Beyond the Last Frame: Temporal Modelling of Fluorescein Angiography for Hyperfluorescence Classification

Tymo van Rijn[1,2], Roberto Pulvirenti[1], Mahshid Shiri[1], Ilenia Meloni[3], Oscar Jiménez[1], Mattia Tomasoni[3], Florence Hoogewoud[3], Eirini Kaisari[3], and André Anjos[1]

[1] Idiap Research Institute, Martigny, Switzerland
tymovanrijn@gmail.com
[2] Rotterdam University of Applied Sciences, Rotterdam, The Netherlands
[3] Department of Ophthalmology, University of Lausanne, Fondation Asile des Aveugles, Jules Gonin Eye Hospital, Lausanne, Switzerland

Abstract. Fluorescein angiography (FA) captures the passage of dye through the retinal vasculature over time, and the resulting hyperfluorescence patterns are central to assessing retinal vascular and inflammatory disease. Yet existing automated grading methods reduce each examination to a single frame, discarding the temporal evolution that clinicians rely on to distinguish these patterns. We investigate whether temporal modelling improves automated hyperfluorescence grading beyond single-frame approaches by representing each FA examination as a sequence of frames. We partition each sequence into Early, Mid, and Late phases using a data-driven procedure based on linear discriminant analysis of frame embeddings. Frames from each phase are encoded by the RETFound-Green retinal foundation model and aggregated by an attention-pooled gated recurrent unit (GRU) to classify five hyperfluorescence types. To enable temporal analysis, we reconstruct and release per-frame acquisition times in the public AngioReport FA dataset. Evaluated on held-out patients over five random seeds, our pipeline consistently improves discrimination over the state-of-the-art single-frame approach across all five hyperfluorescence types, raising the overall AUROC from 0.82 to 0.87, with the largest gains for types defined by their temporal evolution. Modelling the temporal dimension of FA, rather than a single frame, thus consistently improves grading, and our released timestamped dataset opens temporal FA analysis to the community. Code and the reconstructed timestamps are available at https://gitlab.idiap.ch/medai/software/paper/miccai-omia-beyond-the-last-frame.

Keywords: Fluorescein angiography · Hyperfluorescence classification · Temporal modelling · Foundation models · Retinal imaging

1 Introduction

Fluorescein angiography (FA) is a dynamic imaging modality used to assess retinal vascular and inflammatory disease. Following injection of fluorescein dye, a

sequence of retinal images is acquired as the dye traverses the ocular vasculature. This temporal evolution provides critical diagnostic information: clinically important hyperfluorescence patterns, including leakage, staining, and pooling, are distinguished not only by their appearance but by how fluorescence changes over time. For example, early leakage and late staining may appear similar in an isolated frame yet exhibit markedly different temporal behaviour throughout the angiographic transit (Fig. 1b). However, FA examinations pose additional challenges for temporal modelling: irregular temporal sampling, variable sequence length across examinations, and non-stationary frame appearance over the dye transit.

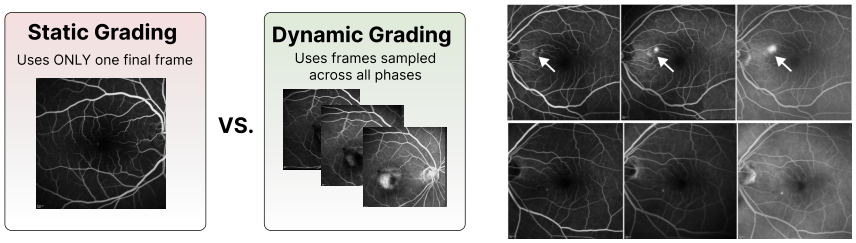
Despite the inherently dynamic nature of FA, most recent automated analysis methods treat an examination as a static classification problem. Existing approaches typically select a single representative frame, often the late-phase image, and perform grading using a single frame [1,19,20,11]. While effective for several tasks, this formulation discards the temporal information that clinicians routinely use during interpretation.

Consequently, the contribution of temporal dynamics to automated FA grading remains largely unexplored. In particular, it remains unclear whether temporal modelling improves grading performance beyond single-frame approaches and whether such benefits are consistent across different hyperfluorescence types. To address this limitation, this paper investigates two core questions: **(1)** Does sequential temporal modelling improve grading over single-frame models? **(2)** Which hyperfluorescence types benefit most from temporal context?

We answer these questions by revisiting FA grading from a temporal perspective. Using the AngioReport dataset [18], to our knowledge the only publicly available FA dataset with temporal data, we recover per-frame acquisition times to reconstruct the temporal progression of each examination. To facilitate temporal modelling, examinations are partitioned into clinically interpretable Early, Mid, and Late phases using a data-driven procedure based on Fisher's Linear Discriminant Analysis (LDA). These sequences are then processed by a lightweight framework combining a retinal foundation model encoder [16] with a gated recurrent unit (GRU) and attention pooling.

Across five hyperfluorescence types, incorporating temporal context consistently improves grading over strong single-frame baselines, suggesting that temporal dynamics are an underutilised source of diagnostic information in automated FA analysis.

The main contributions are: **(1) Dataset Enrichment:** Reconstructing and releasing the per-frame acquisition time of the public AngioReport FA dataset [18]. **(2) Temporal Grading Framework:** A temporal modelling pipeline that combines a retinal foundation model encoder [16] with GRU and attention-based aggregation to capture hyperfluorescence dynamics. **(3) Data-driven Phase Partitioning:** A parameter-free procedure that partitions examinations into clinically interpretable Early/Mid/Late phases without manual annotations, aligning with clinical FA interpretation. **(4) Experimental Analysis:** Study of temporal dynamics in hyperfluorescence grading, demonstrating



(a) (b)

Fig. 1: (a) Hyperfluorescence is defined by temporal evolution. (b) FA samples from early to late phases (left to right): (top) leakage, where the arrow marks the origin and hyperfluorescence progressively spreads and diffuses; (bottom) staining, showing a gradual increase in hyperfluorescence.

consistent overall AUROC gains (from $0.82 \rightarrow 0.87$) and clarifying which types benefit most from temporal context.

2 Related Work

Automated FA analysis. Clinical interpretation of FA follows standard grading of hyperfluorescence patterns (e.g., leakage, staining, pooling) [14,7], and recent studies have explored automated FA analysis to reduce the burden and variability of manual assessment. Most existing FA grading methods operate on individual frames rather than complete examinations. Prior work has addressed inflammatory findings in widefield FA [1], vascular leakage in uveitis [19], neovascular leakage in diabetic retinopathy [20], and retinal foundation-model representations [11], but all rely primarily on single-frame predictions. While Shen et al. [13] demonstrated the value of temporal fluorescence patterns in diabetic retinopathy, their task-specific approach does not generalise to examination-level hyperfluorescence-type grading. As a result, the role of temporal context in FA hyperfluorescence grading remains unclear.

Retinal foundation models. Recent retinal foundation models show strong transferability across ophthalmic imaging tasks [21,16], providing robust visual representations under limited supervision. Their computational efficiency further enables frame-wise processing of complete FA examinations, making sequence-level modelling feasible at low computational cost.

Temporal modelling in medical imaging. Recurrent architectures such as LSTMs [6] and GRUs [4], as well as attention-based approaches [2,15], have been successfully applied to dynamic imaging tasks including echocardiography [9] and surgical video analysis [12]. These methods demonstrate that modelling temporal evolution can improve performance compared with frame-wise analysis, whereas FA research has largely remained focused on static-image formulations.

3 Methods

3.1 Dataset Enrichment and Problem Formulation

Experiments are conducted on the public AngioReport FA dataset [18], augmented with recovered per-frame temporal information. Frame acquisition times are extracted using a prompt-engineered Optical Character Recognition (OCR) pipeline based on Gemini 2.5 Flash [5], which reads the burned-in timestamps present in each image and excludes frames identified as indocyanine-green angiography (ICGA). To verify the recovered timing, we manually annotated the elapsed acquisition time for 150 randomly selected frames and compared them against the OCR output, which matched the manual ground truth on all 150 frames (100% agreement). These frames were drawn uniformly at random from the full corpus. The recovered timestamps are released publicly as a static annotation file, so reproducing our experiments requires neither access to nor re-execution of the OCR pipeline. The final dataset contains 997 patients, 1,877 examinations (some being unilateral), and 31,239 timestamped FA frames, with a median of 12 frames per examination. Hyperfluorescence labels are those distributed with AngioReport, where images were annotated by retinal specialists [18]; inter-observer agreement is not reported by the dataset authors, and any residual label noise affects all compared methods equally. Data are split at the patient level into training, validation, and test sets using a 70/15/15 ratio, ensuring that both eyes from a patient remain in the same split.

Following the state-of-the-art (SOTA) [11], frames are resized to 224×224. Hyperfluorescence (HyperF) grading is formulated as an examination-level classification task with label space $\mathcal{Y} = \{\text{None}, \text{Leakage}, \text{Pooling}, \text{Staining}, \text{Window Defect}\}$. An examination is a variable-length temporal sequence $X = \big((x_i, t_i)\big)_{i=1}^{N}$, where x_i denotes the i-th FA frame, t_i its acquisition time, and N the sequence length. Each examination is associated with a HyperF label $y \in \mathcal{Y}$.

3.2 Data-driven phase partitioning

We partition each examination into three dye-transit phases $c \in \{\text{Early, Mid, Late}\}$ so that the temporal structure of an examination is represented uniformly regardless of its length. Grouping by phase ensures every stage of the dye transit is represented, which matters for long examinations exceeding one hundred frames. The boundaries are data-driven, as the dataset provides no clinical-phase annotations. We define a finite set of candidate boundary pairs by sampling $Q = 50$ equally-spaced quantiles of the empirical frame-timestamp distribution between the 5th and 95th percentiles. Each ordered pair $(a < b)$ drawn from this grid partitions every frame by the phase — Early, Mid, or Late — into which its timestamp falls. Each frame is encoded independently by a frozen RETFound-Green [16], denoted f, yielding a frame-level embedding $z_i = f(x_i) \in \mathbb{R}^{384}$. We score each candidate partition with Fisher's discriminant criterion J, the ratio of between- to within-phase scatter of the embeddings [3], and select the time

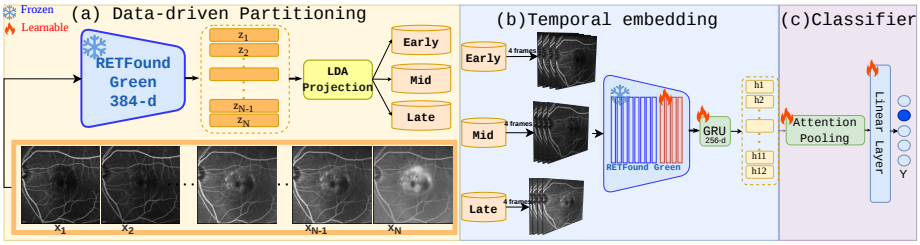


Fig. 2: Overview of the proposed framework. **(a)** FA examinations are partitioned into Early, Mid, and Late phases. **(b)** Up to four frames per phase are encoded by RETFound-Green and sequentially processed by a GRU to capture temporal dynamics. **(c)** Attention pooling aggregates the hidden states produced by the GRU, and a linear classifier predicts the hyperfluorescence grade.

borders from dye injection that maximise it,

$$(a^\star, b^\star) = \arg\max_{a<b} \; J(a, b). \tag{1}$$

The selected phases tend to align with the points of greatest change in the embedding trajectory across the dye transit (Fig. 2).

3.3 Temporal grading pipeline

Our pipeline encodes the selected frames, aggregates their embeddings as an ordered sequence, and classifies that sequence into a HyperF type. We sample up to four frames from each phase; phases with fewer frames contribute all of theirs, so each examination yields a sequence of $L \leq 12$ frames. Each frame is encoded by the RETFound-Green retinal model, denoted by f_{θ_f,θ_t}, where θ_f is the frozen parameters and θ_t the trainable parameters corresponding to the last four Transformer blocks. These embeddings are passed in temporal order to a single-layer, unidirectional GRU with a 256-dimensional hidden state [4]; sequences shorter than twelve frames are zero-padded and packed with `pack_padded_sequence`, so the recurrence is computed only over the L real frames. Instead of relying only on the final hidden state, which may place greater emphasis on later frames, we aggregate the per-frame hidden states $\mathbf{h}_1, \ldots, \mathbf{h}_L \in \mathbb{R}^{256}$ with an attention-pooling layer. A learned vector $\mathbf{w}$ assigns a score to each state, and a softmax over the sequence converts these scores into per-frame attention weights α_k; padded positions are masked out, so the softmax normalises only over the L real frames. The weighted sum forms the sequence representation $\mathbf{c}$,

$$\alpha_k = \frac{\exp(\mathbf{w}^\top \mathbf{h}_k)}{\sum_{j=1}^{L} \exp(\mathbf{w}^\top \mathbf{h}_j)}, \qquad \mathbf{c} = \sum_{k=1}^{L} \alpha_k \, \mathbf{h}_k. \tag{2}$$

A linear head maps $\mathbf{c}$ to the five HyperF-type logits, so the model learns how much each frame contributes to the prediction. During training, the parameters θ_t are optimised jointly with the GRU and classification head.

3.4 Baselines

To evaluate the performance gains of the proposed framework, we compare it against both SOTA and an ablation configuration: **SOTA (Static):** To evaluate our first question (the value of temporal vs. static tracking), we train a single-frame model [11]. This setup processes only the final (late-phase) frame of the examination using an identical RETFound-Green encoder and linear head, without the GRU module. **Phase Partitioning Strategies:** To isolate the impact of our partitioning approach, we evaluate our framework against two alternative frame-selection rules: *(1) Fixed equal-time thirds*, which splits each examination into three segments of equal temporal duration based purely on total elapsed time, sampling four frames uniformly within each third; and *(2) Uniform-by-index*, a structural baseline that ignores both timestamps and embedding content, selecting evenly spaced frames across the raw frame array length.

3.5 Training and Evaluation Protocol

Training and implementation. We use PyTorch [10] with `timm` image backbones [17]. We optimise with AdamW [8] using two parameter groups—a learning rate of 9.68×10^{-6} for the ViT encoder and 9.68×10^{-5} for the GRU and linear head—with a weight decay of 5×10^{-3} for both, a cross-entropy loss, a batch size of 16, and up to 40 epochs, retaining the checkpoint with the highest validation AUROC.

Evaluation metrics. In line with existing SOTA, we evaluate on the held-out test split using macro-averaged AUROC and F1, and report per-class AUROC and F1 (both one-vs-rest) for type-specific effects relevant to the second question. The overall macro F1 is computed over the five-way (argmax) prediction and therefore differs from the mean of the per-class one-vs-rest scores. The five types are markedly imbalanced (Table 1); we therefore report macro-averaged metrics, which weight every type equally regardless of prevalence, together with the per-type n_{test}. Results are averaged over five training seeds on a fixed train/validation/test partition; we report mean $\pm$ standard deviation. Checkpoints are selected by validation macro AUROC.

4 Results

4.1 Hyperfluorescence Classification Performance

As summarised in Table 1, incorporating full dye-transit temporal data consistently outperforms the single-frame baseline across overall and class-specific performance. Our framework raises the macro-averaged AUROC from 0.82 ± 0.02 to 0.87 ± 0.01, alongside an overall macro F1 score improvement ($0.48 \rightarrow 0.53$).

When evaluating specific hyperfluorescence types, the gains are clearly class-dependent. The largest AUROC improvements are observed for *Pooling* ($+0.07$) and *Window defect* ($+0.07$). This aligns with clinical expectations: both types are defined by a characteristic temporal evolution that a single late frame cannot

Table 1: Per-type and overall grading performance against the single-frame baseline [11]. n_{test} is the number of test examinations of each type. Values are mean ± std AUROC and F1 over five seeds; the better method per type and metric is in bold.

		AUROC		F1	
HyperF type	n_{test}	Last-frame	Ours	Last-frame	Ours
None	69	0.95 ± 0.01	$\mathbf{0.97 \pm 0.00}$	0.81 ± 0.02	$\mathbf{0.83 \pm 0.02}$
Leakage	107	0.87 ± 0.02	$\mathbf{0.89 \pm 0.00}$	$\mathbf{0.75 \pm 0.02}$	0.73 ± 0.03
Pooling	29	0.76 ± 0.02	$\mathbf{0.83 \pm 0.01}$	0.33 ± 0.02	0.33 ± 0.02
Staining	54	0.75 ± 0.03	$\mathbf{0.79 \pm 0.01}$	$\mathbf{0.44 \pm 0.04}$	0.35 ± 0.06
Window defect	19	0.78 ± 0.04	$\mathbf{0.85 \pm 0.01}$	0.29 ± 0.05	$\mathbf{0.39 \pm 0.04}$
Overall	278	0.82 ± 0.02	$\mathbf{0.87 \pm 0.01}$	0.48 ± 0.02	$\mathbf{0.53 \pm 0.01}$

capture — pooling by progressive accumulation of dye in a fluid space, and window defect by early choroidal transmission that fades in later phases. On F1 the picture is mixed: *None* and *Window defect* improve and *Pooling* is unchanged, whereas *Leakage* and *Staining* regress. The drop is most pronounced for *Staining*, whose F1 falls from 0.44 ± 0.04 to 0.35 ± 0.06 even as its AUROC rises from 0.75 ± 0.03 to 0.79 ± 0.01.

4.2 Ablation study

Data-driven Phase Partitioning. Under an identical architecture and a fixed budget of up to 12 frames (Table 2), the proposed data-driven partitioning and simple uniform-by-index sampling achieve comparable performance, with differences within one standard deviation both overall and across examination-length strata. This indicates that the main benefit comes from incorporating temporal information rather than the exact placement of phase boundaries. In contrast, fixed equal-time partitioning consistently results in lower performance across all strata. We retain the proposed strategy as it is parameter-free, gives a clinically meaningful Early/Mid/Late representation, and avoids the degradation seen with fixed boundaries.

Aggregation strategy. To disentangle whether the gain comes from using more frames or from temporal modelling, we replaced the GRU and attention pooling with mean pooling over the 12 frames per examination. Mean pooling matches the last-frame baseline (0.82 ± 0.02 AUROC), so aggregating frames without temporal modelling yields no improvement. The gain is attributable entirely to temporal modelling: a plain GRU reaches 0.84 ± 0.01 ($+0.02$), and adding attention pooling over the hidden states reaches 0.87 ± 0.01 ($+0.05$ over the baseline). The recurrence and the attention-based readout, not the additional frames, therefore drive the improvement.

Table 2: Evaluation of phase partitioning strategies with a fixed 12-frame budget. AUROC and F1 are mean ± std over five seeds; right columns report AUROC across examination-length bins (n test examinations each). Best values in bold.

	Overall		AUROC by length (#frames)		
Frame selection	AUROC	F1	≤12 (n=135)	13–24 (n=79)	>24 (n=64)
Uniform-by-index	0.86 ± 0.01	$\mathbf{0.53 \pm 0.02}$	$\mathbf{0.87 \pm 0.01}$	0.82 ± 0.01	0.85 ± 0.01
Fixed equal-time thirds	0.85 ± 0.01	0.53 ± 0.03	0.86 ± 0.01	0.81 ± 0.02	0.84 ± 0.01
Data-driven phases (Ours)	$\mathbf{0.87 \pm 0.01}$	$\mathbf{0.53 \pm 0.01}$	$\mathbf{0.87 \pm 0.01}$	$\mathbf{0.83 \pm 0.01}$	$\mathbf{0.86 \pm 0.01}$

5 Discussion

Our results answer the two questions posed in the Introduction. First, sequential modelling of the dye transit consistently improves grading relative to a single-frame formulation, confirming that temporal information provides complementary diagnostic signal. The ablation indicates that modelling temporal relationships is necessary to benefit from additional frames. Second, the largest gains are observed for Pooling and Window defect, indicating that types with a distinctive temporal signature — progressive accumulation for Pooling, early transmission that fades for Window defect — benefit most. In contrast, Staining shows improved AUROC but reduced F1. We hypothesise this reflects an operating-point effect rather than weaker discrimination: F1 is computed at the argmax boundary while AUROC is threshold-free, and since checkpoints are selected on validation macro AUROC, the resulting threshold need not be optimal for a minority type such as Staining (n=54), whose appearance also overlaps with Leakage and may draw borderline cases across the decision boundary.

The gains are largely robust to the frame-selection strategy, indicating that preserving temporal progression matters more than the precise placement of phase boundaries.

Our results are established on what is, to the best of our knowledge, the only public FA dataset with recoverable temporal information; now that timestamped FA data are available, validation across further acquisition protocols and devices is a natural next step. The temporal readout was kept deliberately lightweight to isolate the effect of temporal modelling itself, and richer sequence encoders or explicit modelling of inter-frame changes may extract more from long examinations. Relating the learned attention weights to the Early, Mid, and Late phases is a promising route to identifying which part of the dye transit each hyperfluorescence type depends on.

6 Conclusion

We asked whether temporal modelling improves automated fluorescein angiography grading, and which hyperfluorescence types benefit most. Modelling the full dye transit consistently outperforms a single-frame approach, raising overall

AUROC from 0.82 to 0.87 and improving discrimination across all five types, with the largest gains for Pooling and Window defect — the types defined by their temporal evolution.

Beyond the modelling results, we reconstruct and release per-frame acquisition times for the public AngioReport dataset, enabling temporally ordered analysis of FA examinations. Our findings indicate that the temporal dimension of FA carries diagnostic information that static approaches discard, and provide a foundation for future work on temporal angiography modelling.

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

References

1. Amiot, V., Pulvirenti, R., Jiménez-del Toro, O., et al.: UveAI: clinic-ready scoring of retinal inflammation in uveitis on widefield fluorescein angiography using AI. Scientific Reports (2026). https://doi.org/10.1038/s41598-026-46069-w
2. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: International Conference on Learning Representations (ICLR) (2015), https://arxiv.org/abs/1409.0473
3. Bishop, C.M.: Pattern Recognition and Machine Learning. Information Science and Statistics, Springer, New York, NY (2006)
4. Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1724–1734 (2014). https://doi.org/10.3115/v1/D14-1179
5. Google DeepMind: Gemini 2.5 flash. https://deepmind.google/technologies/gemini/ (2025)
6. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
7. Jabs, D.A., Nussenblatt, R.B., Rosenbaum, J.T., Standardization of Uveitis Nomenclature (SUN) Working Group: Standardization of uveitis nomenclature for reporting clinical data. Results of the First International Workshop. American Journal of Ophthalmology **140**(3), 509–516 (2005). https://doi.org/10.1016/j.ajo.2005.03.057
8. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (ICLR) (2019), https://openreview.net/forum?id=Bkg6RiCqY7
9. Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C.P., Bhatt, A.B., Lundberg, M., Lee, R.T., Liang, D.H., Quertermous, T., Zou, J.Y.: Video-based AI for beat-to-beat assessment of cardiac function. Nature **580**, 252–256 (2020). https://doi.org/10.1038/s41586-020-2145-8
10. Paszke, A., Gross, S., Massa, F., et al.: PyTorch: An imperative style, high-performance deep learning library (2019), https://pytorch.org
11. Pulvirenti, R.: When specialization helps (and hurts): Cross-modality transfer in ophthalmic imaging with foundation models. In: Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI) (2026). https://doi.org/10.1109/ISBI61048.2026.11515636
12. Ross, P., Reinke, A., Full, P.M., Wiesenfarth, M., Engelhardt, S., Bodenstedt, S., Speidel, S., Kopp-Schneider, A., Maier-Hein, L.: Robust medical instrument segmentation challenge 2019. arXiv preprint (2021)
13. Shen, H.Q., Wang, J., Niu, T., Chen, J.L., Xu, X.: Dynamic versus static ultra-widefield fluorescein angiography in eyes with diabetic retinopathy: a pilot prospective cross-sectional study. International Journal of Ophthalmology **14**(3), 409 (2021)
14. Tugal-Tutkun, I., Herbort, C.P., Khairallah, M., the Angiography Scoring for Uveitis Working Group (ASUWOG): Scoring of dual fluorescein and ICG inflammatory angiographic signs for the grading of posterior segment inflammation (dual fluorescein and ICG angiographic scoring system for uveitis). International Ophthalmology **30**, 539–552 (2010). https://doi.org/10.1007/s10792-008-9263-x

15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 30 (2017)
16. Wang, J.E., MacGillivray, T., Bernabeu, M.O.: Training a high-performance retinal foundation model with half-the-data and 400 times less compute. arXiv preprint (2024), https://arxiv.org/abs/2405.00117
17. Wightman, R.: PyTorch Image Models (timm). https://github.com/huggingface/pytorch-image-models (2019). https://doi.org/10.5281/zenodo.4414861
18. Xu, P., Chotcomwongse, P., Zhang, W., Chen, X., Wu, X., Chung, F.H.T., Zhang, X., He, M., Shi, D., Ruamviboonsuk, P.: Angioreport: dataset and baseline methods for fundus angiography report generation. British Journal of Ophthalmology **109**(11), 1283–1288 (2025). https://doi.org/10.1136/bjo-2024-327006, https://bjo.bmj.com/content/109/11/1283
19. Young, L.H., Kim, J., Yakin, M., Lin, H., Dao, D.T., Kodati, S., Sharma, S., Lee, A.Y., Lee, C.S., Sen, H.N.: Automated detection of vascular leakage in fluorescein angiography—a proof of concept. Translational Vision Science & Technology **11**(7), 19 (2022). https://doi.org/10.1167/tvst.11.7.19
20. Zhao, P.Y., Bommakanti, N., Yu, G., Aaberg, M.T., Patel, T.P., Paulus, Y.M.: Deep learning for automated detection of neovascular leakage on ultra-widefield fluorescein angiography in diabetic retinopathy. Scientific Reports **13**(1), 9165 (2023). https://doi.org/10.1038/s41598-023-36327-6
21. Zhou, Y., Chia, M.A., Wagner, S.K., Ayhan, M.S., Williamson, D.J., Struyven, R.R., Liu, T., Xu, M., Lozano, M.G., Woodward-Court, P., Kihara, Y., Altmann, A., Lee, A.Y., Topol, E.J., Denniston, A.K., Alexander, D.C., Keane, P.A.: A foundation model for generalizable disease detection from retinal images. Nature **622**(7981), 156–163 (2023). https://doi.org/10.1038/s41586-023-06555-x