bohra, M., Shaheer, N., Jayaprakash, N., Joshi, P., Balasubramaniam, A., Murali, K., Kwatra, N., Jain, M.: Smartkc: Smartphone-based corneal topographer for keratoconus detection. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **5**(4) (Dec 2022). https://doi.org/10.1145/3494982
8. Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F.A., Brendel, W.: Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness (2022), https://arxiv.org/abs/1811.12231
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015), https://arxiv.org/abs/1512.03385
10. Jacobsen, B., Lahham, S., Lahham, S., Patel, A., Spann, S., Fox, J.C.: Retrospective review of ocular point-of-care ultrasound for detection of retinal detachment. West. J. Emerg. Med. **17**(2), 196–200 (Mar 2016)
11. Jiao, J., Zhou, J., Li, X., Xia, M., Huang, Y., Huang, L., Wang, N., Zhang, X., Zhou, S., Wang, Y., Guo, Y.: Usfm: A universal ultrasound foundation model generalized to tasks and organs towards label efficient image analysis. Medical Image Analysis **96**, 103202 (2024). https://doi.org/10.1016/j.media.2024.103202
12. Kumar, A., Raghunathan, A., Jones, R., Ma, T., Liang, P.: Fine-tuning can distort pretrained features and underperform out-of-distribution (2022), https://arxiv.org/abs/2202.10054
13. Li, Z., Yang, J., Wang, X., Zhou, S.: Establishment and evaluation of intelligent diagnostic model for ophthalmic ultrasound images based on deep learning. Ultrasound in Medicine & Biology **49**(8), 1760–1767 (2023). https://doi.org/10.1016/j.ultrasmedbio.2023.03.022
14. Mueller, S., Sachdeva, B., Prasad, S.N., Lechtenboehmer, R., Holz, F.G., Finger, R.P., Murali, K., Jain, M., Wintergerst, M.W.M., Schultz, T.: Phase recognition in manual small-incision cataract surgery with ms-tcn++ on the novel sics-105 dataset. Scientific Reports **15**(1), 16886 (May 2025). https://doi.org/10.1038/s41598-025-00303-z
15. Müller, S., Jain, M., Sachdeva, B., Shah, P.N., Holz, F.G., Finger, R.P., Murali, K., Wintergerst, M.W.M., Schultz, T.: Artificial intelligence in cataract surgery: A systematic review. Transl Vis Sci Technol **13**(4), 20 (Apr 2024)
16. Naseer, M., Ranasinghe, K., Khan, S., Hayat, M., Khan, F., Yang, M.H.: Intriguing properties of vision transformers. In: Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems (2021), https://openreview.net/forum?id=o2mbl-Hmfgd
17. Navard, P., Ozkut, Y., Adhikari, S., Situ-LaCasse, E., Acuña, J., Yarnish, A., Yilmaz, A.: Erdes: A benchmark video dataset for retinal detachment and macular status classification in ocular ultrasound (2025), https://arxiv.org/abs/2508.04735
18. Qiu, J., Wu, J., Wei, H., Shi, P., Zhang, M., Sun, Y., Li, L., Liu, H., Liu, H., Hou, S., Zhao, Y., Shi, X., Xian, J., Qu, X., Zhu, S., Pan, L., Chen, X., Zhang, X., Jiang, S., Wang, K., Yang, C., Chen, M., Fan, S., Hu, J., Lv, A., Miao, H., Guo, L., Zhang, S., Pei, C., Fan, X., Lei, J., Wei, T., Duan, J., Liu, C., Xia, X., Xiong, S., Li, J., Lam, K., Lo, B., Tham, Y.C., Wong, T.Y., Wang, N., Yuan, W.: Development and validation of a multimodal multitask vision foundation model for generalist ophthalmic artificial intelligence. NEJM AI **1**(12), AIoa2300221 (2024). https://doi.org/10.1056/AIoa2300221
19. Ramjee, P., Sachdeva, B., Golechha, S., Kulkarni, S., Fulari, G., Murali, K., Jain, M.: Cataractbot: An llm-powered expert-in-the-loop chatbot for cataract patients. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **9**(2) (Jun 2025). https://doi.org/10.1145/3729479

20. Ruffini, F., et al.: Benchmarking foundation models and parameter-efficient fine-tuning for prognosis prediction in medical imaging. Computer Methods and Programs in Biomedicine p. 109196 (2025)
21. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: Imagenet large scale visual recognition challenge (2015), https://arxiv.org/abs/1409.0575
22. Sachdeva, B., Akash, N., Ashraf, T., Müller, S., Schultz, T., Wintergerst, M.W.M., Singri, N., Murali, K., Jain, M.: Phase-informed tool segmentation for manual small-incision cataract surgery. In: MICCAI. pp. 446–455. Springer Nature (2025)
23. Sachdeva, B., Kumari, S., Agarwal, R., Kumaraswamy, S., Prasad, N.S., Mueller, S., Lechtenboehmer, R., Wintergerst, M.W.M., Schultz, T., Murali, K., Jain, M.: Cataractcompdetect: Intraoperative complication detection in cataract surgery (2025), https://arxiv.org/abs/2511.18968
24. Sanghvi, A., Akash, N., Imam, R., Sharma, A., Jain, M.: Medxagent: Multi-agent consultation for interactive medical diagnosis (2026), https://arxiv.org/abs/2606.03416
25. Sim, D., Hussain, A., Tebbal, A., Daly, S., Pringle, E., Ionides, A.: National survey of the management of eye emergencies in the accident and emergency departments by senior house officers: 10 years on–has anything changed? Emerg. Med. J. **25**(2), 76–77 (Feb 2008)
26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2015), https://arxiv.org/abs/1409.1556
27. Tan, M., Le, Q.: EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th Int Conf on Machine Learning. Proceedings of Machine Learning Research, vol. 97, pp. 6105–6114. PMLR (09–15 Jun 2019), https://proceedings.mlr.press/v97/tan19a.html
28. Uhr, J.H., Governatori, N.J., Zhang, Q.E., Hamershock, R., Radell, J.E., Lee, J.Y., Tatum, J., Wu, A.Y.: Training in and comfort with diagnosis and management of ophthalmic emergencies among emergency medicine physicians in the united states. EYE **34**(9), 1504–1511 (Sep 2020)
29. Vrablik, M.E., Snead, G.R., Minnigan, H.J., Kirschner, J.M., Emmett, T.W., Seupaul, R.A.: The diagnostic accuracy of bedside ocular ultrasonography for the diagnosis of retinal detachment: a systematic review and meta-analysis. Ann. Emerg. Med. **65**(2), 199–203.e1 (Feb 2015)
30. Wang, Y., Xu, Z., Dan, R., Yao, C., Shao, J., Sun, Y., Wang, Y., Ye, J.: Automated classification of multiple ophthalmic diseases using ultrasound images by deep learning. British Journal of Ophthalmology **108**(7), 999–1004 (2024). https://doi.org/10.1136/bjo-2022-322953
31. Zheng, X., Chen, X., Rauf, A., Fu, Q., Monosi, B., Rivellese, F., Lewis, M.J., Gong, S., Slabaugh, G.: Openus: A fully open-source foundation model for ultrasound image analysis via self-adaptive masked contrastive learning (2025), https://arxiv.org/abs/2511.11510