Enabling Cross-Device Retinal Biomarker Alignment for Generalizable Oculomics using Supervised Space Translation

Sparsh Rastogi
srastogi_be22@thapar.edu
Thapar Institute of Engineering and Technology
Patiala, Punjab, India

Sahil Thakur
sahil.thakur@mediwhale.com
Mediwhale
South Korea
Singapore Eye Research Institute
Singapore

Abstract

Oculomics leverages retinal vascular biomarkers to noninvasively assess systemic disease burden, offering a scalable and cost effective alternative to conventional diagnostics. However, the real world clinical utility of oculomics is hindered by variability in retinal features extracted from different imaging devices, which limits reproducibility and comparability. In this study, we propose a regression based framework to standardize vascular biomarker measurements across devices by translating features into a common reference space. Using paired retinal images from two fundus cameras, Topcon Triton and Topcon Maestro2, sourced from the AIREADI dataset (N=764), we define the Triton as the standard camera space and develop both univariate and multivariate regression models to map features extracted from Maestro2 into this space. Vascular features were extracted using the AutoMorph pipeline, and rigorous left right eye validation was employed to ensure subject independence during training and evaluation. A range of supervised learning algorithms, including linear, kernel based, and ensemble models, were evaluated. Our results demonstrate that the proposed translation approach significantly reduces inter-device variability and enables consistent and comparable quantification of retinal biomarkers. This work highlights the importance of addressing device induced variability and presents a scalable solution for harmonizing retinal feature analysis in large scale oculomics studies.

CCS Concepts

• **Applied computing → Consumer health**; • **Computing methodologies → Supervised learning by regression**.

Keywords

Retinal Imaging, Machine Learning, Oculomics, Space Translation, Automorph

ACM Reference Format:

Sparsh Rastogi and Sahil Thakur. 2025. Enabling Cross-Device Retinal Biomarker Alignment for Generalizable Oculomics using Supervised Space Translation. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (KDD UMC'25).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

KDD UMC'25, Toronto, Canada

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Oculomics is an emerging field which utilizes retinal biomarkers to non-invasively assess systemic disease burden, offering a portable and cost-effective alternative to conventional diagnostic techniques. The retina, as a direct extension of the central nervous and vascular systems, provides a unique and non-invasive window into microvascular health. Several studies have demonstrated the strong diagnostic and prognostic relevance of retinal biomarkers such as vessel tortuosity, fractal dimension, and the arteriovenous ratio as predictors of systemic diseases. These vascular features capture microvascular changes that have been consistently associated with a range of chronic conditions, including chronic kidney disease [19], type 2 diabetes [14], cardiovascular events [4], and neurological disorders [7]. For example, lower fractal dimension has been linked to an increased risk of stroke and mortality [8, 9], while elevated vessel tortuosity has been associated with early manifestations of atherosclerosis [15]. In addition, a reduced arteriovenous ratio has been shown to predict the likelihood of stroke and coronary artery disease. These associations are further supported by evidence of shared genetic markers between retinal vascular complexity and coronary artery disease [17]. Collectively, this evidence highlights the value of retinal imaging as a promising non-invasive tool for systemic disease assessment.

While oculomics presents significant potential, initial methods for retinal feature extraction heavily relied on manual analysis, making overall process slow, labor-intensive, and prone to subjective bias, thus limiting its real-world clinical application. Although recent advances in vessel segmentation and feature extraction, including deep learning methods [6, 13, 16], have improved accuracy in controlled settings, their integration into large-scale clinical research and practice has faced persistent limitations. A primary challenge is the narrow scope of many existing tools, which often perform only isolated tasks and exhibit poor generalizability when applied to diverse clinical datasets containing variations in imaging devices and pathologies [2, 5, 10, 11, 20]. Furthermore, widespread platforms like RA, IVAN [1], and VAMPIRE [12] remain semi-automated, requiring substantial manual review that introduces subjective bias and limits scalability [1, 3, 12, 18]. Consequently, despite the established strong associations between retinal vascular features and systemic conditions, its full potential remains under-utilized in large-scale clinical settings. To address these methodological limitations, a fully automated deep learning

based framework AutoMorph [21] was proposed in 2022, integrating image preprocessing, quality control, anatomical segmentation, and comprehensive extraction of morphological features into a unified pipeline. This end-to-end system facilitates precise and scalable analysis of retinal vascular structures, thereby improving the translational potential of oculomics in both research and clinical settings.

While AutoMorph significantly advances automated retinal vascular feature extraction via a standardized pipeline [21], its performance is sensitive to variability across imaging devices, modalities, and acquisition conditions. This inherent variability introduces inconsistencies in extracted features when applied to data from cameras different from its training set, challenging the validation and reliable application in diverse real-world clinical settings. Therefore, addressing this camera-induced variability is crucial for reliable interpretation of retinal biomarkers across diverse clinical settings. This study proposes a framework for aligning features extracted from different imaging systems by establishing one camera system as a "standard space" and framing the problem as a space translation task. Using regression-based mapping, we translate AutoMorph extracted features from other camera systems into this standardized camera space. Analogous to converting measurements from disparate units into a single standard unit, this approach utilizes camera-specific mapping models for each retinal feature to translate its value into the standard space, thereby enabling consistent and comparable retinal biomarker analysis across diverse imaging platforms, irrespective of the source camera. We conducted our analyses using data from two retinal imaging devices, Topcon Triton and Topcon Maestro2, involving 764 individuals sourced from the AIREADI dataset. Standard supervised learning models were trained using vascular parameters extracted from the left eye, while corresponding features from the right eye were reserved for evaluation, thereby ensuring independence between training and testing sets. Since we are focusing on vascular features such as vessel caliber, fractal dimension, and tortuosity which are influenced by systemic factors like blood pressure and auto-regulatory mechanisms, the left–right eye split serves as an effective strategy to prevent data leakage. It enables testing on a cohort with a similar physiological distribution while ensuring that the evaluation remains independent of the training data, thus preserving the integrity of the model validation process. The major contributions of this paper could be summarized as follows:

- We propose a regression-based framework to standardize retinal vascular features, addressing device-specific variability in automated oculomics by projecting measurements into a unified space.
- Using real-world retinal images from two different imaging systems and a rigorous left-right eye validation protocol, we demonstrate the effectiveness of our approach in achieving consistent and comparable biomarker quantification across devices.

2 Methodology

2.1 Problem Definition

Given a paired retinal dataset $\mathcal{D} = \{\mathbf{X}_S, \mathbf{X}_T\}$, where $\mathbf{X}_S \in \mathbb{R}^{N\times d}$ denotes vascular features extracted from the **source camera** and $\mathbf{X}_T \in \mathbb{R}^{N\times d}$ represents the corresponding features obtained from a **standard target camera**, the objective is to learn a supervised regression-based mapping from the source feature space to the target space. Each row in $\mathbf{X}_S$ and $\mathbf{X}_T$ corresponds to a d-dimensional vector of retinal vascular measurements such as central retinal arteriolar equivalent (CRAE), central retinal venular equivalent (CRVE), arteriovenous ratio (AVR), vessel tortuosity, and fractal dimensions.

$$\hat{\mathbf{x}}_T^{(i)} = f_\theta(\mathbf{x}_S^{(i)}) \tag{1}$$

To perform this translation, we learn a parametric function f_θ, with learnable parameters θ, that maps each input feature vector from the source domain to the target domain. The predicted target feature vector for the i-th subject is given by Eq. 1. The model is trained to minimize the discrepancy between the predicted and true target feature vectors. We adopt the root mean squared error (RMSE) loss as the objective function, which penalizes larger errors more heavily and encourages consistency in the learned mapping across the dataset. This loss function is defined in Eq. 2. Minimizing this objective enables the model to perform accurate space translation, effectively mitigating device-induced variability in retinal measurements. Such standardization is essential for ensuring reliable downstream analysis across datasets collected using heterogeneous imaging systems.

$$\mathcal{L}_{\text{RMSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\| f_\theta(\mathbf{x}_S^{(i)}) - \mathbf{x}_T^{(i)} \right\|_2^2} \tag{2}$$

2.2 Supervised Space Translation

To accomplish the translation of retinal vascular features from the source space to the standard target space, a range of regression-based supervised learning algorithms were employed. These included both linear and nonlinear models such as simple linear regression, ridge regression, polynomial regression, support vector regression (SVR), decision trees (DT), random forest (RF), XGBoost, and gradient boosting regression. The goal was to evaluate the effectiveness of each method in learning the mapping between feature spaces derived from different retinal imaging devices. All models were evaluated under two distinct settings: *univariate* and *multivariate*. In the univariate setting, each target feature was predicted independently using only its corresponding feature from the source stream. In contrast, the multivariate setting leveraged the entire set of source features to predict each target feature, in order to examine the effect of auxiliary biomarkers on predictive performance.

2.2.1 Univariate Regression. In the univariate framework, each retinal feature dimension was modeled separately. For a given target feature $x_{T,j}$, only the corresponding source feature $x_{S,j}$ was used as input to the regression model. This approach prioritized interpretability and isolated the device-induced variability at individual feature level, enabling a direct analysis of how each feature behaves across different devices. Despite its simplicity, this method serves as a robust and interpretable baseline for assessing the performance of feature-wise translation.

2.2.2 Multivariate Regression. In the multivariate framework, the complete vector of source features $\mathbf{x}_S \in \mathbb{R}^d$ was utilized to predict each individual target feature $x_{T,j}$. This approach enables the model to capture potential interdependencies and shared variance across retinal biomarkers, which may contribute valuable contextual information for more accurate translation. It leverages the possibility that certain features may act as auxiliary predictors for others, enhancing the model's ability to handle redundancy and noise in real-world data. By incorporating the entire source feature set, this setting evaluates whether leveraging the joint representation space improves regression performance relative to the univariate baseline. For instance, features such as CRAE, CRVE, and AVR—each indicative of vessel calibre are physiologically related and may influence one another, thus a multivariate approach seems more suitable to capture such cross-feature dependencies and physiological coherence.

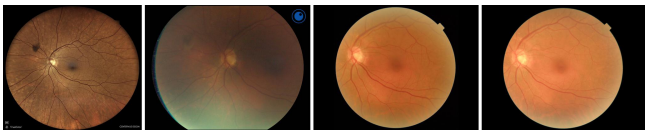

Figure 1: Retinal fundus images of the same patient captured using four different cameras from the AIREADI dataset: (a) iCare Eidon, (b) Optomed Aurora, (c) Topcon Triton, and (d) Topcon Maestro2

2.3 Hyperparameter Optimization

To ensure fair and representative performance comparisons across the diverse set of regression models, we conducted systematic hyperparameter optimization for each algorithm under both univariate and multivariate settings. Given the heterogeneity in model structures and their sensitivity to factors such as initialization, regularization, and kernel configuration, a rigorous tuning protocol was essential. We adopted a combination of grid search and five-fold cross-validation on the training set to identify optimal hyperparameters for each model. For linear models such as ridge and polynomial regression, tuning focused on the regularization coefficient and polynomial degree, respectively. In the case of support vector regression (SVR), key hyperparameters including the penalty parameter C, kernel type, and kernel-specific parameters (e.g., RBF bandwidth) were systematically explored to balance model complexity and generalization.

For tree-based methods like decision trees, random forests, XGBoost, and gradient boosting regression—we varied parameters such as maximum tree depth, number of estimators, and minimum samples per leaf. This optimization process was performed in conjunction with training, with the final configuration for each model selected based on its average cross-validation performance. The chosen hyperparameters were then fixed and applied to the held-out test set for evaluation. This structured and consistent tuning strategy ensured that all models were assessed under conditions best suited to their respective capacities, mitigating biases due to suboptimal settings and enabling a reliable comparison of regression performance across modeling choices.

3 Experimentation

3.1 Dataset

This study uses data from the AIREADI (Artificial Intelligence Ready and Equitably Accessible Disease Insights) initiative, a large-scale, multi-center research program aimed at improving our understanding of ocular and systemic health. The full cohort includes around 4,000 participants enrolled across three academic clinical centers, with extensive clinical, physiological, and imaging data collected for each individual. For this analysis, we used the currently released subset comprising 1,067 participants. Retinal color fundus images were captured using four non-mydriatic imaging devices: Topcon Maestro2, Topcon Triton, Optomed Aurora, and iCare Eidon. These devices differ in imaging protocols, field of view, and sensor characteristics, leading to systematic variations in the extracted retinal features, as illustrated in Fig. 1. A subset of participants had images captured using more than one device, enabling the creation of paired samples across imaging sources. This paired design made it possible to directly learn supervised regression models that translate feature representations between devices, helping to address variability caused by differences in hardware and acquisition settings.

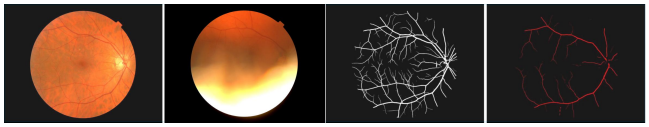

Figure 2: Illustration of outputs across different stages of the AutoMorph pipeline: (a) original CFP, (b) images filtred during quality control, (c) binary vessel segmentation showing the retinal vasculature, and (d) arteriolar–venular segmentation separating arteries and veins.

3.2 Data Preparation

Retinal feature extraction for all participants and imaging devices was performed using the AutoMorph pipeline, an open-source framework for automated analysis of color fundus photographs. AutoMorph comprises a series of modules for image quality grading, anatomical segmentation, and quantitative vascular feature extraction (Fig. 2. In this study, the official implementation[1] was executed on an NVIDIA H100 GPU to enable efficient, large-scale processing. Although images from all four devices (Topcon Maestro2, Topcon Triton, Optomed Aurora, and iCare Eidon) were processed through the pipeline, a considerable number of images from Optomed Aurora and iCare Eidon did not pass the image quality control stage due to poor visual quality. In the case of iCare Eidon, its wide field of view often included peripheral retinal regions with reduced clarity, which led to frequent exclusion by the image quality control module. Overcoming this limitation would require fine-tuning AutoMorph on a large, annotated dataset specifically containing wide-field fundus photographs, which was beyond the scope of this study. Similarly, images acquired using Optomed Aurora, a handheld mobile device, were frequently affected by motion blur and image noise resulting from unstable capture conditions, further

[1] https://github.com/rmaphoh/AutoMorph

limiting the number of high-quality samples available for analysis. To ensure consistency and reliability in vascular feature extraction as well as in the downstream space translation task, the final analysis was restricted to high-quality paired images obtained from the Topcon Maestro2 and Topcon Triton cameras, both of which consistently produced standardized outputs that aligned well with AutoMorph's pretrained configuration.

3.3 Training and Evaluation

Among the available imaging modalities, we used 3D Macula CFPs from the Topcon Maestro2 and Macula 6×6 CFPs from the Topcon Triton, as these were the most consistently acquired across participants and provided reliable views of the retinal vasculature. Retinal feature extraction was performed using AutoMorph, an open-source pipeline for automated analysis of fundus images. From the full set of outputs, a subset of vascular features was selected to retain clinical relevance while minimizing redundancy. Features such as CRAE, CRVE, and AVR (all based on the Knudtson formula) were included as they represent vessel calibre, offering insights into the relative widths of arterioles and venules—key indicators of systemic vascular health. Fractal dimension and vessel density were selected to capture the geometric complexity and spatial distribution of the vascular network, both of which are known to reflect microvascular integrity. Distance tortuosity and tortuosity density were included to quantify vessel curvature and its distribution, traits often associated with hypertension, diabetes, and neurological conditions. To further capture vessel-specific patterns, we included artery and vein-level measurements of fractal dimension and distance tortuosity, enabling more granular assessment of structural variation across vessel types. All features were extracted from Zone C, corresponding to the macular region, which is commonly used for retinal vascular analysis due to its dense vascularity, clinical importance, and reduced sensitivity to peripheral imaging artifacts.

To ensure robustness in downstream analyses, outlier detection was performed on each feature individually using the interquartile range (IQR) method. Participants with extreme values in three or more features, as well as those with missing values, were excluded. Following this filtering procedure, the final dataset comprised 764 participants with complete data for the left eye and 769 participants for the right eye. To perform supervised space translation, we treated the Maestro2 feature space as the source domain and the Triton feature space as the target domain, aiming to learn a regression-based mapping that converts Maestro2-derived features into their corresponding Triton representations. A range of models were evaluated in both univariate and multivariate settings, as described in Section 2. Model training was conducted using left-eye data, with an 80-20 split for training and validation. Final evaluation was carried out on a held-out test set comprising right-eye data from the same participants. This left–right eye protocol was chosen to assess the generalizability of the learned mappings under anatomical symmetry. Since both eyes of an individual tend to exhibit similar vascular characteristics, this approach enables subject-independent evaluation while preserving distributional similarity. Model performance was assessed using the coefficient of determination (R^2) and root mean squared error (RMSE), which together provide complementary insights into both the consistency and accuracy of the predicted mappings across imaging sources.

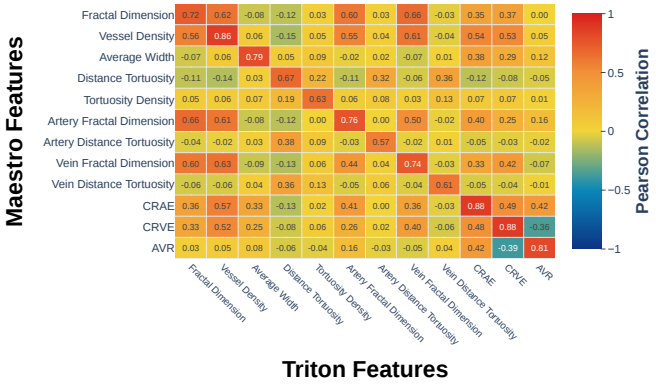

Figure 3: Cross-Dataset Correlation Matrix between Maestro and Triton Features

4 Results and Discussion

4.1 Univariate Analysis

We began our analysis by evaluating feature-wise translation using univariate regression, where each target feature was predicted solely from its corresponding source-domain variable. This approach is grounded in the assumption that several vascular features exhibit device-independent physiological relationships. The cross-domain correlation matrix (Fig. 3) provided preliminary support for this, showing strong positive correlations for features such as CRAE, CRVE, AVR, and vessel density, each exceeding a Pearson correlation of 0.75. These features are known to be stable markers of vascular calibre and structure, which may be less sensitive to imaging protocol differences.

Table 1: R^2 scores from univariate regression models for cross-device retinal feature translation

Feature	Linear	Polynomial	Ridge	SVR	Decision Tree	Random Forest	Gradient Boosting	XGBoost
AVR (Knudtson)	0.45	0.60	0.51	0.64	0.62	0.64	**0.66**	**0.66**
Artery Distance Tortuosity	0.21	0.22	0.21	0.21	0.11	0.09	0.21	**0.23**
Artery Fractal Dimension	**0.17**	0.15	-0.75	-0.79	-0.03	0.13	-0.11	-0.11
Average Width	**0.70**	**0.70**	**0.70**	0.62	0.65	0.64	0.68	0.66
CRAE (Knudtson)	0.49	0.59	0.49	0.72	0.71	0.72	**0.74**	**0.74**
CRVE (Knudtson)	**0.73**	**0.73**	**0.73**	0.59	0.68	0.67	0.71	0.72
Distance Tortuosity	0.36	0.36	0.36	**0.41**	0.29	0.21	0.35	0.37
Fractal Dimension	**-0.24**	-0.82	-2.43	-3.70	-0.97	-0.89	-1.57	-1.48
Tortuosity Density	**0.31**	**0.31**	0.24	0.24	0.29	0.28	**0.31**	0.30
Vein Distance Tortuosity	**0.05**	**0.05**	**0.05**	0.03	-0.09	-0.13	-0.02	0.04
Vein Fractal Dimension	**-0.23**	-0.26	-1.26	-1.10	-0.53	-0.49	-0.60	-0.63
Vessel Density	**0.64**	**0.64**	-0.08	-0.20	0.57	0.57	0.63	0.62

The regression outcomes in Table 1 validate this assumptions across several retinal features. High R^2 values were achieved for CRVE (0.74), CRAE (0.74), and average width (0.70), across both linear and ensemble-based models, confirming their strong transferability between Maestro2 and Triton domains. Ensemble methods, especially XGBoost and Gradient Boosting, often outperformed their linear counterparts, capturing mild non-linearities while maintaining generalization. However, several localized features, particularly artery and vein fractal dimensions and tortuosity-based metrics, exhibited low or negative R^2 scores, indicating poor predictive performance. These results suggest that such complex features are

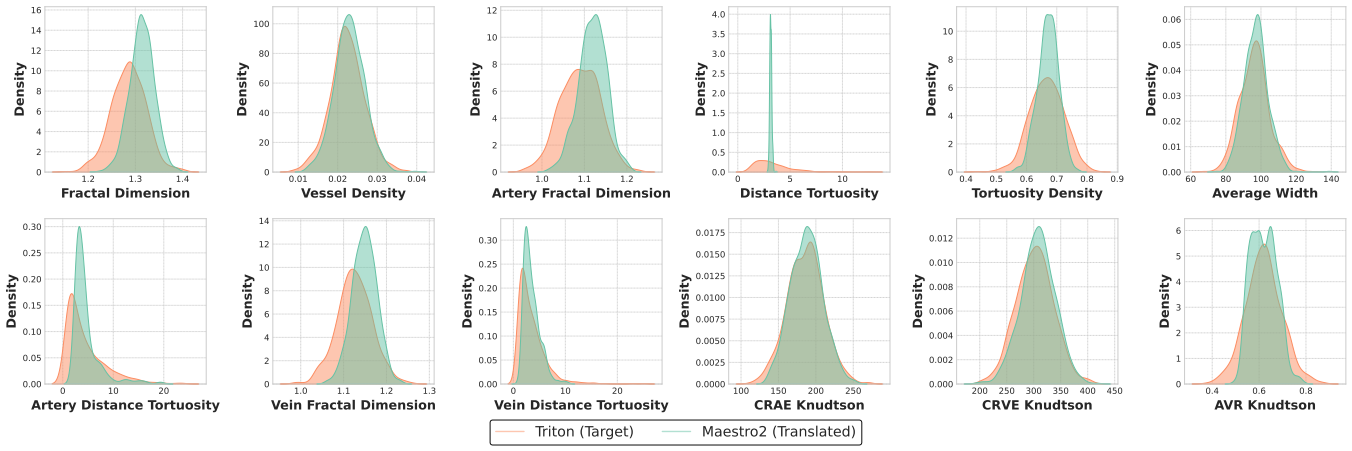

Figure 4: Comparison of translated feature distributions from Maestro2 with ground truth Triton distributions

highly sensitive to segmentation inconsistencies and device-specific imaging characteristics, limiting their learnability in a univariate framework. The overall pattern indicates that simple one-to-one mappings are sufficient only for structurally stable, globally descriptive features.

4.2 Multivariate Analysis

To address the limitations of univariate models, we extended the analysis to multivariate regression, where each target feature was predicted using the full set of source-domain variables. This approach is particularly suited for capturing inter-feature dependencies and compensating for potential noise in any single predictor. The cross-domain correlation matrix revealed several off-diagonal correlations, suggesting that multiple auxiliary features may contribute to more accurate translation beyond direct one-to-one mappings. As shown in Table 2, multivariate models achieved improved or comparable performance across nearly all features. Notably, CRAE and CRVE reached R^2 values of 0.81 and 0.87 respectively using Ridge regression, surpassing their univariate baselines. Surprisingly, even basic linear models performed competitively when sufficient predictive redundancy was present in the source feature space. Ensemble models, particularly XGBoost, consistently performed well, achieving top scores for AVR (0.66), CRAE (0.79), and average width (0.66), benefiting from their ability to capture complex interactions. The XGBoost feature importance radar plot (Fig. 5) reinforces this view, confirming that while the corresponding source feature often plays a major role in prediction, auxiliary inputs contribute meaningfully for moderately stable targets.

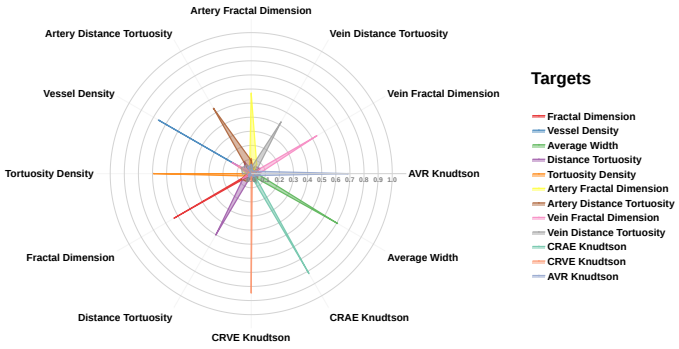

Figure 5: Normalized feature importance for XGBoost models across targets

Poorly performing features in the univariate setting showed moderate gains with multivariate inputs. For example, artery distance tortuosity showed modest improvement, reaching an R^2 of 0.23, but artery and vein fractal dimensions remained difficult to predict even with multivariate inputs. This persistent underperformance highlights a combination of factors. On the data side, these features are highly sensitive to image quality, segmentation fidelity, and vessel-type labeling, all of which are prone to variability across imaging devices. Such noise reduces the signal consistency needed for effective learning. While multivariate inputs help leverage correlated features, they cannot fully overcome cases where the primary signal is unstable or poorly defined. Ultimately, no model structure was sufficient to resolve the challenges posed by high-variance, morphology-dependent features.

Table 2: R^2 scores from multivariate regression models for cross-device retinal feature translation

Feature	Linear	Polynomial	Ridge	SVR	Decision Tree	Random Forest	Gradient Boosting	XGBoost
AVR (Knudtson)	0.72	-5.27	0.72	0.65	0.58	0.69	0.67	**0.71**
Artery Distance Tortuosity	0.42	-10.14	0.42	-0.06	0.32	0.40	0.43	**0.44**
Artery Fractal Dimension	**0.58**	-3.44	0.48	0.13	0.50	0.58	0.57	0.57
Average Width	**0.59**	-2.03	0.60	0.15	0.51	0.57	0.55	0.58
CRAE (Knudtson)	**0.81**	-0.81	0.81	0.36	0.71	0.76	0.78	0.79
CRVE (Knudtson)	**0.86**	-2.72	0.87	0.42	0.81	0.83	0.80	0.82
Distance Tortuosity	**0.37**	-7.38	0.39	0.06	0.19	0.32	0.38	0.35
Fractal Dimension	**0.51**	-4.86	0.45	0.00	0.41	0.48	0.49	0.51
Tortuosity Density	**0.48**	-7.79	0.38	0.00	0.42	0.42	0.42	0.42
Vein Distance Tortuosity	0.24	-7.70	**0.25**	-0.06	0.23	0.26	0.27	0.25
Vein Fractal Dimension	**0.52**	-6.60	0.46	0.11	0.44	0.49	0.49	0.51
Vessel Density	**0.80**	-5.12	0.59	0.00	0.68	0.77	0.78	0.79

This trend is further reflected in the regression plots (Fig. 6), where global vascular features such as CRVE, CRAE, and AVR show strong alignment with the identity line, indicating consistent relationships and successful cross-device translation. In contrast, features like vein fractal dimension display scattered predictions with no clear structure, reflecting a combination of model uncertainty

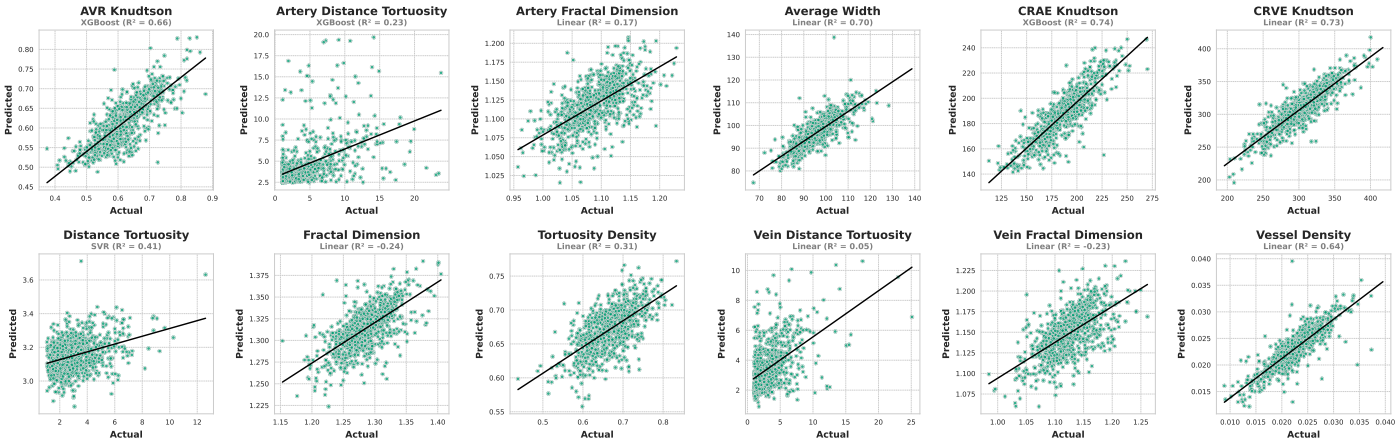

Figure 6: Regression plots illustrating the model fit for each of the 12 retinal features using the best-performing multivariate regression model for the respective feature

and weak underlying patterns. The same distinction is evident in the feature distribution plots (Fig. 4), where well-performing features yield overlapping predicted and true distributions, while poorly performing ones exhibit skew, dispersion, or systematic shifts. In particular, distributions with long tails or asymmetric profiles often indicate high inter-subject variability or segmentation artifacts that vary significantly across devices. Such irregular distributions reduce the effective signal-to-noise ratio and lead to unstable regression fits, as the model struggles to learn consistent mappings. These discrepancies suggest that the observed performance gaps arise from both high intrinsic data variance and limited model fit, especially in the presence of segmentation noise or inconsistent feature extraction.

Taken together, these results demonstrate that global vascular features such as CRAE, CRVE, AVR, vessel density, and average width are reliably transferable across imaging devices. These metrics exhibit high cross-domain correlation, robust regression performance, and consistent distributional alignment, reflecting both physiological stability and technical reproducibility. For such features, the learned regression models can be distilled into transformation equations or coefficient sets that may be embedded into preprocessing pipelines. This enables real-time harmonization of retinal features across heterogeneous imaging systems, reducing device-induced variability and supporting standardized analysis in multi-center clinical studies and large-scale screening programs. In contrast, features derived from localized vessel morphology particularly tortuosity and fractal dimension remain challenging to translate due to their sensitivity to noise, segmentation variability, and inconsistent vessel labeling. Even with multivariate modeling, these features show limited generalizability. Their successful standardization may require upstream steps such as denoising, refined anatomical segmentation, or vessel-type correction to stabilize feature values before translation. Nonetheless, the proposed supervised space translation framework offers a flexible, lightweight, and interpretable solution for harmonizing retinal biomarkers, producing reusable mappings that support scalable and device-agnostic analysis across clinical workflows.

5 Future Directions

While this study establishes a regression-based framework for translating retinal vascular features across devices, future work could explore advanced learning strategies such as autoencoders, manifold alignment, or contrastive learning to derive robust, device-invariant embeddings. These may improve generalizability, particularly for morphology-sensitive features. Expanding the feature set to include optic disc, macular, or texture-based metrics could further enhance prediction and provide a more complete assessment of retinal health. Evaluating model performance across diverse cohorts stratified by age, ethnicity, and disease status is also important for ensuring fairness and clinical utility. Additionally, incorporating longitudinal data may support temporal modeling of disease progression and response to treatment. Together, these directions can significantly strengthen the scalability and clinical relevance of retinal biomarker translation for population level oculomics.

6 Conclusion

This study proposed a supervised regression-based framework for translating retinal vascular features across imaging devices using paired Maestro2 and Triton data. Global vascular metrics such as CRAE, CRVE, AVR, and vessel density were found to be highly transferable, showing strong cross-domain correlations and predictive performance. Multivariate models, especially ensemble methods like XGBoost, outperformed univariate baselines by leveraging auxiliary feature dependencies and mitigating device-induced variability. Feature importance analysis confirmed that the corresponding source feature remained the strongest predictor, while related biomarkers contributed to improved accuracy for moderately stable targets. In contrast, morphology-specific features such as tortuosity and fractal dimensions remained difficult to translate due to high noise and segmentation inconsistency. For stable features, the learned mappings can be embedded as standardized transformations within preprocessing pipelines, enabling scalable, device-agnostic retinal biomarker translation for oculomics applications across diverse clinical and research settings.

References

[1] Ciprian Danielescu, Marius Gabriel Dabija, Alin Horatiu Nedelcu, Vasile Valeriu Lupu, Ancuta Lupu, Ileana Ioniuc, Georgiana-Emmanuela Gîlcă-Blanariu, Vlad-Constantin Donica, Maria-Luciana Anton, and Ovidiu Musat. 2023. Automated retinal vessel analysis based on fundus photographs as a predictor for non-ophthalmic diseases-evolution and perspectives. *J. Pers. Med.* 14, 1 (Dec. 2023).

[2] M M Fraz, R A Welikala, A R Rudnicka, C G Owen, D P Strachan, and S A Barman. 2015. QUARTZ: Quantitative Analysis of Retinal Vessel Topology and size – An automated system for quantification of retinal vessels morphology. *Expert Syst. Appl.* 42, 20 (Nov. 2015), 7221–7234.

[3] M M Fraz, R A Welikala, A R Rudnicka, C G Owen, D P Strachan, and S A Barman. 2015. QUARTZ: Quantitative Analysis of Retinal Vessel Topology and size – An automated system for quantification of retinal vessels morphology. *Expert Syst. Appl.* 42, 20 (Nov. 2015), 7221–7234.

[4] Yuechuan Fu, Mayinuer Yusufu, Yueye Wang, Mingguang He, Danli Shi, and Ruobing Wang. 2023. Association of retinal microvascular density and complexity with incident coronary heart disease. *Atherosclerosis* 380 (Sept. 2023), 117196.

[5] Joseph Futoma, Morgan Simons, Trishan Panch, Finale Doshi-Velez, and Leo Anthony Celi. 2020. The myth of generalisability in clinical research and machine learning in health care. *Lancet Digit. Health* 2, 9 (Sept. 2020), e489–e492.

[6] Fan Huang, Behdad Dashtbozorg, and Bart M Ter Haar Romeny. 2018. Artery/vein classification using reflection features in retina fundus images. *Mach. Vis. Appl.* 29, 1 (Jan. 2018), 23–34.

[7] Sophie Lemmens, Astrid Devulder, Karel Van Keer, Johan Bierkens, Patrick De Boever, and Ingeborg Stalmans. 2020. Systematic review on fractal dimension of the retinal vasculature in neurodegeneration and stroke: Assessment of a potential biomarker. *Front. Neurosci.* 14 (Jan. 2020), 16.

[8] Gerald Liew, Bamini Gopinath, Andrew J White, George Burlutsky, Tien Yin Wong, and Paul Mitchell. 2021. Retinal vasculature fractal and stroke mortality. *Stroke* 52, 4 (April 2021), 1276–1282.

[9] Gerald Liew, Bamini Gopinath, Andrew J White, George Burlutsky, Tien Yin Wong, and Paul Mitchell. 2021. Retinal vasculature fractal and stroke mortality. *Stroke* 52, 4 (April 2021), 1276–1282.

[10] Gustav Mårtensson, Daniel Ferreira, Tobias Granberg, Lena Cavallin, Ketil Oppedal, Alessandro Padovani, Irena Rektorova, Laura Bonanni, Matteo Pardini, Milica G Kramberger, John-Paul Taylor, Jakub Hort, Jón Snædal, Jaime Kulisevsky, Frederic Blanc, Angelo Antonini, Patrizia Mecocci, Bruno Vellas, Magda Tsolaki, Iwona Kłoszewska, Hilkka Soininen, Simon Lovestone, Andrew Simmons, Dag Aarsland, and Eric Westman. 2020. The reliability of a deep learning model in clinical out-of-distribution MRI data: A multicohort study. *Med. Image Anal.* 66, 101714 (Dec. 2020), 101714.

[11] A Perez-Rovira, T MacGillivray, E Trucco, K S Chin, K Zutis, C Lupascu, D Tegolo, A Giachetti, P J Wilson, A Doney, and B Dhillon. 2011. VAMPIRE: Vessel assessment and measurement platform for images of the REtina. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Boston, MA). IEEE.

[12] A Perez-Rovira, T MacGillivray, E Trucco, K S Chin, K Zutis, C Lupascu, D Tegolo, A Giachetti, P J Wilson, A Doney, and B Dhillon. 2011. VAMPIRE: Vessel assessment and measurement platform for images of the REtina. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Boston, MA). IEEE.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Lecture Notes in Computer Science*. Springer International Publishing, Cham, 234–241.

[14] Emmanuel Sandoval-Garcia, Stela McLachlan, Anna H Price, Thomas J MacGillivray, Mark W J Strachan, James F Wilson, and Jackie F Price. 2021. Retinal arteriolar tortuosity and fractal dimension are associated with long-term cardiovascular outcomes in people with type 2 diabetes. *Diabetologia* 64, 10 (Oct. 2021), 2215–2227.

[15] Muhammad Bayu Sasongko, Tien Yin Wong, Kim C Donaghue, Ning Cheung, Alicia J Jenkins, Paul Benitez-Aguirre, and Jie Jin Wang. 2012. Retinal arteriolar tortuosity is associated with retinopathy and early kidney dysfunction in type 1 diabetes. *Am. J. Ophthalmol.* 153, 1 (Jan. 2012), 176–83.e1.

[16] Chetan L Srinidhi, Aparna P, and Jeny Rajan. 2019. Automated method for retinal artery/vein separation via graph search metaheuristic approach. *IEEE Trans. Image Process.* 28, 6 (Jan. 2019), 2705–2718.

[17] Ana Villaplana-Velasco, Marie Pigeyre, Justin Engelmann, Konrad Rawlik, Oriol Canela-Xandri, Claire Tochel, Frida Lona-Durazo, Muthu Rama Krishnan Mookiah, Alex Doney, Esteban J Parra, Emanuele Trucco, Tom MacGillivray, Kristiina Rannikmae, Albert Tenesa, Erola Pairo-Castineira, and Miguel O Bernabeu. 2023. Fine-mapping of retinal vascular complexity loci identifies Notch regulation as a shared mechanism with myocardial infarction outcomes. *Commun. Biol.* 6, 1 (May 2023), 523.

[18] Tien Yin Wong, F M Amirul Islam, Ronald Klein, Barbara E K Klein, Mary Frances Cotch, Cecilia Castro, A Richey Sharrett, and Eyal Shahar. 2006. Retinal vascular caliber, cardiovascular risk factors, and inflammation: the multi-ethnic study of atherosclerosis (MESA). *Invest. Ophthalmol. Vis. Sci.* 47, 6 (June 2006), 2341–2350.

[19] Linbin Wu, Xia Gong, Wei Wang, Lei Zhang, Jiachen Zhou, Xi Ming, Meng Yuan, Wenyong Huang, and Lanhua Wang. 2022. Association of retinal fractal dimension and vessel tortuosity with impaired renal function among healthy Chinese adults. *Front. Med. (Lausanne)* 9 (Sept. 2022), 925756.

[20] Yukun Zhou, Zailiang Chen, Hailan Shen, Xianxian Zheng, Rongchang Zhao, and Xuanchu Duan. 2021. A refined equilibrium generative adversarial network for retinal vessel segmentation. *Neurocomputing* 437 (May 2021), 118–130.

[21] Yukun Zhou, Siegfried K Wagner, Mark A Chia, An Zhao, Peter Woodward-Court, Moucheng Xu, Robbert Struyven, Daniel C Alexander, and Pearse A Keane. 2022. AutoMorph: Automated retinal vascular morphology quantification via a deep learning pipeline. *Transl. Vis. Sci. Technol.* 11, 7 (July 2022), 12.

A CodeReproducibility

To facilitate the reproduction of our results, we have made our code and trained models publicly available. The repository includes a Jupyter Notebook that offers a comprehensive, step-by-step guide from data acquisition to the final results. You can access the repository here: https://github.com/SparshRastogi/KDD-UMC-25

B Data Reproducibility

The data used in this study is sourced from the open-source, publicly available AIREADI dataset. Access to this dataset can be requested via the following link: https://aireadi.org/