



# A comparative study of deep learning-based retinal image registration methods

Thenuka Dharmaseelan<sup>1†</sup> , Neelabh Sinha<sup>1†</sup> , Samyyia Ashraf<sup>1</sup> , Kimia Daneshvar<sup>2</sup> , Amit John<sup>1</sup> , Periklis Giannakis<sup>1,3</sup> , Yik Ting Chan<sup>1,4</sup> , Yiu Wai Chan<sup>1</sup> , Nikolas Pontikos<sup>1\*</sup> 

<sup>1</sup>UCL Institute of Ophthalmology, EC1V 9EL London, United Kingdom

<sup>2</sup>Translational Ophthalmology Research Center, Farabi Eye Hospital, Tehran University of Medical Sciences, Tehran 13366 16351, Iran

<sup>3</sup>Faculty of Medicine and Dentistry, Queen Mary University of London, E1 2AD London, United Kingdom

<sup>4</sup>Darent Valley Hospital, Dartford and Gravesham NHS Trust, DA2 8DA Dartford, United Kingdom

<sup>†</sup>These authors share the first authorship.

\***Correspondence:** Nikolas Pontikos, UCL Institute of Ophthalmology, 11-43 Bath Street, EC1V 9EL London, United Kingdom. [n.pontikos@ucl.ac.uk](mailto:n.pontikos@ucl.ac.uk)

**Academic Editor:** Anastasios Koulaouzidis, University of Southern Denmark (SDU), Denmark

**Received:** November 30, 2025 **Accepted:** March 18, 2026 **Published:** May 24, 2026

**Cite this article:** Dharmaseelan T, Sinha N, Ashraf S, Daneshvar K, John A, Giannakis P, et al. A comparative study of deep learning-based retinal image registration methods. *Explor Digit Health Technol.* 2026;4:101194. <https://doi.org/10.37349/edht.2026.101194>

## Abstract

**Aim:** To benchmark three deep learning-based retinal image registration methods RetinaRegNet, EyeLiner, and GeoFormer on the Fundus Image Registration (FIRE) dataset to compare registration accuracy and computational efficiency using mean landmark error (MLE) as the primary outcome measure.

**Methods:** The three image registration approaches were evaluated using the FIRE dataset under consistent conditions across varying image overlap conditions (Classes S, A, and P). These included: (a) RetinaRegNet, which incorporates diffusion features, dual keypoint sampling through Scale-Invariant Feature Transform (SIFT) and random, two-stage outlier removal, and a multilevel registration hierarchy progressing from homography to polynomial transforms; (b) EyeLiner, which integrates anatomical segmentation with SuperPoint feature extraction, LightGlue matching, and thin-plate spline warping; (c) GeoFormer, which builds on Local Feature Transformers (LoFTR) through cross-attention mechanisms and Random Sampling Consensus (RANSAC)-based refinement. Registration performance was quantified using MLE.

**Results:** Across all 134 FIRE image pairs, RetinaRegNet achieved the lowest overall MLE (3.12 pixels), outperforming EyeLiner (3.81 pixels) and GeoFormer (6.06 pixels). Class-specific analysis showed that RetinaRegNet delivered the highest accuracy in Class S images (1.70 pixels), competitive performance in Class A (5.24 pixels), and the strongest results in the most challenging Class P cases (4.57 pixels). GeoFormer demonstrated the shortest processing time at 0.32 seconds per image pair, compared with 4.92 seconds for EyeLiner and 31.23 seconds for RetinaRegNet. In Class P, RetinaRegNet achieved a 59.2% improvement in accuracy relative to GeoFormer (4.57 vs 11.20 pixels). The code is available at: [https://github.com/ThenukaDharmaseelan/image\\_registration](https://github.com/ThenukaDharmaseelan/image_registration).

© The Author(s) 2026. This is an Open Access article licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



**Conclusions:** Overall, the evaluation reveals a clear trade-off between registration precision and computational speed. RetinaRegNet achieves the lowest MLE for complex clinical cases despite higher computational cost. EyeLiner balances precision and speed for routine use, while GeoFormer prioritizes rapid throughput where processing speed is critical.

## Keywords

retinal image registration, fundus photography, deep learning, EyeLiner, GeoFormer, RetinaRegNet

---

## Introduction

Retinal image registration is the process of aligning images across timepoints or across imaging modalities to make them comparable. This area of computer vision represents a cornerstone of modern ophthalmic research and clinical care, facilitating objective assessment of disease progression and evaluation of treatment efficacy [1, 2]. Retinal imaging modalities include colour fundus photography, fundus autofluorescence, infrared reflectance, optical coherence tomography (OCT), fluorescein angiography, and scanning laser ophthalmoscopy, which capture complementary anatomical and functional information, essential for comprehensive disease assessment. These medical images are widely used in longitudinal studies to monitor disease progression and are therefore ideally suited for image alignment and comparative analysis. However, effective utilisation of retinal imaging data depends on robust registration frameworks capable of accurately aligning images despite substantial variations in resolution, contrast, illumination, and spectral characteristics across imaging modalities.

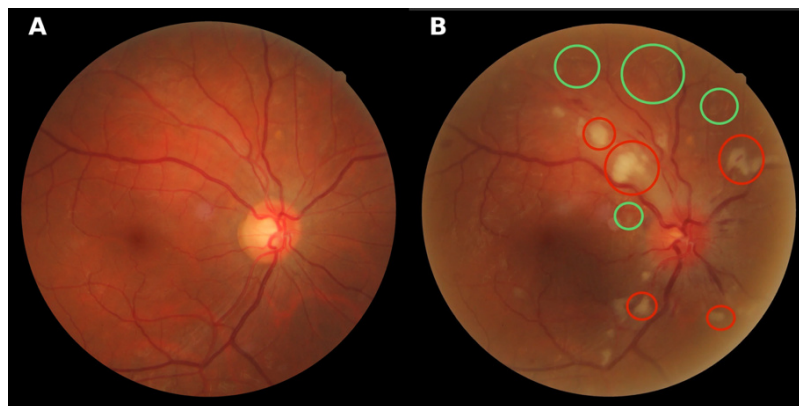
The advent of deep learning precipitated a paradigm shift, initially utilising convolutional neural networks (CNNs) to predict alignment directly or via coarse-to-fine strategies [3–5]. As the field matured, research pivoted toward addressing specific limitations of early CNNs. Style-transfer frameworks were introduced to unify multimodal representations, enabling robust vessel segmentation without pixel-wise annotations [6, 7]. To eliminate the dependency on ground-truth transformations, unsupervised deformable networks employing Spatial Transformer architectures were developed [8], while multi-scale frameworks incorporating edge similarity losses addressed optical distortions in fundus-OCT alignment [9]. Despite these sophisticated advances, the demand for greater robustness and interpretability has led to the divergence of three specialized architectural families (Figure S1). First, anatomically guided models began leveraging structural priors; early work explored structure-driven regression [10], while keypoint-based methods focused on explicit landmark matching [11, 12]. Second, transformer-based architectures and detector-free frameworks were adopted to capture long-range spatial dependencies via attention mechanisms [13, 14]. Finally, generative diffusion models, built on foundational probabilistic synthesis research, emerged to synthesize robust, illumination-invariant feature representations [15, 16].

However, a clear consensus on the optimal balance between computational efficiency and registration accuracy across these families remains elusive. To address this gap, we conduct a focused comparative analysis of three state-of-the-art frameworks on the FIRE dataset: RetinaRegNet, a zero-shot diffusion-based model leveraging hierarchical outlier suppression [17]; EyeLiner, an anatomically guided pipeline combining vessel landmark extraction with transformer matching [18]; and GeoFormer, a geometry-aware transformer incorporating Local Feature Transformer (LoFTR)-style correspondence detection with Random Sample Consensus (RANSAC) filtering for spatial consistency [19]. By evaluating these methods under varying degrees of overlap and anatomical distortion using an established classification system, this study provides a systematic comparative assessment of their accuracy, efficiency, and clinical suitability for large-scale retinal image analysis that has not been previously performed. A comprehensive comparison of the architectural approaches used by these three methods across key registration stages is provided in Table S1.

## Materials and methods

### Dataset

The FIRE dataset is widely recognized as a benchmark for evaluating retinal image registration algorithms. It comprises 134 colour fundus image pairs derived from 129 individual images, each captured at a resolution of  $2,912 \times 2,912$  pixels with a  $45^\circ$  field of view using a NIDEK AFC-210 fundus camera [20]. To enable systematic performance assessment, the dataset is divided into three difficulty categories based on field-of-view overlap and anatomical variation: Class S (71 pairs) represents easy cases with substantial image overlap ( $> 75\%$ ) without anatomical changes between images; Class A (14 pairs) consists of cases with  $> 75\%$  overlap that additionally exhibit anatomical changes as shown in Figure 1, and Class P (49 pairs) comprises the most challenging cases with limited overlap ( $< 75\%$ ) without anatomical differences between images. Each image pair includes 10 manually annotated ground-truth landmarks, typically located at vessel bifurcations and crossings, which provide a reference standard for quantitative evaluation of registration accuracy [20]. A visual overview of representative image pairs from each FIRE class is shown in Figure 2.



**Figure 1. Class A anatomical changes on the FIRE dataset. A:** Reference retinal image showing normal fundus appearance; **B:** Test image with pathological features highlighted. Red circles indicate hard exudates; green circles denote cotton wool spots. Prominent optic disc swelling is observed in the test image.

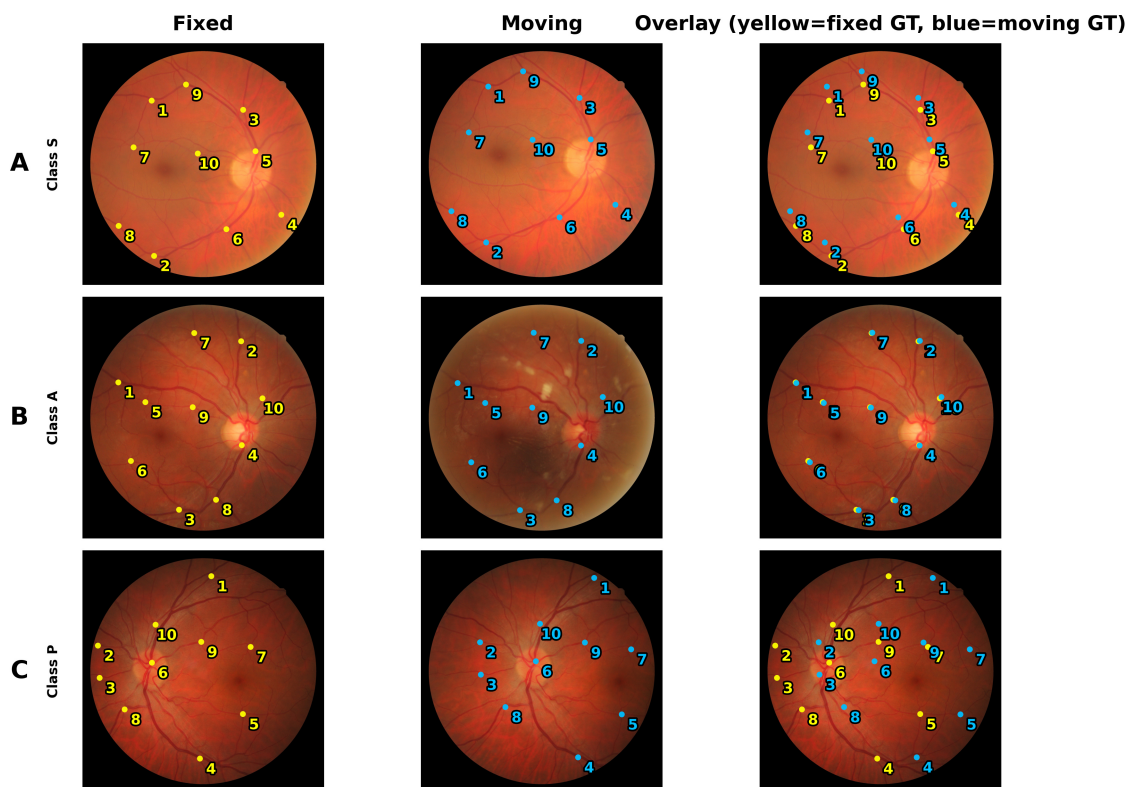
### RetinaRegNet

RetinaRegNet is a zero-shot framework designed to register pairs of retinal images by identifying correspondences within semantic diffusion features and subsequently warping one image onto the other. The process follows a broadly three-stage pipeline encompassing feature extraction, correspondence refinement, and hierarchical transformation (Figure 3).

First, rich feature maps are extracted from both the fixed and moving images using a pretrained Stable Diffusion model. Each image is passed through the model at a low noise step, producing intermediate feature tensors that capture both vessel patterns and broader retinal anatomy while remaining robust to variations in illumination and contrast.

With these features in hand, the method selects a balanced set of control points, approximately half from SIFT (to capture textured vessel regions) and half sampled uniformly at random (to cover smooth areas), ensuring that correspondence estimation is not biased toward densely textured regions. It then computes correspondences by cosine similarity in feature space so that a point  $p$  in the fixed image is paired with the most similar location  $q$  in the moving image. This approach provides invariance to photometric variations and tolerates moderate geometric drift.

Matched pairs are then refined through a two-step filtering process that eliminates outliers. First, a forward–backward (inverse consistency) check ensures that correspondences agree in both directions; mismatches that fail this test are discarded. Second, a geometric filtering step removes any pairs that deviate significantly from a globally consistent transformation. This combination of semantic matching and



**Figure 2. Pre-registration visual summary of the FIRE dataset.** Columns show the fixed (reference) image, the moving (test) image, and their unregistered overlay. Ground-truth landmarks are shown as yellow (fixed) and blue (moving) points.

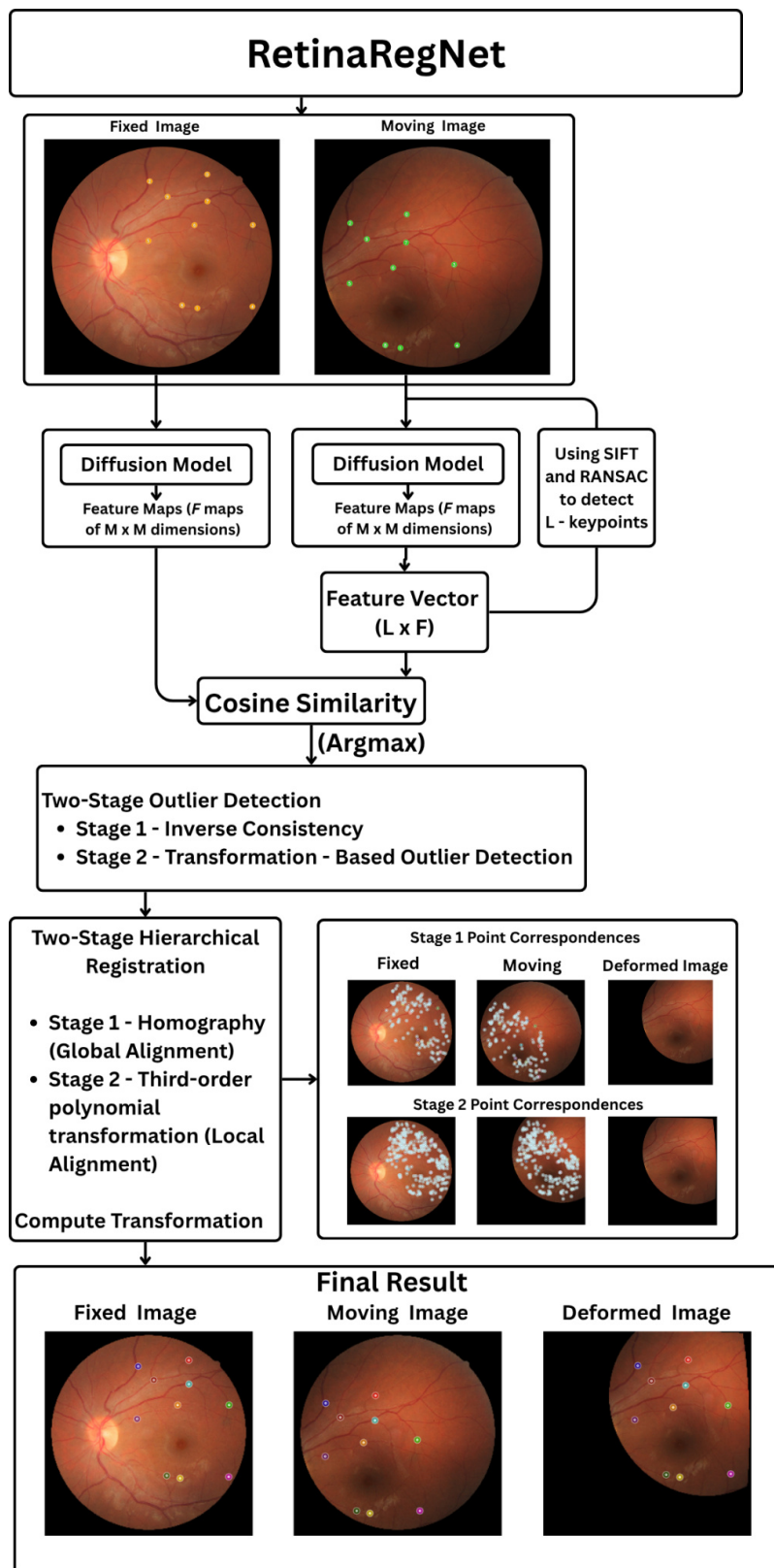
hierarchical outlier rejection produces a dense but reliable correspondence field that can support both global and local alignment.

Finally, RetinaRegNet performs a coarse-to-fine warp in two stages. Stage 1 estimates a global homography to absorb overall eye or camera motion, followed by Stage 2, a smooth local polynomial warp to correct small residual deformations. In practice, this combination of semantic feature matching, rigorous yet efficient filtering, and sequential global-to-local warping achieves high registration accuracy, particularly in low-overlap scenarios. The diffusion features enable recognition of vascular and anatomical patterns even where traditional pixel-based methods fail, while the two-stage transformation mitigates overfitting and preserves anatomical coherence.

### EyeLiner

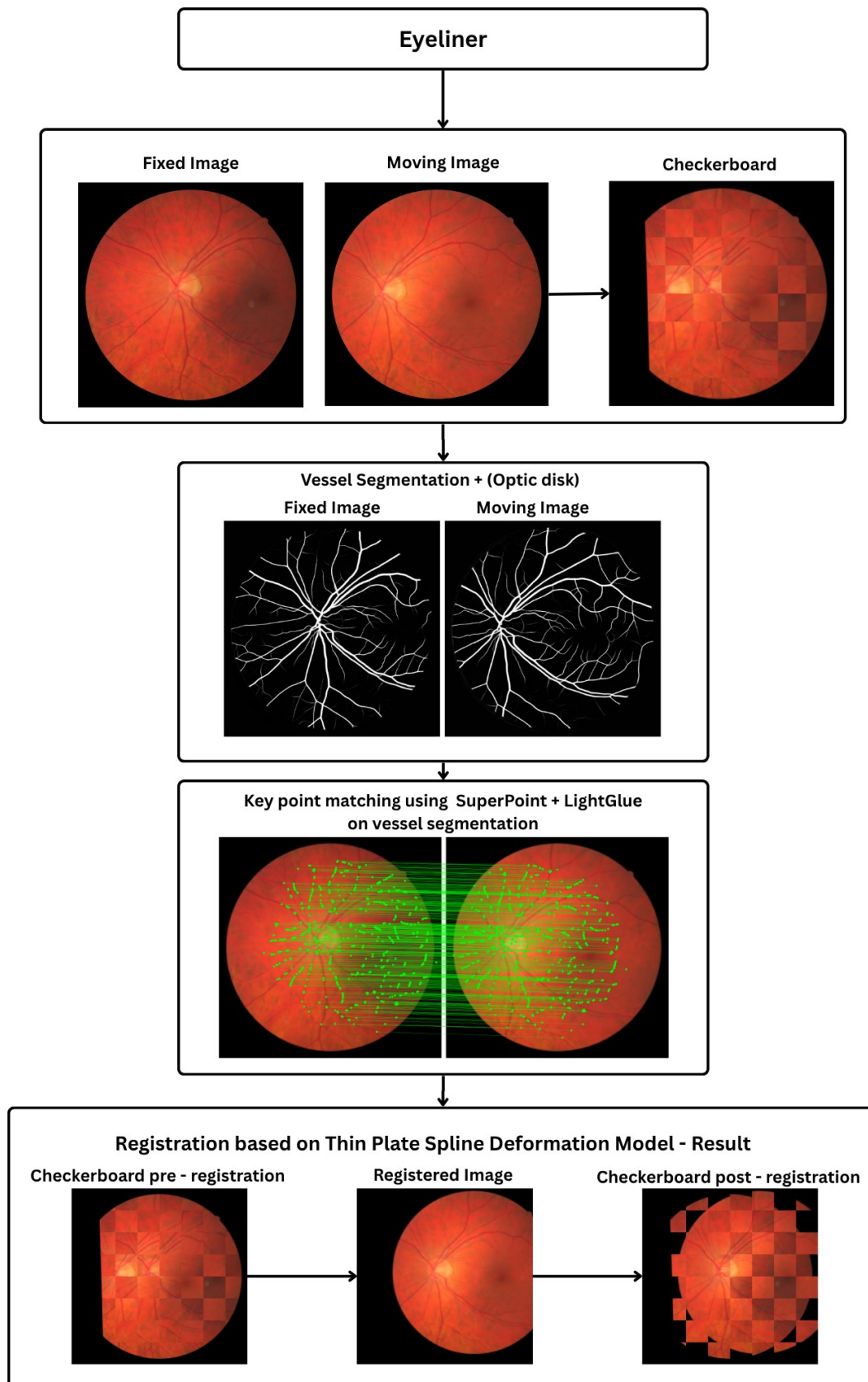
EyeLiner replicates clinical assessment patterns through modular anatomical structure analysis. The pipeline can be understood as a four-stage process that algorithmically mimics how clinicians track pathological changes relative to stable anatomical landmarks. An overview of the EyeLiner pipeline is shown in Figure 4. The image pairs are first segmented to extract the retinal vasculature and optic disc. A standard U-Net architecture was used for training on established vessel segmentation datasets to generate vascular masks for both the fixed and moving images. For reproducibility, we relied on AutoMorph [21] instead, a publicly available deep learning pipeline for vessel segmentation. Optic disc segmentation is performed using MaskFormer [22], a transformer-based architecture for general-purpose image segmentation. The resulting disc mask can be used to filter out keypoints detected (in the following step) within the optic disc region, as vessels inside the disc are subject to biological motion and thus are less reliable for geometric alignment.

Once the vascular regions are defined, EyeLiner detects and matches distinctive vessel features in a unified process. The SuperPoint [23] network is used to identify salient keypoints along the segmented vessels and to compute descriptor vectors that capture the local appearance of each point. These descriptors are then passed directly to LightGlue [24], a lightweight transformer that performs feature matching. Rather than relying on direct numerical comparison of descriptor values, LightGlue interprets the



**Figure 3. Overview of the RetinaRegNet registration pipeline.** Moving and fixed fundus images are processed using diffusion-based feature extraction. Keypoints are detected with SIFT and RANSAC, correspondences established via cosine similarity, and outliers removed through two-stage filtering. Two-stage hierarchical registration applies homography for global alignment and third-order polynomial transformation for local alignment, producing the final registered output.

geometric layout and contextual relationships among vessels in both images, producing more reliable correspondences even when illumination, scale, or focus differ between acquisitions. This combination of SuperPoint and LightGlue therefore establishes anatomically grounded, context-aware correspondences that are robust to the variations that typically challenge conventional intensity-based registration methods.



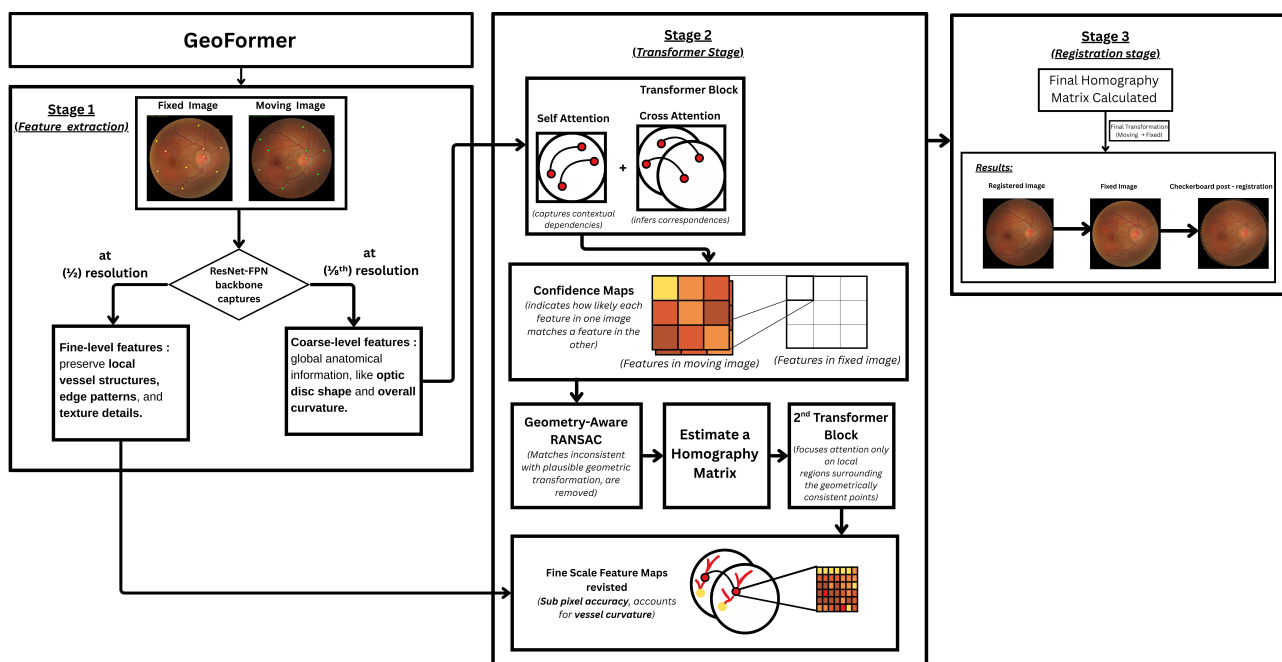
**Figure 4. Overview of the EyeLiner registration pipeline.** Fixed and moving fundus images are first compared using a checkerboard overlay. Vessel and optic disc regions are segmented to guide correspondence estimation. Keypoints are detected with SuperPoint and matched using LightGlue along vascular structures. A Thin-Plate Spline model then warps the moving image, producing the registered output and improved checkerboard alignment.

The final stage of the EyeLiner pipeline performs the actual alignment by warping the moving image to match the fixed reference. Once correspondences are confirmed, a thin-plate spline transformation is used

to model the deformation between the two images. This smooth, flexible transformation captures both global displacement and local non-rigid motion in the retinal vasculature while maintaining overall anatomical plausibility. In practice, the combination of anatomically guided segmentation, context-aware matching, and smooth deformation modelling enables EyeLiner to produce clinically interpretable and geometrically stable registrations, offering a strong balance between accuracy, computational efficiency, and biological realism.

## Geoformer

GeoFormer builds upon the existing LoFTR framework and employs a geometry-aware transformer to learn dense correspondences between retinal images without relying on explicit keypoint detectors. The method can be broadly divided into three stages. A schematic overview of the GeoFormer framework is presented in Figure 5. GeoFormer begins with a ResNet-FPN backbone [25] that extracts multiscale feature maps from the fixed and moving retinal images. The model captures coarse-level features at one-eighth resolution, encoding global anatomical information such as optic disc shape and overall curvature, and fine-level features at one-half resolution, which preserve local vessel structures, edge patterns, and texture details. This multi-level representation ensures that both global context and fine anatomical information are passed to the transformer network.



**Figure 5. GeoFormer architecture and methodology.** Stage 1: ResNet-FPN extracts multi-scale features at 1/2 and 1/8 resolution. Stage 2: Transformer blocks perform self-attention and cross-attention to generate confidence maps, followed by geometry-aware RANSAC and refinement for sub-pixel accuracy. Stage 3: The final homography matrix is computed and applied to produce the registered output.

The extracted features are processed through transformer blocks that perform self-attention within each image to capture contextual dependencies and cross-attention between the image pair to infer correspondences. This produces a confidence map that indicates how likely each feature in one image matches a feature in the other. The resulting matches are filtered using a geometry-aware RANSAC procedure, which removes correspondences inconsistent with a plausible geometric transformation and estimates an initial homography matrix. These verified matches are then refined through a second transformer stage that focuses attention only on local regions surrounding the geometrically consistent points, improving computational efficiency and accuracy in vessel-dense areas. Finally, the model revisits the fine-scale feature maps to achieve sub-pixel correspondence accuracy, accommodating vessel curvature and illumination variations that coarse features cannot fully capture.

The refined correspondences are used to compute a final homography matrix that maps coordinates from the moving image onto the fixed image. The moving image is then warped according to this transformation, completing the registration process.

### Evaluation metrics

MLE in pixels for each evaluated method. MLE measures registration accuracy by calculating the average distance between the corresponding anatomical landmarks and their estimated locations obtained from the registration method, with lower values indicating better alignment:

$$MLE = \frac{1}{N} \sum_{i=1}^N \sqrt{(x_i'' - x_i')^2 + (y_i'' - y_i')^2}$$

where  $(x_i', y_i')$  represent manually annotated ground truth landmarks and  $(x_i'', y_i'')$  are transformed coordinates.

In addition to the mean landmark error, we evaluate registration performance using the success rate. A registration is considered successful if a transformed landmark lies within a predefined pixel threshold of its corresponding ground-truth location. For each image pair, this criterion is applied independently to the ten annotated landmarks, yielding a binary outcome (success or failure) per point. The success rate is then computed as the proportion of successful landmarks out of the ten ground-truth points for that image pair. The threshold  $\tau$  is set to 12.5 pixels, following [17], to ensure comparability with prior results.

$$SR = \frac{1}{N} \sum_{i=1}^N I(d_i \leq \tau),$$

where  $I(\cdot)$  is the indicator function,  $d_i = \sqrt{(x_i' - x_i'')^2 + (y_i' - y_i'')^2}$  and  $\tau = 12.5$  pixels.

Finally, we quantify overall registration robustness using the Area Under the Curve (AUC) of the cumulative success-rate curve. For a range of pixel thresholds, the success rate is computed as the proportion of landmarks whose registration error falls below each threshold. The resulting curve characterises how registration performance degrades as the tolerance increases. The AUC provides a single scalar summary of this behaviour, with higher values indicating consistently higher success rates across a wide range of error thresholds:

$$AUC = \frac{1}{T_{max}} \int_0^{T_{max}} SR(t) dt,$$

where  $T_{max}$  is the maximum error threshold, set to 25 pixels for the FIRE dataset.

We also evaluated registration results using Normalized Cross Correlation (NCC), an intensity-based metric that quantifies linear correlation between image intensities. In retinal imaging, NCC reflects global photometric consistency after warping but is influenced by acquisition-related factors and appearance changes due to disease progression. As NCC is computed over the full image domain, the metric does not explicitly encode pointwise anatomical correspondence. NCC is therefore reported as a complementary measure of global intensity agreement rather than a primary indicator of anatomical registration accuracy.

$$NCC = \frac{\sum (I_f(x, y) - \mu_f)(I_m(x, y) - \mu_m)}{\sqrt{\sum (I_f(x, y) - \mu_f)^2 \sum (I_m(x, y) - \mu_m)^2}},$$

Where  $\sum$  is 2-dimensional summation across  $(x, y)$ ,  $I_f(x, y)$  is the intensity of the fixed image at location  $(x, y)$  and  $I_m(x, y)$  is the intensity of the moving image at the location  $I_m(x, y)$ , and  $\mu_f$  and  $\mu_m$  are the mean intensities of fixed and moving images, respectively.

### Experimental setup

All three registration pipelines were reproduced and evaluated in the same environment to ensure consistency and comparability across methods. All experiments were executed on a workstation equipped with an NVIDIA GeForce RTX 3090 GPU (24 GB VRAM) running CUDA 12.6. The implementations used Python 3.11.13, PyTorch 2.2.2, and torchvision 0.17.2 as the primary deep learning frameworks. For methods requiring vessel segmentation (EyeLiner), AutoMorph was used for reproducibility. All runtime

and accuracy measurements reported in Results section reflect executions under this standardised environment. The code for data processing is available at: [https://github.com/ThenukaDharmaseelan/image\\_Registration](https://github.com/ThenukaDharmaseelan/image_Registration).

Exact parametric configurations used for each registration pipeline are provided in the Tables S2, 3, 4.

## Results

### Overall performance analysis

As shown in Table 1 and Figure 6, RetinaRegNet achieved the lowest overall MLE of 3.12 ( $\pm 2.43$  pixels) across the FIRE dataset. A Friedman test confirmed significant differences between the three methods ( $\chi^2(2) = 134.37$ ,  $P = 2 \times 10^{-16}$ ). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction demonstrated that RetinaRegNet achieved significantly lower MLE than both EyeLiner ( $3.81 \pm 3.13$  pixels,  $P = 1 \times 10^{-4}$ ) and GeoFormer ( $6.06 \pm 4.86$  pixels,  $P = 2 \times 10^{-21}$ ), while EyeLiner also significantly outperformed GeoFormer (Table 2).

**Table 1. Performance Comparison on FIRE Dataset using Mean Landmark Error (MLE), Success Rate (SR), Area Under Curve (AUC), and Normalized Cross Correlation (NCC).**

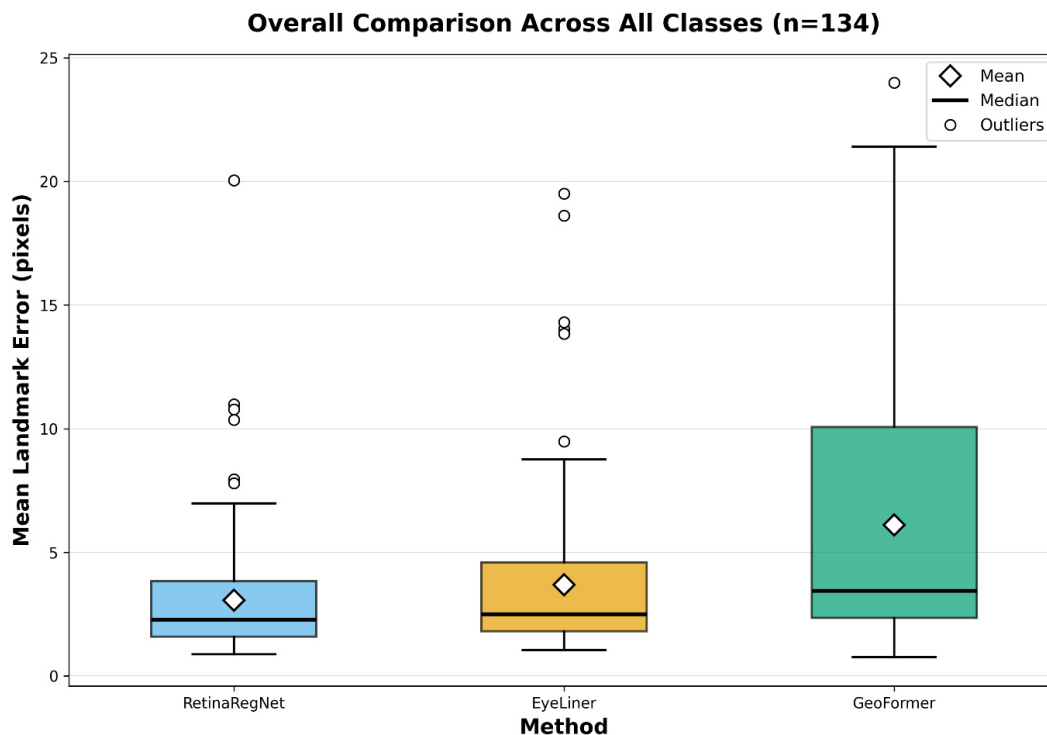
| Method       | Overall<br>( <i>n</i> = 134) | Class S<br>( <i>n</i> = 71) | Class A<br>( <i>n</i> = 14) | Class P<br>( <i>n</i> = 49) |
|--------------|------------------------------|-----------------------------|-----------------------------|-----------------------------|
|              | <b>MLE</b>                   | <b>MLE</b>                  | <b>MLE</b>                  | <b>MLE</b>                  |
| RetinaRegNet | 3.12 $\pm$ 2.43              | 1.70 $\pm$ 0.54             | 5.24 $\pm$ 2.64             | 4.57 $\pm$ 2.80             |
| EyeLiner     | 3.81 $\pm$ 3.13              | 1.80 $\pm$ 0.40             | 4.87 $\pm$ 3.05             | 6.01 $\pm$ 3.75             |
| GeoFormer    | 6.06 $\pm$ 4.86              | 2.42 $\pm$ 0.77             | 6.55 $\pm$ 4.71             | 11.20 $\pm$ 3.25            |
|              | <b>SR</b>                    | <b>SR</b>                   | <b>SR</b>                   | <b>SR</b>                   |
| RetinaRegNet | 97.76%                       | 100%                        | 92.86%                      | 95.92%                      |
| EyeLiner     | 97.01%                       | 100%                        | 92.86%                      | 93.88%                      |
| GeoFormer    | 88.06%                       | 100%                        | 92.86%                      | 69.39%                      |
|              | <b>AUC</b>                   | <b>AUC</b>                  | <b>AUC</b>                  | <b>AUC</b>                  |
| RetinaRegNet | 0.89                         | 0.95                        | 0.79                        | 0.85                        |
| EyeLiner     | 0.85                         | 0.93                        | 0.80                        | 0.76                        |
| GeoFormer    | 0.76                         | 0.91                        | 0.74                        | 0.54                        |
|              | <b>NCC</b>                   | <b>NCC</b>                  | <b>NCC</b>                  | <b>NCC</b>                  |
| RetinaRegNet | 0.56                         | 0.75                        | 0.60                        | 0.29                        |
| EyeLiner     | 0.56                         | 0.74                        | 0.63                        | 0.28                        |
| GeoFormer    | 0.64                         | 0.72                        | 0.62                        | 0.53                        |

**Table 2. Overall statistical comparison of MLE across all methods (*n* = 134).**

| Comparison                | Test        | Statistic | Raw<br><i>P</i> -value | Bonferroni <i>P</i> -value | Mean difference<br>(A-B) |
|---------------------------|-------------|-----------|------------------------|----------------------------|--------------------------|
| Friedman                  | $\chi^2(2)$ | 134.37    | $2 \times 10^{-16}$    | –                          | –                        |
| RetinaRegNet vs EyeLiner  | Wilcoxon V  | 2653      | $3 \times 10^{-5}$     | $1 \times 10^{-4}$ ***     | -0.626                   |
| RetinaRegNet vs GeoFormer | Wilcoxon V  | 198       | $8 \times 10^{-22}$    | $2 \times 10^{-21}$ ***    | -3.050                   |
| EyeLiner vs GeoFormer     | Wilcoxon V  | 544       | $1 \times 10^{-18}$    | $3 \times 10^{-18}$ ***    | -2.424                   |

Asterisks denote Bonferroni-adjusted significance: \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ .

Box plot analysis in Figure 6 reveals distinct distributional characteristics. RetinaRegNet demonstrates the most compact error distribution with a median of approximately 2.0 pixels and relatively few outliers. EyeLiner shows a slightly wider distribution with a median around 2.5 pixels and a comparable outlier pattern. In contrast, GeoFormer exhibits substantially higher variability, with a median of approximately 4 pixels and a significantly larger interquartile range.



**Figure 6. Overall comparison of registration methods across all classes on the FIRE dataset ( $n = 134$ ).** Box plots show the distribution of MLE in pixels for the three methods.

Success rate analysis (Table 1) showed RetinaRegNet achieved 97.76%, EyeLiner 97.01%, and GeoFormer 88.06% overall, demonstrating RetinaRegNet’s superior reliability in completing registration tasks successfully.

AUC analysis (Table 1) confirmed RetinaRegNet’s superior performance with an overall AUC of 0.89, compared to EyeLiner (0.85) and GeoFormer (0.76), representing a 5% improvement over EyeLiner and 17% over GeoFormer.

NCC analysis (Table 1) revealed contrasting patterns. GeoFormer achieved the highest overall NCC (0.64), followed by RetinaRegNet and EyeLiner (both 0.56). Notably, NCC rankings contradicted landmark-based metrics. Despite GeoFormer achieving the highest intensity correlation, it exhibited the poorest landmark accuracy (MLE: 6.06 pixels), lowest AUC (0.76), and poorest success rate (88.06%). These findings indicate that intensity-based similarity does not reliably reflect geometric alignment accuracy.

Overall, RetinaRegNet demonstrates lower mean error, higher success rate, and superior AUC compared with EyeLiner and GeoFormer across all landmark-based metrics, while GeoFormer exhibits the widest error distribution despite achieving higher intensity-based correlation.

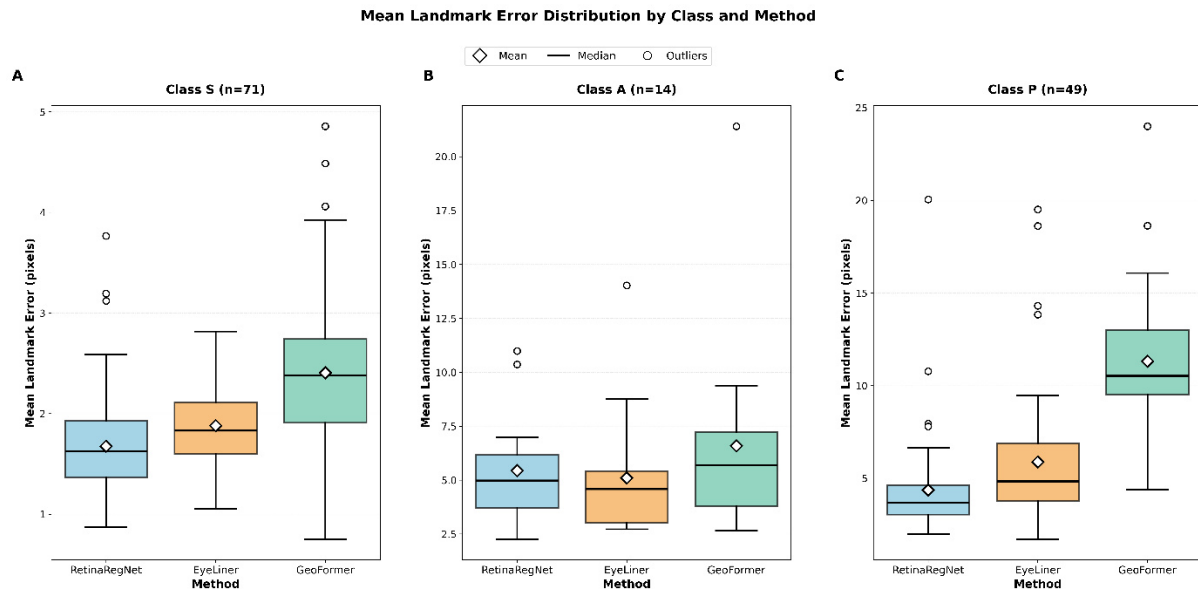
Across difficulty classes, RetinaRegNet achieves superior overall performance (3.12 pixels), with the lowest MLE in Classes S and P and mean performance in Class A that is close to EyeLiner. EyeLiner attains the lowest mean MLE in Class A, although its advantage over RetinaRegNet is not statistically significant in this subset. Both methods consistently outperform GeoFormer, which shows particularly large errors in Class P. These findings underscore the importance of class specific evaluation in assessing the robustness and generalization capabilities of retinal registration methods.

### Class specific performance analysis

#### Class S performance

Class S, representing the largest subset with 71 samples, proves to be the least challenging for all methods (Figure 7A). RetinaRegNet obtained the best performance with an MLE of (1.70 pixels), outperforming EyeLiner (1.80 pixels) and GeoFormer (2.42 pixels) by 6% and 42%, respectively. The consistently low error rates across all methods indicate that Class S images possess favourable characteristics for vessel

landmark localisation. These differences were statistically significant, as shown in Table 3 (Friedman  $P = 2 \times 10^{-14}$ ; all pairwise comparisons reached significance after Bonferroni correction with  $P \leq 8 \times 10^{-4}$ ). AUC analysis (Table 1) demonstrated RetinaRegNet’s superior performance (0.95), followed by EyeLiner (0.93) and GeoFormer (0.91). All three methods achieved 100% success rates (Table 1), confirming the relative ease of Class S registration tasks. NCC values (Table 1) were comparable across methods (RetinaRegNet: 0.75, EyeLiner: 0.74, GeoFormer: 0.72), indicating similar intensity-based alignment in high-overlap scenarios. Overall, RetinaRegNet provides consistently lower landmark error and higher AUC than EyeLiner and GeoFormer in this class.



**Figure 7. Mean Landmark Error distribution by class and method on the FIRE dataset.** Boxplots show the performance comparison of the three methods across the three classes. Diamond markers indicate mean values, horizontal lines show medians, boxes represent interquartile ranges (IQR), whiskers extend to  $1.5 \times$  IQR, and red circles represent outliers. Lower values indicate better registration accuracy.

**Table 3. Statistical comparison of MLE for Class S images ( $n = 71$ ).**

| Comparison                | Test        | Statistic | Raw P-value         | Bonferroni P-value      | Mean Difference (A-B) |
|---------------------------|-------------|-----------|---------------------|-------------------------|-----------------------|
| Friedman                  | $\chi^2(2)$ | 63.577    | $2 \times 10^{-14}$ | –                       | –                     |
| RetinaRegNet vs EyeLiner  | Wilcoxon V  | 640       | $2 \times 10^{-4}$  | $8 \times 10^{-4}$ ***  | -0.204                |
| RetinaRegNet vs GeoFormer | Wilcoxon V  | 51        | $2 \times 10^{-12}$ | $6 \times 10^{-12}$ *** | -0.729                |
| EyeLiner vs GeoFormer     | Wilcoxon V  | 339       | $8 \times 10^{-8}$  | $2 \times 10^{-7}$ ***  | -0.524                |

Asterisks denote Bonferroni-adjusted significance: \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ .

### Class A performance

Class A, representing moderate cases with anatomical change (14 samples), shows intermediate difficulty (Figure 7B). EyeLiner attained the lowest mean MLE in this class (4.87 pixels), followed closely by RetinaRegNet (5.24 pixels); however, this difference is not statistically significant, as shown in Table 4 (adjusted  $P = 0.776$ ). GeoFormer showed higher error at 6.55 pixels and performed significantly worse than EyeLiner (adjusted  $P = 0.012$ ). AUC analysis (Table 1) showed EyeLiner achieved the highest performance (0.80), followed by RetinaRegNet (0.79) and GeoFormer (0.74). All three methods achieved identical 92.86% success rates (Table 1). NCC values (Table 1) showed minimal variation across methods (RetinaRegNet: 0.60, EyeLiner: 0.63, GeoFormer: 0.62), indicating comparable intensity-based similarity. These results indicate that performance differences between the methods are less pronounced in this subset, reflecting the moderate difficulty of Class A cases. However, these results should be interpreted with caution due to the limited sample size.

**Table 4. Statistical comparison of MLE for Class A images (n = 14).**

| Comparison                | Test        | Statistic | Raw P-value | Bonferroni P-value | Mean difference (A-B) |
|---------------------------|-------------|-----------|-------------|--------------------|-----------------------|
| Friedman                  | $\chi^2(2)$ | 8.143     | 0.017       | –                  | –                     |
| RetinaRegNet vs EyeLiner  | Wilcoxon V  | 71        | 0.259       | 0.776              | 0.340                 |
| RetinaRegNet vs GeoFormer | Wilcoxon V  | 18        | 0.033       | 0.098              | -1.147                |
| EyeLiner vs GeoFormer     | Wilcoxon V  | 6         | 0.004       | 0.012*             | -1.487                |

Asterisks denote Bonferroni-adjusted significance: \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ .

### Class P performance

Performance degradation occurs across all methods on Class P samples (49 samples), the most challenging category (Figure 7C). RetinaRegNet maintains relatively strong performance with an MLE of 4.57 pixels, while EyeLiner records 6.01 pixels. GeoFormer experiences substantial difficulty in this class, with its error increasing to 11.20 pixels (145% increase compared to RetinaRegNet). This pronounced performance gap reveals that GeoFormer is particularly sensitive to the challenging characteristics present in low-overlap images. All pairwise differences in Class P were statistically significant after Bonferroni correction, as shown in Table 5, with RetinaRegNet significantly outperforming both EyeLiner (adjusted  $P = 6 \times 10^{-4}$ ) and GeoFormer (adjusted  $P = 4 \times 10^{-9}$ ), and EyeLiner also significantly outperforming GeoFormer (adjusted  $P = 4 \times 10^{-8}$ ). AUC analysis (Table 1) demonstrated RetinaRegNet's substantial advantage (0.85), compared to EyeLiner (0.76) and GeoFormer (0.54), representing a 57% relative improvement over GeoFormer. Success rate analysis (Table 2) showed RetinaRegNet achieved 95.92%, EyeLiner 93.88%, while GeoFormer dropped to 69.39%, indicating frequent registration failures in challenging low-overlap scenarios. Notably, NCC values (Table 1) revealed contrasting patterns: GeoFormer maintained the highest NCC (0.53) while RetinaRegNet and EyeLiner decreased to 0.29 and 0.28, respectively. This discordance between high intensity correlation and poor landmark accuracy in GeoFormer highlights that NCC does not reliably reflect geometric alignment quality in challenging registration scenarios.

**Table 5. Statistical comparison of MLE for Class P images (n = 49).**

| Comparison                | Test        | Statistic | Raw P-value         | Bonferroni P-value     | Mean difference (A-B) |
|---------------------------|-------------|-----------|---------------------|------------------------|-----------------------|
| Friedman                  | $\chi^2(2)$ | 71.306    | $3 \times 10^{-16}$ | –                      | –                     |
| RetinaRegNet vs EyeLiner  | Wilcoxon V  | 238       | $2 \times 10^{-4}$  | $6 \times 10^{-4}$ *** | -1.513                |
| RetinaRegNet vs GeoFormer | Wilcoxon V  | 1         | $1 \times 10^{-9}$  | $4 \times 10^{-9}$ *** | -6.957                |
| EyeLiner vs GeoFormer     | Wilcoxon V  | 41        | $1 \times 10^{-8}$  | $4 \times 10^{-8}$ *** | -5.444                |

Asterisks denote Bonferroni-adjusted significance: \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ .

### Runtime analysis

Table 6 presents the runtime performance analysis of the evaluated methods. GeoFormer demonstrates the fastest processing speed at 0.32 seconds per image, offering real-time processing capability. EyeLiner achieves moderate computational efficiency at 4.92 seconds per image. This includes 4.22 seconds for AutoMorph vessel segmentation and 0.70 seconds for the registration process itself, making it approximately 15 times slower than GeoFormer. RetinaRegNet exhibits the highest computational cost at 31.23 seconds per image, representing 98 times increase compared to GeoFormer and 6.4 times increase compared to EyeLiner.

These computational differences highlight the trade-offs between different architectural approaches. EyeLiner's reliance on explicit vessel segmentation via AutoMorph adds substantial preprocessing overhead compared to end-to-end approaches. Due to its complex architecture, RetinaRegNet's multi-stage registration pipeline with iterative refinement steps results in considerably longer processing times. In contrast, GeoFormer's transformer-based architecture enables direct processing of raw retinal images without intermediate segmentation steps, resulting in significantly faster inference times suitable for

**Table 6. Runtime performance comparison.**

| Method       | Time per image (seconds) |
|--------------|--------------------------|
| RetinaRegNet | 31.23                    |
| EyeLiner     | 4.92                     |
| GeoFormer    | 0.32                     |

clinical deployment. The theoretical computational complexity and scalability characteristics of each pipeline across individual processing stages are analysed in [Table S5](#).

## Discussion

Our study highlights how the different design choices of the three methods influence their performance across the FIRE dataset. RetinaRegNet achieves the lowest overall MLE and consistently high success rate and AUC, indicating strong capability in aligning anatomical landmarks even under challenging imaging conditions but it takes the longest to complete registration. RetinaRegNet and EyeLiner performed similarly well in Class A, where the small sample size limits the ability to resolve subtle performance differences. In contrast, in Class P, which involves reduced overlap, RetinaRegNet demonstrates a clear advantage, indicating greater robustness under challenging alignment conditions. This was not the case for GeoFormer, whose reduced performance in difficult scenarios can largely be attributed to its architectural reliance on a global homography-only transformation, although insufficient fine-scale deformation modelling or dataset-specific effects may also contribute. EyeLiner produces competitive accuracy with relatively stable behaviour across the three difficulty classes at a moderate speed GeoFormer offers very fast processing but displays substantially higher error, particularly in cases with reduced overlap or anatomical distortion. Notably, this trend contrasts with intensity-based similarity measures such as NCC, where GeoFormer attains higher scores despite poorer landmark alignment, reinforcing that NCC reflects global photometric agreement rather than precise anatomical correspondence. These findings confirm that architectural choices, such as the use of local deformation models and anatomically guided correspondence strategies, play an important role in determining landmark-based registration accuracy and robustness.

The observed accuracy differences have direct implications for clinical deployment of retinal registration pipelines. RetinaRegNet's strong performance in challenging cases supports its use in longitudinal monitoring, where precise alignment is critical for detecting subtle structural change over time. EyeLiner offers a practical compromise for routine follow-up and vessel-centric analysis by balancing competitive accuracy in moderate-difficulty settings with lower computational cost and clear anatomical interpretability. In contrast, GeoFormer's rapid inference makes it suitable for high-throughput applications such as screening or acquisition quality control, provided that difficult cases are identified and redirected to a more robust registration pipeline. Method selection therefore, depends on the required balance between registration precision, robustness to failure, and computational efficiency.

Although RetinaRegNet exhibits the highest computational cost among the evaluated methods, this runtime is most relevant to offline longitudinal analysis rather than real-time clinical deployment. In many longitudinal retinal imaging workflows, such as disease progression studies or retrospective cohort analyses, robustness and alignment accuracy are prioritised over per-image latency, making runtimes on the order of tens of seconds acceptable. The current implementation does not support straightforward batching across image pairs due to its sequential correspondence filtering and transformation stages; improving throughput through architectural or implementation-level optimisation remains an avenue for future work.

Our results confirm the findings of previous work [17], showing that incorporating local deformation, such as polynomial warping in RetinaRegNet and thin-plate spline warping in EyeLiner, improves landmark-based alignment accuracy, as reflected in the superior performance in both methods. Inverse consistency in RetinaRegNet helps mitigate inconsistent keypoint matches in regions where the underlying diffusion model exhibits low confidence. However, it may become less effective when the model is highly

confident, which is, in practice, typically the case for well-pretrained models. Consequently, its contribution is less critical in this setting as EyeLiner is able to achieve comparable performance without imposing strong consistency constraints. Anatomy-guided filtering in EyeLiner encourages focus on vascular structure, but can be disadvantageous when vessel morphology changes drastically, leaving large regions without matched keypoints, as also noted in prior anatomy-guided approaches. RetinaRegNet alleviates this overemphasis on edges and high contrast regions through random keypoint sampling, promoting coverage across nearly the whole image. However, this strategy may mislead the warping algorithm in the presence of anatomical changes, resulting in slightly inferior performance in Class A.

This study is limited by its reliance on the FIRE dataset, which contains only colour fundus photographs captured with a single 45-degree field-of-view camera at a fixed resolution. Results may differ for other imaging modalities such as autofluorescence, infrared, fluorescein angiography, or widefield imaging. Class A is relatively small in size ( $n = 14$ ), which may limit statistical power to detect subtle performance differences between top-performing methods; future studies using larger, more balanced datasets are required to confirm these trends. Evaluation was based on ten manually annotated vascular landmarks per image pair, which capture alignment accuracy at a limited number of locations and may not fully reflect performance across the entire retina, particularly near lesions or in peripheral regions. Pixel-based errors were not converted into physical units, which limits direct clinical interpretation across devices with different magnifications or sensor characteristics. Finally, the lack of validation across multiple independent datasets limits the generalisability of the findings, and further studies on diverse datasets are required to address this limitation.

This study presents an in-depth comparative evaluation of three deep learning-based retinal registration pipelines using the FIRE dataset. RetinaRegNet achieves the highest accuracy across the full dataset and in the most challenging low-overlap cases. EyeLiner provides reliable performance with moderate computational cost through anatomically guided matching and flexible warping. GeoFormer offers the fastest processing speed, but at the expense of reduced robustness in difficult scenarios. These findings underscore the value of combining global alignment with locally adaptive deformation when precise registration is required and highlight the importance of selecting a registration method that aligns with the clinical or research needs of the intended application. Future work should evaluate these approaches across diverse imaging modalities and explore optimisation strategies to improve the computational efficiency of high-accuracy models.

## Abbreviations

AUC: Area Under the Curve

CNNs: convolutional neural networks

IQR: Interquartile Range

LoFTR: Local Feature Transformer

MLE: mean landmark error

NCC: Normalized Cross Correlation

OCT: optical coherence tomography

RANSAC: Random Sample Consensus

SIFT: Scale-Invariant Feature Transform

## Supplementary materials

The supplementary materials for this article are available at: [https://www.explorationpub.com/uploads/Article/file/101194\\_sup\\_1.pdf](https://www.explorationpub.com/uploads/Article/file/101194_sup_1.pdf).

## Declarations

### Author contributions

TD: Conceptualization, Methodology, Data curation, Formal analysis, Project administration, Writing—original draft, Writing—review & editing, Visualization, Writing—review & editing, Funding acquisition. NS: Conceptualization, Investigation, Methodology, Formal analysis, Data curation, Visualization, Writing—original draft, Writing—review & editing. YWC: Supervision, Conceptualization, Methodology, Validation, Writing—review & editing, Funding acquisition. SA: Supervision, Writing—review & editing, Writing—original draft. KD: Investigation, Writing—original draft, Writing—review & editing. PG: Writing—review & editing. YTC: Writing—review & editing. AJ: Validation. NP: Supervision, Project administration, Validation, Writing—review & editing, Funding acquisition. All authors read and approved the submitted version.

### Conflicts of interest

NP is a patent holder of PCT/EP2023/076614 filed by UCL Business. The authors declare that they have no other relevant conflicts of interest.

### Ethical approval

This study utilised the FIRE dataset, which is publicly available, so it involves no new data collection or direct interaction with human participants. All data are fully anonymised. Therefore, ethical approval was not required.

### Consent to participate

This study involved secondary analysis of publicly available, anonymised retinal image data and did not involve direct participation of human subjects. Accordingly, consent to participate was not required.

### Consent to publication

No identifiable individual data are presented in this study. Therefore, consent to publication was not required.

### Availability of data and materials

The FIRE dataset analysed during this study is publicly available at: <https://projects.ics.forth.gr/cvrl/fire/> and was used in accordance with the terms of use. No additional datasets were generated or analysed. Code used for analysis is available from the corresponding author upon reasonable request.

### Funding

The research was supported by a grant from the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at MEH NHS Foundation Trust and UCL Institute of Ophthalmology (grant no. NIHR203322). N.P. is funded by an Artificial Intelligence in Health and Care Award (NIHR AI Award grant no. AI\_AWARD02488). The Artificial Intelligence in Health and Care Award is part of the NHS AI Laboratory, which has made funding available to accelerate the testing and evaluation of artificial intelligence technologies that meet the aims set out in the NHS Long Term Plan. The NHS AI Laboratory is a joint unit of teams from the Department of Health and Social Care and NHS England, driving forward the digital transformation of health and social care (<https://transform.england.nhs.uk/ai-lab/>). N.P. is also funded by Sight Research UK (grant no. TRN004). N.P. and Y.W.C are also funded by Medical Research Foundation and Moorfields Eye Charity (grant no. MRF-JF-EH-23-122) and Fight for Sight (grant no. RESSGA250). T.D. and N.P. are also funded by NIHR i4i THRIVE (grant no. NIHR505133). N.P. was also previously funded by Retina UK as part of the UK IRD Consortium, Moorfields Eye Charity Career Development Award (grant no. R190031A), HDRUK (grant no. MC\_PC\_18036) and by a Translational Innovation grant awarded by the UCL Translational Research Office, which has seed funded this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Copyright

© The Author(s) 2026.

## Publisher's note

Open Exploration maintains a neutral stance on jurisdictional claims in published institutional affiliations and maps. All opinions expressed in this article are the personal views of the author(s) and do not represent the stance of the editorial team or the publisher.

## References

1. Ramsey DJ, Sunness JS, Malviya P, Applegate C, Hager GD, Handa JT. Automated image alignment and segmentation to follow progression of geographic atrophy in age-related macular degeneration. *Retina*. 2014;34:1296–307. [DOI] [PubMed]
2. Hussain MA, Govindaiah A, Souied E, Smith RT, Bhuiyan A. Automated tracking and change detection for age-related macular degeneration progression using retinal fundus imaging. 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR); 2018 Jun 25–29; Kitakyushu, Japan. IEEE; pp. 394–8. [DOI]
3. Balakrishnan G, Zhao A, Sabuncu MR, Gutttag J, Dalca AV. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE Trans Med Imaging*. 2019. [DOI] [PubMed]
4. Sokooti H, de Vos B, Berendsen F, Lelieveldt BPF, Išgum I, Staring M. Nonrigid Image Registration Using Multi-scale 3D Convolutional Neural Networks. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins D, Duchesne S, editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*. MICCAI 2017. Cham: Springer; 2017. [DOI]
5. Zhang J, Wang Y, Dai J, Cavichini M, Bartsch DG, Freeman WR, et al. Two-Step Registration on Multi-Modal Retinal Images via Deep Neural Networks. *IEEE Trans Image Process*. 2022;31:823–38. [DOI] [PubMed] [PMC]
6. Zhang J, Wang Y, Bartsch DG, Freeman WR, Nguyen TQ, An C. Perspective Distortion Correction for Multi-Modal Registration between Ultra-Widefield and Narrow-Angle Retinal Images. *Annu Int Conf IEEE Eng Med Biol Soc*. 2021;2021:4086–91. [DOI] [PubMed] [PMC]
7. Zhang J, An C, Dai J, Amador M, Bartsch DU, Borooah S. Joint Vessel Segmentation and Deformable Registration on Multi-Modal Retinal Images Based on Style Transfer. 2019 IEEE International Conference on Image Processing (ICIP); 2019 Sep 22–25; Taipei, Taiwan. IEEE; 2019. pp. 839–43. [DOI]
8. Benvenuto GA, Colnago M, Casaca W. Unsupervised Deep Learning Network for Deformable Fundus Image Registration. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2022 May 23–27; Singapore, Singapore. IEEE; 2022. pp. 1281–5. [DOI]
9. Tian Y, Hu Y, Ma Y, Hao H, Mou L, Yang J, et al. Multi-scale U-net with Edge Guidance for Multimodal Retinal Image Deformable Registration. *Annu Int Conf IEEE Eng Med Biol Soc*. 2020;2020:1360–3. [DOI] [PubMed]
10. Zou B, He Z, Zhao R, Zhu C, Liao W, Li S. Non-rigid retinal image registration using an unsupervised structure-driven regression network. *Neurocomputing*. 2020;404:14–25. [DOI]
11. Liu J, Li X, Wei Q, Xu J, Ding D. Semi-Supervised Keypoint Detector and Descriptor for Retinal Image Matching. In: Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T, editors. *Computer Vision – ECCV 2022*. ECCV 2022. Lecture Notes in Computer Science. 2022. pp. 593–609. [DOI]
12. Wang AQ, Yu EM, Dalca AV, Sabuncu MR. A robust and interpretable deep learning framework for multi-modal registration via keypoints. *Med Image Anal*. 2023;90:102962. [DOI] [PubMed] [PMC]
13. Chen J, Frey EC, He Y, Segars WP, Li Y, Du Y. TransMorph: Transformer for unsupervised medical image registration. *Med Image Anal*. 2022;82:102615. [DOI] [PubMed] [PMC]

14. Sun J, Shen Z, Wang Y, Bao H, Zhou X. LoFTR: Detector-Free Local Feature Matching with Transformers. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2021 Jun 20–25; Nashville, USA. IEEE; 2021. pp. 8918–27. [DOI]
15. Ho J, Jain A, Abbeel P. Denoising Diffusion Probabilistic Models. Adv Neural Inf Process Syst. 2020.
16. Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-Resolution Image Synthesis with Latent Diffusion Models. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2022 Jun 18–24; New Orleans, USA. IEEE; 2022. pp. 10674–85. [DOI]
17. Sivaraman VB, Imran M, Wei Q, Muralidharan P, Tamplin MR, Grumbach IM, et al. RetinaRegNet: A zero-shot approach for retinal image registration. Comput Biol Med. 2025;186:109645. [DOI] [PubMed] [PMC]
18. Veturi YA, McNamara S, Kinder S, Clark CW, Thakuria U, Bearce B, et al. EyeLiner: A Deep Learning Pipeline for Longitudinal Image Registration Using Fundus Landmarks. Ophthalmol Sci. 2024;5:100664. [DOI] [PubMed] [PMC]
19. Liu J, Li X. Geometrized Transformer for Self-Supervised Homography Estimation. Proceedings of the IEEE International Conference on Computer Vision. 2023 October 1–6; Paris, France. IEEE; 2023. pp. 9522–31. [DOI]
20. Hernandez-Matas C, Zabulis X, Triantafyllou A, Anyfanti P, Douma S, Argyros AA. FIRE: Fundus Image Registration dataset. Model Artif Intell Ophthalmol. 2017;1:16–28. [DOI]
21. Zhou Y, Wagner SK, Chia MA, Zhao A, Woodward-Court P, Xu M, et al. AutoMorph: Automated Retinal Vascular Morphology Quantification Via a Deep Learning Pipeline. Transl Vis Sci Technol. 2022;11:12. [DOI] [PubMed] [PMC]
22. Cheng B, Schwing AG, Kirillov A. Per-Pixel Classification is Not All You Need for Semantic Segmentation. Adv Neural Inf Process Syst. 2021;22:17864–75.
23. Detone D, Malisiewicz T, Rabinovich A. SuperPoint: Self-Supervised Interest Point Detection and Description. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 2018 Jun 18–22; Salt Lake City, USA. IEEE; 2018. pp. 337–49. [DOI]
24. Lindenberger P, Sarlin PE, Pollefeys M. LightGlue: Local Feature Matching at Light Speed. 2023 IEEE/CVF International Conference on Computer Vision (ICCV); 2023 Oct 1–6; Paris, France. IEEE; 2023. pp. 17581–92. [DOI]
25. Lin TY, Dollar P, Girshick R, He K, Hariharan B, Belongie S. Feature Pyramid Networks for Object Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21–26; Honolulu, HI, USA. IEEE; 2017. pp. 936–44. [DOI]